THE UNIVERSITY OF MANITOBA

Post-Data Pivotal Inference in Balanced Random Effects Models

by

Dennis John Murphy

A Thesis

submitted to the Faculty of Graduate Studies

in partial fulfillment of the requirements for the Degree of

Doctor of Philosophy

Department of Statistics

Winnipeg, Manitoba

0-612-51659-8

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

*****

COPYRIGHT PERMISSION PAGE

Post-Data Pivotal Inference in Balanced Random Effects Models

BY

Dennis John Murphy

A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of Manitoba in partial fulfillment of the requirements of the degree

of

Doctor of Philosophy

Abstract

Post-data pivotal (PDP) inference is a conditional sampling-theory oriented mode of statistical inference designed for classes of problems with ordered or truncated parameter spaces. For example, in balanced random effects models with nonnegative variance components, the orderings of the expected mean squares constrain the joint parameter space. Under the usual model, the PDP approach replaces the usual pivotal distributions by conditional analogues, where the conditioning event is a subset of the usual pivotal space, incorporating information from both the data and the parameter constraints. Point and interval estimators are obtained by applying the usual procedures with respect to a family of PDP-modified distributions. As a result, 'good' frequentist procedures are modified to achieve good conditional performance. In contrast, classical inference procedures that fail to take account of existing parameter constraints are capable of poor post-data performance, e. g., a negative variance estimate.

PDP methodology is largely developed through the balanced one-way random effects model, although the three-stage nested model is also examined. PDP distribution theory is shown to be connected to the Bayesian posterior distributions of Box and Tiao (1973) and the fiducial distributions of Wild (1981). In this sense, the PDP approach is unified with other post-data schools of inference. In the case of the one-way model, PDP procedures have guaranteed conditional acceptability. Point estimators, and intervals with so-called 'fixed length', have frequentist properties at least as good as the usual estimators, along with a generalized Bayes property. These results extend the work of Brewster and Zidek (1974) in the corresponding fixed effects problem.

Contents

List of Figures

Acknowledgements

In the course of my education, I've been blessed with the acquaintance and friendship of many wonderful, talented people. I would like to take this opportunity to acknowledge some of these individuals in print.

First of all, I would like to thank my examining committee, whose comments and encouragement have greatly enhanced this work: Dr. Mirek Pawlak and Dr. Bruce Johnston of the University of Manitoba, and Dr. George Casella of Cornell University. I also thank Dr. Craig Platt of the University of Manitoba for chairing my thesis defense.

I would also like to thank my first advisor, Dr. Lai Chan of the City University of Hong Kong, and Dr. Frederick Spiring of Pollard Banknote, who served on my advisory committee. Their friendship and advice over the years is deeply appreciated.

Several other individuals are acknowledged for their contributions to this work: Dr. James Fu, who assisted in a couple of proofs; Dr. Michael Doob, who patiently answered many silly (and some not so silly) questions concerning LaTeX; and Dr. Russell Wolfinger of SAS Institute, who helped with a SAS PROC MIXED problem during my candidacy exam.

The support of the faculty, students and secretarial staff in the Department of

Statistics has been excellent, and I thank them all, past and present, for enhancing my professional and personal life.

I would also like to acknowledge several fine gentlemen from my years at the University of Wisconsin and Iowa State University who have had a deep impact on my life, professionally and socially: Dr. James Wendelberger, Dr. Spencer Graves, Dr. Stephen Rathbun, Dr. Conrad Fung, Dr. Arthur Fries, Dr. Steve Verrill, Dr. Robert Wardrop and Gregory Hudak. I am also deeply indebted to Dr. Norman Draper for the best advice I have ever received.

Special thanks are due to Dr. James Evans of the Forest Products Laboratory, who taught me the joy of discovery through statistics, and to my advisor at Iowa State, Dr. W. Robert Stephenson, who taught me the fundamentals of research, with emphasis on fun.

I cannot adequately thank my thesis advisor, Dr. John Brewster, for what he has meant to me over the past few years. I thank him for this thesis topic, which has been the most fascinating toy I've ever had; for his unfailing optimism and patience; for always being there when needed; and most of all, for being a true mentor. I couldn't ask for a better advisor, and I hope that I have been a worthy student. No one, apart from my parents, has ever taught me more.

My family has been a constant source of love and support. In particular, I would like to thank my parents, Elsie and the late Timothy Murphy, my sister Jean Clark and my niece Jessica Clark for being the backbone of my existence. I'd also like to thank my longest and dearest friends—Mike Chybowski, Jim Kastenholz and Joe Ziman—for putting up with me for all these years.

Finally, I would like to thank my wife, Dr. Lena Horne, for her steadfast love and emotional/financial support. Her forbearance during this process has been remarkable, and I am grateful to her for the freedom to pursue this dream.

Chapter 1

Introduction

1.1 Overview

This thesis introduces a conditional, sampling-theory based approach to statistical inference called the *post-data pivotal* (or PDP) method. PDP inference is intended for use in location-scale problems involving one or more constraints on the model parameters. It is designed to synthesize information from both the data and the constrained parameter(s) in forming a conditioning event on a pivotal space. Applying this event to the usual pivotals of the model modifies the distribution theory, which in turn modifies the 'usual' frequentist procedures. The result is that PDP-modified estimators always exhibit reasonable conditional behavior, and in most circumstances, possess good frequentist properties as well. In contrast, classical frequentist procedures that fail to account for constraints on the parameter space sometimes produce undesirable inferences or estimates. In this thesis, we consider applications of PDP methodology to simple balanced random effects models, where the purpose is to make inferences about (functions of) the variance components.

PDP inference touches on several foundational topics in statistical inference. It bears directly on both conditional and unconditional sampling theory inference, and is closely tied to Bayesian inference based on Jeffreys noninformative priors. In some problems, PDP distributions may also be connected to fiducial distributions; the one-way random model of Chapter 2 is one such case. Hence, PDP inference is connected to other post-data schools of statistical inference in this restricted class of problems through its distribution theory.

PDP inference is also related to some degree to other major schools of statistical thought. The frequentist properties of PDP procedures are established through (frequentist) decision theory, and likelihood concepts can be exploited in PDP point estimation. There is also some resemblance in principle to the generalized p-value method of Weerahandi (1994), and more tenuously, to structural inference (Fraser, 1968, 1979, 1983) and pivotal inference (Barnard, 1995). Although the PDP approach is similar in spirit to Weerahandi's work, its implementation is radically different.

The remainder of this chapter examines several topics that are relevant to PDP inference. The ideas of initial and final precision, along with those of frequentist and conditional acceptability, provide a reference point for examining the behavior of PDP methods; they are the subject of section 1.3. The Stein-Brown-Brewster-Zidek (SBBZ) process for improving upon the best $\mathscr{G}$-equivariant estimator of the normal variance is discussed in section 1.4. It plays an important role in establishing a template for frequentist risk assessment of certain types of PDP estimators in Chapter 2.

1.2 Basics of PDP inference

As noted above, post-data pivotal inference is applicable in classes of problems involving pivotal distributions and constrained parameter spaces. To illustrate how it works, we consider two relatively simple statistical problems that have attracted some attention in the literature. The first example is the fuel of the Mayo-Seidenfeld debate (Seidenfeld, 1979, 1981; Mayo, 1981), reexamined by Casella (1988) and Wolpert (1988). The second example concerns inference about a normal mean bounded on one side, considered by Box and Tiao (1973, ch. 2), among others.

Example 1.1. Suppose $X \sim \text{Uniform}(0, \theta)$, where θ is the unknown parameter of interest. Suppose further that it is known *a priori* that $\theta \leq 15$. The statistical problem is to obtain a log-shortest confidence interval for θ based on a random sample of size $n = 1$.

Prior to data collection, the frequentist statistician determines that a log-shortest $1 - \alpha$ confidence procedure for θ is $(X, X/\alpha)$, where $1 - \alpha$ denotes the confidence coefficient. A random experiment is then conducted; suppose $x = 10$ is the realization of X. From this datum, the frequentist concludes that (10, 200) is a log-shortest 95% confidence interval for θ; however, θ *must* lie in this interval, given prior knowledge that $\theta \leq 15$, so one should attach 100% conditional confidence to it!

On the other hand, a Bayesian statistician can accommodate the parameter constraint $\theta \leq \theta^*$ in the inference process, where in this problem, $\theta^* = 15$. Following Wolpert (1988), place a Bernardo reference prior on θ and use Bayes' theorem to obtain the posterior distribution of $\theta|x$, where $\theta \in (0, \theta^*]$. A log-shortest 95% credible set for θ based on its posterior distribution has the form

$$\left(x, \frac{x}{1 - (1 - \alpha)(1 - x/\theta^*)}\right). \tag{1.1}$$

For $x = 10$ and $\theta^* = 15$, the resulting interval is (10, 14.634), which lies completely within the constrained parameter space of θ.

The PDP approach to this problem is as follows. Since $X \sim \text{Uniform}(0, \theta)$, let $Y = X/\theta$. Then $Y \sim \text{Uniform}(0, 1)$, so Y is a pivotal quantity. Given that $\theta \leq \theta^*$, one can assert pre-data that $Y \geq Y^*$, where $Y^* = X/\theta^*$. The left hand side of the inequality is a pivotal quantity, while the right hand side is a random lower bound on Y; in other words, the event $Y \geq Y^*$ is a random reference set for inference about θ from a pre-data perspective. Given a realization x of X, a specific reference set $Y \geq y$ becomes manifest, where $y = x/\theta^*$. The distribution of $Y|Y \geq y$ is then Uniform$(y, 1)$. A log-shortest $1-\alpha$ PDP-CI (PDP confidence interval) estimate of θ is constructed from the conditional probability statement

$$\begin{aligned}
1-\alpha &= P(c \leq Y \leq 1|Y \geq y) \\
&= P\Big(c \leq \frac{X}{\theta} \leq 1|Y \geq y\Big) \\
&= P\Big(1 \leq \frac{\theta}{X} \leq \frac{1}{c}|Y \geq y\Big) \\
&= P(x \leq \theta \leq xc^{-1}|Y \geq y),
\end{aligned}$$

obtained from the distribution of $Y|Y \geq y$. By construction, c is the lower α quantile of the conditional distribution, which is expressible as

$$c = 1-(1-\alpha)(1-y) = 1-(1-\alpha)(1-x/\theta^*).$$

Therefore, a log-shortest $1-\alpha$ PDP-CI estimate of θ is of the form

$$\Big(x, \frac{x}{1-(1-\alpha)(1-x/\theta^*)}\Big),$$

which is precisely Wolpert's interval (1.1). Therefore, in this example, the log-shortest PDP-CI estimate coincides with the Bayesian credible interval.

Discussion. This problem exposes how widely pre-data and post-data inference rules can diverge in the presence of a parameter constraint. Frequentist rules are functions of the data alone, so it is not unusual for a procedure with 'good' frequentist properties to occasionally exhibit poor conditional performance when parameter constraints exist and are unaccounted for in the procedure. The extent to which a frequentist procedure retains its properties in post-data assessment depends on its ability to *anticipate* different partitions of the sample space $\mathcal{X}$ (Brown, 1990) and to behave sensibly on each partition. 'Nonsensical' post-data behavior occurs, for example, when a point estimate of a parameter θ assumes a value outside its parameter space Θ, or a confidence interval estimate covers values outside Θ, or the post-data 'confidence' that one would attach to an interval estimate does not agree with the pre-data confidence report (as in this example).

The Mayo-Seidenfeld controversy dealt with frequentist vs. post-data inference *per se*, i. e., whether one should be concerned with initial precision (pre-data) or final precision (post-data); see section 1.3.1 for further discussion. Mayo (1981) correctly noted that the only claimed properties of a frequentist procedure are those achieved over repeated sampling; e. g., the level of a confidence procedure is a statement of the asymptotic relative frequency of covering the true parameter value in a sequence of hypothetical random experiments from the same population, *prior to actual observation of data.* Seidenfeld, on the other hand, advocated the standard Bayesian position that final precision is what matters (e. g., Savage, 1962), i. e., the inference produced from the data actually observed:

> ... it seems reasonable to say, *before knowing the data*, that the [frequentist] interval will cover the true value of the unknown parameter with probability $(1-\alpha)$; however, *having seen the value* x, it may be unreasonable to maintain the probability statement or use it to express a degree of confidence in the interval generated by the [frequentist] rule. (Seidenfeld, 1979, pp. 56-57).

Until the mid-1970s, little work had been done to quantify both the pre-data and post-data 'accuracy' of a frequentist procedure. Kiefer's theory of conditional confidence was a noteworthy contribution, if only because it raised awareness that both frequentist and conditional properties of an inference (procedure) matter (Kiefer 1975, 1976, 1977). The original ideas have expanded into a theory of *frequentist post-data* *inference*, in which the goal is to produce statistical procedures that can be recommended from both frequentist and conditional viewpoints (Casella, 1988; Goutis and Casella, 1995). This subject is addressed in greater detail in section 1.3.4.

How does all of this pertain to the PDP approach? Casella's (1988) position is that a statistical procedure should be both *frequentist acceptable* (i. e., possess good pre-data properties) and *conditionally acceptable* (i. e., have good conditional properties). Casella's objection to Wolpert's interval is that it has unconditional coverage probability 0 at $\theta = 15$, so the interval is not frequentist acceptable. From our perspective, however, the frequentist acceptability criterion is overly strict, and in some sense unfair, when the parameter space of an inference problem is constrained. Post-data inference techniques, such as Bayesian credible intervals, PDP intervals, or even fiducial intervals, are capable of adapting in the presence of a parameter constraint, since the probability measure can be (re)defined over a constrained region of the parameter space or pivotal space. Conversely, most frequentist procedures are not designed to perform similar *post-data* adaptations.

Casella (1992) suggests that conditional acceptability of a frequentist procedure has primacy over frequentist acceptability when a choice must be made between them. In our view, this is a reasonable compromise in the types of problems under consideration. The conflict between initial and final precision of an inference (procedure) is relevant to this topic, and is discussed more fully in section 1.3.

Example 1.2. Suppose $X \sim N(\mu, 1)$, where $\mu \geq 0$ is the parameter of interest. The statistical problem is to find a 95% CI for μ based on a random sample of size $n = 1$. A frequentist cites the 'usual' shortest interval procedure

$$X \pm z_{1-\alpha/2} = X \pm 1.96.$$

Suppose a random experiment produces the outcome $x = -2$. The realized frequentist interval would then be $(-3.96, -0.04)$, which lies entirely outside the constrained parameter space.

The PDP approach operates as follows. Since $X \sim N(\mu, 1)$, the quantity $Z = X - \mu$ has a standard normal distribution, so Z is a pivotal. Define $Z^* = X - \mu^*$, where μ^* is the lower boundary of the constrained parameter space. Therefore, $Z \leq Z^*$ pre-data, where Z^* is a random upper bound on Z. Once $X = x$ is observed, $z^* = x - \mu^*$ is a realization of Z^*; the density of the post-data pivotal $Z|Z \leq z^*$ is

$$g^*(z|z^*) = \frac{\phi(z)}{\Phi(z^*)}, \qquad -\infty < z \leq z^*; \tag{1.2}$$

i. e., $Z|Z \leq z^*$ follows a truncated normal distribution.

Point and interval estimates of μ can be obtained from this PDP distribution. For example, noting that

$$E(Z|Z \leq z^*) = \frac{-\phi(z^*)}{\Phi(z^*)},$$

one can deduce that the 'PDP-unbiased' estimate of μ is given by

$$x + \frac{\phi(x - \mu^*)}{\Phi(x - \mu^*)};$$

the corresponding estimator $X + \phi(X - \mu^*)/\Phi(X - \mu^*)$ is the generalized Bayes estimator of μ under quadratic loss with an improper uniform prior on the parameter space of μ. For this datum, the estimate is equal to 0.3732.

Pivotal confidence intervals are obtained in the usual way, but based on the truncated normal density (1.2). For the realized datum, the shortest 95% PDP-CI of μ is $(0, 1.0511)$, which falls within the constrained parameter space. Some may argue that one should change the 'confidence coefficient' as a function of the data in problems such as this. The key point, however, is that the interval should be based on the PDP distribution rather than the unconditional distribution.

The lesson from the above examples is that the PDP approach, as well as other post-data modes of inference, can produce sensible inferences in constrained parameter problems. In contrast, frequentist procedures may yield specific inferences that lie (partially) outside the intended parameter space if the parameter constraint is sufficiently tight.

1.3 Topics in post-data inference

What separates a pre-data inference procedure from a post-data inference statement is that the former is concerned with average performance over the collection of all potential random experiments of some fixed size n, while the latter is concerned with the goodness of an inference in the experiment actually performed. Pre-data inference is the sole province of the frequentist school of inference; nearly all others perform post-data inference, including Bayesian, likelihood, fiducial-based[1] and conditional schools.

Several books investigate the features and differences among these various modes of inference (e. g., Barnett, 1982; Welsh, 1996). Our concern is twofold: the difference between pre-data and post-data measures of 'accuracy', and the distinction between frequentist and conditional inference, both of which are based on sampling theory.

The discussion is relevant to the present work because PDP inference is a specialized form of conditional sampling-theory inference. Its implementation, however, is quite different from traditional conditional inference, particularly in the choice of conditioning event. The material presented in this section provides a foundation for the results in subsequent chapters.

---

[1] By fiducial-based, we mean Fisher's fiducial inference and its descendants, pivotal and structural inference.

1.3.1 Initial and final precision

The notions of initial and final precision[2], terms coined by Savage (1962), are intended to distinguish the meaning of pre-data and post-data 'accuracy' in estimation, respectively. Initial precision refers to quantitative assessment of a statistical procedure in terms of its average performance among all prospective random samples of some fixed size n from a stated family of distributions (Barnett, 1982). Consequently, initial precision refers to accuracy from a frequentist perspective. Mechanisms for assessing initial precision include risk performance of point estimators, average length or coverage probability of interval estimators, power of statistical tests, and goodness of fit criteria for statistical models. Use of decision theory principles to determine 'best' estimators or tests within a particular class of candidate procedures is an extension of this basic idea.

Frequentist procedures are usually based on observables alone; thus, criteria for assessing initial precision entail averaging, in some well-defined sense, over a probability measure defined on the sample space. To a frequentist, initial precision is what matters in the assessment of a statistical procedure, since from it, one can determine

(a) the long-run average performance among all potential random samples of size n (i. e., some guarantee of repeatability of performance); and

(b) a means of determining the sampling effort required to achieve a prespecified level of precision (Casella, 1988).

Final precision, on the other hand, refers to post-data assessment of the 'accuracy' or 'precision' of an inference in the particular case under study; that is, it aims to

[2] The terms initial and final precision are somewhat of a misnomer. 'Initial precision' refers to average performance in repeated sampling, which actually represents *accuracy* (closeness to target) rather than *precision* (variation about the target or mean). In certain contexts, such as unbiasedness, the accuracy measure is fixed, in which case precision is used to 'order' the estimators. Similarly, 'final precision' may refer to post-data accuracy or precision in specific problems.

quantify the 'goodness' of an inference based on the experiment actually realized. However, the various post-data schools of thought operationalize final precision in different ways. For example, a Bayesian or fiducialist might quantify final precision in terms of (posterior) probabilities, while a likelihood advocate may express it in terms of likelihood intervals or likelihood ratios. Some conditional measures are discussed in section 1.3.4.

A critical stakeholder in the debate over initial and final precision of an inference (procedure) is the practitioner of statistical methods. Although comforted by the long-run performance of frequentist procedures, it can be easily argued that the primary concern of a practitioner or consumer is the final precision of an inference.

Although it is easy to turn a discussion of initial and final precision into a debate over frequentist and Bayesian inference, we take the position that both initial and final precision are important in the development and practice of statistical methods. Conformance to this ideal implies that statistical procedures should have certain performance guarantees in repeated sampling, but should also have performance guarantees in individual cases. Frequentist post-data inference and PDP methodology attempt to achieve these goals, but by different means.

1.3.2 Conditional inference

The genesis of conditional inference stemmed from the observation that procedures with good repeated sampling properties do not necessarily use all of the information available about a parameter θ in specific realizations (Welsh, 1996). In other words, a frequentist procedure does not necessarily take advantage of 'lucky' or 'unlucky' samples (Kiefer, 1977).

An important aspect of conditional inference lies in the notion of a *reference set*, which can be interpreted as a subset of the inference space to which the scope of an inference should be limited. Conversely, a *conditioning event* is a subset or partition of the inference space upon which a conditional inference is made. Fisher (1934, 1956, 1959) was interested in which types of conditioning events lead to good conditional performance of a procedure, and those which lead to poor conditional performance; i. e., which conditioning events are capable of acting as reference sets.

Mixture experiments and conditioning. Conditional inference is readily motivated through mixture experiments. Let $\mathscr{E} = \{\mathscr{E}_z : z \in \mathscr{Z}\}$ denote a collection of potential experiments. If different experiments $\mathscr{E}_z$ vary in their information content about the parameter of interest θ, a question arises: should one perform a frequentist analysis, or should one perform analysis based on the experiment actually performed?

Example 1.3. This example is taken from Berger and Wolpert (1984). Suppose X_1 and X_2 are independent random variables such that

$$P_\theta(X_i = \theta - 1) = P_\theta(X_i = \theta + 1) = \frac{1}{2},$$

$i = 1, 2$, where $\theta \in \Re$ is the parameter of interest. A 75% confidence set of smallest size for θ comprises the points

$$C(X_1, X_2) = \begin{cases} (X_1 + X_2)/2 & \text{if } X_1 \neq X_2, \\ X_1 - 1 & \text{if } X_1 = X_2. \end{cases}$$

If we observe $x_1 \neq x_2$, it is *certain* that $\theta = (x_1 + x_2)/2$, whereas if $x_1 = x_2$ is observed, it is equally uncertain whether $\theta = x_1 - 1$ or $\theta = x_1 + 1$. Thus, the 'conditional confidence' of $C(x_1, x_2)$ is either 1 or 0.5, depending on what we happen to observe.

This example shows that additional information supplied by observed data can modify our relative certainty in the proposition $C(x_1, x_2)$ post-data. It raises a troubling question: "does it make sense to report a pre-experimental measure when it is known to be misleading after seeing the data?" (Berger and Wolpert, 1984). A number of like examples exists in the literature; see, for example, Berger and Wolpert (1984), Casella (1988) or Goutis and Casella (1995).

Example 1.3 entails two potential experiments: $\mathscr{E}_1 = \{(x_1, x_2) : x_1 = x_2\}$ and $\mathscr{E}_2 = \{(x_1, x_2) : x_1 \neq x_2\}$. However, it is possible (and rather convenient) to consider a continuum of potential experiments $\mathscr{E} = \{\mathscr{E}_z : z \in \mathscr{Z}\}$, where a known probability distribution exists on an index set $\mathscr{Z}$. In this context, one can consider the traditional mode of conditional inference based on an ancillary statistic Z.

Conditioning on an ancillary statistic. Let $\mathscr{E} = \{\mathscr{E}_z : z \in \mathscr{Z}\}$ denote a collection of potential random experiments, of which one is randomly selected according to a known probability distribution $F(Z)$ on $\mathscr{Z}$. Given $Z = z$, the experiment $\mathscr{E}_z$ consists of observing a random sample of values of a random variable X with distribution $P_\theta(\cdot|z)$. This describes a general two-stage experiment, where Z is observed first, and conditional on $Z = z$, X is observed (Lehmann, 1986, ch. 10). For simplicity, assume that Z is continuous.

A statistic is said to be *ancillary* if its probability distribution is independent of the parameter θ. Suppose that Z is an ancillary statistic with cdf $F(\cdot)$; the two-stage experiment in this situation can then be described as follows:

(i) observe $Z = z$, drawn randomly from the distribution F;

(ii) observe X with distribution $P_\theta(\cdot|z)$.

When a statistical model admits this type of construction, the prescription is to perform inference conditional on the observed value of the ancillary Z, by the

Conditionality principle:

If an ancillary statistic Z exists for a family of models $\mathscr{P}_\theta$, then inference for the model should be made conditional on the ancillary statistic. (Welsh, 1996)

Fisher (1934) developed the theory of conditional inference based on ancillaries. An ancillary statistic Z contains no information about the parameter of interest θ *per se*; more specifically, it contains no information about the *position* or *location* of θ. However, the value of Z does provide information about the (final) *precision* of the estimate $T(x)$ of θ, so that $\text{Var}_\theta(T(x)|Z = z)$ may vary as a function of z (Kiefer, 1983).

Many problems in conditional inference can be handled by conditioning on a (maximal) ancillary. However, a general theory of conditional inference based on ancillary statistics is not feasible for the following reasons:

(i) ancillary statistics do not always exist;

(ii) when they do exist, there may be more than one (maximal) ancillary, and the question is then how to choose among them (Cox, 1971; Becker and Gordon, 1983).

(iii) there is no universally accepted definition of a partial ancillary statistic in problems containing nuisance parameters;

(iv) there have been a number of attempts to define and utilize approximate ancillary statistics in problems for which no ancillary statistic exists, but again, there is no agreement on when it is appropriate to use approximate ancillaries.

The point of conducting inference conditional on ancillary statistics is that such procedures have been shown to have 'good' conditional properties, which apply quite generally in location-scale families of distributions. On the other hand, a 'good' conditional procedure does not always have good frequentist properties. Hence, there is no compelling reason to restrict conditional inference to situations admitting ancillary statistics. In the next subsection, we consider the problem of conditioning on more general subsets of the sample space $\mathcal{X}$.

1.3.3 Conditional evaluations of statistical procedures

Another use of conditional inference is to assess whether or not the claimed repeated sampling properties of a frequentist procedure can be materially altered by conditioning. Fisher (1956), in a criticism of the Welch-Aspin solution to the Behrens-Fisher problem, observed that when conditioning on the event $S_1^2 = S_2^2$, the type I error rate of a test of $\mu_1 = \mu_2$ was bounded below by the nominal significance level. Fisher termed the event $S_1^2 = S_2^2$ a *recognizable subset*, and went on to show that if it is used as a reference set for inference with respect to the t-interval, then its claimed frequentist properties are compromised. With this work, Fisher pioneered the theory of conditional evaluations of frequentist procedures.

A formal structure to conditionally evaluate frequentist procedures was introduced by Buehler (1959). The principal idea is to evaluate procedures on the basis of a two-person betting scenario, where for a specific form of bet, the expected loss (or gain) to the bettor is interpretable as a conditional probability. Buehler's approach was enlarged and formalized by Robinson (1979a, b), so we will briefly consider the salient features of Robinson's work. The underlying purpose of such conditional evaluations is to assess the extent to which a frequentist procedure can withstand criticism, stated in terms of betting strategies (Robinson, 1979a).

Remark 1.1. In the following discussion, 'confidence' is used in a generic sense, and is not limited to Neyman-Pearson confidence procedures.

Let $\langle C(x), \gamma(x) \rangle$ be an ordered pair that denotes a confidence procedure $C(x)$ along with a confidence function $\gamma(x)$. The betting game between two players, Peter and Paul, proceeds as follows. Peter cites a procedure $\langle C(x), \gamma(x) \rangle$ as an inference for some parameter θ; Paul's task is to bet against that procedure by citing a real-valued function $s(x) \in [-1, 1]$ such that

(i) $s(x) > 0$ implies that Paul places a bet of size $s(x)$ that $\theta \in C(x)$, at odds corresponding to Peter's stated confidence $\gamma(x)$ in the procedure $C(x)$;

(ii) $s(x) < 0$ implies that Paul places a bet of size $-s(x)$ that $\theta \notin C(x)$, at odds corresponding to Peter's stated assertion $1 - \alpha(x)$ that $\theta \notin C(x)$;

(iii) $s(x) = 0$ implies that Paul makes no bet.

Robinson (1979a) goes on to describe six classes of betting procedures:

(a) The function $s(x)$ is said to be a *wide-sense betting procedure* if $s(x)$ is allowed to be unbounded, but $E[|s(X)|]$ is bounded. (This is the most general class of 'statistically interesting' betting procedures, but the following class is more relevant to the task.)

(b) The function $s(x)$ is referred to as a *betting procedure* if it is bounded as a function of x. Without loss of generality, we can restrict attention to betting procedures $s(x)$ whose range is bounded on the interval $[-1, 1]$. In this case, the sign of s indicates whether the bet is for or against the procedure.

(c) If $0 \leq s(x) \leq 1$, $s(x)$ is said to be a *positively biased* selection;

(d) If $-1 \leq s(x) \leq 0$, $s(x)$ is called a *negatively biased* selection;

(e) If $s(x) = I_A(x)$ for some subset $A \subset \mathcal{X}$, where $I(\cdot)$ denotes an indicator function, then either A or $s(x)$ is termed a *positively biased subset*;

(f) If $s(x) = -I_A(x)$ for some $A \subset \mathcal{X}$, then A or $s(x)$ is a *negatively biased subset.*

Conditional evaluations of frequentist procedures are generally conducted with betting functions of the form (e) or (f).

A conditional property of a statistical procedure is defined in terms of whether Paul can elicit a winning strategy against Peter's adopted procedure. A betting strategy is said to be

- **semirelevant**, if

$$E[\{I_{C(X)}(\theta) - \gamma(X)\}\, s(X)] \geq 0 \qquad \forall \theta \in \Theta, \tag{1.3}$$

and strictly positive for some $\theta \in \Theta$;

- **relevant**, if for some $\epsilon > 0$,

$$E[\{I_{C(X)}(\theta) - \gamma(X)\}\, s(X) - \epsilon\, |s(X)|] \geq 0 \qquad \forall \theta \in \Theta, \tag{1.4}$$

and strictly positive for some $\theta \in \Theta$.

In terms of betting strategy (e), a relevant betting strategy implies that either

$$P_\theta(\theta \in C(X) | X \in A) \geq 1 - \alpha + \epsilon \tag{1.5}$$

for all $\theta \in \Theta$, if A is a positively biased relevant subset, or

$$P_\theta(\theta \in C(X) | X \in A) \leq 1 - \alpha - \epsilon \tag{1.6}$$

for all θ, if A is a negatively biased relevant subset. If (e) is a semirelevant betting strategy, then remove the ϵ term in (1.5) and (1.6) above. According to Robinson (1979a), existence of a semirelevant betting strategy is mild criticism against a confidence procedure, while existence of a relevant betting strategy is "criticism on a level which seems to be just serious enough to bother about."

If a relevant betting strategy $s(x)$ exists on a set $A \subset \mathcal{X}$, then the confidence behavior is qualitatively changed on A, which raises suspicion about the validity of the procedure C (Casella, 1992). Robinson (1979a) concluded that identification of negatively biased relevant subsets is enough to eliminate a frequentist procedure, but identification of negative semirelevant subsets is too strong a condition to recommend against use. On the other hand, identification of positively biased (semi)relevant subsets corresponds to conservative procedures, which are generally of lesser concern, although situations exist where this direction of error is important (Casella, 1992).

Several researchers have identified relevant or semirelevant subsets in commonly used freqentist procedures. Buehler and Fedderson (1963) showed that the standard one-sample t-interval $\langle \bar{x} \pm t\, s/\sqrt{n}, 1-\alpha \rangle$ admits positively biased relevant subsets of the form $\{|\overline{X}|/S < c'\}$ for some $c' > 0$, when $n = 2$. Brown (1967) extended this to the case of arbitrary n, with some restrictions on the choice of c'. Essentially, on sets that can be interpreted as acceptance regions of a test of $H_0 : \mu = 0$, the conditional confidence of a t-interval may be bounded below by $1 - \alpha$. Robinson (1976) showed that no negatively biased relevant sets exist for the t-interval, but sets with interpretations as rejection regions of a test of $H_0 : \mu = 0$ are negatively biased semirelevant (Maatta and Casella, 1990). Maatta and Casella (1987) considered conditional properties of shortest, log-shortest and equal tailed intervals of the variance σ^2 in sampling from a $N(\mu, \sigma^2)$ population with both parameters unknown. For two-sided intervals, they concluded that

(i) equal tailed intervals admit negatively biased semirelevant subsets;

(ii) shortest intervals admit no negatively biased semirelevant subsets;

(iii) log-shortest intervals are free of relevant and semirelevant subsets (positive or negative).

Thus, log-shortest intervals have the strongest conditional properties in the normal variance problem.

Some work has been done beyond one-sample problems. Olshen (1973) extended the results of Brown (1967) to the one-way fixed effects model. Conditional on rejecting a preliminary test of equal treatment means, Olshen showed that the Scheffé multiple comparisons procedure admitted negatively biased semirelevant subsets. With respect to the Behrens-Fisher problem, Fisher (1956) showed that the Welch-Aspin procedure admitted a negatively biased relevant subset, while Robinson (1979b) showed that the Behrens-Fisher solution allowed no negatively biased relevant subsets.

Conditional evaluation of frequentist rules can be viewed as a diagnostic process that weeds out procedures with poor conditional performance, evidenced by the existence of a negatively biased relevant betting strategy. Conversely, procedures free of semirelevant subsets should have 'good' conditional properties. A key question is: under what conditions can we assert that a frequentist procedure has 'good' conditional properties? At present, there is no result that provides the exact (necessary and sufficient) conditions for the nonexistence of relevant betting strategies against a frequentist procedure. However, it is known that limits of Bayes rules which satisfy

$$\gamma(x) = \lim_{m \to \infty} \int_{\Theta} I(\theta \in C(x)) \, \pi_m(\theta|x) \, d\theta,$$

have no relevant betting strategies against them. In other words, a betting strategy of the form (e) above, whose limiting expected posterior gain is expressible as a conditional coverage probability, is free of relevant betting. An equivalent statement is that a procedure $\langle C(x), \gamma(x) \rangle$ is free of relevant betting if the confidence function $\gamma(x)$ is equal to the limiting posterior probability of $C(x)$ with respect to some (possibly improper) sequence of priors (Maatta and Casella, 1990).

1.3.4 Frequentist post-data inference

Legitimate or not, practitioners often endow real-world applications of frequentist procedures with post-data interpretations. For example, the p-value of a statistical test is often interpreted (inappropriately) as an informal measure of the strength of the sample evidence against a null hypothesis H_0—the smaller the p-value, the stronger the evidence against H_0. Confidence intervals may also be given a conditional interpretation as the range of plausible values of μ for which the observed data is 'typical', obtained from a procedure that covers the true mean $(1 - \alpha)\,100\%$ of the time in repeated sampling from the underlying distribution of the population. Such interpretations conform to the consumer's intuition that the specific evidence at hand is central to inference. Consequently, it is relevant to consider post-data measures of accuracy for frequentist procedures, which has come to be called *frequentist post-data inference.*

The foundations of this subject were set out in a series of papers by Kiefer (1975, 1976, 1977; Brownie and Kiefer, 1977). The importance of this work lies in the development of two post-data measures of accuracy for frequentist procedures: conditional confidence and estimated confidence.

Conditional confidence is expressed as $P(\theta \in C(X)|X \in A)$, where A is a conditioning set derived from a partition of the sample space $\mathcal{X}$ induced by some statistic $T(X)$. If A is a relevant subset of $\mathcal{X}$, the conditional confidence of a frequentist procedure will be be bounded below (above) by $\gamma+\epsilon$ $(\gamma-\epsilon)$ if A is positively (negatively) biased for some $\epsilon > 0$, where γ represents the pre-data confidence report. If A is a semirelevant subset, then γ is the lower (upper) bound on conditional coverage when A is positive (negative) semirelevant.

Estimated confidence, on the other hand, operates as follows: instead of reporting a constant confidence coefficient $1-\alpha$, report a data-dependent measure $1-\alpha(x)$, where x is a (vector) realization of a random variable X. This idea is particularly useful if the coverage probability of a confidence procedure depends on a parameter, in which case an estimator of the coverage probability can be derived. An estimate of the confidence level can then be obtained for each realized x. Evaluation of a class of coverage probability estimators is conducted in terms of decision theory criteria (Berger, 1988).

Remark 1.2. As in Robinson's work (1979a, b), the notion of 'confidence' is meant to be generic. Kiefer applied both conditional and estimated confidence to point estimators and statistical tests in addition to confidence procedures.

The primary shortcoming of Kiefer's theory is that the partition of the sample space $\mathcal{X}$ is arbitrary; hence, different partitions yield different results, all of which are equally valid (Goutis and Casella, 1995). Fortunately, recent developments have filled in most of the gaps in Kiefer's theory.

Conditional and frequentist acceptability. Casella (1988) contended that a commendable statistical procedure should be both 'conditionally acceptable' and 'frequentist acceptable'; i. e., it should have both 'good' pre-data properties and 'good' conditional properties. Some compromise on either or both fronts may be required, but in the end, a reasonable procedure should have some long-run performance guarantee and be free of relevant subsets.

Casella (1988) did not offer a formal criterion of frequentist acceptability, but did produce one for conditional acceptability:

Definition 1.1. A frequentist procedure is said to be *conditionally acceptable* if:

(a) it does not admit negatively biased relevant subsets;

(b) if A is a set on which it is known that a conditional inference may be drawn (i. e., a stopping region), then A cannot be a negatively biased semirelevant subset.

Condition (b) is tantamount to saying that a reference set for inference cannot be a negatively biased semirelevant subset of $\mathcal{X}$. However, this condition is difficult to quantify, since there is no established criterion for deciding which conditioning sets are acceptable as reference sets, and which are not (Berger, 1988).

If we define the confidence function $\gamma(x)$ as

$$\gamma(x) = \frac{\int_{C(x)} f(x|\theta)\,\pi(\theta)\,d\theta}{\int_{\Theta} f(x|\theta)\,\pi(\theta)\,d\theta}, \tag{1.7}$$

where $\pi(\theta)$ is a proper prior, then the procedure $\langle C(x), \gamma(x)\rangle$ is free of semirelevant subsets (Casella, 1992). To strengthen the condition, allow $\pi(\theta)$ to be a possibly improper prior. A sufficient condition for conditional acceptability, under some regularity conditions, is then (Casella, 1987)

$$\gamma(x) = \frac{\int_{C(x)} f(x|\theta)\,\pi(\theta)\,d\theta}{\int_{\Theta} f(x|\theta)\,\pi(\theta)\,d\theta} \geq 1-\alpha, \tag{1.8}$$

uniformly in x. If (1.8) holds, the procedure $\langle C(x), \gamma(x)\rangle$ admits no negatively biased relevant subsets, the first condition of Definition 1.1.

In the discussion of Casella (1988), Berger put forth a measure of frequentist acceptability, stated in terms of estimated confidence:

Definition 1.2. An estimated confidence coefficient $1-\alpha(x)$ is said to have *frequentist validity* if, for all θ,

$$E_\theta[1-\alpha(X)] \leq 1-\alpha^*(\theta) = P_\theta(\theta \in C(X)). \tag{1.9}$$

By this definition, the average reported coverage probability of the confidence procedure $\langle C(x), 1-\alpha(x)\rangle$ is bounded above by the actual coverage (as a function of θ); i. e., the confidence assertion is conservative, on average. Condition (1.9) is compatible with the frequentist approach in two respects:

(i) it provides a stated performance guarantee in repeated sampling;

(ii) it does not depend on θ.

Since there are many estimated confidence functions $1-\alpha(x)$ which satisfy (1.9), Berger suggested that estimated confidence functions be compared and evaluated in terms of risk performance.

In his rejoinder to Berger, Casella (1988) suggested placing a lower bound on (1.9), thus modifying the frequentist validity condition to

$$\inf_\theta P_\theta(\theta \in C(X)) \leq E_\theta[1-\alpha(X)] \leq P_\theta(\theta \in C(X)), \tag{1.10}$$

thereby trapping the expected conditional confidence between the confidence coefficient of the procedure $C(X)$ and the true coverage probability of $C(X)$ at θ.

Having established criteria for frequentist and conditional acceptability, what types of procedures are reasonable candidates for satisfying these dual criteria? Generally speaking, frequentist procedures that are generalized Bayes or limits of Bayes rules are prime candidates, since:

- they are generally free of relevant betting strategies, thus satisfying at least one of the conditional acceptability conditions;

- they are typically smooth procedures, usually capable of maintaining some frequentist guarantee of performance (Maatta and Casella, 1990; Casella, 1992).

Casella (1992, p. 9) cogently summarized the process of determining frequentist and conditionally acceptable procedures in practice:

> ... Considering the theories of relevant subsets, and how confidence sets are used by practitioners, the following strategy seems most reasonable. For a set $C(x)$, assert confidence $\gamma(x)$ where $\gamma(x)$ satisfies [1.7] for some (possibly improper) prior $\pi(\theta)$. This strategy assures us that $\langle C(x), \gamma(x)\rangle$ is conditionally acceptable. Moreover, we would require that $\gamma(x)$ be valid as a measure of frequentist confidence. Ideally, we would require that $\gamma(x)$ satisfy [1.8], which not only renders $\langle C(x), \gamma(x)\rangle$ frequency valid, but also yields the conditional acceptability of $\langle C(x), 1-\alpha\rangle$. However, condition [1.8] may not always be attainable and, in such a case, we would settle for $\gamma(x)$ satisfying a condition such as [1.9]. This would give some frequentist acceptability to the procedure $\langle C(x), \gamma(x)\rangle$.
>
> If neither condition [1.8] nor condition [1.9] can be attained by a $\gamma(x)$ satisfying [1.7], then frequentist acceptability may have to be compromised.

> The frequentist guarantee of the procedure may then be based on quantities such as $E_\theta\gamma(X)$, $\min E_\theta\gamma(X)$, or $\min_x \gamma(x)$ (as long as these last two quantities are positive). The point should be clear. A guaranteed legitimate conditional inference is of primary importance. After that, the frequentist guarantee should be arrived at in *some* reasonable manner.

(Bracketed equation citations are in our system.)

Estimation of indicator functions. A frequentist confidence assertion γ is interpretable as the 'best guess' conditional coverage probability of a confidence procedure $C(X)$ over the coarsest possible partition, the sample space $\mathcal{X}$ itself (Goutis and Casella, 1995). Kiefer's conditional confidence approach was to compute the coverage of $C(X)$ over some subset A of $\mathcal{X}$, where A is a member, or union of members, of some appropriate partitioning of $\mathcal{X}$. Goutis and Casella (1995) went to the opposite extreme, considering estimation of the confidence over the finest possible partition of $\mathcal{X}$, the collection $\{x : x \in \mathcal{X}\}$.

Consider the problem of evaluating the coverage probability of a confidence procedure $C(X)$, conditioned on the event $X = x$:

$$P_\theta(\theta \in C(X)|X = x) = P_\theta(\theta \in C(x)) = I(\theta \in C(x)), \tag{1.11}$$

where

$$I(\theta \in C(x)) = \begin{cases} 1 & \text{if } \theta \in C(x) \\ 0 & \text{if } \theta \notin C(x); \end{cases}$$

i. e., having observed $X = x$, the observed confidence interval $C(x)$ either covers θ or it doesn't.

The quantity $I(\theta \in C(x))$ is a sensible measure of the post-data accuracy of $C(X)$, but it is not observable since θ is unknown. Thus, it is necessary to consider estimators of $I(\theta \in C(x))$. Candidate estimators should be evaluated on the basis of risk performance. Goutis and Casella (1995) advocate use of the quadratic loss function

$$L(\gamma(x); \theta) = [\gamma(x) - I(\theta \in C(x))]^2,$$

since the Bayes rule against this loss function, for some chosen prior $\pi(\theta)$, is the posterior probability of $C(x)$; i. e.

$$\gamma^{\pi}(x) = P(\theta \in C(x)|x) = \int_{C(x)} \pi(\theta|x)\, d\theta,$$

where $\pi(\theta|x)$ is the posterior density of θ. Under quadratic loss, an estimator $\gamma(X)$ of the indicator function $I(\theta \in C(X))$ has risk

$$R(\gamma(X); \theta) = E_{\theta}[\gamma(X) - I(\theta \in C(X))]^2.$$

The advantage of this approach is that it does not depend on the existence of a sample space partition, but if one exists, the method is capable of exploiting it. The primary disadvantage is that decision theory criteria to evaluate $\gamma(X)$ are difficult to apply, since $\gamma(X)$ estimates a random function.

Summary. Frequentist post-data inference is predicated on the idea that an inference procedure should behave sensibly both pre-data and post-data. Evaluation of a procedure should entail measures of both initial and final precision. From the sampling theory viewpoint, final precision is measured in terms of estimates of $I(\theta \in C(x))$, a modification of Kiefer's conditional confidence, and estimated confidence, a la Berger (1988). These measures of final precision complement the ideas of section 1.3.3 in the following way. The dual criteria of frequentist and conditional acceptability limit the class of candidate procedures: given a procedure $C(X)$ and a realization x of X, the estimated confidence $1 - \alpha(x)$ and estimate of $I(\theta \in C(x))$ provide two measures of the post-data accuracy of the realized $C(x)$. If $C(X)$ is both frequentist and conditionally acceptable, we should expect that both the estimated confidence and estimated indicator function of $C(X)$ are bounded below by $1 - \alpha$ in ideal situations.

1.4 Improved estimation of scale parameters

One of the most famous results in statistical theory from the second half of this century is the proof by Stein (1955) that the sample mean $\overline{X}$ is inadmissible as an estimator of a mean vector $\boldsymbol{\mu}$ when the dimension of $\boldsymbol{\mu}$ is at least 3, under (composite) quadratic loss and sampling from a $N_p(\boldsymbol{\mu}, I_p)$ distribution. The area of shrinkage estimation is a direct outgrowth of this work. Stein's contributions to improvement over the sample mean in the location parameter case are well known (e. g., James and Stein, 1961); less appreciated, but equally important, are his contributions to improving upon S^2 as an estimator of the population variance σ^2. This work, and its descendants, provide a theoretical foundation for improving upon standard estimation procedures in the balanced one-way random model. Therefore, it is prudent to review this work in the most well-known case, the normal variance problem. What follows is largely based on the review of this problem by Maatta and Casella (1990), and for more recent developments, on the review paper of Kubokawa (1998a).

1.4.1 Normal variance problem

The problem embarked upon by Stein (1964) is the following. Let X be a $N(\mu, \sigma^2)$ distributed random variable, both parameters unknown, and let $X_1, X_2, \ldots, X_n$ denote a prospective random sample from this population. The joint sufficient statistic for this family of distributions is then $(\overline{X}, T)$, where $T = \sum(X_i - \overline{X})^2$. The objective is to find a point estimator of the population variance σ^2 under the normalized quadratic loss (NQL) function

$$L(d, \sigma^2) = \left(\frac{d - \sigma^2}{\sigma^2}\right)^2 = \left(\frac{d}{\sigma^2} - 1\right)^2, \tag{1.12}$$

where $d(\boldsymbol{X})$ is an estimator of σ^2. Under the group $\mathscr{G}$ of affine transformations, the family $\mathcal{F}$ of normal distributions and the loss function (1.12), the estimation problem is invariant. Furthermore, the group $\mathscr{G}$ is transitive on the parameter space Θ and decision space $\mathcal{D}$.

1.4.2 Stein-Brown-Brewster-Zidek process

Point estimation. Prior to Stein's work, it was known that (Maatta and Casella, 1990):

(i) If $X_1, \ldots, X_n$ is a random sample from a $N(\mu, \sigma^2)$ population with μ known and σ^2 unknown, the estimator

$$d_1(\boldsymbol{X}) = \frac{\sum(X_i - \mu)^2}{n + 2}$$

is admissible for estimating σ^2 under (normalized) quadratic loss (Hodges and Lehmann, 1951; Girschick and Savage, 1951).

(ii) If $X_1, \ldots, X_n$ is a random sample from a $N(\mu, \sigma^2)$ population with both μ and σ^2 unknown, then in the class of estimators of the form kT, the choice of k that minimizes the risk under (1.12) is $k = (n+1)^{-1}$. Since the risk function of this estimator is constant (equal to $2/(n+1)$), it is the *best equivariant estimator* (or BEE) of σ^2 under the group of affine transformations, transitive on Θ. Furthermore, it is minimax with respect to (1.12).

These results provide an important context for Stein's insight. The admissible estimator and the BEE differ in their multipliers by one degree of freedom; if $\mu = 0$ and $\overline{X}$ is 'small', perhaps there is a way to use the information in $\overline{X}$ to 'shrink' the

BEE towards the admissible estimator. Stein's strategy to obtain an estimator that performs such shrinkage was novel in several respects.

With the affine invariant decision problem as a foundation, let $\mathscr{H}$ be the subgroup of the affine group $\mathscr{G}$ consisting of scale and orthogonal transformations; observe that $\mathscr{H}$ does not act transitively on the parameter space. Construct the quantities $U = T/\sigma \sim \chi^2_\nu$ and $V = n\overline{X}^2/\sigma^2 \sim \chi^2_1\,(\eta^2/2)$, where $\nu = n-1$, $\eta = \mu/\sigma$, and $\lambda = \eta^2/2$ is the noncentrality parameter of V. The statistic $Z = \sqrt{n}\,|\overline{X}|/\sqrt{T}$ is the maximal $\mathscr{H}$-invariant on the sample space $\mathcal{X}$, whose distribution depends only on η^2, the maximal $\mathscr{H}$-invariant on the parameter space Θ. In the $\mathscr{H}$-invariant decision problem, $\mathscr{H}$-equivariant estimators of σ^2 are of the form $\phi(Z)\,T$; i.e., the multiplier function $\phi(Z)$ depends on the maximal $\mathscr{H}$-invariant. Furthermore, the maximal invariant Z with respect to $\mathscr{H}$ depends on *both* $\overline{X}$ and T.

Stein's strategy was to use, in essence, a preliminary test estimator. The revolutionary aspect of this work was demonstrating that such estimators can have uniformly smaller risk than the usual estimator of the variance in this one-sample problem, and more generally, in the normal fixed-effects linear model.

To establish Stein's result, we use an alternative proof from Brewster and Zidek (1974). Under NQL, the risk function of an estimator of the form $\phi(Z)\,T$ is given by

$$\begin{aligned} R(\phi(Z)\,T;\sigma^2) &= E_\eta\Big(\frac{\phi(Z)\,T}{\sigma^2} - 1\Big)^2 \\ &= E_\eta(\phi(Z)\,U - 1)^2 \\ &= E_\eta[E_\eta[(\phi(Z)\,U - 1)^2\,|Z]] \\ &= E_\eta[\phi^2(Z)\,E_\eta(U^2\,|Z) - 2\phi(Z)\,E_\eta(U\,|Z) + 1]. \end{aligned}$$

This risk function can be minimized through the inner conditional expectation, one of the important insights of Stein (1964). Taking the derivative of the conditional risk with respect to $\phi(Z)$ and setting the result to zero, the optimal multiplier under

NQL for fixed $\eta \geq 0$ is

$$\phi_\eta(Z) = \frac{E_\eta(U|Z)}{E_\eta(U^2|Z)}.$$

Unfortunately, $\phi_\eta(Z)$ depends on $|\eta|$, so $\phi_\eta(Z)\,T$ is not an estimator. However, Brewster and Zidek (1974) showed that, for all η,

$$\phi_\eta(z) \leq \phi_0(z) = \frac{1+z^2}{n+2} \tag{1.13}$$

pointwise in z, by the increasing monotone likelihood ratio property of noncentral χ^2 distributions in $|\eta|$ for each fixed $z > 0$. Moreover, for $z \leq (n+1)^{-1}$,

$$\phi_0(z) \leq \frac{1}{n+1}.$$

As a result, the estimator $\phi_S(Z)\,T$, where

$$\phi_S(Z) = \min\Big(\frac{1+Z^2}{n+2}, \frac{1}{n+1}\Big) = \min(\phi_0(Z), (n+1)^{-1}) \tag{1.14}$$

has uniformly smaller risk than the BEE. This is precisely Stein's estimator.

Observe that

$$(1+Z^2)\,T = (1+n\overline{X}^2/T)\,T = T + n\overline{X}^2 = \sum X_i^2;$$

therefore, when $Z^2 < (\nu+2)^{-1}$,

$$\phi_S(Z)\,T = \frac{\sum X_i^2}{\nu+3}.$$

Thus, when Z^2 is 'small', Stein's estimator shrinks to the Hodges-Lehmann estimator with $\mu = 0$. This is sensible behavior, as Z^2 small provides sample evidence that μ is (near) zero. Thus, for 'small' values of Z^2, pooling the sum of squares T with $n\overline{X}^2$ enables one to shrink from the (inadmissible) BEE of σ^2 when μ is unknown towards the (admissible) Hodges-Lehmann estimator when $\mu = 0$, in the process gaining up to an additional degree of freedom for estimating σ^2.

Stein's improved estimator (1.14) was an impressive theoretical achievement, but he recognized that it was inadmissible. Brown (1968) extended Stein's work to a wider class of distributions and a wider class of loss functions. In doing so, Brown considered estimators of the form $\phi_{c,d}(Z)T$, where

$$\phi_{c,d}(Z) = \begin{cases} c & \text{if } Z^2 \leq r_1^2 \\ d & \text{if } Z^2 > r_1^2, \end{cases}$$

for some constants c, d and r_1. Under NQL, the risk of this estimator is

$$R(\phi_{c,d}(Z)\,T;\sigma^2) = E_\eta[(cU-1)^2\,|Z^2 \leq r_1^2]\,P_\eta(Z^2 \leq r_1^2) \\ + E_\eta[(dU-1)^2\,|Z^2 > r_1^2]\,P_\eta(Z^2 > r_1^2).$$

Differentiating the risk with respect to c and setting the result to zero yields the 'optimal' c:

$$c_\eta(r_1^2) = \frac{E_\eta(U|Z^2 \leq r_1^2)}{E_\eta(U^2|Z^2 \leq r_1^2)}.$$

Since the family of distributions of $U|Z^2 \leq r_1^2$ has increasing monotone likelihood ratio in $|\eta|$ for fixed $r_1 > 0$, $c_\eta(r_1^2) \leq c_0(r_1^2)$; furthermore, $c_0(r_1^2) < (\nu+2)^{-1}$ for all $r_1 > 0$. Brown's estimator is then of the form $\phi_B(Z)\,T$, where the multiplier function is given by

$$\phi_B(Z) = \begin{cases} c_0(r_1^2) & \text{if } Z^2 \leq r_1^2; \\ 1/(\nu+2) & \text{if } Z^2 > r_1^2. \end{cases} \tag{1.15}$$

Brown's estimator also dominates the best $\mathscr{G}$-equivariant estimator $T/(\nu+2)$ under (1.12) (Brown, 1968). Like Stein's estimator, Brown's estimator of σ^2 is inadmissible under NQL.

Brewster and Zidek (1974) noted that one could obtain an improved version of (1.15) by selecting a point $r_2 < r_1$ and iterating the above process. The resulting estimator $\phi_2(Z)\,T$, where

$$\phi_2(Z) = \begin{cases} c_0(r_2^2) & \text{if } Z^2 \le r_2^2 \\ c_0(r_1^2) & \text{if } r_2^2 < Z^2 \le r_1^2 \\ (\nu+2)^{-1} & \text{if } Z^2 > r_1^2, \end{cases}$$

has lower risk than $\phi_B(Z)\,T$. There is no compelling reason to stop here: one could choose another point $r_3 < r_2$, and by repeating the above process, improve upon $\phi_2(Z)T$. One can obtain successively better estimators by selecting, at stage k, a constant $r_k < r_{k-1}$ and reiterating the argument.

Brewster and Zidek (1974) next constructed a sequence of step-function estimators $\{\phi^{(i)}(Z)T\}$ of σ^2 as follows. Let $\{R_i\}$ denote a sequence of partitions of $(0,\infty)$ such that, for each $i = 1, 2, 3, \ldots$ and $j = 1, 2, ..., n_i$,

(i) $0 = r_{i0} < r_{i1} < r_{i2} < \cdots < r_{i,n_i} = \infty$;

(ii) $\{r_{ij} : j = 1, 2, ..., n_i\} \subset \{r_{i+1,j} : j = 1, 2, ..., n_{i+1}\}$;

(iii) $\lim_{i\to\infty} r_{i,n_i-1} = \infty$;

(iv) $\lim_{i\to\infty} \max_{1<j\le n_i-1} |r_{ij} - r_{i,j-1}| = 0$.

Within a particular partition R_i, apply the inductive process described above to the set of break points in R_i, leading to the step-function estimator $\phi^{(i)}(Z)\,T$, where

$$\phi^{(i)}(Z) = \begin{cases} c_0(r_{ij}^2), & r_{i,j-1}^2 < Z^2 \le r_{ij}^2 \\ (\nu+2)^{-1} & Z^2 > r_{i,n_i-1}^2, \end{cases} \tag{1.16}$$

for $j = 1, 2, \ldots, n_i - 1$. For each i, $\phi^{(i)}(Z)T$ dominates all other step function estimators produced by the iterative process on R_i, as well as the BEE. However, *across different partitions*, none of the $\phi^{(i)}(Z)\,T$ necessarily dominates another in risk, since they are based on different partitions of $(0,\infty)$.

Now, consider the limit of the sequence of step-function estimators $\{\phi^{(i)}(Z)T\}$ as $i \to \infty$. By conditions (iii) and (iv) above and the form of $\phi^{(i)}(Z)$, Brewster and Zidek showed that

$$\lim_{i\to\infty} \phi^{(i)}(Z) = \phi_{BZ}(Z),$$

where, for each fixed $z \in (0, \infty)$,

$$\phi_{BZ}(z) = \frac{E_0(U|Z^2 \le z^2)}{E_0(U^2|Z^2 \le z^2)}. \tag{1.17}$$

The limiting estimator $\phi_{BZ}(Z)\,T$ is called the *Brewster-Zidek estimator* of σ^2 with respect to normalized quadratic loss.

The multiplier function $\phi_{BZ}(Z)$ is monotone increasing and continuous in Z, where $\phi_{BZ}(0) = (\nu+3)^{-1}$ and $\phi_{BZ}(\infty) = (\nu+2)^{-1}$ [under (1.12)]. By Theorem 2.1.3 of Brewster and Zidek (1974), the risk of $\phi_{BZ}(Z)\,T$ dominates that of the BEE. Furthermore, they showed that $\phi_{BZ}(Z)\,T$ is generalized Bayes and admissible in the class of scale equivariant estimators. Proskin (1985) later showed that the BZ estimator is admissible among all estimators of σ^2.

Kubokawa (1998) briefly reviewed two other improved point estimators of σ^2. Strawderman (1974) developed a smooth estimator of σ^2 by decision theory arguments different from those of Brewster and Zidek (1974). Shinozaki (1995) proposed three types of non-equivariant point estimators of σ^2. Under certain conditions on the multiplier function, Shinozaki's estimators have uniformly lower risk than the BEE, and can achieve much larger risk improvements than does Stein's estimator.

Interval estimation. Several researchers have developed improved interval estimators of σ^2 in the normal variance problem. The general form of an (unconditional) interval estimator of σ^2 is $(T/b, T/a)$, where T is the sum of squares, a is a lower quantile of the distribution of a random variable or pivotal U, and b is the corresponding upper quantile, satisfying $P(a \le U \le b) = 1 - \alpha$.

The seminal work in extending the ideas of improved point estimation to the problem of interval estimation was that of Cohen (1972), who considered a class of improved interval estimators of the form $(\phi(Z)\,T, (\phi(Z)+c)\,T)$, where $\phi(Z)$ is a data-dependent multiplier function, c is the length of the 'usual' shortest interval

$$I_U = \left\{\sigma^2 : \frac{T}{a} \leq \sigma^2 \leq \left(\frac{1}{a}+c\right)T\right\},$$

and a is the lower quantile of a χ^2_ν distribution. For any point $r \in (0,1)$, Cohen's interval is defined as

$$I_c(Z,T) = \begin{cases} (\phi_c(r^2)\,T, (\phi_c(r^2)+c)\,T) & \text{if } Z^2 \leq r^2 \\ \left(\frac{T}{a}, \left(\frac{1}{a}+c\right)T\right) & \text{if } Z^2 > r^2, \end{cases}$$

a 'step-function' interval with a discontinuity point at r^2, where $\phi_0(r^2)$ is the unique solution of

$$g_{\nu+4}\left(\frac{1}{\phi}\right) G_1\left(\frac{r^2}{(1-r^2)\,\phi}\right) = g_{\nu+4}\left(\frac{1}{\phi+c}\right) G_1\left(\frac{r^2}{1-r^2}\frac{1}{\phi+c}\right). \tag{1.18}$$

Cohen showed that $\phi_c(r^2)$ can be chosen so that it is never larger than $1/a$, thus providing shrinkage towards zero if $Z^2 \leq r^2$. Furthermore, Cohen proved that this interval has uniformly higher unconditional coverage than the unconditional interval.

The process developed by Brewster (1972) [see also Brewster and Zidek, 1974] provided the mechanism for improving upon Cohen's interval. As in the point estimation case, fix r and pick a value $r' < r$. Construct the interval

$$I_c^{(2)}(Z,T) = \begin{cases} (\phi_c(r'^2)\,T, (\phi_c(r'^2)+c)\,T) & \text{if } Z^2 \leq r'^2 \\ (\phi_c(r^2)\,T, (\phi_c(r^2)+c)\,T) & \text{if } r'^2 < Z^2 \leq r^2 \\ \left(\frac{T}{a}, \left(\frac{1}{a}+c\right)T\right) & \text{if } Z^2 > r^2; \end{cases}$$

by application of Cohen's result, $I_c^{(2)}(\cdot)$ has higher unconditional coverage than both I_U and $I_c(\cdot)$. By taking successively smaller constants $r'', r''', \ldots$ and iterating the argument, one can always produce an interval with higher unconditional coverage than its predecessors.

Now, construct a sequence of partitions $\{R_i\}$ of $[0,1]$ with properties (i)–(iv) discussed earlier in this section; in conditions (i) and (iii), replace ∞ by 1. Within partition R_i, fix r_{i,n_i-1} and produce a finite sequence of successively better step-function intervals of σ^2 by following the process in the preceding paragraph. The best step-function interval in R_i is denoted by $I^{(i)}(Z,T)$, where

$$I^{(i)}(Z,T) = \begin{cases} (\phi_0(r_{ij})\,T, (\phi_0(r_{ij})+c)\,T) & \text{if } r_{i,j-1}^2 < Z^2 \le r_{ij}^2 \\ \left(\frac{T}{a}, \left(\frac{1}{a}+c\right)T\right) & \text{if } Z^2 > r_{i,n_i-1}^2 \end{cases}$$

and $\phi_0(r_{ij})$ is the solution of

$$g_{\nu+4}\Big(\frac{1}{\phi}\Big)\,G_1\Big(\frac{r_{ij}^2}{(1-r_{ij}^2)\,\phi}\Big) = g_{\nu+4}\Big(\frac{1}{\phi+c}\Big)\,G_1\Big(\frac{r_{ij}^2}{1-r_{ij}^2}\,\frac{1}{\phi+c}\Big).$$

for each $j = 1, 2, \ldots, n_i - 1$. By properties (ii)–(iv) of $\{R_i\}$,

$$\lim_{i\to\infty} I^{(i)}(Z,T) = I_{BZ}(Z,T),$$

where

$$I_{BZ}(Z,T) = (\phi_{BZ}(Z)\,T, (\phi_{BZ}(Z)+c)\,T) \tag{1.19}$$

and $\phi_{BZ}(t)$ is the unique solution of

$$g_{\nu+4}\Big(\frac{1}{\phi_{BZ}(t)}\Big)\,G_1\Big(\frac{t}{\phi_{BZ}(t)}\Big) = g_{\nu+4}\Big(\frac{1}{\phi_{BZ}(t)+c}\Big)\,G_1\Big(\frac{t}{\phi_{BZ}(t)+c}\Big) \tag{1.20}$$

for each $t > 0$. Shorrock (1990) showed that the BZ-type interval (1.19) has uniformly higher unconditional coverage than the usual shortest interval, and is generalized Bayes in the class of shortest intervals of the form $(\phi(Z)\,T, (\phi(Z)+c)\,T)$. If, in the foregoing discussion, we replace $\nu+4$ with $\nu+2$ and $\phi_{BZ}(t)+c$ by $c\,\phi_{BZ}(t)$ in (1.20), then by following a similar line of argument, the resulting BZ-type interval is generalized Bayes in the class of log-shortest intervals of σ^2 having the generic form $(\phi(Z)\,T, c\,\phi(Z)\,T)$, and has uniformly higher unconditional coverage than the corresponding 'usual' log-shortest interval. Kubokawa (1994) applied the IERD method (see section 1.4.3) to obtain the same log-shortest BZ-type of σ^2.

A strategy that simultaneously improves upon the length and unconditional coverage of an unconditional shortest interval estimator of σ^2 was described by Goutis and Casella (1991). Their interval is of the form

$$I_{GC}(Z,T) = (\phi_1(Z)\,T, \phi_2(Z)\,T), \tag{1.21}$$

where $\phi_1(t)$ and $\phi_2(t)$ satisfy the conditions

$$g_{\nu+4}\Big(\frac{1}{\phi_1(t)}\Big)\,G_1\Big(\frac{\tau(t)}{\phi_1(t)}\Big) = g_{\nu+4}\Big(\frac{1}{\phi_2(t)}\Big)\,G_1\Big(\frac{\tau(t)}{\phi_2(t)}\Big)$$
$$\Big(\frac{d\phi_1(t)}{dt}\Big)\,g_{\nu+4}\Big(\frac{1}{\phi_1(t)}\Big)\,G_1\Big(\frac{t}{\phi_1(t)}\Big) = \Big(\frac{d\phi_2(t)}{dt}\Big)\,g_{\nu+4}\Big(\frac{1}{\phi_2(t)}\Big)\,G_1\Big(\frac{t}{\phi_2(t)}\Big),$$

and $\tau(t)$ is a positive-valued function satisfying $\tau(t) \geq t$. If $\tau(t) = t$, then the GC interval reduces to Shorrock's interval; but if $\tau(t) > t$, different values of $\tau(t)$ produce different intervals, each of which is shorter than Shorrock's interval, with coverage at least $1-\alpha$ for all $\eta \geq 0$. (Note that changing $\nu+4$ to $\nu+2$ produces similarly improved log-shortest estimators of σ^2.)

The key feature to note in this discussion is that the smooth procedures which improve upon the usual point and interval estimators are obtained by *conditioning* on an interval of Z^2 values.

1.4.3 IERD method

Kubokawa (1994) introduced a process for unifying improved point and interval estimation, termed the "Integrated Expression of Risk Difference", or IERD, method.

The following discussion is limited to its application in the normal variance problem, although it has been usefully exploited in a variety of estimation problems; see Kubokawa (1994, 1997, 1998) for further details.

The central idea behind the IERD method is employment of a definite integral to express the risk difference between a 'standard' and generic 'improved' estimator in order to generate a class of improved estimators of a parameter θ. The method is sufficiently broad to encompass various distributional families and loss functions, thereby generalizing the theory of improved estimation.

Let $\delta_0 = kT$ represent a standard unconditional point estimator of σ^2 in the normal variance problem, and let δ_ϕ denote a proposed improved estimator of the form $\delta_\phi = \phi(Z^2)\,T$, with the property that $\phi(\infty) = k$. Furthermore, let $L(t)$ denote the loss function, where $t = d/\sigma^2$ and d represents an estimator of σ^2, conforming to the notation in Kubokawa (1994). The loss function $L(t)$ is absolutely continuous and strictly bowl-shaped; consequently, $L(t)$ is differentiable almost everywhere. To apply the IERD method in the normal variance problem, the following assumptions are necessary:

1. To guarantee interchange of limits and integrals,

$$\int_0^\infty L(cv)\, g(v)\, dv < \infty \qquad \text{and} \qquad \int_0^\infty |L'(cv)|\, v\, g(v)\, dv < \infty,$$

   where $c > 0$ and $g(\cdot)$ denotes the χ^2 density of U.

2. In addition,

$$\int_0^\infty |L'(cv)|\, v^2\, g(v)\, h(v)\, dv < \infty,$$

   where $h(v) = h(v; 0)$ is the central χ_1^2 density of V when $\mu = 0$.

3. The MLR property of noncentral χ^2 distributions is required:

$$\frac{H(x;\lambda)}{H(x;0)} \text{ is nondecreasing in } x > 0.$$

4. To guarantee a unique solution, we also require that $g(c_1 x)/g(c_2 x)$ is strictly increasing in x for $0 < c_1 < c_2$.

The following result takes advantage of the bowl-shaped property of the loss function (Strawderman, 1974):

Lemma 1.1. *Let $h(x)$ be a nondecreasing, integrable function on an interval (a, b), and let $\nu(\cdot)$ denote a finite measure on (a, b). Let $K(x)$ be an integrable function on (a, b). If there exists a point $x_0 \in (a, b)$ for which $K(x) \leq 0$ when $x \leq x_0$ and $K(x) \geq 0$ when $x \geq x_0$, then*

$$\int_a^b K(x)\, h(x)\, \nu(dx) \geq h(x_0) \int_a^b K(x)\, \nu(dx),$$

where the equality holds iff $h(x)$ is constant a. e.

Kubokawa's main result is now given:

Theorem 1.2. *Assume that the MLR property on $H(\cdot)$ holds; furthermore, assume that*

(a) $\phi(w)$ is nondecreasing, with $\lim_{w \to \infty} \phi(w) = k$;

(b) $\int_0^\infty L'(\phi(w)u)\, u\, g(u)\, H(wu; 0)\, du \geq 0$.

Then δ_ϕ is at least as good as δ_0, where $\delta_\phi = \phi(Z^2)\, T$ and $\delta_0 = k\, T = T/(\nu + 2)$.

An instructive proof of this result is in Kubokawa (1998).

The IERD method provides a means of obtaining a class of improved estimators without explicitly invoking invariance; however, it assumes the same monotone likelihood ratio properties as in the SBBZ process. Thus, in the normal variance problem, the estimators of Stein, Brown and Brewster & Zidek are all members of this improved class. Furthermore, Section 3 of Kubokawa (1994) shows how the IERD method can be used to obtain a class of improved interval estimators; in particular, the intervals of Cohen (1972) and Shorrock (1990) are members of the class of improved shortest interval estimators with length fixed to that of the unconditional shortest interval.

The IERD method has been applied to several types of inference problems in the past few years; see Kubokawa (1997, 1998) for recent reviews of these developments.

1.5 Summary

In this chapter, the primary influences on PDP inference were reviewed: initial and final precision, conditional inference, frequentist post-data inference and estimation in the normal variance problem. A brief discussion of their roles in relation to PDP inference follows.

PDP inference does not view initial and final precision as an either-or proposition. Instead, it starts with a procedure that has good initial precision, and through PDP conditioning, modifies it to have good final precision. We view this as a critical feature of PDP inference: it suggests that in certain circumstances, one can start with recommended frequentist procedures, and improve their estimates 'on-the-fly' based on suitable PDP conditioning. As far as we know, this is a feature unique to PDP inference.

PDP inference is both conditional and sampling-theory oriented, so the review of conditional inference is included to provide some historical context. Conventionally, one conditions on some partition of the sample space $\mathcal{X}$. In contrast, PDP conditioning events are applied on a *pivotal space*. There is no arbitrariness in the choice of a PDP conditioning event; it is driven by the parameter constraint(s) of the underlying model and the observed data through some statistic. Finally, a PDP conditioning event is meant to be a reference set for inference on the pivotal space.

A fundamental goal of frequentist post-data inference is to limit the class of frequentist procedures to those that meet the twin criteria of frequentist and conditional acceptability, as defined by Berger (1988) and Casella (1988), respectively. Measures of final precision, such as estimation of indicator functions or estimated confidence, are designed as aids in determining whether conditional and frequentist acceptability is achieved. In contrast, the PDP approach actively modifies frequentist procedures to achieve conditional acceptability. We then turn around to investigate the frequentist properties of the PDP procedures to assess the extent to which they are frequentist acceptable. Therefore, PDP inference attempts to be compatible with the spirit of frequentist post-data inference, but differs in the implementation.

Chapter 2

One-way random model

2.1 Introduction

The objective of this chapter is to develop the post-data pivotal (PDP) approach to statistical inference in the context of the balanced one-way random effects model. Methodologically, the goal is to develop sampling theory-based procedures that behave sensibly both pre-data and post-data. Starting from the conventional formulation of the balanced one-way random model, the following strategy is employed:

1. Defining σ_e^2 to be the error variance component and σ_α^2 to be the treatment variance, assume that $\sigma_e^2 > 0$ and $\sigma_\alpha^2 \geq 0$, thereby inducing a constraint on the expected mean squares.

2. Using the information from the parameter constraint in (1), construct a conditioning event on the joint pivotal space, which restricts inference to a subregion of the pivotal space.

3. Condition each of the basic pivotals by the event in (2). This modifies the distribution theory of the basic pivotals as a function of the observed data.

4. Conduct point and interval estimation using the modified distributions from (3) as the probabilistic foundation for inference. In other words, conduct the 'usual' analysis, but based on the distributions in (3) rather than the unconditional distributions.

The consequences of following this process are that:

(i) all PDP point and interval estimates assume values within their respective parameter spaces;

(ii) PDP point estimators are at least as good as their unconditional counterparts in frequentist risk performance under a wide class of loss functions;

(iii) PDP interval estimators with the same length as corresponding unconditional intervals are at least as good in terms of unconditional coverage probability;

(iv) PDP distributions are connected to the Bayesian posteriors of Box and Tiao (1973) and the fiducial distributions of Wild (1981) and Venables & James (1978), which in some sense unifies the PDP approach with other post-data modes of inference;

(v) PDP distributions can be obtained by a Brewster-Zidek type of limiting argument, providing a justification for (ii) and (iii);

(vi) the process that leads to (i)–(v) derives from the underlying model and its assumptions.

The approach is based on the premise that conditioning on a (joint) pivotal space allows one to incorporate *both* the data and the constrained parameter(s) to improve statistical inference. When applied to each of the (pivotal) distributions under the entertained model, it forms a basis for post-data inference from the sampling theory perspective that can be directly tied to Bayesian, fiducial and likelihood inference, and unifies conditional sampling theory with these post-data paradigms. Furthermore, the connections to fiducial inference and use of the Invariance Principle in the selection of statistical procedures provides further links to structural inference (Fraser 1968, 1979, 1983) and pivotal inference[1] (Barnard, 1995).

Example 2.1. To illustrate the type of problem that will be addressed in this chapter, consider a balanced one-way random model with $I = 4$ treatments and $J = 2$ replications per treatment. Assume that the standard formulation of the one-way random model is in force (see section 2.2) and the following ANOVA table obtains:

Source	df	SS	MS	E(MS)
Treatment	3	30	10	$\tau_2 = \sigma_e^2 + 2\,\sigma_\alpha^2$
Error	4	60	15	$\tau_1 = \sigma_e^2$
Total	7	90		

The 'usual' estimators of the variance components σ_e^2 and σ_α^2 are the method of moments (or ANOVA) estimators, obtained by equating observed with expected mean squares. The estimates are $\hat{\sigma}_e^2 = 15$ (the MSE) and $\hat{\sigma}_\alpha^2 = -2.5$, where

$$\hat{\sigma}_\alpha^2 = \frac{\text{MS(treatment)} - \text{MSE}}{2}.$$

Since the 'usual' model formulation assumes that both variance components are non-negative, the obtained estimate of σ_α^2 is disconcerting, but not terribly unusual in practice. Whenever the observed mean square ratio is less than one, the realized estimate $\hat{\sigma}_\alpha^2$ will be negative. Thus, the ANOVA method does not always produce a sensible estimate of σ_α^2.

[1] Despite the similarities in names, the PDP approach is not a special form of Barnard's pivotal inference, and should not be interpreted as such. The term 'post-data pivotal' is to emphasize that the basic pivotal quantities of a model are conditioned upon. The effect is to modify the distribution theory in light of both the observed data and the parameter constraint(s), which in turn modifies the standard inference procedures.

ANOVA estimators in balanced random models are 'optimal' in an average sense: taken over the ensemble of all possible random samples of fixed size N, ANOVA estimators turn out to be best, *on average*, in the restricted class of quadratic, unbiased estimators of the variance components. If the normality assumption applies, this result can be extended to the broader class of unbiased estimators. However, 'good' average pre-data performance does not necessarily translate to 'good' post-data performance, as illustrated by this problem. In particular, whenever the observed sums of squares in a random or mixed model do not conform to the ordering implied by the expected mean squares, the 'pre-data optimal' ANOVA estimators exhibit undesirable post-data behavior.

In contrast, the PDP approach modifies unconditional estimators of variance functions by, in essence, adjusting the error and treatment degrees of freedom post-data, the amount of adjustment depending on the magnitude of the sum of squares ratio s (or equivalently, the observed mean square f). This adjustment ensures that estimators of the variance functions lie in the appropriate parameter space, in accordance with the constraint(s) implied by the set of expected mean squares. Another way to phrase this is to say that the degrees of freedom adjustment is a mechanism to bring the ordering of the modified mean squares in line with the ordering of the expected mean squares. This chapter examines the motivations for, and consequences of, such post-data adjustments.

The distribution theory developed in this chapter is similar to the Bayesian theory of Box and Tiao (1973, ch. 5) or the fiducial distributions of Wild (1981) or Venables and James (1978). Essentially, PDP inference in the balanced one-way random model is a sampling-theory analogue of Box and Tiao's Bayesian analysis. The advantage of the Bayesian approach is that probability is defined on the parameter space; in one-way random models, the parameter constraint is handled directly by making it an integration limit when marginalizing over the joint posterior density of the expected mean squares. The fiducial method, on the other hand, integrates the inverted pivotal of a variance function over a constrained region, dictated in part by the parameter constraint (e. g., Venables and James, 1978). In each case, one achieves:

- a mode of inference that directly accounts for the parameter constraint(s);
- point and interval estimates of parametric functions of interest that lie in the appropriate parameter space with probability 1;
- a 'smooth pooling' of error and treatment sums of squares.

The PDP approach is also *model-based*, in that the distribution theory and resulting estimators derive from the model assumptions and the parameter constraint. Thus, the proposed methodology fundamentally depends on the model formulation.

The 'pooling dilemma'. In random and mixed effects models, a problem arises for the frequentist statistician if one fails to reject a null hypothesis that a particular variance component (or set of fixed effects) is zero. In the example considered above, we would not have rejected $H_0 : \sigma_\alpha^2 = 0$ by the F test. Under H_0, the two expected mean squares are equal (i. e., $\tau_1 = \tau_2$). If this is indeed the case, then the observed mean squares m_1 and m_2 yield two independent estimates of the error variance τ_1, so it is reasonable that one should pool this information to improve its estimate. The question is: when should one pool, how, and how much? This has come to be known as the "pooling dilemma".

This is not a simple question for the frequentist, as various attempts to solve this problem attest. One approach is to imbed a 'pooling rule' into the procedure, as in REML estimation of σ_e^2 and σ_α^2 in the one-way random model. When the observed F ratio is less than 1, the rule is to pool both the sums of squares and the degrees of freedom into the estimator of the error variance, and setting the estimator of σ_α^2 to zero. Similarly, preliminary test procedures are usually associated with conditional pooling rules; however, such procedures are typically inadmissible from a decision theory perspective, and none of them 'smoothly' trades off information between treatment and error. One of the compelling features of the PDP method is that it produces statistical procedures that do exhibit smooth pooling, and can be connected to the aforementioned Bayesian and fiducial approaches. Furthermore, these procedures have reasonable pre-data properties, as will be shown in various places throughout this chapter.

We turn next to the foundations required to apply the PDP approach in the one-way random model. We begin with the 'usual' model assumptions, and outline the customary probabilistic structure under this model. We then turn to consideration of the conditioning event that activates the post-data pivotal method; from there, we develop the PDP approach from the ground up. Finally, we compare the resulting PDP procedures with some well-known frequentist procedures and show that, in general, they work quite well.

2.2 Unconditional 1-way random model

The balanced one-way random effects model is expressible in nonmatrix form as

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \tag{2.1}$$

with treatment levels $i = 1, 2, \ldots, I$, and within-cell replicates $j = 1, 2, \ldots, J$, where

- y_{ij} is the observed response of (experimental) unit j in (treatment) group i;
- μ is the unknown grand population mean;
- α_i is the random effect associated with (treatment) group i; and
- ϵ_{ij} is the unobservable random error associated with the j^{th} experimental unit in (treatment) group i.

The model parameters are $(\mu, \sigma_e^2, \sigma_\alpha^2)$.

It is customary to impose the following model assumptions on the random effects:

$$\begin{aligned} \alpha_i &\sim IN(0, \sigma_\alpha^2) & i &= 1, 2, \ldots, I, \\ \epsilon_{ij} &\sim IN(0, \sigma_e^2) & j &= 1, 2, \ldots, J, \\ \text{cov}(\alpha_i, \epsilon_{ij}) &= 0 & &\text{for all } i, j. \end{aligned}$$

Under the above assumptions, it follows that

$$\begin{aligned} \text{cov}(y_{ij}, y_{kl}) &= \text{cov}(\mu + \alpha_i + \epsilon_{ij}, \mu + \alpha_k + \epsilon_{kl}) \\ &= \begin{cases} \sigma_\alpha^2 + \sigma_e^2 & \text{if } i = k, j = l; \\ \sigma_\alpha^2 & \text{if } i = k, j \neq l; \\ 0 & \text{if } i \neq k, j \neq l. \end{cases} \end{aligned}$$

That is, observations within the same (treatment) group are assumed to be positively correlated, but observations from different (treatment) groups are assumed to be uncorrelated, and thus independent under normality. More specifically, the covariance matrix of the response vector $\boldsymbol{y}$ is compound symmetric.

It is explicitly assumed that $\sigma_e^2 > 0$ and $\sigma_\alpha^2 \geq 0$ in this model formulation. Without it, the approach to inference outlined below does not apply, and alternative methods, such as those of Smith and Murray (1984), are more appropriate.

Notation: The notation in this chapter is sufficiently extensive to warrant a short detour. There are three basic categories of notation: entries in the ANOVA table, parametric functions, and pivotal quantities. When subscripts are necessary, the subscript 1 will be associated with the error term, while the subscript 2 corresponds to the treatment factor. The term 'treatment' is used generically in the sense that it represents the 'between group' factor in an experimental design or observational study. Throughout this chapter, upper case letters will represent random variables

(e. g., T_i or M_i), while lower case letters represent their observed (or realized) values; e. g., t_i or m_i.

Entries in the ANOVA table. Under model (2.1), the analysis of variance table is given by

Source	df	SS	MS	Expected MS
Treatment	ν_2	T_2	M_2	$\tau_2 = \sigma_e^2 + J\sigma_\alpha^2$
Error	ν_1	T_1	M_1	$\tau_1 = \sigma_e^2$
Total	$\nu_1 + \nu_2$	$T_1 + T_2$		

where, for $i = 1, 2$,

- T_i is a sum of squares;
- ν_i = degrees of freedom (df), where $\nu_1 = I(J-1)$ is the error df and $\nu_2 = I - 1$ the treatment df;
- $M_i = T_i/\nu_i$ is a mean square;
- τ_i is an expected mean square;

Parametric functions. The parametric functions of interest in the one-way random effects model are:

- the error variance component $\tau_1 = \sigma_e^2$;
- the expected mean square for treatments $\tau_2 = \sigma_e^2 + J\sigma_\alpha^2$;
- the treatment variance component σ_α^2;
- the variance component ratio $\gamma = \sigma_\alpha^2/\sigma_e^2$;
- the expected mean square ratio $\sigma = \tau_2/\tau_1 = 1 + J\,\gamma$; and

- the intraclass correlation

$$\rho = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_e^2}.$$

Point and interval estimators of these quantities are of interest.

Pivotal quantities and their distributions. Based on the model (2.1) and its underlying assumptions, we have

- $U_i = T_i/\tau_i \sim \chi^2_{\nu_i}$, $i = 1, 2$, the unconditional pivotal quantities associated with error and treatment, respectively;
- $\phi_i = \nu_i/2$, $i = 1, 2$;
- $S = T_2/T_1$, the sum of squares ratio;
- $W = U_2/U_1 = S/\sigma$, the ratio of unconditional treatment and error pivotals;
- $Z = S/(1+S)$, the usual R^2 in the one-way model;
- $g_\nu(x)$, a χ^2 density with ν degrees of freedom;
- $G_\nu(x)$, a χ^2 cdf with ν degrees of freedom evaluated at x;
- $I_x(\phi_2, \phi_1)$, a Beta cdf with parameters ϕ_2 and ϕ_1, evaluated at $x \in [0, 1]$; and
- $F = (\nu_1/\nu_2)\, S$, so that $F/\sigma \sim \mathcal{F}_{\nu_2,\nu_1}$.

The pivotal W has a Beta type II distribution, with density

$$h_W(w) = \frac{\Gamma(\phi_1 + \phi_2)}{\Gamma(\phi_1)\,\Gamma(\phi_1)}\,\frac{w^{\phi_2 - 1}}{(1+w)^{\phi_1+\phi_2}}, \tag{2.2}$$

defined for all $w > 0$.

2.3 PDP conditioning argument

By assuming that $\sigma_e^2 > 0$ and $\sigma_\alpha^2 \geq 0$ in model (2.1), the expected mean squares follow the inequality $\tau_2 \geq \tau_1$. This information is available pre-data. Figure 2.1 illustrates this constraint in the joint (τ_1, τ_2) parameter space. Conditioning on some subset of the sample space $\mathcal{X}$, as is the case in conditional inference, fails to take advantage of the parameter constraint. However, if we think about conditioning on a *pivotal space*, then the parameter constraint can be used to advantage.

Under the one-way random model (2.1), the basic pivotals are U_1 and U_2 (along with their ratio W), so consider the joint pivotal space of (U_1, U_2) as the inference space associated with the joint (τ_1, τ_2) parameter space. We will motivate the conditioning event via the pivotal ratio $W = U_2/U_1$, and then translate it back to this joint pivotal space.

The ratio of the $U's$ is expressible as

$$W = \frac{U_2}{U_1} = \frac{T_2/T_1}{\tau_2/\tau_1} = \frac{S}{\sigma}. \tag{2.3}$$

Since $\tau_2 \geq \tau_1$ by the model assumptions, it follows that $\sigma = \tau_2/\tau_1 \geq 1$. Therefore, one can assert *pre-data* that $W \leq S$, where S can be interpreted as a random upper bound on W. When a realization s of S becomes manifest, this additional information is used to specify a particular region $W \leq s$, or equivalently, $U_2 \leq sU_1$, in the joint (U_1, U_2) pivotal space (see Figure 2.2). The resulting region is used as the reference set for PDP inference in the balanced one-way random model; i. e.,

$$C(s) = \{U_2 \leq sU_1\} = \{W \leq s\} \tag{2.4}$$

is the *conditioning event* associated with model (2.1). As s increases, more and more of the unconditional joint pivotal space is covered by the conditioning event; conversely, as $s \to 0$, the conditioning event converges to the line $U_2 = 0$, corresponding to pure pooling of information in both sums of squares towards inference about τ_1.

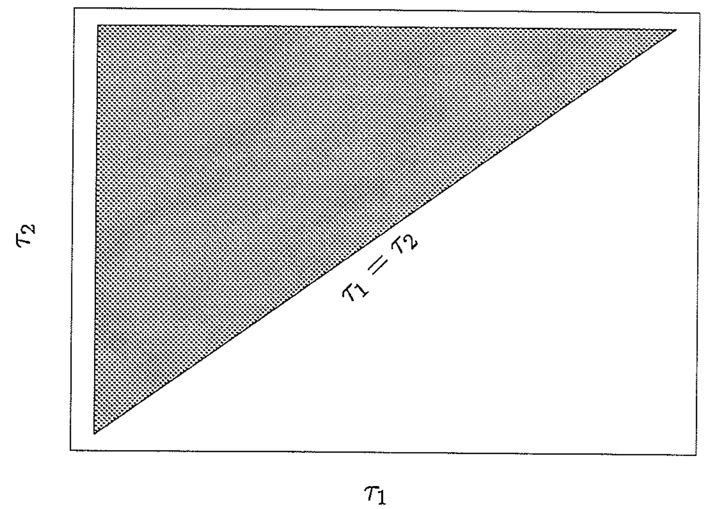

Figure 2.1: Constrained (τ_1, τ_2) parameter space: the shaded area corresponds to the parameter constraint $\tau_2 \geq \tau_1$.

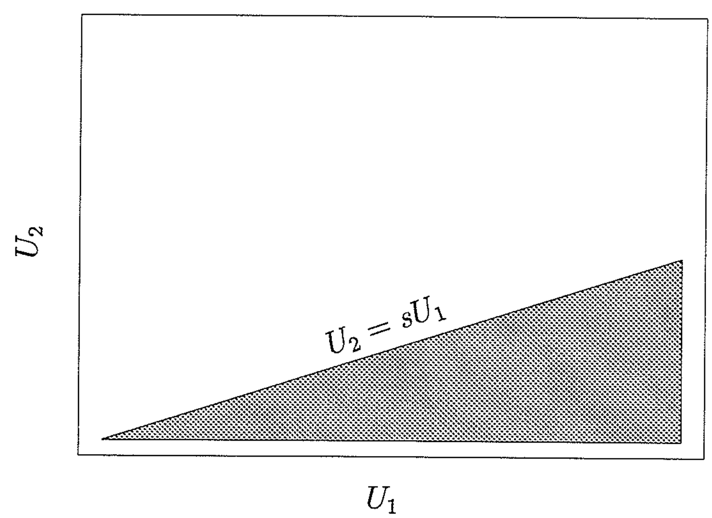

Figure 2.2: Constrained (U_1, U_2) pivotal space: the shaded area corresponds to the pivotal constraint $U_2 \leq sU_1$.

The constraint on σ implies that its marginal parameter space is $[1, \infty)$, with realized reference set $[0, s]$ on the pivotal space of W.

The next step is to apply the conditioning event $C(s)$ to each of the basic pivotals U_1, U_2 and W in the balanced one-way random model. The effect of such conditioning is to modify the distributions of the basic pivotals as a function of s. To clarify the notation, we define, for any fixed $s > 0$,

$$U_1^*(s) = U_1|W \leq s$$
$$U_2^*(s) = U_2|W \leq s$$
$$W^*(s) = W|W \leq s$$

as the 'post-data pivotals' of the model, or PDP's for short. In general, the argument s will be dropped; the starred superscript will implicitly denote dependence on s. The term 'post-data pivotal' is a bit of a misnomer, since the basic pivotals themselves are not modified. However, the *distributions* of the basic pivotals are modified by the conditioning event $W \leq s$; we will call these 'PDP distributions'. The notation $g^*(u_i|s)$ denotes the (conditional) density of the PDP $U_i^* = U_i|W \leq s$, and $h^*(w|s)$ denotes the PDP density of W^*.

The PDP argument is very simple: replace the 'usual' pivotal distributions with their PDP counterparts, and conduct inference in the same manner as before. The benefits are that point and interval estimates will always lie in the appropriate parameter space, while point and interval estimators exhibit good frequentist performance in addition to good conditional behavior.

Thus, what we are proposing is an approach to inference that starts with procedures having good initial precision, and modifying them (by steps 1–4 on p. 39) so that they have good final precision. The key modification is replacement of the usual pivotal distributions with PDP analogues.

2.4 PDP distribution theory

Frequentist point and interval estimation under model (2.1) is founded on the distribution theory of the basic pivotals U_1, U_2 and W. Under the PDP approach, the basic pivotals are replaced with the corresponding post-data pivotals U_1^*, U_2^* and W^* given any fixed $s > 0$. Inference then proceeds by using the PDP distributions as the underlying reference for conditional point and interval estimation of the parametric functions of interest. We begin by considering the joint density of U_1 and U_2 conditioned on $W \leq s$, from which the marginal PDP densities of U_1^*, U_2^* and W^* are easily derived for any fixed s. Properties of the resulting PDP distributions are then investigated.

It should be noted that the minimal sufficient statistic in this problem is $(\overline{Y}, T_1, T_2)$. However, we have decided to restrict attention to T_1 and T_2 because interest lies only in τ_1 and τ_2. On one hand, we are not, strictly speaking, justified in doing so; this is at the root of the decision theoretic results of Stein (1964), Brown (1968) and Brewster & Zidek (1974) in fixed effects models. On the other hand, it simplifies the situation if we ignore the 'fixed effects part' of the data, somewhat analogous to using the restricted (or residual) likelihood rather than the full likelihood.

Under the unconditional one-way random effects model (2.1), it is assumed that $U_1 \sim \chi^2_{\nu_1}$ and $U_2 \sim \chi^2_{\nu_2}$ are independently distributed random variables, where $U_i = T_i/\tau_i$, $i = 1, 2$. The joint density of (U_1, U_2) is then

$$\begin{aligned} g(u_1, u_2) &= g_{\nu_1}(u_1)\, g_{\nu_2}(u_2) \\ &= \frac{u_1^{\phi_1 - 1}\, u_2^{\phi_2 - 1}}{\Gamma(\phi_1)\, \Gamma(\phi_2)\, 2^{\phi_1 + \phi_2}} \exp\{-\tfrac{1}{2}(u_1 + u_2)\} \end{aligned}$$

where $u_1 > 0$ and $u_2 > 0$, respectively, with $\phi_1 = \nu_1/2$, $\phi_2 = \nu_2/2$.

To find the joint PDP density of U_1 and U_2 , observe that the conditioning event

$W \leq s$ is equivalent to $U_2 \leq s\, U_1$. Therefore, for each fixed $s > 0$,

$$g^*(u_1, u_2|s) = \frac{g_{\nu_1}(u_1)\, g_{\nu_2}(u_2)}{\int_0^\infty \int_0^{su_1} g_{\nu_1}(u_1)\, g_{\nu_2}(u_2)\, du_2\, du_1},$$

defined on the region $\{u_1 > 0, 0 \leq u_2 \leq su_1\}$, where

$$\begin{aligned}\int_0^\infty \int_0^{su_1} g_{\nu_1}(u_1)\, g_{\nu_2}(u_2)\, du_2\, du_1 &= \int_0^\infty g_{\nu_1}(u_1) \left[\int_0^{su_1} g_{\nu_2}(u_2)\, du_2 \right] du_1 \\ &= \int_0^\infty g_{\nu_1}(u_1)\, G_{\nu_2}(s\, u_1)\, du_1 \\ &= I_z(\phi_2, \phi_1)\end{aligned}$$

by Lemma A.24, with $z = s/(1+s)$ and $\phi_i = \nu_i/2$, $i = 1, 2$. Thus, the joint PDP density of U_1 and U_2 is

$$g^*(u_1, u_2|s) = \frac{g_{\nu_1}(u_1)\, g_{\nu_2}(u_2)}{I_z(\phi_2, \phi_1)}, \tag{2.5}$$

where $u_1 \geq 0$ and $0 \leq u_2 \leq su_1$. For each fixed (τ_1, τ_2) pair, and fixed degrees of freedom (ν_2, ν_1), this defines a family of joint PDP densities indexed by s.

The joint PDP density (2.5) induces a visual 'partial eclipse' over the corresponding unconditional joint density, the extent of the 'eclipse' depending on the size of s. Smaller values of s induce stronger eclipses; conversely, as s gets larger, less of the unconditional joint density is being excluded by the conditioning event. Thus, as s increases, the joint PDP density converges to the unconditional joint density. Marginal PDP densities exhibit similar asymptotic behavior, which will be reflected in the PDP estimation procedures. Figure 2.3 illustrates this phenomenon for an example with $(\nu_2, \nu_1) = (3, 4)$ degrees of freedom and various choices of s. Figure 2.4 is a contour plot of the joint PDP density of U_1 and U_2 , with the PDP and (asymptotic) unconditional densities of U_1 above the contour plot, and the PDP and unconditional densities of U_2 to the right. The serrated edge of the joint density contour plot is an artifice, a product of sampling a discrete grid of points from a continuous joint density.

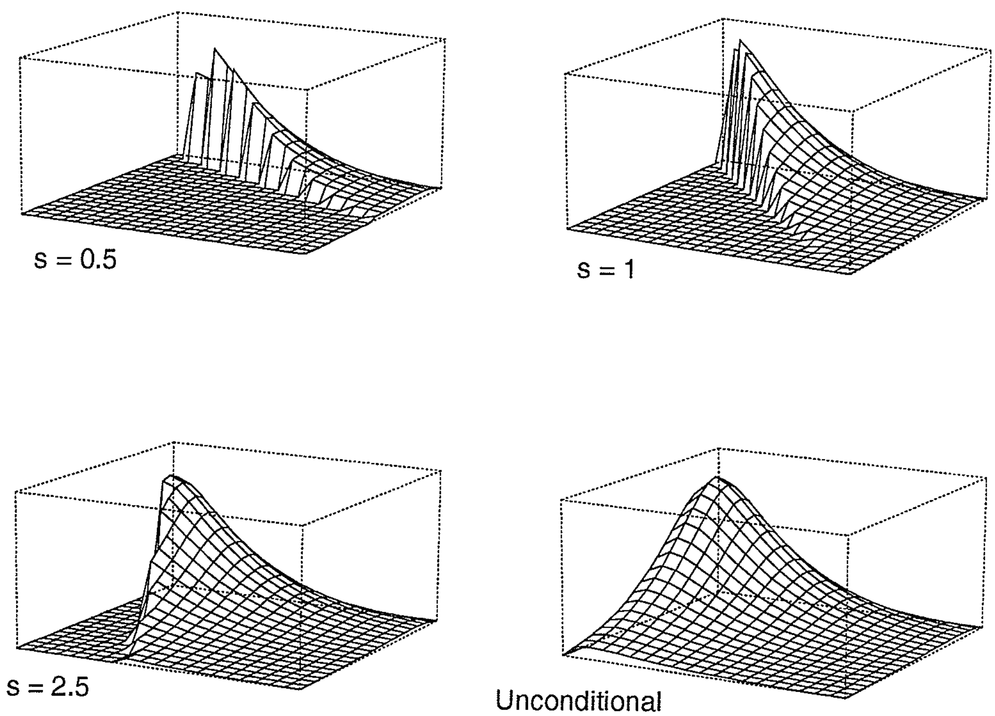

Figure 2.3: 3D-plots of the joint PDP density of U_1 and U_2 for various choices of s, with $(\nu_2, \nu_1) = (3, 4)$. The U_1 axis runs from left to right at the front of each box, while the U_2 axis runs from back to front on the side of each box. As $s \to 0$, the joint PDP density becomes degenerate at $(0, 0)$; as $s \to \infty$, the PDP joint density converges to the unconditional joint density. Note that the graphs are not on the same vertical scale—as s decreases, the height of the peak increases, as the density is apportioned over a smaller wedge.

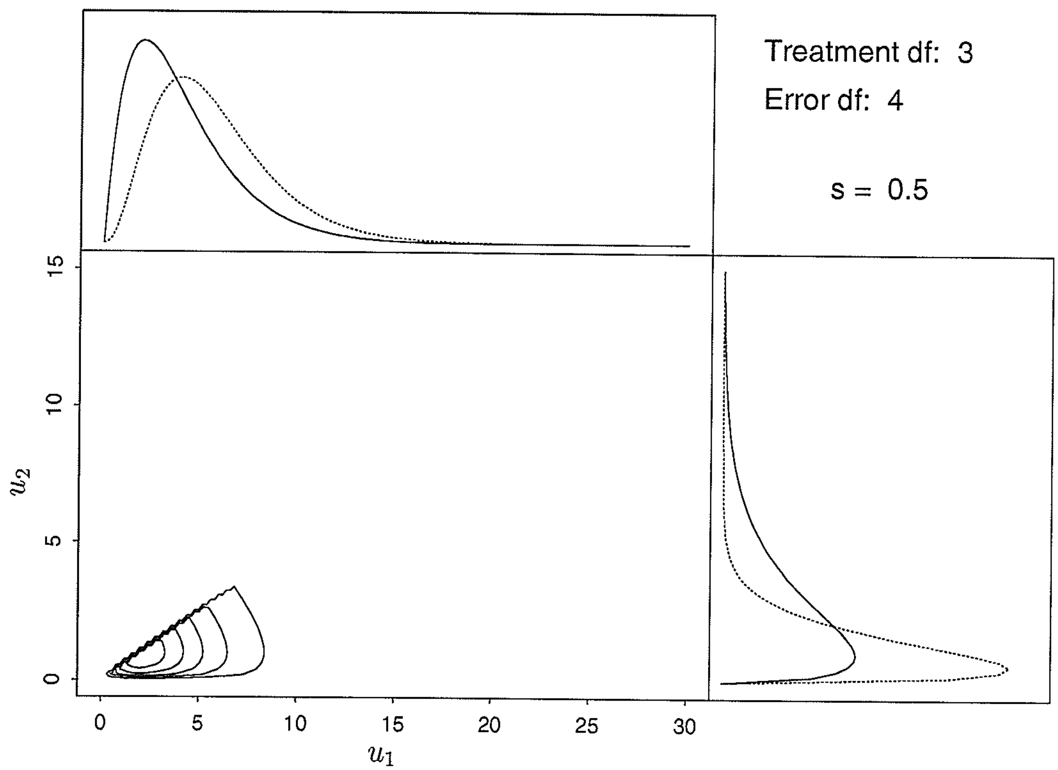

Figure 2.4: Contour plot of the PDP joint density of U_1 and U_2 with accompanying PDP and unconditional marginals of U_1 (top) and U_2 (right, rotated 90 degrees), with $(\nu_2, \nu_1) = (3, 4)$. Unconditional marginal densities correspond to the solid curve.

Marginal density of U_1^*. Integrating (2.5) with respect to U_2,

$$g^*(u_1|s) = \frac{\int_0^{su_1} g_{\nu_1}(u_1)\, g_{\nu_2}(u_2)\, du_2}{I_z(\phi_2, \phi_1)} = \frac{g_{\nu_1}(u_1) \int_0^{su_1} g_{\nu_2}(u_2)\, du_2}{I_z(\phi_2, \phi_1)}.$$

Hence, given any fixed $s > 0$,

$$g^*(u_1|s) = \frac{g_{\nu_1}(u_1)\, G_{\nu_2}(s\, u_1)}{I_z(\phi_2, \phi_1)}, \tag{2.6}$$

for $u_1 > 0$. For each fixed τ_1, a family of U_1^* densities of the form (2.6) is generated as a function of s.

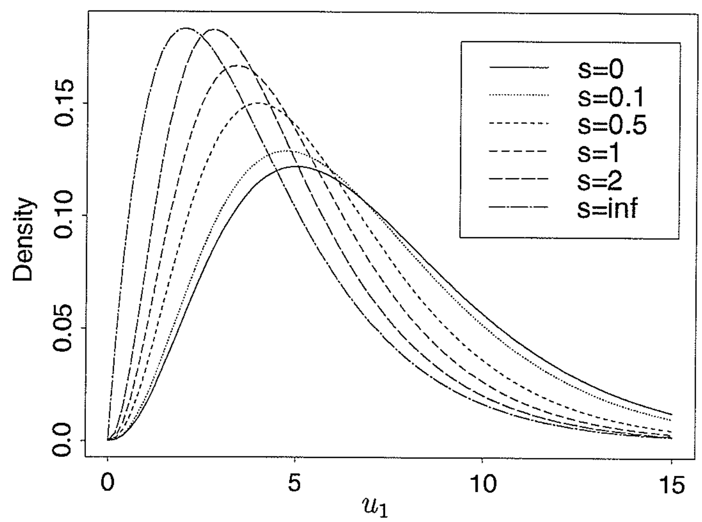

Figure 2.5: Graphs of the marginal density of U_1^* for various choices of s (see legend), with $(\nu_2, \nu_1) = (3, 4)$. As $s \to 0$, the PDP density approaches a χ_7^2; as $s \to \infty$, the limiting density is the 'usual' χ_4^2. Visually, as s increases, the density shifts to the left, corresponding to less pooling of information from treatments.

Functions of this form will be termed 'shifted-χ^2' densities. The 'shifting' is driven by $G_{\nu_2}(s\, u_1)/I_z(\phi_2, \phi_1)$, the ratio of a chi-square cdf to a Beta cdf, which represents the contribution of the treatment SS to the estimation of the error variance $\tau_1 = \tau_1$. This cdf ratio has an interesting frequentist interpretation (Box & Tiao, 1973, p. 261), which will be taken up later in section 2.4.9. Its effect with respect to the conditional density (2.6) is to 'shift' the density of U_1 somewhere between a $\chi^2_{\nu_1}$ and a $\chi^2_{\nu_1+\nu_2}$ density, where the amount of 'shift' depends on the observed sum of squares ratio s. See Figure 2.5 for an illustration in the case where $(\nu_2, \nu_1) = (3, 4)$.

The density of U_1^* can be viewed as a sampling theory analogue of Box and Tiao's posterior density of τ_1 for any fixed $s > 0$, and by extension, to the fiducial density of τ_1 (Wild, 1981). These connections are addressed in section 2.4.9.

Marginal density of U_2^*. To obtain the marginal density of $U_2^*(s)$ for any fixed $s > 0$, integrate the joint conditional density (2.5) with respect to U_1. In this case, the conditioning event $W \leq s$ is equivalent to $U_1 \geq U_2/s$, so that

$$g^*(u_2|s) = \frac{\int_{u_2/s}^{\infty} g_{\nu_1}(u_1)\, g_{\nu_2}(u_2)\, du_1}{I_z(\phi_2, \phi_1)} = \frac{g_{\nu_2}(u_2)}{I_z(\phi_2, \phi_1)} \int_{u_2/s}^{\infty} g_{\nu_1}(u_1)\, du_1,$$

or

$$g^*(u_2|s) = g_{\nu_2}(u_2)\, \frac{[1 - G_{\nu_1}(u_2/s)]}{I_z(\phi_2, \phi_1)}. \tag{2.7}$$

The CDF ratio in (2.7) 'shifts' the conditional density somewhere between a $\chi^2_{\nu_2}$ and a degenerate distribution at zero, depending on the size of s. For each value of τ_2, a family of U_2^* distributions is produced as a function of s. See Figure 2.6 for an illustration in the case where $(\nu_2, \nu_1) = (3, 4)$.

Marginal density of W^*. For any fixed $s > 0$, the marginal PDP density of W^* is derived as follows. Starting from the joint PDP density (2.5), let

$$W = \frac{U_2}{U_1}, \qquad T = U_1,$$

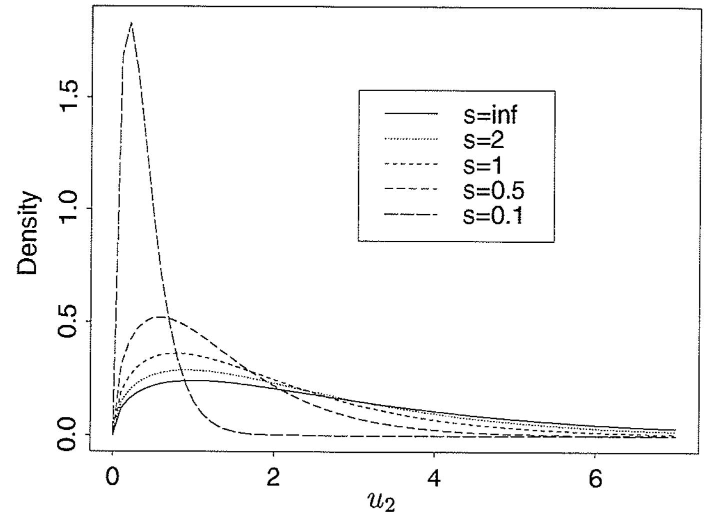

Figure 2.6: Graphs of the density of U_2^* for various choices of s (see legend), with $(\nu_2, \nu_1) = (3, 4)$. As $s \to 0$, the PDP density becomes degenerate at 0; as $s \to \infty$, the limiting density is the 'usual' χ_3^2. As s decreases, the density shifts to the left.

with back transform

$$U_2 = WT, \qquad U_1 = T,$$

and Jacobian T. The bivariate mapping $(U_1, U_2) \longmapsto (W, T)$ transforms the area of support from the triangular region $\{U_1 > 0, 0 < U_2 \leq sU_1\}$ to the rectangular region $\{T > 0, 0 < W < s\}$. The joint PDP density of W and T is then

$$g^*(w, t|s) = \frac{1}{\Gamma(\phi_1)\,\Gamma(\phi_2) 2^{\phi_1+\phi_2}} \frac{w^{\phi_2-1}\, t^{\phi_1+\phi_2-1}\, e^{-t(1+w)/2}}{I_z(\phi_2, \phi_1)}.$$

To get the marginal density of W^*, rearrange some terms and integrate over T:

$$g^*(w|s) = \frac{\Gamma(\phi_1+\phi_2)}{\Gamma(\phi_1)\Gamma(\phi_2)} \frac{w^{\phi_2-1}}{I_z(\phi_2, \phi_1)} \int_0^\infty \frac{t^{\phi_1+\phi_2-1}\, e^{-t(1+w)/2}}{2^{\phi_1+\phi_2}\Gamma(\phi_1+\phi_2)}\, dt.$$

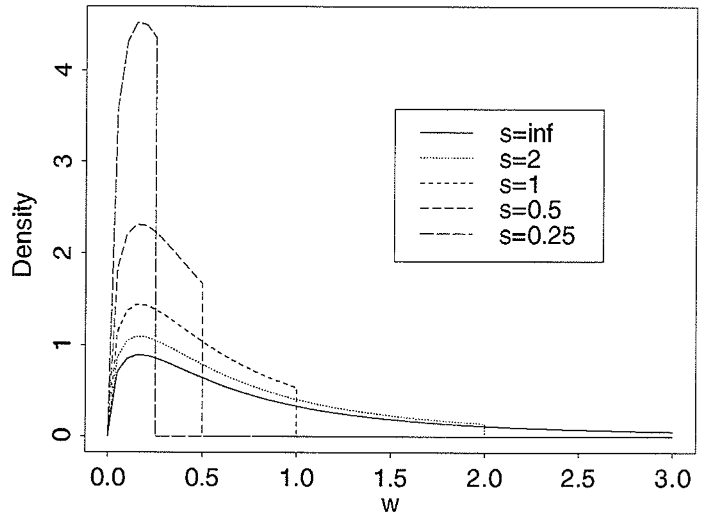

Figure 2.7: Graphs of the density of W^* for various choices of truncation points s (see legend), with $(\nu_2, \nu_1) = (3, 4)$. As $s \to 0$, the PDP density becomes degenerate at 0; as $s \to \infty$, the limiting density is the unconditional Beta$_{II}$ density of W.

By making the transformation $v = t(1+w)$ inside the above integral and simplifying, we get

$$g^*(w|s) = \frac{\Gamma(\phi_1 + \phi_2)}{\Gamma(\phi_1)\Gamma(\phi_2)} \frac{w^{\phi_2 - 1}(1+w)^{-(\phi_1+\phi_2)}}{I_z(\phi_2, \phi_1)} \chi_{[0,s]}(w), \tag{2.8}$$

a truncated Beta type II density with parameters ϕ_2 and ϕ_1, defined on the interval $[0, s]$ (See Figure 2.7).

2.4.1 Properties of the PDP distributions

Using results from Appendix A, the following properties are established for each of the PDP distributions derived above:

- moment generating functions of U_1^* and U_2^*;

- the first two (central) moments;
- the limiting moment generating functions of U_1^* and U_2^*;
- the cumulative distribution function; and
- the monotone likelihood ratio property.

2.4.1.1 Properties of U_1^*.

Derivation of the moments. Using Lemmas A.23 and A.25 and Theorem A.26, the moment generating function of U_1^* is, for any fixed $s > 0$,

$$M_{U_1^*}(t|s) = \frac{(1-2t)^{-\phi_1}\, I_{z^*}(\phi_2, \phi_1)}{I_z(\phi_2, \phi_1)}, \tag{2.9}$$

where $z^* = s/(1-2t+s)$ and $|t| < 1/2$. The mgf of U_1^* differs from that of U_1 by the ratio of Beta cdfs in (A.24). The moments of U_1^* can be derived from Theorem A.28 or by successively differentiating the mgf with respect to t and evaluating at $t = 0$. In particular, for any fixed $s > 0$,

$$E(U_1^*) = \nu_1 \frac{I_z(\phi_2, \phi_1 + 1)}{I_z(\phi_2, \phi_1)}, \tag{2.10}$$

$$E(U_1^{*2}) = \nu_1 (\nu_1 + 2) \frac{I_z(\phi_2, \phi_1 + 2)}{I_z(\phi_2, \phi_1)}, \tag{2.11}$$

by Corollary A.28.1, which implies that

$$\text{Var}(U_1^*) = \nu_1 \frac{I_z(\phi_2, \phi_1 + 1)}{I_z(\phi_2, \phi_1)} \left[(\nu_1 + 2) \frac{I_z(\phi_2, \phi_1 + 2)}{I_z(\phi_2, \phi_1 + 1)} - \nu_1 \frac{I_z(\phi_2, \phi_1 + 1)}{I_z(\phi_2, \phi_1)} \right]. \tag{2.12}$$

By Corollaries A.2.1 and A.28.2, the mean and variance of U_1^* can be simplified to

$$E(U_1^*) = \nu_1 + 2\,\delta_z(\phi_2, \phi_1) \tag{2.13}$$

$$\text{Var}(U_1^*) = 2\left[(\nu_1 + 2\delta_z(\phi_2, \phi_1)) + \psi_z(\phi_2, \phi_1)\right], \tag{2.14}$$

where

$$\delta_z(\phi_2, \phi_1) = \frac{\Gamma(\phi_1 + \phi_2)}{\Gamma(\phi_2)\,\Gamma(\phi_1)} \frac{z^{\phi_2}\,(1-z)^{\phi_1}}{I_z(\phi_2, \phi_1)} \tag{2.15}$$

and

$$\psi_z(\phi_2, \phi_1) = \delta\,\left[(1-z)\,(\nu_2 - 2\delta) - z\,(\nu_1 + 2\delta)\right], \tag{2.16}$$

with $\delta = \delta_z(\phi_2, \phi_1)$ in (2.16). Since the $\delta(\cdot)$ and $\psi(\cdot)$ terms affect the moments of both U_1^* and U_2^*, further discussion follows in section 2.4.3.

Limiting m.g.f. From the moment generating function (2.9), the limiting behavior of U_1^* as $s \to 0$ and as $s \to \infty$ is of interest. Theorem A.27 shows that $U_1^* \xrightarrow{d} \chi^2_{\nu_1}$ as $s \to \infty$; conversely, as $s \to 0$, $U_1^* \xrightarrow{d} \chi^2_{\nu_1+\nu_2}$. In other words, as s gets large, the PDP distribution converges to the unconditional distribution of U_1, while as s gets small, it converges to the distribution of $U_1 + U_2$ with $\nu_1 + \nu_2$ degrees of freedom, the 'purely pooled' distribution.

Cumulative distribution function. The cdf of U_1^* at some value $x \in (0, \infty)$, for any fixed $s > 0$, is given by

$$\begin{aligned} G_{U_1^*}(x|s) &= \int_0^x g_{U_1^*}(u_1|s)\,du_1 \\ &= \frac{\int_0^x g_{\nu_1}(u_1)\,G_{\nu_2}(su_1)\,du_1}{I_z(\phi_2, \phi_1)}. \end{aligned}$$

Since the integral has no closed form solution, it requires numerical evaluation. A computer algebra system such as Maple or Mathematica can be used for this task.

Monotone likelihood ratio property. A necessary condition for the validity of certain results involving point and interval estimation in this chapter is the monotone likelihood ratio property (MLRP). An important consequence of this property is that any family of distributions indexed by a parameter θ with the MLRP is stochastically ordered in θ (Lehmann, 1986).

We are interested in establishing the following MLR properties with respect to the family of $U_1|S \leq r$ distributions. This construction is necessary in order to express the MLRP in terms of the parametric function σ. Noting that $S \leq r$ is equivalent to $W \leq r/\sigma$, we essentially substitute r/σ for s in the PDP density of U_1 to obtain

$$g^*(u_1|r,\sigma) = \frac{1}{\sigma}\frac{g_{\nu_1}(u_1)\, G_{\nu_2}(ru_1/\sigma)}{I_{z^*}(\phi_2, \phi_1)},$$

where $z^* = r/(\sigma + r)$. Theorem A.14 establishes the following results:

(i) for fixed $r > 0$, $U_1|S \leq r$ has increasing MLR in σ;

(ii) for fixed $\sigma = 1$, $U_1|S \leq r$ has decreasing MLR in r;

(iii) for fixed $\sigma = 1$, the density ratio $g_{U_1|S\leq r}(u_1|r)/g_{\nu_1}(u_1)$ is increasing in u_1.

2.4.1.2 Properties of U_2^*

Generation of the moments. By Theorem A.31, the m.g.f. of U_2^* is, for fixed $s > 0$,

$$M_{U_2^*}(t|s) = (1-2t)^{-\phi_2}\,\frac{I_{z^*}(\phi_2, \phi_1)}{I_z(\phi_2, \phi_1)}, \tag{2.17}$$

where $s^* = (1-2t)s$ and $z^* = s^*/(1+s^*)$. By Theorem A.33, or by successively differentiating (2.17) with respect to t and evaluating at $t = 0$, the first two moments of $U_2^*(s)$ are given by

$$E(U_2^*(s)) = \nu_2\,\frac{I_z(\phi_2+1, \phi_1)}{I_z(\phi_2, \phi_1)}, \tag{2.18}$$

$$E(U_2^{*\,2}(s)) = \nu_2(\nu_2+2)\,\frac{I_z(\phi_2+2, \phi_1)}{I_z(\phi_2, \phi_1)}, \tag{2.19}$$

for any fixed $s > 0$. Hence,

$$\mathrm{Var}(U_2^*) = \nu_2 \frac{I_z(\phi_2+1,\phi_1)}{I_z(\phi_2,\phi_1)} \left[(\nu_2+2) \frac{I_z(\phi_2+2,\phi_1)}{I_z(\phi_2+1,\phi_1)} - \nu_2 \frac{I_z(\phi_2+1,\phi_1)}{I_z(\phi_2,\phi_1)} \right]. \qquad (2.20)$$

By Corollaries A.3.1, A.33.1 and A.33.2, the mean and variance of U_2^* reduce to

$$E(U_2^*(s)) = \nu_2 - 2\,\delta_z(\phi_2,\phi_1) \qquad (2.21)$$

$$\mathrm{Var}(U_2^*(s)) = 2\,\left[(\nu_2 - 2\,\delta_z(\phi_2,\phi_1)) + \psi_z(\phi_2,\phi_1)\right], \qquad (2.22)$$

where $\delta(\cdot)$ and $\psi(\cdot)$ are defined by (2.15) and (2.16), respectively.

Limiting m.g.f. By Theorem A.32, we have that (a) $U_2^* \xrightarrow{d} \chi^2_{\nu_2}$ as $s \to \infty$, and (b) as $s \to 0$, the limiting distribution of U_2^* becomes degenerate at 0.

Cumulative distribution function of U_2^*. The cdf of U_2^* at some value $x \in (0,\infty)$ is given by

$$\begin{aligned} G_{U_2^*}(x) &= \int_0^x g^*(u_2|s)\,du_2 \\ &= \frac{\int_0^x g_{\nu_2}(u_2)\,[1 - G_{\nu_1}(u_2/s)]\,du_2}{I_z(\phi_2,\phi_1)} \end{aligned}$$

for any fixed $s > 0$. This integral has no closed form solution (as was the case for the cdf of U_1^*), so it too must be evaluated numerically.

Monotone likelihood ratio property. Analogous to $U_1|S \le r$, the following MLR properties for $U_2|S \le r$ are established by Theorem A.14:

(i) $U_2|S \le r$ has the decreasing MLRP in σ for fixed $r > 0$;

(ii) for fixed $\sigma = 1$, $U_2|S \le r$ has the increasing MLRP in r;

(iii) the density ratio $g_{U_2|S\le r}(u_2|r)/g_{\nu_2}(u_2)$ is decreasing in u_2 for fixed $\sigma = 1$ and any $r > 0$.

2.4.2 Nonindependence of U_1^* and U_2^*

Unconditionally, U_1 and U_2 are independently distributed χ^2 random variables. However, when one conditions on the set $W \leq s$, the resulting PDPs $U_1^*(s)$ and $U_2^*(s)$ covary for any fixed $s > 0$, as is evident from the non-rectangular support of the joint density. By the definition of covariance,

$$\mathrm{cov}(U_1, U_2|W \leq s) = E(U_1 U_2|W \leq s) - E(U_1^*(s))\, E(U_2^*(s)),$$

where $E(U_1^*) = \nu_1 + 2\delta$ and $E(U_2^*) = \nu_2 - 2\delta$ by Corollaries A.2.1 and A.3.1, respectively, $\delta = \delta_z(\phi_2, \phi_1)$ is defined by (2.15), and $E(U_1 U_2|W \leq s)$ is

$$E(U_1 U_2|W \leq s) = \nu_1\, \nu_2\, \frac{I_z(\phi_2 + 1, \phi_1 + 1)}{I_z(\phi_2, \phi_1)}, \tag{2.23}$$

by Theorem A.35. Consequently, the conditional covariance is

$$\mathrm{cov}(U_1, U_2|W \leq s) = \nu_1\, \nu_2 \left[\frac{I_z(\phi_2 + 1, \phi_1 + 1)}{I_z(\phi_2, \phi_1)} - \frac{I_z(\phi_2, \phi_1 + 1)}{I_z(\phi_2, \phi_1)}\, \frac{I_z(\phi_2 + 1, \phi_1)}{I_z(\phi_2, \phi_1)}\right]. \tag{2.24}$$

It is of interest to simplify this expression in much the same way as was done for the means of U_1^* and U_2^*. In fact, the covariance can be reduced to

$$\mathrm{cov}(U_1, U_2|W \leq s) = -2\psi_z(\phi_2, \phi_1) \tag{2.25}$$

for any fixed $s > 0$, by Theorem A.36.

The limiting behavior of (2.25) is rather interesting. As $s \to \infty$, the asymptotic covariance is zero since $\delta \xrightarrow{s \to \infty} 0$. This corresponds to the fact that U_1^* and U_2^* tend to statistical independence as s gets large. Conversely, as $s \to 0$, $\psi \to 0$ since $\delta \to \phi_2$. In this case, since the distribution of U_2^* becomes degenerate as $s \to 0$, the limiting covariance is essentially that of a random variable and a constant, which is zero by definition. In between these extremes, the covariance between U_1^* and U_2^* is positive, since $\psi_z(\phi_2, \phi_1)$ is nonpositive over $(0, \infty)$ by Lemma A.7 (see Figure 2.8).

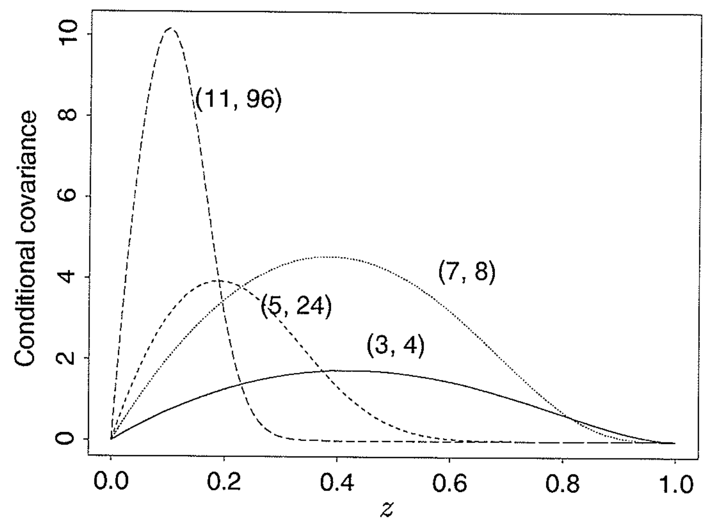

Figure 2.8: Graphs of the covariance of U_1^* and U_2^* as a function of z, for several degree of freedom pairs.

Now, consider the correlation between U_1^* and U_2^*, defined as

$$\begin{aligned}\text{corr}(U_1, U_2|W \leq s) &= \frac{\text{cov}(U_1, U_2|W \leq s)}{\left[\text{Var}(U_1^*(s))\ \text{Var}(U_2^*(s))\right]^{1/2}} \\ &= -\psi\left[(\nu_1 + 2\delta + \psi)(\nu_2 - 2\delta + \psi)\right]^{-1/2}.\end{aligned}$$

Figures 2.9 and 2.10 show plots of the conditional correlation of U_1^* and U_2^* as a function of z. Each indicates an ordering of the correlation functions when one of ν_1 or ν_2 is fixed. Figure 2.9 shows that when ν_1 is fixed, an increase in the number of treatment levels increases the correlation between U_1^* and U_2^* for fixed z. Conversely, by fixing the treatment df ν_2, the correlation decreases as the number of replicates per cell increases (or equivalently, as ν_1 increases). See Figure 2.10. Furthermore, as s (or z) increases, the correlation between U_1^* and U_2^* monotonically decreases to zero in all cases, as one would expect. In a sense, the correlation function describes the extent that U_1^* and U_2^* rely on one another to effect 'smooth pooling' in PDP estimation procedures for any fixed $s > 0$.

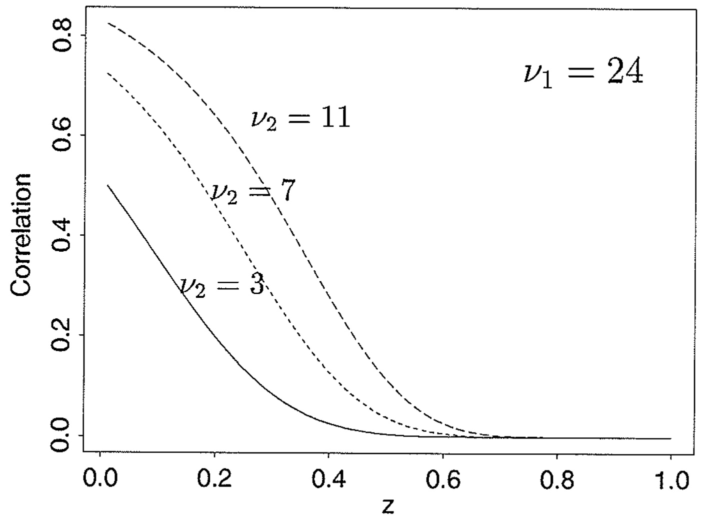

Figure 2.9: Graph of the correlation function for fixed $\nu_1 = 24$ and various choices of ν_2.

2.4.3 Interpretation of $\delta(\cdot)$ and $\psi(\cdot)$

In the above discussion, it was found that $E(U_1^*(s)) = \nu_1 + 2\delta_z(\phi_2, \phi_1)$ for any fixed $s > 0$, where $\delta(\cdot)$ is defined by (2.15). The function $\delta(\cdot)$ can be interpreted as a 'smoothing function' for the PDP-modified degrees of freedom of a one-way random model. By Theorem A.4, as $z = s/(1+s)$ tends to zero, 2δ tends to ν_2; as z tends to 1, δ tends to zero. In other words, δ performs the post-data adjustment of error and treatment degrees of freedom, operationalizing the 'smooth pooling' of degrees of freedom between treatment and error. As a function of z, $\delta(\cdot)$ is continuous and monotonically decreasing.

The term $\psi(\cdot)$ appears in the variances of U_1^* and U_2^*, as well as their covariance.

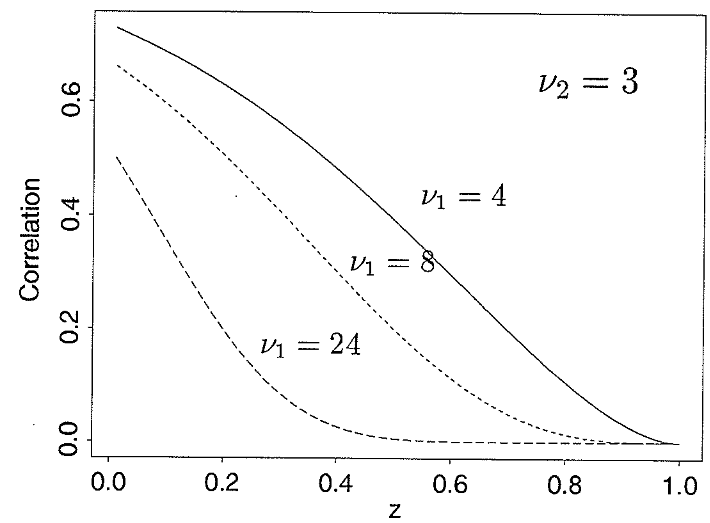

Figure 2.10: Graph of the correlation function for fixed $\nu_2 = 3$ with varying ν_1.

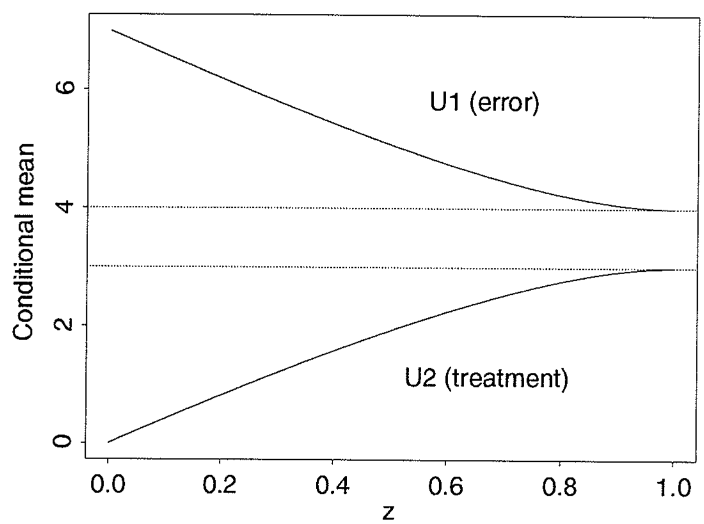

Figure 2.11: Graphs of the conditional means of U_1 and U_2 when $(\nu_1, \nu_2) = (3, 4)$. This graph illustrates: (i) that the sum of the conditional means preserves the total df for all z, and (ii) the duality in behavior between the two PDPs.

Observe that $\psi(\cdot)$ is strictly a function of $z = s/(1+s)$, since the degrees of freedom are fixed post-data. However, different choices of ϕ_1 and ϕ_2 lead to different $\psi(\cdot)$ functions. Lemma A.6 shows that the partial derivative of $E(U_1^*)$ with respect to z is $\psi_z(\cdot)/z(1-z)$; hence, the rate of change of the conditional mean is functionally dependent on $\psi_z(\cdot)$.

2.4.4 Comparison of distributions of U_1^* and U_2^*

There is clearly a duality in the behavior of the distributions of U_1^* and U_2^*. The tradeoff is manifested through the identity

$$E(U_1^*) + E(U_2^*) = (\nu_1 + 2\delta) + (\nu_2 - 2\delta) = \nu_1 + \nu_2,$$

which states that the 'post-data adjusted' degrees of freedom preserve the total df. Figure 2.11 illustrates this for the case where $(\nu_2, \nu_1) = (3, 4)$. There is a consistency among the limiting conditional means and limiting moment generating functions: the limiting conditional means are exactly the degrees of freedom of the asymptotic distributions both as $s \to 0$ and as $s \to \infty$. Furthermore, nonzero conditional correlation between U_1^* and U_2^* is a necessary condition for the 'smooth pooling' of degrees of freedom post-data, for without it, there is no way to 'borrow' degrees of freedom from treatment and reallocate it into the error term. The function $\psi(\cdot)$ is common to the conditional variances of U_1^* and U_2^*, their conditional covariance, and the rate of change of $\delta(\cdot)$ with respect to z.

The distributions of U_1^* and U_2^* modify the pre-data distributions of U_1 and U_2 in light of the parameter constraint $\sigma \geq 1$ and the observed data through the observed sum of squares ratio s. In effect, U_1^* and U_2^* 'update' U_1 and U_2 in the presence of this additional information, analogous to a Bayesian posterior distribution. Furthermore, the families of U_1^* and U_2^* densities can be anticipated prior to the experiment, although the actual choice of distributions is only possible post-data.

2.4.5 Properties of W^*.

Unlike the densities of U_1^* and U_2^*, which 'shift' the usual unconditional χ^2 densities as a function of s, the density of W^* is truncated on the right at s. In a sense, the truncated density is a compromise between (i) the 'push and pull' of the densities of U_1^* and U_2^*, and (ii) fealty to the parameter constraint. Proofs of the results given below are provided in section A.5.

Moments of W^*. By definition, the moment generating function of $W^*(s)$ for each fixed, finite $s > 0$ is given by

$$M_{W^*}(t|s) = \int_0^s \frac{\Gamma(\phi_1+\phi_2)}{\Gamma(\phi_1)\,\Gamma(\phi_2)} \frac{e^{tx}\, x^{\phi_2-1}\,(1+x)^{-(\phi_1+\phi_2)}}{I_z(\phi_2,\phi_1)}\,dx. \tag{2.26}$$

However, this expression is not particularly easy to work with. Fortunately, the first two moments of $W^*(s)$ are easily found for any fixed s by Theorem A.37:

$$E(W^*) = \frac{\nu_2}{\nu_1-2}\,\frac{I_z(\phi_2+1,\phi_1-1)}{I_z(\phi_2,\phi_1)}, \qquad \nu_1 > 2, \tag{2.27}$$

$$E(W^2|W\le s) = \frac{\nu_2\,(\nu_2+2)}{(\nu_1-2)(\nu_1-4)}\,\frac{I_z(\phi_2+2,\phi_1-2)}{I_z(\phi_2,\phi_1)}, \qquad \nu_1 > 4. \tag{2.28}$$

The first two moments of W^* differ from the corresponding unconditional moments by the Beta cdf ratios.

For $\nu_1 > 4$, it follows by the definition of variance and Theorem A.37 that

$$\mathrm{Var}(W^*) = \frac{\nu_2}{(\nu_1-2)}\,\frac{I_z(\phi_2+1,\phi_1-1)}{I_z(\phi_2,\phi_1)} \times \left[\frac{\nu_2+2}{\nu_1-4}\,\frac{I_z(\phi_2+2,\phi_1-2)}{I_z(\phi_2+1,\phi_1-1)} - \frac{\nu_2}{\nu_1-2}\,\frac{I_z(\phi_2+1,\phi_1-1)}{I_z(\phi_2,\phi_1)}\right]. \tag{2.29}$$

Limiting distributions. As $s \to 0$, it is easily seen that the distribution of W^* becomes degenerate at zero in the limit; conversely, as $s \to \infty$, it converges to the unconditional Beta type II distribution of W with parameters ϕ_2 and ϕ_1, respectively.

Cumulative distribution function. Unlike U_1^* and U_2^*, the c.d.f. of W^* has a simple form. For any fixed $s > 0$ and $x \in (0, s]$,

$$H^*(x|s) = \frac{I_{x^*}(\phi_2, \phi_1)}{I_z(\phi_2, \phi_1)},$$

where $x^* = x/(1+x)$.

Monotone likelihood ratio property. The MLR properties associated with W^* are pretty easy to establish. By Theorem A.15, it follows that:

(i) for fixed $r > 0$, $W|W \leq r/\sigma$ has decreasing MLR in σ;

(ii) for fixed $\sigma = 1$, $S|S \leq r$ has increasing MLR in r;

(iii) for fixed $\sigma = 1$, $g_{S|S\leq r}(s|r)/g_S(s)$ is decreasing in s.

2.4.6 Confidence distribution for $J\sigma_\alpha^2$

It is possible to express $J\sigma_\alpha^2$ as a function of the sums of squares T_1 and T_2, along with the pivotals U_1 and U_2, as follows:

$$\begin{aligned} J\sigma_\alpha^2 &= \tau_2 - \tau_1 \\ &= \frac{T_2\tau_2}{T_2} - \frac{T_1\tau_1}{T_1} \\ &= \frac{T_2}{U_2} - \frac{T_1}{U_1} \\ &= V, \end{aligned}$$

say. The pivotals U_1 and U_2 have χ^2 distributions, which carry the probabilistic information in V. Given the data, if one substitutes the realizations t_1 of T_1 and t_2 of T_2 into the above expression, one obtains the 'potential pivotal quantity' (Weerahandi, 1994, p. 255)

$$V' = \frac{t_2}{U_2} - \frac{t_1}{U_1}.$$

In this form, the random variable V' is defined over the entire real line, since it is entirely possible that $t_2/U_2 < t_1/U_1$. Fisher (1935) derived V' as the fiducial pivotal for $J\sigma_\alpha^2$, but essentially ignored that part of the distribution to the left of zero. Healy (1963) formalized this idea by attaching an atom of probability equal to $\Pr(V' \leq 0)$ at the point $V' = 0$. The resulting distribution has become known as the Fisher-Healy distribution.

Fisher (1935) justified the transition from V to V' by noting that t_1 and t_2 are fixed sample quantities given the data, so that, conditionally, the uncertainty in V' pertains to uncertainty in $J\sigma_\alpha^2$ by the fiducial argument. Box and Tiao (1973) used a Bayesian analogue of this transformation to obtain a posterior density of σ_α^2. Use of t_1 and t_2 as fixed sample quantities post-data is justified from a Bayesian perspective, since observed data are treated as fixed. More recently, Weerahandi (1994) used the pivotal V' to construct confidence intervals for σ_α^2 by a generalized p-value method. He, too, treats t_1 and t_2 as fixed sample quantities rather than random variables.

Our approach to this problem is to condition V' on $W \leq s$, producing the conditioned random variable

$$V^*(t_1, t_2) = \frac{t_2}{U_2} - \frac{t_1}{U_1} \mid W \leq s.$$

Given the data, replacement of T_1 and T_2 by t_1 and t_2, respectively, has a fiducial flavor in that the observed sums of squares are treated as fixed constants, while the pivotals U_1 and U_2 carry the probabilistic information. An important byproduct is that the support of V^* is $[0, \infty)$, the parameter space of $J\sigma_\alpha^2$; in contrast, the support of the potential pivotal V' can include negative values if t_2 is sufficiently small relative to t_1.

The purpose of using V^* and its distribution in PDP inference is to provide a means of producing confidence intervals for $J\sigma_\alpha^2$, *a la* Weerahandi (1994, section 9.4).

2.4.7 Properties of V^*

To find the density of V^* for any fixed (t_1, t_2) pair, we begin with the joint PDP density (2.5):

$$\begin{aligned} g^*(u_1, u_2|s) &= \frac{g_{\nu_1}(u_1)\, g_{\nu_2}(u_2)}{I_z(\phi_2, \phi_1)} \\ &= \frac{u_1^{\phi_1-1}\, u_2^{\phi_2-1}\, e^{-(u_1+u_2)/2}}{\Gamma(\phi_1)\,\Gamma(\phi_2)\, 2^{\phi_1+\phi_2} I_z(\phi_2, \phi_1)}, \end{aligned}$$

where $u_1 \geq 0$ and $0 \leq u_2 \leq su_1$. Make the transformation

$$V_1 = \frac{t_1}{U_1} \quad \text{and} \quad V_2 = \frac{t_2}{U_2} - \frac{t_1}{U_1},$$

with inverse transform

$$U_1 = \frac{t_1}{V_1} \quad \text{and} \quad U_2 = \frac{t_2}{V_1 + V_2}$$

and Jacobian $t_1 t_2/(V_1(V_1+V_2))^2$. The constraint $W \leq s$ becomes

$$W = \frac{U_2}{U_1} = \frac{t_2}{V_1+V_2}\frac{V_1}{t_1} = \frac{s\,V_1}{V_1+V_2} \leq s,$$

which implies $V_1 \leq V_1 + V_2$, or $V_2 \geq 0$. The transformation maps the bivariate region $\{U_1 > 0, 0 < U_2 \leq sU_1\}$ into $\{V_1 > 0, V_2 \geq 0\}$. The transformed joint density is then

$$\begin{aligned} g^*(v_1, v_2|t_1, t_2) &= K^{-1}\Big(\frac{t_1}{v_1}\Big)^{\phi_1-1}\Big(\frac{t_2}{v_1+v_2}\Big)^{\phi_2-1}\exp\Big\{-\frac{1}{2}\Big(\frac{t_1}{v_1} + \frac{t_2}{v_1+v_2}\Big)\Big\}\frac{t_1 t_2}{v_1^2(v_1+v_2)^2} \\ &= \frac{1}{t_1 t_2 K}\Big(\frac{t_1}{v_1}\Big)^{\phi_1+1}\Big(\frac{t_2}{v_1+v_2}\Big)^{\phi_2+1}\exp\Big\{-\frac{1}{2}\Big(\frac{t_1}{v_1} + \frac{t_2}{v_1+v_2}\Big)\Big\}, \end{aligned} \tag{2.30}$$

where $K = \Gamma(\phi_1)\,\Gamma(\phi_2)\, 2^{\phi_1+\phi_2}\, I_z(\phi_2, \phi_1)$ is the normalizing constant.

2.4.7.1 Density of V^*

The marginal density of $V^* = V_2$ is obtained by integrating (2.30) over V_1; i.e.,

$$g^*(v_2|t_1, t_2) = \frac{1}{t_1 t_2 K}\int_0^\infty \Big(\frac{t_1}{v_1}\Big)^{\phi_1+1}\Big(\frac{t_2}{v_1+v_2}\Big)^{\phi_2+1} \exp\Big\{-\frac{1}{2}\Big(\frac{t_1}{v_1} + \frac{t_2}{v_1+v_2}\Big)\Big\}\, dv_1, \tag{2.31}$$

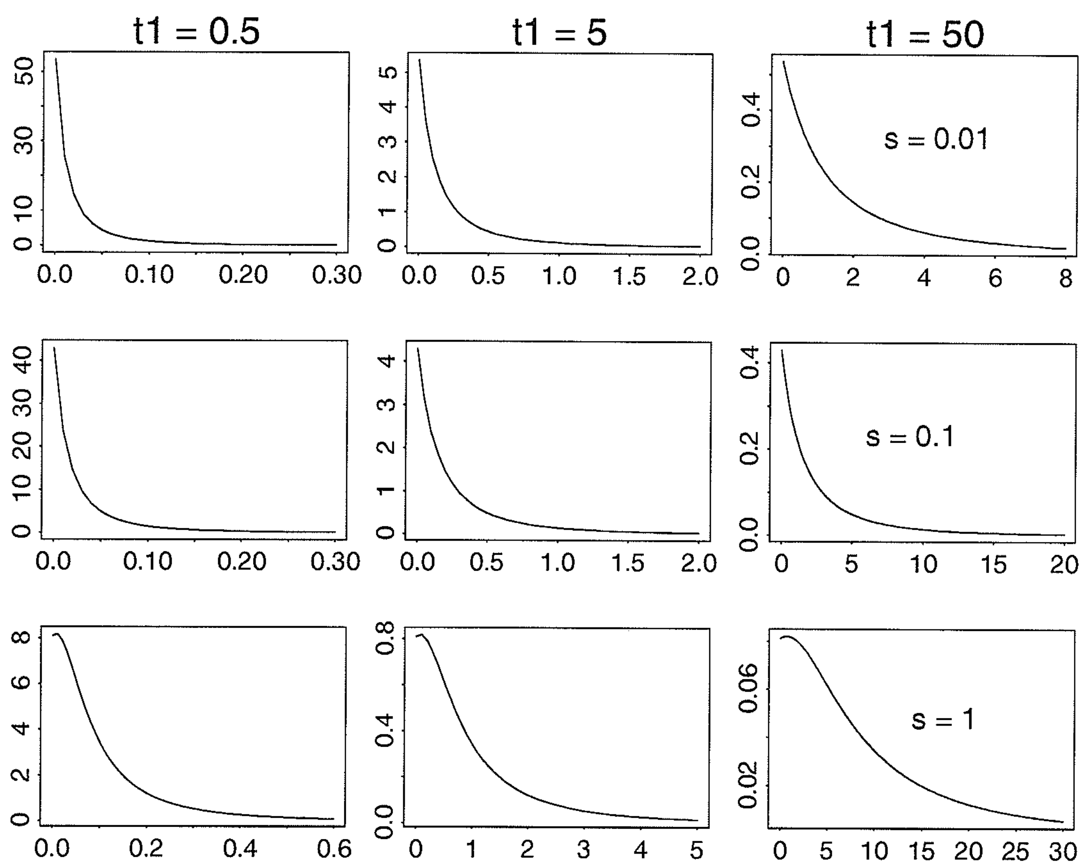

Figure 2.12: Graphs comparing the density of $V^*(t_1, t_2)$ for the case where $(\nu_2, \nu_1) = (5, 6)$. The rows correspond to a fixed value of $s = t_2/t_1$, while the columns are associated with fixed values of t_1. (See also Figure 2.13.)

which has no closed form solution. This density is a function of *both* t_1 and t_2, and not solely on their ratio s; hence, different values of t_1 and t_2, even with the same ratio s, will yield different densities. The expression (2.31) is mathematically identical (up to the scale factor J) to the posterior density of σ_α^2 obtained by Box and Tiao, although its interpretation is quite different.

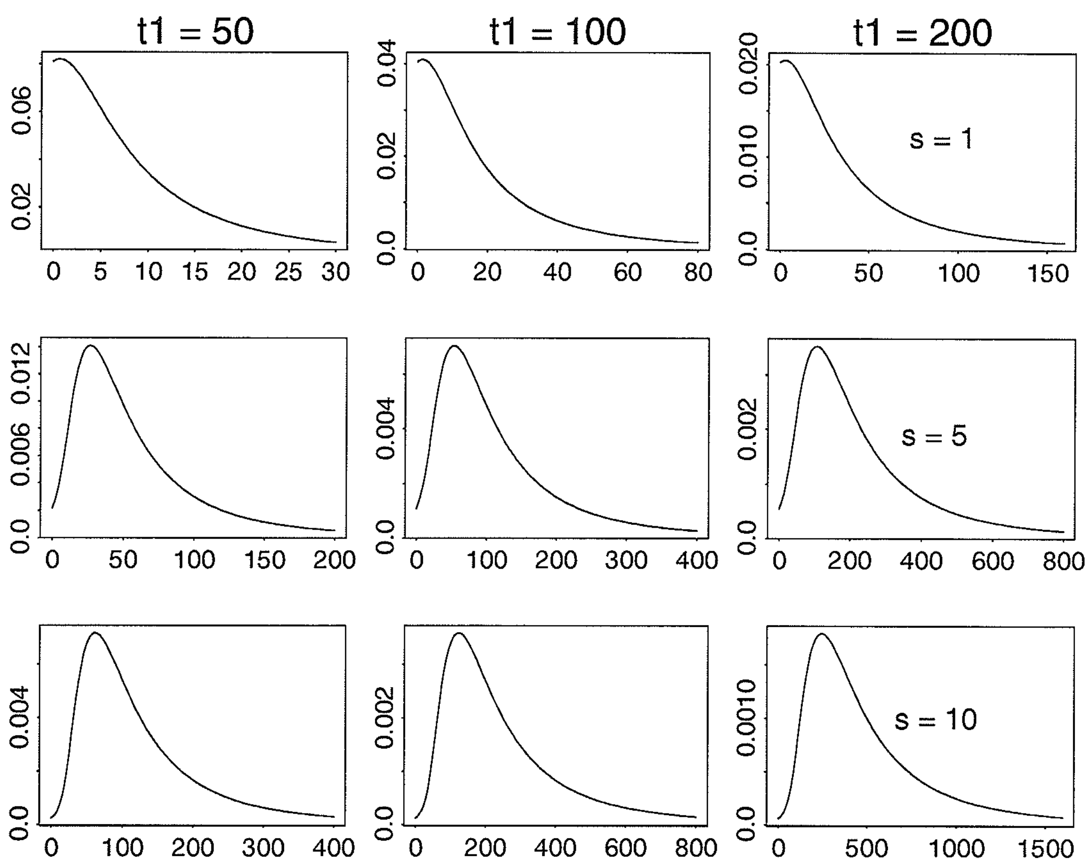

Figure 2.13: More graphs of the density of $V^*(t_1, t_2)$ when $(\nu_2, \nu_1) = (5, 6)$. The rows correspond to fixed values of $s = t_2/t_1$, while the columns are associated with fixed values of t_1.

This process defines a family of V^* distributions pre-data; given t_1 and t_2 from observed data, a particular member is singled out for inference about $J\sigma^2_\alpha$. Figures 2.12 and 2.13 illustrate several members of the family of V^* distributions for varying (t_1, t_2) pairs when $(\nu_2, \nu_1) = (5, 6)$. When s is fixed across rows, the domain and range of the density changes with t_1 and t_2, but its shape is the same. This feature has interesting ramifications for interval estimation of $J\sigma^2_\alpha$ from this family of densities.

2.4.7.2 Moments of V^*.

By Theorem A.40, the expected value of V^* is given by

$$\begin{aligned} E(V^*|t_1, t_2) &= \frac{t_2}{2}\left[\frac{\Gamma(\phi_2 - 1)}{\Gamma(\phi_2)}\frac{I_z(\phi_2 - 1, \phi_1)}{I_z(\phi_2, \phi_1)} - \frac{1}{s}\frac{\Gamma(\phi_1 - 1)}{\Gamma(\phi_1)}\frac{I_z(\phi_2, \phi_1 - 1)}{I_z(\phi_2, \phi_1)}\right] \\ &= \frac{t_2}{\nu_2 - 2}\frac{I_z(\phi_2 - 1, \phi_1)}{I_z(\phi_2, \phi_1)} - \frac{t_1}{\nu_1 - 2}\frac{I_z(\phi_2, \phi_1 - 1)}{I_z(\phi_2, \phi_1)}. \end{aligned} \tag{2.32}$$

This result also follows from

$$\begin{aligned} E(V^*|t_1, t_2) &= E\Big(\frac{t_2}{U_2} - \frac{t_1}{U_1}\;\Big|\; W \le s\Big) \\ &= t_2\, E\Big(\frac{1}{U_2^*}\Big) - t_1\, E\Big(\frac{1}{U_1^*}\Big), \end{aligned} \tag{2.33}$$

where the expectations in brackets are the first negative moments of U_2^* and U_1^* (Theorems A.33 and A.28, respectively). This expectation exists if $\nu_1 > \nu_2 > 2$. The limiting behavior of $E(V^*)$ is evaluated with respect to t_1 and t_2 in section A.8.2; essentially, one of these has to be fixed while the other goes to zero in order for $E(V^*)$ to exist in the limit as $s \to 0$ or $s \to \infty$.

Determination of the second moment of V^* follows similar lines. Since

$$V^{*2}(t_1, t_2) = \Big(\frac{t_2}{U_2} - \frac{t_1}{U_1}\Big)^2 \Big|\; W \le s,$$

by using Theorems A.28 and A.33, it is rather straightforward to show that

$$E(V^{*2}|t_1,t_2) = \left(\frac{t_2}{2}\right)^2 \left[\frac{\Gamma(\phi_2-2)I_z(\phi_2-2,\phi_1)}{\Gamma(\phi_2)I_z(\phi_2,\phi_1)} - \frac{2}{s}\frac{\Gamma(\phi_2-1)\Gamma(\phi_1-1)I_z(\phi_2-1,\phi_1-1)}{\Gamma(\phi_2)\Gamma(\phi_1)I_z(\phi_2,\phi_1)} + \frac{1}{s^2}\frac{\Gamma(\phi_1-2)I_z(\phi_2,\phi_1-2)}{\Gamma(\phi_1)I_z(\phi_2,\phi_1)}\right] \quad (2.34)$$

whenever $\nu_1 > \nu_2 > 4$.

In general, the r^{th} moment of V^* is given by

$$\mu_r = \left(\frac{t_2}{2}\right)^r \sum_{k=0}^{r} \binom{r}{k}\left(-\frac{1}{s}\right)^k \frac{I_z(\phi_2-r+k,\phi_1-k)}{I_z(\phi_2,\phi_1)}\,\frac{\Gamma(\phi_2-r+k)\Gamma(\phi_1-k)}{\Gamma(\phi_2)\Gamma(\phi_1)}, \quad (2.35)$$

which holds whenever $\nu_1 > \nu_2 > 2r$, as shown by Box and Tiao (1973, p. 296)[2]. Thus, the first moment exists only if $\nu_2 \geq 3$, whereas the second moment, and hence, the variance, exists only if $\nu_2 \geq 5$.

2.4.7.3 cdf of V^*

Using a derivation of Weerahandi (1994), the cdf of V^* at some value $c > 0$ is given by

$$P(V^* \leq c|t_1,t_2) = 1 - \frac{\int_0^z G_{\nu_1+\nu_2}\left[\frac{1}{c}\left(\frac{t_2}{v} - \frac{t_1}{1-v}\right)\right] g_V(v)\,dv}{I_z(\phi_2,\phi_1)}. \quad (2.36)$$

Derivation of this expression is given in section A.8.3 for any fixed (t_1,t_2) pair.

2.4.8 Related distributions

This subsection deals with properties of the distribution of $U_i|S=s$, $i=1,2$, which arise in the construction of Stein-like point estimators in section 2.5. The conditional

[2] Two things should be noted in comparing (2.35) with expression (A.5.3.1) in Box & Tiao. Firstly, we use $1/s$ rather than s because we are using chi-square rather than inverse chi-square densities; secondly, Box and Tiao meant to use r in their moment formula rather than n, so there is a minor typo.

moments of $U_i|S = s$, $i = 1, 2$, are easily found by observing that

$$(1 + S/\sigma)U_1 = U_1 + U_2 = (1 + \sigma/S)U_2. \tag{2.37}$$

Since $U_1+U_2 \sim \chi^2_{\nu_1+\nu_2}$, it follows that $(1+S/\sigma)U_1$ and $(1+\sigma/S)U_2$ are also distributed $\chi^2_{\nu_1+\nu_2}$. Hence,

$$E_\sigma[(1 + S/\sigma)U_1] = \nu_1 + \nu_2 = E_\sigma[(1 + \sigma/S)U_2]$$

and

$$E_\sigma[(1 + S/\sigma)^2 U_1^2] = (\nu_1 + \nu_2)(\nu_1 + \nu_2 + 2) = E_\sigma[(1 + \sigma/S)^2 U_2^2].$$

Since $U_1 + U_2$ and W are independently distributed, conditioning on W (or S) does not affect the above expectations, so that

$$E_\sigma[U_1|S] = \frac{\nu_1 + \nu_2}{1 + S/\sigma}, \tag{2.38}$$

$$E_\sigma[U_1^2|S] = \frac{(\nu_1 + \nu_2)(\nu_1 + \nu_2 + 2)}{(1 + S/\sigma)^2}, \tag{2.39}$$

$$E_\sigma[U_2|S] = \frac{\nu_1 + \nu_2}{1 + \sigma/S}, \tag{2.40}$$

$$E_\sigma[U_2^2|S] = \frac{(\nu_1 + \nu_2)(\nu_1 + \nu_2 + 2)}{(1 + \sigma/S)^2}. \tag{2.41}$$

2.4.9 Relations among PDP, Bayesian and fiducial distributions

In this subsection, we show how PDP distribution theory in the balanced one-way random model is related to the Bayesian posteriors derived by Box and Tiao (1973, ch. 5) and the fiducial distributions of Venables and James (1978) and Wild (1981). We begin by outlining Box and Tiao's Bayesian approach, showing how their posterior distributions connect to PDP distributions. Since Wild (1981) showed how to relate Box and Tiao's distributions with the fiducial distributions through a conditioning argument, an appeal to his results shows the connection between the fiducial and Bayesian results. Hence, the ensuing discussion is geared towards 'unification' of these three modes of post-data inference in terms of their distribution theory.

2.4.9.1 Bayesian posteriors

Under the balanced one-way random model (2.1), Box and Tiao (1973, section 5.2) derived posterior distributions of the parameters of interest over the constrained parameter space $\tau_2 \geq \tau_1$. Under the distributional assumptions of model (2.1) and Jeffreys' rule, the joint prior of μ, τ_1 and τ_2 is given by

$$p(\mu, \tau_1, \tau_2) \propto \tau_1^{-1}\, \tau_2^{-1}, \tag{2.42}$$

while the corresponding joint likelihood function is

$$\ell(\mu, \tau_1, \tau_2 | \bar{y}_{..}, t_1, t_2) \propto \tau_1^{-\phi_1}\, \tau_2^{-(\phi_2+1/2)} \exp\left\{ -\frac{1}{2} \left[\frac{N(\bar{y}_{..} - \mu)^2}{\tau_2} + \frac{t_2}{\tau_2} + \frac{t_1}{\tau_1} \right] \right\},$$

where $N = IJ$ is the total number of observations. The joint posterior density of (μ, τ_1, τ_2) is then

$$p(\mu, \tau_1, \tau_2 | \bar{y}_{..}, t_1, t_2) \propto \tau_1^{-(\phi_1+1)}\, \tau_2^{-(\phi_2+1+1/2)} \exp\left\{ -\frac{1}{2} \left[\frac{N(\bar{y}_{..} - \mu)^2}{\tau_2} + \frac{t_2}{\tau_2} + \frac{t_1}{\tau_1} \right] \right\}, \tag{2.43}$$

where $\mu \in \Re$, $0 < \tau_1 < \tau_2$. By integrating over μ, the joint posterior of τ_1 and τ_2 over the region $0 < \tau_1 < \tau_2$ becomes

$$p(\tau_1, \tau_2 | t_1, t_2) = \frac{J}{t_1\, t_2} \frac{g^*_{\nu_1}(\tau_1/t_1)\, g^*_{\nu_2}(\tau_2/t_2)}{I_z(\phi_2, \phi_1)}, \tag{2.44}$$

where $g^*(\cdot)$ denotes an inverse χ^2 density. Since probability is defined on the parameter space under the Bayesian approach, the τ_i are considered random while the corresponding t_i are presumed fixed, $i = 1, 2$. Box and Tiao (1973, p. 254) show that

$$P(\sigma \geq 1) = I_z(\phi_2, \phi_1); \tag{2.45}$$

inclusion of this term in the joint posterior of (τ_2, τ_1) is necessary since the posterior is defined over $\tau_1 > 0, \tau_2 > \tau_1$. Without this term, the joint posterior (2.44) would be defined over the region $\tau_1 > 0, \tau_2 > 0$.

The marginal posterior density of τ_1 is obtained by integrating (2.44) over τ_2, yielding

$$p(\tau_1|t_1,t_2) = \frac{1}{t_1}\frac{g^*_{\nu_1}(\tau_1/t_1)\,G_{\nu_2}(t_2/\tau_1)}{I_z(\phi_2,\phi_1)} = \frac{1}{t_1}\,g^*_{\nu_1}(\tau_1/t_1)\,f(\tau_1), \tag{2.46}$$

where

$$f(\tau_1) = \frac{G_{\nu_2}(t_2/\tau_1)}{I_z(\phi_2,\phi_1)}$$

and $G(\cdot)$ denotes a χ^2 cdf. Box and Tiao (p. 259) view $f(\tau_1)$ as representing the additional information about τ_1 that comes from the treatment sum of squares t_2: the numerator represents the probability of the constrained event $\sigma \geq 1$ for a specific value of τ_1, whereas the denominator represents the probability that the constraint is true over all values of τ_1. The denominator $I_z(\phi_2,\phi_1)$ is independent of τ_1 and is simply a normalizing constant. The shape of $f(\tau_1)$ depends on the observed F ratio m_2/m_1, and therefore operationalizes the 'smooth pooling' principle in this family of posteriors. When this ratio is large, $f(\tau_1)$ is a relatively flat function, taking values close to 1; this corresponds to $\tau_1|t_1,t_2 \to \chi^{-2}_{\nu_1}$ as s gets large. Conversely, as $s \to 0$, $f(\tau_1)$ increases in such a way that $\tau_1|t_1,t_2 \to \chi^{-2}_{\nu_1+\nu_2}$. In general, f is a decreasing function of τ_1 (cf. Figures 5.2.3 and 5.2.4 in Box and Tiao, pp. 259-260).

From a sampling theory perspective, one can write

$$f(\tau_1) = \frac{G_{\nu_2}(t_2/\tau_1)}{I_z(\phi_2,\phi_1)} = \frac{1-\alpha(\tau_1)}{1-\alpha},$$

where $\alpha(\tau_1)$ is the p-value of an F test of $H_0 : \tau_1 = \tau_2$ vs. $H_1 : \tau_2 > \tau_1$ for a specific value of τ_1, and α is the p-value of an overall F test of H_0 where τ_1 is unspecified (Box and Tiao, p. 261).

The posterior density of τ_2 is not explicitly derived in Box and Tiao because it is of minor interest to a Bayesian, but it can be obtained by integrating the joint posterior of (τ_1, τ_2) over τ_1:

$$p(\tau_2|t_1,t_2) = \frac{1}{t_2}\frac{g^*_{\nu_2}(\tau_2/t_2)\,[1-G_{\nu_1}(t_1/\tau_2)]}{I_z(\phi_2,\phi_1)}, \tag{2.47}$$

where $\tau_2 > 0$. Analogous to the PDP density of U_2, this posterior ranges between a $\chi^{-2}_{\nu_2}$ as s gets large, and a degenerate distribution at 0 as $s \to 0$.

The posterior density of $\sigma = 1 + J\gamma$ is obtained by performing a bivariate transformation $(\tau_1, \tau_2) \mapsto (\tau_1, \sigma)$ on the joint posterior density (2.44); integrating over τ_1,

$$p(\sigma|f) = \frac{1}{f}\frac{h_{\nu_1,\nu_2}(\sigma/f)}{I_z(\phi_2,\phi_1)} \tag{2.48}$$

where $f = m_2/m_1$ and $h(\cdot)$ denotes an $\mathcal{F}_{\nu_1,\nu_2}$ density with support $\sigma \geq 1$. Hence, the posterior density of σ is truncated from below at 1.

Finally, the marginal posterior of σ^2_α is obtained by performing a bivariate transformation $(\tau_1, \tau_2) \mapsto (\tau_1, \sigma^2_\alpha)$ on the joint posterior (2.44); integrating over τ_1,

$$p(\sigma^2_\alpha|t_1,t_2) = \frac{J}{t_1\,t_2\,I_z(\phi_2,\phi_1)}\int_0^\infty g^*_{\nu_2}\Big(\frac{\tau_1 + J\sigma^2_\alpha}{t_2}\Big)\,g^*_{\nu_1}\Big(\frac{\tau_1}{t_1}\Big)\,d\tau_1,$$

with support $\sigma^2_\alpha > 0$.

We are now in position to demonstrate the connection between the PDP densities and the above posterior densities. Firstly, we summarize the above posteriors as a lemma. A simple transformation expresses these posteriors in pivotal form; from there, the connection can be made by a straightforward argument.

Lemma 2.1. *If the (improper) prior density of (μ, τ_1, τ_2) is given by*

$$\pi(\mu,\tau_1,\tau_2) = \frac{1}{\tau_1\,\tau_2} \qquad -\infty < \mu < \infty,\ 0 < \tau_1 < \tau_2 < \infty,$$

then

(a) the joint posterior density function of (μ, τ_1, τ_2) is

$$h(\mu,\tau_1,\tau_2|\bar{y}_{..},t_1,t_2) \propto \tau_1^{-(\phi_1+1)}\,\tau_2^{-(\phi_2+1+1/2)}\,\exp\Big\{-\frac{1}{2}\Big[\frac{N(\bar{y}_{..}-\mu)^2}{\tau_2} + \frac{t_2}{\tau_2} + \frac{t_1}{\tau_1}\Big]\Big\},$$

where $\mu \in \Re$, $\tau_1 > 0$ *and* $\tau_2 > 0$;

(b) the joint posterior density of (τ_1, τ_2) *is*

$$h(\tau_1, \tau_2|t_1, t_2) = \frac{J}{t_1\, t_2} \frac{g^*_{\nu_1}(\tau_1/t_1)\, g^*_{\nu_2}(\tau_2/t_2)}{I_z(\phi_2, \phi_1)}, \qquad 0 < \tau_1 < \tau_2 < \infty,$$

where $g^*(\cdot)$ *denotes an inverse* χ^2 *density;*

(c) the posterior density functions of τ_1, τ_2 *and* σ *are given by*

$$\begin{aligned} h(\tau_1|s, t_1) &= \frac{1}{t_1} \frac{g^*_{\nu_1}(\tau_1/t_1)\, G_{\nu_2}(t_2/\tau_1)}{I_z(\phi_2, \phi_1)} && \tau_1 > 0,\\ h(\tau_2|s, t_1) &= \frac{1}{t_2} \frac{g^*_{\nu_2}(\tau_2/t_2)\, [1 - G_{\nu_1}(t_1/\tau_2)]}{I_z(\phi_2, \phi_1)} && \tau_2 > 0,\\ h(\sigma|s) &= \frac{1}{f} \frac{p_{\nu_1,\nu_2}(\sigma/f)}{I_z(\phi_2, \phi_1)} && \sigma > 1, \end{aligned}$$

respectively, where $f = \nu_1\, s/\nu_2$ *and* $p(\cdot)$ *denotes an* F *density.*

Corollary 2.1.1. *If* $U_1 = t_1/\tau_1$, $U_2 = t_2/\tau_2$ *and* $W = s/\sigma$, *then*

(a) the joint posterior density of (U_1, U_2) *is*

$$h(u_1, u_2|t_1, t_2) = \frac{g_{\nu_1}(u_1)\, g_{\nu_2}(u_2)}{I_z(\phi_2, \phi_1)}, \qquad 0 < u_2 < s\, u_1 < \infty,$$

where $g(\cdot)$ *denotes a* χ^2 *density;*

(b) the posterior density functions of U_1, U_2 *and* W *are given by*

$$\begin{aligned} h(u_1|s, t_1) &= \frac{g_{\nu_1}(u_1)\, G_{\nu_2}(s\, u_1)}{I_z(\phi_2, \phi_1)} && u_1 > 0,\\ h(u_2|s, t_1) &= \frac{g_{\nu_2}(u_2)\, [1 - G_{\nu_1}(u_2/s)]}{I_z(\phi_2, \phi_1)} && u_2 > 0,\\ h(w|s) &= \frac{p(w; \phi_2, \phi_1)}{I_z(\phi_2, \phi_1)} && 0 < w < s, \end{aligned}$$

respectively, where $p(\cdot)$ *denotes a Beta type II density. In other words, the above posterior densities agree with the corresponding PDP densities.*

Proof. If we write the joint posterior density of (τ_1, τ_2) given in Lemma 2.1(b) in terms of chi-square pdf's rather than inverse chi-squares, we have

$$h(\tau_1, \tau_2 | t_1, t_2) = K^{-1}\, \tau_1^{-2}\, \tau_2^{-2}\, g_{\nu_1}\!\left(\frac{t_1}{\tau_1}\right) g_{\nu_2}\!\left(\frac{t_2}{\tau_2}\right), \qquad 0 < \tau_1 < \tau_2, \tag{2.49}$$

where

$$K = \int_0^\infty \int_{\tau_1}^\infty \tau_1^{-2}\, \tau_2^{-2}\, g_{\nu_1}\!\left(\frac{t_1}{\tau_1}\right) g_{\nu_2}\!\left(\frac{t_2}{\tau_2}\right) d\tau_2\, d\tau_1.$$

Make the transformation $(\tau_1, \tau_2) \mapsto (U_1, U_2)$; then

$$d\tau_2 = \frac{t_2}{u_2^2}\, du_2 \qquad \text{and} \qquad d\tau_1 = \frac{t_1}{u_1^2}\, du_1.$$

At the same time,

$$\begin{aligned} \tau_2 > \tau_1 \quad &\Longrightarrow \quad \tau_2^{-1} < \tau_1^{-1} \\ &\Longrightarrow \quad \frac{t_2}{\tau_2} < \frac{t_2}{\tau_1} = s\,\frac{t_1}{\tau_1} \\ &\Longrightarrow \quad U_2 < s\, U_1, \end{aligned}$$

which shows the connection between the constrained parameter space and the PDP conditioning event on the pivotal space. With these changes,

$$\begin{aligned} K &= \int_0^\infty \int_0^{s u_1} \frac{1}{t_1\, t_2}\, g_{\nu_1}(u_1)\, g_{\nu_2}(u_2)\, du_2\, du_1 \\ &= \frac{1}{t_1\, t_2} \int_0^\infty g_{\nu_1}(u_1)\, G_{\nu_2}(s u_1)\, du_1 \\ &= \frac{1}{t_1\, t_2}\, I_z(\phi_2, \phi_1), \end{aligned}$$

the last result due to Lemma A.24. Plugging this into the joint posterior density (2.49) of (τ_1, τ_2), a slight rearrangement yields

$$h(\tau_1, \tau_2 | t_1, t_2) = \frac{1}{I_z(\phi_2, \phi_1)} \prod_{i=1}^{2} \frac{1}{\tau_i}\, \frac{t_i}{\tau_i}\, g_{\nu_i}\!\left(\frac{t_i}{\tau_i}\right) \qquad 0 < \tau_1 < \tau_2.$$

Transforming to $u_i = t_i/\tau_i$, $i = 1, 2$, so that $\tau_i = t_i/u_i$ and $d\tau_i = t_i/u_i^2\, du_i$, yields

$$h(u_1, u_2 | t_1, t_2) = \frac{1}{I_z(\phi_2, \phi_1)} \prod_{i=1}^{2} \left(\frac{u_i}{t_i}\right) u_i\, g_{\nu_i}(u_i)\, \frac{t_i}{u_i^2}$$

$$= \frac{g_{\nu_1}(u_1)\, g_{\nu_2}(u_2)}{I_z(\phi_2, \phi_1)} \qquad u_2 < s\, u_1.$$

Part (a) of the corollary is now established. The results in part (b) follow by marginalization and/or change of variables, as in the derivation of Lemma 2.1. □

Corollary 2.1.1 establishes the important connection between PDP and Bayesian distribution theory. This corollary essentially maps the posterior distributions from their respective parameter spaces to the corresponding pivotal spaces, *keeping the t_i fixed and the τ_i random.* Once in the pivotal space, U_i is a random variable whether t_i is fixed and τ_i is random *or vice versa.* Lemma 2.1 and its corollary show that the Box and Tiao posterior densities, when transformed according to Corollary 2.1.1, have the same form as the PDP densities and are defined over the same regions in the joint pivotal space. The critical difference between them is the interpretation of the densities when the t_i are fixed versus when the T_i are random.

Corollary 2.1.1 establishes the connection between the marginal posteriors of the τ_i and the PDP distributions of U_i^*, as well as the connection between the posterior of σ and the PDP distribution of $W^*(s)$. A further set of transformations connects the posterior of $J\sigma_\alpha^2$ to the distribution of $V^*(t_1, t_2)$.

2.4.9.2 Fiducial densities

Fiducial densities used for inference in the one-way random model that account for the parameter constraint $\sigma \geq 1$ have been considered by Venables and James (1978) and Wild (1981). Let $\omega = J\gamma$, where $\gamma = \sigma_\alpha^2/\tau_1$ is the variance component ratio.

Venables and James begin by considering the pivotal quantities

$$\left(t_1 + \frac{t_2}{\sigma}\right)\frac{\omega}{J\sigma_\alpha^2} = \left(t_1 + \frac{t_2}{\sigma}\right)\frac{1}{\tau_1} \sim \chi^2_{\nu_1+\nu_2}. \tag{2.50}$$

Pre-data, the t's above are actually T's, since they are considered random variables for fixed, unknown values of τ_1 and σ_α^2. Once the data are observed, the realizations t_1 and t_2 are considered fixed, so the probability content in the pivotals now becomes associated with the parameters. The pivotals are then inverted so that, post-data,

$$\tau_1 \sim \left(t_1 + \frac{t_2}{\sigma}\right)\chi^{-2}_{\nu_1+\nu_2},$$
$$J\sigma_\alpha^2 \sim \left(t_1 + \frac{t_2}{\sigma}\right)\omega\chi^{-2}_{\nu_1+\nu_2}.$$

Wild (1981) starts slightly differently, beginning with the two independent pivotals in (2.50) along with

$$\frac{f}{\sigma} \sim \mathcal{F}_{\nu_2,\nu_1},$$

where f is the (observed) F ratio; the inverted pivotal is $\sigma \sim f^{-1}\mathcal{F}_{\nu_1,\nu_2}$.

In each paper, the authors are concerned with modifying the fiducial distributions in the presence of the parameter constraint $\sigma = 1 + \omega \geq 1$ (or $\omega \geq 0$). Wild (1981) observes that there are two ways to handle this constraint. Letting $Y = 1 + \omega$, one approach is to attach an atom of probability equal to $P(Y \leq 1)$ to the event $1 + \omega = 1$. The method of Venables and James (1978) is of this type, leading to the Fisher-Healy distribution for σ_α^2 (Healy, 1963). Fiducial distributions of this type are also confidence distributions, and resulting fiducial intervals are also confidence intervals (Wild, 1981).

Another tack is to condition on the event $Y \geq 1$. Then, Wild shows that the fiducial distributions for τ_1, σ_α^2 and σ are all mathematically the same as the corresponding posterior densities of Box and Tiao (1973, section 5.2). As a result, one can connect the fiducial, Bayesian and PDP distributions in the balanced one-way random model.

2.5 Point estimation

Having developed PDP distribution theory, the next step is to discuss various types of point estimators of the parametric functions of interest in the balanced one-way random model. The present task is to apply the PDP argument to produce point estimators of the relevant parametric functions, and to evaluate their frequentist properties with respect to a specific bowl-shaped loss function.

2.5.1 Frequentist point estimators

Point estimation of (functions of) the variance components in a balanced one-way random model with normal errors typically entails discussion of one or more of the following methods: ANOVA (or method of moments), unbiasedness, or (restricted) maximum likelihood.

2.5.1.1 ANOVA and unbiased estimators

ANOVA estimators for τ_i are of the form T_i/ν_i, $i = 1, 2$. Both are uniformly minimum variance unbiased (UMVU) in the class of quadratic estimators, and under normality, are UMV in the wider class of unbiased estimators of τ_i. Thus, unbiased ANOVA estimators of τ_1 and τ_2 have 'good' pre-data properties. The ANOVA estimator of $J\sigma_\alpha^2 = \tau_2 - \tau_1$,

$$\bar{\sigma}_\alpha^2 = \frac{T_2}{\nu_2} - \frac{T_1}{\nu_1}, \tag{2.51}$$

is also unbiased, since $J\sigma_\alpha^2$ is a linear function of the expected mean squares; moreover, it is UMV in the class of unbiased estimators of $J\sigma_\alpha^2$. However, when realizations t_1 and t_2 arise such that $m_2 < m_1$, the estimate of $J\sigma_\alpha^2$ through the rule (2.51) turns out to be negative.

ANOVA estimators of σ and $J\gamma$ are plug-ins, based on the ANOVA estimators of τ_1 and τ_2. Neither estimator is unbiased, and both can yield estimates outside the appropriate parameter space. One alternative is to use an unbiased estimator of σ:

$$\bar{\sigma} = \frac{\nu_2}{\nu_1 - 2} S. \tag{2.52}$$

Since $J\gamma = \sigma - 1$, plugging in (2.52) for σ produces an unbiased estimator. When a realization s is smaller than $\nu_2/(\nu_1 - 2)$, however, estimates of σ and $J\gamma$ based on the above unbiased rules lie outside their respective parameter spaces. *Thus, neither ANOVA nor unbiased estimators of σ, $J\sigma_\alpha^2$, $J\gamma$ (or the intraclass correlation coefficient ρ) exhibit consistently good post-data behavior.*

2.5.1.2 REML estimators of τ_1 and τ_2

Under the balanced one-way random model (2.1) with parameter constraint $\sigma \geq 1$, REML estimators of τ_1, τ_2 and $J\sigma_\alpha^2$ are given by

$$\hat{\tau}_1^R = \min\left[\frac{T_1}{\nu_1}, \frac{T_1 + T_2}{\nu_1 + \nu_2}\right] \tag{2.53}$$

$$\hat{\tau}_2^R = \max\left[\frac{T_2}{\nu_2}, \frac{T_1 + T_2}{\nu_1 + \nu_2}\right] \tag{2.54}$$

$$J\hat{\sigma}_\alpha^{2\,R} = \max\left[0, \frac{T_2}{\nu_2} - \frac{T_1}{\nu_1}\right]. \tag{2.55}$$

The rationale for, and derivation of, these estimators is given in Searle, Casella and McCulloch (1992, pp. 90–92). Basically, solutions of the REML equations that lie outside the joint $(\sigma_e^2, \sigma_\alpha^2)$ parameter space are pulled to the boundary. From (2.53)–(2.55), the plug-in principle produces REML estimators of σ and $J\gamma$:

$$\hat{\sigma}^R = \max[1, F], \tag{2.56}$$

$$J\hat{\gamma}^R = \max[0, F - 1], \tag{2.57}$$

where F denotes the usual F ratio. The advantage of REML (and ML) estimators is that they always assume values within the appropriate joint or marginal parameter space. ANOVA estimators provide no such guarantee.

2.5.2 PDP point estimators

In section 2.5.1, the unconditional unbiased (and ANOVA) estimators of τ_i, $i = 1, 2$, were found to be T_i/ν_i, where $\nu_i = E(U_i)$, the mean of the corresponding unconditional pivotal. As mentioned earlier, these estimators have good pre-data properties, but there are conditions under which unbiased or ANOVA estimates of $J\sigma_\alpha^2$, σ and $J\gamma$ exhibit undesirable post-data behavior under model (2.1) with $\sigma_\alpha^2 \geq 0$.

The PDP approach attempts to ameliorate such problems by restricting inference to the region $C(s) = \{U_2 \leq sU_1\}$ in the joint (U_1, U_2) pivotal space. In the process, the basic pivotals U_1, U_2 and W are conditioned by $C(s)$, yielding the post-data pivotals $U_1^*(s)$, $U_2^*(s)$ and $W^*(s)$ for a realized $s > 0$. We now consider how to conduct point estimation when the PDPs form the probabilistic foundation for inference.

2.5.2.1 PDP-unbiased point estimators

PDP-unbiased estimator of τ_1. Let $s > 0$ be given, so that $U_1^*(s)$ is a PDP for inference about τ_1. Instead of using the unconditional estimate $t_1/E(U_1)$ of τ_1, replace the unconditional mean $E(U_1)$ with the PDP mean $E[U_1^*(s)]$ given by (2.10), yielding the PDP estimate

$$\tilde{\tau}_1^{**}(s, t_1) = \frac{t_1}{E(U_1^*(s))} = \frac{t_1}{\nu_1} \frac{I_z(\phi_2, \phi_1)}{I_z(\phi_2, \phi_1 + 1)}.$$

Since this holds for any $s > 0$, the ensemble of solutions $\{\tilde{\tau}_1^{**}(s) : s > 0\}$ produces the PDP-unbiased estimator of τ_1:

$$\tilde{\tau}_1^{**}(S, T_1) = \frac{T_1}{\nu_1} \frac{I_Z(\phi_2, \phi_1)}{I_Z(\phi_2, \phi_1 + 1)}. \tag{2.58}$$

Its limiting behavior is described in the following theorem.

Theorem 2.2. *The PDP-unbiased estimator of τ_1, $\tilde{\tau}_1^{**}(S, T_1)$, has the following limiting behavior for fixed t_1:*

(a) $\lim_{s\to 0} \tilde{\tau}_1^{**}(s, t_1) = \dfrac{t_1}{\nu_1 + \nu_2}$;

(b) $\lim_{s\to\infty} \tilde{\tau}_1^{**}(s, t_1) = \dfrac{t_1}{\nu_1}$.

Proof. Let $z = s/(1+s)$. By Theorem A.2,

$$\frac{I_z(\phi_2, \phi_1 + 1)}{I_z(\phi_2, \phi_1)} = 1 + \frac{\delta_z(\phi_2, \phi_1)}{\phi_1}.$$

By Theorem A.4, $\delta_z(\phi_2, \phi_1) \to 0$ as $z \to 1$ (or $s \to \infty$), while $\delta_z(\phi_2, \phi_1) \to \phi_2$ as $s \to 0$. Consequently,

$$\lim_{s\to 0} E(U_1^*(s)) = \nu_1 + \nu_2;$$
$$\lim_{s\to\infty} E(U_1^*(s)) = \nu_1,$$

from which the result follows. □

The limiting values of the PDP-unbiased estimator of τ_1 are the same as the limits of the REML estimator $\hat{\tau}_1^R$.

PDP-unbiased estimator of τ_2. The PDP-unbiased estimator of τ_2 is produced by a similar line of argument. Given $s > 0$, $U_2^*(s)$ is a PDP for inference about τ_2. Instead of using the unconditional estimate $t_2/E(U_2)$, replace the denominator with the PDP mean $E[U_2^*(s)]$ given by (2.18), yielding the PDP estimate

$$\tilde{\tau}_2^{**}(s, t_2) = \frac{t_2}{E(U_2^*(s))} = \frac{t_2}{\nu_2} \frac{I_z(\phi_2, \phi_1)}{I_z(\phi_2 + 1, \phi_1)}.$$

Since this holds for any $s > 0$, the ensemble of solutions $\{\tilde{\tau}_2^{**}(s) : s > 0\}$ produces the PDP-unbiased estimator of τ_2:

$$\tilde{\tau}_2^{**}(S, T_2) = \frac{T_2}{\nu_2} \frac{I_Z(\phi_2, \phi_1)}{I_Z(\phi_2 + 1, \phi_1)}. \tag{2.59}$$

Its limiting behavior is described in the following theorem.

Theorem 2.3. *The PDP-unbiased estimator of τ_2, $\tilde{\tau}_2^{**}(S, T_2)$, has the following limiting behavior:*

(a) $\lim_{s\to 0} \tilde{\tau}_2^{**}(s, t_2) = \dfrac{\nu_2 + 2}{\nu_2(\nu_1 + \nu_2)}\, t_1$ *for fixed* t_1*;*

(b) $\lim_{s\to \infty} \tilde{\tau}_2^{**}(s, t_2) = \dfrac{t_2}{\nu_2}$ *for fixed* t_2*.*

Proof. To prove part (a), observe that

$$\tilde{\tau}_2^{**}(s, t_2) = \frac{t_2}{E(U_2^*(s))} = \frac{s\, t_1}{E(U_2^*(s))}.$$

For any fixed $s > 0$,

$$\nu_2 \frac{s}{E(U_2^*(s))} = \frac{z\, I_z(\phi_2, \phi_1)}{(1-z)\, I_z(\phi_2 + 1, \phi_1)}, \tag{2.60}$$

since $z = s/(1+s)$ implies that $s = z/(1-z)$. With respect to the right hand side term of (2.60), it follows by Fact A.4 that

$$\begin{aligned}\frac{z\, I_z(\phi_2, \phi_1)}{(1-z)\, I_z(\phi_2 + 1, \phi_1)} &= \frac{I_z(\phi_2 + 1, \phi_1) - (1-z)\, I_z(\phi_2 + 1, \phi_1 - 1)}{(1-z)\, I_z(\phi_2 + 1, \phi_1)} \\ &= \frac{1}{1-z} - \frac{I_z(\phi_2 + 1, \phi_1 - 1)}{I_z(\phi_2 + 1, \phi_1)}.\end{aligned}$$

By Theorem A.2,

$$\begin{aligned} &= \frac{1}{1-z} - \left(1 + \frac{\delta_z(\phi_2 + 1, \phi_1 - 1)}{\phi_1 - 1}\right)^{-1} \\ &= \frac{1}{1-z} - \frac{\phi_1 - 1}{\phi_1 - 1 + \delta_z(\phi_2 + 1, \phi_1 - 1)}.\end{aligned}$$

As $z \to 0$, it follows by Theorem A.4 that

$$\begin{aligned}\lim_{z\to 0} \frac{z\, I_z(\phi_2, \phi_1)}{(1-z)\, I_z(\phi_2 + 1, \phi_1)} &= 1 - \frac{\phi_1 - 1}{\phi_1 + \phi_2} \\ &= \frac{\nu_2 + 2}{\nu_1 + \nu_2}.\end{aligned}$$

Therefore, the limiting value of the estimate $\tilde{\tau}_2^{**}(s, t_2)$ is, for fixed t_1,

$$\lim_{z \to 0} \frac{t_1}{\nu_2} \frac{z\, I_z(\phi_2, \phi_1)}{(1-z)\, I_z(\phi_2 + 1, \phi_1)} = \frac{\nu_2 + 2}{\nu_2\,(\nu_1 + \nu_2)}\, t_1.$$

This proves (a). To prove part (b), fix t_2 and rewrite the estimate $\tilde{\tau}_2^{**}(s, t_2)$ as

$$\tilde{\tau}_2^{**}(s, t_2) = \frac{t_2}{\nu_2 - 2\delta_z(\phi_2, \phi_1)}.$$

As $z \to 1$, $\delta_z(\phi_2, \phi_1) \to 0$ by Theorem A.4, from which the result follows. □

From this result, the limit of $\tilde{\tau}_2^{**}(s, t_2)$ as s gets large is the same as the corresponding limit of unbiased and ANOVA estimates of τ_2. However, as $s \to 0$, PDP-unbiased estimates of τ_2 do not quite shrink to the limiting REML estimate $t_1/(\nu_1 + \nu_2)$.

PDP-unbiased estimator of σ. The usual unbiased estimator of σ is $S/E(W)$. Given $s > 0$, $W^*(s)$ is the PDP for inference about σ. Replacing $E(W)$ with $E[W^*(s)]$ (given by (2.27)) in the unconditional estimator of σ, the PDP estimate is

$$\tilde{\sigma}^{**}(s) = \frac{s}{E(W^*(s))} = \frac{\nu_1 - 2}{\nu_2} \frac{I_z(\phi_2, \phi_1)\, s}{I_z(\phi_2 + 1, \phi_1 - 1)}.$$

Over all $s > 0$, the ensemble of estimates $\{\tilde{\sigma}^{**}(s) : s > 0\}$ produces the PDP-unbiased estimator $\tilde{\sigma}^{**}(S) = \phi^{**}(S)\, S$ of σ, where

$$\tilde{\sigma}^{**}(S) = \frac{\nu_1 - 2}{\nu_2} \frac{I_Z(\phi_2, \phi_1)}{I_Z(\phi_2 + 1, \phi_1 - 1)}\, S. \tag{2.61}$$

Because $\tilde{\sigma}^{**}(S)$ is of the form $\phi^{**}(S)\, S$, we need to be somewhat careful in expressing its limiting behavior.

Theorem 2.4. *The PDP-unbiased estimator of σ, $\tilde{\sigma}^{**}(S) = \phi^{**}(S)\, S$, has the following limiting behavior:*

(a) $\lim_{s \to 0} \tilde{\sigma}^{**}(s) = \dfrac{\nu_2 + 2}{\nu_2}$;

(b) $\lim_{s \to \infty} \dfrac{\tilde{\sigma}^{**}(s)}{s} = \lim_{s \to \infty} \phi^{**}(s) = \dfrac{\nu_1 - 2}{\nu_2}$.

Proof. To prove (a), begin with

$$\frac{I_z(\phi_2,\phi_1)\,s}{I_z(\phi_2+1,\phi_1-1)} = \frac{z\,I_z(\phi_2,\phi_1)}{(1-z)\,I_z(\phi_2+1,\phi_1-1)}$$
$$= \frac{I_z(\phi_2+1,\phi_1)-(1-z)\,I_z(\phi_2+1,\phi_1-1)}{(1-z)\,I_z(\phi_2+1,\phi_1-1)}$$

by Fact A.4. Then,

$$= \frac{1}{1-z}\,\frac{I_z(\phi_2+1,\phi_1)}{I_z(\phi_2+1,\phi_1-1)} - 1$$
$$= \frac{1}{1-z}\left(1+\frac{\delta_z(\phi_2+1,\phi_1-1)}{\phi_1-1}\right) - 1$$

by Theorem A.2. As $z \to 0$, the limiting value of the above expression is

$$\lim_{z\to 0}\frac{z\,I_z(\phi_2,\phi_1)}{(1-z)\,I_z(\phi_2+1,\phi_1-1)} = \left(1+\frac{\phi_2+1}{\phi_1-1}\right) - 1 = \frac{\nu_2+2}{\nu_1-2}.$$

Therefore,

$$\lim_{s\to 0}\frac{\nu_1-2}{\nu_2}\,\frac{I_z(\phi_2,\phi_1)\,s}{I_z(\phi_2+1,\phi_1-1)} = \frac{\nu_2+2}{\nu_2}.$$

This proves (a). To prove part (b), note that, for any fixed $s > 0$,

$$\phi^{**}(s) = \frac{\nu_1-2}{\nu_2}\,\frac{I_z(\phi_2,\phi_1)}{I_z(\phi_2+1,\phi_1-1)}.$$

As $z \to 1$, both Beta cdf's go to 1, leaving

$$\lim_{s\to\infty}\phi^{**}(s) = \frac{\nu_1-2}{\nu_2}.$$

This completes the proof. □

In this case, the limiting PDP estimate of σ does not shrink as far as the adjusted unbiased estimator as $s \to 0$. However, as s gets large, PDP-unbiased estimates converge to the unbiased estimates, as expected.

PDP-unbiased estimators of $J\sigma_\alpha^2$ and $J\gamma$. One can apply the plug-in principle to produce PDP-unbiased estimators of $J\sigma_\alpha^2$ and $J\gamma$. In the former case,

$$\begin{aligned} J\tilde{\sigma}_\alpha^2(T_1, T_2) &= \tilde{\tau}_2^{**}(S, T_2) - \tilde{\tau}_1^{**}(S, T_1) \\ &= \frac{T_2}{\nu_2 - 2\,\delta_Z(\phi_2, \phi_1)} - \frac{T_1}{\nu_1 + 2\,\delta_Z(\phi_2, \phi_1)}. \end{aligned} \tag{2.62}$$

Similarly, the PDP-unbiased estimator of $J\gamma = \sigma - 1$ is

$$J\tilde{\gamma}^{**}(S) = \tilde{\sigma}^{**}(S) - 1; \tag{2.63}$$

The limiting behavior of $J\tilde{\sigma}_\alpha^2$ and $J\tilde{\gamma}^{**}$ can be deduced from Theorems 2.2–2.4.

2.5.2.2 PDP-MLEs of τ_1 and τ_2

In this subsection, maximum likelihood estimators based on the ensemble of marginal PDP distributions of U_1^* and U_2^* are developed. The motivation for investigating such estimators is predicated on the idea that these marginal PDP densities, in some sense, account for the parameter constraint of the model. The resulting estimators are termed PDP-MLEs, and will be informally compared to REML estimators and PDP-unbiased estimators.

The PDP-MLEs derived below are actually REML in nature, in that the PDP distributions from which they emerge are derived solely from the random effects portion of the data represented by (T_1, T_2), and do not utilize the fixed effects portion of the data represented by $\overline{Y}_{...}$

PDP-MLE of τ_1. Let $U_1 = T_1/\tau_1$. Changing variables in the marginal PDP density (2.6) of U_1^* yields

$$g^*(t_1\,|\,s; \tau_1) = \frac{1}{\tau_1}\frac{g_{\nu_1}(t_1/\tau_1)\,G_{\nu_2}(st_1/\tau_1)}{I_z(\phi_2, \phi_1)}. \tag{2.64}$$

Treating (2.64) as a function of τ_1 for fixed t_1 and s produces a marginal (conditional) PDP likelihood function

$$L(\tau_1|t_1, s) \propto \frac{1}{\tau_1} g_{\nu_1}(t_1/\tau_1)\, G_{\nu_2}(st_1/\tau_1).$$

Since the logarithm is a 1–1 monotone function, it is convenient to work with the PDP log likelihood. Expanding $g(\cdot)$ and taking logs of the above expression yields

$$\log L(\tau_1|t_1, s) = -\phi_1 \log \tau_1 - \frac{t_1}{2\,\tau_1} + \log G_{\nu_2}\left(\frac{st_1}{\tau_1}\right), \tag{2.65}$$

where the constant terms are ignored. To find the maximum of (2.65), we first differentiate it with respect to τ_1:

$$\frac{d}{d\tau_1} \log L(\tau_1|t_1, s) = -\frac{\nu_1}{2\,\tau_1} + \frac{\beta}{2\,\tau_1} - \frac{s\beta}{\tau_1} \frac{g_{\nu_2}(s\beta)}{G_{\nu_2}(s\beta)},$$

where $\beta = t_1/\tau_1$.

Noting that $x g_\nu(x) = \nu\, g_{\nu+2}(x)$ when $g(\cdot)$ is a χ^2 density (by Lemma A.20), multiplying the above equation through by $-2\tau_1$ and setting the result to zero produces the estimating equation

$$\nu_1 - \hat{\beta} + 2\nu_2 \frac{g_{\nu_2+2}(s\hat{\beta})}{G_{\nu_2}(s\hat{\beta})} = 0. \tag{2.66}$$

This nonlinear equation is solved numerically for $\hat{\beta}$ at each fixed s. Over all $s > 0$, the ensemble of solutions $\{\hat{\beta}^{-1}(s) : s > 0\}$ produces a multiplier function $\phi_1^+(S)$ which, when multiplied by T_1, yields the PDP-MLE of τ_1. Its limiting values are the same as those of the REML and PDP-unbiased estimators of τ_1; the derivations are found in section A.7.

PDP-MLE of τ_2. Starting from the density (2.7) of $U_2(s)$ for some fixed $s > 0$, let $U_2 = T_2/\tau_2$. Changing variables in (2.7) yields

$$g^*(t_2|s; \tau_2) = \frac{1}{\tau_2} \frac{g_{\nu_2}(t_2/\tau_2)\,[1 - G_{\nu_1}(t_2/s\tau_2)]}{I_z(\phi_2, \phi_1)}. \tag{2.67}$$

Treating (2.67) as a function of τ_2 for fixed t_2 and s, a marginal PDP likelihood function is produced, of the form

$$L(\tau_2|t_2, s) \propto \frac{1}{\tau_2}\, g_{\nu_2}(t_2/\tau_2)\,[1 - G_{\nu_1}(t_2/s\tau_2)].$$

Taking the logarithm of the above expression and ignoring the constant terms, the PDP log likelihood of τ_2 is, for fixed $s > 0$,

$$\log L(\tau_2|t_2, s) = -\phi_2 \log \tau_2 - \frac{t_2}{2\,\tau_2} + \log\left[1 - G_{\nu_1}\left(\frac{t_2}{s\tau_2}\right)\right], \tag{2.68}$$

To find the maximum, we first differentiate (2.68) with respect to τ_2, yielding

$$\frac{d}{d\tau_2} \log L(\tau_2|t_2, s) = -\frac{\phi_2}{\tau_2} + \frac{\lambda}{2\,\tau_2} - \frac{w^*}{\tau_2}\,\frac{g_{\nu_1}(w^*)}{1 - G_{\nu_1}(w^*)},$$

where $\lambda = t_2/\tau_2$ and $w^* = \lambda/s$. Setting the right hand side to zero and multiplying through by $-2\tau_2$ yields the estimating equation

$$0 = \nu_2 - \hat{\lambda} + 2\nu_1\,\frac{g_{\nu_1+2}(\hat{\lambda}/s)}{1 - G_{\nu_1}(\hat{\lambda}/s)}, \tag{2.69}$$

which is solved numerically for $\hat{\lambda}$ at any fixed $s > 0$. The ensemble of solutions $\{\hat{\lambda}^{-1}(s) : s > 0\}$ produces a multiplier function $\phi_2^+(S)$, from which the PDP-MLE of τ_2 is expressible as

$$\hat{\tau}_2(S) = \phi_2^+(S)\,T_2 = \phi_2^+(S)\,ST_1. \tag{2.70}$$

Its limiting behavior is discussed in section A.7.

2.5.2.3 Other PDP-MLEs

The PDP-MLEs of τ_1 and τ_2 can be used to determine PDP-MLEs of σ, $J\gamma$ and $J\sigma_\alpha^2$ by the plug-in principle. We have the following:

(i) the PDP-MLE of $J\sigma_\alpha^2$ is given by

$$J\hat{\sigma}_\alpha^2(S) = \phi_2^+(S)\,T_2 - \phi_1^+(S)\,T_1; \tag{2.71}$$

(ii) the PDP-MLE of σ is

$$\hat{\sigma}(S) = \frac{\phi_2^+(S)}{\phi_1^+(S)} S; \tag{2.72}$$

(iii) the PDP-MLE of $J\gamma$ is

$$J\hat{\gamma}(S) = \hat{\sigma}(S) - 1. \tag{2.73}$$

Note that as s gets large, the PDP-ML estimators of σ and $J\gamma$ behave like the corresponding ANOVA estimators. In addition, PDP-MLEs can be viewed as 'smooth versions' of REML estimators in the balanced one-way random model.

2.5.2.4 Properties of multiplier functions

Before proceeding to the next subsection, we cite some simple results to establish the properties of the multiplier functions of REML and PDP point estimators.

Theorem 2.5. *Consider estimators of τ_1 of the form $\phi(S)\,T_1$; then,*

(a) the multiplier function $\phi_1^*(S)$ corresponding to the REML estimator is continuous and monotone nondecreasing in s;

(b) the multiplier function $\phi_1^+(S)$ corresponding to the PDP-MLE is continuous and monotone increasing in s;

(c) the multiplier function $\phi_1^{**}(S)$ corresponding to the PDP-unbiased estimator is continuous and monotone increasing in s.

Proof. To prove (a), observe that the multiplier function of the REML estimator of τ_1 can be written as

$$\phi_1^*(s) = \begin{cases} \dfrac{1+s}{\nu_1+\nu_2} & \text{if } s < \dfrac{\nu_2}{\nu_1}, \\[2ex] \dfrac{1}{\nu_1} & \text{if } s \geq \dfrac{\nu_2}{\nu_1}. \end{cases}$$

It is clear that $(1+s)/(\nu_1+\nu_2)$ is monotone increasing and continuous on $[0, \nu_2/\nu_1]$ and constant at ν_1^{-1} for all $s \geq \nu_2/\nu_1$. Since $s = \nu_2/\nu_1$ is a continuity point of ϕ_1, the continuity of $\phi_1^*(s)$ is established.

For (b), monotonicity of ϕ_1^+ follows from Theorem A.14(b), while its continuity follows from the continuity of the estimating equation (2.66) as a function of s.

Finally, to prove (c), the monotonicity of the PDP-unbiased multiplier $\phi_1^{**}(s)$ follows from Theorems A.13(b) and A.14(b), while its continuity follows from continuity of the Beta cdf ratio in (2.10). □

Corollary 2.5.1. *The function $\delta_z(\phi_2, \phi_1)$ is continuous and monotone decreasing in s (through z).*

Proof. By Theorem 2.5(c), $\phi_1^{**}(s) = 1/E_1(U_1|S \leq s)$ is monotone increasing and continuous for each fixed $s > 0$. By Corollary A.2.1, $E(U_1^*(s)) = E_1(U_1|S \leq s) = \nu_1 + 2\delta_z(\phi_2, \phi_1)$. Since $\delta(\cdot)$ is the only term in the above expectation that depends on s, it follows that $\delta(\cdot)$ is continuous and monotone decreasing in s through z. □

Theorem 2.6. *Consider estimators of τ_2 of the form $\phi(S)\,T_2$; then*

(a) the multiplier function $\phi_2^*(S)$ corresponding to the REML estimator is continuous and monotone nonincreasing in s;

(b) the multiplier function $\phi_2^+(S)$ corresponding to the PDP-MLE is continuous and monotone decreasing in s;

(c) the multiplier function $\phi_2^{**}(S)$ corresponding to the PDP-unbiased estimator is continuous and monotone decreasing in s.

Theorem 2.7. *Among estimators of σ of the form $\phi(S)\,S$,*

(a) the multiplier function $\phi^*(S)$ corresponding to the REML estimator is continuous and monotone nonincreasing in s;

(b) the multiplier function $\phi^{+}(S)$ corresponding to the PDP-MLE is continuous and monotone decreasing in s;

(c) The multiplier function $\phi^{**}(S)$ corresponding to the PDP-unbiased estimator is continuous and monotone decreasing in s.

The proofs of Theorems 2.6 and 2.7 are parallel to the proof of Theorem 2.5. One of the deductions that follow from these results is that PDP-MLEs can be viewed as 'smooth versions' of REML estimators in the balanced one-way random model.

2.5.3 Unconditional evaluation of point estimators

Having developed two types of PDP point estimators for τ_1, τ_2 and σ, we turn to evaluation of their unconditional risk performance in relation to the corresponding unconditional unbiased estimators. Entropy (or Kullback-Leibler) loss is chosen as the basis of comparison.

With respect to an invariant decision problem, we begin by showing that the usual unbiased estimators are best in the class of $\mathscr{G}$-equivariant estimators, where $\mathscr{G}$ is the group of affine transformations. Adaptation of Stein's argument to the one-way random model yields the REML estimators of τ_1 and τ_2 as *Stein-like* estimators. Comparison of risk functions with respect to entropy loss will show that the REML estimators dominate the usual unbiased estimators in risk performance.

From our perspective, however, the critical observation is that PDP-unbiased and PDP-ML estimators also outperform the usual unbiased estimators in terms of risk. Moreover, the proof of this claim, and the construction of the estimators, is predicated on a Brewster-Zidek type of argument. *In particular, for the PDP-unbiased estimators, it is important to observe that the conditioning event that results from the B-Z argument is precisely the same as that obtained from the PDP argument.*

Definition 2.1. Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two estimators of a parameter θ, and suppose their risk function depends on τ. We say that

(a) $\hat{\theta}_2$ is *at least as good as* $\hat{\theta}_1$ if $R(\hat{\theta}_2;\tau) \leq R(\hat{\theta}_1;\tau)$ for all τ;

(b) $\hat{\theta}_2$ *dominates* $\hat{\theta}_1$ if $R(\hat{\theta}_2;\tau) \leq R(\hat{\theta}_1;\tau)$ for all τ, and $R(\hat{\theta}_2;\tau) < R(\hat{\theta}_1;\tau)$ for at least one value of τ; and

(c) $\hat{\theta}_2$ *strictly dominates* $\hat{\theta}_1$ if $R(\hat{\theta}_2;\tau) < R(\hat{\theta}_1;\tau)$ for all τ.

Loss functions of interest. Point estimators of τ_1, τ_2 and σ are often compared with respect to one of the following bowl-shaped loss functions: normalized quadratic loss (NQL), defined as

$$L(d,\theta) = \left(\frac{d}{\theta} - 1\right)^2, \tag{2.74}$$

and entropy loss, defined as

$$L(d,\theta) = \frac{d}{\theta} - 1 - \log\frac{d}{\theta}. \tag{2.75}$$

Although NQL has been traditionally used to compare variance estimators, we restrict attention in this thesis to evaluations with respect to entropy loss, a member of the family of bowl-shaped loss functions. Moreover, this loss function is convex, as evidenced by Figure 2.14 and proved in Lemma A.18. Brown (1968) showed that entropy loss *uniquely* produces the best unbiased estimator as the best $\mathscr{G}$-equivariant estimator. In the balanced one-way random model, entropy loss is employed to search for improvements to the usual unbiased estimators. We will be concerned with point estimators of the general form $\phi(S)\,T$, where $\phi(S)$ is a multiplier function that depends on the sum of squares ratio S, and T is a statistic on which the usual estimators of τ_1, τ_2 or σ are based. The choice of an (improved) estimator in a given problem is tied to the group structure of the invariant decision problem, the MLR properties

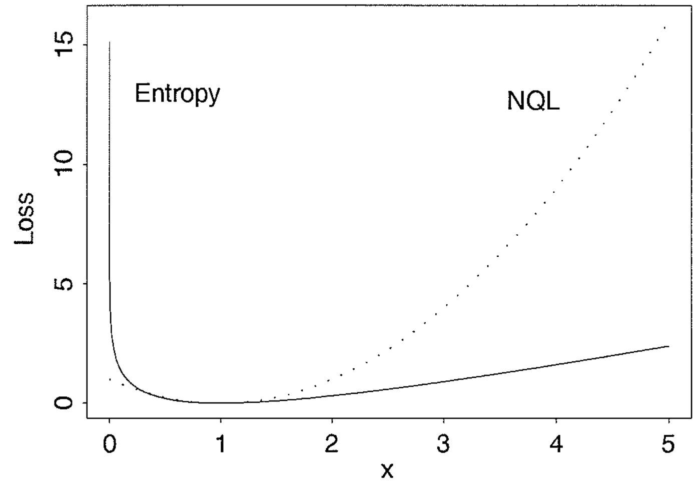

Figure 2.14: Graphs of the normalized quadratic loss (NQL) and entropy loss functions, as a function of $x = d/\theta$.

of the appropriate family of reference distributions, and the bowl-shaped property of the risk function.

2.5.4 Point estimation of τ_1

Best $\mathscr{G}$-equivariant estimator of τ_1. Consider point estimation of τ_1 under the entropy loss function (2.75). We begin with the invariant decision problem under $\mathscr{G}$, the group of affine transformations. $\mathscr{G}$-equivariant estimators of τ_1 are of the form kT_1, the risk functions of such estimators are constant, and the best $\mathscr{G}$-equivariant estimator corresponds to that value of k which minimizes the risk. Differentiating the risk with respect to k yields

$$\begin{aligned}\frac{d}{dk}R(kT_1;\tau_1) &= \frac{d}{dk}\left(E(kU_1) - 1 - E(\log kU_1)\right)\\ &= \frac{d}{dk}\left[kE(U_1) - 1 - \log k - E(\log U_1)\right]\\ &= E(U_1) - 1/k.\end{aligned}$$

Setting the last result equal to zero, the solution is $k = 1/E(U_1) = 1/\nu_1$, which is easily seen to be a minimum since the second derivative of the risk function is k^{-2}, which is positive for all $k > 0$. Therefore, the best $\mathscr{G}$-equivariant estimator of τ_1 under entropy loss is precisely the unbiased estimator $\tilde{\tau}_1 = T_1/\nu_1$.

Stein-like estimator of τ_1. Let $\mathscr{H}$ be the scale subgroup of $\mathscr{G}$, and consider $\mathscr{H}$-equivariant estimators of the form $\phi(S)T_1$. The optimal estimator (were it to exist) would minimize the risk

$$R(\phi(S)T_1;\tau_1,\sigma) = E_\sigma[E_\sigma[(\phi(S)\,U_1 - 1 - \log\phi(S) - \log U_1)|S]].$$

If the true value of σ were available, the optimal multiplier function would obtain by minimizing the conditional risk. Thus, for each fixed s, $\phi_\sigma(s)$ is the value of k that minimizes

$$E_\sigma[k\,U_1 - 1 - \log k - \log U_1|S = s]. \tag{2.76}$$

But

$$\frac{d}{dk}\,E_\sigma[k\,U_1 - 1 - \log k - \log U_1|S = s] = E_\sigma[U_1|S = s] - k^{-1};$$

setting the right hand side of this equation to zero yields as its solution

$$\phi_\sigma(s) = \frac{1}{E_\sigma(U_1|S = s)} = \frac{1 + s/\sigma}{\nu_1 + \nu_2}.$$

There are two key points to be made about this multiplier function:

1. Since $\phi_\sigma(s)$ is a function of σ, it is not possible to obtain a best $\mathscr{H}$-equivariant estimator of τ_1.

2. $\phi_\sigma(s)$ is linearly increasing in s for fixed σ, and decreasing in σ for fixed s. Thus, for each fixed $s \geq 0$ and all $\sigma \geq 1$, we have $\phi_\sigma(s) \leq \phi_1(s)$. Moreover, if $s < \nu_2/\nu_1$, then $\phi_1(s) < \nu_1^{-1}$. Since the conditional risk (2.76) is a bowl-shaped function of k (in fact, convex), if we define

$$\phi_1^*(S) = \min\left[\frac{1+S}{\nu_1+\nu_2}, \frac{1}{\nu_1}\right] = \begin{cases} \dfrac{1+S}{\nu_1+\nu_2} & \text{if } S < \dfrac{\nu_2}{\nu_1}, \\ \dfrac{1}{\nu_1} & \text{if } S \geq \dfrac{\nu_2}{\nu_1}, \end{cases} \tag{2.77}$$

then $\phi_1^*(S)T_1$ dominates T_1/ν_1, where the estimator can be written as

$$\phi_1^*(S)\,T_1 = \min\left[\frac{T_1+T_2}{\nu_1+\nu_2}, \frac{T_1}{\nu_1}\right] = \begin{cases} \dfrac{T_1+T_2}{\nu_1+\nu_2} & \text{if } \dfrac{T_2}{\nu_2} \leq \dfrac{T_1}{\nu_1}, \\ \dfrac{T_1}{\nu_1} & \text{if } \dfrac{T_2}{\nu_2} > \dfrac{T_1}{\nu_1}. \end{cases} \tag{2.78}$$

Comparison with section 3.8a-iii of Searle, Casella and McCulloch (1992, p. 92) reveals that this *Stein-like* estimator is precisely the REML estimator $\hat{\tau}_1^R$ of τ_1 in the balanced one-way random model, given by (2.53).

This derivation of the dominance of the REML estimator parallels the argument used by Brewster and Zidek (1974) to show that Stein's estimator dominates the best affine equivariant estimator of the normal variance. Hence, what has been shown is an illustration of Brewster and Zidek's first method of improving estimators. In summary, we have proved the following:

Theorem 2.8. *The REML estimator of τ_1 strictly dominates the best unbiased estimator of τ_1 under entropy loss.*

Step-function estimators of τ_1. Again, consider the class of $\mathscr{H}$-equivariant estimators of the form $\phi(S)T_1$. Following the argument of Brown (1968), consider an estimator of the form $\phi_c(S)\,T_1$, where for an arbitrarily chosen $r > 0$,

$$\phi_c(S) = \begin{cases} c & \text{if } S \le r, \\ \nu_1^{-1} & \text{if } S > r. \end{cases} \tag{2.79}$$

Under entropy loss, the risk of such an estimator is

$$\begin{aligned} R(\phi_c(S)\,T_1; \sigma) &= E_\sigma[cU_1 - 1 - \log c - \log U_1 | S \le r]\, P_\sigma(S \le r) \\ &\quad + E_\sigma[U_1/\nu_1 - 1 + \log \nu_1 - \log U_1 | S > r]\, P_\sigma(S > r). \end{aligned} \tag{2.80}$$

Differentiating with respect to c yields

$$\frac{d}{dc} R(\phi_c(S)\,T_1; \sigma) = [E_\sigma(U_1 | S \le r) - c^{-1}]\, P_\sigma(S \le r).$$

Setting the right hand side of the above expression to zero yields the optimum multiplier

$$c_\sigma(r) = \frac{1}{E_\sigma(U_1 | S \le r)}.$$

Since $c_\sigma(r)$ depends on σ, however, it cannot be used as a multiplier in (2.79).

For any fixed $r > 0$, it follows that $c_\sigma(r) \le c_1(r)$ for all $\sigma \ge 1$ by Theorem A.14(a). Moreover, by Theorem A.14(c), $c_1(r) < \nu_1^{-1}$ for any $r > 0$. Therefore, for fixed $r > 0$,

$$c_\sigma(r) \le c_1(r) < \frac{1}{\nu_1} \tag{2.81}$$

uniformly in σ. The *Brown-like estimator* of τ_1 is then $\phi_B(S)\,T_1$, where

$$\phi_B(S) = \begin{cases} c_1(r) & \text{if } S \le r \\ \nu_1^{-1} & \text{if } S > r. \end{cases} \tag{2.82}$$

As in the previous theorem, since the conditional risk is a bowl-shaped (in fact, convex) function by Lemma A.19, we have proved:

Theorem 2.9. *For any fixed $r > 0$, the Brown-like estimator of τ_1 strictly dominates the best unbiased estimator of τ_1 under entropy loss.*

Keeping r fixed, one can choose $r' < r$ and repeat the above argument to produce an estimator $\phi_1^{(2)}(S)\,T_1$, where

$$\phi_1^{(2)}(S) = \begin{cases} c_1(r') & \text{if } S \le r' \\ c_1(r) & \text{if } r' < S \le r \\ \nu_1^{-1} & \text{if } S > r. \end{cases}$$

This estimator strictly dominates both $\phi_B(S)\,T_1$ and T_1/ν_1. This process can be iterated by selecting successively smaller constants $r'', r''', r^{(4)}, \ldots$ to produce estimators $\phi_1^{(k)}(S)\,T_1$, $k = 2, 3, 4, \ldots$, with progressively better risk performance than their predecessors.

BZ-like estimator of τ_1. Following the process in section 1.4.2, construct a sequence of partitions $\{R_i\}_{i=1}^{\infty}$ of $[0, \infty)$ which have the properties described on page 30. In partition R_i, construct the step-function estimator $\phi^{(i)}(S)\,T_1$, where

$$\phi^{(i)}(S) = \begin{cases} c_1(r_{ij}) & \text{if } r_{i,j-1} \le S \le r_{ij} \\ \nu_1^{-1} & \text{if } S > r_{i,n_i-1}, \end{cases}$$

$j = 1, 2, \ldots, n_i - 1$ and $c_1(r_{ij}) = 1/E(U_1|S \le r_{ij})$. By Theorem 2.9 and the argument in the preceding paragraph, the estimator $\phi^{(i)}(S)\,T_1$ strictly dominates the best unbiased estimator T_1/ν_1 for each i. However, the estimators $\phi^{(i)}(S)\,T_1$ do not necessarily dominate one another since they are based on different partitions of $[0, \infty)$. Now, under conditions (i)–(iv) on page 30, it follows that

$$\lim_{i \to \infty} \phi^{(i)}(s) = \phi_1^{**}(s),$$

where $\phi_1^{**}(s) = 1/E_1(U_1|W \le s)$, pointwise in s. The resulting estimator $\phi_1^{**}(S)\,T_1$ is called the *Brewster-Zidek-like* estimator of τ_1 under entropy loss. For each $s > 0$, its multiplier function is the reciprocal of $E[U_1^*(s)]$, the PDP mean function of U_1^*. *In other words, under entropy loss, the BZ-like estimator* $\phi_1^{**}(S)\,T_1$ *of* τ_1*, produced by decision theoretic arguments, is precisely the PDP-unbiased estimator of* τ_1*!*

Theorem 2.10. *The PDP-unbiased estimator of* τ_1 *is at least as good as the best unbiased estimator of* τ_1 *under entropy loss.*

Proof. As shown above, the PDP-unbiased estimator of τ_1 is equal to $\phi_1^{**}(S)\,T_1$, where $\phi_1^{**}(s) = \lim_{i\to\infty}\phi^{(i)}(s)$ pointwise in s. Therefore, by Fatou's lemma, for each $\sigma \geq 1$,

$$\begin{aligned} R(\phi_1^{**}(S)\,T_1;\sigma) &= R(\lim_{i\to\infty}\phi_1^{(i)}(S)\,T_1;\sigma) \\ &\leq \liminf_{i\to\infty} R(\phi_1^{(i)}(S)\,T_1;\sigma) \\ &\leq R(T_1/\nu_1;\sigma), \end{aligned}$$

where the last inequality follows because, for each i, $\phi_1^{(i)}(S)\,T_1$ strictly dominates the best unbiased estimator of τ_1. □

The method of proof in Theorem 2.10 does not guarantee that $\phi^{**}(S)\,T_1$ strictly dominates T_1/ν_1. However, we would be extremely surprised if the risk functions of the two (very different) estimators were identically equal to one another for all $\sigma \geq 1$. In particular examples, it is easy to see that when $\sigma > 1$, the risk function of the PDP-unbiased estimator of τ_1 lies strictly below that of the unbiased estimator of τ_1.

Figure 2.15 compares multiplier functions of the four estimators of τ_1 derived under entropy loss for the case $(\nu_2, \nu_1) = (3, 4)$. The multiplier of the unconditional estimator is the reciprocal of its mean $E(U_1)$. In contrast, the multiplier function of the PDP-unbiased estimator is the reciprocal of $E(U_1^*(s))$ for each $s \geq 0$. The multiplier function of the REML estimator increases linearly from $(\nu_1 + \nu_2)^{-1}$ at $s = 0$ to ν_1^{-1} at $s = \nu_2/\nu_1$, after which it remains constant. This behavior reflects the pooling of sums of squares and degrees of freedom to estimation of σ_e^2 when the observed F ratio is less than one. Finally, the multiplier function of the PDP-MLE of τ_1 appears to compromise between the REML and PDP-unbiased multiplier functions.

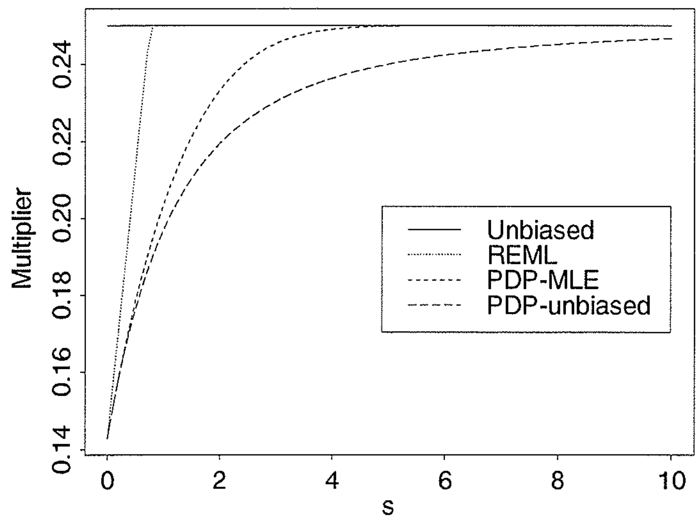

Figure 2.15: Multiplier functions of the best unbiased, REML, PDP-MLE and PDP-unbiased estimators of τ_1, $(\nu_2, \nu_1) = (3, 4)$.

Figure 2.16 compares the risk functions of the best unbiased, REML, PDP-MLE and PDP-unbiased estimators for a particular example in which $(\nu_1, \nu_1) = (3, 4)$. The (Stein-like) REML estimator, PDP-MLE and (BZ-like) PDP-unbiased estimators all have risk functions below that of the best unbiased estimator of τ_1. The risk function of the PDP-MLE falls roughly between those of the PDP-unbiased and REML estimators, tending closer to the PDP-unbiased estimator in this example. The risk of the PDP-unbiased estimator at $\sigma = 1$ is the same as the risk of T_1/ν_1, but has strictly lower risk for $\sigma > 1$ in this example; in the limit as $\sigma \rightarrow \infty$, the two risks converge. Although all three alternative estimators dominate $\hat{\tau}_1$ in risk performance in this example, it is apparent that the PDP-MLE, REML and PDP-unbiased estimators do not dominate one another.

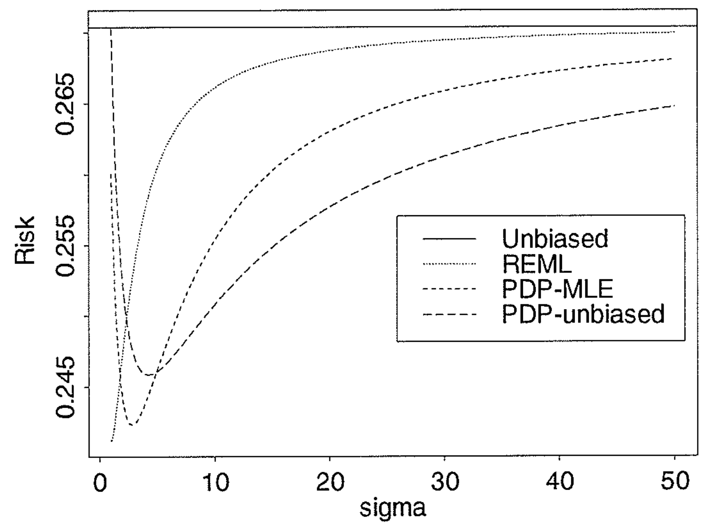

Figure 2.16: Risk functions of the best unbiased, REML, PDP-ML and PDP-unbiased estimators of τ_1 under entropy loss, $(\nu_2, \nu_1) = (3, 4)$.

We have not formally proven that the PDP-MLE of τ_1 is at least as good as the best unbiased estimator in risk, but the evidence in Figure 2.16 suggests that this should be the case. In fact, as long as one can show that $\phi_1^{**}(s) \leq \phi_1^{\circ}(s) < \nu_1^{-1}$ pointwise in s, where $\phi_1^{\circ}(S)$ is a monotone, continuous multiplier function, the method of proof used for the PDP-unbiased estimator could be extended to produce similar results for estimators such as the PDP-MLE. Furthermore, we have not proved dominance of the PDP point estimators of τ_1 over the best unbiased estimator, but Figure 2.16 suggests that this should hold as well.

2.5.5 Point estimation of τ_2

The process of obtaining point estimators of τ_2 is essentially the same as for τ_1. The invariant decision problem is nearly identical, except that the underlying families of distributions pertain to those associated with pivotals involving U_2 rather than U_1. As the process of obtaining improved point estimators is virtually unchanged, the focus is to report the estimators and the decision theory results.

Best $\mathscr{G}$-equivariant estimator of τ_2. Consider the invariant decision problem with respect to point estimation of τ_2 under entropy loss. Begin with the full affine transformation group $\mathscr{G}$ and the family of $\chi^2_{\nu_2}$ distributions. Following the argument on p. 98, the best $\mathscr{G}$-equivariant estimator of τ_2 under entropy loss is the unbiased estimator $\hat{\tau}_2 = T_2/\nu_2$, where $k = 1/E(U_2) = 1/\nu_2$ is the minimum risk multiplier.

Stein-like estimator of τ_2. Restricting attention to the scale subgroup $\mathscr{H}$, scale-equivariant estimators of τ_2 are of the form $\phi(S)T_2$. The optimal multiplier, which depends on σ, would minimize the risk

$$R(\phi(S)T_2; \tau_1, \sigma) = E_\sigma[E_\sigma[\phi(S)\,U_2 - 1 - \log\phi(S) - \log U_2|S]].$$

Following the argument starting on p. 99, the optimum multiplier function would be

$$\phi_\sigma(S) = \frac{1}{E_\sigma(U_2|S)} = \frac{1+\sigma/S}{\nu_1+\nu_2},$$

which is increasing in σ for fixed $s > 0$ and decreasing in s for fixed $\sigma \geq 1$. By Theorem A.14(d), it follows that $\phi_\sigma(s) \geq \phi_1(s)$ uniformly in σ for each fixed s. Moreover, if $s < \nu_2/\nu_1$, then $\phi_1(s) > \nu_2^{-1}$. Hence, if we define

$$\phi_2^*(S) = \begin{cases} \dfrac{1+1/S}{\nu_1+\nu_2} & S \leq \dfrac{\nu_2}{\nu_1} \\ \dfrac{1}{\nu_2} & S > \dfrac{\nu_2}{\nu_1}, \end{cases} \tag{2.83}$$

then $\phi_2^*(S)\,T_2$, a Stein-like estimator of τ_2, strictly dominates the best unbiased estimator of τ_2; as in the τ_1 case, $\phi_2^*(S)\,T_2$ is precisely the REML estimator of τ_2.

Theorem 2.11. *The REML estimator of τ_2 strictly dominates the best unbiased estimator of τ_2 under entropy loss.*

Step-function estimators of τ_2. The line of argument follows that developed for estimation of τ_1. In the class of $\mathscr{H}$-equivariant estimators of τ_2 of the form $\phi(S)T_2$, begin by considering those of the form $\phi_c(S)\,T_2$, where $\phi_c(S)$ is of the form (2.79) with ν_2 replacing ν_1. Its risk function is

$$\begin{aligned} R(\phi_c(S)\,T_2;\sigma) &= E_\sigma[cU_2 - 1 - \log c - \log U_2 | S \le r]\,P_\sigma(S \le r) \\ &\quad + E_\sigma[U_2/\nu_2 - 1 + \log \nu_2 - \log U_2 | S > r]\,P_\sigma(S > r). \end{aligned} \tag{2.84}$$

Differentiating this risk with respect to c yields

$$\frac{d}{dc} R(\phi_c(S)\,T_2;\sigma) = [E_\sigma(U_2|S \le r) - c^{-1}]\,P_\sigma(S \le r);$$

setting the right hand side of this result to zero, the solution $c = c^*$ is the optimum (i. e., minimum risk) multiplier

$$c^*_\sigma(r) = \frac{1}{E_\sigma(U_2|S \le r)}.$$

By Theorem A.14(d), it follows that $c^*_\sigma(r) \ge c^*_1(r)$ for any fixed $r > 0$. By Theorem A.14(f), $c^*_1(r) > \nu_2^{-1}$ for any $r > 0$. Therefore,

$$c^*_\sigma(r) \ge c^*_1(r) > \frac{1}{\nu_2}. \tag{2.85}$$

Since (2.84) is bowl-shaped, the Brown-type estimator $\phi^*_B(S)\,T_2$ of τ_2, where

$$\phi^*_B(S) = \begin{cases} c^*_1(r) & \text{if } S \le r \\ \nu_2^{-1} & \text{if } S > r, \end{cases}$$

strictly dominates the best $\mathscr{G}$-equivariant estimator. We have therefore proved:

Theorem 2.12. *For any fixed $r > 0$, the Brown-like estimator of τ_2 strictly dominates the best unbiased estimator of τ_2 under entropy loss.*

BZ-like estimator of τ_2. Following the argument starting on p. 102, construct a sequence of partitions $\{R_i\}_{i=1}^{\infty}$ of $[0,\infty)$. In each partition R_i, construct the step-function estimator $\phi^{(i)}(S)\,T_2$, where

$$\phi^{(i)}(S) = \begin{cases} c_1^*(r_{ij}) & \text{if } r_{i,j-1} \le S \le r_{ij} \\ \nu_2^{-1} & \text{if } S > r_{i,n_i-1}, \end{cases} \tag{2.86}$$

with $j = 1, 2, \ldots, n_i - 1$ and $c_1^*(r_{ij}) = 1/E(U_2|S \le r_{ij})$. In this situation, $\phi^{(i)}(s) > \nu_2^{-1}$ pointwise in s for each $i = 1, 2, 3, \ldots$ by Theorem A.14(f). As before, the estimators $\phi^{(i)}(S)\,T_2$ do not dominate one another, but each of them strictly dominates the best unbiased estimator T_2/ν_2. Now, under the conditions (i)–(iv) on page 30,

$$\lim_{i\to\infty} \phi^{(i)}(S) = \phi_2^{**}(S),$$

where $\phi_2^{**}(s) = 1/E_1(U_2|W \le s)$ pointwise in s. The limiting estimator $\phi_2^{**}(S)\,T_2$ is the *Brewster-Zidek-like* estimator of τ_2 under entropy loss, which is precisely the PDP-unbiased estimator of τ_2.

Theorem 2.13. *The PDP-unbiased estimator of τ_2 is at least as good as the best unbiased estimator of τ_2 under entropy loss.*

The proof is analogous to that of Theorem 2.10.

Multiplier functions of the four estimators considered above are shown in Figure 2.17 for the case $(\nu_2, \nu_1) = (3, 4)$. In this case, the multiplier functions of the REML and PDP estimators inflate towards infinity as $s \to 0$. Once again, the multiplier function of the PDP-MLE tends to fall somewhere between those of the REML and PDP-unbiased estimators.

The risk functions corresponding to these estimators are displayed in Figure 2.18 for the (3,4) case. The similarity in shape between the risks of these estimators with the corresponding τ_1 estimators in Figure 2.16 is noteworthy.

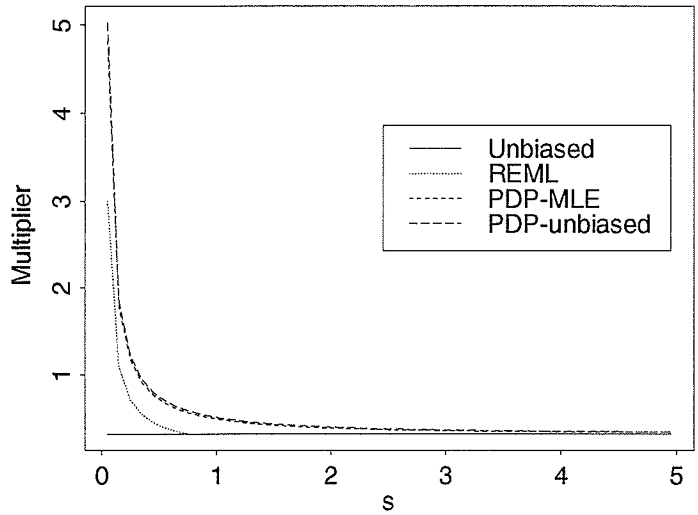

Figure 2.17: Multiplier functions of the best unbiased, REML PDP-ML and PDP-unbiased estimators of τ_2, $(\nu_2, \nu_1) = (3, 4)$.

2.5.6 Point estimation of σ

The point estimation problem with regard to σ is somewhat different from the τ_1 and τ_2 cases. To begin with, the parameter space associated with σ is $\Theta = [1, \infty)$. We will also find that estimators with improved risk properties are driven by the parameter constraint $\sigma \geq 1$ in some fashion. Since PDP-MLEs of σ are biased, they will not be included in this subsection.

Best unbiased estimator of σ. Applying the invariant decision problem once again, begin with the affine group $\mathscr{G}$. A $\mathscr{G}$-equivariant estimator of σ is of the form $k\,S$, with (constant) risk function

$$R(k\,S; \sigma) = E_\sigma[k\,W - 1 - \log k - \log W].$$

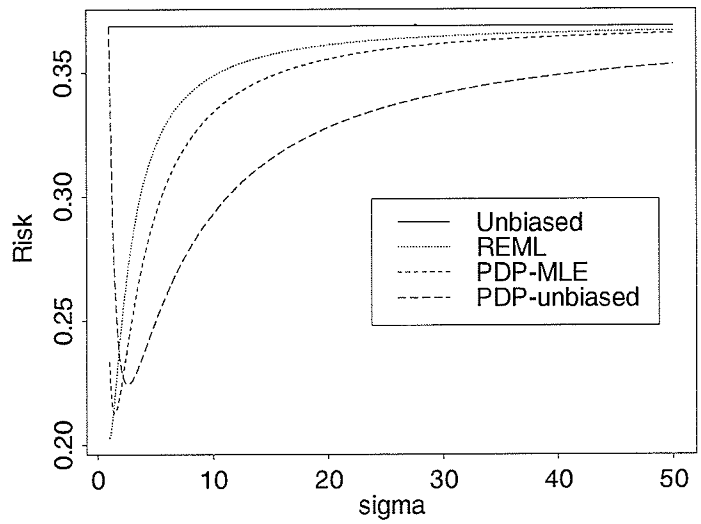

Figure 2.18: Risk functions of the best unbiased, REML, PDP-ML and PDP-unbiased estimators of τ_2 under entropy loss, when $(\nu_2, \nu_1) = (3, 4)$.

Following the argument on p. 98, the best $\mathscr{G}$-equivariant estimator of σ under entropy loss is the unbiased estimator $\hat{\sigma} = (\nu_1 - 2)S/\nu_2$, where $k = 1/E(W) = (\nu_1 - 2)/\nu_2$ is the minimum risk multiplier. Observe that this estimator can assume values less than 1 if $S < \nu_2/(\nu_1 - 2)$.

Adjusted unbiased estimator of σ. Next, consider the class of $\mathscr{H}$-equivariant estimators of the form $\phi(S)S$. A simple means of improving upon the best unbiased estimator of σ is to modify it so that it always assumes values no less than 1, analogous to restricted maximum likelihood. The 'adjusted unbiased estimator' of σ is then of the form $\phi^*(S)\,S$, where

$$\phi^*(S)\,S = \max\left[\frac{\nu_1 - 2}{\nu_2}\,S,\, 1\right]. \tag{2.87}$$

Thus, for all $s > 0$, $\phi^*(s) \geq (\nu_1-2)/\nu_2$, with strict inequality whenever $s < \nu_2/(\nu_1-2)$. It is clear that such an estimator will dominate the best unbiased estimator under entropy loss, since for any $s < \nu_2/(\nu_1-2)$, the estimator is moved to the boundary of the parameter space $[1, \infty)$. Since the risk function based on entropy loss is strictly bowl-shaped, we have proven that:

Theorem 2.14. *The 'adjusted unbiased estimator' of σ strictly dominates the best unbiased estimator of σ under entropy loss.*

Step-function estimators of σ. In the class of $\mathscr{H}$-equivariant estimators of σ, begin by considering those of the form $\phi_c(S)\,S$, where for any fixed $r > 0$, $\phi_c(S)$ is described by (2.79), with ν_1^{-1} being replaced by $(\nu_1-2)/\nu_2$. The risk function of this estimator is then

$$\begin{aligned} R(\phi_c(S)\,S;\sigma) &= E_\sigma\Big[c\,W - 1 - \log c - \log W \,\Big|\, W \leq \frac{r}{\sigma}\Big]\, P_\sigma\left(W \leq \frac{r}{\sigma}\right) \\ &\quad + E_\sigma\Big[\frac{(\nu_1-2)\,W}{\nu_2} - 1 - \log\frac{(\nu_1-2)}{\nu_2} - \log W \,\Big|\, W > \frac{r}{\sigma}\Big]\, P_\sigma\left(W > \frac{r}{\sigma}\right). \end{aligned} \tag{2.88}$$

This follows since the event $S \leq r$ is equivalent to $W \leq r/\sigma$; furthermore, the MLR properties in this case are correctly associated with the distribution of $W|W \leq r/\sigma$. Now, differentiating this risk with respect to c yields

$$\frac{d}{dc}\,R(\phi_c(S)\,S;\sigma) = [E_\sigma(W|W \leq r/\sigma) - c^{-1}]\,P(W \leq r/\sigma).$$

Setting the right hand side of this result to zero yields the optimum multiplier

$$c_\sigma^*(r) = \frac{1}{E_\sigma(W|W \leq r/\sigma)}.$$

By Theorem A.15(a), it follows that $c_\sigma^*(r) \geq c_1^*(r)$ for any fixed $r > 0$. By Theorem A.15(c), $c_1(r) > (\nu_1-2)/\nu_2$ for any $r > 0$. Therefore,

$$c_\sigma^*(r) \geq c_1^*(r) > \frac{\nu_1-2}{\nu_2} \tag{2.89}$$

for all $r > 0$ uniformly in σ. Since (2.88) is bowl-shaped, the Brown-type estimator $\phi_B^*(S)\,S$ of σ, where

$$\phi_B^*(S) = \begin{cases} c_1^*(r) & \text{if } S \leq r \\ \dfrac{\nu_1 - 2}{\nu_2} & \text{if } S > r, \end{cases}$$

and $c_1^*(r) = 1/E_1(W|W \leq r/\sigma) = 1/E(S|S \leq r)$, strictly dominates the best unbiased estimator of σ.

Theorem 2.15. *For any fixed $r > 0$, the Brown-like estimator of σ strictly dominates the best unbiased estimator under entropy loss.*

BZ-like estimator of σ. The process of developing a sequence of step-function estimators whose limit leads to the BZ-like estimator of σ is very similar to that shown for the τ_2 estimator. Let $\{R_i\}_{i=1}^{\infty}$ denote a sequence of partitions of $[0, \infty)$ with properties (i)–(iv) listed on p. 30. In partition R_i, the multiplier function of the step-function estimator is of the form $\phi^{(i)}(S)\,S$, where

$$\phi^{(i)}(S) = \begin{cases} c_1^*(r_{ij}) & \text{if } r_{i,j-1} < S \leq r_{ij} \\ \dfrac{\nu_1 - 2}{\nu_2} & \text{if } S > r_{i,n_i-1}, \end{cases}$$

where

$$c_1(r_{ij}) = \frac{1}{E(S|S \leq r_{ij})},$$

$j = 1, \dots, n_i - 1$, and $(\nu_1 - 2)/\nu_2$ if $S > r_{i,n_i-1}$. The orderings of multipliers and distributions are the same as in the τ_2 case. So, by inductive application of Theorem 2.15, $\phi^{(i)}(S)\,S$ strictly dominates the best unbiased estimator of σ on each partition R_i, $i = 1, 2, 3, \dots$ Finally, taking limits as $i \to \infty$,

$$\lim_{i \to \infty} \phi^{(i)}(S) = \phi^{**}(S),$$

where $\phi^{**}(s) = 1/E(S|S \leq s)$ pointwise in s. The limiting estimator $\phi^{**}(S)\,S$ is the *Brewster-Zidek-like* estimator of σ under entropy loss, which is precisely the PDP-unbiased estimator of σ.

Theorem 2.16. *The PDP-unbiased estimator of σ is at least as good as the best unbiased estimator of σ under entropy loss.*

Multiplier functions of the above estimators of σ are graphed for $(\nu_2, \nu_1) = (3, 5)$ in Figure 2.19, with the corresponding risk functions in Figure 2.20.

2.5.7 Generalized Bayes PDP-unbiased estimators

We begin by establishing the form of the Bayes estimator of a parameter θ under entropy loss, given in the following result.

Theorem 2.17. *Under the entropy loss function*

$$L(\hat{\theta}; \theta) = \frac{\hat{\theta}}{\theta} - 1 - \log \hat{\theta} + \log \theta, \tag{2.90}$$

the Bayes estimate of θ is

$$\hat{\theta} = \frac{1}{E(\theta^{-1}|x)}, \tag{2.91}$$

where x represents the observed response vector.

Proof. Let $\hat{\theta} = \hat{\theta}(x)$ be an estimate of θ, and let $\pi(\theta)$ denote a proper prior density on θ. By definition, the Bayes risk is given by

$$\begin{aligned} r_\pi(\hat{\theta}) &= \int_\Theta R(\hat{\theta}; \theta)\, \pi(\theta)\, d\theta \\ &= \int_\Theta \int_\mathcal{X} L(\hat{\theta}(x); \theta)\, f(x|\theta)\, \pi(\theta)\, dx\, d\theta \\ &= \int_\mathcal{X} \int_\Theta L(\hat{\theta}(x); \theta)\, \pi(\theta|x)\, h(x)\, d\theta\, dx, \end{aligned}$$

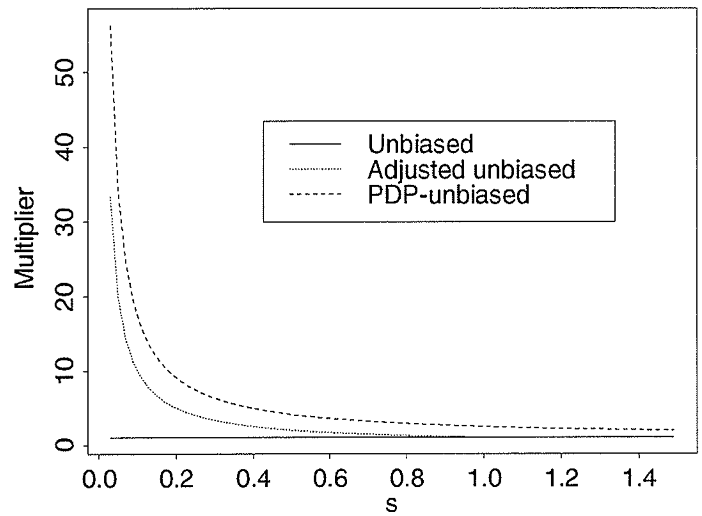

Figure 2.19: Multiplier functions of the best unbiased, adjusted best unbiased and PDP-unbiased estimators of σ, $(\nu_2, \nu_1) = (3, 5)$.

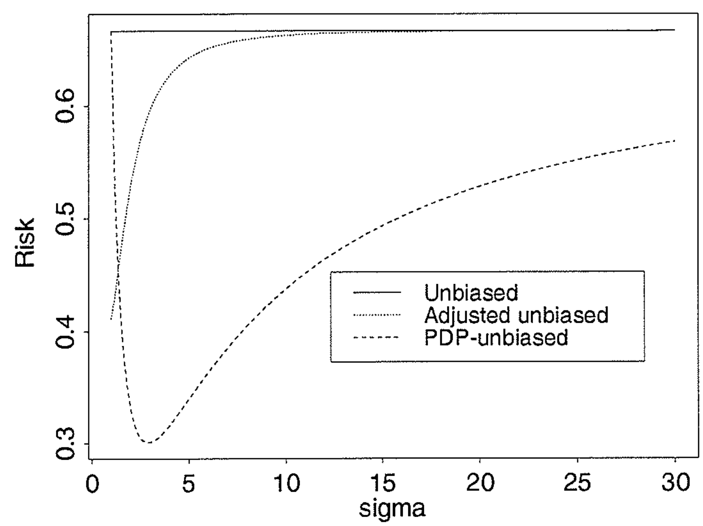

Figure 2.20: Risk functions of the best unbiased, adjusted best unbiased and PDP-unbiased estimators of σ under entropy loss, $(\nu_2, \nu_1) = (3, 5)$.

where $\pi(\theta|x)$ is the posterior density of θ, $h(x)$ is the marginal density of X, and the conditions allowing interchange of order of integration are in force. The last integral above can be expressed as

$$\int_{\mathcal{X}} \Big[\int_{\Theta} L(\hat{\theta}(x); \theta) \, \pi(\theta|x) \, d\theta \Big] h(x) \, dx,$$

the integral within brackets corresponding to the posterior risk, or expected posterior loss. For each x, the Bayes estimate can be found by minimizing this term. Under (2.90), the expected posterior loss is

$$\int_{\Theta} \Big(\frac{\hat{\theta}(x)}{\theta} - 1 - \log \hat{\theta}(x) + \log \theta \Big) \, \pi(\theta|x) \, d\theta.$$

Differentiating this with respect to $\hat{\theta}(x)$,

$$\frac{d}{d\hat{\theta}(x)} \int_{\Theta} \Big(\frac{\hat{\theta}(x)}{\theta} - 1 - \log \hat{\theta}(x) + \log \theta \Big) \, \pi(\theta|x) \, d\theta =$$

$$\int_{\Theta} \frac{d}{d\hat{\theta}(x)} \Big(\frac{\hat{\theta}(x)}{\theta} - 1 - \log \hat{\theta}(x) + \log \theta \Big) \, \pi(\theta|x) \, d\theta,$$

where the conditions allowing interchange of derivative and integral are assumed to hold. The resulting derivative is

$$\int_{\Theta} \Big(\frac{1}{\theta} - \frac{1}{\hat{\theta}(x)} \Big) \pi(\theta|x) \, d\theta.$$

Setting the above integral to zero and solving for $\hat{\theta}$ yields the solution

$$\frac{1}{\hat{\theta}(x)} = \int_{\Theta} \frac{1}{\theta} \, \pi(\theta|x) \, d\theta = E(\theta^{-1}|x),$$

or

$$\hat{\theta}(x) = \frac{1}{E(\theta^{-1}|x)}.$$

□

Since this is true for each $x \in \mathcal{X}$, the decision rule $\hat{\theta}(X)$ is the Bayes estimator of θ under entropy loss.

If $\pi(\theta)$ is an improper prior, then $\hat{\theta}(X)$ given by (2.91) is referred to as a *generalized Bayes estimator*, as long as the posterior risk is finite. The connections between the PDP distributions and the Bayesian posterior distributions were established in section 2.4.9.1. Our final step is to show that the PDP estimators are generalized Bayes with respect to the improper prior associated with the Box-Tiao posterior.

Theorem 2.18. *Under entropy loss and the (improper) prior* (2.42),

(a) the generalized Bayes estimator of τ_1 *is the PDP-unbiased estimator* (2.58)*;*

(b) the generalized Bayes estimator of τ_2 *is the PDP-unbiased estimator* (2.59)*;*

(c) the generalized Bayes estimator of σ *is the PDP-unbiased estimator* (2.61).

Proof. Assume $\pi(\mu, \tau_1, \tau_2) = (\tau_1\,\tau_2)^{-1}$. To prove (a), we want to show that for any fixed $t_1, t_2 > 0$,

$$E(\tau_1^{-1}|t_1, t_2) = \frac{E(U_1^*(s))}{t_1} = \frac{\nu_1}{t_1}\,\frac{I_z(\phi_2, \phi_1 + 1)}{I_z(\phi_2, \phi_1)}. \tag{2.92}$$

Now,

$$E(\tau_1^{-1}|t_1, t_2) = \int_0^\infty \frac{1}{\tau_1}\, g^*(\tau_1|t_1, t_2)\, d\tau_1,$$

where

$$g^*(\tau_1|t_1, t_2) = \frac{1}{t_1}\,\frac{g^*_{\nu_1}\!\left(\frac{\tau_1}{t_1}\right) G_{\nu_2}\!\left(\frac{t_2}{\tau_1}\right)}{I_z(\phi_2, \phi_1)}$$

and $g^*(\cdot)$ denotes an inverse χ^2 density. Make the transformation $u = t_1/\tau_1$; then,

$$\begin{aligned} E(\tau_1^{-1}|t_1, t_2) &= \int_0^\infty u\,\frac{1}{t_1}\,\frac{g_{\nu_1}(u)\, G_{\nu_2}(su)}{I_z(\phi_2, \phi_1)}\, du \\ &= \frac{1}{t_1}\int_0^\infty u\,\frac{g_{\nu_1}(u)\, G_{\nu_2}(su)}{I_z(\phi_2, \phi_1)}\, du. \end{aligned}$$

The above integral is precisely $E[U_1^*(s)]$, so we have established (2.92). Therefore, the generalized Bayes estimator of τ_1 with respect to the prior $(\tau_1\,\tau_2)^{-1}$ is

$$\tilde{\tau}_1 = \frac{1}{E(\tau_1^{-1}|T_1, T_2)} = \frac{T_1}{\nu_1}\,\frac{I_Z(\phi_2, \phi_1)}{I_Z(\phi_2, \phi_1 + 1)},$$

the PDP-unbiased estimator of τ_1.

To prove (b), we begin with

$$E(\tau_2^{-1}|t_1, t_2) = \int_0^\infty \frac{1}{\tau_2}\, g^*(\tau_2|t_1, t_2)\, d\tau_2,$$

where

$$g^*(\tau_2|t_1, t_2) = \frac{1}{t_2}\,\frac{g_{\nu_2}^*\left(\frac{\tau_2}{t_2}\right)\left[1 - G_{\nu_1}\left(\frac{t_2}{s\tau_2}\right)\right]}{I_z(\phi_2, \phi_1)}.$$

Make the transformation $u = t_2/\tau_2$, so that

$$E(\tau_2^{-1}|t_1, t_2) = \frac{1}{t_2}\int_0^\infty u\,\frac{g_{\nu_2}(u)\,[1 - G_{\nu_1}(u/s)]}{I_z(\phi_2, \phi_1)}\,du.$$

But the integral term is precisely $E[U_2^*(s)]$, so we have established that the generalized Bayes estimator of τ_2 under the prior $(\tau_1\,\tau_2)^{-1}$ is the PDP-unbiased estimator (2.59).

Finally, to prove (c), we need to show that

$$E(\sigma^{-1}|s) = \frac{1}{s}\,\frac{\nu_2}{(\nu_1 - 2)}\,\frac{I_z(\phi_2 + 1, \phi_1 - 1)}{I_z(\phi_2, \phi_1)}. \tag{2.93}$$

But

$$E(\sigma^{-1}|s) = \int_1^\infty \frac{1}{\sigma}\,p(\sigma|s)\,d\sigma,$$

where

$$p(\sigma|s) = \frac{\Gamma(\phi_1 + \phi_2)}{\Gamma(\phi_1)\,\Gamma(\phi_2)}\,\frac{1}{s}\,\frac{(\sigma/s)^{\phi_1 - 1}}{(1 + \sigma/s)^{\phi_1 + \phi_2}}\,\frac{1}{I_z(\phi_2, \phi_1)}, \tag{2.94}$$

defined for $\sigma \geq 1$, is the posterior density of σ. Make the transformation $u = s/\sigma$, for which $d\sigma = (s/u^2)\,du$. Then

$$p(u|s) = \frac{\Gamma(\phi_1 + \phi_2)}{\Gamma(\phi_1)\,\Gamma(\phi_2)}\,\frac{u^{\phi_2 - 1}}{(1 + u)^{\phi_1 + \phi_2}}\,\frac{1}{I_z(\phi_2, \phi_1)},$$

$\sigma^{-1} = u/s$ and the integration limits transform from $(1, \infty)$ to $(0, s)$. Next, make the transformation $w = u/(1+u)$, for which $du = dw/(1-w)^2$. This transforms $p(u|s)$ from a Beta type II density into a Beta type I density; in addition, the original σ^{-1} is transformed into $w/[(1-w)s]$ and the integration limits become $(0, z)$, where $z = s/(1+s)$. As a result,

$$E(\sigma^{-1}|s) = \int_0^z \frac{w}{(1-w)s}\, p(w|s)\, dw,$$

where $p(w|s)$ is a Beta(ϕ_2, ϕ_1) density divided by $I_z(\phi_2, \phi_1)$. Absorbing the terms in w into $p(w|s)$, we obtain

$$\begin{aligned} E(\sigma^{-1}|s) &= \frac{1}{s}\int_0^z \frac{\Gamma(\phi_1+\phi_2)}{\Gamma(\phi_1)\,\Gamma(\phi_2)} \frac{w^{(\phi_2+1)-1}\,(1-w)^{(\phi_1-1)-1}}{I_z(\phi_2, \phi_1)}\, dw \\ &= \frac{1}{s}\,\frac{\nu_2}{\nu_1-2}\,\frac{I_z(\phi_2+1, \phi_1-1)}{I_z(\phi_2, \phi_1)}, \end{aligned}$$

the reciprocal of the PDP-unbiased estimator of σ. Therefore, the generalized Bayes estimator of σ with respect to the prior $(\tau_1\,\tau_2)^{-1}$ is the PDP-unbiased estimator (2.61). This completes the proof. □

2.5.8 Joint estimation of $(\sigma_e^2, J\sigma_\alpha^2)$

2.5.8.1 Improved estimation of $(\sigma_e^2, J\sigma_\alpha^2)$: IERD method

Kubokawa (1994) applied the Integrated Expression of Risk Difference (IERD) method to unify point and interval estimation in the normal variance problem. Recently, Kubokawa, Saleh and Konno (1998) adapted this method to the problem of jointly estimating the variance components $(\sigma_e^2, \sigma_\alpha^2)$ in the balanced one-way random model under the composite entropy loss function

$$L(\hat\sigma_e^2, \hat\sigma_\alpha^2; \sigma_e^2, \sigma_\alpha^2) = \nu_1\left\{\frac{\hat\sigma_e^2}{\sigma_e^2} - 1 - \log\frac{\hat\sigma_e^2}{\sigma_e^2}\right\} + \nu_2\left\{\frac{\hat\sigma_e^2 + J\hat\sigma_\alpha^2}{\sigma_e^2 + J\sigma_\alpha^2} - 1 - \log\frac{\hat\sigma_e^2 + J\hat\sigma_\alpha^2}{\sigma_e^2 + J\sigma_\alpha^2}\right\}, \tag{2.95}$$

a linear combination of entropy loss functions for estimating τ_1 and τ_2, the expected mean squares of the model.

By the nature of the estimators under consideration, it follows that

$$\hat{\tau}_1 = \hat{\sigma}_e^2 = \frac{T_1}{\nu_1}$$

$$\hat{\tau}_2 = \hat{\sigma}_e^2 + J\hat{\sigma}_\alpha^2 = \frac{T_2}{\nu_2}$$

are best $\mathscr{G}$-equivariant estimators of τ_1 and τ_2, respectively, under (marginal) entropy loss, as seen by our earlier results. Their difference,

$$J\hat{\sigma}_\alpha^2 = \frac{T_2}{\nu_2} - \frac{T_1}{\nu_1},$$

is the unbiased (ANOVA) estimator of $J\sigma_\alpha^2 = \tau_2 - \tau_1$. Kubokawa *et al.* (1998) show that the joint estimator $(\hat{\sigma}_e^2, J\hat{\sigma}_\alpha^2)$ is minimax under (2.95) [Proposition 1 of their paper].

The question, then, is how to improve upon the joint ANOVA estimator. Kubokawa *et al.* (1998) consider joint estimators of the form

$$(\tilde{\sigma}_e^2, \tilde{\sigma}_\alpha^2) = (\phi_1(S)T_1, \psi(S)T_1),$$

where $\psi(S) = \phi_2(1/S)S - \phi_1(S)$, in the class of minimax estimators of $(\sigma_e^2, \sigma_\alpha^2)$. This form of estimator improves upon the unconditional joint ANOVA estimator whenever

(a) $\phi_1(w)$ is nondecreasing, and $\lim_{w\to\infty} \phi_1(w) = \nu_1^{-1}$;

(b) $\phi_1(w) \geq \phi_1^0(w)$, where

$$\phi_1^0(w) = \frac{1}{\nu_1}\frac{I_c(\phi_2, \phi_1)}{I_c(\phi_2, \phi_1 + 1)} = \frac{1}{\nu_1 + 2\delta_c(\phi_2, \phi_1)},$$

and $c = w/(1+w)$;

(c) $\phi_2(t)$ is nondecreasing, and $\phi_2(0) = \nu_2^{-1}$;

(d) $\phi_2(t) \leq \phi_2^0(t)$, where

$$\phi_2^0(t) = \frac{1}{\nu_2} \frac{I_{1-d}(\phi_2, \phi_1)}{I_{1-d}(\phi_2 + 1, \phi_1)}$$

and $d = t/(1+t)$.

Conditions (a) and (b) are familiar: the multiplier function $\phi_1(S)$ is bounded above by the 'usual' multiplier ν_1^{-1} as s gets large, and is bounded below by the multiplier of the BZ-like estimator (the reciprocal of $E(U_1^*)$). Kubokawa *et al.*'s conditions (c) and (d) are designed to be compatible with IERD theory: rather than working with $\phi_2(S)$, a monotone decreasing function of S, they work with $\phi_2(1/S)$, which is strictly increasing with lower bound ν_2^{-1}. Essentially, the PDP multiplier function is the reflection of the KSK multiplier function on $(0, \infty)$.

Corollary 1 of Kubokawa *et al.* (1998) established that the class of improved joint estimators satisfying the above conditions is minimax. Their next step was to find empirical Bayes and generalized Bayes procedures within this class of improved joint estimators: empirical Bayes estimators have the general form

$$\tilde{\sigma}_e^2(a) = \min\left\{\frac{T_1}{\nu_1}, \frac{T_1 + aT_2}{\nu_1 + \nu_2}\right\}$$

$$J\tilde{\sigma}_\alpha^2(a) = \max\left\{\frac{T_2}{\nu_2}, \frac{T_1/a + T_2}{\nu_1 + \nu_2} - \tilde{\sigma}_e^2(a)\right\},$$

whereas generalized Bayes estimators are of the form

$$\bar{\sigma}_e^2(b) = \frac{T_1}{\nu_1} \frac{I_{Z^*}(\phi_2, \phi_1)}{I_{Z^*}(\phi_2, \phi_1 + 1)}$$

$$\bar{\sigma}_\alpha^2(b) = \frac{T_2}{\nu_2} \frac{I_{Z^*}(\phi_2, \phi_1)}{I_{Z^*}(\phi_2 + 1, \phi_1)} - \bar{\sigma}_e^2(b),$$

where $Z^* = bS/(1 + bS)$.

When $\nu_2/(\nu_2 + 2) \leq a \leq 1$, conditions (a)–(d) above on the multiplier functions $\phi_1(S)$ and $\phi_2(S)$ are satisfied, establishing the class of improved empirical Bayes procedures. In particular, when $a = 1$, the joint estimator $(\tilde{\sigma}_e^2(1), J\tilde{\sigma}_\alpha^2(1))$ reduces to the joint REML estimator, which is both minimax and empirical Bayes under (2.95). Whenever $b \geq 1$, generalized Bayes estimators satisfy conditions (a)–(d); in particular, when $b = 1$, the joint estimator $(\bar{\sigma}_e^2(1), J\bar{\sigma}_\alpha^2(1))$ is the joint PDP-unbiased estimator of $(\sigma_e^2, J\sigma_\alpha^2)$ under (2.95), which is both minimax and generalized Bayes.

2.6 Interval estimation

The objective of this section is to develop confidence intervals for τ_1, τ_2 and σ based on the PDP distributions. These will be of two types: intervals with fixed conditional coverage probability, and intervals whose "length" is fixed to that of the corresponding unconditional interval. In the latter case, we will show that the PDP intervals have at least nominal unconditional coverage uniformly in σ; furthermore, we will develop a Brewster-Zidek argument for fixed length intervals of τ_1 and τ_2 and show that they are generalized Bayes with respect to the Jeffreys prior, analogous to the point estimation problem. From intervals for σ, one can obtain intervals for $J\gamma$ and ρ. PDP intervals for $J\sigma_\alpha^2$ are considered at the end of this section.

We outline three criteria below for producing confidence intervals for scale parameters in unimodal families of distributions. We will focus on one of these, the log-shortest (or shortest unbiased) criterion, as it corresponds nicely with unbiased point estimators. Similar work has been done with respect to minimum length intervals, and will be made available as a separate technical report.

2.6.1 Criteria for interval estimators

In a scale family of continuous distributions, an unconditional pivotal confidence interval for a scale parameter θ has the general form

$$C_\theta(\boldsymbol{X}) = \left(\frac{T}{b}, \frac{T}{a}\right), \tag{2.96}$$

where T/θ is a pivotal quantity with continuous cdf $F(\cdot)$, a is a lower quantile of F, and b an upper quantile of F. Tate and Klett (1959) established several optimality criteria for obtaining a and b when the underlying distribution is unimodal. Three such criteria are given below.

1. An *equal-tailed* interval satisfies the twin conditions

$$F(a) = \frac{\alpha}{2} \qquad \text{and} \qquad F(b) = 1 - \frac{\alpha}{2}, \tag{2.97}$$

   where $\gamma = 1 - \alpha$ is the confidence coefficient. Equal tailed intervals often have good frequentist properties in location problems with symmetric distributions; however, their utility in scale estimation problems is suspect, since in most cases $F(\cdot)$ is asymmetric.

2. A *shortest* (pivotal) confidence interval has minimum length for fixed coverage probability γ. The quantiles a and b of a shortest interval are obtained from constrained minimization of

$$g(a, b, \lambda; \gamma) = \left(\frac{1}{a} - \frac{1}{b}\right) + \lambda\,(F(b) - F(a) - \gamma).$$

   The gradient vector of this objective function has the following components:

$$\frac{\partial g}{\partial a} = -\frac{1}{a^2} - \lambda f(a),$$

$$\frac{\partial g}{\partial b} = \frac{1}{b^2} + \lambda f(b),$$

$$\frac{\partial g}{\partial \lambda} = F(b) - F(a) - \gamma.$$

Setting each of these equations equal to zero, the 'solution' is the nonlinear system of equations

$$a^2 f(a) = b^2 f(b), \tag{2.98}$$

$$F(b) - F(a) = \gamma, \tag{2.99}$$

where $f(x) = F'(x)$ is a continuous density function. The nonlinear system (2.98) is solved numerically for a and b to produce the quantiles for a shortest confidence interval of θ with confidence coefficient γ.

3. A *log-shortest* (pivotal) interval has minimum log length (or equivalently, minimum endpoint ratio) for fixed coverage probability γ . Quantiles a and b are obtained from constrained minimization of

$$h(a, b, \lambda; \gamma) = [\log b - \log a] + \lambda\,(F(b) - F(a) - \gamma),$$

where log $b -$ log a is the logarithm of the endpoint ratio. The gradient vector consists of the components

$$\frac{\partial h}{\partial a} = -\frac{1}{a} - \lambda f(a),$$

$$\frac{\partial h}{\partial b} = \frac{1}{b} + \lambda f(b),$$

$$\frac{\partial h}{\partial \lambda} = F(b) - F(a) - \gamma.$$

Setting each of these equations equal to zero, the solution is the nonlinear system of equations

$$a\,f(a) = b\,f(b), \tag{2.100}$$

$$F(b) - F(a) = \gamma, \tag{2.101}$$

which is solved numerically for a and b to produce the quantiles of a log-shortest confidence interval of θ with confidence level γ.

Log-shortest intervals minimize the probability of false coverage in the class of unbiased intervals of a parameter θ; i. e., they are uniformly most accurate unbiased. Since the ratio of endpoints is an invariant measure of size, log-shortest intervals are optimal in the class of scale invariant intervals (Tate and Klett, 1959; Casella and Berger, 1990, ch. 9), analogous to the use of entropy loss rather than (normalized) quadratic loss for point estimation of scale parameters. This criterion penalizes movement of the lower endpoint towards zero because small movement of the lower endpoint may require much larger movement of the upper endpoint to maintain a constant ratio.

The coverage probability of a confidence interval $C(\boldsymbol{X})$ is defined as $P_\theta(\theta \in C(\boldsymbol{X}))$. Since the coverage probability can vary for different values of θ, it is customary to express coverage in terms of a *confidence coefficient* γ, defined as

$$\gamma = \inf_{\theta \in \Theta} P_\theta(\theta \in C(\boldsymbol{X})).$$

We are concerned with the unconditional coverage of intervals based on PDP distributions, whether they have fixed conditional coverage probability or fixed log length. Average log length is also a consideration for intervals with fixed conditional coverage.

2.6.2 PDP confidence intervals

Notation. In the balanced one-way random effects model, we are primarily interested in CI's for scale parameters τ_1, τ_2 and σ. In this case, an unconditional γ level confidence interval for a scale parameter θ has the general form (2.96), where T is a statistic and a and b are appropriate quantiles of a unimodal distribution that satisfy the probability statement

$$P_\sigma\left(\frac{T}{b} \le \theta \le \frac{T}{a}\right) = P_\sigma(a \le U \le b) = \gamma$$

by any selected optimality criterion, where $U = T/\theta$ is a pivotal quantity. For any fixed $s > 0$, the basic idea is to replace the unconditional distribution of U with that of $U^*(s) = U|W \le s$, and consider interval estimation of θ with respect to the PDP.

In the sequel, we will make reference to three related functions associated with PDP confidence intervals: quantile functions, multiplier functions and endpoint functions. If $P(\xi_1(S) \le U^* \le \xi_2(S)|W \le s) = \gamma$ for some PDP U^*, then the $\xi_i(S)$ are called *quantile functions*, as they pertain to the choice of quantiles from the reference distribution of U^* that achieves the desired coverage. If $U = T/\theta$, a transformation of the probability statement given above is

$$P\Big(\frac{T}{\xi_2(S)} \le \theta \le \frac{T}{\xi_1(S)}\Big| W \le s\Big)$$

with respect to the distribution of U^*. Then, $\psi_1(S) = 1/\xi_2(S)$ and $\psi_2(S) = 1/\xi_1(S)$ are called the *multiplier functions* of the γ level PDP-CI of θ. Multiplying the $\psi_i(\cdot)$ by T yields the *endpoint functions* of the interval. These terms are used rather freely throughout this section.

2.6.2.1 PDP-CIs of τ_1: fixed conditional coverage

To motivate the idea of a PDP-CI, choose some $s > 0$, which educes a PDP $U_1^*(s)$ with density (2.6). To find a confidence interval for τ_1 based on $U_1^*(s)$, apply some criterion for producing quantiles $\xi_1(s)$ and $\xi_2(s)$ such that $P(\xi_1(s) \le U_1^*(s) \le \xi_2(s)) = \gamma$, where γ is the conditional coverage probability. The resulting PDP confidence interval estimate of τ_1 is then

$$\Big(\frac{t_1}{\xi_2(s)}, \frac{t_1}{\xi_1(s)}\Big).$$

This process can be iterated for any $s > 0$, so over all values of s, a continuous, monotone function $\psi_1(S)$ represents the lower multiplier function and a continuous, monotone function $\psi_2(S)$ represents the upper multiplier function associated with the family of PDP confidence intervals for τ_1 with fixed conditional coverage.

For fixed $\sigma \geq 1$, a PDP-CI of τ_1 has the general form

$$C_\sigma(\boldsymbol{X}|S) = (\psi_1(S)\, T_1, \psi_2(S)\, T_1), \tag{2.102}$$

where

$$\psi_1(S) = \frac{1}{\xi_2(S)} \qquad \text{and} \qquad \psi_2(S) = \frac{1}{\xi_1(S)}$$

are the lower and upper multiplier functions of the PDP-CI of τ_1. If each endpoint is multiplied by $1/\tau_1$, a *normalized* interval $(\psi_1(S)\, U_1, \psi_2(S)\, U_1)$ is produced, which is convenient for comparing average lengths.

A log-shortest $1-\alpha$ PDP-CI of τ_1 is obtained by substituting $\psi_2^{-1}(s)$ for a and $\psi_1^{-1}(s)$ for b in (2.100) and (2.101), and setting $f(x) = g_{\nu_1}(x)\, G_{\nu_2}(sx)$ in (2.100). The resulting system, expressed in terms of the multiplier functions, is

$$\frac{1}{\psi_2(s)}\, g_{\nu_1}\left(\frac{1}{\psi_2(s)}\right) G_{\nu_2}\left(\frac{s}{\psi_2(s)}\right) = \frac{1}{\psi_1(s)}\, g_{\nu_1}\left(\frac{1}{\psi_1(s)}\right) G_{\nu_2}\left(\frac{s}{\psi_1(s)}\right) \tag{2.103}$$

$$F\left(\frac{1}{\psi_1(s)}\right) - F\left(\frac{1}{\psi_2(s)}\right) = 1-\alpha, \tag{2.104}$$

where $F(\cdot)$ is the cdf of $U_1^*(s)$. This system can be solved numerically for $\psi_1(s)$ and $\psi_2(s)$ at any fixed $s > 0$.

Figure 2.21 shows the multiplier functions of the log shortest PDP-CI of τ_1 for the case $(\nu_2, \nu_1) = (3, 4)$. These multipliers get progressively shorter as $s \to 0$, converging to the corresponding unconditional multipliers with $\nu_1 + \nu_2$ degrees of freedom in the limit. Conversely, as $s \to \infty$, they converge to the unconditional multipliers with ν_1 degrees of freedom. The shape of the multiplier functions (and their monotonicity) is a consequence of Theorem A.14(a)–(c); observe that the behavior in Figure 2.21 mirrors that of the PDP point estimators.

The following result yields the unconditional coverage probability of any type of PDP-CI of τ_1 (fixed conditional coverage or fixed length) for any $\sigma \geq 1$.

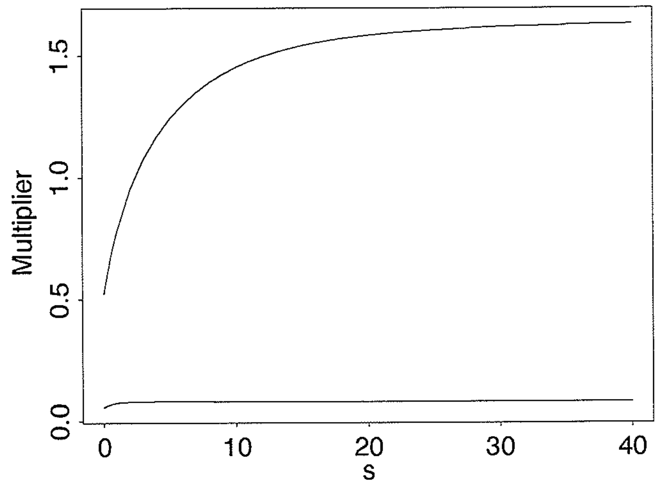

Figure 2.21: Multiplier functions of the log-shortest conditional 95% PDP-CI of τ_1 with $(\nu_2, \nu_1) = (3, 4)$.

Theorem 2.19. *Let $\xi_1(S)$ and $\xi_2(S)$ denote the lower and upper quantiles of a PDP-CI of τ_1. Then, for any fixed $\sigma \geq 1$, the unconditional coverage probability of the PDP-CI is given by*

$$\int_0^1 \left[G_{\nu_1+\nu_2}\Big(\frac{\xi_2(c_\sigma(v))}{1-v}\Big) - G_{\nu_1+\nu_2}\Big(\frac{\xi_1(c_\sigma(v))}{1-v}\Big)\right] g_V(v)\, dv, \tag{2.105}$$

where

$$v = \frac{s/\sigma}{1+s/\sigma} \tag{2.106}$$

and

$$c_\sigma(v) = \frac{\sigma v}{1-v}. \tag{2.107}$$

Proof. Let $\sigma \geq 1$ be given, and define $p(S) = 1 + S/\sigma$. Recall that $p(S)\, U_1 \sim \chi^2_{\nu_1+\nu_2}$, where U_1 and S are independently distributed. The unconditional coverage probability of a family of PDP-CIs of τ_1 with fixed conditional coverage $1-\alpha$ is then

$$\begin{aligned} P_\sigma(\xi_1(S) \leq U_1 \leq \xi_2(S)) &= E_\sigma\big[P_\sigma(\xi_1(S) \leq U_1 \leq \xi_2(S)|S)\big] \\ &= E_\sigma[P_\sigma(\xi_1(S)\,p(S) \leq U_1 + U_2 \leq \xi_2(S)\,p(S)|S)] \end{aligned}$$

where $T_1 = 1$ without loss of generality. Since $(1+S/\sigma)\,U_1|S \sim \chi^2_{\nu_1+\nu_2}$, it follows that

$$\begin{aligned} P_\sigma(\xi_1(S) \leq U_1 \leq \xi_2(S)) &= E_\sigma\big[P_\sigma\left[p(S)\,\xi_1(S) \leq p(S)\,U_1 \leq (S)\,\xi_2(S)|S\right]\big] \\ &= E_\sigma\big[G_{\nu_1+\nu_2}\left[p(S)\,\xi_2(S)\right] - G_{\nu_1+\nu_2}\left[p(S)\,\xi_1(S)\right]\big] \\ &= \int_0^\infty \big[G_{\nu_1+\nu_2}\left[p(s)\,\xi_2(s)\right] - G_{\nu_1+\nu_2}\left[p(s)\,\xi_1(s)\right]\big]\,h_S(s)\,ds, \end{aligned} \tag{2.108}$$

where $G(\cdot)$ is a χ^2 cdf and $h(\cdot)$ is the unconditional density of S.

For numerical stability, it is convenient to make the transformation v, with

$$dv = \frac{1}{\sigma}\left(\frac{1}{1+s/\sigma}\right)^2 ds; \tag{2.109}$$

then, the transformed random variable V has a Beta(ϕ_2, ϕ_1) distribution. Note that the inverse transform back to s in terms of v is (2.107), while $p(s)$ transforms to $1/(1-v)$. Under this transformation, the integral can be expressed as

$$\int_0^1 \left[G_{\nu_1+\nu_2}\Big(\frac{\xi_2(c_\sigma(v))}{1-v}\Big) - G_{\nu_1+\nu_2}\Big(\frac{\xi_1(c_\sigma(v))}{1-v}\Big)\right] g_V(v)\,dv.$$

This completes the proof. □

Figure 2.22 shows the unconditional coverage of log-shortest 95% PDP-CIs of τ_1 (as a function of σ) when $(\nu_2, \nu_1) = (3, 4)$. In this case, the unconditional coverage is bounded above by the nominal 0.95 coverage. This performance is not surprising since these PDP intervals can be much shorter than their unconditional counterparts, particularly for small s and σ close to 1. The benefit of shorter average log length (see below) of PDP-CIs with fixed conditional coverage $1-\alpha$ is associated with the cost of lower than nominal unconditional coverage.

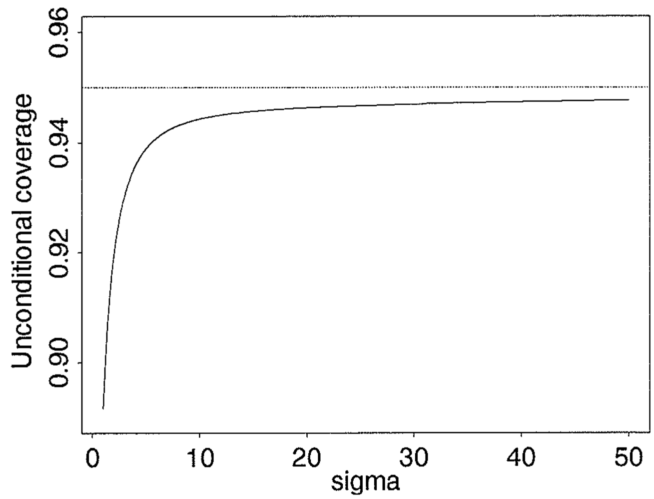

Figure 2.22: Unconditional coverage probabilities of 95% log-shortest PDP-CI of τ_1 as a function of σ, when $(\nu_2, \nu_1) = (3, 4)$.

Theorem 2.20. *For any fixed $\sigma \geq 1$, the average log length of a log-shortest $1 - \alpha$ PDP-CI of τ_1 with form (2.102) is*

$$E_\sigma(\textit{log-length}) = \int_0^1 \left[\log\, \psi_2(c_\sigma(v)) - \log\, \psi_1(c_\sigma(v))\right] g_V(v)\, dv. \tag{2.110}$$

Proof. Fix $\sigma \geq 1$. The interval is of the form $T_1(\psi_1(S), \psi_2(S))$; its log length is

$$\log\left(\frac{T_1\psi_2(S)}{T_1\psi_1(S)}\right) = \log \psi_2(S) - \log \psi_1(S).$$

Therefore,

$$\begin{aligned} E_\sigma(\text{log-length}) &= E_\sigma[\log \psi_2(S) - \log \psi_1(S)] \\ &= \int_0^\infty [\log \psi_2(s) - \log \psi_1(s)]\, h_S(s)\, ds. \end{aligned}$$

Applying the transformation (2.106) for V, we obtain

$$E_\sigma(\text{log-length}) = \int_0^1 \left[\log\, \psi_2(c_\sigma(v)) - \log\, \psi_1(c_\sigma(v))\right] g_V(v)\, dv,$$

where $c_\sigma(v)$ is given by (2.107). This completes the proof. □

Since $\log \psi_i(S) = -\log \xi_{3-i}(S)$, $i = 1, 2$, we may also express the average log length function as

$$E_\sigma(\text{log-length}) = \int_0^1 \left[\log\, \xi_1(c_\sigma(v)) - \log\, \xi_2(c_\sigma(v))\right] g_V(v)\, dv. \tag{2.111}$$

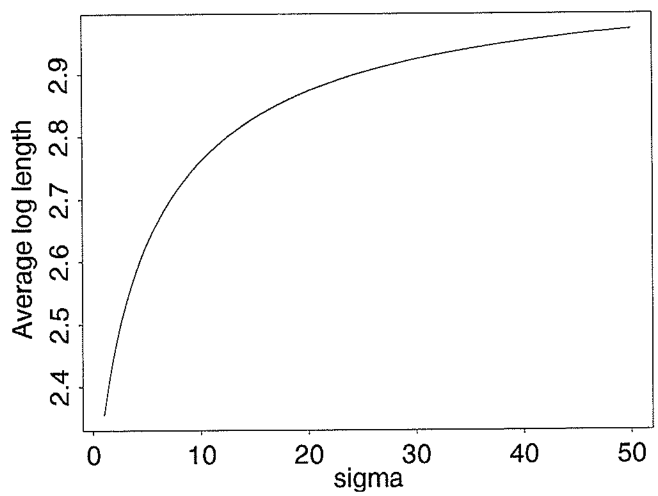

Figure 2.23: Expected log length of 95% PDP-CIs of τ_1, when $(\nu_2, \nu_1) = (3, 4)$.

Average log lengths of PDP intervals with fixed conditional coverage 0.95 are shown for $(\nu_2, \nu_1) = (3, 4)$ in Figure 2.23. The upper asymptote corresponds to the average log-length of the unconditional 95% log-shortest CI of τ_1, and we find that significant reduction in average log length can be achieved by the PDP shrinkage intervals, especially for small values of σ.

2.6.2.2 PDP-CIs of τ_2: fixed conditional coverage

The behavior of post-data intervals for τ_2 mirrors that of point estimators, just as they did for τ_1. In this case, the PDP multiplier functions will inflate as $s \to 0$, and converge to the unconditional multipliers as s gets large. The setup is very similar to the τ_1 case. Unconditional CIs of τ_2 have the general form

$$C_2(\boldsymbol{X}) = \left(\frac{T_2}{b}, \frac{T_2}{a}\right),$$

where T_2 is the treatment sum of squares, and a and b are appropriate quantiles of a $\chi^2_{\nu_2}$ distribution that satisfy $P_\sigma(a \le U_2 \le b) = 1 - \alpha$ by any selected length criterion.

To construct a PDP interval for τ_2 with fixed conditional coverage $1-\alpha$, define two continuous and monotone quantile functions $\zeta_1(S)$ and $\zeta_2(S)$, such that for each $s > 0$ and each fixed $\sigma \ge 1$,

$$P_\sigma(\zeta_1(s) \le U_2 \le \zeta_2(s) | W \le s) = 1 - \alpha.$$

The lower quantile function $\zeta_1(S)$ is the function derived from the ensemble of lower quantiles $\{\zeta_1(s|\sigma) : s > 0\}$ from all $U_2^*(s)$ densities; the upper quantile function $\zeta_2(S)$ is obtained in a parallel fashion.

A normalized PDP-CI for τ_2 has the general form

$$C_\sigma(\boldsymbol{X}|S) = U_2\left(\psi_1^*(S), \psi_2^*(S)\right), \tag{2.112}$$

where

$$\psi_1^*(S) = \frac{1}{\zeta_2(S)} \qquad \text{and} \qquad \psi_2^*(S) = \frac{1}{\zeta_1(S)}$$

are the lower and upper multiplier functions of the PDP-CI of τ_2, respectively.

A log shortest $1-\alpha$ CI of τ_2 is of the form (2.112), where $\psi_1^*(\cdot)$ and $\psi_2^*(\cdot)$ are solutions of the nonlinear system

$$\frac{1}{\psi_2^*(s)}\, g_{\nu_2}\left(\frac{1}{\psi_2^*(s)}\right)\left[1 - G_{\nu_1}\left(\frac{1}{s\,\psi_2^*(s)}\right)\right] = \frac{1}{\psi_1^*(s)}\, g_{\nu_2}\left(\frac{1}{\psi_1^*(s)}\right)\left[1 - G_{\nu_1}\left(\frac{1}{s\,\psi_1^*(s)}\right)\right] \tag{2.113}$$

$$F^*\left(\frac{1}{\psi_1^*(s)}\right) - F^*\left(\frac{1}{\psi_2^*(s)}\right) = 1 - \alpha, \tag{2.114}$$

for any fixed $s > 0$, where $F^*(\cdot)$ is the cdf of $U_2^*(s)$.

Multiplier functions of 95% log-shortest PDP-CIs of τ_2 are shown in Figure 2.24 when $(\nu_2, \nu_1) = (3, 4)$. As in the point estimation problem, both multipliers inflate sharply towards infinity as $s \to 0$. (Note: Figure 2.24 is drawn to avoid visual clutter near $s = 0$, so is somewhat misleading to the extent that the trend toward $+\infty$ is not as obvious for the lower multiplier function as it should be.)

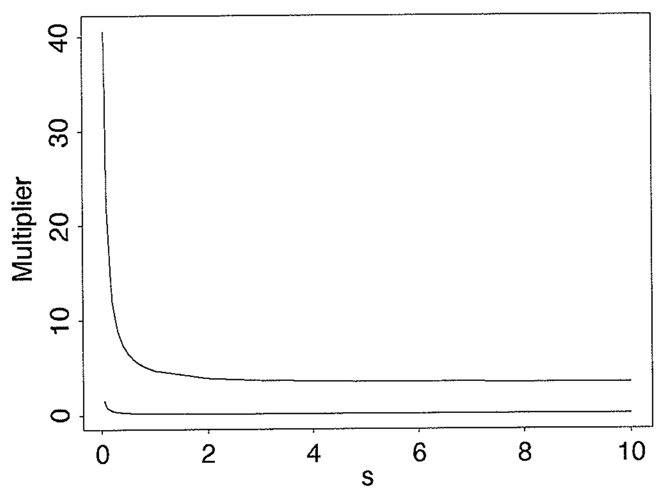

Figure 2.24: Multiplier functions of log-shortest 95% PDP-CIs of τ_2 when $(\nu_2, \nu_1) = (3, 4)$.

Theorem 2.21. *For any fixed $\sigma \geq 1$, the unconditional coverage probability of a PDP-CI of τ_2 is given by*

$$P_\sigma(\zeta_1(S) \leq U_2 \leq \zeta_2(S)) =$$

$$\int_0^1 \left[G_{\nu_1+\nu_2}\left(\frac{\zeta_2(c_\sigma(v))}{v}\right) - G_{\nu_1+\nu_2}\left(\frac{\zeta_1(c_\sigma(v))}{v}\right)\right] g_V(v)\, dv, \quad (2.115)$$

where $c_\sigma(v)$ is defined by (2.107).

Proof. In Theorem 2.19, set $\zeta_i(\cdot) = \xi_i(\cdot)$ and replace $p(s)$ with $q(s) = 1 + \sigma/s$ so that $q(S)\, U_2 = U_1 + U_2$. The transformation from s to v then transforms $q(s)$ to $1/v$, from which the result follows. □

Unconditional coverage probabilities of 95% log-shortest PDP-CIs (as a function of σ) are shown in Figure 2.25 for the case $(\nu_2, \nu_1) = (3, 4)$. These are at least the nominal level 0.95 for all σ, which is expected since PDP intervals for τ_2 are wider than the unconditional interval, especially when s is small.

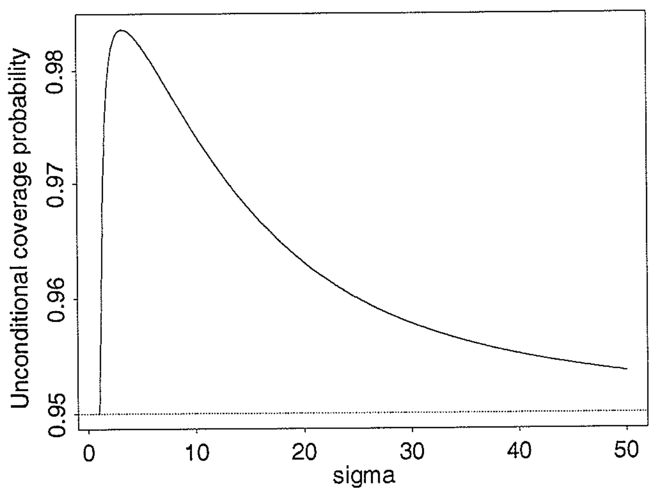

Figure 2.25: Unconditional coverage probabilities of 95% log-shortest PDP-CIs of τ_2, when $(\nu_2, \nu_1) = (3, 4)$.

Theorem 2.22. *For any fixed* $\sigma \geq 1$, *the average log length of a log-shortest* $1 - \alpha$ *PDP-CI of* τ_2 *is given by*

$$E_\sigma(\textit{log-length}) = \int_0^1 [\log \psi_2^*(c_\sigma(v)) - \log \psi_1^*(c_\sigma(v))]\, g_V(v)\, dv, \tag{2.116}$$

where $c_\sigma(v)$ *is defined by* (2.107).

Proof. Set $\psi_i^*(\cdot) = \psi_i(\cdot)$ in the proof of Theorem 2.20. □

The average log length function of a log-shortest PDP-CI of τ_2 is illustrated in Figure 2.26 in the case where $(\nu_2, \nu_1) = (3, 4)$. Interestingly, the average log length is an increasing function of σ even though the multiplier functions of the PDP intervals are inflating as s gets small.

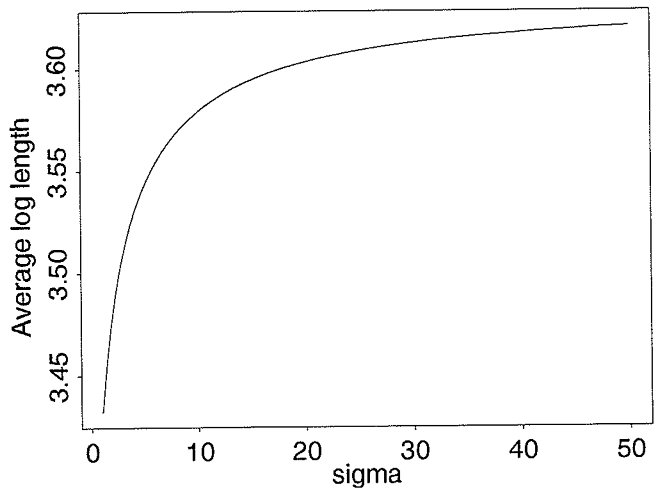

Figure 2.26: Average log lengths of 95% log-shortest PDP-CIs of τ_2 when $(\nu_2, \nu_1) = (3, 4)$.

2.6.2.3 PDP-CIs of σ: fixed conditional coverage

The process of constructing confidence intervals for σ is a bit different from that used for establishing PDP-CIs of τ_1 and τ_2. To begin with, consider the problem of constructing an unconditional CI for σ with confidence coefficient $1 - \alpha$. It is convenient to use the $\mathcal{F}$ distribution as a reference, associated with the pivotal

$$\frac{F}{\sigma} = \frac{\nu_1}{\nu_2} W = \frac{\nu_1}{\nu_2} \frac{S}{\sigma} \sim \mathcal{F}_{\nu_2,\nu_1}.$$

Let a and b denote the lower and upper quantiles, respectively, of a $1 - \alpha$ level CI of σ. Then, prior to sampling,

$$1 - \alpha = P\Big(a \leq \frac{F}{\sigma} \leq b\Big) = P\bigg(\frac{\nu_1 S}{\nu_2 b} \leq \sigma \leq \frac{\nu_1 S}{\nu_2 a}\bigg).$$

Therefore, letting $F = \nu_1 S/\nu_2$ denote the observed F ratio, a $1-\alpha$ unconditional CI of σ is $(F/b, F/a)$. However, the unconditional $1-\alpha$ CI of σ based on the $\mathcal{F}$ distribution may not necessarily yield sensible results. To illustrate the situation, consider the following example.

Example 2.2. Let $(\nu_2, \nu_1) = (3, 4)$, and let $f = 0.5$ be the observed F ratio. Consider a shortest 95% CI of σ, whose lower and upper quantiles are given by $a = 0.109676$ and $b = 901.412$, respectively. A shortest 95% CI of σ is then

$$0.5\left(\frac{1}{901.412}, \frac{1}{0.109676}\right) = (0.00055, 4.559).$$

Over 20% of the width of this interval lies outside $[1, \infty)$, the parameter space of σ. To guarantee that the lower bound of this interval lies in the parameter space of σ, the observed F ratio would have to be larger than $b = 901.412$ in this example ... a fairly steep requirement.

Admittedly, this is a rather extreme example, but it does clarify that the observed F ratio must be larger than the upper quantile b in order for the interval to lie entirely in the parameter space of σ.

We begin with the problem of determining log-shortest confidence intervals of σ with fixed conditional coverage $1-\alpha$, based on the family of truncated $\mathcal{F}$ distributions associated with the PDP $F^*(f) = F/\sigma | F/\sigma \le f$. Let $h(x)$ denote the pdf and $H(x)$ the cdf of an $\mathcal{F}_{\nu_2,\nu_1}$ distribution; then, the pdf and cdf of $F^*(f)$ are given by

$$h^*(x) = \frac{h(x)}{H(f)} \chi_{[0,f]}(x)$$

and

$$H^*(x|f) = \begin{cases} 0 & x \le 0 \\ \dfrac{H(x)}{H(f)} & 0 < x \le f \\ 1 & x > f. \end{cases}$$

Log-shortest PDP-CIs of σ. A PDP confidence interval of σ is of the general form $(\psi_1(F)\,F, \psi_2(F)\,F)$. Given $f > 0$, the multiplier functions $\psi_1(F)$ and $\psi_2(F)$ of a $1-\alpha$ log-shortest PDP-CI of σ are pointwise solutions of the nonlinear system

$$\frac{1}{\psi_2(f)}\,h^*\left(\frac{1}{\psi_2(f)}\right) = \frac{1}{\psi_1(f)}\,h^*\left(\frac{1}{\psi_1(f)}\right), \tag{2.117}$$

$$H^*\left(\frac{1}{\psi_1(f)}\right) - H^*\left(\frac{1}{\psi_2(f)}\right) = 1-\alpha, \tag{2.118}$$

restricted to the domain $[0, f]$. If the truncation point f is sufficiently small, the above system has no solution in $[0, f]$. In this event, the upper quantile ξ_2 is set to the terminus f, which in turn means that the lower quantile ξ_1 is set to the α quantile of F^*, denoted by $H_*^{-1}(\alpha|f)$, where the subscripted star identifies the quantile with a truncated $\mathcal{F}$ distribution, whose value depends on f. Consequently, when (2.117) and (2.118) admit no solution, the log-shortest $1-\alpha$ PDP-CI of σ is $(1, f/H_*^{-1}(\alpha|f))$. The next question to address is: what is the minimum value of the terminus f that admits a two-sided PDP-CI for σ?

To answer this, set $\psi_1(f) = 1/f'$ and $\psi_2(f) = 1/H_*^{-1}(\alpha|f')$ in (2.117) and (2.118), and solve for f'. The solution yields the maximum value of the truncation point for which a one-sided interval is necessary; for any $f > f'$, a unique pair of solutions ψ_1 and ψ_2 to the system (2.117) and (2.118) obtain, and a two-sided interval is constructed.

If we define $(\varphi_1(F), \varphi_2(F)) = (\psi_1(F)\,F, \psi_2(F)\,F)$, then the endpoint functions φ_1 and φ_2 of a $1-\alpha$ log-shortest PDP-CI of σ are given by

$$\varphi_1(F) = \begin{cases} 1 & f \le f' \\ \psi_1(F)\,F & f > f', \end{cases}$$

and

$$\varphi_2(F) = \begin{cases} f/H_*^{-1}(\alpha|f) & f \leq f' \\ \psi_2(F)\,F & f > f'. \end{cases}$$

This notation is rather unusual, but necessary since the truncation point f plays into the interval construction process.

It is of interest to see what happens to these intervals in the cases where $f \to \infty$ and $f \to 0$. In the former case, the distribution of F^* converges to the unconditional $\mathcal{F}$ distribution, so the conditional $1-\alpha$ intervals converge to the unconditional $1-\alpha$ intervals as f gets large. In particular, the multiplier functions $\psi_1(F)$ and $\psi_2(F)$ converge to the unconditional multipliers as f gets large. On the other hand, as $f \to 0$, the density of F^* converges to a degenerate distribution at 0, which raises concern about the limit of the endpoint function $\varphi_2(F)$ as $f \to 0$, i. e.,

$$\lim_{f\to 0} \frac{f}{H_*^{-1}(\alpha|f)}.$$

Now, given $\alpha \in (0,1)$, we can write

$$\alpha = \frac{H(x)}{H(f)}$$

for some x in $(0, f)$. Since f is known post-data, so is $H(f)$, so

$$H(x) = \alpha H(f),$$

which implies that the α quantile of F^* is

$$x = H^{-1}[\alpha H(f)].$$

Now, it is possible in Mathematica to determine x when $\alpha H(f) \geq 10^{-15}$, or

$$f \geq H^{-1}(10^{-15}/\alpha).$$

The .05 quantile of the truncated $\mathcal{F}$ distribution can be found for all $f \geq 2 \times 10^{-13}$, while the .01 quantile can be computed for all $f \geq 10^{-13}$. Thus, the limiting α quantile as $f \to 0$ will be approximated by $H^{-1}(10^{-15})$, corresponding to $f = H^{-1}(10^{-15}/\alpha)$; i. e., the limiting upper endpoint φ_2 as $f \to 0$ is approximately

$$\lim_{f \to 0} \frac{f}{H_*^{-1}(\alpha|f)} \approx \frac{H^{-1}(10^{-15}/\alpha)}{H^{-1}(10^{-15})}. \tag{2.119}$$

This approximation should be very close to the actual limit, and should have at most a negligible effect on the computation of unconditional properties of intervals for σ (i. e., unconditional coverage probabilities and expected lengths).

Truncating the $\mathcal{F}$ distribution from above at the observed F ratio f guarantees that (one-sided) PDP-CIs of σ have lower endpoints ≥ 1 for all $f > 0$, thus ensuring that intervals for σ lie in the parameter space $[1, \infty)$. Figure 2.27 displays the endpoint functions of a 95% log-shortest PDP-CI of σ with $(\nu_2, \nu_1) = (3, 4)$. The horizontal axis corresponds to values of f, while the vertical axis corresponds to values of σ.

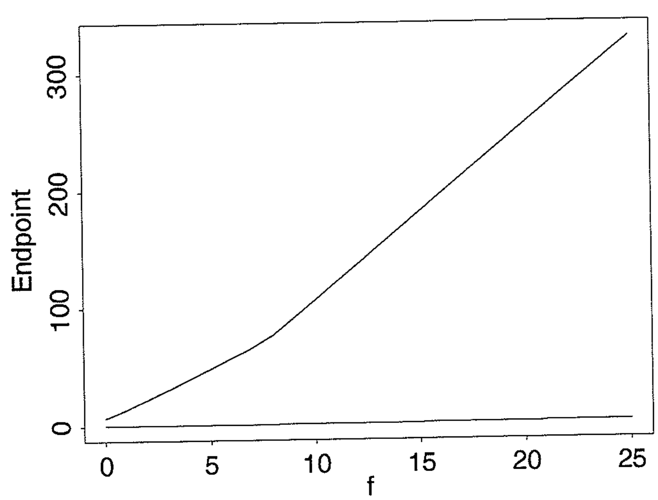

Figure 2.27: Endpoint functions of log-shortest 95% PDP-CIs of σ with $(\nu_2, \nu_1) = (3, 4)$.

To find the unconditional coverage probabilities of these intervals as a function of σ, view each interval as a set of σ values, and treat the vertical axis of an endpoint function graph as the σ axis. With respect to the upper endpoint function $\varphi_2(F)$, let σ^* denote the limiting upper endpoint (2.119) as $f \to 0$. The unconditional coverage at a fixed value of $\sigma \geq 1$ is then obtained as follows. Choose a value of σ, and draw a horizontal line through that value. If $1 \leq \sigma < \sigma^*$, the horizontal line intersects the endpoint function $\varphi_1(F)$. If $\sigma \geq \sigma^*$, the line intersects both endpoint functions at distinct F values. Let $\varphi_2^{-1}(\sigma)$ and $\varphi_1^{-1}(\sigma)$ denote the inverse functions of $\varphi_1(F)$ and $\varphi_2(F)$, respectively, at a fixed value of σ, where φ_1 is monotone on $[f', \infty)$ and φ_2 is monotone on $(0, \infty)$.

Theorem 2.23. *The unconditional coverage probability of a log-shortest* $1 - \alpha$ *PDP-CI of* σ *is*

$$P(\varphi_1(F) \leq \sigma \leq \varphi_2(F)) = \begin{cases} H\Big(\frac{\varphi_1^{-1}(\sigma)}{\sigma}\Big) & 1 \leq \sigma \leq \sigma^* \\ H\Big(\frac{\varphi_1^{-1}(\sigma)}{\sigma}\Big) - H\Big(\frac{\varphi_2^{-1}(\sigma)}{\sigma}\Big) & \sigma > \sigma^*, \end{cases} \tag{2.120}$$

where

$$\sigma^* = \lim_{f \to 0} \frac{f}{H_*^{-1}(\alpha | f)},$$

and $H(\cdot)$ *is the unconditional cdf of an* F *distribution.*

Proof.

$$\begin{aligned} P(\varphi_1(F) \leq \sigma \leq \varphi_2(F)) &= P(\varphi_2^{-1}(\sigma) \leq F \leq \varphi_1^{-1}(\sigma)) \\ &= P\Big(\frac{\varphi_2^{-1}(\sigma)}{\sigma} \leq \frac{F}{\sigma} \leq \frac{\varphi_1^{-1}(\sigma)}{\sigma}\Big) \\ &= \begin{cases} H\Big(\frac{\varphi_1^{-1}(\sigma)}{\sigma}\Big) & 1 \leq \sigma \leq \sigma^* \\ H\Big(\frac{\varphi_1^{-1}(\sigma)}{\sigma}\Big) - H\Big(\frac{\varphi_2^{-1}(\sigma)}{\sigma}\Big) & \sigma > \sigma^*, \end{cases} \end{aligned}$$

since $\varphi_2^{-1}(\sigma) = 0$ when $1 \leq \sigma \leq \sigma^*$. □

Figure 2.28 illustrates the unconditional coverage probabilities as a function of σ for the case where $(\nu_2, \nu_1) = (3, 4)$, corresponding to the endpoint functions given in Figure 2.27. Unlike any of the coverage probability graphs seen to date, the

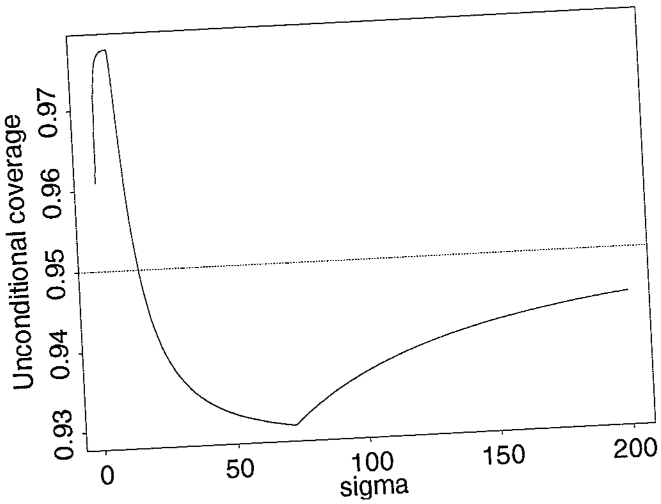

Figure 2.28: Unconditional coverage probabilities of log-shortest 95% PDP-CIs of σ with $(\nu_2, \nu_1) = (3, 4)$.

unconditional coverage of 95% log-shortest PDP-CIs in Figure 2.28 go both above and below the nominal 0.95 coverage.

Theorem 2.24. *The average log-length of a* $1-\alpha$ *log-shortest PDP-CI is, for fixed* $\sigma \geq 1$,

$$E_\sigma(\textit{log-length}) = \int_0^\infty [\log \psi_2(f) - \log \psi_1(f)] h_F(f)\, df, \tag{2.121}$$

where $\psi_1(F)$ *is a lower quantile function and* ψ_2 *an upper quantile function of the PDP-CI.*

The proof follows from the definition of log length.

For the case $(\nu_2, \nu_1) = (3, 4)$, average log-lengths are plotted in Figure 2.29. As with other log-shortest PDP-CIs, the average log-length is increasing in σ.

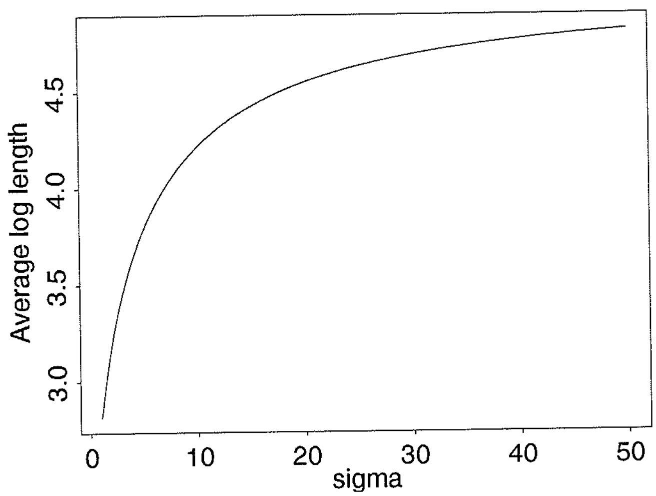

Figure 2.29: Average log-lengths of log-shortest 95% PDP-CI's of σ with $(\nu_2, \nu_1) = (3, 4)$.

2.6.3 Fixed length PDP-CIs

In this section, we consider PDP-CIs for τ_1, τ_2 and σ with the same log length as the usual interval for all $s > 0$. The results presented here are reminiscent of those of Shorrock (1990) or Brewster and Zidek (1974) in the normal variance problem.

Log-shortest fixed length PDP-CIs of τ_1. Consider the problem of finding a log-shortest fixed length PDP-CI of τ_1 based on the distribution of U_1^*. Let $k = b/a$ denote the endpoint ratio of the unconditional log-shortest CI of τ_1 with confidence coefficient $1 - \alpha$. For fixed s, the problem is to find quantiles ξ_1 and ξ_2, functions of s, such that

$$\xi_1 f(\xi_1) = \xi_2 f(\xi_2), \tag{2.122}$$

$$\xi_2/\xi_1 = k, \tag{2.123}$$

where $f(\cdot)$ is the density of $U_1^*(s)$. Thus, $\xi_1(S) = \xi_2(S)/k$, which reduces the problem to one variable. Letting $\psi(S) = \xi_2^{-1}(S)$, the system (2.122) and (2.123) reduces to

$$g_{\nu_1}\left(\frac{1}{\psi(s)}\right) G_{\nu_2}\left(\frac{s}{\psi(s)}\right) = \frac{1}{k}\, g_{\nu_1}\left(\frac{1}{k\,\psi(s)}\right) G_{\nu_2}\left(\frac{s}{k\,\psi(s)}\right). \qquad (2.124)$$

The normalized fixed length log-shortest PDP-CI of τ_1 is then $U_1(\psi(S), k\psi(S))$, where $\psi(s)$ is the pointwise solution of (2.124) which maximizes the conditional coverage

$$P(\psi(S) \le U_1^{-1} \le k\,\psi(S) | S \le s).$$

A graph of the multiplier functions for $(\nu_2, \nu_1) = (3, 4)$ is shown in Figure 2.30; the sharp drop in the multiplier function as s approaches 0 is a consequence of maintaining a fixed endpoint ratio: a small drop in the lower multiplier causes a much sharper drop in the upper multiplier to maintain the constant ratio.

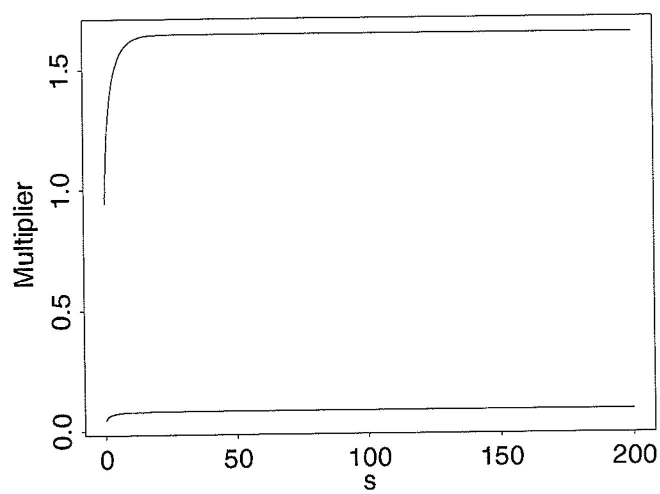

Figure 2.30: Multiplier functions of a fixed length log-shortest PDP-CI of τ_1 with log-length equal to that of the corresponding unconditional log-shortest 95% CI, with $(\nu_2, \nu_1) = (3, 4)$.

The unconditional coverage of this interval is given by (2.105), with $\xi_1(S) = \xi_2(S)/k$. Unconditional coverage probabilities of log-shortest fixed length PDP-CIs of τ_1 are shown in Figure 2.31 for the $(\nu_2, \nu_1) = (3, 4)$ case, from which we find that the nominal 0.95 coverage supplies a lower bound. We will show that this result holds for (at least) log-shortest fixed length PDP-CIs of τ_1, τ_2 and σ.

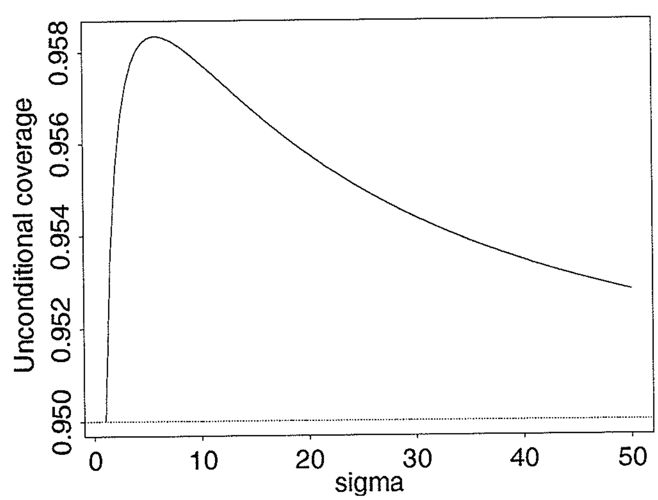

Figure 2.31: Unconditional coverage probabilities of fixed length PDP-CIs of τ_1 as a function of σ by each length criterion, when $(\nu_2, \nu_1) = (3, 4)$.

Log-shortest fixed length PDP-CIs for τ_2. The criteria for producing fixed length PDP-CIs of τ_2 are similar to those for τ_1, the difference being that the underlying family of distributions is now associated with the PDP U_2^*. Substituting the density $f(x) = g_{\nu_2}(x)\,[1 - G_{\nu_1}(x/s)]$ into (2.122), the objective function for obtaining the multiplier $\phi(s)$ of a log-shortest PDP-CI of τ_2 with length $k = b/a$ is

$$g_{\nu_2}\Big(\frac{1}{\psi(s)}\Big)\Big[1 - G_{\nu_1}\Big(\frac{1}{s\,\psi(s)}\Big)\Big] = \frac{1}{k}\,g_{\nu_2}\Big(\frac{1}{k\,\psi(s)}\Big)\Big[1 - G_{\nu_1}\Big(\frac{1}{ks\,(\psi(s))}\Big)\Big]. \qquad (2.125)$$

The unconditional coverage probability of a fixed length PDP-CI is now based on (2.115), with $\zeta_2(S) = [k\psi(S)]^{-1}$ and $\zeta_1(S) = \psi^{-1}(S)$.

Multipliers of fixed length log-shortest PDP-CIs for τ_2 when $(\nu_2, \nu_1) = (3, 4)$ are illustrated in Figure 2.32. Since the endpoint ratio is fixed, both multipliers inflate to infinity as s approaches 0, but the graph is curtailed for small s. Unconditional coverage probabilities of fixed length PDP-CIs of τ_2 are presented in Figure 2.33; once again, the unconditional coverage of these intervals is bounded below by the nominal coverage 0.95.

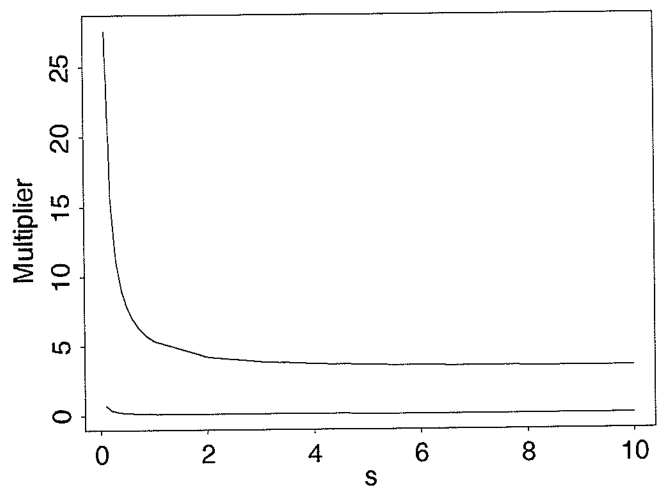

Figure 2.32: Multiplier functions of log-shortest fixed length PDP-CI of τ_2 with nominal coverage 0.95 and $(\nu_2, \nu_1) = (3, 4)$.

Log-shortest fixed length PDP-CIs of σ. As in the case of fixed coverage PDP-CIs of σ, we consider situations where the interval can be two-sided and where it must be one-sided due to the size of the truncation point f. Consider the class of log-shortest intervals of σ of the form $(\psi(F)\, F, k\, \psi(F)\, F)$. The multiplier function $\psi = \psi(F)$ is the unique solution of

$$\frac{1}{k}\, h^*\!\left(\frac{1}{k\,\psi}\right) = h^*\left(\frac{1}{\psi}\right), \tag{2.126}$$

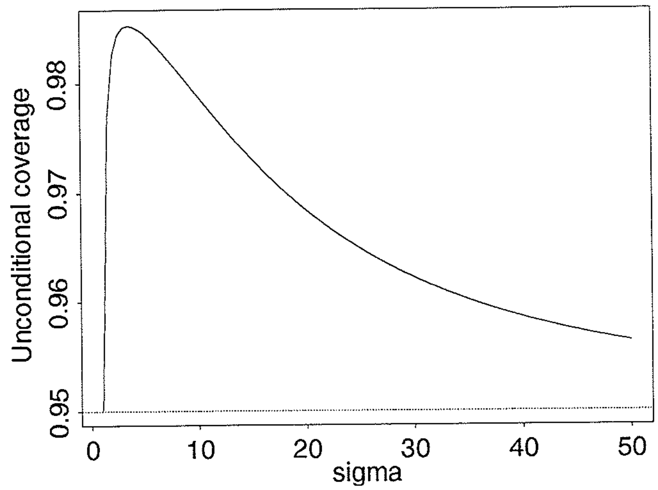

Figure 2.33: Unconditional coverage probabilities of log-shortest fixed length PDP-CI of τ_2 with nominal coverage 0.95 and $(\nu_2, \nu_1) = (3, 4)$, as a function of σ.

where $k = b/a$ is the length of the unconditional log-shortest $1-\alpha$ CI of σ and $h^*(\cdot)$ is the density of an $\mathcal{F}_{\nu_2,\nu_1}$ random variable truncated above at f. When the truncation point f is too small to allow solution of (2.126), we set $\psi = 1/f$. In this case, a fixed length log-shortest PDP-CI of σ is $(1, k)$. As f increases, there exists a value f' such that for all $f > f'$, the system (2.126) can be solved uniquely for $\psi(F)$.

The two-sided log-shortest fixed length PDP-CI of σ has the general form

$$(\psi(F)\, F, k\, \psi(F)\, F) = F\, \psi(F)\, (1, k).$$

The endpoint functions of the PDP-CI of σ with fixed length $k = b/a$ amalgamate both the one-sided and two-sided versions of the interval, and are piecewise continuous with change point k.

The lower and upper endpoint functions of this PDP-CI are then

$$\varphi_1(F) = \begin{cases} 1 & f \le f', \\ \psi(F)\, F & f > f'; \end{cases}$$

$$\varphi_2(F) = \begin{cases} k & f \le f', \\ k\psi(F)\, F & f > f'. \end{cases}$$

This interval becomes the 'usual' one if $\psi(F) = b^{-1}$; since $\psi(\cdot)$ is a monotone nonincreasing function, it follows that $b^{-1} = \inf \psi(F)$. However, when $f < b$, $\psi(F) = 1/f$, so the fixed length interval reduces to

$$(\varphi_1(F), \varphi_2(F)) = \begin{cases} (1, \frac{b}{a}) & f < b, \\ (\frac{F}{b}, \frac{F}{a}) & f \ge b; \end{cases}$$

i. e., the interval is constant at $(1, k)$ whenever the truncation point f is less than b, and becomes the 'usual' interval thereafter. This behavior has interesting ramifications with respect to unconditional coverage. For the (3, 4) case, the endpoint functions of a log-shortest fixed length PDP-CI of σ are shown in Figure 2.34.

Theorem 2.25. *For any $\sigma \ge 1$, the unconditional coverage of a log-shortest fixed length PDP-CI of σ is given by*

$$P_\sigma[\varphi_1(F) \le \sigma \le \varphi_2(F)] = \begin{cases} H(b) & 1 \le \sigma < b/a \\ H(b) - H(a) & \sigma \ge b/a, \end{cases} \tag{2.127}$$

where $k = b/a$ is the endpoint ratio of the usual $1 - \alpha$ log-shortest interval of σ.

Proof. Since each endpoint function is continuous and monotone increasing (φ_1 on $[b, \infty)$ and φ_2 on $[0, \infty)$), their inverse functions exist and can be determined analytically for this type of interval. Beginning with solutions for the lower endpoint function $\varphi_1(F)$,

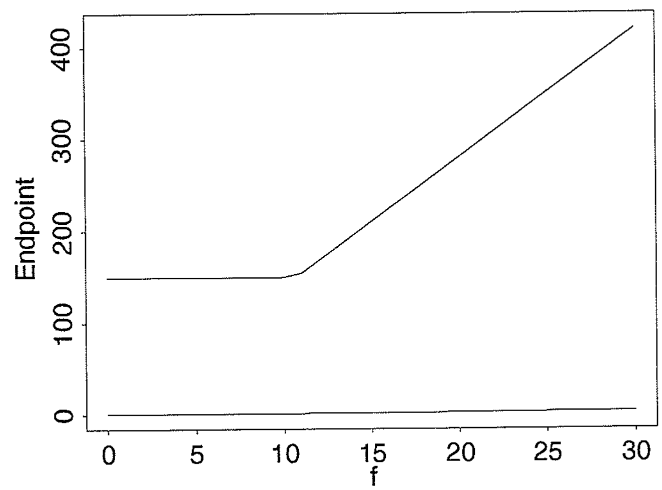

Figure 2.34: Endpoint functions of log-shortest fixed length PDP-CI of σ with $(\nu_2, \nu_1) = (3, 4)$.

- for all $F \leq b$, $\varphi_1(F) = 1$; hence, $\varphi_2^{-1}(1) = \sup_{F \leq b} F = b$.
- for all $F > b$, we have $\varphi_1(F) = F/b$ (i. e., $\sigma = F/b$); then, $F = \varphi_2^{-1}(\sigma) = b\sigma$, so

$$\frac{\varphi_2^{-1}(\sigma)}{\sigma} = b \qquad \text{for all } \sigma \geq 1.$$

Turning to solutions of the upper endpoint function $\varphi_2(F)$, observe that:

- for $1 \leq \sigma \leq b/a$, the endpoint function φ_2 does not intersect σ; hence, we define $\varphi_1^{-1}(\sigma) = 0$ for all σ in this range.
- when $\sigma = b/a$, the endpoint function $\varphi_2(F)$ is intersected for all $f \in [0, b]$; hence, $\varphi_1^{-1}(b/a) = b$, which implies

$$\frac{\varphi_1^{-1}(b/a)}{b/a} = \frac{b}{b/a} = a.$$

- for all $\sigma > b/a$, $\sigma = \varphi_2(F) = F/a$; solving for F yields $\varphi_1^{-1}(\sigma) = a\sigma$, so that

$$\frac{\varphi_1^{-1}(\sigma)}{\sigma} = \frac{a\sigma}{\sigma} = a.$$

Therefore,

$$\frac{\varphi_1^{-1}(\sigma)}{\sigma} = \begin{cases} 0 & \sigma < b/a, \\ a & \sigma \geq b/a. \end{cases}$$

The unconditional coverage probability function for fixed length log-shortest PDP-CIs of σ is then

$$\begin{aligned} P_\sigma[\varphi_1(F) \leq \sigma \leq \varphi_2(F)] &= P_\sigma(\varphi_1^{-1}(\sigma) \leq F \leq \varphi_2^{-1}(\sigma)) \\ &= P_\sigma\left(\frac{\varphi_1^{-1}(\sigma)}{\sigma} \leq \frac{F}{\sigma} \leq \frac{\varphi_2^{-1}(\sigma)}{\sigma}\right) \\ &= \begin{cases} H(b) & 1 \leq \sigma < b/a \\ H(b) - H(a) & \sigma \geq b/a, \end{cases} \end{aligned}$$

where $H(b) - H(a) = 1 - \alpha$, the nominal unconditional coverage. □

For the (3, 4) case, unconditional coverage probabilities of fixed length log-shortest 95% PDP-CIs of σ are displayed in Figure 2.35. More generally, we have the following

Theorem 2.26. *The unconditional coverage of a log-shortest fixed length PDP-CI of σ of the form* (2.6.3) *is* $\geq 1 - \alpha$ *for all* $\sigma \geq 1$.

Proof. Let $1 - \alpha = H(b) - H(a)$, where $H(\cdot)$ is the cdf of an $\mathcal{F}_{\nu_2,\nu_1}$ distribution. By Theorem 2.25, the unconditional coverage of a fixed length log-shortest PDP-CI (2.6.3) is

$$P_\sigma[\psi(F)\, F \leq \sigma \leq k\,\psi(F)\, F] = \begin{cases} H(b) & 1 \leq \sigma < b/a \\ 1 - \alpha & \sigma \geq b/a. \end{cases}$$

Since $1 - \alpha = H(b) - H(a)$, it follows that $H(b) = 1 - \alpha + H(a)$. Since $a = b/k > 0$ (where b is the solution of (2.126) with respect to the unconditional $\mathcal{F}$ density), $H(a) > 0$, from which it follows that $H(b) > 1 - \alpha$ when $1 \leq \sigma < b/a$. This establishes the proof. □

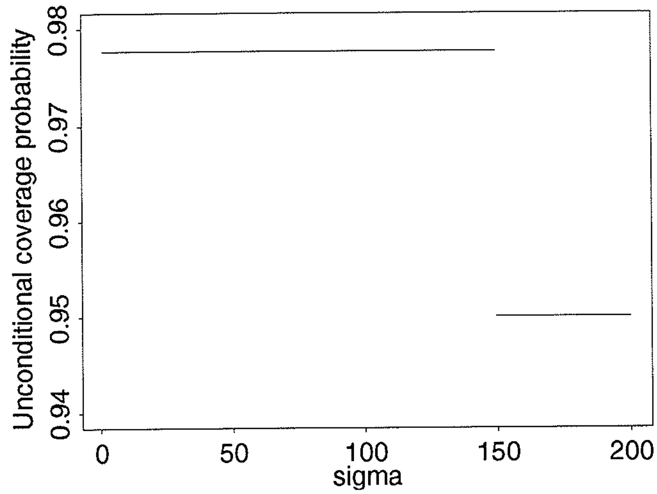

Figure 2.35: Unconditional coverage probabilities of log-shortest fixed length PDP-CIs of σ with $(\nu_2, \nu_1) = (3, 4)$.

2.6.4 Dominance results: fixed length PDP-CIs

2.6.4.1 PDP-CIs of τ_1 and τ_2

The present objectives are to show that the fixed length PDP-CIs of τ_1 and τ_2 are interval analogues of Brewster-Zidek-like point estimators, and to establish that these intervals have higher than nominal unconditional coverage. The following lemma is a modification of Lemma 3.1 in Shorrock (1990).

Lemma 2.27. *Let f and g be unimodal densities, and consider intervals with multipliers of the form $(\psi, k\,\psi)$, where $k = b/a$ is the endpoint ratio. Let ψ_f maximize $\int_{\psi}^{k\psi} f(x)\,dx$, and let ψ_g maximize $\int_{\psi}^{k\psi} g(x)\,dx$. If f/g is increasing (decreasing) in x, then $\psi_f > \psi_g$ ($\psi_f < \psi_g$).*

The usual unconditional interval of a scale parameter θ is of the form $(T/b, T/a)$, where T is a statistic used for estimating θ, and a and b are quantiles of the distribution of $U = T/\theta$ obtained by one of the Tate-Klett length criteria. For an unconditional $1-\alpha$ log-shortest interval, define $\psi = b^{-1}$; the interval is then of the form $(\psi, k\,\psi)\,T$, where $k = b/a$. Fixing a log-shortest PDP-CI to have length k, a solution $\psi = \psi_1$ of (2.124) yields an interval of the form $(\psi_1(S), k\psi_1(S))\,T_1$. The corresponding fixed length PDP-CI for τ_2 is obtained from the solution $\psi = \psi_2$ of (2.125).

The next step is to show that the fixed-length PDP-CIs of τ_1 and τ_2 are of the Brewster-Zidek type. To do this, we first construct a Brown-like fixed-length interval and show that it has uniformly higher than nominal coverage for all $\sigma \geq 1$. We then go through the Brewster-Zidek argument, analogous to the point estimation case, and show that the fixed length PDP-CIs are BZ-like. We then show that the unconditional coverage of BZ-like intervals is at least the nominal coverage uniformly in σ.

2.6.4.2 Brown-like fixed-length intervals

Log-shortest fixed length interval of τ_1. Select an arbitrary $r > 0$, and consider intervals of the form $(\psi, k_1\,\psi)\,T_1$, where $k_1 = b_1/a_1$ is the endpoint ratio of the usual log-shortest interval of τ_1. Let $\psi = \psi_1(r)$ maximize the conditional probability

$$P(\psi\,T_1 \leq \tau_1 \leq k_1\,\psi\,T_1 | S \leq r) = P(\psi \leq U_1^{-1} \leq k_1\,\psi | S \leq r),$$

which is proportional to

$$\int_{\psi}^{k_1\,\psi} g_{\nu_1+2}\Big(\frac{1}{y}\Big)\,G_{\nu_2}\Big(\frac{r}{y}\Big)\,dy,$$

Then, $\psi_1(r)$ is the unique solution of (2.124), with r replacing s; the resulting Brown-like log-shortest fixed length interval of τ_1 is then

$$I_1^L(T_1, S; r) = \begin{cases} (\psi_1(r)\,T_1, k_1\,\psi_1(r)\,T_1) & \text{if } S \leq r, \\ (T_1/b_1, T_1/a_1) & \text{if } S > r. \end{cases} \tag{2.128}$$

Log-shortest fixed length interval of τ_2. Select $r > 0$. Among intervals of the form $(\psi, k_2\,\psi)\,T_2$, where $k_2 = b_2/a_2$ is the endpoint ratio of the usual log-shortest interval of τ_2, let $\psi = \psi_2(r)$ maximize the conditional probability

$$P(\psi\,T_2 \le \tau_2 \le k_2\,\psi\,T_2|S \le r) = P(\psi \le U_2^{-1} \le k_2\,\psi|S \le r),$$

which is proportional to

$$\int_{\psi}^{k_2\,\psi} g_{\nu_2+2}\Big(\frac{1}{y}\Big)\left[1 - G_{\nu_1}\Big(\frac{1}{ry}\Big)\right] dy.$$

Then, $\psi_2(r)$ is the unique solution of (2.125), substituting r for s; the resulting Brown-like log-shortest fixed length interval of τ_2 is then

$$I_2^L(T_2, S; r) = \begin{cases} (\psi_2(r)\,T_2, k_2\,\psi_2(r)\,T_2) & \text{if } S \le r, \\ (T_2/b_2, T_2/a_2) & \text{if } S > r. \end{cases} \tag{2.129}$$

Theorem 2.28. *For any fixed $r > 0$, each of the Brown-type fixed length intervals of τ_1 and τ_2 has unconditional coverage $> 1 - \alpha$ for all $\sigma \ge 1$, where $1 - \alpha$ is the confidence level of the corresponding unconditional interval.*

Proof. Let $r > 0$ be arbitrarily chosen, and define $I_p^L = (T_p/b_p, T_p/a_p)$ to be the usual shortest interval of τ_p, $p = 1, 2$, with length $k_p = b_p/a_p$.

First, consider log-shortest fixed length intervals of τ_1. We need to show that

$$P(\tau_1 \in I_1^L(T_1, S; r)) > P(\tau_1 \in I_1^L) \tag{2.130}$$

for all $r > 0$. Comparing (2.128) and I_1^L, it is seen that the two intervals differ only over the interval $[0, r]$. By construction, the multiplier $\psi = \psi_1(r)$ of (2.128) maximizes

$$P(\psi\,T_1 \le \tau_1 \le k_1\,\psi\,T_1|S \le r) = P(\psi \le U_1^{-1} \le k_1\,\psi|S \le r) \tag{2.131}$$

By Theorem A.14(c), $g^*(u_1|r)/g(u_1)$ is monotone increasing in u_1; by Lemma A.9, $g^*(u_1^{-1}|r)/g(u_1^{-1})$ is monotone decreasing. Since $g(\cdot)$ and $g^*(\cdot)$ are unimodal densities, it follows by Lemma 2.27 that $\psi_1(r) < b_1^{-1}$. Therefore, when r is fixed,

$$\begin{aligned} P(\psi_1(r)\, T_1 \leq \tau_1 \leq k_1\, \psi_1(r)\, T_1 | S \leq r) &= P(\psi_1(r) \leq U_1^{-1} \leq k_1\, \psi_1(r) | S \leq r) \\ &> P(b_1^{-1} \leq U_1^{-1} \leq a_1^{-1} | S \leq r), \qquad (2.132) \end{aligned}$$

which in turn implies (2.130). Since r was arbitrarily chosen, the result follows for intervals of τ_1.

For log-shortest fixed length intervals of τ_2, we need to show that

$$P(\tau_2 \in I_2^L(T_2, S; r)) > P(\tau_2 \in I_2^L)$$

for all $r > 0$. We follow the same line of proof as above: first, find $\psi_2(r)$ as the solution to (2.125), which maximizes the conditional coverage

$$P(\psi\, T_2 \leq \tau_2 \leq k_2\, \psi\, T_2 | S \leq r) = P(\psi \leq U_2^{-1} \leq k_2\, \psi | S \leq r);$$

since the MLR properties are opposite in direction in this case, Theorem A.14(e) and (f), along with Lemmas A.9 and 2.27, are relevant in justification of the result. □

2.6.4.3 BZ-like fixed length intervals

The argument used in section 2.5 to improve upon Brown-like point estimators applies equally to the case of interval estimation. That is, by keeping r fixed, one can select successively smaller constants $r' > r'' > r''' > \ldots$ and construct extended Brown-type step-function intervals with fixed length k which have higher unconditional coverage than all their predecessors, as well as the usual interval.

With this in mind, construct a sequence of partitions $\{R_i\}$ of $[0, \infty)$ with properties (i)–(iv) on p. 30. Within partition R_i, let $E_{ij} = (r_{i,j-1}, r_{ij}]$, $j = 1, 2, \ldots, n_i - 1$, denote the sets over which the step functions are defined. Following the process described in the previous paragraph within each R_i, a best step-function interval estimator $I_i(\cdot)$ can be determined.

To illustrate the Brewster-Zidek construction process for interval estimation, consider log-shortest intervals of τ_1 of the form $(\psi T_1, k_1 \psi T_1)$, where $k_1 = b_1/a_1$. On each partition R_i, construct a step-function interval of the form

$$I_1^{(i)}(T_1, S) = \begin{cases} (\psi_1(r_{ij})\, T_1, k_1\, \psi_1(r_{ij})\, T_1) & r_{i,j-1} < S < r_{ij} \\ (T_1/b_1, T_1/a_1) & S > r_{i,n_i-1}, \end{cases} \tag{2.133}$$

where $\psi_1(r_{ij})$ is the solution of (2.124) on each of the $n_i - 1$ subintervals $(r_{i,j-1}, r_{ij}]$ of R_i, where r_{ij} is substituted for s in the objective function. By repeated application of Theorem 2.28 it follows that

$$P(\tau_1 \in I_1^{(i)}(T_1, S)) > P(\tau_1 \in I_1^L)$$

on each partition R_i.

As $i \to \infty$, the step-function intervals $I_1^{(i)}(T_1, S)$ converge to a limiting interval

$$I_1^*(T_1, S) = (\psi_1(S)\, T_1, k_1\, \psi_1(S)\, T_1), \tag{2.134}$$

where $\psi_1(S)$ is continuous and monotone increasing in S. The interval $I_1^*(T_1, S)$ is a Brewster-Zidek-like log-shortest interval of τ_1 with length k_1, and it is readily shown that $\psi_1(s)$ is the pointwise solution of (2.124) for each $s > 0$; therefore, $I_1^*(T_1, S)$ is precisely the log-shortest fixed-length PDP-CI of τ_1. A similar type of argument shows that the log-shortest PDP-CI of τ_2 with fixed length k_2 is also BZ-like.

Theorem 2.29. *Each fixed-length PDP-CI of τ_1 and τ_2 has unconditional coverage $\geq 1 - \alpha$ for all $\sigma \geq 1$, where $1 - \alpha$ denotes the nominal coverage of the corresponding unconditional interval.*

Proof. We establish the case for the log-shortest fixed length PDP-CIs of τ_1; the τ_2 case follows a parallel line of proof.

Let $\{R_i\}$ denote a sequence of partitions of $[0, \infty)$ with properties (i)–(iv) listed on p. 30, and consider intervals of the form $(\psi T_1, k_1 \psi T_1)$, where k_1 is the length of the usual log-shortest CI. By the argument preceding this this theorem, the sequence of step function intervals $\{I_1^{(i)}(S, T_1)\}$ converges to the Brewster-Zidek-like interval $I_1^*(S, T_1)$ as $i \to \infty$, which coincides with the log-shortest PDP-CI of τ_1 with length k_1 given by (2.134). The multiplier function $\psi(S)$ is monotone increasing and continuous in s: the monotonicity follows by Theorem A.14(b), Lemma A.9 and Lemma 2.27, while the continuity follows from the continuity of the objective function (2.124). Hence,

$$\begin{aligned} P_\sigma(\tau_1 \in I_1^*(T_1, S)) &= P_\sigma(\tau_1 \in \lim_{i\to\infty} I_1^{(i)}(T_1, S)) \\ &\geq \liminf_{i\to\infty} P_\sigma(\tau_1 \in I_1^{(i)}(T_1, S)) && \text{by Fatou's lemma} \\ &\geq P(\tau_1 \in (T_1/b_1, T_1/a_1)) && \text{by Theorem 2.28} \\ &= 1 - \alpha. \end{aligned}$$

□

Remark 2.1. Shorrock (1990) was able to establish strict uniform improvement in unconditional coverage of shortest fixed length BZ-type intervals of σ^2 in the normal variance problem by establishing a completeness theorem that takes advantage of an auxiliary Poisson random variable. In this problem, there is no comparable result, so we cannot formally prove uniform improvement in coverage of the fixed length PDP-CIs at this point. However, Figure 2.31 shows that the unconditional coverage of fixed length PDP-CIs of τ_1 with nominal coverage 0.95 is always 0.95 or larger for all σ. Similar results hold for fixed length PDP-CIs of τ_2. Hence, there is empirical evidence that fixed length PDP-CIs dominate the usual intervals in unconditional coverage, but no formal statement to confirm it as yet.

The next result shows that the fixed length PDP-CIs of τ_1 and τ_2 are generalized Bayes with respect to the prior $(\tau_1\,\tau_2)^{-1}$.

Theorem 2.30. *Under the joint prior $\pi(\mu, \tau_1, \tau_2)$ of Lemma 2.1, the log-shortest fixed length PDP-CIs of τ_1 and τ_2 are log-shortest credible intervals.*

Proof. Let $\pi(\mu, \tau_1, \tau_2) = (\tau_1, \tau_2)^{-1}$. Multiplying this by the joint likelihood function of $(\tau_1, \tau_2)|y$ yields the joint density of (τ_1, τ_2). Derivations of marginal posterior densities based on this prior are given in section 2.4.9.1 and are not reproduced here.

Consider a log-shortest credible interval for τ_1 with fixed length k_1 based on the posterior density (2.46) of τ_1. The multiplier function ψ that produces a shortest interval of length k_1 is the solution $\psi = \psi_1(s)$ of

$$q^*_{\nu_1+2}(\psi)\, G_{\nu_2}\left(\frac{s}{\psi}\right) = q^*_{\nu_1+2}(k_1\,\psi)\, G_{\nu_2}\left(\frac{s}{k_1\,\psi}\right), \tag{2.135}$$

where $q^*(\cdot)$ is an inverse χ^2 density. This is readily seen to be proportional to (2.124), so for each fixed s, the solution $\psi_1(s)$ is the same in each case. Therefore, the log-shortest fixed length credible interval of τ_1 based on the prior $(\tau_1\,\tau_2)^{-1}$ coincides with the log-shortest fixed length PDP-CI of τ_1 pointwise in s, so the PDP-CI is generalized Bayes in the class of intervals of the form $(\psi_1(S)\,T_1, k_1\,\psi_1(S)\,T_1)$ with respect to this prior. A similar line of argument for the τ_2 case establishes the result. □

2.6.5 Interval estimation of ρ and γ

Parametric functions $\gamma = \sigma^2_\alpha/\sigma^2_e$, the variance component ratio, and $\rho = \sigma^2_\alpha/(\sigma^2_\alpha+\sigma^2_e)$, the intraclass correlation coefficient, are easily derived as functions of σ:

$$\sigma = 1 + J\gamma \quad\Longrightarrow\quad \gamma = \frac{\sigma - 1}{J} \tag{2.136}$$

$$\rho = \frac{\sigma^2_\alpha}{\sigma^2_\alpha + \sigma^2_e} = \frac{\gamma}{1+\gamma} = \frac{\sigma - 1}{\sigma + (J-1)} \tag{2.137}$$

Since $\sigma \geq 1$ by the model formulation, it follows that both ρ and γ are nonnegative.

2.6.5.1 PDP-CIs for γ

As shown in Searle, Casella and McCulloch (1992, p. 66), it is possible to obtain interval estimates of γ and ρ based on intervals for σ. Using the $\mathcal{F}$ distribution as reference, an unconditional $1-\alpha$ CI of the variance component ratio γ is

$$P\big(a \leq \frac{F}{\sigma} \leq b\big) = P\big(\frac{F}{b} \leq \sigma \leq \frac{F}{a}\big).$$

Since $\sigma = 1 + J\gamma$,

$$\begin{aligned} &= P\big(\frac{F}{b} \leq 1 + J\gamma \leq \frac{F}{a}\big) \\ &= P\big(\frac{F}{b} - 1 \leq J\gamma \leq \frac{F}{a} - 1\big). \end{aligned}$$

Therefore, an unconditional $1-\alpha$ CI for γ is

$$\frac{1}{J}\left(\frac{F}{b} - 1, \frac{F}{a} - 1\right).$$

A PDP-CI for γ obtains by replacing F/a and F/b with the appropriate endpoint functions from PDP-CIs of σ. That is, a PDP-CI of γ has the general form

$$\frac{1}{J}\left(\varphi_1(F) - 1, \varphi_2(F) - 1\right),$$

where $\varphi_i(F)$ is an endpoint function from a PDP-CI of σ. Any concerns or problems with PDP intervals of σ will propagate to PDP intervals for γ and ρ.

Log-shortest PDP-CI with fixed conditional coverage.

$$\frac{1}{J}\left(\varphi_1(F) - 1, \varphi_2(F) - 1\right) = \begin{cases} \frac{1}{J}\left(0, \frac{f}{H_*^{-1}(\alpha|f)} - 1\right) & \text{if } f \leq f^+ \\ \frac{1}{J}\left(\frac{F}{\xi_2^*(F)} - 1, \frac{F}{\xi_1^*(F)} - 1\right) & \text{if } f > f^+, \end{cases}$$

where $\xi_i^*(\cdot)$ are quantile functions associated with a log-shortest $1-\alpha$ PDP-CI of σ, and f^+ is the transition point between a one-sided and two-sided log-shortest PDP-CI of σ.

Log-shortest fixed length PDP-CI. To obtain a log-shortest fixed length PDP-CI of γ, let $k_2 = b/a$, the endpoint ratio of the unconditional log-shortest CI of σ with nominal coverage $1-\alpha$. The desired PDP-CI is then

$$\frac{1}{J}\left(\varphi_1^*(F)-1, \varphi_2^*(F)-1\right) = \begin{cases} \frac{1}{J}\left(0, \frac{b}{a}-1\right) & \text{if } f \leq b, \\ \frac{1}{J}\left(\frac{F}{b}-1, \frac{F}{a}-1\right) & \text{if } f > b. \end{cases}$$

Observe that the 'usual' CI of γ obtains if $F > b$.

2.6.5.2 PDP-CIs for ρ

The unconditional $1-\alpha$ CI of ρ, the intraclass correlation, is derived in Searle, Casella and McCulloch (1992, ch. 3) as

$$\left(\frac{F-b}{F+(J-1)b}, \frac{F-a}{F+(J-1)a}\right) = \left(\frac{F/b-1}{F/b+(J-1)}, \frac{F/a-1}{F/a+(J-1)}\right),$$

where a and b are the lower and upper quantiles associated with an unconditional $1-\alpha$ interval of σ based on the $\mathcal{F}$ distribution, f is the observed mean square ratio, and J is the number of replicate measurements per cell. As before, we replace a and b with the quantile functions $\xi_1(F)$ and $\xi_2(F)$, respectively, from a PDP-CI of σ. We begin with PDP-CIs of ρ with fixed conditional coverage $1-\alpha$.

PDP-CIs with fixed conditional coverage. A log-shortest $1-\alpha$ PDP-CI of ρ has the form

$$\left(\frac{F-\xi_2(F)}{F+(J-1)\xi_2(F)}, \frac{F-\xi_1(F)}{F+(J-1)\xi_1(F)}\right).$$

Let f' denote the transition point between a one-sided and two-sided interval; the interval can then be expressed as

$$\begin{cases} \left(0, \frac{f-H_*^{-1}(\alpha|f)}{f+(J-1)H_*^{-1}(\alpha|f)}\right) & \text{if } f \leq f'; \\ \left(\frac{F-\xi_2(F)}{F+(J-1)\xi_2(F)}, \frac{F-\xi_1(F)}{F+(J-1)\xi_1(F)}\right) & \text{if } f > f', \end{cases}$$

where $\xi_i(F)$, $i = 1, 2$ is a quantile function associated with a log-shortest PDP-CI of σ with fixed conditional coverage $1 - \alpha$ and $H_*^{-1}(\alpha|f)$ is the α quantile of an $\mathcal{F}_{\nu_2,\nu_1}$ distribution truncated above at f.

PDP-CIs with fixed length. Let $k_2 = b/a$ denote the endpoint ratio of an unconditional log-shortest $1 - \alpha$ CI of σ. A log-shortest, fixed length PDP-CI of ρ is then

$$\left(\frac{\varphi_1(F) - 1}{\varphi_1(F) + J - 1}, \frac{\varphi_2(F) - 1}{\varphi_2(F) + J - 1}\right) = \begin{cases} \left(0, \dfrac{k_2 - 1}{J + k_2 - 1}\right) & \text{if } f \leq b; \\ \left(\dfrac{F - b}{F + (J-1)b}, \dfrac{F - a}{F + (J-1)a}\right) & \text{if } f > b. \end{cases} \tag{2.138}$$

2.6.6 Interval estimation of $J\sigma_\alpha^2$

The distribution of V^* can be viewed as a conditional confidence distribution with regard to interval estimation of $J\sigma_\alpha^2$; i. e., a $1 - \alpha$ PDP confidence interval can be obtained directly from the distribution of V^* for each fixed s. The influences in this case are fiducial and Bayesian; there is no frequentist equivalent. As a result, we do not attempt frequentist evaluation of conditional CIs of $J\sigma_\alpha^2$ from the distribution of V^*.

Two types of interval estimators of $J\sigma_\alpha^2$ will be discussed in this subsection: (i) an interval derived from the density of V^*, and (ii) an alternative to the Tukey-Williams interval. With respect to case (ii), the basic idea is to replace the unconditional quantiles with corresponding PDP counterparts.

2.6.6.1 Intervals based on density of V^*

The density of V^* is unimodal, but may be J-shaped with mode zero. Box and Tiao (1973, Appendix A5.3) showed that the density of V^*, when viewed as a Bayesian posterior, has mode zero whenever

$$s < \frac{\nu_2 + 2}{\nu_1 + \nu_2 + 4};$$

in this case, only one-sided intervals of $J\sigma_\alpha^2$ are possible. When

$$s > \frac{\nu_2 + 2}{\nu_1 + 2},$$

the mode is bounded away from zero, and lies somewhere in the interval

$$\Big(\frac{(\nu_1+2)t_2 - (\nu_2+2)t_1}{(\nu_2+2)(\nu_1+\nu_2+10)}, \big(\frac{t_2}{\nu_2+2} - \frac{t_1}{\nu_1+\nu_2+4}\big)\Big);$$

in this case, two-sided intervals are obtainable.

Since an interval for $J\sigma_\alpha^2$ can be obtained directly from the density of V^*, there is an easy way to determine whether a two-sided interval is obtainable. Let $q_{1-\alpha}$ denote the $1-\alpha$ quantile of V^*, and consider the density ratio

$$d = \frac{g_{V^*}(q_{1-\alpha})}{g_{V^*}(0)}.$$

If $d > 1$, a two-sided interval is obtainable; otherwise, a one-sided interval is necessary.

One-sided intervals. We first consider one-sided $1-\alpha$ PDP-CIs of $J\sigma_\alpha^2$. This case is particularly simple, as 0 is taken as the lower endpoint, and the $1-\alpha$ quantile of V^* is the upper endpoint. To find the latter, determine the value q such that $1-\alpha = P(V^* \le q)$; i. e., q is the root of the nonlinear integral equation

$$\alpha = \frac{\int_0^z G_{\nu_1+\nu_2}\big[\frac{1}{q}\big(\frac{t_2}{v} - \frac{t_1}{1-v}\big)\big]\, g_V(v)\, dv}{I_z(\phi_2, \phi_1)}$$

for given ν_1, ν_2, t_1 and t_2 from the observed data. Derivation of this result is given in section A.8.3 of Appendix A.

Two-sided intervals. For a two-sided PDP-CI of $J\sigma_\alpha^2$ from the distribution of V^*, the following approach can be used to obtain an interval with fixed conditional coverage $1-\alpha$. Let $\psi_1(T_1, T_2)$ and $\psi_2(T_1, T_2)$ denote a lower and upper quantile, respectively, of V^*. To find the values of $\psi_1(\cdot)$ and $\psi_2(\cdot)$, we need to solve a pair of nonlinear equations (one per endpoint). After establishing the preliminary condition that $g_{V^*}(q_{1-\alpha}) > g_{V^*}(0)$, one equation to solve is

$$1-\alpha = \frac{\int_0^z \left[G_{\nu_1+\nu_2}\left(\frac{1}{\psi_1}\left(\frac{t_2}{r} - \frac{t_1}{1-r}\right)\right) - G_{\nu_1+\nu_2}\left(\frac{1}{\psi_2}\left(\frac{t_2}{r} - \frac{t_1}{1-r}\right)\right)\right] g_R(r)\,dr}{I_z(\phi_2, \phi_1)}, \tag{2.139}$$

derived in section A.8.3. The other condition depends on the type of interval desired. For a shortest $1-\alpha$ PDP-CI, the second condition is

$$g_{V^*}(\psi_1) = g_{V^*}(\psi_2),$$

whereas for a log-shortest interval, the second condition would be

$$\psi_1 g_{V^*}(\psi_1) = \psi_2 g_{V^*}(\psi_2).$$

In section 2.4.6, it was noted that the shape of the density of V^* is the same for any fixed s. This suggests constructing a standardized interval estimate of $J\sigma_\alpha^2$ from the density of $V^*(t_1, t_2)$ with $t_1 = 1$ and $t_2 = s$ by any selected optimality criterion; multiplying both endpoints of the standardized interval by t_1 produces the desired interval estimate for any $t_1 > 0$. Experiments in Mathematica verify that such a procedure works in practice.

2.6.6.2 Tukey-Williams intervals and PDP modifications

The Tukey-Williams interval is probably the most popular choice among frequentists for interval estimation of $J\sigma_\alpha^2$. Tukey (1951) observed that one could indirectly obtain an interval for $J\sigma_\alpha^2$ by intersecting intervals of $J\gamma$ with those of τ_2, where both are parametric functions of σ_e^2 and σ_α^2. Williams (1962) refined the method, and obtained a lower bound on the coverage probability of the resulting interval based on the Bonferroni inequality.

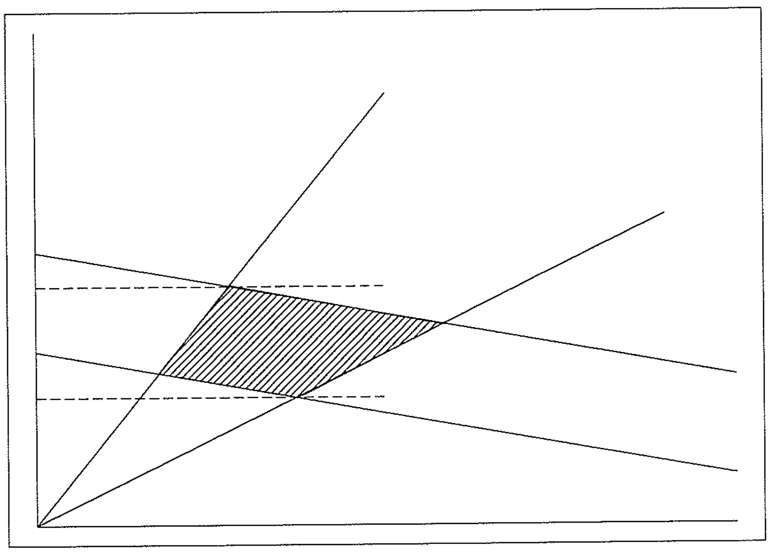

Figure 2.36: Graph of the Tukey-Williams interval of σ_α^2. The intersection of the two intervals induces a rhombus in $(\sigma_e^2, \sigma_\alpha^2)$ space; the Tukey-Williams interval is obtained by projecting this rhombus onto the vertical (σ_α^2) axis. The dotted lines represent the endpoints of the T-W interval; for explanation of the solid lines, refer to the surrounding text.

The idea behind the Tukey-Williams interval is both simple and inspired. One begins by obtaining a joint confidence region for σ_e^2 and $J\sigma_\alpha^2$ by considering the intersection of intervals for both τ_2 and $J\gamma$, as functions of σ_e^2 and $J\sigma_\alpha^2$. Let (l, u) denote the endpoints of a $1-\alpha$ CI of $J\gamma$, and let (L, U) be the corresponding endpoints of a $1-\alpha$ CI of τ_2.

Working in $(\sigma_e^2, J\sigma_\alpha^2)$ space, first consider the positively sloped lines in Figure 2.36:

the line to the left is $J\sigma_\alpha^2 = u\sigma_e^2$, while the other is $J\sigma_\alpha^2 = l\sigma_e^2$. The negatively sloped lines correspond to the endpoints of a τ_2 interval: the lower line is $J\sigma_\alpha^2 = -\sigma_e^2 + L$, while the upper one is $J\sigma_\alpha^2 = -\sigma_e^2 + U$. Intersecting the four lines produces a rhombus in the joint $(\sigma_e^2, J\sigma_\alpha^2)$ parameter space. The desired interval for $J\sigma_\alpha^2$ is obtained by projecting the region of intersection onto the $J\sigma_\alpha^2$ axis.

The standard form of the Tukey-Williams interval is given by

$$\Big(\frac{T_2}{\chi^2_{\nu_2}(1-\alpha/2)}\big[1 - \frac{F_{\nu_2,\nu_1}(1-\beta/2)}{F}\big], \frac{T_2}{\chi^2_{\nu_2}(\alpha/2)}\big[1 - \frac{F_{\nu_2,\nu_1}(\beta/2)}{F}\big]\Big),$$

where $1-\alpha$ is the unconditional coverage of the interval for τ_2, $1-\beta$ is the coverage of the $J\gamma$ interval, $\chi^2(\gamma)$ denotes the γ quantile of the χ^2 distribution with subscripted degrees of freedom, and $F(\eta)$ is the η quantile of the $\mathcal{F}$ distribution with the subscripted pair of degrees of freedom. Typically, one selects $\alpha = \beta$, in which case the unconditional coverage of the Tukey-Williams interval is at least $1-2\alpha$, by the Bonferroni inequality. Simulation studies of Boardman (1974) and Wang (1990), along with theoretical work by Meng (1994), suggest that the actual coverage is actually much closer to $1-\alpha$ if α is taken to be 'small' (say, ≤ 0.1).

Despite the encouraging frequentist properties of the Tukey-Williams interval, its post-data behavior is not as compelling. Assuming an equal tailed interval (as shown above), the lower endpoint becomes negative whenever $F < \mathcal{F}_{\nu_2,\nu_1}(1-\alpha/2)$, which means that with probability $1-\alpha/2$, the lower endpoint will be negative. Similarly, the upper endpoint (and hence, the entire interval estimate) will be negative whenever $F < \mathcal{F}_{\nu_2,\nu_1}(\alpha/2)$, i. e., with $\alpha/2$ probability. Clearly, such intervals are not conditionally acceptable, which provides an opportunity to apply the PDP approach to this procedure to see if its conditional properties can be improved without great loss of pre-data coverage.

The PDP approach modifies the Tukey-Williams interval by replacing the unconditional χ^2 and $\mathcal{F}$ quantiles with their PDP counterparts. The result is an interval for $J\sigma_\alpha^2$ that is guaranteed to have nonnegative endpoints, with conditional coverage at least $1 - 2\alpha$. The unconditional coverage properties of the PDP version of the Tukey-Williams interval are not investigated in this thesis, but 'fixed length' versions should have higher than nominal unconditional coverage, consistent with earlier results in this chapter. On the other hand, it is not at all clear what the frequentist performance of fixed coverage PDP-CIs would be in this case, given the frequentist performance of log-shortest intervals for σ (and therefore, $J\gamma$). Simulation studies would be necessary to answer these questions.

The Tukey-Williams interval for $J\sigma_\alpha^2$ is obtained from intervals for τ_2 and $J\gamma$, both of which have been considered by the PDP approach. The tangible differences between unconditional and PDP intervals lie in the choice of quantiles, and it is no different in this case. The table below summarizes the differences in quantiles for intervals about τ_2 with fixed conditional coverage:

	Unconditional	PDP version
Lower quantile	χ_L^2	$\xi_1(S)$
Upper quantile	χ_U^2	$\xi_2(S)$
Lower endpoint of CI	T_2/χ_U^2	$\psi_1(S)\,T_2$
Upper endpoint of CI	T_2/χ_L^2	$\psi_2(S)\,T_2$,

where χ_L^2 generalizes $\chi_{\alpha/2}^2$ and χ_U^2 generalizes $\chi_{1-\alpha/2}^2$, so that shortest or log-shortest intervals, for example, can be accommodated under this setup.

A similar comparison applies to unconditional and PDP confidence intervals of $J\gamma$ with fixed conditional coverage:

	Unconditional	PDP version
Lower quantile	F_L	$\eta_1(F)$
Upper quantile	F_U	$\eta_2(F)$
Lower endpoint of CI	$\dfrac{F}{F_U} - 1$	$\varphi_1(F) - 1$
Upper endpoint of CI	$\dfrac{F}{F_L} - 1$	$\varphi_2(F) - 1$,

where $\eta_i(F)$ denote the quantile functions of a PDP-CI of σ, and $\varphi_2(F) > \varphi_1(F) \geq 1$.

Making the appropriate substitutions, the PDP-modified version of the Tukey-Williams interval for $J\sigma_\alpha^2$ is

$$\Big(T_2\,\psi_1(S)\,(1 - \frac{1}{\varphi_1(F)}), T_2\,\psi_2(S)\,(1 - \frac{1}{\varphi_2(F)})\Big),$$

with conditional coverage $\geq 1 - 2\alpha$. This version guarantees that the lower endpoint of the interval is nonnegative and the upper endpoint is positive, thus lying completely in the parameter space of $J\sigma_\alpha^2$. (To obtain an interval for σ_α^2, divide both endpoints above by J). Some special cases of the PDP-modified Tukey-Williams interval are given below.

Log-shortest fixed CP. Let f^+ denote the transition point from a one-sided to a two-sided log-shortest PDP-CI of σ with fixed conditional coverage $1 - \alpha$. The corresponding PDP-modified Tukey-Williams interval for $J\sigma_\alpha^2$ is then

$$\begin{cases} \Big(0, T_2\,\psi_2(S)\,(1 - \frac{H_*^{-1}(\alpha|f)}{f})\Big) & \text{if } f \leq f^+, \\ \Big(T_2\,\psi_1(S)\,(1 - \frac{1}{\varphi_1(F)}), T_2\,\psi_2(S)\,(1 - \frac{1}{\varphi_2(F)})\Big) & \text{if } f > f^+. \end{cases}$$

Log-shortest fixed length. Let l_1 denote the length of a log-shortest unconditional $1 - \alpha$ CI of τ_2, and let $l_2 = b/a$ denote the corresponding length of an unconditional $1-\alpha$ CI of σ (again having the same length as a $1-\alpha$ CI for $J\gamma$). The log-shortest PDP-modified 'fixed length' Tukey-Williams interval for $J\sigma_\alpha^2$ is then

$$\begin{cases} \Big(0, l_1\,T_2\,\psi(S)\,(1 - \frac{1}{l_2})\Big) & \text{if } f \leq b, \\ \Big(T_2\,\psi(S)\,(1 - \frac{b}{F}), l_1\,T_2\,\psi(S)\,(1 - \frac{a}{F})\Big) & \text{if } f > b, \end{cases}$$

where a and b denote the respective lower and upper quantiles of a $1-\alpha$ CI for σ.

2.7 Summary

To apply the PDP approach in the one-way random model, the following elements must be in place:

- the treatment variance σ_α^2 must be nonnegative, so that the joint parameter space of the expected mean squares is constrained to the subregion $\tau_2 \geq \tau_1$;
- the probability mass must be distributed over a (joint) pivotal space;
- the data must be balanced, for otherwise the exact distribution theory does not hold.

The PDP conditioning argument is applied on the joint (U_1, U_2) pivotal space, which allows information from both the data and the constrained (τ_1, τ_2) parameter space to combine in forming a conditional reference set for inference. Applying the conditioning event $U_2 \leq sU_1$ to each of the basic pivotals (U_1, U_2 and W) of the model modifies their distributions as a function of the observed SS ratio s.

Distribution theory. The joint PDP density of (U_1, U_2) is defined on the region $\{U_1 > 0, 0 < U_2 \leq sU_1\}$ in the joint (U_1, U_2) pivotal space. The marginal PDP densities of U_1^* and U_2^* are obtained by integration over the joint PDP density, indexed by s for fixed degrees of freedom ν_1 and ν_2.

The families of distributions of U_1^* and U_2^* are complementary to one another as a function of s in the following sense:

- $U_1^*(s)$ and $U_2^*(s)$ are positively correlated for any fixed $s > 0$: the correlation function is monotone decreasing in s, and tends to zero as $s \to \infty$;
- for any fixed $s > 0$, $E(U_1^*(s)) + E(U_1^*(s)) = \nu_1 + \nu_2$, showing that the total degrees of freedom are preserved by PDP conditioning;

- the MLR properties of the families of U_1^* and U_2^* distributions are in opposite directions (Theorem A.14);
- the smooth pooling principle applies to the families of U_1^* and U_2^* distributions as a function of s.

The conditioning event $U_2 \leq sU_1$ implies that $W \leq s$, where $W = U_2/U_1$. Hence, the family of W^* distributions is right truncated at s, whose MLR properties are in the same direction as the family of U_2^* distributions.

Section 2.4.9 demonstrated the connections between the PDP densities and the posterior densities of Box and Tiao (1973, section 5.2) based on a Jeffreys noninformative prior. The results of Wild (1981) extend these connections to corresponding families of fiducial distributions, thereby connecting the three post-data modes of inference through their distribution theory. The density of V^*, the confidence distribution for inference about $J\sigma_\alpha^2$, can be similarly connected to the corresponding Bayesian and fiducial densities. It is motivated by an argument from fiducial inference; given the conditioning event, the distribution varies as a function of both t_1 and t_2 rather than s alone.

Point estimation. Two types of PDP point estimators were introduced in this chapter: PDP-unbiased estimators, obtained by applying the 'usual' unbiased rules to the PDP distributions, and PDP-MLEs, obtained by maximizing a marginal conditional likelihood of τ_1 or τ_2 for any fixed $s > 0$. In terms of risk performance, the PDP-unbiased estimators of τ_1, τ_2 and σ were shown to be at least as good as the corresponding unbiased estimators under entropy loss. Furthermore, these estimators are generalized Bayes with respect to the prior $\pi(\mu, \tau_1, \tau_2)$ given in Lemma 2.1. The generalized Bayes property leaves open the possibility that the PDP-unbiased estimators are admissible, at least in the class of scale equivariant estimators based on

T_1 and T_2. In addition, PDP-unbiased estimators were shown to be of the Brewster-Zidek type. As a result, PDP-unbiased estimators can be justified either by the PDP argument or the Brewster-Zidek argument.

PDP-MLEs can be viewed as 'smooth versions' of REML estimators, since their multiplier functions are differentiable for all $s > 0$. Both types of PDP estimators are amenable to plug-ins: since $J\sigma_\alpha^2$ and $J\gamma$ are linear parametric functions, PDP-unbiased plug-ins are also PDP-unbiased and PDP-ML plug-ins are themselves PDP-ML. PDP plug-in estimators of the intraclass correlation ρ are guaranteed to be nonnegative, but are biased (as are REML and ANOVA plug-ins). The risk performance of the PDP estimators of ρ remains an open question.

Interval estimation. Applying the 'usual' rules for interval estimators to PDP distributions leads to PDP confidence intervals. Particular focus was placed on development of log-shortest PDP-CIs, with either fixed conditional coverage $1-\alpha$ or with its length fixed to that of the corresponding unconditional interval. One general conclusion is that interval estimation of τ_1 and τ_2 is a different problem from interval estimation of σ, γ or ρ.

There is a certain duality between log-shortest PDP intervals of τ_1 and τ_2 with fixed conditional coverage. First of all, $1-\alpha$ PDP-CIs of τ_1 have multiplier functions that shrink as $s \to 0$, while multipliers of τ_2 inflate as $s \to 0$, in concert with the multiplier functions of their respective PDP point estimators. A similar duality occurs with respect to the unconditional coverage of these intervals: those of τ_1 are bounded above by $1-\alpha$, while those of τ_2 are bounded below by $1-\alpha$. Hence, the PDP conditioning event is a negative semirelevant subset with respect to interval estimation of τ_1 with fixed conditional coverage, but is a positive semirelevant subset for similar estimation of τ_2.

Fixed length log-shortest PDP-CIs of τ_1 and τ_2 each have unconditional coverage bounded below by $1 - \alpha$, are of the Brewster-Zidek type, and are generalized Bayes with respect to the prior $\pi(\mu, \tau_1, \tau_2)$ of Lemma 2.1. In this sense, log-shortest fixed length PDP-CIs are at least as good as the corresponding unconditional intervals in terms of unconditional coverage, uniformly in σ.

The performance of log-shortest PDP-CIs for σ appears to be another matter. Application of the log-shortest criterion of Tate and Klett (1959) to a family of truncated distributions creates some difficulties. In particular, the unconditional coverage of $1 - \alpha$ log-shortest PDP-CIs is above $1 - \alpha$ for a certain range of σ values, then drops below $1 - \alpha$, and after reaching some minimum, rises back to $1 - \alpha$ in the limit. It is hard to recommend an interval estimator with such performance, even if it is always conditionally acceptable. For different reasons, the log-shortest fixed length PDP-CI of σ is difficult to recommend. Although its unconditional coverage is at least $1 - \alpha$ uniformly in σ, the form of the interval itself is problematic: given $k = b/a$, the endpoint ratio of the 'usual' log-shortest interval, the 'rule' is to maintain the same one-sided interval $(1, k)$ until the truncation point f reaches the upper quantile b of the unconditional interval, at which point one should switch to the 'usual' interval. Despite its unconditional performance, such an estimator is not intuitively sensible. The problem is rather acute, as intervals for σ can be transformed into intervals for $J\gamma$ and ρ.

Finally, we investigated two approaches to interval estimation of $J\sigma_\alpha^2$: $1 - \alpha$ log-shortest intervals from the distribution of the PDP V^*, analogous to the associated Bayesian credible interval from the posterior of $J\sigma_\alpha^2$ due to Box and Tiao (1973), and PDP analogues to log-shortest versions of the Tukey-Williams interval, for both fixed conditional coverage and fixed length. The advantage of the PDP intervals is that they all yield nonnegative interval estimates of $J\sigma_\alpha^2$; on the other hand, their unconditional properties are as yet an open question.

Chapter 3

Three-stage nested model

The next step in the development of PDP inference for balanced random effects models is to consider the completely random three-stage nested model. In this chapter, we will focus on the PDP-modified distribution theory of the basic (joint) pivotals, and show how PDP point and interval estimators are produced. The essential purpose is to illustrate how PDP inference can be extended beyond the one-way random model. Unconditional evaluations of PDP procedures are not addressed in this chapter.

Let I denote the number of experimental or sampling units at the top stage, J the number of subunits associated with each top stage unit, and K the number of subunits associated with each second-stage unit. Observations are taken at the bottom stage, so there are IJK observational units altogether.

Completely random three-stage nested model. The nonmatrix form of a balanced, nested three-stage random effects model is given by

$$y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \epsilon_{k(ij)}, \tag{3.1}$$

where $i = 1, 2, \ldots, I$, $j = 1, 2, \ldots, J$ and $k = 1, 2, \ldots, K$, and $N = IJK$ denotes the number of observations. The terms of the model are:

- y_{ijk}, the response on the k^{th} subunit of subunit j, nested within unit i;
- μ, the unknown common population mean;
- α_i, the random effect associated with unit i;
- $\beta_{j(i)}$, the random effect associated with subunit j nested within unit i;
- $\epsilon_{k(ij)}$, the random error associated with observation k on subunit j of unit i.

The distributional assumptions underlying this model are:

$$\begin{aligned} \alpha_i &\sim IN(0, \sigma_\alpha^2) \qquad i = 1, 2, \ldots, I; \\ \beta_{j(i)} &\sim IN(0, \sigma_\beta^2) \qquad j = 1, 2, \ldots, J; \\ \epsilon_{k(ij)} &\sim IN(0, \sigma_e^2) \qquad k = 1, 2, \ldots, K, \end{aligned}$$

which implies that

$$\begin{aligned} \text{Var}(y_{ijk}) &= \sigma_\alpha^2 + \sigma_\beta^2 + \sigma_e^2 \\ \text{cov}(y_{ijk}, y_{ijl}) &= \sigma_\alpha^2 + \sigma_\beta^2 \qquad k \neq l \\ \text{cov}(y_{ijk}, y_{iml}) &= \sigma_\alpha^2 \qquad j \neq m, k \neq l \\ \text{cov}(y_{ijk}, y_{nml}) &= 0 \qquad i \neq n, j \neq m, k \neq l. \end{aligned}$$

The covariance matrix of the IJK responses is then of compound symmetric form

$$\boldsymbol{V} = \sigma_e^2(\boldsymbol{I}_I \otimes \boldsymbol{I}_J \otimes \boldsymbol{I}_K) + \sigma_\beta^2(\boldsymbol{I}_I \otimes \boldsymbol{I}_J \otimes \boldsymbol{J}_K) + \sigma_\alpha^2(\boldsymbol{I}_I \otimes \boldsymbol{J}_J \otimes \boldsymbol{J}_K),$$

where $\boldsymbol{I}_p$ is a $p \times p$ identity matrix and $\boldsymbol{J}_q$ is a $q \times q$ matrix of ones.

The analysis of variance table associated with model (3.1) is

Source	df	SS	MS	Expected MS
A	ν_3	T_3	M_3	$\tau_3 = \sigma_e^2 + K\sigma_\beta^2 + JK\sigma_\alpha^2$
B(A)	ν_2	T_2	M_2	$\tau_2 = \sigma_e^2 + K\sigma_\beta^2$
Error	ν_1	T_1	M_1	$\tau_1 = \sigma_e^2$
Total	$\nu_1 + \nu_2 + \nu_3$	$T_1 + T_2 + T_3$		

We assume that each of the variance components σ_e^2, σ_β^2 and σ_α^2 is nonnegative, thus ordering the expected mean squares, which induces the parameter constraint

$$\tau_3 \geq \tau_2 \geq \tau_1;$$

estimation procedures should be constructed to obey this set of constraints. Based on a prospective data vector $\boldsymbol{Y}$, the joint minimal sufficient statistic is $T(\boldsymbol{Y}) = (\overline{Y}, T_1, T_2, T_3)$, but we will focus only on (T_1, T_2, T_3), as in the one-way model.

We begin the process of obtaining PDP point and interval estimators by deriving the PDP reference distributions, whose starting point is the set of basic pivotals and their distributions:

$$U_i = \frac{T_i}{\tau_i} \sim \chi^2_{\nu_i} \qquad i = 1, 2, 3,$$

and

$$W_j = \frac{U_{j+1}}{U_j} \sim \text{Beta}_{II}(\phi_{j+1}, \phi_j) \qquad j = 1, 2,$$

where $\phi_j = \nu_j/2$.

The parameter constraints imply that $W_1 \leq S_1$ and $W_2 \leq S_2$ pre-data, where S_1 and S_2 are interpretable as random upper bounds on the pivotals W_1 and W_2, respectively. Given the data, realizations s_1 of S_1 and s_2 of S_2 become available, and the wedge $\{U_2 \leq s_1 U_1, U_3 \leq s_2 U_2\}$ in the joint (U_1, U_2, U_3) pivotal space becomes the reference set for inference about the variance functions of interest. Each of the basic pivotals is conditioned on this reference set, so that point and interval estimates of the variance functions are based on the set of PDP-conditioned distributions with fixed (s_1, s_2). PDP procedures are obtained by evaluating the conditional rules over all (s_1, s_2) pairs.

In the three-stage nested random model, the composite conditioning event is denoted as

$$C(s_1, s_2) = \{U_2 \leq s_1 U_1, U_3 \leq s_2 U_2\} \tag{3.2}$$

for any fixed $s_1, s_2 > 0$, where $s_j = t_{j+1}/t_j$, $j = 1, 2$. Since $U_3 = U_1 U_2$, it follows by extension that $U_3 \leq s_1 s_2 U_1$. (In general, for hierarchical sets of expected mean squares whose variance components are nonnegative, there is one constraint per prospective F test, and one component conditioning event per constraint.)

3.1 PDP distribution theory

Joint PDP density of (U_1, U_2, U_3). For any fixed $s_1, s_2 > 0$, the joint density of $U_1, U_2, U_3 \,|\, C$ is given by

$$g^*(u_1, u_2, u_3 | s_1, s_2) = \frac{g_{\nu_1}(u_1)\, g_{\nu_2}(u_2)\, g_{\nu_3}(u_3)}{P(W_1 \leq s_1, W_2 \leq s_2)}, \tag{3.3}$$

defined on the region $\{U_1 > 0, U_2 \leq s_1 U_1, U_3 \leq s_2 U_2\}$, the region in the positive octant of $\Re^3$ bounded by the U_1 axis, the plane $U_2 = s_1 U_1$ and the plane $U_3 = s_2 U_2$. (The parameter space lies to the left of the plane $\tau_2 = \tau_1$, to the right of the τ_2 axis, and above the plane $\tau_3 = \tau_2$.). The starred superscript indicates that conditioning on the set C has taken place for fixed (s_1, s_2).

Normalizing constant. The normalizing constant of the joint density is the denominator term in (3.3), expressible as

$$P(W_1 \leq s_1, W_2 \leq s_2) = \int_0^\infty \int_0^{s_1 u_1} \int_0^{s_2 u_2} g_{\nu_1}(u_1)\, g_{\nu_2}(u_2)\, g_{\nu_3}(u_3)\, du_3\, du_2\, du_1. \tag{3.4}$$

This probability can be obtained in several different ways, depending on one's choice of transformations and the direction one chooses to integrate.

Two derivations of this probability statement are given in section B.1, which become relevant when deriving the moments of the marginal PDP densities. One form of the normalizing constant is given by

$$P(W_1 \leq s_1, W_2 \leq s_2) = \int_0^{s_1} \int_0^{s_2} \frac{\Gamma(\phi_1 + \phi_2 + \phi_3)}{\Gamma(\phi_1)\Gamma(\phi_2)\Gamma(\phi_3)} \frac{v_2^{\phi_2+\phi_3-1}\, v_3^{\phi_3-1}}{(1 + v_2 + v_2 v_3)^{-(\phi_1+\phi_2+\phi_3)}}\, dv_3\, dv_2; \tag{3.5}$$

the preferred version is a representation of a Dirichlet cdf, given by

$$I_{z_1,z_2}(\phi_3, \phi_2, \phi_1) = \int_0^{z_1} \int_0^{z_2} \frac{\Gamma(\phi_1 + \phi_2 + \phi_3)}{\Gamma(\phi_1)\Gamma(\phi_2)\Gamma(\phi_3)} v_2^{\phi_2 - 1} v_3^{\phi_3 - 1} (1 - v_2 - v_3)^{\phi_1 - 1} \, dv_3 \, dv_2. \tag{3.6}$$

See Theorems B.1 and B.2, respectively. Equations (3.5) and (3.6) are equivalent in the sense that the integrands (and integration limits) are 1–1 transformations of one another:

$$w_1 = \frac{v_2}{1 + v_2 + v_2 v_3} \qquad v_2 = \frac{w_1}{1 - w_1 - w_2}$$

$$w_2 = \frac{v_2 v_3}{1 + v_2 + v_2 v_3} \qquad v_3 = \frac{w_2}{1 - w_1 - w_2}$$

3.1.1 PDP distributions of the U_i^*

All of the following densities and moments exist for any $s_1, s_2 > 0$; over all (s_1, s_2) pairs, a family of densities is produced for each fixed (σ_1, σ_2) pair, where $\sigma_1, \sigma_2 \geq 1$. The conditioning set (3.2) will be denoted by C, where the argument in parentheses is generally dropped.

Density of $U_1, U_2|C$. By integrating (3.3) over U_3,

$$\begin{aligned} g^*(u_1, u_2|s_1, s_2) &= \int_0^{s_2 u_2} \frac{g_{\nu_1}(u_1)\, g_{\nu_2}(u_2)\, g_{\nu_3}(u_3)\, du_3}{K} \\ &= \frac{g_{\nu_1}(u_1)\, g_{\nu_2}(u_2)\, G_{\nu_3}(s_2 u_2)}{K}, \end{aligned}$$

where K is the normalizing constant (3.6); this density is defined over the region $\{u_1 > 0, 0 < u_2 \leq s_1 u_1\}$.

Density of $U_2, U_3|C$. To integrate (3.3) over U_1, we need to use the fact that $U_1 \geq U_2/s_1$:

$$g^*(u_2, u_3|s_1, s_2) = \int_{u_2/s_1}^{\infty} \frac{g_{\nu_1}(u_1)\, g_{\nu_2}(u_2)\, g_{\nu_3}(u_3)\, du_1}{K}$$

$$= \frac{[1 - G_{\nu_1}(u_2/s_1)]\, g_{\nu_2}(u_2)\, g_{\nu_3}(u_3)}{K},$$

defined over the region $\{u_2 > 0, 0 < u_3 \leq s_2 u_2\}$.

Density of $U_1, U_3|C$. To integrate over U_2, we need to bear in mind that U_2 is constrained from both sides: i.e., $U_2 \leq s_1 U_1$ and $U_3 \leq s_2 U_2$. Hence, $U_3/s_2 \leq U_2 \leq s_1 U_1$. Integrating (3.3) over this region, we obtain

$$g^*(u_1, u_3|s_1, s_2) = \int_{u_3/s_2}^{s_1 u_1} \frac{g_{\nu_1}(u_1)\, g_{\nu_2}(u_2)\, g_{\nu_3}(u_3)\, du_2}{K}$$

$$= \frac{g_{\nu_1}(u_1)\, g_{\nu_3}(u_3)\, [G_{\nu_2}(s_1 u_1) - G_{\nu_2}(u_3/s_2)]}{K},$$

defined over the region $\{u_1 > 0, 0 < u_3 \leq s_1 s_2 u_1\}$.

Marginal densities of the $U_i^* = U_i|C$ can be obtained in three different ways: double integration over the joint density (3.3), or integration over the two bivariate densities that involve U_i^*, $i = 1, 2, 3$.

Marginal density of $U_1|C$. From the joint density (3.3), the marginal density of $U_1^* = U_1|C$ is given by

$$g^*(u_1|s_1, s_2) = \int_0^{s_1 u_1} \int_0^{s_2 u_2} \frac{g_{\nu_1}(u_1)\, g_{\nu_2}(u_2)\, g_{\nu_3}(u_3)\, du_3\, du_2}{K}$$

$$= g_{\nu_1}(u_1) \int_0^{s_1 u_1} \int_0^{s_2 u_2} \frac{g_{\nu_2}(u_2)\, g_{\nu_3}(u_3)\, du_3\, du_2}{K}$$

$$= g_{\nu_1}(u_1) \int_0^{s_1 u_1} \frac{g_{\nu_2}(u_2)\, G_{\nu_3}(s_2 u_2)\, du_2}{K}.$$

From the bivariate density of $U_1, U_2|C$, the marginal density of U_1^* is obtained as

$$\begin{aligned} g^*(u_1|s_1,s_2) &= \int_0^{s_1u_1} \frac{g_{\nu_1}(u_1)\,g_{\nu_2}(u_2)\,G_{\nu_3}(s_2u_2)\,du_2}{K} \\ &= g_{\nu_1}(u_1) \int_0^{s_1u_1} \frac{g_{\nu_2}(u_2)\,G_{\nu_3}(s_2u_2)\,du_2}{K}, \end{aligned}$$

identical to that obtained from the joint density.

From the bivariate density of $U_1, U_3|C$, the marginal density of U_1^* is given by

$$\begin{aligned} g^*(u_1|s_1,s_2) &= \int_0^{s_1s_2u_1} \frac{g_{\nu_1}(u_1)\,g_{\nu_3}(u_3)\,[G_{\nu_2}(s_1u_1) - G_{\nu_2}(u_3/s_2)]\,du_3}{K} \\ &= g_{\nu_1}(u_1) \int_0^{s_1s_2u_1} \frac{g_{\nu_3}(u_3)\,[G_{\nu_2}(s_1u_1) - G_{\nu_2}(u_3/s_2)]\,du_3}{K} \\ &= \frac{g_{\nu_1}(u_1)\left[G_{\nu_2}(s_1u_1)\,G_{\nu_3}(s_1s_2u_1) - \int_0^{s_1s_2u_1} g_{\nu_3}(u_3)\,G_{\nu_2}(u_3/s_2)\right] du_3}{K}, \end{aligned}$$

from which one can deduce that

$$G_{\nu_2}(s_1u_1)\,G_{\nu_3}(s_1s_2u_1) = \int_0^{s_1u_1} g_{\nu_2}(u_2)\,G_{\nu_3}(s_2u_2)\,du_2 + \int_0^{s_1s_2u_1} g_{\nu_3}(u_3)\,G_{\nu_2}(u_3/s_2)\,du_3.$$

Marginal density of U_2^*. From the joint density (3.3), the density of U_2^* is obtained as follows:

$$\begin{aligned} g^*(u_2|s_1,s_2) &= \int_{u_2/s_1}^{\infty} \int_0^{s_2u_2} \frac{g_{\nu_1}(u_1)\,g_{\nu_2}(u_2)\,g_{\nu_3}(u_3)\,du_3\,du_1}{K} \\ &= \frac{[1 - G_{\nu_1}(u_2/s_1)]\,g_{\nu_2}(u_2)\,G_{\nu_3}(s_2u_2)}{K}. \end{aligned}$$

Integrating over each bivariate density involving U_2 yields the same result.

Marginal density of U_3^*. From (3.3), the marginal density of U_3^* is

$$g^*(u_3|s_1,s_2) = \int_{u_2/s_1}^{\infty}\int_{u_3/s_2}^{\infty} \frac{g_{\nu_1}(u_1)\, g_{\nu_2}(u_2)\, g_{\nu_3}(u_3)\, du_2\, du_1}{K}$$

$$= g_{\nu_3}(u_3) \int_{u_3/s_2}^{\infty}\int_{u_2/s_1}^{\infty} \frac{g_{\nu_1}(u_1)\, g_{\nu_2}(u_2)\, du_1\, du_2}{K}$$

$$= g_{\nu_3}(u_3) \int_{u_3/s_2}^{\infty} \frac{g_{\nu_2}(u_2)\, [1 - G_{\nu_1}(u_2/s_1)]\, du_2}{K}$$

$$= \frac{g_{\nu_3}(u_3) \left[(1 - G_{\nu_2}(u_3/s_2)) - \int_{u_3/s_2}^{\infty} g_{\nu_2}(u_2)\, G_{\nu_1}(u_2/s_1)\, du_2\right]}{K}.$$

By integrating over the joint density of $U_2, U_3|C$, the same result obtains. On the other hand, integrating over the joint density of $U_1, U_3|C$ produces

$$g^*(u_3|s_1,s_2) = \frac{g_{\nu_3}(u_3) \int_{u_3/s_1 s_2}^{\infty} g_{\nu_1}(u_1)\, [G_{\nu_2}(s_1 u_1) - G_{\nu_2}(u_3/s_2)]\, du_1}{K}$$

$$= \frac{g_{\nu_3}(u_3) \left[\int_{u_3/s_1 s_2}^{\infty} g_{\nu_1}(u_1)\, G_{\nu_2}(s_1 u_1)\, du_1 - G_{\nu_2}(u_3/s_2)\, [1 - G_{\nu_1}(u_3/s_1 s_2)]\right]}{K}.$$

Comparing the two expressions for the density of U_3^*, it follows that

$$1 - G_{\nu_2}(u_3/s_2)\, G_{\nu_1}(u_3/s_1 s_2) = \int_{u_3/s_1 s_2}^{\infty} g_{\nu_1}(u_1)\, G_{\nu_2}(s_1 u_1)\, du_1 + \int_{u_3/s_2}^{\infty} g_{\nu_2}(u_2)\, G_{\nu_1}(u_2/s_1)\, du_2. \quad (3.7)$$

3.1.2 Moments of the U_i^*

In the balanced one-way random model, it was shown that the sum of the conditional means preserved the total degrees of freedom; i. e.,

$$E(U_1^*) + E(U_2^*) = \nu_1 + \nu_2 = E(U_1) + E(U_2),$$

where $U_i^* = U_i|U_2 \leq sU_1$, $i = 1, 2$. In section B.2 of Appendix B, it is shown that this result generalizes to the present model (Lemma B.3), and more generally, to any balanced normal-theory random effects model under appropriate (PDP) conditioning (Theorem B.4 and the subsequent remarks).

The following results are a consequence of Theorem B.11, which establishes a general expression for product moments of U_1, U_2 and U_3.

Lemma 3.1. *For any fixed $s_1, s_2 > 0$, the k^{th} moments of $U_i|C$ are given by*

$$E(U_1^{*\,k}) = 2^k \frac{\Gamma(\phi_1 + k)}{\Gamma(\phi_1)} \frac{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1 + k)}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)}$$

$$E(U_2^{*\,k}) = 2^k \frac{\Gamma(\phi_2 + k)}{\Gamma(\phi_2)} \frac{I_{z_1,z_2}(\phi_3, \phi_2 + k, \phi_1)}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)}$$

$$E(U_3^{*\,k}) = 2^k \frac{\Gamma(\phi_3 + k)}{\Gamma(\phi_3)} \frac{I_{z_1,z_2}(\phi_3 + k, \phi_2, \phi_1)}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)},$$

respectively, for $k = 1, 2, 3, \ldots$. In particular, the means correspond to $k = 1$ and the second moments to $k = 2$; therefore, the variances are given by

$$\mathrm{Var}(U_1|C) = \nu_1(\nu_1 + 2) \frac{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1 + 2)}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)} - \nu_1^2 \frac{I^2_{z_1,z_2}(\phi_3, \phi_2, \phi_1 + 1)}{I^2_{z_1,z_2}(\phi_3, \phi_2, \phi_1)} \tag{3.8}$$

$$\mathrm{Var}(U_2|C) = \nu_2(\nu_2 + 2) \frac{I_{z_1,z_2}(\phi_3, \phi_2 + 2, \phi_1)}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)} - \nu_2^2 \frac{I^2_{z_1,z_2}(\phi_3, \phi_2 + 1, \phi_1)}{I^2_{z_1,z_2}(\phi_3, \phi_2, \phi_1)} \tag{3.9}$$

$$\mathrm{Var}(U_3|C) = \nu_3(\nu_3 + 2) \frac{I_{z_1,z_2}(\phi_3 + 2, \phi_2, \phi_1)}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)} - \nu_3^2 \frac{I^2_{z_1,z_2}(\phi_3 + 1, \phi_2, \phi_1)}{I^2_{z_1,z_2}(\phi_3, \phi_2, \phi_1)} \tag{3.10}$$

3.1.3 Covariances among the U_i^*

Lemma 3.2. *For any fixed $s_1, s_2 > 0$,*

$$E(U_1U_2|C) = \nu_1\,\nu_2 \frac{I_{z_1,z_2}(\phi_3, \phi_2 + 1, \phi_1 + 1)}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)}, \tag{3.11}$$

$$E(U_1U_3|C) = \nu_1\,\nu_3 \frac{I_{z_1,z_2}(\phi_3 + 1, \phi_2, \phi_1 + 1)}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)}, \tag{3.12}$$

$$E(U_2U_3|C) = \nu_2\,\nu_3 \frac{I_{z_1,z_2}(\phi_3 + 1, \phi_2 + 1, \phi_1)}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)}. \tag{3.13}$$

Proof. Apply Theorem B.11 with $(k_3, k_2, k_1) = (0,1,1)$, $(1,0,1)$ and $(1,1,0)$, in order. □

The covariance terms, conditional on the event C, are given as follows:

Lemma 3.3. *For any fixed $s_1, s_2 > 0$, covariances among the pairs U_i, U_j, conditioned on the event C, for $i < j$, $i = 1,2$, $j = 2,3$, are given by*

$$\operatorname{cov}(U_1, U_2|C) = \nu_1\nu_2\left[\frac{I_{z_1,z_2}(\phi_3, \phi_2+1, \phi_1+1)}{I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)} - \frac{I_{z_1,z_2}(\phi_3,\phi_2,\phi_1+1)I_{z_1,z_2}(\phi_3,\phi_2+1,\phi_1)}{I^2_{z_1,z_2}(\phi_3,\phi_2,\phi_1)}\right] \tag{3.14}$$

$$\operatorname{cov}(U_1, U_3|C) = \nu_1\nu_3\left[\frac{I_{z_1,z_2}(\phi_3+1, \phi_2, \phi_1+1)}{I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)} - \frac{I_{z_1,z_2}(\phi_3,\phi_2,\phi_1+1)I_{z_1,z_2}(\phi_3+1,\phi_2,\phi_1)}{I^2_{z_1,z_2}(\phi_3,\phi_2,\phi_1)}\right] \tag{3.15}$$

$$\operatorname{cov}(U_2, U_3|C) = \nu_2\nu_3\left[\frac{I_{z_1,z_2}(\phi_3+1, \phi_2+1, \phi_1)}{I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)} - \frac{I_{z_1,z_2}(\phi_3,\phi_2+1,\phi_1)I_{z_1,z_2}(\phi_3+1,\phi_2,\phi_1)}{I^2_{z_1,z_2}(\phi_3,\phi_2,\phi_1)}\right]. \tag{3.16}$$

The proof follows from Lemmas 3.2 and 3.1, along with the definition of covariance.

3.1.4 Distribution theory of the $W_i|C$

The pivotal ratios W_i are defined as

$$W_1 = \frac{U_2}{U_1}, \qquad W_2 = \frac{W_3}{W_2}, \qquad W_3 = W_1W_2 = \frac{U_3}{U_1}.$$

Ascertainment of the relevant densities and moments among the $W_i|C$ involves much more detail than do the densities and moments of the PDP's $U_i|C$. Consequently, these results are detailed in Appendix B. Section B.3 is concerned with results involving the joint and marginal densities of the $W_i|C$, while section B.5.2 is concerned with the (product) moments and covariances among the $W_i|C$.

3.1.5 Confidence distributions for $K\sigma_\beta^2$ and $JK\sigma_\alpha^2$

In the balanced three-stage nested model, $K\sigma_\beta^2 = \tau_2 - \tau_1$ and $JK\sigma_\alpha^2 = \tau_3 - \tau_2$. Analogous to the case of the one-way random model, define two quantities V_1^* and V_2^* (where the starred superscript denotes conditioning on the event C), such that

$$V_1^*(t_1, t_2, t_3) = \frac{t_2}{U_2} - \frac{t_1}{U_1}\Big|W_1 \le s_1, W_2 \le s_2 \tag{3.17}$$

is used for inference about $K\sigma_\beta^2$, while

$$V_2^*(t_1, t_2, t_3) = \frac{t_3}{U_3} - \frac{t_2}{U_2}\Big|W_1 \le s_1, W_2 \le s_2 \tag{3.18}$$

is used for inference about $JK\sigma_\alpha^2$.

3.1.5.1 Density and moments of V_1^*

Density of V_1^*. Following the derivation in Appendix B.6, the density of V_1^* is

$$g^*(v_1|\boldsymbol{t}) = \int_0^\infty \frac{\left(\frac{t_1}{v_2}\right)^{\phi_1+1}\left(\frac{t_2}{v_1+v_2}\right)^{\phi_2+1}\exp\Big\{-\frac{1}{2}\Big(\frac{t_1}{v_2}+\frac{t_2}{v_1+v_2}\Big)\Big\}\,G_{\nu_3}\Big(\frac{s_2 t_2}{v_1+v_2}\Big)}{t_1 t_2\,\Gamma(\phi_1)\Gamma(\phi_2)\,2^{\phi_1+\phi_2}\,I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)}\,dv_2, \tag{3.19}$$

where $\boldsymbol{t} = (t_1, t_2, t_3)$. As was the case in the one-way random model, the marginal density of V_1^* does not admit a closed form solution.

Moments of V_1^*. From section B.6.1 of Appendix B, it is shown that

$$E(V_1^*|\boldsymbol{t}) = \frac{t_2}{2}\Big[\frac{\Gamma(\phi_2-1)}{\Gamma(\phi_2)}\,\frac{I_{z_1,z_2}(\phi_3,\phi_2-1,\phi_1)}{I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)} - \frac{1}{s_1}\,\frac{\Gamma(\phi_1-1)}{\Gamma(\phi_1)}\,\frac{I_{z_1,z_2}(\phi_3,\phi_2,\phi_1-1)}{I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)}\Big]. \tag{3.20}$$

Since

$$E(V_1^{*2}|\boldsymbol{t}) = E\Big[\Big(\frac{t_2}{U_2} - \frac{t_1}{U_1}\Big)^2 \Big| W_1 \le s_1, W_2 \le s_2\Big],$$

expansion of the integral leads to

$$E(V_1^{*2}|\boldsymbol{t}) = \Big(\frac{t_2}{2}\Big)^2 \Bigg[\frac{\Gamma(\phi_2-2)}{\Gamma(\phi_2)} \frac{I_{z_1,z_2}(\phi_3,\phi_2-2,\phi_1)}{I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)} - \frac{2}{s_1}\frac{\Gamma(\phi_1-1)\Gamma(\phi_2-1)}{\Gamma(\phi_1)\Gamma(\phi_2)} \frac{I_{z_1,z_2}(\phi_3,\phi_2-1,\phi_1-1)}{I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)} + \frac{1}{s_1^2}\frac{\Gamma(\phi_1-2)}{\Gamma(\phi_1)} \frac{I_{z_1,z_2}(\phi_3,\phi_2,\phi_1-2)}{I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)}.\Bigg] \tag{3.21}$$

Compare $E(V_1^*|\boldsymbol{t})$ and $E(V_1^{*2}|\boldsymbol{t})$ with (2.32) and (2.34), respectively, from Chapter 2.

Extending (2.35) to the r^{th} moment of $V_1^*|\boldsymbol{t}$, we obtain

$$\Big(\frac{t_2}{2}\Big)^r \sum_{k=0}^{r} \binom{r}{k} \Big(-\frac{1}{s_1}\Big)^k \frac{\Gamma(\phi_1-k)\,\Gamma(\phi_2-r+k)}{\Gamma(\phi_1)\,\Gamma(\phi_2)} \frac{I_{z_1,z_2}(\phi_3,\phi_2-r+k,\phi_1-k)}{I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)}, \tag{3.22}$$

denoted by $E(V_1^{*r}|\boldsymbol{t})$.

cdf of V_1^*. The cdf of V_1^* is obtained by integration over (3.19). The recipe for deriving tail probabilities or cumulative probabilities over finite intervals follows exactly the same process as that for the cdf of V^* in the one-way random model; the details are shown in section A.8.3 of Appendix A. The same process applies equally to the cdf of V_2^* below.

3.1.5.2 Density and moments of V_2^*

As shown in Appendix B.6.2, the marginal density of $V_2^*|\boldsymbol{t}$ is

$$\int_0^\infty \frac{\Big(\frac{t_2}{w_2}\Big)^{\phi_2+1}\Big(\frac{t_3}{w_2+w_3}\Big)^{\phi_3+1}\exp\Big\{-\frac{1}{2}\Big(\frac{t_2}{w_2}+\frac{t_3}{w_2+w_3}\Big)\Big\}\,[1-G_{\nu_1}(t_1/w_2)]}{t_2 t_3\,\Gamma(\phi_2)\Gamma(\phi_3)\,2^{\phi_2+\phi_3}\,I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)}\,dw_2,$$

denoted by $g^*(v_2|\boldsymbol{t})$, and defined over the positive reals.

In the same section of Appendix B, it is shown that

$$E(V_2^*|\boldsymbol{t}) = \frac{t_3}{2}\Big[\frac{\Gamma(\phi_3-1)}{\Gamma(\phi_3)}\,\frac{I_{z_1,z_2}(\phi_3-1,\phi_2,\phi_1)}{I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)} - \frac{1}{s_2}\,\frac{\Gamma(\phi_2-1)}{\Gamma(\phi_2)}\,\frac{I_{z_1,z_2}(\phi_3,\phi_2-1,\phi_1)}{I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)}\Big]; \tag{3.23}$$

in general, the r^{th} moment of $V_2^*|\boldsymbol{t}$ is

$$\left(\frac{t_3}{2}\right)^r \sum_{k=0}^{r} \binom{r}{k} \left(-\frac{1}{s_2}\right)^k \frac{\Gamma(\phi_2-k)\,\Gamma(\phi_3-r+k)}{\Gamma(\phi_2)\,\Gamma(\phi_3)}\,\frac{I_{z_1,z_2}(\phi_3-r+k,\phi_2-k,\phi_1)}{I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)}.$$

3.2 Unbiased and PDP point estimation

In this section, we review the 'usual' unbiased point estimators of the variance functions of interest, and obtain corresponding PDP point estimators. The essential difference between PDP point estimation in the three-stage nested model relative to the one-way random model is that monotone likelihood ratio (MLR) properties are no longer guaranteed to hold. As a result, we can no longer assert that PDP estimators are 'BZ-like', since MLR properties are essential to the limiting argument leading to BZ-like estimators.

3.2.1 Parametric functions

In one sense, the parametric functions of interest in the three-stage nested model are a straightforward extension of those considered in the one-way random model. However, the relationships among certain parametric functions are more complex in the nested model.

We begin with the expected mean squares:

$$\tau_3 = \sigma_e^2 + K\sigma_\beta^2 + JK\sigma_\alpha^2$$
$$\tau_2 = \sigma_e^2 + K\sigma_\beta^2$$
$$\tau_1 = \sigma_e^2.$$

Define the following variance component ratios:

$$\gamma_1 = \frac{\sigma_\beta^2}{\sigma_e^2} \qquad \gamma_2 = \frac{\sigma_\alpha^2}{\sigma_\beta^2} \qquad \gamma_3 = \frac{\sigma_\alpha^2}{\sigma_e^2};$$

Then, $\gamma_3 = \gamma_1\gamma_2$. Next, define the expected mean square ratios as

$$\sigma_1 = \frac{\tau_2}{\tau_1} = 1 + K\gamma_1$$
$$\sigma_2 = \frac{\tau_3}{\tau_2} = 1 + \frac{JK\sigma_\alpha^2}{\sigma_e^2 + K\sigma_\beta^2}$$
$$\sigma_3 = \frac{\tau_3}{\tau_1} = 1 + K\gamma_1 + JK\gamma_3.$$

Expressing the variance component ratios in terms of the σ's, we have that

$$K\gamma_1 = \sigma_1 - 1 \tag{3.24}$$
$$JK\gamma_3 = \sigma_3 - \sigma_1. \tag{3.25}$$

Next, observe that

$$JK\gamma_3 = JK\gamma_1\gamma_2 = J\gamma_2(\sigma_1 - 1);$$

thus, after a little algebra, we have

$$J\gamma_2 = \frac{\sigma_3 - 1}{\sigma_1 - 1} - 1. \tag{3.26}$$

The intraclass correlations are

$$\rho_i = \frac{\sigma_i^2}{\sigma_e^2 + \sigma_\beta^2 + \sigma_\alpha^2},$$

$i \in I$, where $I = \{e, \beta, \alpha\}$ is an index set.

3.2.2 Unconditional point estimators

The 'usual' unbiased estimators of the expected mean squares τ_i in the balanced, three-stage, nested random model are given by $\hat{\tau}_i = T_i/\nu_i$, $i = 1, 2, 3$. Pre-data, these estimators are UMVU for their respective estimands under normality. By the plug-in principle, unbiased estimators of $K\sigma_\beta^2$ and $JK\sigma_\alpha^2$ are given by

$$J\hat{\sigma}_\beta^2 = \frac{T_2}{\nu_2} - \frac{T_1}{\nu_1};$$

$$JK\hat{\sigma}_\alpha^2 = \frac{T_3}{\nu_3} - \frac{T_2}{\nu_2}.$$

These estimators are also UMVU for their respective variance components under normality (Graybill, 1976).

Unbiased estimators of the expected mean square ratios are given by

$$\hat{\sigma}_1 = \frac{\nu_1 - 2}{\nu_2} S_1,$$

$$\hat{\sigma}_2 = \frac{\nu_2 - 2}{\nu_3} S_2,$$

$$\hat{\sigma}_3 = \frac{\nu_1 - 2}{\nu_3} S_1 S_2.$$

Consequently, unbiased estimators of the variance component ratios $K\gamma_1$ and $JK\gamma_3$ are obtainable by the plug-in principle:

$$K\hat{\gamma}_1 = \hat{\sigma}_1 - 1 = \frac{\nu_1 - 2}{\nu_2} S_1 - 1$$

$$JK\hat{\gamma}_3 = \hat{\sigma}_3 - \hat{\sigma}_1 = (\nu_1 - 2)\, S_1 \left[\frac{S_2}{\nu_3} - \frac{1}{\nu_2}\right].$$

On the other hand, the plug-in estimator of $J\gamma_2$ is

$$J\hat{\gamma}_2 = \frac{JK\hat{\gamma}_3}{K\hat{\gamma}_1} = \frac{(\nu_1 - 2)S_1(\nu_2 S_2 - \nu_3)}{\nu_3((\nu_1 - 2)S_1 - \nu_2)},$$

a biased estimator of $J\gamma_2$.

PDP point estimators of these variance functions are straightforward extensions of the methodology developed in the one-way random model of Chapter 2. We consider estimators of τ_i of the form $\phi_i(S_1, S_2)\, T_i$, $i = 1, 2, 3$, which yields

$$\hat{\tau}_1^* = \frac{T_1}{\nu_1} \frac{I_{Z_1,Z_2}(\phi_3, \phi_2, \phi_1)}{I_{Z_1,Z_2}(\phi_3, \phi_2, \phi_1 + 1)}$$

$$\hat{\tau}_2^* = \frac{T_2}{\nu_2} \frac{I_{Z_1,Z_2}(\phi_3, \phi_2, \phi_1)}{I_{Z_1,Z_2}(\phi_3, \phi_2 + 1, \phi_1)}$$

$$\hat{\tau}_3^* = \frac{T_3}{\nu_3} \frac{I_{Z_1,Z_2}(\phi_3, \phi_2, \phi_1)}{I_{Z_1,Z_2}(\phi_3 + 1, \phi_2, \phi_1)}.$$

Each of these estimators is conditionally unbiased for its estimand with respect to its corresponding PDP distribution. Applying the plug-in principle, the PDP point estimators of $K\sigma_\beta^2$ and $JK\sigma_\alpha^2$ are, respectively,

$$K\tilde{\sigma}_\beta^2 = \hat{\tau}_2^* - \hat{\tau}_1^*,$$

$$JK\tilde{\sigma}_\alpha^2 = \hat{\tau}_3^* - \hat{\tau}_2^*;$$

these estimators are always nonnegative, in concert with results from the balanced one-way random model.

PDP point estimators of the σ_i are given by

$$\hat{\sigma}_1^* = \frac{\nu_1 - 2}{\nu_2} \frac{I_{Z_1,Z_2}(\phi_3, \phi_2, \phi_1)}{I_{Z_1,Z_2}(\phi_3, \phi_2 + 1, \phi_1 - 1)} S_1$$

$$\hat{\sigma}_2^* = \frac{\nu_2 - 2}{\nu_3} \frac{I_{Z_1,Z_2}(\phi_3, \phi_2, \phi_1)}{I_{Z_1,Z_2}(\phi_3 + 1, \phi_2 - 1, \phi_1)} S_2$$

$$\hat{\sigma}_3^* = \frac{\nu_1 - 2}{\nu_3} \frac{I_{Z_1,Z_2}(\phi_3, \phi_2, \phi_1)}{I_{Z_1,Z_2}(\phi_3 + 1, \phi_2, \phi_1 - 1)} S_1 S_2;$$

these, too, are conditionally unbiased for their estimands with respect to their corresponding PDP distributions, and each $\hat{\sigma}_i^*$ assumes values in the interval $[1, \infty)$, the (marginal) parameter space.

Applying the plug-in principle once again, we find that

$$K\hat{\gamma}_1 = \hat{\sigma}_1^* - 1$$
$$JK\hat{\gamma}_3 = \hat{\sigma}_3^* - \hat{\sigma}_1^*$$

are conditionally unbiased for $K\gamma_1$ and $JK\gamma_3$, respectively. Consequently, the PDP plug-in estimator of $J\gamma_2$ is

$$J\hat{\gamma}_2 = \frac{JK\hat{\gamma}_3}{K\hat{\gamma}_1} = \frac{\hat{\sigma}_3^* - \hat{\sigma}_1^*}{\hat{\sigma}_1^* - 1}.$$

Decision theory aspects of PDP point estimators. The natural approach to justifying PDP estimators in the balanced three-stage nested model is to extend the ideas of Brewster and Zidek (1974) and Kubokawa, et al (1998). However, an implicit assumption of both methods is that the appropriate monotone likelihood ratio properties hold. Unfortunately, MLR properties do not extend to the three-stage nested model. To see this, consider estimation of τ_2: the distribution of $U_2|C$ is affected by T_3 from 'above' (in an ANOVA table sense) and by T_1 from 'below'. As a result, there is no guarantee that monotonicity can be maintained since the sizes of T_1 and T_3 affect the (conditional) distribution of $U_2|C$, and may pull it in different directions. Consequently, we cannot claim that the PDP estimators described above are of the Brewster-Zidek type. On the other hand, we *can* appeal to the decision theory properties of PDP estimators in the balanced one-way random model as an indication of the soundness of PDP methodology as a general inferential device when the expected mean squares are constrained by the model formulation. No claims of uniform improvement over unconditional unbiased estimators are made at this stage; this will be left as one topic for future research.

3.3 Interval estimation

Interval estimators for the parametric functions of interest in the balanced three-stage nested model are obtained by straightforward extension of the methodology developed in Chapter 2, with some notable exceptions: (a) the variance component ratios γ_2 and γ_3, and (b) the intraclass correlations ρ_i, $i = 1, 2, 3$. Interval estimators of the variance components $K\sigma_\beta^2$ and $JK\sigma_\alpha^2$ will be derived from their respective (families of) confidence distributions. Estimators of interest are once again log-shortest intervals. We restrict attention to intervals with fixed conditional coverage; intervals with fixed length can be obtained in a manner analogous to that described in section 2.6.

3.3.1 Interval estimation of τ_i

The 'usual' interval for τ_i, $i = 1, 2, 3$, is of the form $(T_i/b_i, T_i/a_i)$, where a_i and b_i are lower and upper quantiles of a χ^2 distribution with ν_i degrees of freedom, respectively, such that the probability content between a_i and b_i is $1 - \alpha$. A PDP-CI of τ_i with fixed conditional coverage $1 - \alpha$ is of the general form $(\psi_i^{(1)}(S_1, S_2)T_i, \psi_i^{(2)}(S_1, S_2)T_i)$, where $\psi_i^{(1)}(\cdot)$ and $\psi_i^{(2)}(\cdot)$ are multiplier functions that depend on both S_1 and S_2, with probability $1 - \alpha$ between their reciprocals for each fixed (s_1, s_2) pair, i. e.,

$$\begin{aligned}1 - \alpha &= P(\xi_i^{(1)}(S_1, S_2) \le U_i \le \xi_i^{(2)}(S_1, S_2) | C) \\ &= P\Big(\frac{1}{\psi_i^{(2)}(S_1, S_2)} \le U_i \le \frac{1}{\psi_i^{(1)}(S_1, S_2)}\Big| C\Big)\end{aligned}$$

where

$$\xi_i^{(1)}(S_1, S_2) = \frac{1}{\psi_i^{(2)}(S_1, S_2)} \qquad \text{and} \qquad \xi_i^{(2)}(S_1, S_2) = \frac{1}{\psi_i^{(1)}(S_1, S_2)}$$

are the quantile functions of the PDP-CI evaluated with respect to the PDP density of U_i conditioned on the event C.

Let $g_i^*(u_i|s_1, s_2)$ denote the PDP density of U_i conditioned on the event C, and let $G_i^*(u_i|s_1, s_2)$ denote its corresponding CDF. Then, $1 - \alpha$ log-shortest PDP-CIs of τ_i are obtained by solving the nonlinear system

$$c_i g_i^*(c_i|s_1, s_2) = d_i g_i^*(d_i|s_1, s_2) \quad (3.27)$$

$$G_i^*(d_i|s_1, s_2) - G_i^*(c_i|s_1, s_2) = 1 - \alpha \quad (3.28)$$

for $c_i = \xi_i^{(1)}(s_1, s_2)$ and $d_i = \xi_i^{(2)}(s_1, s_2)$, with s_1 and s_2 fixed. A computer algebra system such as Mathematica or Maple can be used for this purpose.

3.3.2 Interval estimation of σ_i

The PDP distributions $W_i|W_1 \leq s_1, W_2 \leq s_2$, $i = 1, 2, 3$, are all right truncated, so the methodology from Chapter 2 regarding interval estimation of an expected mean square ratio can be extended to the balanced nested model. Again, attention is limited to intervals with fixed conditional coverage.

As in the one-way random model, it is convenient to work with truncated $\mathcal{F}$ distributions, so define

$$\frac{F_1}{\sigma_1} = \frac{\nu_1 S_1}{\nu_2 \sigma_1}, \qquad \frac{F_2}{\sigma_2} = \frac{\nu_2 S_2}{\nu_3 \sigma_2}, \qquad \frac{F_3}{\sigma_3} = \frac{\nu_1 S_1 S_2}{\nu_3 \sigma_1 \sigma_2},$$

to be the reference pivotals, with observed values

$$f_1 = \frac{\nu_1 s_1}{\nu_2}, \qquad f_2 = \frac{\nu_2 s_2}{\nu_3}, \qquad f_3 = \frac{\nu_1 s_1 s_2}{\nu_3}.$$

The conditioning event is then $C^*(f_1, f_2) = \{F_1/\sigma_1 \leq f_1, F_2/\sigma_2 \leq f_2\}$.

In general, deriving a PDP-CI for σ_i with fixed conditional coverage $1 - \alpha$ entails the following considerations:

1. When f_1 or f_2 are sufficiently small, the Tate-Klett criteria applied to a truncated F distribution may not be capable of providing a two-sided interval (as

in the one-way case). The family of densities associated with $F_i/\sigma_i|C^*$ are right truncated, so one can find a value f_i^*, the minimum observed F_i ratio capable of producing a log-shortest two-sided PDP-CI for σ_i, $i = 1, 2, 3$. Separate numerical root finding routines can be developed to find f_i^* for each distribution.

2. When $f_i < f_i^*$, a lower one-sided interval is only obtainable with lower endpoint 1, as in the one-way problem.

3.3.2.1 Log-shortest PDP-CIs for σ_i

Let $H_i^{-1}(\alpha|F_1, F_2)$ denote the lower α quantile of the PDP density of F_i/σ_i conditioned on the event $C^*(f_1, f_2)$ (C^* for short), truncated on the right at f_i. The quantile functions of a shortest $1-\alpha$ PDP-CI of σ_i are given by

$$\zeta_i^{(1)}(F_1, F_2) = \begin{cases} H_i^{-1}(\alpha|f_1, f_2) & f_i < f_i^* \\ \xi_i^{(1)}(F_1, F_2) & f_i \geq f_i^* \end{cases}$$

and

$$\zeta_i^{(2)}(F_1, F_2) = \begin{cases} f_i & f_i < f_i^* \\ \xi_i^{(2)}(F_1, F_2) & f_i \geq f_i^*, \end{cases}$$

where $\xi_i^{(1)}(\cdot)$ and $\xi_i^{(2)}(\cdot)$ are solutions c_i and d_i, respectively, of the nonlinear system

$$c_i\, h_i^*(c_i|f_1, f_2) = d_i\, h_i^*(d_i|f_1, f_2) \tag{3.29}$$

$$H_i^*(d_i|f_1, f_2) - H_i^*(c_i|f_1, f_2) = 1 - \alpha, \tag{3.30}$$

taken over all possible (F_1, F_2) pairs. The term $h_i^*(\cdot)$ denotes the density function of $F_i/\sigma_i|C^*$, with $H_i^*(\cdot)$ designating the corresponding cdf.

The above notation is rather messy and potentially confusing, so let's pause a moment and consider it in greater detail. The quantile functions for a PDP-CI of

σ_i are based on the distribution of F_i/σ_i conditioned on the event C^*, truncated on the right at f_i. The shape of these PDP distributions is affected by the conditioning event C^*, so quantile and endpoint functions of PDP-CIs for σ_i are dependent on both truncation points f_1 and f_2.

When the distribution is truncated at an f_i below the threshold f_i^* required to achieve a two-sided interval, the quantiles of the lower one-sided interval are the lower α quantile $H_i^{-1}(\alpha|f_1, f_2)$ and the terminus f_i, since there is conditional probability content $1-\alpha$ between them. Conversely, if the truncation point f_i is larger than the minimum threshold f_i^* required to achieve a two-sided interval, then the Tate-Klett criteria (3.29) and (3.30) can be applied to yield a lower quantile function $\xi_i^{(1)}(F_1, F_2)$ and an upper quantile function $\xi_i^{(2)}(F_1, F_2)$, which are determined pointwise from all possible (F_1, F_2) pairs. The $\zeta_i(\cdot)$'s, then, combine quantiles from both one-sided and two-sided versions of a PDP-CI for σ_i, $i = 1, 2, 3$.

Over all (F_1, F_2) pairs, the $\zeta_i(\cdot)$'s are 'envelopes' in the positive octant of $\Re^3$, like the quantiles of the τ_i estimators earlier in this chapter.

Endpoint functions of a log-shortest $1-\alpha$ PDP-CI of σ_i are of the form

$$(\varphi_1(F_1, F_2), \varphi_2(F_1, F_2)\,) = \left(\frac{F_i}{\zeta_i^{(2)}(F_1, F_2)}, \frac{F_i}{\zeta_i^{(1)}(F_1, F_2)}\right),$$

so that

$$\varphi_1(F_1, F_2) = \begin{cases} 1 & f_i < f_i^* \\ F_i/\xi_i^{(2)}(F_1, F_2) & f_i \geq f_i^*, \end{cases}$$

and

$$\varphi_2(F_1, F_2) = \begin{cases} f_i/H_i^{-1}(\alpha|f_1, f_2) & f_i < f_i^* \\ F_i/\xi_i^{(1)}(F_1, F_2) & f_i \geq f_i^*. \end{cases}$$

As in the one-way problem, one can approximate the limiting values of the one-sided upper endpoints $f_i/H_i^{-1}(\alpha|f_1, f_2)$ as $f_i \to 0$, $i = 1, 2, 3$.

3.3.3 Interval estimation of $K\sigma_\beta^2$ and $JK\sigma_\alpha^2$

Since the distributions of V_1^* and V_2^* can be interpreted as confidence distributions, the approach described in Chapter 2 for interval estimation of $J\sigma_\alpha^2$ based on the distribution of V^* carries over directly to interval estimation of $K\sigma_\beta^2$ from V_1^* and $JK\sigma_\alpha^2$ from V_2^*. In particular, derivation of the cdf's is a direct translation of the process described in section A.8.3.

The major difference between interval estimation for variance components in the nested model from that in the one-way random model is the absence of a Tukey-Williams type frequentist confidence procedure. In nested models, several types of frequentist interval procedures have been catalogued in Burdick and Graybill (1992). Like Tukey-Williams intervals, these procedures generally do not have good conditional properties.

3.4 Summary

The process of PDP inference established in the one-way random model has been shown to extend rather readily to the balanced, normal-theory, three stage nested model (3.1). Assuming that the variance components σ_α^2 and σ_β^2 are nonnegative, and σ_e^2 is positive, the joint (τ_1, τ_2, τ_3) parameter space is constrained to the wedge $\{\tau_1 > 0, \tau_2 \geq \tau_1, \tau_3 \geq \tau_2\}$. Given the observed sums of squares (t_1, t_2, t_3) and the constrained parameter space, the PDP conditioning argument produces the event (3.2) in the joint (U_1, U_2, U_3) pivotal space. This event C is the reference set for conditional inference by the PDP approach with respect to the nested model (3.1).

Distribution theory. PDP distributions for the $U_i^* = U_i|C$ and the W_i^* are derived rather easily in the three-stage nested model. There are more distributions to deal with, but the ideas from the one-way random model extend in a straightforward manner. In particular, the consonance between PDP densities and the posterior densities of Box and Tiao (1973) is retained under this model. Moment formulas for each of the U_i^* and W_i^* turn out to be constant multiples of ratios of Dirichlet cdf's.

Distributions of V_1^* and V_2^* are conditional confidence distributions for inference about $K\sigma_\beta^2$ and $JK\sigma_\alpha^2$, respectively. The distribution theory basically follows the script established in Chapter 2 with respect to the distribution of V^*. The moments and cdf's of V_1^* and V_2^* are quite similar in form to that of V^* and are derived by similar means. In addition, the distributions of V_1^* and V_2^* dovetail with the posteriors of $K\sigma_\beta^2$ and $JK\sigma_\alpha^2$ obtained by Box and Tiao.

Inference for intraclass correlations in the balanced three-stage nested model has not been dealt with in this chapter. Box and Tiao (1973, section 5.3) derived a joint posterior density for the intraclass correlations, so it should be possible to translate it to the PDP approach to the problem. This is another topic for future research.

PDP point estimation. In Chapter 2, best equivariant estimators under entropy loss, along with corresponding PDP estimators, had the property that the multiplier of the estimator was the reciprocal of the expected value of its reference distribution. The same idea is applied to point estimation under the balanced nested model (3.1). Multipliers of unconditional, unbiased estimators of τ_i are of the form $T_i/E(U_i)$, while unbiased estimators of the σ_i are $S_i/E(W_i)$, $i = 1, 2, 3$. PDP point estimators are unbiased with respect to the PDP distributions; thus, a PDP-unbiased estimator of τ_i is $T_i/E(U_i^*)$, whereas a PDP-unbiased estimator of σ_i is $S_i/E(W_i^*)$.

One advantage of PDP-unbiased estimators over unbiased estimators is that they are guaranteed to lie within the appropriate parameter space. Since PDP estimators are conditionally unbiased, one can apply the plug-in principle to obtain PDP estimators of each variance component and the variance component ratio $K\gamma_1$; these will always assume nonnegative values.

Not all of the news is positive, though. For example, we cannot claim that PDP estimators are Brewster-Zidek-like, due to the lack of consistent monotone likelihood ratio properties. Loss of this property fundamentally changes the approach by which unconditional evaluations of conditional point estimators take place.

Despite this problem, the results from Chapter 2 inspire confidence that PDP estimators have good statistical properties in more complex ANOVA models. Comparisons of risk performance along the lines of Chapter 2 can be done, but in this case, the risk functions depend on both σ_1 and σ_2. So the question is: to what extent do REML and PDP estimators dominate unconditional unbiased estimators with respect to (joint) entropy loss when the risk function is a surface in $\Re^3$?

PDP interval estimation. It is fairly easy to produce PDP-CIs for the τ_i and σ_i, as the methodology is a straightforward extension of the work on interval estimation in Chapter 2. However, unconditional coverage probabilities and average lengths of PDP-CIs of the τ_i are functions of both σ_1 and σ_2. The unconditional properties of these procedures are still an open question.

Unfortunately, interval estimation of the variance component ratios and intraclass correlations is not a simple matter in the balanced nested model. Clearly, further work in this direction is necessary as well.

Chapter 4

Summary and future directions

The theory of post-data pivotal inference has been developed in detail for the balanced one-way random model, and extended to the three-stage nested model. The methodology is straightforward to apply, and the dictum "do what you normally would, but use the PDP distributions instead of the usual ones" is a simple way to describe the method in the class of problems to which it applies. In this final chapter, some issues and opportunities associated with PDP inference are discussed.

4.1 Single-parameter problems

The PDP method was introduced through two simple examples that have created some controversy in the literature. In both problems, (i) a constraint was imposed on the parameter space and (ii), the underlying distribution could be converted to pivotal form so that the parameter of interest was embedded in the pivotal. To apply the PDP approach to such problems, both elements must be in place.

Uniform $(0, \theta)$ **problem.** In this case, the goal was to find a log-shortest 95% CI of θ. The PDP solution was identical to the Bayesian solution of Wolpert (1988), yielding an interval that lies completely in the parameter space with 95% conditional coverage. As noted by Casella (1988), the unconditional coverage of this procedure is bounded above by 0.95, and tends to zero as $\theta \to \theta^*$, where θ^* is the upper boundary on the parameter space. Hence, the PDP conditioning event is a negative semirelevant set for intervals with fixed conditional coverage. We have seen in Chapter 2 that PDP intervals with fixed length have uniformly higher than nominal unconditional coverage, but in this problem, the rationale for fixed length intervals is suspect since the parameter space $(0, \theta^*)$ is bounded.

$N(\mu, 1)$ problem with $\mu \geq 0$. In this case, we found that the PDP point estimator of μ was generalized Bayes. It is also rather easy to show that intervals with fixed conditional coverage are bounded above by the nominal coverage $1 - \alpha$, so again the issue of negative semirelevant subsets arises with respect to interval estimation.

Since the PDP conditioning event is a fixed set in the (joint) pivotal space post-data, it can act as a negative or positive semirelevant set with respect to interval estimation depending on the type of interval desired and the monotone likelihood ratio properties of the underlying family of PDP distributions. This topic has not been given much attention in this thesis, but it is pertinent since the PDP approach is a form of conditional inference.

4.2 Motivations for the PDP approach

In the normal variance problem, Brewster and Zidek (1974) showed that a point estimator of σ^2 of the form $\phi^{**}(Z)\,T$ dominates the usual unbiased estimator under entropy loss, where

$$\phi^{**}(Z) = \frac{1}{E_1(U|Z \leq z)}.$$

This estimator is also generalized Bayes with respect to a prior that is a mixture of Gaussian pdf's. A similar type of result holds for a wide class of loss functions, in that the dominating estimator is based on the conditional distribution of T given $Z = z$ and $\eta = 0$, rather than the unconditional distribution of T. If we go beyond the estimation problem, two interesting questions arise:

- why should one condition on $Z \leq z$?
- why is the B-Z estimator generalized Bayes under a prior that is a mixture of Gaussian pdf's?

These questions have never been directly addressed to our knowledge, but they have played an important role in the development of PDP inference in the balanced one-way random model. In particular, it impelled us to consider the problem from its origins in the model, which eventually forced us to consider how, and why, it relates to other forms of post-data inference. The result can be summarized by the 'PDP argument', which is stated below for balanced random effects models:

1. Start with the usual normal-theory model, explicitly assuming that $\sigma_e^2 > 0$ and the other variance components are nonnegative. This induces an ordering on the expected mean squares, constraining the joint parameter space of the τ_i's to some nonnegative subregion.
2. Construct a conditioning event in the joint pivotal space of the U_i's that incorporates the parameter constraints on the expected mean squares and the data through ratios of sums of squares. The resulting event is a subregion of the joint pivotal space which serves as a reference set for inference.
3. Condition each of the basic pivotals of the model by the event in (2). The effect is to modify the distribution theory to be data-dependent.

4. Conduct inference in the 'usual' manner, but base the procedures on the families of PDP distributions rather than the usual unconditional distributions.

The above prescription for inference is very simple, and can be logically deduced from the model assumptions. In the absence of nonnegativity constraints on the variance components, the methods of Smith and Murray (1984) or the AVE method of Hocking (1996) are more likely to be appropriate. In such cases, the PDP method does not apply.

4.3 PDP approach: one-way random model

With respect to the balanced one-way random model, implementation of the PDP argument yields the following results:

(i) The PDP distributions have direct connections to the Bayesian distributions of Box and Tiao (1973, section 5.2) as well as the fiducial distributions of Wild (1981) and Venables & James (1978). This connects the PDP approach to other forms of post-data inference, and in some sense, 'unifies' them. In particular, the smooth pooling principle applies to all three methods above.

(ii) PDP-unbiased point estimators of τ_1 and τ_2, which can be deduced directly from the PDP distributions, can be independently derived by a Brewster-Zidek type of argument on decision theoretic grounds. These estimators dominate the usual unbiased estimators, and are also generalized Bayes with respect to the Jeffreys prior. Hence, there is a dual justification for PDP-unbiased point estimators; even if one is not entirely convinced by the PDP argument, the decision theory results demonstrate the viability of PDP-unbiased estimators in their own right.

(iii) Similar remarks apply to PDP interval estimators of τ_1 and τ_2 with length fixed to that of the usual interval: as a function of σ, the unconditional coverage of these PDP intervals is bounded below by the nominal coverage, are also derived by a Brewster-Zidek type of argument, and are generalized Bayes intervals with respect to the Jeffreys prior.

(iv) The PDP estimation process can be viewed as one which starts with the 'usual' estimators, having good frequentist properties, and modifies them through the PDP distributions so that they are guaranteed to behave well conditionally, and exhibit good frequentist behavior. In other words, the PDP approach takes a procedure with good initial precision and modifies it so that it has both good initial precision and good final precision. In this sense, PDP estimators satisfy the twin goals of frequentist and conditional acceptability espoused by Casella (1988, 1992) and Berger (1988).

(v) The PDP method also considered point estimators of the parametric functions of interest based on marginal conditional likelihoods associated with the families of U_1^* and U_2^* distributions. The resulting PDP-MLEs can be interpreted as 'smooth versions' of the REML estimators, and can also be compared favorably to ANOVA estimators. In a sense, the PDP-MLEs are to REML estimators what the PDP-unbiased estimators are to the best unbiased estimators.

(vi) The plug-in principle is liberally applied in the conventional approach to point estimation. Both types of PDP estimators can be used as plug-ins to estimate $J\sigma_\alpha^2$, $J\gamma$ and ρ with appropriate post-data performance, and pre-data performance at least comparable to that of REML estimators.

(vii) REML estimators are not admissible because their multiplier functions are not differentiable for all $s > 0$. In contrast, PDP point estimators do have differentiable multiplier functions for all $s > 0$, which are a direct result of the 'smooth pooling' property of PDP distributions. In combination with the frequentist properties of PDP-unbiased estimators cited in point (2) above, it is reasonable to suggest that PDP-unbiased estimators have the *potential* to be admissible, at least in the class of scale-equivariant estimators based on (T_1, T_2). It is unclear at this juncture whether PDP-MLEs are potentially admissible for τ_1 and τ_2, but if so, that would be an added bonus.

The connections between PDP, Bayesian and fiducial inference are rather interesting. In the one-way random model, we explicitly assumed that $\sigma_e^2 > 0$ and $\sigma_\alpha^2 \geq 0$, which induced a constraint in the joint (τ_1, τ_2) parameter space. We then used the data, through the observed sum of squares ratio s, to limit the reference set for conditional inference to the region $\{U_1 > 0, U_2 \leq sU_1\}$ in (U_1, U_2) pivotal space. The conditioning event was then applied to each of the basic pivotals U_1, U_2 and W, which modified their distributions post-data. For inference about σ_e^2 and σ_α^2 (and selected functions thereof), the PDP distributions were projected onto the sample space for conditional sampling theory to apply, where the distributions are those associated with T_i and S, $i = 1, 2$.

A fiducialist armed with this reference set would invert the PDPs and project them onto the parameter space, where they would be distributions of τ_1, τ_2 and σ, or other relevant parametric functions. The distributions described by Wild (1981), for example, would apply in this case, and inference would follow from the fiducial argument. In contrast, Bayesians would work solely in the context of the constrained parameter space and bring in the data through the likelihood. Marginal posteriors would be obtained by integration over the constrained parameter space.

One connection between PDP distribution theory and Box & Tiao's Bayesian posteriors is that the pivotal space conditioning event $U_2 \leq sU_1$ is complementary to the parameter space constraint $\sigma \geq 1$. The domain of the fiducial distribution for σ would also be constrained to the interval $\sigma \geq 1$. Connections between PDP and fiducial inference are manifested by their mutual reliance on the pivotal space as the foundation for inference. Another connection is through the 'smooth pooling' between the distributions of U_1^* and U_2^* on the pivotal space *vis à vis* the Bayesian and fiducial complements of these distributions.

The connections between Bayesian and PDP inference go even further. The PDP-unbiased estimators in the one-way random model are generalized Bayes by virtue of the fact that they minimize the posterior risk with respect to Jeffreys' prior under entropy loss. Kubokawa, Saleh and Konno (1998) obtained this same result from a different angle, but the perspective gained from the PDP approach illuminates the Bayesian connection with frequentist risk properties.

It is interesting to note that the prior obtained by Brewster and Zidek (1974) to establish their estimator as generalized Bayes in the normal variance problem is a scale mixture of normal densities. This suggests that the fixed effects problem, from a Bayesian perspective is, in a sense, 'embedded' in a random effects problem, which may provide some insight to aid in the development of a model-based justification for 'improved' estimation procedures in fixed effects models. The PDP approach, as described in this thesis, cannot apply in fixed effects models; in particular, the distribution of T_2/τ_1 in a one-way fixed effects model is noncentral χ^2, so it is not a pivotal quantity. However, as far as inference concerning the error variance σ^2 is concerned, it is interesting to note that the 'answers' coincide in the balanced one-way fixed and random effects models. That is, the improved inference procedures for σ^2 are based on the same conditioning event, although the rationale for this event cannot be obtained by the PDP argument in the fixed effects problem. This is discussed in somewhat more detail in the next section.

The temptation is to judge the PDP approach on the basis of the decision theory results. These are important, and put the exclamation point on the approach, but it is equally important to recognize that the motivation for PDP estimators can be traced back to the model assumptions. We have developed an argument for the conditioning event that is rooted in the model, independent of the decision theoretic argument leading to Brewster-Zidek like estimators conditioned on the same event. The distribution theory based on PDP conditioning (discussed in point (1) above), in combination with the preceding statement, connects the Jeffreys prior to the generalized Bayes property of PDP estimators under any bowl-shaped loss function. Thus, we have established a web of connections that reveals the origins of properties associated with Brewster-Zidek type and generalized Bayes estimators in the balanced one-way random model.

4.4 Extensions to other models

Balanced one-way fixed effects model. The normal variance problem is a special case of the one-way fixed effects model. We believe the PDP argument has something useful to offer in this problem, although some modification is required. In this case, the parameters of interest are the treatment means $\mu_1, \ldots, \mu_I$ and the variance σ^2; the joint sufficient statistic is $(\overline{Y}_{1\cdot}, \ldots, \overline{Y}_{I\cdot}, T_1)$, where T_1 is the error SS.

The ANOVA table associated with the fixed effects version of model (2.1) is the same except for the expected mean square of treatments. The distribution theory of the one-way fixed effects model is as follows. The pivotal U_1 associated with error is unchanged; i. e., $U_1 \sim \chi^2_{\nu_1}$. For treatments, however, we have $U_2 = T_2/\tau_1$, which follows a $\chi^2_{\nu_2}(\lambda_1)$ distribution, where

$$\lambda_1 = \frac{J\sum(\mu_i - \mu)^2}{I-1}$$

is the noncentrality parameter; hence, U_2 is not a pivotal, so the PDP method as described in Chapter 2 cannot apply. Moreover, the expected mean square for treatments is $\tau_2 = \sigma_e^2 + \lambda_1$.

We could also consider the sum of squares due to the mean, $N\overline{Y}_{..}^2$. The distribution of $N\overline{Y}_{..}^2/\tau_1$ is $\chi^2_1(\lambda_2)$, where $\lambda_2 = \mu^2$ is its noncentrality parameter. From the work on the normal variance problem, we should be using this information, since $\overline{Y}_{..}$ is a function of the joint sufficient statistic. However, it is more convenient to ignore this information for reasons similar to those given for the one-way random model.

The random effects and fixed effects problems are similar in the following respects:

- there exists an ordering on the expected mean squares;
- the conditioning event is the same;
- the conditioned distributions have MLR properties.

The ordering on the expected mean squares is due to the presence of the noncentrality parameter in U_2; i. e., $\tau_2 = \sigma_e^2 + \lambda_1 \geq \sigma_e^2 = \tau_1$.

The conditioning event is motivated as follows. The usual F ratio is the test statistic for $H_0 : \mu_1 = \cdots = \mu_I$. If the observed F ratio is small, then H_0 is not rejected, which implies that the observed mean squares m_1 and m_2 yield two independent estimates of τ_1, just as in the random effects model. Thus, as f gets smaller, there should be some smooth tradeoff in information from m_2 towards estimation of τ_1. However, in the fixed effects model, there are two elements at play: estimation of the means and estimation of the error variance. A 'smooth pooling' approach to inference would suggest that there are two complementary actions taking place as f decreases: (i) an improved variance estimate, and (ii) shrinkage of the sample means towards $\overline{Y}_{..}$, analogous to a modified James-Stein type of estimator. Thus, conditioning on $F \leq f$, which is equivalent to the PDP conditioning event in the corresponding random model, is a reasonable thing to do in the fixed effects model as well.

By performing 'PDP-type conditioning', the resulting families of conditional distributions have MLR properties similar to those described in Theorems A.14 and A.15, where the noncentrality parameter λ_1 is substituted for σ. One then performs inference with respect to the conditional distributions, modifying the usual inference rules by applying them to these distributions. One outcome is that the resulting estimator of τ_1 is the Brewster-Zidek estimator.

The above comments indicate that the philosophy behind PDP inference is not necessarily limited to PDP inference *per se.* Although the structure of the problem is rather different, there are elements that can be drawn from the PDP argument to improve estimation in fixed and mixed effects models.

Balanced nested random model. The major purpose of Chapter 3 was to show that PDP methodology could be extended beyond the one-way random model. Effort was directed towards derivation of the distribution theory and construction of point estimators. Interval estimation and evaluation of frequentist properties has not been addressed in this work. Of the results from that chapter, the following appear to be most salient:

1. The distribution theory no longer enjoys monotone likelihood ratio properties, which implies that PDP estimators can no longer be of the Brewster-Zidek type. This is related to the presence of two or more nuisance parameters for inference concerning a parameter of interest; for example, an inference about τ_1 entails the nuisance parameters τ_2 and τ_3 (or equivalently, σ_1 and σ_2). On the other hand, PDP estimators continue to assume values in the parameter space and the multiplier functions of PDP estimators continue to exhibit the smooth pooling principle. Moreover, connections to Bayesian posteriors still exist. Hence, PDP estimators should still possess 'good' frequentist properties and have a conditional performance guarantee.

2. The parametric functions of interest are more complex in form relative to the one-way random model, so reliance on the plug-in principle takes on even greater importance. We know that plug-ins of PDP point estimators into parametric functions of interest maintain appropriate conditional performance, but their

frequentist performance is an open question. We would claim that the risk properties of PDP estimators are at least as good as those of frequentist estimators, as in the one-way random model. However, the risk functions are dependent on σ_1 and σ_2 in the three-stage random model, and will depend on more parametric functions as the dimensionality of the model increases. Hence, the meaning of terms such as 'at least as good as' or 'dominates' becomes cloudier in more complex problems, and is far more difficult to establish as the dimension of the problem increases. In the absence of MLR properties, the task becomes nearly impossible theoretically. As a result, the properties of PDP estimators developed in Chapter 2 provide a context for their use in more complex models.

3. We were able to show that the total degrees of freedom is preserved by PDP conditioning for any classification model. One of the immediate consequences of this result is that the mean of a PDP distribution is a multiple of some ratio of Dirichlet cdf's under a normal-theory classification model. Hence, the construction of PDP-unbiased point estimators is not a particularly difficult task in any given model; computation, on the other hand, may be. In the special case of the one-way model, the Dirichlet cdf's reduce to Beta cdf's.

The balanced nested model reveals some the opportunities and challenges that are likely to emerge in multifactor random effects models.

Balanced two-way random model. The various manifestations of classification models in two-factor problems will be important to consider in the development of PDP-based inference. The most critical of these is the balanced two-way random effects model, with or without interaction. An important application of this model is the gauge repeatability and reproducibility problem from manufacturing. It is not at all uncommon for one or two of the estimated variance components to be negative, which has a deleterious impact on estimates of parametric functions such as reproducibility. The plug-in principle is likely to play a prominent role in point estimation; furthermore, nearly all types of interval estimators employed in this problem are subject to poor conditional performance, so the PDP approach should have something of value to offer in this problem.

Satterthwaite approximation to one-way random model. Box and Tiao (1973, ch. 5) discussed at length certain approximations to the posterior densities of τ_1 and τ_2 based on scaled χ^2 distributions of the form $p\chi^2_q$, where the scale factor p and approximate degrees of freedom q are derived as functions of the first two moments of the posterior distribution. We have applied a similar type of approach by performing inference using scaled χ^2 densities in place of PDP densities for estimation of τ_1 and τ_2. The resulting point and interval estimators closely approximate the exact PDP procedures.

An analysis of this type is actually a Satterthwaite approximation, since Satterthwaite (1946) applied the same idea of approximating an exact (but messy) distribution with a density whose first two moments were identical. One would then perform inference with respect to the approximate distributions. This method was originally designed to apply to unbalanced ANOVA models, but it has been used to advantage in many types of problems since then. This type of approximation to PDP inference has several interesting features; for example, the MLE of τ_i in the one-way random model by this approximate method turns out to be PDP-unbiased, which is obvious in retrospect. However, this finding encourages use of PDP-unbiased estimators of τ_1 and τ_2 as plug-ins for other parametric functions of interest, which is not so obvious in the original PDP problem.

4.5 Limitations of PDP inference

Considerable space has been devoted to the advantages of PDP methodology, but there are some limitations that need to be mentioned:

1. In random effects models, the PDP method entails an assumption of nonnegative variance components beyond the positive error variance σ_e^2. It also requires continuous parameter and pivotal spaces, along with balanced data structures. These preconditions limit the range of problems to which PDP inference applies.

2. The parameter space must be constrained in order to apply PDP inference.

3. Computation is a concern, especially beyond the one-way random model. In this event, PDP point estimators are proportional to a ratio of Dirichlet cdf's, which no current statistical package or programming environment contains as part of its base distribution. This suggests that perhaps simulation based approaches such as (Markov Chain) Monte Carlo or importance sampling could be of assistance in making the techniques more widely accessible in practice.

4. As the model increases in complexity, so too does the number and complexity of PDP distributions. Unfortunately, no general theory of PDP inference in random effects models exists at present; for now, the theory must be developed on a model by model basis.

5. Unconditional estimators of variance components are known for their lack of robustness when the normality assumption is violated. This problem is likely to extend to PDP estimators, but one should investigate this by simulation.

4.6 Future directions

The principles of post-data pivotal inference should be adaptable to other types of problems involving constrained parameter spaces. The exact details for adapting the PDP approach to fixed or mixed effects models are unclear at this point, but enough elements are in place to suggest that a solution is tenable. In this section, some ideas for future research are outlined; some of the unfinished work in the balanced one-way and nested random models is listed below, along with some other topics of interest.

One-way random model. In Chapter 2, several topics were either left incomplete or were not addressed in the thesis. These include the following:

1. Comparative risk performance of point estimators of $J\gamma$ was not addressed in this work. There are two primary issues that pertain to this problem:

   (a) choice of a suitable loss function;

   (b) whether to use a univariate or bivariate approach to the decision problem.

   The form of the univariate entropy loss function prevents evaluation of estimators of $J\gamma$; however, it may be possible to handle matters by evaluating joint estimators with respect to composite entropy loss, *a la* Kubokawa, et al. (1998).

2. We have used entropy loss to search for improvements to unbiased point estimators in the one-way random model. This created a problem when evaluating estimators for σ, since the REML estimator (which we would like to have compared) is biased for σ. This led us to the realization that perhaps we should have been looking for improvements to ANOVA estimators instead, since the plug-in principle has been applied with impunity to such estimators. This suggests that one should look for a loss function which yields the ANOVA estimator as the best $\mathscr{G}$-equivariant estimator, and then go through the SBBZ process as usual.

This is not likely to be easy, as it should be able to handle differences and ratios of estimators seamlessly.

3. Point estimation of the intraclass correlation ρ was not addressed in this thesis. Plug-in estimators, whether based on ANOVA, REML, PDP-ML or PDP-unbiased estimators, turn out to be biased for ρ, so the utility of entropy loss as a quality criterion is in question. In this problem, normalized quadratic loss may be a better choice for evaluating estimators.

4. PDP-modified versions of Tukey-Williams intervals of $J\sigma_\alpha^2$ were derived in section 2.6.6.2, but their frequentist coverage probabilities are unknown. Simulation studies of their properties need to be conducted to assess their frequentist performance. Average length is also an issue.

5. Admissibility of PDP-unbiased point estimators and fixed length confidence intervals has not been considered. Since these estimators have smooth multiplier functions and are generalized Bayes with respect to the Jeffreys prior, it is entirely possible that these estimators are admissible, at least in the class of scale-equivariant estimators based on (T_1, T_2).

6. The important problem of predicting random effects has not been dealt with in this work. This problem is complicated because it involves several plug-in estimators, including that of μ, which has also been ignored in this thesis. Comparisons to BLUPs and related quantities are also required.

7. The interpretation of residuals in PDP inference has not been undertaken here. The interpretation and diagnostic value of PDP-modified residuals is especially important in underspecified models, since in an overspecified model, the method is 'self-correcting'; e. g., if a term has no effect, the PDP method will adjust by shifting the information towards error as a function of the data.

Three-stage nested model. The primary goal of Chapter 3 was to establish that the PDP approach could be usefully extended beyond the one-way random model. We believe this goal has been partially achieved, but much work was left undone in the process. Some of the open problems include the following:

1. Evaluation of risk properties of Stein-like and PDP estimators of the τ's and variance components under (composite) entropy loss needs to be done. This should not be particularly difficult, as it is a simple extension of the work in the one-way random model; for evaluation of joint estimators, the method of Kubokawa, *et al.* (1998) can be adapted. However, the risk function is now a function of both σ_1 and σ_2. Establishing conditions for the existence of generalized Bayes, empirical Bayes and minimax estimators in the class of estimators of the form $\phi_i(S_1, S_2)T_i$, $i = 1, 2, 3$, is also an interesting challenge.

2. Derivation of PDP-MLEs of the τ_i is needed; this too should not be difficult.

3. PDP-CIs of the τ_i and σ_i with length fixed to that of the corresponding frequentist intervals need to be constructed, and their unconditional coverage probabilities need to be derived.

4. Unconditional coverage probabilities and average lengths of PDP-CIs with fixed conditional coverage still need to be done.

5. Predicted random effects and residuals also need to be dealt with in this problem.

Extensions. One of the major hurdles in PDP inference at this point is the need to reinvent the wheel for every new model. A general approach to PDP modeling is necessary in the long term, but this is probably a ways off until some general patterns emerge across different models.

To some degree, the fixed effects problem can be embedded in a random effects problem, analogous to a hierarchical Bayes model. This may be an interesting angle to pursue in the analysis of fixed and mixed effects models.

4.7 Foundational issues

The broad scope of this thesis engenders a variety of foundational issues. Some that have emerged in the course of this work include the following:

1. **Conditioning on a pivotal space.** As far as we know, the novelty of PDP inference lies in conditioning on a pivotal space to incorporate both parameter constraints and the data. In the limited class of problems considered, the approach yields sensible inferences that have both pre-data and post-data appeal. We have also been able to connect this type of inference with other modes of post-data inference. An important foundational question is: under what conditions can post-data modes of inference be unified, and how does this unification take place?

2. **Initial and final precision.** Traditionally, the issue has been initial precision *versus* final precision, or equivalently, pre-data vs. post-data inference. The PDP method has shown that it is possible to begin with procedures having good initial precision, and through suitable conditioning, modify the procedures (as a function of the data) so that the initial precision is at least as good and the final precision is guaranteed. The question is then: to what extent can this type of modification be applied, and under what conditions is it valid?

3. **Smooth pooling principle.** One of the linchpins of PDP inference is manifestation of the smooth pooling principle via a family of conditional distributions, indexed by one or more functions of the data. Unification of the Bayesian,

fiducial and PDP distributions is, in no small part, due to the fact that all three share this feature. It also underlies the continuous multiplier functions of PDP estimators. The role of smooth pooling in post-data inference needs to be investigated more deeply.

4. **Marginalization paradoxes.** One of the interesting features of PDP distributions in random effects models is that the effect of conditioning on a subregion of a pivotal space is felt in the marginal PDP distributions, but not the joint PDP distributions. PDP-MLEs have some appeal because of the 'shifting' in the families of marginal PDP distributions indexed by one or more observed sum of squares ratios. How is this property of PDP distributions related to the marginalization paradoxes described in Dawid, Stone and Zidek (1973)?

5. **Loss functions and marginal vs. joint estimation.** We have tried, as far as possible, to adhere to a single univariate loss function to evaluate various point estimators under a given model. However, this doesn't always work; for example, point estimators of $J\sigma_\alpha^2$ are differences of estimators of τ_1 and τ_2. Thus, one eventually needs to consider some type of composite loss function to evaluate joint estimators. Part of the question is: which one? It would be convenient if the selected loss function could handle both univariate and joint estimators.

6. **Order-restricted inference.** PDP methodology can be interpreted as conditional order-restricted inference. The traditional methodology in order-restricted problems is isotonic inference, which has not been mentioned in this thesis until now, and has had no direct influence on this work. However, at some point it will become necessary to investigate the similarities and differences between these methods.

Appendix A

Complements: 1-way random model

This appendix provides background information, preliminary lemmas and results associated with material presented in Chapter 2. The following topics are addressed with respect to PDP distribution theory:

- moment generating functions;
- the $\delta(\cdot)$ and $\psi(\cdot)$ functions, and their properties;
- means and variances;
- monotone likelihood ratio properties.

Additional derivations associated with limiting values of PDP estimators and distributional properties of V^* (the confidence distribution for inference about $J\sigma_\alpha^2$), are also presented in this appendix.

A.1 Distribution theory

Let $U_1^*(s)$ be a random variable whose density function has the form

$$f(u_1; \nu_2, \nu_1, s) = \frac{g_{\nu_1}(u_1)\, G_{\nu_2}(su_1)}{I_z(\phi_2, \phi_1)} \tag{A.1}$$

for any fixed $s > 0$, where

- $g_{\nu_1}(u_1)$ is a $\chi^2_{\nu_1}$ density;
- $G_{\nu_2}(su_1)$ is a 'shifted' $\chi^2_{\nu_2}$ cdf;
- $I_z(\phi_2, \phi_1)$ is a Beta cdf evaluated at $z = s/(1+s)$, with parameters $\phi_i = \nu_i/2$, $i = 1, 2$; and
- s and u_1 are both defined on the positive real line.

Correspondingly, let $U_2(s)$ denote a random variable with density

$$h(u_2; \nu_2, \nu_1, s) = \frac{g_{\nu_2}(u_2)\,[1 - G_{\nu_1}(u_2/s)]}{I_z(\phi_2, \phi_1)}, \tag{A.2}$$

where $g(\cdot)$, $G(\cdot)$ and $I(\cdot)$ are defined as above, with $u_2 > 0$.

A.1.1 Facts concerning Beta distributions

Certain results that pertain to Beta distributions are relevant to discussions involving PDP-MLEs of τ_1 and τ_2 and the moments of $U_1^*(s)$ and $U_2^*(s)$. The following facts are taken from Abramowitz and Stegun (1964).

Fact A.1. Let $X \sim \chi^2_\nu$ with cdf $G_\nu(x)$, $Q_\nu(x) = 1 - G_\nu(x)$, and $\phi = \nu/2$. Then,

$$Q_{\nu+2}(x) = Q_\nu(x) + \frac{(x/2)^\phi e^{-x/2}}{\Gamma(\phi+1)}.$$

Fact A.2. Let $X \sim \text{Beta}(a, b)$ with cdf $I_x(a, b)$, $x \in [0, 1]$. Then,

$$I_x(a, b) = 1 - I_{1-x}(b, a).$$

Fact A.3.

$$(a + b)\, I_x(a, b) = a\, I_x(a + 1, b) + b\, I_x(a, b + 1).$$

Fact A.4.

$$x\,I_x(a,b) = I_x(a+1,b) - (1-x)\,I_x(a+1,b-1).$$

Fact A.5.

$$I_x(a,b) = \frac{\Gamma(a+b)}{\Gamma(a+1)\,\Gamma(b)}\,x^a\,(1-x)^{b-1} + I_x(a+1,b-1).$$

Fact A.6.

$$I_x(a,b) = \frac{\Gamma(a+b)}{\Gamma(a+1)\,\Gamma(b)}\,x^a\,(1-x)^{b} + I_x(a+1,b).$$

Fact A.7.

$$I_x(a,b) = x\,I_x(a-1,b) + (1-x)\,I_x(a,b-1).$$

Lemma A.1. *Under the conditions of Fact A.1,*

$$G_\nu(x) - G_{\nu+2}(x) = 2\,g_{\nu+2}(x). \tag{A.3}$$

Proof. Let $Q_\nu(x) = 1 - G_\nu(x)$, where $G_\nu(x)$ is the cdf of X evaluated at x. A rearrangement of Fact 1 yields

$$Q_{\nu+2}(x) - Q_\nu(x) = \frac{x^\phi\,e^{-x/2}}{2^\phi\,\Gamma(\phi+1)};$$

since

$$Q_{\nu+2}(x) - Q_\nu(x) = (1 - G_{\nu+2}(x)) - (1 - G_\nu(x)) = G_\nu(x) - G_{\nu+2}(x)$$

and

$$\frac{x^\phi\,e^{-x/2}}{2^\phi\,\Gamma(\phi+1)} = 2\,\frac{x^{(\phi+1)-1}\,e^{-x/2}}{2^{\phi+1}\Gamma(\phi+1)} = 2\,g_{\nu+2}(x),$$

the result follows. □

Corollary A.1.1.

$$\frac{g_{\nu+2}(x)}{G_\nu(x)} = \frac{1}{2}\,\big(1 - \frac{G_{\nu+2}(x)}{G_\nu(x)}\big). \tag{A.4}$$

Proof. Divide both sides of (A.3) by $2G_\nu(x)$. □

A.1.2 Results concerning $\delta(\cdot)$ and $\psi(\cdot)$

The functions $\delta_z(\phi_2, \phi_1)$ and $\psi_z(\phi_2, \phi_1)$ that appear in the moments of U_1^* and U_2^* in Chapter 2 are important quantities in the PDP approach. The function $\delta(\cdot)$ operationalizes the 'smooth pooling' of treatment and error degrees of freedom post-data, while $\psi(\cdot)$ appears in the variances of U_1^* and U_2^* as well as their covariance. In this subsection, we derive several results concerning the properties of these functions. The terms $z = s/(1+s)$, $\phi_1 = \nu_1/2$ and $\phi_2 = \nu_2/2$ appear regularly, where s is the realized sum of squares ratio in the one-way random model (2.1). We begin by defining $\delta(\cdot)$ and $\psi(\cdot)$:

Definition A.1. Given $\phi_1, \phi_2 > 0$ and $z \in [0, 1]$, define

$$\delta_z(\phi_2, \phi_1) = \frac{\Gamma(\phi_2 + \phi_1)}{\Gamma(\phi_2)\Gamma(\phi_1)} \frac{z^{\phi_2}(1-z)^{\phi_1}}{I_z(\phi_2, \phi_1)}. \tag{A.5}$$

Definition A.2. Given $z \in [0, 1]$ and $\phi_2, \phi_1 > 0$,

$$\psi_z(\phi_2, \phi_1) = \delta\left[(1-z)(\nu_2 - 2\delta) - z(\nu_1 + 2\delta)\right], \tag{A.6}$$

where $\delta = \delta_z(\phi_2, \phi_1)$.

The following pair of results are relevant in the determination of the mean of $U_1^*(s)$.

Theorem A.2. *If $x \in [0, 1]$ and $a, b > 0$,*

$$\frac{I_x(a, b+1)}{I_x(a, b)} = 1 + \frac{\delta_x(a, b)}{b}. \tag{A.7}$$

Proof. To simplify notation, let $\delta_x(a, b) = \delta$. We prove the result by appealing to the above identities from Abramowitz and Stegun (1964). From Fact A.3,

$$(a+b)\, I_x(a, b) = a\, I_x(a+1, b) + b\, I_x(a, b+1).$$

Rearranging terms,

$$a\left[I_x(a, b) - I_x(a+1, b)\right] + b\, I_x(a, b) = b\, I_x(a, b+1).$$

By Fact A.6,

$$a\left[\frac{\Gamma(a+b)}{\Gamma(a+1)\,\Gamma(b)}\,x^a\,(1-x)^b\right]+b\,I_x(a,b)=b\,I_x(a,b+1),$$

which implies

$$b\,I_x(a,b+1)=b\,I_x(a,b)+\frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)}\,x^a\,(1-x)^b.$$

Dividing both sides through by $b\,I_x(a,b)$, one obtains

$$\frac{I_x(a,b+1)}{I_x(a,b)}=1+\frac{\delta_x(a,b)}{b}$$

by Definition A.1, completing the proof. □

Corollary A.2.1. *Substituting z for x, ϕ_2 for a and ϕ_1 for b in Theorem A.2, where $\phi_i=\nu_i/2$, $i=1,2$, and multiplying both sides of* (A.7) *by ν_1 yields*

$$E[U_1^*(s)]=\nu_1\,\frac{I_z(\phi_2,\phi_1+1)}{I_z(\phi_2,\phi_1)}=\nu_1+2\delta_z(\phi_2,\phi_1). \tag{A.8}$$

The next pair of results correspond to $U_2^*(s)$, complementing the above pair.

Theorem A.3. *If $x\in[0,1]$ and $a,b>0$,*

$$\frac{I_x(a+1,b)}{I_x(a,b)}=1-\frac{\delta_x(a,b)}{a}. \tag{A.9}$$

Proof. By Fact A.4,

$$\begin{aligned}x\,I_x(a,b)&=I_x(a+1,b)-(1-x)\,I_x(a+1,b-1)\\&=I_x(a+1,b)-(1-x)\left[I_x(a,b)-\frac{\Gamma(a+b)}{\Gamma(a+1)\,\Gamma(b)}\,x^a\,(1-x)^{b-1}\right],\end{aligned}$$

by Fact A.5. Thus,

$$=I_x(a+1,b)-(1-x)\,I_x(a,b)+\frac{1}{a}\,\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\,x^a\,(1-x)^b.$$

Dividing both sides through by $I_x(a,b)$ along with a little rearrangement of terms yields

$$\frac{I_x(a+1,b)}{I_x(a,b)} = 1 - \frac{\delta_x(a,b)}{a},$$

which completes the proof. □

Corollary A.3.1. *If one substitutes z for x, ϕ_2 for a and ϕ_1 for b in Theorem A.3 and multiplies* (A.9) *through by ν_2, then*

$$E[U_2^*(s)] = \nu_2 \frac{I_z(\phi_2+1,\phi_1)}{I_z(\phi_2,\phi_1)} = \nu_2 - 2\delta_z(\phi_2,\phi_1). \tag{A.10}$$

An important result concerning δ is given below; it describes the asymptotic behavior of $\delta(\cdot)$ in the balanced one-way random model.

Theorem A.4. *The limiting behavior of $\delta(\cdot)$ is given as follows:*

(a) $\lim_{x\to 1} \delta_x(a,b) = 0$;

(b) $\lim_{x\to 0} \delta_x(a,b) = a$.

Proof. Starting from

$$\delta_x(a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{x^a\,(1-x)^b}{I_x(a,b)},$$

it suffices to observe that the numerator tends to zero and the denominator to one as $x \to 1$ in order to prove (a).

To prove (b), both the numerator and denominator of δ tend to zero as $x \to 0$. Appealing to L'Hôpital's rule,

$$\begin{aligned}
\lim_{x\to 0} \delta_x(a,b) &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \lim_{x\to 0} \frac{\frac{d}{dx}\, x^a\,(1-x)^b}{\frac{d}{dx}\, I_x(a,b)} \\
&= \lim_{x\to 0} \frac{ax^{a-1}(1-x)^b + bx^a(1-x)^{b-1}}{x^{a-1}(1-x)^{b-1}} \\
&= \lim_{x\to 0} \frac{x^{a-1}(1-x)^{b-1}\,\{a(1-x)+bx\}}{x^{a-1}(1-x)^{b-1}} \\
&= \lim_{x\to 0}\, [a + x(b-a)] \\
&= a.
\end{aligned}$$

□

The following result is relevant to certain point estimation results in Chapter 2.

Lemma A.5.

$$\frac{z\, I_z(\phi_2,\phi_1)}{I_z(\phi_2+1,\phi_1)} = 1 - \frac{(1-z)(\nu_1-2)}{(\nu_1-2) + 2\delta_z(\phi_2+1,\phi_1-1)}. \tag{A.11}$$

Proof. By Fact A.4, we have

$$zI_z(\phi_2,\phi_1) = I_z(\phi_2+1,\phi_1) - (1-z)\, I_z(\phi_2+1,\phi_1-1),$$

so

$$\frac{z\, I_z(\phi_2,\phi_1)}{I_z(\phi_2+1,\phi_1)} = 1 - \frac{(1-z)I_z(\phi_2+1,\phi_1-1)}{I_z(\phi_2+1,\phi_1)}.$$

By Theorem A.2, one can deduce that

$$\frac{I_z(\phi_2+1,\phi_1)}{I_z(\phi_2+1,\phi_1-1)} = 1 + \frac{\delta_z(\phi_2+1,\phi_1-1)}{\phi_1-1},$$

so that

$$\begin{aligned}
\frac{z\, I_z(\phi_2,\phi_1)}{I_z(\phi_2+1,\phi_1)} &= 1 - \frac{(1-z)(\phi_1-1)}{(\phi_1-1) + \delta_z(\phi_2+1,\phi_1-1)} \\
&= 1 - \frac{(1-z)(\nu_1-2)}{(\nu_1-2) + 2\delta_z(\phi_2+1,\phi_1-1)}
\end{aligned}$$

□

In the next few results, it is convenient to express the conditional means (A.8) and (A.10) as functions of z for fixed arguments ϕ_2 and ϕ_1:

Definition A.3. Given $\phi_1, \phi_2 > 0$ and $z \in [0, 1]$, define

$$J_z(\phi_2, \phi_1) = \nu_1 \frac{I_z(\phi_2, \phi_1 + 1)}{I_z(\phi_2, \phi_1)} = \nu_1 + 2\delta_z(\phi_2, \phi_1)$$
$$H_z(\phi_2, \phi_1) = \nu_2 \frac{I_z(\phi_2 + 1, \phi_1)}{I_z(\phi_2, \phi_1)} = \nu_2 - 2\delta_z(\phi_2, \phi_1).$$

Then, for each fixed $z \in [0, 1]$, it follows that

$$H_z(\phi_2, \phi_1) + J_z(\phi_2, \phi_1) = \nu_1 + \nu_2; \tag{A.12}$$

i. e., the total degrees of freedom are preserved by conditioning.

Corollary A.5.1. *If $H(\cdot)$ and $J(\cdot)$ are given as in Definition A.3, then*

$$\frac{d}{dz}[H_z(\phi_2, \phi_1) + J_z(\phi_2, \phi_1)] = 0. \tag{A.13}$$

Proof. Since ν_1 and ν_2 are constant, taking the derivative of (A.12) with respect to z leads immediately to the result. □

An interesting consequence of the above corollary is that the rates of change of $J(\cdot)$ and $H(\cdot)$ have the same magnitude but are opposite in direction. The following lemma shows how the function $\psi(\cdot)$ arises in this problem.

Lemma A.6. *Let $\delta = \delta_z(\phi_2, \phi_1)$. If $J(\cdot)$ is given as in Definition A.3, then*

$$\frac{d}{dz} J_z(\phi_2, \phi_1) = \frac{\delta}{z(1-z)} [(1-z)(\nu_2 - 2\delta) - z(\nu_1 + 2\delta)] = \frac{\psi_z(\phi_2, \phi_1)}{z(1-z)}.$$

Proof. Since

$$J_z(\phi_2, \phi_1) = \nu_1 \frac{I_z(\phi_2, \phi_1 + 1)}{I_z(\phi_2, \phi_1)},$$

the derivative of J with respect to z is

$$\begin{aligned}
\frac{d\,J_z(\phi_2,\phi_1)}{dz} &= \nu_1 \frac{I_z(\phi_2,\phi_1)\,\frac{d}{dz}I_z(\phi_2,\phi_1+1) - I_z(\phi_2,\phi_1+1)\,\frac{d}{dz}I_z(\phi_2,\phi_1)}{\left[I_z(\phi_2,\phi_1)\right]^2} \\
&= \nu_1 \frac{\frac{d}{dz}I_z(\phi_2,\phi_1+1)}{I_z(\phi_2,\phi_1)} - J_z(\phi_2,\phi_1)\,\frac{\frac{d}{dz}I_z(\phi_2,\phi_1)}{I_z(\phi_2,\phi_1)} \\
&= \nu_1 \frac{\Gamma(\phi_2+\phi_1+1)}{\Gamma(\phi_2)\Gamma(\phi_1+1)}\,\frac{z^{\phi_2-1}(1-z)^{\phi_1}}{I_z(\phi_2,\phi_1)} \\
&\quad - \frac{\Gamma(\phi_2+\phi_1)}{\Gamma(\phi_2)\Gamma(\phi_1)}\,\frac{z^{\phi_2-1}(1-z)^{\phi_1-1}}{I_z(\phi_2,\phi_1)}\,J_z(\phi_2,\phi_1) \\
&= \nu_1 \frac{\phi_1+\phi_2}{\phi_1 z}\,\delta_z(\phi_2,\phi_1) - \frac{J_z(\phi_2,\phi_1)}{z(1-z)}\,\delta_z(\phi_2,\phi_1) \\
&= \frac{2(\phi_1+\phi_2)}{z}\,\delta_z(\phi_2,\phi_1) - \frac{J_z(\phi_2,\phi_1)}{z(1-z)}\,\delta_z(\phi_2,\phi_1).
\end{aligned}$$

Letting $\delta = \delta_z(\phi_2,\phi_1)$,

$$\begin{aligned}
&= \frac{\delta}{z(1-z)}\,\left[2(\phi_1+\phi_2)(1-z) - J_z(\phi_2,\phi_1)\right] \\
&= \frac{\delta}{z(1-z)}\,\left[(\nu_1+\nu_2)(1-z) - J_z(\phi_2,\phi_1)\right].
\end{aligned}$$

By (A.12),

$$\begin{aligned}
&= \frac{\delta}{z(1-z)}\,\left[(1-z)(H_z(\phi_2,\phi_1)+J_z(\phi_2,\phi_1)) - J_z(\phi_2,\phi_1)\right] \\
&= \frac{\delta}{z(1-z)}\,\left[(1-z)H_z(\phi_2,\phi_1) - zJ_z(\phi_2,\phi_1)\right] \\
&= \frac{\delta}{z(1-z)}\,\left[(1-z)(\nu_2-2\delta) - z\,(\nu_1+2\delta)\right].
\end{aligned}$$

Applying the definition of $\psi(\cdot)$ in the above expression yields the result. □

Corollary A.6.1. *Given $\delta_z(\phi_2,\phi_1)$ defined for $z \in [0,1]$,*

$$\frac{d\,\delta}{dz} = \frac{\psi_z(\phi_2,\phi_1)}{2\,z\,(1-z)}. \tag{A.14}$$

Proof.

$$\frac{d\,J}{dz} = \frac{d}{dz}(\nu_1 + 2\delta_z(\phi_2,\phi_1)) = 2\frac{d\,\delta}{dz}.$$

The result now follows from Theorem A.6. □

The following lemma establishes that $\psi(\cdot)$ is uniformly nonpositive in z.

Lemma A.7. *If $\phi_1, \phi_2 > 0$ and $z \in [0,1]$, then $\psi_z(\phi_2, \phi_1) \leq 0$ uniformly in z.*

Proof. By Corollary 2.5.1, $\delta_z(\phi_2, \phi_1)$ is monotone decreasing in $z \in [0,1]$; therefore, $d\delta/dz \leq 0$ uniformly in z. But this implies $\psi_z(\phi_2, \phi_1) \leq 0$ by Corollary A.6.1. □

The following result involves a limit that arises in point estimation of the expected mean square ratio σ.

Lemma A.8. *For $z \in [0,1]$,*

$$\lim_{z \to 0} \frac{\nu_1 + 2\delta_z(\phi_2, \phi_1)}{\nu_2 - 2\delta_z(\phi_2, \phi_1)} \frac{z}{1-z} = 1 + \frac{2}{\nu_2}. \tag{A.15}$$

Proof. By Definition A.3, the left hand side of (A.15) can be written as

$$\frac{\nu_1 + 2\delta_z(\phi_2, \phi_1)}{\nu_2 - 2\delta_z(\phi_2, \phi_1)} \frac{z}{1-z} = \frac{\nu_1}{\nu_2} \frac{z\, I_z(\phi_2, \phi_1 + 1)}{(1-z)\, I_z(\phi_2 + 1, \phi_1)}.$$

By Fact A.4,

$$= \frac{\nu_1}{\nu_2} \frac{I_z(\phi_2 + 1, \phi_1 + 1) - (1-z)\, I_z(\phi_2 + 1, \phi_1)}{(1-z)\, I_z(\phi_2 + 1, \phi_1)}$$

$$= \frac{\nu_1}{\nu_2} \left[\frac{I_z(\phi_2 + 1, \phi_1 + 1)}{(1-z)\, I_z(\phi_2 + 1, \phi_1)} - 1 \right]$$

By Theorem A.2,

$$= \frac{\nu_1}{\nu_2} \left[\frac{\phi_1 + \delta_z(\phi_2 + 1, \phi_1)}{\phi_1(1-z)} - 1 \right].$$

As $z \to 0$, it follows by Theorem A.4 that

$$\lim_{z \to 0} \frac{\nu_1 + 2\delta_z(\phi_2, \phi_1)}{\nu_2 - 2\delta_z(\phi_2, \phi_1)} s = \frac{\nu_1}{\nu_2} \left[\frac{\phi_1 + \phi_2 + 1}{\phi_1} - 1 \right]$$

$$= \frac{\nu_1}{\nu_2} \frac{\nu_2 + 2}{\nu_1}$$

$$= \frac{\nu_2 + 2}{\nu_2}.$$

□

A.1.3 Monotone likelihood ratio

The monotone likelihood ratio property (MLRP) is central to the theory of point and interval estimation in Chapter 2. The ability to order distributions (and hence their expectations) in r or in σ greatly simplifies the task of finding suitable estimators of the variance functions of interest. Furthermore, it allows us to use results from Brown (1968), Brewster and Zidek (1974), Shorrock (1990) and Kubokawa (1994), all of which are predicated on existence of the MLRP. In this subsection, definitions and results pertaining to monotone likelihood ratio are presented, after which derivations of the MLRP are developed for $U_1^*(s)$ and $U_2^*(s)$ in r and σ. We begin with some definitions.

Definition A.4. A family of pdf's/pmf's $\{f(t|\theta) : \theta \in \Theta \subset \Re\}$ associated with a univariate random variable T, indexed by a real-valued parameter θ, is said to have the increasing (decreasing) *monotone likelihood ratio property* (MLRP) in θ if, for every $\theta_2 > \theta_1$, the distribution functions $F(t|\theta_2)$ and $F(t|\theta_1)$ exist and are distinct, and the density ratio $f(t|\theta_2)/f(t|\theta_1)$ is a nondecreasing (nonincreasing) function of t on the set $\{t : f(t|\theta_1) > 0 \text{ or } f(t|\theta_2) > 0\}$. If $c > 0$, then $c/0 = \infty$ by definition.

Definition A.5. A cdf F_X is said to be *stochastically greater* than a cdf F_Y if $F_X(t) \leq F_Y(t)$ for all t, and $F_X(t) < F_Y(t)$ for some t. This is equivalent to saying that $\Pr(X \geq t) \geq \Pr(Y \geq t)$ for all t, with strict inequality for some t. In other words, X tends to have larger values than Y.

Definition A.6. A family of cdf's $\{F(x|\theta), \theta \in \Theta\}$ is said to be *stochastically increasing* in θ if $\theta_1 > \theta_2$ implies that the $F(x|\theta_i)$ are distinct, $i = 1, 2$, and $F(x|\theta_1)$ is stochastically greater than $F(x|\theta_2)$ for all x.

Definition A.7. A family of cdf's $\{F(x|\theta), \theta \in \Theta\}$ is said to be *stochastically decreasing* in θ if $\theta_1 > \theta_2$ implies that the $F(x|\theta_i)$ are distinct, $i = 1, 2$, and $F(x|\theta_2)$ is stochastically greater than $F(x|\theta_1)$ for all x.

Lemma A.9. *Let $\mathcal{X} = [0, \infty)$. If $X \sim F_X(x|\theta)$ and $F_X(x|\theta)$ is stochastically increasing (decreasing) in θ, then for $Y = 1/X$, $F_Y(y|\theta)$ is stochastically decreasing (increasing) in θ.*

Lemma A.10. *If $X \sim F_X(x|\theta)$, where $F_X(x|\theta)$ is stochastically increasing (decreasing) in θ and $\theta > 0$, then $F_X(x|\theta^{-1})$ is stochastically decreasing (increasing) in θ.*

The above pair of lemmas are stated as problems in Lehmann (1986). The next pair of results are also taken from Lehmann (pp. 84-85), where the proofs can be found.

Theorem A.11. *Let F_0 and F_1 denote two cdf's defined on $\Re$. Then $F_1(x) \leq F_0(x)$ for all $x \in \Re$ iff there exist two nondecreasing functions f_0 and f_1, and a random variable V, such that*

(a) $f_0(v) \leq f_1(v)$ for all v;

(b) the distributions of $f_0(V)$ and $f_1(V)$ are F_0 and F_1, respectively.

Theorem A.12. *Let $\{f(x|\theta) : \theta \in \Theta \subset \Re\}$ be a family of densities with increasing MLR in θ. Then*

(i) If $\psi(\cdot)$ is a nondecreasing function of x, then $E_\theta[\psi(x)]$ is a nondecreasing function of θ; if $X_1, \ldots, X_n$ are independently distributed random variables with density $f(x|\theta)$, and ψ' is a function of $\boldsymbol{x} = (x_1, \ldots, x_n)'$, which is nondecreasing in each of its arguments, then $E_\theta[\psi'(\boldsymbol{X})]$ is a nondecreasing function of θ.

(ii) For any $\theta_1 < \theta_2$, the cdfs of X under θ_1 and θ_2 satisfy $F_{\theta_2}(x) \leq F_{\theta_1}(x)$ for all x.

(iii) Let $\psi(\cdot)$ be a function with a single sign change. More specifically, suppose there is a value x_0 such that $\psi(x) \leq 0$ for all $x < x_0$, and $\psi(x) \geq 0$ for all $x \geq x_0$. Then, there exists a value $\theta_0 \in \Theta$ such that $E_\theta[\psi(x)] \leq 0$ for $\theta < \theta_0$ and

$E_\theta[\psi(x)] \geq 0$ for $\theta \geq \theta_0$, unless $E_\theta[\psi(x)]$ is either positive for all θ or negative for all θ.

(iv) Suppose that (a) $f(x|\theta)$ is positive for all θ and all x, (b) $f(x|\theta_2)/f(x|\theta_1)$ is strictly increasing in x for $\theta_2 > \theta_1$, and (c) $\psi(x)$ is defined as in (iii) and is nonzero with positive probability. If $E_{\theta_0}[\psi(x)] = 0$, then $E_\theta[\psi(x)] < 0$ for $\theta < \theta_0$ and $E_\theta[\psi(x)] > 0$ for $\theta > \theta_0$.

Part (ii) of this theorem states that any family of distributions with the increasing (decreasing) MLRP in x is stochastically increasing (decreasing) in x. The converse of this statement is not true: a counterexample is the Cauchy location-scale family, whose cdf's are stochastically increasing in the location parameter θ, but the likelihood ratio is not monotone.

Finally, the following theorem connects the MLRP with stochastic ordering, and relates these ideas to ordering of expected values.

Theorem A.13. *If $f(x)/g(x)$ is increasing in x, $x > 0$, then:*

(a) $F(x) \leq G(x)$ for all x;

(b) $E_F(X) \geq E_G(X)$;

(c) $\dfrac{E_F(X^2)}{E_F(X)} \geq \dfrac{E_G(X^2)}{E_G(X)}$.

Proof. Let $f(x)/g(x)$ be increasing in x for all $x > 0$, where f and g denote two densities (which may be from the same family). The proof of part (a) establishes that monotone likelihood ratio implies stochastic ordering. To that end, choose a point $x_0 \in \Re^+$, and construct the sets $A = \{x : f(x) < g(x)\}$ and $B = \{x : f(x) > g(x)\}$. Define the functions

$$\psi(x) = \begin{cases} 1 & \text{if } x > x_0; \\ 0 & \text{if } x \leq x_0, \end{cases}$$

and $h(x) = f(x)/g(x)$, and let

$$a = \sup_A \psi(x) \qquad \text{and} \qquad b = \inf_B \psi(x).$$

Finally, define $x^* = h^{-1}(1)$, where h^{-1} is the inverse function of h. We can then reexpress the sets A and B as $A = (0, x^*)$ and $B = (x^*, \infty)$. We then have three possible cases for determining a and b:

(i) if $x_0 < x^*$, then $x_0 \in A$, so that $a = 1$, and hence $b = 1$;

(ii) if $x_0 > x^*$, then $a = 0$; further, there is an $x \in B$ such that $\psi(x) = 0$, so that $b = 0$;

(iii) if $x_0 = x^*$, then $\psi(x) = 0$ on all of A and $\psi(x) = 1$ on all of B, which means $a = 0$ and $b = 1$.

In all cases, $b - a \geq 0$. Consider the integral

$$\begin{aligned} \int \psi(x)\,[f(x) - g(x)]\,dx &= \int_A \psi(x)\,[f(x) - g(x)]\,dx + \int_B \psi(x)\,[f(x) - g(x)]\,dx \\ &\geq a \int_A [f(x) - g(x)]\,dx + b \int_B [f(x) - g(x)]\,dx. \end{aligned}$$

In the former integral, $f(x) - g(x) < 0$ over A, while in the latter integral, b represents the infimum of $\psi(\cdot)$ over the set B. Now,

$$\int_A [f(x) - g(x)]\,dx = -\int_B [f(x) - g(x)]\,dx,$$

where the left hand integral is nonnegative. Then

$$\begin{aligned} \int \psi(x)\,[f(x) - g(x)]\,dx &= -a \int_B [f(x) - g(x)]\,dx + b \int_B [f(x) - g(x)]\,dx, \\ &= (b - a) \int_B [f(x) - g(x)]\,dx \\ &\geq 0. \end{aligned}$$

Having established that the integral is nonnegative, we have

$$\begin{aligned}
\int \psi(x)\,[f(x)-g(x)]\,dx &= \int I_{\{x>x_0\}}(x)\,[f(x)-g(x)]\,dx \\
&= \int_{x_0}^{\infty} [f(x)-g(x)]\,dx \\
&= (1-F(x_0)) - (1-G(x_0)) \\
&= G(x_0) - F(x_0) \\
&\geq 0.
\end{aligned}$$

Since x_0 was arbitrarily chosen on $(0,\infty)$, we have that

$$F(x) \leq G(x) \qquad \forall x > 0.$$

This establishes that $F(\cdot)$ is stochastically larger than $G(\cdot)$.

To prove (b), recall that for a nonnegative random variable X, the expectation is defined as

$$E_F(X) = \int_0^{\infty} [1 - F(x)]\,dx.$$

By part (a), $F(x) \leq G(x)$, which implies, for any $x > 0$, that $1 - F(x) \geq 1 - G(x)$; integrating both sides from 0 to ∞ with respect to x, we get

$$E_F(X) = \int_0^{\infty} [1-F(x)]\,dx \geq \int_0^{\infty} [1-G(x)]\,dx = E_G(X).$$

Finally, we need to show that

$$\frac{E_F(X^2)}{E_F(X)} \geq \frac{E_G(X^2)}{E_G(X)}.$$

Observe that

$$\frac{E_F(X^2)}{E_F(X)} = \frac{\int_0^{\infty} x^2 f(x)\,dx}{\int_0^{\infty} x\,f(x)\,dx} = \int_0^{\infty} x\,\frac{x\,f(x)}{\int_0^{\infty} x\,f(x)\,dx}\,dx.$$

Define

$$f^*(x) = \frac{x\,f(x)}{\int_0^{\infty} x\,f(x)\,dx} = \frac{x\,f(x)}{E_F(X)};$$

then, $f^*(x)$ is a pdf. Hence, we can rewrite $E_F(X^2)/E_F(X)$ as

$$\frac{E_F(X^2)}{E_F(X)} = \int_0^\infty x \frac{x\, f(x)}{\int_0^\infty x\, f(x)\, dx}\, dx = \int_0^\infty x\, f^*(x)\, dx = E_{F^*}(X).$$

The same argument applies to $E_G(X^2)/E_G(X)$, so it is sufficient to prove that $E_{F^*}(X) \geq E_{G^*}(X)$, or equivalently, that $F^*(x) \leq G^*(x)$ for all $x > 0$. Now,

$$\frac{f^*(x)}{g^*(x)} = \frac{f(x)}{g(x)} \frac{E_G(X)}{E_F(X)}.$$

On the right hand side of this expression, the expectation ratio is bounded between 0 and 1; furthermore, both expectations are constant in X. Letting k denote the expectation ratio, we then have that

$$\frac{f^*(x)}{g^*(x)} = k\, \frac{f(x)}{g(x)}.$$

Since $f(x)/g(x)$ is increasing by hypothesis, it follows that $f^*(x)/g^*(x)$ is increasing. Now apply parts (a) and (b) of the theorem with respect to $f^*(x)$ and $g^*(x)$. □

Corollary A.13.1. *If $f(x)/g(x)$ is decreasing in x on $(0, \infty)$, then the signs in parts (a), (b) and (c) of Theorem A.13 are reversed.*

The proof follows by observing that $g(x)/f(x)$ is increasing.

A.1.3.1 MLR properties for $U_1|S \leq r$ and $U_2|S \leq r$

The following discussion is concerned with establishing MLR properties with respect to the distributions of $U_1|S \leq r$ and $U_2|S \leq r$, where $r > 0$ is a constant.

Theorem A.14. *With respect to the family of distributions of $U_1|S \leq r$, we have the following:*

(a) if $r > 0$ is fixed, then $U_1|S \leq r$ has increasing MLR in σ;

(b) if $\sigma = 1$ is fixed, then $U_1|S \leq r$ has decreasing MLR in r;

(c) if $\sigma = 1$ is fixed, then the density ratio $g_{U_1|S\leq r}(u_1|r)/g_{\nu_1}(u_1)$ is increasing in u_1.

With respect to the family of distributions of $U_2|S \leq r$, we have the following:

(d) if $r > 0$ is fixed, then $U_2|S \leq r$ has decreasing MLR in σ;

(e) if $\sigma = 1$ is fixed, then $U_2|S \leq r$ has increasing MLR in r;

(f) if $\sigma = 1$ is fixed, then the density ratio $g_{U_2|S\leq r}(u_2|r)/g_{\nu_2}(u_2)$ is decreasing in u_2.

Proof. Define $\theta = r/\sigma$. If we fix $r > 0$ and let $1 \leq \sigma_1 < \sigma_2 < \infty$, then $\theta_1 > \theta_2$; conversely, if $\sigma = 1$ is fixed and $0 < r_1 < r_2 < \infty$, then $\theta_1 < \theta_2$, or equivalently, $\theta_1^{-1} > \theta_2^{-1}$. By the duality between the PDP distributions of $U_1|S \leq r$ and $U_2|S \leq r$, on the one hand, and Theorem A.12(ii) on the other, it is sufficient to prove any one of items (a), (b), (d) or (e) to establish all four. Item (b) is selected for the task.

Let $\sigma = 1$ and let $0 < r_1 < r_2 < \infty$. We want to show that $U_1|S \leq r$ has decreasing MLR in r, or equivalently, we want to show that the density ratio

$$\frac{g^*(u_1|r_2)}{g^*(u_1|r_1)} = \frac{g_{\nu_1}(u_1)\, G_{\nu_2}(r_2 u_1)/I_{z_2}(\phi_2, \phi_1)}{g_{\nu_1}(u_1)\, G_{\nu_2}(r_1 u_1)/I_{z_1}(\phi_2, \phi_1)}$$

is monotone decreasing in u_1, where $z_i = r_i/(1 + r_i)$, $i = 1, 2$. Clearly, the $g(\cdot)$ terms cancel out; furthermore, the Beta cdf's $I_{z_i}(\cdot)$ are constant as a function of u_1, so they are immaterial to the discussion. What remains, then, is to show that $G_{\nu_2}(r_2 u_1)/G_{\nu_2}(r_1 u_1)$ is monotone decreasing in u_1, or equivalently, that its derivative is nonpositive uniformly in u_1.

The derivative of the above χ^2 cdf ratio with respect to u_1 is

$$\begin{aligned}
\frac{d}{du_1} \frac{G_{\nu_2}(r_2 u_1)}{G_{\nu_2}(r_1 u_1)} &= \frac{G_{\nu_2}(r_1 u_1)\, g_{\nu_2}(r_2 u_1)\, r_2 - G_{\nu_2}(r_2 u_1)\, g_{\nu_2}(r_1 u_1)\, r_1}{G_{\nu_2}^2(r_1 u_1)} \\
&= r_2 \frac{g_{\nu_2}(r_2 u_1)}{G_{\nu_2}(r_1 u_1)} - r_1 \frac{G_{\nu_2}(r_2 u_1)}{G_{\nu_2}(r_1 u_1)} \frac{g_{\nu_2}(r_1 u_1)}{G_{\nu_2}(r_1 u_1)}
\end{aligned}$$

After a little rearrangement of terms, we obtain

$$= \frac{G_{\nu_2}(r_2u_1)}{G_{\nu_2}(r_1u_1)}\left[r_2\frac{g_{\nu_2}(r_2u_1)}{G_{\nu_2}(r_2u_1)} - r_1\frac{g_{\nu_2}(r_1u_1)}{G_{\nu_2}(r_1u_1)}\right].$$

It is sufficient to show that the bracketed term above is nonpositive for all $u_1 > 0$. With another slight rearrangement, what we want to prove is

$$\frac{G_{\nu_2}(r_2u_1)}{G_{\nu_2}(r_1u_1)} \geq \frac{r_2}{r_1}\frac{g_{\nu_2}(r_2u_1)}{g_{\nu_2}(r_1u_1)}.$$

Now, the density ratio on the right hand side of the above inequality is

$$\frac{g_{\nu_2}(r_2u_1)}{g_{\nu_2}(r_1u_1)} = \frac{(r_2u_1)^{\phi_2-1}e^{-r_2u_1/2}}{(r_1u_1)^{\phi_2-1}e^{-r_1u_1/2}}$$

$$= \left(\frac{r_2}{r_1}\right)^{\phi_2-1}\exp\{(r_1-r_2)u_1/2\}.$$

Therefore, what we need to prove is

$$\frac{G_{\nu_2}(r_2u_1)}{G_{\nu_2}(r_1u_1)} \geq \left(\frac{r_2}{r_1}\right)^{\phi_2}\exp\{(r_1-r_2)u_1/2\}; \tag{A.16}$$

the exponential term is monotone decreasing in u_1 since $r_1 - r_2 < 0$.

To prove (A.16), we have

$$\frac{G_{\nu_2}(r_2u_1)}{G_{\nu_2}(r_1u_1)} = \frac{\int_0^{r_2u_1} u_2^{\phi_2-1}e^{-u_2/2}\,du_2}{\int_0^{r_1u_1} u_2^{\phi_2-1}e^{-u_2/2}\,du_2}.$$

On the right hand side, make the transformation $y = u_2/r_2$ in the numerator, and $y = u_2/r_1$ in the denominator. Then,

$$\frac{G_{\nu_2}(r_2u_1)}{G_{\nu_2}(r_1u_1)} = \frac{\int_0^{u_1} r_2\,(r_2y)^{\phi_2-1}\,e^{-r_2y/2}\,dy}{\int_0^{u_1} r_1\,(r_1y)^{\phi_2-1}\,e^{-r_1y/2}\,dy}$$

$$= \left(\frac{r_2}{r_1}\right)^{\phi_2}\frac{\int_0^{u_1} y^{\phi_2-1}\,e^{(r_1-r_2)y/2}\,e^{-r_1y/2}\,dy}{\int_0^{u_1} y^{\phi_2-1}\,e^{-r_1y/2}\,dy}$$

The term inserted into the integrand of the numerator is monotone decreasing in u_1; therefore,

$$\frac{G_{\nu_2}(r_2u_1)}{G_{\nu_2}(r_1u_1)} \geq \left(\frac{r_2}{r_1}\right)^{\phi_2}\exp\{(r_1-r_2)u_1/2\},$$

and the proof is complete for (a), (b), (d) and (e).

To prove (c), fix $\sigma = 1$ and consider the density ratio

$$\begin{aligned} \frac{g_{U_1|S\leq r}(u_1|r)}{g_{U_1}(u_1)} &= \frac{g_{\nu_1}(u_1)\, G_{\nu_2}(ru_1)/I_z(\phi_2,\phi_1)}{g_{\nu_1}(u_1)} \\ &= \frac{G_{\nu_2}(ru_1)}{I_z(\phi_2,\phi_1)} \\ &\propto G_{\nu_2}(ru_1), \end{aligned}$$

where $z = r/(1+r)$ and $I(\cdot)$ is independent of u_1. This is clearly increasing in u_1 for any $r > 0$, so (c) is proved.

To show (f), fix $\sigma = 1$. Then

$$\begin{aligned} \frac{g_{U_2|S\leq r}(u_2|r)}{g_{U_2}(u_2)} &= \frac{1 - G_{\nu_1}(u_2/r)}{I_z(\phi_2,\phi_1)} \\ &\propto 1 - G_{\nu_1}(u_2/r), \end{aligned}$$

since $I(\cdot)$ is independent of u_2. For any fixed $r > 0$, this function is decreasing in u_2, which establishes the desired result. □

A.1.3.2 MLR properties of $W|W \leq r/\sigma$.

We begin by noting that the density of $W|W \leq r/\sigma$ is, for $r > 0$,

$$g_{W^*}(w|r,\sigma) = \frac{\Gamma(\phi_1+\phi_2)}{\Gamma(\phi_1)\,\Gamma(\phi_2)}\,\frac{1}{I_{z^*}(\phi_2,\phi_1)}\,\frac{w^{\phi_2-1}}{(1+w)^{\phi_1+\phi_2}}\,\chi_{[0,r/\sigma]}(w), \qquad \text{(A.17)}$$

where $z^* = r/(\sigma + r)$ and $\chi(\cdot)$ is the indicator function

$$\chi_{[0,t]}(w) = \begin{cases} 1 & \text{if } 0 \leq w \leq t, \\ 0 & \text{otherwise.} \end{cases}$$

Theorem A.15. *With respect to the distribution of* $W|W \leq r/\sigma$,

(i) if $r > 0$ *is fixed, then* $W|W \leq r/\sigma$ *has decreasing MLR in* σ;

(ii) if $\sigma = 1$ is fixed, then $S|S \leq r$ has increasing MLR in r;

(iii) if $\sigma = 1$ is fixed, then $g_{S|S\leq r}(s|r)/g_S(s)$ is decreasing in s.

Proof. To prove (i), fix $r > 0$ and let $1 \leq \sigma_1 < \sigma_2 < \infty$. It is sufficient to show that the density ratio $g_{S|S\leq r}(s|\sigma_2)/g_{S|S\leq r}(s|\sigma_1)$ is decreasing in s. Then

$$\frac{g_{W^*}(r/\sigma_2|r,\sigma)}{g_{W^*}(r/\sigma_1|r,\sigma)} = \frac{\chi_{[0,r/\sigma_2]}(w)\, I_{z'}(\phi_2,\phi_1)}{\chi_{[0,r/\sigma_1]}(w)\, I_{z''}(\phi_2,\phi_1)},$$

where $z' = r/(\sigma_2 + r)$ and $z'' = r/(\sigma_1 + r)$. Since the Beta cdf ratio is independent of w, it plays no role here, so can be ignored. This leaves us with

$$\frac{\chi_{[0,r/\sigma_2]}(w)}{\chi_{[0,r/\sigma_1]}(w)} = \begin{cases} 1 & \text{if } w \leq r/\sigma_2, \\ 0 & \text{if } \dfrac{r}{\sigma_2} < w \leq \dfrac{r}{\sigma_1}. \end{cases}$$

Since this function is decreasing in σ when r is fixed, the result is established.

For (ii), fix $\sigma = 1$, and let $r_2 > r_1 > 0$. Then, it is easily shown that

$$\frac{g_{S|S\leq r_2}(s|r_2)}{g_{S|S\leq r_1}(s|r_1)} = \frac{\chi_{[0,r_2]}(s)}{\chi_{[0,r_1]}(s)}\, \frac{I_{z_2}(\phi_2,\phi_1)}{I_{z_1}(\phi_2,\phi_1)},$$

where $z_i = r_i/(1 + r_i)$, for $i = 1, 2$. Since the Beta cdf ratio is independent of s, the above density ratio is proportional to the indicator function ratio, which, as a function of s, is

$$\frac{\chi_{[0,r_2]}(s)}{\chi_{[0,r_1]}(s)} = \begin{cases} 1 & 0 \leq s \leq r_1; \\ \infty & r_1 < s \leq r_2. \end{cases}$$

The above ratio is defined on $[0, r_2]$; for $s > r_2$, the indicator function ratio is undefined (of the form $0/0$). Over the domain of definition, this ratio is a nondecreasing function of s; therefore, $S|S \leq r$ has increasing MLR in r for fixed $\sigma = 1$.

To establish (iii), the density ratio of interest is

$$\begin{aligned} \frac{g_{S|S\leq r}(s|r)}{g_S(s)} &= \frac{s^{\phi_2-1}/(1+s)^{\phi_1+\phi_2}}{s^{\phi_2-1}/(1+s)^{\phi_1+\phi_2}}\, \frac{1}{I_z(\phi_2,\phi_1)}\, \frac{\chi_{[0,r]}(s)}{\chi_{[0,\infty]}(s)} \\ &\propto \frac{\chi_{[0,r]}(s)}{\chi_{[0,\infty]}(s)}, \end{aligned}$$

since the unconditional density of S is defined on the nonnegative reals. This ratio is decreasing in s because

$$\frac{\chi_{[0,r]}(s)}{\chi_{[0,\infty]}(s)} = \begin{cases} 1 & \text{if } 0 \le s \le r; \\ 0 & \text{if } s > r. \end{cases}$$

□

A.2 Properties of bowl-shaped loss functions

Bowl-shaped loss functions are fundamental to establishment of risk dominance results in the normal variance problem, and can be applied to the balanced one-way random or fixed effects models as well. Brown (1968) formalized the relationships between monotone likelihood ratio, bowl-shaped loss functions and the (conditional) risk properties of various types of point estimators. The present objective is to provide some definitions and basic results concerning bowl-shaped or convex risk functions. To that end, let $L(t)$ represent a loss function which is (absolutely) continuous and nonnegative on the positive reals.

Definition A.8. $L(t)$ is said to be *bowl-shaped* if L is nondecreasing for all $t \ge 1$ and nonincreasing for all $t \in (0,1]$. $L(t)$ is said to be *strictly bowl-shaped* if L is strictly increasing for all $t \ge 1$ and strictly decreasing for all $t \in (0,1]$.

If $L(t)$ is (strictly) bowl-shaped, it is differentiable almost everywhere on $(0,\infty)$. It is assumed that the regularity conditions that allow interchange of derivative and integral are in force for integrals involving $L(t)$. When a scale family of distributions has the monotone likelihood ratio property, the following result holds, due to Brewster and Zidek (1974):

Lemma A.16. *If f is a density on $(0,\infty)$ and $\{f(x/c) : c > 0\}$ has monotone likelihood ratio, then $\int x\, L'(cx)\, f(x)\, dx$ has at most one sign change and $\int L(cx)\, f(x)\, dx$ is strictly bowl-shaped (or monotone).*

Under the conditions of this lemma, if the loss function $L(\cdot)$ is strictly bowl-shaped, then so is the risk function.

Convex loss or risk functions are a subset of the collection of bowl-shaped loss (or risk) functions. Of particular interest in this thesis is the entropy loss function $L(t) = t - 1 - \log t$. We will verify that this loss function is convex.

Definition A.9. A set $E \subset \Re^n$ is said to be convex if the line segment connecting any two points of E is contained in E.

Definition A.10. Let f be a real-valued function defined on a convex set E. If, for every $x_1, x_2 \in E$ and $t \in [0, 1]$, we have

$$f(tx_1 + (1-t)x_2) \leq t\, f(x_1) + (1-t)\, f(x_2),$$

then f is a convex function on E. If the inequality is strict, then f is said to be strictly convex on E.

The following is a standard result for verifying convexity of a function.

Lemma A.17. *Let f be a twice differentiable function on an interval $E \subset \Re$. If*

(a) f' is nondecreasing (increasing) on E with at most one sign change, and

(b) $f''(x) \geq 0$ for all $x \in E$,

then f is a convex function on E.

Lemma A.18. *The entropy loss function (2.75) is convex on the positive reals.*

Proof. The interval $(0, \infty)$ is a convex set by Definition A.9. We will show that the conditions of Lemma A.17 hold. Since $L(t) = t - 1 - \log t$, differentiation with respect to t yields $L'(t) = 1 - 1/t$, which is strictly increasing in t, negative for all $t < 1$ and positive for all $t > 1$. The second derivative is $L''(t) = t^{-2}$, which is positive for all $t > 0$ (and strictly decreasing). Since both conditions of the lemma are satisfied, the result follows. □

Lemma A.19. *If* $L(t) = t - 1 - \log t$ *and* f *is a density on* $(0, \infty)$ *such that* $\{f(x/c) : c > 0\}$ *has monotone likelihood ratio, then* $\int L(cx)\, f(x)\, dx$ *is strictly convex on* $(0, \infty)$.

The proof follows from Lemmas A.16 and A.18.

A.3 Results involving moments of $U_1^*(s)$ and $U_2^*(s)$

Lemma A.20. *Let* $g(\cdot)$ *represent a* χ^2 *density. Then*

$$x\, g_\nu(x) = \nu\, g_{\nu+2}(x). \tag{A.18}$$

Proof. Let $\phi = \nu/2$. Then

$$\begin{aligned}
x\, g_\nu(x) &= x\, \frac{x^{\phi-1}\, e^{-x/2}}{\Gamma(\phi)\, 2^\phi} \\
&= \frac{x^{(\phi+1)-1}\, e^{-x/2}}{\Gamma(\phi)\, 2^\phi} \\
&= \frac{\Gamma(\phi+1) \cdot 2}{\Gamma(\phi)}\, \frac{x^{(\phi+1)-1}\, e^{-x/2}}{\Gamma(\phi+1)\, 2^{\phi+1}} \\
&= \nu\, g_{\nu+2}(x).
\end{aligned}$$

□

Lemma A.21. *Under the conditions of Lemma A.20,*

$$x^2\, g_\nu(x) = \nu\, (\nu+2)\, g_{\nu+4}(x). \tag{A.19}$$

Proof. Again, let $\phi = \nu/2$. Then

$$\begin{aligned} x^2\, g_\nu(x) &= x^2\, \frac{x^{\phi-1}\, e^{-x/2}}{\Gamma(\phi)\, 2^\phi} \\ &= \frac{x^{(\phi+2)-1}\, e^{-x/2}}{\Gamma(\phi)\, 2^\phi} \\ &= \frac{2^2\, \Gamma(\phi+2)}{\Gamma(\phi)}\, \frac{x^{(\phi+2)-1}\, e^{-x/2}}{\Gamma(\phi+2)\, 2^{\phi+2}} \\ &= \nu\ (\nu+2)\ g_{\nu+4}(x). \end{aligned}$$

□

This idea can be generalized to arbitrary k, as shown in the following theorem:

Theorem A.22. *If $g(\cdot)$ denotes a χ^2 density and $\phi = \nu/2$, then for $k = 1, 2, 3, \ldots,$*

$$x^k\, g_\nu(x) = \frac{2^k\, \Gamma(\phi+k)}{\Gamma(\phi)}\, g_{\nu+2k}(x). \tag{A.20}$$

Proof. Lemmas A.20 and A.21 establish the $k = 1$ and $k = 2$ cases. To complete the proof by induction, assume that the hypothesis is true for a generic integer k; it suffices to show that the result must also hold for $k+1$. To that end,

$$\begin{aligned} x^{k+1}\, g_\nu(x) &= x\, \left(x^k\, g_\nu(x)\right) \\ &= x \left(\frac{2^k\, \Gamma(\phi+k)}{\Gamma(\phi)}\, g_{\nu+2k}(x) \right) \\ &= \frac{2^k\, \Gamma(\phi+k)}{\Gamma(\phi)}\, x \left(\frac{x^{\phi+k-1}\, e^{-x/2}}{\Gamma(\phi+k)\, 2^{\phi+k}} \right) \\ &= \frac{2^k\, \Gamma(\phi+k)}{\Gamma(\phi)}\, \frac{\Gamma(\phi+(k+1))\, 2^{\phi+(k+1)}}{\Gamma(\phi+k)\, 2^{\phi+k}} \left(\frac{x^{(\phi+(k+1))-1}\, e^{-x/2}}{\Gamma(\phi+(k+1))\, 2^{\phi+(k+1)}} \right) \\ &= \frac{2^{k+1}\, \Gamma(\phi+(k+1))}{\Gamma(\phi)}\, g_{\nu+2(k+1)}(x). \end{aligned}$$

This completes the proof. □

A.3.1 Moments of $U_1^*(s)$

Moment generating function. We begin with two preliminary lemmas, followed by a derivation of the mgf of $U_1^*(s)$ for any fixed $s > 0$.

Lemma A.23. *Let $g_\nu(\cdot)$ denote a χ^2_ν pdf, and let $\phi = \nu/2$. Then*

$$e^{tx} g_\nu(x) = (1-2t)^{-\phi}\, g_\nu\left[(1-2t)x\right]. \tag{A.21}$$

Proof.

$$\begin{aligned} e^{tx}\, g_\nu(x) &= e^{tx}\left[\frac{x^{\phi-1}\, e^{-x/2}}{\Gamma(\phi)2^\phi}\right] \\ &= \frac{x^{\phi-1}\, e^{-\frac{x}{2}(1-2t)}}{\Gamma(\phi)2^\phi}. \end{aligned}$$

Let $y = (1-2t)x$, so that $dy = (1-2t)\, dx$. Then, substituting for x,

$$\begin{aligned} &= \frac{(y/(1-2t))^{\phi-1} e^{-y/2}(1-2t)^{-1}}{\Gamma(\phi)2^\phi} \\ &= (1-2t)^{-\phi}\, g_\nu(y) \\ &= (1-2t)^{-\phi}\, g_\nu\big[(1-2t)x\big]. \end{aligned}$$

□

Lemma A.24. *Let $g_\nu(\cdot)$ denote a χ^2_ν pdf, $G(\cdot)$ its cdf and let $\phi = \nu/2$. Then*

$$\int_0^\infty g_{\nu_1}(x)\, G_{\nu_2}(sx)\, dx = I_z(\phi_2, \phi_1), \tag{A.22}$$

where $I(\cdot)$ denotes a Beta cdf evaluated at the subscripted value.

Proof.

$$\int_0^\infty g_{\nu_1}(x)\, G_{\nu_2}(sx)\, dx = \int_0^\infty g_{\nu_1}(x) \int_0^{sx} g_{\nu_2}(y)\, dy\, dx,$$

where X and Y have independent χ^2 distributions with ν_1 and ν_2 degrees of freedom, respectively. Now,

$$= \int_0^\infty \int_0^{sx} \frac{1}{\Gamma(\phi_1)\Gamma(\phi_2)2^{\phi_1+\phi_2}} x^{\phi_1-1} y^{\phi_2-1} e^{-(x+y)/2}\, dy\, dx.$$

Make the bivariate transformation

$$V = \frac{Y}{X}, \qquad T = X,$$

with inverse transformation

$$Y = VT, \qquad X = T,$$

and Jacobian T. Note that $t > 0$ and $0 < v \leq s$. Under this transformation, we have

$$\int_0^\infty g_{\nu_1}(x)\, G_{\nu_2}(sx)\, dx = \int_0^\infty \int_0^s \frac{1}{\Gamma(\phi_1)\Gamma(\phi_2)2^{\phi_1+\phi_2}} v^{\phi_1-1} t^{\phi_1+\phi_2-1}\, e^{-\frac{t}{2}(1+v)}\, dt\, dv.$$

Interchanging the order of integration,

$$= \int_0^s \frac{v^{\phi_2-1}}{\Gamma(\phi_1)\Gamma(\phi_2)} \int_0^\infty \frac{1}{2^{\phi_1+\phi_2}} t^{\phi_1+\phi_2-1} \exp\{-\tfrac{t}{2}(1+v)\}\, dt\, dv.$$

Now, make the transformation $w = t(1+v)$; then,

$$t = \frac{w}{1+v}, \qquad dt = \frac{dw}{1+v},$$

so that

$$\begin{aligned}\int_0^\infty g_{\nu_1}(x)\, G_{\nu_2}(sx)\, dx &= \int_0^s \frac{v^{\phi_2-1}}{\Gamma(\phi_1)\Gamma(\phi_2)} \left[\int_0^\infty \frac{1}{2^{\phi_1+\phi_2}} \left(\frac{w}{1+v}\right)^{\phi_1+\phi_2-1} e^{-w/2}\, \frac{dw}{1+v}\right] dv \\ &= \int_0^s \frac{\Gamma(\phi_2+\phi_1)}{\Gamma(\phi_2)\Gamma(\phi_1)} v^{\phi_2-1} (1+v)^{-(\phi_1+\phi_2)}\, dv.\end{aligned}$$

Finally, make the transformation $r = v/(1+v)$ with differential $dr = dv/(1+v)^2$. Then, using $z = s/(1+s)$, we have

$$\begin{aligned}&= \int_0^z \frac{\Gamma(\phi_2+\phi_1)}{\Gamma(\phi_2)\Gamma(\phi_1)} r^{\phi_2-1} (1-r)^{\phi_1-1}\, dx \\ &= I_z(\phi_2, \phi_1),\end{aligned}$$

which completes the proof. □

Lemma A.25. *If* $z^* = s/(1-2t+s)$, $\phi_i = \nu_i/2$, $i = 1,2$, *and if* $g(\cdot)$ *and* $G(\cdot)$ *are defined as in Lemma A.24, then*

$$\int_0^\infty g_{\nu_1}\left[(1-2t)x\right] G_{\nu_2}(sx)\,dx = I_{z^*}(\phi_2, \phi_1). \tag{A.23}$$

Proof. We have

$$g_{\nu_1}\left[(1-2t)x\right] G_{\nu_2}(sx)\,dx = \frac{\left[(1-2t)x\right]^{\phi_1-1} e^{-(1-2t)x/2}}{\Gamma(\phi_1)2^{\phi_1}} G_{\nu_2}(sx)\,(1-2t)\,dx.$$

Let $y = (1-2t)x$, with $dy = (1-2t)\,dx$. Making the appropriate substitution,

$$= \frac{y^{\phi_1-1}e^{-y/2}}{\Gamma(\phi_1)2^{\phi_1}} G_{\nu_2}\left(\frac{sy}{1-2t}\right).$$

Letting $r = s/(1-2t)$, the above expression is of the form $g_{\nu_1}(y)\,G_{\nu_2}(ry)$, so that by Lemma A.24,

$$\int_0^\infty g_{\nu_1}(y) G_{\nu_2}(ry)\,dy = I_{r^*}(\phi_2, \phi_1),$$

where we define $r^* = r/(1+r)$. Then

$$r^* = \frac{r}{1+r} = \frac{s/(1-2t)}{1+(s/(1-2t))} = \frac{s}{1-2t+s} = z^*,$$

from which the result follows. □

The following result establishes the form of the moment generating function of $U_1^*(s)$ for any fixed $s > 0$.

Theorem A.26. *The moment generating function of* $U_1^*(s)$ *is, for any fixed* $s > 0$,

$$M_{U_1^*}(t) = (1-2t)^{-\phi_1}\frac{I_{z^*}(\phi_2, \phi_1)}{I_z(\phi_2, \phi_1)}, \tag{A.24}$$

where $z^* = s/(1-2t+s)$ *and* $\phi_i = \nu_i/2$, $i = 1,2$.

Proof. By definition, the mgf of $U_1^*(s)$ is, for any arbitrary $s > 0$,

$$M_{U_1^*}(t) = \frac{\int_0^\infty e^{tx} g_{\nu_1}(x) G_{\nu_2}(sx)\, dx}{I_z(\phi_2, \phi_1)}.$$

By Lemma A.23,

$$= (1-2t)^{-\phi_1} \frac{\int_0^\infty g_{\nu_1}\left[(1-2t)x\right] G_{\nu_2}(sx)\, dx}{I_z(\phi_2, \phi_1)},$$

and by Lemma A.25,

$$= (1-2t)^{-\phi_1} \frac{I_{z^*}(\phi_2, \phi_1)}{I_z(\phi_2, \phi_1)}.$$

□

Limiting m.g.f. From the moment generating function (A.24), we wish to examine the limiting behavior of $U_1^*(s)$ as $s \to 0$ and as $s \to \infty$. Let $z^* = s/(1-2t+s)$.

Theorem A.27. *With respect to the moment generating function* (A.24) *of* $U_1^*(s)$,

(a) $\lim_{s\to\infty} M_{U_1^*}(t|s) = (1-2t)^{-\phi_1}$; *i. e.,* $U_1^*(s) \xrightarrow{d} \chi^2_{\nu_1}$;

(b) $\lim_{s\to 0} M_{U_1^*}(t|s) = (1-2t)^{-\phi_1+\phi_2}$; *i. e.,* $U_1^*(s) \xrightarrow{d} \chi^2_{\nu_1+\nu_2}$.

Proof. To prove (a),

$$\begin{aligned}\lim_{s\to\infty} M_{U_1^*}(t|s) &= \lim_{s\to\infty} (1-2t)^{-\phi_1} \frac{I_{z^*}(\phi_2, \phi_1)}{I_z(\phi_2, \phi_1)} \\ &= (1-2t)^{-\phi_1} \lim_{s\to\infty} \frac{I_{z^*}(\phi_2, \phi_1)}{I_z(\phi_2, \phi_1)}.\end{aligned}$$

Consider the limiting behavior of the arguments z^* and $z = s/(1+s)$ in the two Beta cdf's above:

$$\lim_{s\to\infty} \frac{s}{1-2t+s} = \lim_{s\to\infty} \frac{1}{1+(1-2t)/s} = 1,$$

and

$$\lim_{s\to\infty}\frac{s}{1+s}=\lim_{s\to\infty}\frac{1}{1+(1/s)}=1;$$

thus, as $s\to\infty$, $\lim_{s\to\infty} I_{z^*}(\phi_2,\phi_1)=1$ and $\lim_{s\to\infty} I_z(\phi_2,\phi_1)=1$. Consequently,

$$\lim_{s\to\infty} M_{U_1^*}(t)=(1-2t)^{-\phi_1}; \tag{A.25}$$

i. e., as $s\to\infty$, $U_1^*(s)\xrightarrow{d}\chi^2_{\nu_1}$.

To prove (b), it is clear that both $z^*=s/(1-2t+s)$ and $z=s/(1+s)$ tend to zero as $s\to 0$, leaving the limiting m.g.f. in the indeterminate form 0/0. Applying L'Hôpital's rule,

$$\lim_{s\to 0} M_{U_1^*}(t)=\lim_{s\to 0}(1-2t)^{-\phi_1}\frac{\frac{\partial}{\partial s}I_{z^*}(\phi_2,\phi_1)}{\frac{\partial}{\partial s}I_z(\phi_2,\phi_1)}.$$

Ignoring the constant term in the numerator for the nonce,

$$\frac{\partial}{\partial s}I_{z^*}(\phi_2,\phi_1)=\frac{\partial}{\partial s}\int_0^{z^*} v^{\phi_2-1}(1-v)^{\phi_1-1}\,dv.$$

By Leibnitz's rule, we have, for $z^*=s/(1-2t+s)$,

$$\begin{aligned}\frac{\partial}{\partial s}I_{z^*}(\phi_2,\phi_1)&=\left(\frac{s}{1-2t+s}\right)^{\phi_2-1}\left(1-\frac{s}{1-2t+s}\right)^{\phi_1-1}\left(\frac{1-2t}{(1-2t+s)^2}\right)\\&=\frac{s^{\phi_2-1}(1-2t)^{\phi_1}}{(1-2t+s)^{\phi_1+\phi_2}}.\end{aligned}$$

By a similar argument, the denominator is, for $z=s/(1+s)$,

$$\begin{aligned}\frac{\partial}{\partial s}I_z(\phi_2,\phi_1)&=\frac{\partial}{\partial s}\int_0^{z} w^{\phi_2-1}(1-w)^{\phi_1-1}\,dv\\&=\frac{s^{\phi_2-1}}{(1+s)^{\phi_1+\phi_2}}.\end{aligned}$$

Therefore,

$$\begin{aligned}\lim_{s\to 0} M_{U_1^*(s)}(t) &= (1-2t)^{-\phi_1} \lim_{s\to 0}\left[\left(\frac{s^{\phi_2-1}\,(1-2t)^{\phi_1}}{(1-2t+s)^{\phi_1+\phi_2}}\right)\Big/\left(\frac{s^{\phi_2-1}}{(1+s)^{\phi_1+\phi_2}}\right)\right]\\ &= (1-2t)^{-\phi_1} \lim_{s\to 0}\frac{(1+s)^{\phi_1+\phi_2}\,(1-2t)^{\phi_1}}{(1-2t+s)^{\phi_1+\phi_2}}\\ &= \frac{(1-2t)^{-\phi_1}\,(1-2t)^{\phi_1}}{(1-2t)^{\phi_1+\phi_2}}\\ &= (1-2t)^{-(\phi_1+\phi_2)}.\end{aligned}$$

In other words, as $s \to 0$, $U_1^*(s) \xrightarrow{d} \chi^2_{\nu_1+\nu_2}$. This completes the proof. □

Moments of U_1^*. The following result produces the moments of $U_1^*(s)$ for any arbitrarily chosen $s > 0$:

Theorem A.28. *For $k = 1, 2, 3, \ldots$ and any fixed $s > 0$,*

$$E(U_1^{*\,k}) = \frac{2^k\Gamma(\phi_1+k)}{\Gamma(\phi_1)}\,\frac{I_z(\phi_2,\phi_1+k)}{I_z(\phi_2,\phi_1)}. \tag{A.26}$$

Remark A.1. As long as $\phi_1 - k > 0$ for integer k, the k^{th} negative moment $E(U_1^{*\,-k})$ also exists and has the form (A.26), where $-k$ is substituted for k on the right hand side of (A.26).

Proof. Since U_1^* has a 'shifted' χ^2 distribution with (ν_2, ν_1) degrees of freedom, its k^{th} moment is defined as (for $k = 1, 2, 3, \ldots$)

$$E(U_1^{*\,k}) = \frac{\int_0^\infty x^k\, g_{\nu_1}(x) G_{\nu_2}(sx)\, dx}{I_z(\phi_2,\phi_1)}.$$

By Theorem A.22,

$$= \frac{\Gamma(\phi_1+k)\,2^k}{\Gamma(\phi_1)}\,\frac{\int_0^\infty g_{\nu_1+2k}(x)\,G_{\nu_2}(sx)\,dx}{I_z(\phi_2,\phi_1)}.$$

By Lemma A.24,

$$= \frac{\Gamma(\phi_1+k)\,2^k}{\Gamma(\phi_1)}\,\frac{I_z(\phi_2,\phi_1+k)}{I_z(\phi_2,\phi_1)}.$$

□

Corollary A.28.1. *For any fixed $s > 0$, the first two moments of $U_1^*(s)$ are given by*

$$E[U_1^*] = \nu_1 \frac{I_z(\phi_2, \phi_1 + 1)}{I_z(\phi_2, \phi_1)},$$
$$E[U_1^{*2}] = \nu_1 (\nu_1 + 2) \frac{I_z(\phi_2, \phi_1 + 2)}{I_z(\phi_2, \phi_1)}.$$

From Corollary A.28.1 and the definition of variance, we have

$$\mathrm{Var}[U_1^*] = \nu_1 \frac{I_z(\phi_2, \phi_1 + 1)}{I_z(\phi_2, \phi_1)} \left[(\nu_1 + 2) \frac{I_z(\phi_2, \phi_1 + 2)}{I_z(\phi_2, \phi_1 + 1)} - \nu_1 \frac{I_z(\phi_2, \phi_1 + 1)}{I_z(\phi_2, \phi_1)} \right]. \quad \text{(A.27)}$$

For any fixed $s > 0$, the mean of U_1^* is shown by Corollary A.2.1 to be

$$E[U_1^*] = J_z(\phi_2, \phi_1) = \nu_1 + 2\delta_z(\phi_2, \phi_1).$$

It is of interest to derive a simpler expression for the variance of U_1^*.

Corollary A.28.2. *For any fixed $s > 0$, the variance of $U_1^*(s)$ can be written as*

$$\mathrm{Var}[U_1^*(s)] = 2[\nu_1 + 2\delta_z(\phi_2, \phi_1) + \psi_z(\phi_2, \phi_1)], \quad \text{(A.28)}$$

where $\delta_z(\phi_2, \phi_1)$ is defined by (A.5) *and $\psi_z(\phi_2, \phi_1)$ by* (A.6).

Proof. Start with the expression (A.27) of $\mathrm{Var}[U_1^*(s)]$ for a given $s > 0$. By Corollary A.2.1, one can write

$$\nu_1 + 2\delta_z(\phi_2, \phi_1) = \nu_1 \frac{I_z(\phi_2, \phi_1 + 1)}{I_z(\phi_2, \phi_1)},$$
$$(\nu_1 + 2) + 2\delta_z(\phi_2, \phi_1 + 1) = (\nu_1 + 2) \frac{I_z(\phi_2, \phi_1 + 2)}{I_z(\phi_2, \phi_1 + 1)}.$$

Then, equation (A.27) reduces to

$$\mathrm{Var}[U_1^*(s)] = 2 (\nu_1 + 2\delta_z(\phi_2, \phi_1)) [1 + \delta_z(\phi_2, \phi_1 + 1) - \delta_z(\phi_2, \phi_1)].$$

Define $\Delta(z) = \delta_z(\phi_2, \phi_1 + 1) - \delta_z(\phi_2, \phi_1)$. Then

$$\begin{aligned} \Delta(z) &= \frac{\Gamma(\phi_1 + \phi_2 + 1)}{\Gamma(\phi_2)\,\Gamma(\phi_1 + 1)} \frac{z^{\phi_2} (1 - z)^{\phi_1 + 1}}{I_z(\phi_2, \phi_1 + 1)} - \frac{\Gamma(\phi_1 + \phi_2)}{\Gamma(\phi_2)\,\Gamma(\phi_1)} \frac{z^{\phi_2} (1 - z)^{\phi_1}}{I_z(\phi_2, \phi_1)} \\ &= \frac{\Gamma(\phi_1 + \phi_2)}{\Gamma(\phi_2)\,\Gamma(\phi_1)} \frac{z^{\phi_2} (1 - z)^{\phi_1}}{I_z(\phi_2, \phi_1)} \left[\frac{\phi_1 + \phi_2}{\phi_1} \frac{I_z(\phi_2, \phi_1)}{I_z(\phi_2, \phi_1 + 1)} (1 - z) - 1 \right] \\ &= \delta_z(\phi_2, \phi_1) \left[\frac{(\nu_1 + \nu_2)(1 - z)}{\nu_1 + 2\delta_z(\phi_2, \phi_1)} - 1 \right]. \end{aligned}$$

Now abbreviating $\delta_z(\phi_2, \phi_1)$ by δ, and substituting the above term in the variance formula yields

$$\begin{aligned}\text{Var}[U_1^*] &= 2\,(\nu_1 + 2\delta)\left[1 + \delta\left(\frac{(\nu_1+\nu_2)(1-z)}{\nu_1 + 2\delta} - 1\right)\right] \\ &= 2[(\nu_1 + 2\delta) + \delta(\nu_1+\nu_2)(1-z) - \delta(\nu_1 + 2\delta)].\end{aligned}$$

By (A.12), $\nu_1 + \nu_2 = (\nu_1 + 2\delta) + (\nu_2 - 2\delta)$. Using this result, along with the trick that $1 = z + (1-z)$, we obtain

$$\begin{aligned}&= 2\,[\nu_1 + 2\delta + \delta(1-z)\{(\nu_1+2\delta) + (\nu_2 - 2\delta)\} - \delta(z + (1-z))(\nu_1 + 2\delta)] \\ &= 2\,[\nu_1 + 2\delta + \delta\{(\nu_2 - 2\delta)(1-z) - z(\nu_1 + 2\delta)\}] \\ &= 2\,(\nu_1 + 2\delta + \psi),\end{aligned}$$

where $\psi = \psi_z(\phi_2, \phi_1)$. This completes the proof. □

Finally, the following result provides a lower bound for $E(U_1^*(s))$ for any fixed $s > 0$.

Lemma A.29. *Let $U_1 \sim \chi^2_{\nu_1}$ with mean ν_1. Then, for any $s > 0$,*

$$\frac{E[U_1^*(s)]}{E(U_1)} \geq 1.$$

Proof.

$$\begin{aligned}\frac{E[U_1^*(s)]}{E(U_1)} &= \frac{\nu_1 + 2\delta}{\nu_1} \\ &= 1 + \frac{2\delta}{\nu_1}.\end{aligned}$$

Since $2\delta(\cdot)$ is bounded between 0 and ν_2 by Theorem A.4 for all $s > 0$, the result is immediate. □

The relevance of this result is that the mean of U_1^* is always at least as large as the mean of U_1, the equality coming in the limit as $s \to \infty$.

A.3.2 Moments of U_2^*

This subsection essentially mimics the previous one except that the results pertain to the random variable U_2^* rather than U_1^*. We begin with the mgf of U_2^*, and proceed to results involving the moments of U_2^*. The notation used in the previous subsection carries over.

Moment generating function of U_2^*.

Lemma A.30. *Let $g(\cdot)$ denote a χ^2 pdf and $G(\cdot)$ its cdf. Then,*

$$\int_0^\infty g_{\nu_2}(y)\,[1 - G_{\nu_1}(y/s)]\;dy = I_z(\phi_2, \phi_1), \tag{A.29}$$

where $I(\cdot)$ is a Beta cdf.

Proof.

$$\int_0^\infty g_{\nu_2}(y)\,[1 - G_{\nu_1}(y/s)]\;dy = 1 - \int_0^\infty g_{\nu_2}(y) G_{\nu_1}(y/s)\;dx$$

Let $r = 1/s$. Then

$$r^* = \frac{r}{1+r} = \frac{1/s}{1+1/s} = \frac{1}{1+s} = 1 - z;$$

thus, by Lemma A.24,

$$\begin{aligned}\int_0^\infty g_{\nu_2}(y) G_{\nu_1}(ry)\;dx &= I_{r^*}(\phi_1, \phi_2)\\ &= I_{1-z}(\phi_1, \phi_2),\end{aligned}$$

so that by Fact A.2,

$$\begin{aligned}\int_0^\infty g_{\nu_2}(y)\,[1 - G_{\nu_1}(ry)]\;dy &= 1 - \int_0^\infty g_{\nu_2}(y) G_{\nu_1}(ry)\;dy\\ &= 1 - I_{1-z}(\phi_1, \phi_2)\\ &= I_z(\phi_2, \phi_1).\end{aligned}$$

□

Theorem A.31. *For any fixed $s > 0$, the moment generating function of $U_2^*(s)$ is given by*

$$M_{U_2^*}(t|s) = (1-2t)^{-\phi_2} \frac{I_{z^+}(\phi_2, \phi_1)}{I_z(\phi_2, \phi_1)} \tag{A.30}$$

where $z^+ = (1-2t)s/(1+(1-2t)s)$.

Proof. By definition,

$$\begin{aligned} M_{U_2^*}(t) &= \int_0^\infty e^{ty} h(y; \nu_2, \nu_1, s)\, dy \\ &= \frac{\int_0^\infty e^{ty} g_{\nu_2}(y) \left[1 - G_{\nu_1}(y/s)\right] dy}{I_z(\phi_2, \phi_1)}. \end{aligned}$$

Working with the integral in the numerator (since the denominator is fixed as a function of t),

$$\int_0^\infty e^{ty} g_{\nu_2}(y) \left[1 - G_{\nu_1}(y/s)\right] dy = \int_0^\infty e^{ty} g_{\nu_2}(y)\, dy - \int_0^\infty e^{ty} g_{\nu_2}(y) G_{\nu_1}(y/s)\, dy.$$

Since $g_{\nu_2}(y)$ is a $\chi^2_{\nu_2}$ density, the first integral in the above equation is the m.g.f. $(1-2t)^{-\phi_2}$. The second integral follows from Lemma A.23:

$$= (1-2t)^{-\phi_2} \left(1 - \int_0^\infty g_{\nu_2}\left[(1-2t)y\right] G_{\nu_1}(y/s)\, dy\right).$$

By Lemma A.25, with $r = 1/s$, one obtains

$$= (1-2t)^{-\phi_2} \left(1 - I_{s^*}(\phi_1, \phi_2)\right),$$

where $s^* = (1/s)/((1-2t) + (1/s)) = (1 + (1-2t)s)^{-1}$. Then, by Fact A.2,

$$= (1-2t)^{-\phi_2} I_{z^+}(\phi_2, \phi_1),$$

where $z^+ = 1 - s^*$. Having solved the integral in the numerator, the moment generating function of U_2^* follows directly. □

Limiting m.g.f.

Theorem A.32. *With respect to the moment generating function* (A.30) *of $U_2^*(s)$,*

(a) $\lim_{s\to\infty} M_{U_2^*}(t|s) = (1-2t)^{-\phi_2}$, *i. e.,* $U_2^*(s) \xrightarrow{d} \chi^2_{\nu_2}$;

(b) $\lim_{s\to 0} M_{U_2^*}(t|s) = 1$, *which implies that $U_2^*(s)$ becomes degenerate at 0 as $s \to 0$.*

Proof. Consider first the case as $s \to \infty$:

$$\lim_{s\to\infty} M_{U_2^*}(t) = \lim_{s\to\infty}(1-2t)^{-\phi_2}\frac{I_{z^*}(\phi_2,\phi_1)}{I_z(\phi_2,\phi_1)}.$$

As $s \to \infty$,

$$z^* = \frac{(1-2t)s}{1+(1-2t)s} = \frac{1-2t}{1/s+(1-2t)} \xrightarrow{s\to\infty} 1,$$

so that the ratio of both Beta cdf's tends to one in the limit. Therefore,

$$\lim_{s\to\infty} M_{U_2^*}(t|s) = (1-2t)^{-\phi_2},$$

or equivalently, $U_2^*(s) \xrightarrow{d} \chi^2_{\nu_2}$ as $s \to \infty$. This proves (a).

Next, consider the limiting mgf of U_2^* as $s \to 0$. In this case, for $s^* = (1-2t)s$, both $z^* = s^*/(1+s^*)$ and $z = s/(1+s)$ tend to zero as $s \to 0$; thus, the ratio of Beta cdf's in the mgf of U_2^* goes to 0/0, so L'Hôpital's rule is applied. Since we are differentiating with respect to s, it is sufficient to restrict attention to the Beta cdf's:

$$\begin{aligned}
\frac{\frac{\partial}{\partial s} I_{z^*}(\phi_2,\phi_1)}{\frac{\partial}{\partial s} I_z(\phi_2,\phi_1)} &= \frac{\frac{\partial}{\partial s}\int_0^{z^*} v^{\phi_2-1}(1-v)^{\phi_1-1}\,dv}{\frac{\partial}{\partial s}\int_0^{z} w^{\phi_2-1}(1-w)^{\phi_1-1}\,dw} \\
&= \frac{(z^*)^{\phi_2-1}(1-z^*)^{\phi_1-1}}{z^{\phi_2-1}(1-z)^{\phi_1-1}(1+s)^{-2}}\left(\frac{1-2t}{(1-2t+s)^2}\right) \\
&= \frac{\left(\frac{(1-2t)s}{1+(1-2t)s}\right)^{\phi_2-1}\left(1-\frac{(1-2t)s}{1+(1-2t)s}\right)^{\phi_1-1}\frac{1-2t}{(1+(1-2t)s)^2}}{\left(\frac{s}{1+s}\right)^{\phi_2-1}\left(\frac{1}{1+s}\right)^{\phi_1-1}\left(\frac{1}{(1+s)^2}\right)} \\
&= (1-2t)^{\phi_2}\left(\frac{1+s}{1+(1-2t)s}\right)^{\phi_1+\phi_2},
\end{aligned}$$

after some algebraic reduction. Bringing things together, the limiting mgf of $U_2^*(s)$ as $s \to 0$ is

$$\begin{aligned}
\lim_{s\to 0} M_{U_2^*}(t) &= \lim_{s\to 0}(1-2t)^{-\phi_2}(1-2t)^{\phi_2}\left(\frac{1+s}{1+(1-2t)s}\right)^{\phi_1+\phi_2} \\
&= \lim_{s\to 0}\left(\frac{1+s}{1+(1-2t)s}\right)^{\phi_1+\phi_2} \\
&= 1.
\end{aligned}$$

Since the limiting mgf of $U_2^*(s)$ as $s \to 0$ is a constant, its limiting distribution becomes degenerate at 0. This completes the proof. □

Moments of U_2^*. For any fixed $s > 0$, the moments of $U_2^*(s)$ can be obtained from the following result.

Theorem A.33. *For $k = 1, 2, 3, \ldots$ and any fixed $s > 0$,*

$$E[U_2^{*\,k}] = \frac{2^k\Gamma(\phi_2+k)}{\Gamma(\phi_2)}\frac{I_z(\phi_2+k,\phi_1)}{I_z(\phi_2,\phi_1)}. \tag{A.31}$$

Remark A.2. As long as $\phi_2 - k > 0$ for integer k, the k^{th} negative moment $E(U_2^{*\,-k})$ exists and has the form (A.31), where $-k$ substitutes for k on the right side of (A.31).

Proof. With respect to the density (A.2) of U_2^*, the k^{th} moment is expressible as

$$\begin{aligned}
E[U_2^{*\,k}] &= \int_0^\infty y^k h(y;\nu_2,\nu_1,s)\,dy \\
&= \frac{\int_0^\infty y^k g_{\nu_2}(y)\,[1-G_{\nu_1}(y/s)]\,dy}{I_z(\phi_2,\phi_1)}.
\end{aligned}$$

By Theorem A.22,

$$y^k g_{\nu_2}(y) = \frac{2^k\Gamma(\phi_2+k)}{\Gamma(\phi_2)}\,g_{\nu_2+2k}(y),$$

so that

$$\begin{aligned}\frac{\int_0^\infty y^k g_{\nu_2}(y)\,[1-G_{\nu_1}(y/s)]\,dy}{I_z(\phi_2,\phi_1)} &= \frac{2^k\Gamma(\phi_2+k)}{\Gamma(\phi_2)}\frac{\int_0^\infty g_{\nu_2+2k}(y)\,[1-G_{\nu_1}(y/s)]\,dy}{I_z(\phi_2,\phi_1)}\\ &= \frac{2^k\Gamma(\phi_2+k)}{\Gamma(\phi_2)}\frac{\left[1-\int_0^\infty g_{\nu_2+2k}(y)G_{\nu_1}(y/s)\,dy\right]}{I_z(\phi_2,\phi_1)}\\ &= \frac{2^k\Gamma(\phi_2+k)}{\Gamma(\phi_2)}\frac{\left[1-I_{1-z}(\phi_1,\phi_2+k)\right]}{I_z(\phi_2,\phi_1)},\end{aligned}$$

from the proof of Lemma A.30. Thus,

$$= \frac{2^k\Gamma(\phi_2+k)}{\Gamma(\phi_2)}\frac{I_z(\phi_2+k,\phi_1)}{I_z(\phi_2,\phi_1)}.$$

□

Corollary A.33.1. *For any fixed $s>0$, the first two moments of $U_2^*(s)$ are given by*

$$\begin{aligned}E[U_2^*(s)] &= \nu_2\,\frac{I_z(\phi_2+1,\phi_1)}{I_z(\phi_2,\phi_1)},\\ E[U_2^{*\,2}(s)] &= \nu_2\,(\nu_2+2)\,\frac{I_z(\phi_2+2,\phi_1)}{I_z(\phi_2,\phi_1)}.\end{aligned}$$

For any fixed $s>0$, the variance of $U_2^*(s)$ is

$$\mathrm{Var}[U_2^*(s)] = \nu_2\,\frac{I_z(\phi_2+1,\phi_1)}{I_z(\phi_2,\phi_1)}\left[(\nu_2+2)\,\frac{I_z(\phi_2+2,\phi_1)}{I_z(\phi_2+1,\phi_1)} - \nu_2\,\frac{I_z(\phi_2+1,\phi_1)}{I_z(\phi_2,\phi_1)}\right]. \quad \text{(A.32)}$$

By Corollary A.3.1, the mean of U_2^* can be written as $E[U_2^*]=\nu_2-2\delta_z(\phi_2,\phi_1)$. Having found a simplified expression for $\mathrm{Var}[U_1^*]$, we would like to find a similar expression for $\mathrm{Var}[U_2^*]$.

Corollary A.33.2. *The variance of $U_2^*(s)$ for any fixed $s>0$ can be written as*

$$\mathrm{Var}[U_2^*(s)] = 2\,[(\nu_2-2\delta_z(\phi_2,\phi_1)+\psi_z(\phi_2,\phi_1)]. \quad \text{(A.33)}$$

Proof. Starting from (A.32) and applying Corollary A.2.1, one can write

$$\mathrm{Var}[U_2^*(s)] = (\nu_2-2\delta_z(\phi_2,\phi_1))[(\nu_2+2)-2\delta_z(\phi_2+1,\phi_1)-(\nu_2-2\delta_z(\phi_2,\phi_1))]. \quad \text{(A.34)}$$

Define $\nabla(z) = \delta_z(\phi_2+1,\phi_1) - \delta_z(\phi_2,\phi_1)$. Then

$$\begin{aligned}\nabla(z) &= \frac{\Gamma(\phi_1+\phi_2+1)}{\Gamma(\phi_2+1)\,\Gamma(\phi_1)}\,\frac{z^{\phi_2+1}\,(1-z)^{\phi_1}}{I_z(\phi_2+1,\phi_1)} - \frac{\Gamma(\phi_1+\phi_2)}{\Gamma(\phi_2)\,\Gamma(\phi_1)}\,\frac{z^{\phi_2}\,(1-z)^{\phi_1}}{I_z(\phi_2,\phi_1)} \\ &= \frac{\Gamma(\phi_1+\phi_2)}{\Gamma(\phi_2)\,\Gamma(\phi_1)}\,\frac{z^{\phi_2}\,(1-z)^{\phi_1}}{I_z(\phi_2,\phi_1)}\left[\frac{\phi_1+\phi_2}{\phi_2}\,\frac{I_z(\phi_2,\phi_1)}{I_z(\phi_2+1,\phi_1)}\,z - 1\right] \\ &= \delta_z(\phi_2,\phi_1)\left[\frac{(\nu_1+\nu_2)\,z}{\nu_2 - 2\delta_z(\phi_2,\phi_1)} - 1\right]\end{aligned}$$

Substituting this result into (A.34),

$$\begin{aligned}\mathrm{Var}[U_2^*] &= (\nu_2 - 2\delta)\left[(\nu_2+2-\nu_2) - 2\delta\left(\frac{(\nu_1+\nu_2)\,z}{\nu_2-2\delta} - 1\right)\right] \\ &= 2\left[(\nu_2-2\delta) - \delta[(\nu_1+\nu_2)z - (\nu_2-2\delta)]\right] \\ &= 2\left[(\nu_2-2\delta)+\psi\right],\end{aligned}$$

where $\delta = \delta_z(\phi_2,\phi_1)$ and $\phi = \phi_z(\phi_2,\phi_1)$. This completes the proof. □

And finally, we show that for $U_2 \sim \chi^2_{\nu_2}$, $E[U_2^*(s)]/E(U_2)$ is uniformly bounded between 0 and 1.

Lemma A.34. *Let $U_2 \sim \chi^2_{\nu_2}$. Then for all $s > 0$,*

$$0 \le \frac{E[U_2^*]}{E(U_2)} \le 1.$$

Proof. By Theorem A.4, $0 \le 2\delta \le \nu_2$ for all $s \ge 0$. Hence,

$$\frac{E[U_2^*]}{E(U_2)} = \frac{\nu_2 - 2\delta}{\nu_2} = 1 - 2\delta/\nu_2 \le 1$$

for all $s \ge 0$. However, it is also easy to see that $1 - 2\delta/\nu_2 \ge 0$ for all s. □

A.4 Covariance of U_1^* and U_2^*

Unconditionally, U_1 and U_2 would be independently distributed χ^2 random variables with parameters ν_1 and ν_2, respectively. However, when conditioned on the event

$W \leq s$, the random variables $U_1^*(s)$ and $U_2^*(s)$ are correlated. By definition,

$$\operatorname{cov}(U_1^*, U_2^*) = E[U_1^* U_2^*] - E[U_1^*]\, E[U_2^*],$$

where $E[U_1^*] = \nu_1 + 2\delta_z(\phi_2, \phi_1)$ and $E[U_2^*] = \nu_2 - 2\delta_z(\phi_2, \phi_1)$. We need to determine the quantity $E[U_1^* U_2^*]$.

Theorem A.35. *For any fixed $s > 0$,*

$$E(U_1^*(s)\, U_2^*(s)) = \nu_1\, \nu_2\, \frac{I_z(\phi_2 + 1, \phi_1 + 1)}{I_z(\phi_2, \phi_1)}. \tag{A.35}$$

Proof. By definition,

$$E[U_1^*(s)\, U_2^*(s)] = \int_0^\infty \int_0^{sx} x\, y\, g_{U_1^*(s), U_2^*(s)}(x, y)\, dy\, dx,$$

where $y \leq sx$ by the constraint. Substituting in the joint PDP density of U_1^* and U_2^* for any fixed $s > 0$,

$$= \int_0^\infty \int_0^{sx} \frac{x\, y\, g_{\nu_1}(x) g_{\nu_2}(y)}{I_z(\phi_2, \phi_1)}\, du_2\, du_1.$$

Since the Beta cdf in the denominator is constant in both x and y, we shall only be concerned with the numerator term. We have

$$\begin{aligned}\int_0^\infty \int_0^x x\, y\, g_{\nu_1}(x) g_{\nu_2}(y)\, dy\, dx &= \int_0^\infty x\, g_{\nu_1}(x) \left[\int_0^{sx} y\, g_{\nu_2}(y) dy\right] dx \\ &= \int_0^\infty x\, g_{\nu_1}(x) \left[\int_0^{sx} \nu_2\, g_{\nu_2+2}(y)\, dy\right] dx,\end{aligned}$$

by Lemma A.20. Since the integral in brackets evaluates to $\nu_2\, G_{\nu_2+2}(sx)$,

$$\begin{aligned} &= \nu_2 \int_0^\infty x\, g_{\nu_1}(x)\, G_{\nu_2+2}(sx)\, dx \\ &= \nu_1\, \nu_2 \int_0^\infty g_{\nu_1+2}(x)\, G_{\nu_2+2}(sx)\, dx \\ &= \nu_1\, \nu_2\, I_z(\phi_2 + 1, \phi_1 + 1).\end{aligned}$$

Therefore,

$$E(U_1^* U_2^*) = \nu_1 \nu_2 \frac{I_z(\phi_2 + 1, \phi_1 + 1)}{I_z(\phi_2, \phi_1)},$$

thereby completing the proof. □

Theorem A.36. *For any fixed $s > 0$,* $\mathrm{cov}(U_1^*, U_2^*) = -2\psi_z(\phi_2, \phi_1)$.

Proof. By Theorem A.35 and the definition of covariance,

$$\mathrm{cov}(U_1^*, U_2^*) = \nu_1 \nu_2 \left[\frac{I_z(\phi_2 + 1, \phi_1 + 1)}{I_z(\phi_2, \phi_1)} - \frac{I_z(\phi_2, \phi_1 + 1)}{I_z(\phi_2, \phi_1)} \frac{I_z(\phi_2 + 1, \phi_1)}{I_z(\phi_2, \phi_1)} \right]. \quad \text{(A.36)}$$

To simplify this expression, the primary step is to simplify the ratio of Beta cdf's in (A.35). Starting from formula 26.5.10 of Abramowitz and Stegun (1964), we have

$$I_z(\phi_2 + 1, \phi_1 + 1) = z\, I_z(\phi_2, \phi_1 + 1) + (1 - z)\, I_z(\phi_2 + 1, \phi_1).$$

Dividing through this equation by $I_z(\phi_2, \phi_1)$,

$$\begin{aligned} \frac{I_z(\phi_2 + 1, \phi_1 + 1)}{I_z(\phi_2, \phi_1)} &= z \frac{I_z(\phi_2, \phi_1 + 1)}{I_z(\phi_2, \phi_1)} + (1 - z) \frac{I_z(\phi_2 + 1, \phi_1)}{I_z(\phi_2, \phi_1)} \\ &= z \left(1 + \frac{\delta}{\phi_1}\right) + (1 - z) \left(1 - \frac{\delta}{\phi_2}\right) \end{aligned}$$

by Theorems A.2 and A.3, with $\delta = \delta_z(\phi_2, \phi_1)$. Multiplying both sides of the equation by $\nu_1 \nu_2$ then yields

$$\nu_1 \nu_2 \frac{I_z(\phi_2 + 1, \phi_1 + 1)}{I_z(\phi_2, \phi_1)} = \nu_2 z(\nu_1 + 2\delta) + \nu_1(1 - z)(\nu_2 - 2\delta).$$

Substituting this term into the right hand side of (A.36), we obtain

$$\begin{aligned} \mathrm{cov}(U_1^*, U_2^*) &= \nu_2 z(\nu_1 + 2\delta) + \nu_1(1 - z)(\nu_2 - 2\delta) - (\nu_1 + 2\delta)(\nu_2 - 2\delta) \\ &= \nu_1 \nu_2(z + (1 - z) - 1) + 2\delta \left[\nu_2 z - \nu_1(1 - z) + \nu_1 - \nu_2 + 2\delta\right] \\ &= 2\delta \left[z\nu_1 - (1 - z)\nu_2 + 2\delta\right] \\ &= 2\delta \left[z\,(\nu_1 + 2\delta) - (1 - z)\,(\nu_2 - 2\delta)\right] \\ &= -2\,\psi, \end{aligned}$$

where $\psi = \psi_z(\phi_2, \phi_1)$ is given by (2.16). This completes the proof. □

A.5 Moments of W^*

The theorem below provides a simple means of arriving at the first two moments of $W^*(s)$ for each fixed s:

Theorem A.37. *Given $s > 0$, the first two moments of $W^*(s)$ are*

$$E(W^*) = \frac{\nu_2}{\nu_1 - 2} \frac{I_z(\phi_2 + 1, \phi_1 - 1)}{I_z(\phi_2, \phi_1)}, \tag{A.37}$$

$$E(W^{*2}) = \frac{\nu_2 (\nu_2 + 2)}{(\nu_1 - 2)(\nu_1 - 4)} \frac{I_z(\phi_2 + 2, \phi_1 - 2)}{I_z(\phi_2, \phi_1)}, \tag{A.38}$$

where $E(W^*)$ exists if $\nu_1 > 2$ and $E(W^{*2})$ exists if $\nu_1 > 4$.

Proof. Given $s > 0$, the expectation of $W^*(s)$ is

$$\begin{aligned} E(W^*) &= \int_0^s \frac{w g_W(w)}{I_z(\phi_2, \phi_1)} \, dw \\ &= \frac{\Gamma(\phi_1 + \phi_2)}{\Gamma(\phi_1)\,\Gamma(\phi_2)} \frac{1}{I_z(\phi_2, \phi_1)} \int_0^s \frac{w^{(\phi_2+1)-1}}{(1+w)^{\phi_1+\phi_2}} \, dw. \end{aligned}$$

Let $v = w/(1 + w)$; this transforms the Beta type II density into a Beta type I, and transforms the upper integration limit from s to z. Then

$$\begin{aligned} E(W^*) &= \frac{\Gamma(\phi_1 + \phi_2)}{\Gamma(\phi_1)\,\Gamma(\phi_2)} \frac{1}{I_z(\phi_2, \phi_1)} \int_0^z v^{(\phi_2+1)-1} (1 - v)^{(\phi_1-1)-1} \, dv \\ &= \frac{\Gamma(\phi_1 + \phi_2)}{\Gamma(\phi_1)\,\Gamma(\phi_2)} \frac{\Gamma(\phi_2 + 1)\Gamma(\phi_1 - 1)}{\Gamma(\phi_1 + \phi_2)} \frac{I_z(\phi_2 + 1, \phi_1 - 1)}{I_z(\phi_2, \phi_1)} \\ &= \frac{\phi_2}{\phi_1 - 1} \frac{I_z(\phi_2 + 1, \phi_1 - 1)}{I_z(\phi_2, \phi_1)} \\ &= \frac{\nu_2}{\nu_1 - 2} \frac{I_z(\phi_2 + 1, \phi_1 - 1)}{I_z(\phi_2, \phi_1)}, \end{aligned}$$

which exists whenever $\nu_1 > 2$.

By a similar argument, the second moment of W^* is given by

$$\begin{aligned}
E(W^{*2}) &= \int_0^s \frac{w^2 g_W(w)}{I_z(\phi_2, \phi_1)}\, dw \\
&= \frac{\Gamma(\phi_1+\phi_2)}{\Gamma(\phi_1)\,\Gamma(\phi_2)} \frac{\Gamma(\phi_2+2)\Gamma(\phi_1-2)}{\Gamma(\phi_1+\phi_2)} \frac{I_z(\phi_2+2, \phi_1-2)}{I_z(\phi_2, \phi_1)} \\
&= \frac{\phi_2(\phi_2+1)}{(\phi_1-1)(\phi_1-2)} \frac{I_z(\phi_2+2, \phi_1-2)}{I_z(\phi_2, \phi_1)} \\
&= \frac{\nu_2(\nu_2+2)}{(\nu_1-2)(\nu_1-4)} \frac{I_z(\phi_2+2, \phi_1-2)}{I_z(\phi_2, \phi_1)},
\end{aligned}$$

which exists whenever $\nu_1 > 4$. This completes the proof. □

The next result is a general expression for the k^{th} moment of $W^*(s)$.

Theorem A.38. *Given $s > 0$, the k^{th} moment of $W^*(s)$ is given by*

$$E(W^{*k}) = \frac{\Gamma(\phi_2+k)\,\Gamma(\phi_1-k)}{\Gamma(\phi_2)\,\Gamma(\phi_1)} \frac{I_z(\phi_2+k, \phi_1-k)}{I_z(\phi_2, \phi_1)}, \tag{A.39}$$

for $\nu_1 > 2k$.

The proof follows by induction.

A.6 Risk functions of point estimators

Using the entropy loss function (2.75) as a basis for comparing point estimators, we consider how to compute risk functions for estimators of τ_1, τ_2 and σ. In some cases, a closed form expression can be derived.

A.6.1 Unbiased estimators

For $i = 1, 2$, the risk of the best unbiased estimators T_i/ν_i are given by

$$R\left(\frac{T_i}{\nu_i};\sigma\right) = E_\sigma\left[\frac{T_i}{\nu_i\tau_i} - 1 - \log\frac{T_i}{\tau_i} + \log\nu_i\right]$$
$$= E_\sigma\left[\frac{U_i}{\nu_i} - 1 - \log U_i + \log\nu_i\right]$$
$$= 1 - 1 - E_\sigma(\log U_i) + \log\nu_i.$$

To find $E(\log U_i)$, we need the following result.

Lemma A.39. *Let U be a random variable with a χ^2_ν distribution, and define $W = \log U$. Then, $E(W) = \log 2 + \Psi(\phi)$, where $\phi = \nu/2$ and $\Psi(\cdot)$ is the digamma function.*

Proof. By hypothesis, $U \sim \chi^2_\nu$. Let $T = U/2$; then T has a standard gamma distribution with parameter ν. Make the transformation $x = \log t$; then, $t = e^x$ and $dt = e^x\,dx$, so that the density of X is

$$g_X(x) = \frac{\exp(\phi x - e^x)}{\Gamma(\phi)},$$

where $\phi = \nu/2$. X has a log-gamma distribution, whose moment generating function is given by (Lawless, 1982, p. 21)

$$M_X(t) = \frac{\Gamma(\phi + t)}{\Gamma(\phi)}.$$

Therefore, $E(X) = \Psi(\phi)$, where $\Psi(\cdot)$ is the digamma function. Setting $U = 2T$,

$$E(\log U) = E(\log 2T) = E(\log 2) + E(X) = \log 2 + \Psi(\phi).$$

□

Using the above lemma, the risk of the best unbiased estimator of τ_i is

$$R(\tilde{\tau}_i;\tau_i) = \log\phi_i - \Psi(\phi_i), \tag{A.40}$$

which is constant in σ for $i = 1, 2$.

The risk function of the best unbiased estimator of σ is given by

$$R\left(\frac{\nu_1 - 2}{\nu_2} S; \sigma\right) = E_\sigma\left[\frac{\nu_1 - 2}{\nu_2}\frac{S}{\sigma} - 1 - \log\frac{\nu_1 - 2}{\nu_2} - \log\frac{S}{\sigma}\right]$$

$$= -\log\frac{\nu_1 - 2}{\nu_2} - E_\sigma(\log W).$$

Since $W = U_2/U_1$, $\log W = \log U_2 - \log U_1$. By Lemma A.39, it follows that $E_\sigma(\log W) = \Psi(\phi_2) - \Psi(\phi_1)$. Therefore,

$$R\left(\frac{\nu_1 - 2}{\nu_2} S; \sigma\right) = E_\sigma(\log U_1) - E_\sigma(\log U_2) - \log\frac{\nu_1 - 2}{\nu_2}$$

$$= \Psi(\phi_1) - \Psi(\phi_2) - \log(\nu_1 - 2) + \log \nu_2. \tag{A.41}$$

A.6.2 Risk functions of REML estimators

In this case, the general form of the risk function under entropy loss associated with estimators of the form $\phi(S)\,T$, where T/ω has a χ^2 distribution, is given by

$$R(\phi(S)\,T; \sigma) = E_\sigma\left[E_\sigma\left[\phi(S)\frac{T}{\omega} - 1 - \log\phi(S) - \log\frac{T}{\omega}\Big|S\right]\right]$$

$$= E_\sigma\left[\phi(S)\,E\left(\frac{T}{\omega}\Big|S\right) - 1 - E_\sigma(\log\phi(S)) - E_\sigma\left(\log\frac{T}{\omega}\Big|S\right)\right]. \tag{A.42}$$

This can be viewed as a template for finding the risk functions of REML and PDP estimators.

REML estimator of τ_1. The REML estimator of τ_1 is of the form $\phi_1^*(S)\,T_1$, where

$$\phi_1^*(S) = \begin{cases} \dfrac{1+S}{\nu_1 + \nu_2} & \text{if } S < \dfrac{\nu_2}{\nu_1}, \\[2ex] \dfrac{1}{\nu_1} & \text{if } S \geq \dfrac{\nu_2}{\nu_1}. \end{cases}$$

The risk function of this estimator is then

$$R(\phi_1^*(S)\,T;\sigma) = E_\sigma[\phi_1^*(S)\,E(U_1|S) - 1 - \log\phi_1^*(S) - E_\sigma(\log U_1|S)].$$

To find $E(\log U_1|S)$, observe that $U_1 + U_2 = (1 + S/\sigma)\,U_1$; hence,

$$\log(U_1 + U_2) = \log(1 + S/\sigma) + \log U_1.$$

Conditioning this expression on S and taking expectations yields

$$E_\sigma[\log(U_1 + U_2)|S] = \log(1 + S/\sigma) + E_\sigma(\log U_1|S).$$

Since $U_1 + U_2$ is distributed independently of S, a slight rearrangement produces

$$E_\sigma(\log U_1|S) = E_\sigma[\log(U_1 + U_2)] - \log(1 + S/\sigma);$$

but by the above lemma, $E_\sigma[\log(U_1 + U_2)] = \log 2 + \Psi(\phi_1 + \phi_2)$, so

$$E_\sigma(\log U_1|S) = \log 2 + \Psi(\phi_1 + \phi_2) - \log(1 + S/\sigma). \tag{A.43}$$

Next, let $m = \nu_2/\nu_1$, and consider the term

$$E_\sigma[\phi_1^*(S)\,E(U_1|S)] = \int_0^m \frac{1+s}{\nu_1+\nu_2}\,\frac{\nu_1+\nu_2}{1+s/\sigma}\,g_S(s)\,ds + \frac{1}{\nu_1}\int_m^\infty g_S(s)\,ds,$$

where $g_S(s)$ is the density function of S with parameters ϕ_2 and ϕ_1 given by

$$g_S(s) = \frac{\Gamma(\phi_1+\phi_2)}{\Gamma(\phi_1)\,\Gamma(\phi_2)}\,\frac{1}{\sigma}\,\frac{(s/\sigma)^{\phi_2-1}}{(1+s/\sigma)^{\phi_1+\phi_2}}. \tag{A.44}$$

By making the transformation

$$V = \frac{S/\sigma}{1+S/\sigma} \qquad \text{with back transform} \qquad S = \frac{\sigma V}{1-V}, \tag{A.45}$$

the density of V is Beta type I with the same parameters as S. In the transformation, the term m gets transformed to $t(\sigma)$, where

$$t(\sigma) = \frac{\nu_2/(\nu_1\sigma)}{1+\nu_2/(\nu_1\sigma)} = \frac{\nu_2}{\nu_1\,\sigma+\nu_2}; \tag{A.46}$$

then, after some straightforward algebra,

$$E_\sigma[\phi_1^*(S)\,E(U_1|S)] = \frac{\nu_1}{\nu_1+\nu_2}\,I_{t(\sigma)}(\phi_2+1,\phi_1) + \frac{\nu_2\,\sigma}{\nu_1+\nu_2}\,I_{t(\sigma)}(\phi_2,\phi_1+1) + \frac{1}{\nu_1}\,(1-I_{t(\sigma)}(\phi_2,\phi_1)). \tag{A.47}$$

The last term that needs to be calculated is

$$E_\sigma[\log\phi_1^*(S)] = \int_0^m \log(1+s)\,g_s(s)\,ds - \log(\nu_1+\nu_2)\int_0^m g_S(s)\,ds - \log\nu_1\int_m^\infty g_S(s)\,ds;$$

Under the transformation (A.45), this evaluates to

$$E_\sigma[\log\phi_1^*(S)] = \int_0^{t(\sigma)} \log\left(1+\frac{\sigma v}{1-v}\right)\,g_V(v)\,dv - \log(\nu_1+\nu_2)\,I_{t(\sigma)}(\phi_2,\phi_1) - \log\nu_1(1-I_{t(\sigma)}(\phi_2,\phi_1)). \tag{A.48}$$

Combining (A.43), (A.47) and (A.48), the risk function of the REML estimator of τ_1 is

$$\begin{aligned} R(\phi_1^*(S)\,T_1;\sigma) = {} & \frac{\nu_1}{\nu_1+\nu_2}\,I_{t(\sigma)}(\phi_2+1,\phi_1) + \frac{\nu_2\,\sigma}{\nu_1+\nu_2}\,I_{t(\sigma)}(\phi_2,\phi_1+1) + E_\sigma(\log(1-V)) \\ & - 1 - \log 2 - \Psi(\phi_1+\phi_2) - \int_0^{t(\sigma)} \log\left(1+\frac{\sigma v}{1-v}\right)\,g_V(v)\,dv \\ & + \log(\nu_1+\nu_2)\,I_{t(\sigma)}(\phi_2,\phi_1) + \left(\frac{1}{\nu_1}+\log\nu_1\right)\,(1-I_{t(\sigma)}(\phi_2,\phi_1)), \end{aligned} \tag{A.49}$$

where $1+S/\sigma$ maps to $(1-V)^{-1}$ under the transformation (A.45).

REML estimator of τ_2. The process in this case is very similar to that for τ_1 above. The differences are that:

1. the REML estimator of τ_2 is $\phi_2^*(S)\,T_2$, where

$$\phi_2^*(S) = \begin{cases} \dfrac{1+\sigma/S}{\nu_1+\nu_2} & S \le \dfrac{\nu_2}{\nu_1} \\[2ex] \dfrac{1}{\nu_2} & S > \dfrac{\nu_2}{\nu_1}; \end{cases}$$

2. $U_1 + U_2 = (1 + \sigma/S)\, U_2$.

Applying the same basic argument as above, the risk function of the REML estimator of τ_2 under entropy loss is

$$\begin{aligned} R(\phi_2^*(S)\,T_2;\sigma) = {} & \frac{\nu_2}{\nu_1+\nu_2}\, I_{t(\sigma)}(\phi_2+1,\phi_1) + \frac{1}{\sigma}\frac{\nu_1}{\nu_1+\nu_2}\, I_{t(\sigma)}(\phi_2,\phi_1+1) + E_\sigma(\log V) \\ & - 1 - \log 2 - \Psi(\phi_1+\phi_2) - \int_0^{t(\sigma)} \log\left(1+\frac{1-v}{\sigma v}\right) g_V(v)\, dv \\ & + \log(\nu_1+\nu_2)\, I_{t(\sigma)}(\phi_2,\phi_1) + \left(\frac{1}{\nu_2} + \log\nu_2\right)(1 - I_{t(\sigma)}(\phi_2,\phi_1)), \end{aligned} \tag{A.50}$$

where $1 + \sigma/S$ maps to V^{-1} under the transformation (A.45).

Adjusted unbiased estimator of σ. The 'REML-inspired' adjustment of the best unbiased estimator of σ is given by $\phi^*(S)\,S$, where

$$\phi^*(S) = \begin{cases} \dfrac{1}{S} & S \le \dfrac{\nu_2}{\nu_1 - 2} \\ \dfrac{\nu_1 - 2}{\nu_2} & S > \dfrac{\nu_2}{\nu_1 - 2}. \end{cases}$$

The risk function of this estimator under entropy loss is

$$R[\phi^*(S)\,S;\sigma] = E_\sigma[\phi^*(S)\,W - 1 - \log\phi^*(S) - \log W].$$

Now,

$$\begin{aligned} E_\sigma(\phi^*(S)\,W) &= \frac{1}{\sigma}\int_0^m g_S(s)\,ds + \frac{\nu_1-2}{\nu_2}\int_m^\infty \frac{s}{\sigma}\, g_S(s)\, ds \\ &= \frac{1}{\sigma} I_p(\phi_2,\phi_1) + \frac{\nu_1 - 2}{\nu_1+\nu_2}(1 - I_p(\phi_2+1,\phi_1)), \end{aligned}$$

where $m = \nu_2/(\nu_1 - 2)$ and

$$p = \frac{\nu_2/(\nu_1-2)}{1+\nu_2/(\nu_1-2)} = \frac{\nu_2}{\nu_1+\nu_2-2}.$$

Furthermore,

$$E(\log \phi^*(S)) = -\int_0^m \log s\, g_S(s)\, ds + \log\left(\frac{\nu_1 - 2}{\nu_2}\right)(1 - I_p(\phi_2, \phi_1)).$$

Since $E(\log W) = \Psi(\phi_2) - \Psi(\phi_1)$, it follows that

$$\begin{aligned} R[\phi^*(S)\, S; \sigma] = {} & \frac{1}{\sigma} I_p(\phi_2, \phi_1) + \frac{\nu_1 - 2}{\nu_1 + \nu_2}(1 - I_p(\phi_2 + 1, \phi_1)) - 1 - \Psi(\phi_2) + \Psi(\phi_1) \\ & + \int_0^p \log\left(\frac{\sigma v}{1 - v}\right) g_V(v)\, dv + \log\left(\frac{\nu_2}{\nu_1 - 2}\right)(1 - I_p(\phi_2, \phi_1)). \end{aligned} \tag{A.51}$$

A.6.3 Risk functions of PDP estimators

PDP estimators do not have a simple form; PDP-unbiased estimators have multiplier functions that involve a ratio of Beta cdf's, while the multiplier function of a PDP-MLE is expressible as a spline. As a result, the risk function of a PDP estimator essentially follows (A.42). It is convenient to express the multiplier function of a PDP estimator as $\phi^\circ(S)$ so that a single expression holds for both estimators.

PDP estimators of τ_1. Let $\phi_1^\circ(S)\, T_1$ be a PDP estimator of τ_1. By (A.42), the risk function is

$$\begin{aligned} R(\phi_1^\circ(S)\, T_1; \sigma) &= E_\sigma[E_\sigma[\phi_1^\circ(S)\, U_1 - 1 - \log \phi_1^\circ(S)\, U_1 | S]] \\ &= E_\sigma[\phi_1^\circ(S)\, E_\sigma(U_1|S)] - 1 - E_\sigma[\log \phi_1^\circ(S)] - E_\sigma[E_\sigma(\log U_1|S)]. \end{aligned}$$

Now,

$$\begin{aligned} E_\sigma[\phi_1^\circ(S)\, E_\sigma(U_1|S)] &= (\nu_1 + \nu_2)\int_0^\infty \frac{\phi_1^\circ(s)}{1 + s/\sigma} g_S(s)\, ds \\ &= \nu_1 \int_0^\infty \phi_1^\circ(s)\, g_S(s; \phi_2, \phi_1 + 1)\, ds. \end{aligned}$$

Making the transformation from S to V yields

$$E_\sigma[\phi_1^\circ(S)\, E_\sigma(U_1|S)] = \nu_1 \int_0^1 \phi_1^\circ\left(\frac{\sigma v}{1 - v}\right) g_V(v; \phi_2, \phi_1 + 1)\, dv.$$

The above transformation also takes $1 + S/\sigma$ into $(1 - V)^{-1}$; in combination with (A.43), the risk function of $\phi_1^\circ(S)\, T_1$ is expressible as

$$\begin{aligned} R(\phi_1^\circ(S)\, T_1; \sigma) = \nu_1 \int_0^1 \phi_1^\circ\left(\frac{\sigma\, v}{1-v}\right)\, g_V(v; \phi_2, \phi_1 + 1)\, dv - 1 - \log 2 \\ - \Psi(\phi_1 + \phi_2) - E_\sigma\Big(\log \phi_1^\circ\left(\frac{\sigma V}{1-V}\right)\Big) + E_\sigma(\log(1-V)). \end{aligned} \tag{A.52}$$

Substituting $\phi_1^+(S)$ for $\phi_1^\circ(S)$ in the above expression yields the risk function for the PDP-MLE of τ_1, while substitution of $\phi_1^{**}(S)$ yields the risk of the PDP-unbiased estimator.

PDP estimators of τ_2. The argument is quite similar to the τ_1 case above, except that $1 + \sigma/S$ substitutes for $1 + S/\sigma$. The transformation to V takes $1 + \sigma/S$ into V^{-1}, so the risk function for a PDP estimator $\phi_2^\circ(S)\, T_2$ of τ_2 is

$$\begin{aligned} R(\phi_2^\circ(S)\, T_2; \sigma) = \nu_2 \int_0^1 \phi_2^\circ\left(\frac{\sigma\, v}{1-v}\right)\, g_V(v; \phi_2 + 1, \phi_1)\, dv - 1 - \log 2 \\ - \Psi(\phi_1 + \phi_2) - E_\sigma\Big(\log \phi_2^\circ\left(\frac{\sigma V}{1-V}\right)\Big) + E_\sigma(\log V). \end{aligned} \tag{A.53}$$

Substitution of $\phi_2^+(S)$ for $\phi_2^\circ(S)$ in (A.53) yields the risk function of the PDP-MLE of τ_2, while substituting $\phi_2^{**}(S)$ yields the risk of the PDP-unbiased estimator.

PDP-unbiased estimator of σ. The PDP-unbiased estimator of σ is $\phi^{**}(S)\, S$, whose risk function is given by

$$\begin{aligned} R(\phi^{**}(S)\, S; \sigma) &= E_\sigma[\phi^{**}(S)\, W - 1 - \log \phi^{**}(S) - \log W] \\ &= \int_0^\infty \phi^{**}(s) \frac{s}{\sigma}\, g_S(s)\, ds - 1 - E_\sigma(\log \phi^{**}(S)) + \Psi(\phi_1) - \Psi(\phi_2). \end{aligned}$$

Making the transformation from S to V, this evaluates to

$$\begin{aligned} R(\phi^{**}(S)\, S; \sigma) = \frac{\nu_2}{\nu_1 + \nu_2} \int_0^1 \phi^{**}\left(\frac{\sigma\, v}{1-v}\right)\, g_V(v; \phi_2 + 1, \phi_1 - 1)\, dv - 1 \\ + \Psi(\phi_1) - \Psi(\phi_2) - E_\sigma\Big(\log \phi^{**}\left(\frac{\sigma V}{1-V}\right)\Big). \end{aligned} \tag{A.54}$$

A.7 Limiting behavior of PDP-MLEs

PDP-MLE of τ_1. A conditional likelihood function of τ_1 for each fixed $s > 0$ can be constructed by treating the PDP density of $U_1^*(s)$ as a function of τ_1 given s. By making the transformation $u_1 = t_1/\tau_1$ in the density formula (A.1), the maximum likelihood estimator based on the PDP log-likelihood, called the PDP-MLE of τ_1, is expressible as $\hat{\tau}_1(S) = \phi_1^+(S)\,T_1$, where $\phi_1^+(\cdot)$ is the reciprocal of the unique solution $\beta = \tilde{\beta}$ of the estimating equation

$$\nu_1 - \beta + 2\nu_2 \frac{g_{\nu_2+2}(s\beta)}{G_{\nu_2}(s\beta)} = 0,$$

evaluated pointwise in s on the nonnegative reals. The set $\{\tilde{\beta}^{-1}(s) : s \geq 0\}$ of pointwise solutions comprises the multiplier function $\phi_1^+(S)$.

The limiting behavior of the PDP-MLE of τ_1 is obtained as follows. Firstly, consider the limit as $s \to \infty$. Noting that $G(\cdot) \to 1$ and $g(\cdot) \to 0$ as $s \to \infty$, what remains is that $\tilde{\beta} \to \nu_1$ in the limit, from which one concludes that $\phi_1^+(S)T_1 \to T_1/\nu_1$ as s gets large. On the other hand, as $s \to 0$, we want to find

$$\lim_{s\to 0} \tilde{\beta}(s) = \nu_1 + 2\nu_2 \lim_{s\to 0} \frac{g_{\nu_2+2}(s\tilde{\beta})}{G_{\nu_2}(s\tilde{\beta})}.$$

By Lemma A.1, the numerator can be reexpressed inside the limit, so that

$$\begin{aligned} &= \nu_1 + \nu_2 \lim_{s\to 0} \frac{G_{\nu_2}(s\tilde{\beta}) - G_{\nu_2+2}(s\tilde{\beta})}{G_{\nu_2}(s\tilde{\beta})} \\ &= \nu_1 + \nu_2 \lim_{s\to 0}\left(1 - \frac{G_{\nu_2+2}(s\tilde{\beta})}{G_{\nu_2}(s\tilde{\beta})}\right) \end{aligned}$$

As $s \to 0$, the cdf ratio goes to $0/0$, so we apply L'Hôpital's rule:

$$\begin{aligned}\lim_{s\to 0}\frac{G_{\nu_2+2}(s\tilde{\beta})}{G_{\nu_2}(s\tilde{\beta})} &= \lim_{s\to 0}\frac{\frac{\partial}{\partial s}G_{\nu_2+2}(s\tilde{\beta})}{\frac{\partial}{\partial s}G_{\nu_2}(s\tilde{\beta})}\\ &= \frac{g_{\nu_2+2}(s\tilde{\beta})}{g_{\nu_2}(s\tilde{\beta})}\\ &= \frac{\Gamma(\phi_2)\,2^{\phi_2}}{\Gamma(\phi_2+1)\,2^{\phi_2+1}}\,s\tilde{\beta}\\ &= \frac{s\tilde{\beta}}{\nu_2}.\end{aligned}$$

Therefore, when s is close to zero, the estimating equation can be expressed as

$$\tilde{\beta} = \nu_1 + \nu_2 - s\tilde{\beta},$$

from which

$$\lim_{s\to 0}\tilde{\beta}(s) = \lim_{s\to 0}\frac{\nu_1+\nu_2}{1+s} = \nu_1+\nu_2.$$

Since $\phi_1^+(S) = \{\beta^{-1}(s) : s \geq 0\}$, it follows that $\phi_1^+(S)T_1 \to T_1/(\nu_1+\nu_2)$ as $s \to 0$.

PDP-MLE of τ_2. A conditional likelihood function of τ_2 for each fixed $s > 0$ can be constructed by treating the PDP density of $U_2^*(s)$ as a function of τ_2 given s by making the transformation $u_2 = t_2/\tau_2$ in the density formula (A.2). For any fixed $s > 0$, differentiating the logarithm of the PDP likelihood of τ_2 and setting the result equal to zero produces an estimating equation of the form

$$0 = \nu_2 - \hat{\lambda} + 2\nu_1\,\frac{g_{\nu_1+2}(\hat{\lambda}/s)}{1-G_{\nu_1}(\hat{\lambda}/s)}, \tag{A.55}$$

which is solved numerically for $\hat{\lambda}$. The PDP-MLE of τ_2 is then

$$\hat{\tau}_2(S) = \phi_2^+(S)\,T_2 = \phi_2^+(S)\,ST_1,$$

where $\phi_2^+(S)$ is the set of pointwise solutions $\{1/\hat{\lambda}(s) : s \geq 0\}$ to (A.55), expressed as a function of S.

To show the limiting case as $s \to 0$, observe that for fixed λ in (A.55), the ratio on the right hand side goes to $0/0$ as $s \to 0$, so once again L'Hôpital's rule is applied. With respect to the ratio,

$$\frac{\frac{\partial}{\partial s} g_{\nu_1+2}(\hat{\lambda}/s)}{\frac{\partial}{\partial s}[1 - G_{\nu_1}(\hat{\lambda}/s)]} = -\frac{\Gamma(\phi_1)\, 2^{\phi_1}}{\Gamma(\phi_1+1)\, 2^{\phi_1+1}} \frac{\frac{\partial}{\partial s}\{(\hat{\lambda}/s)^{\phi_1} \exp(-\hat{\lambda}/2s)\}}{(\hat{\lambda}/s)^{\phi_1-1} \exp(-\hat{\lambda}/2s)\,(-\hat{\lambda}/s^2)}.$$

The derivative in the numerator evaluates to

$$\phi_1(\hat{\lambda}/s)^{\phi_1-1} \exp(-\hat{\lambda}/2s)\,(-\hat{\lambda}/s^2) + (\hat{\lambda}/s)^{\phi_1} \exp(-\hat{\lambda}/2s)\,(\hat{\lambda}/2s^2),$$

which can be reduced to

$$(\hat{\lambda}/s)^{\phi_1-1} \exp(-\hat{\lambda}/2s)\,(\hat{\lambda}/s^2)\left[-\phi_1 + \frac{\hat{\lambda}}{2s}\right].$$

After some cancellation,

$$\frac{\frac{\partial}{\partial s} g_{\nu_1+2}(\hat{\lambda}/s)}{\frac{\partial}{\partial s}[1 - G_{\nu_1}(\hat{\lambda}/s)]} = \frac{1}{\nu_1}\left[-\phi_1 + \frac{\hat{\lambda}}{2s}\right] = -\frac{1}{2} + \frac{\hat{\lambda}}{2\nu_1 s}.$$

Thus, when s is close to zero,

$$\begin{aligned}\hat{\lambda} &= \nu_2 - 2\nu_1\left(-\frac{1}{2} + \frac{\hat{\lambda}}{2\nu_1 s}\right)\\ &= \nu_1 + \nu_2 - \frac{\hat{\lambda}}{s}.\end{aligned}$$

Consequently, $(1 + 1/s)\,\hat{\lambda} = \nu_1 + \nu_2$, and since $\hat{\lambda}$ is the reciprocal of the multiplier of the PDP-MLE for τ_2,

$$\begin{aligned}\hat{\tau}_2 &= \frac{T_2}{\hat{\lambda}}\\ &= \frac{(1+1/S)}{\nu_1+\nu_2} T_2\\ &= \frac{(1+1/S)\,S}{\nu_1+\nu_2} T_1\\ &= \frac{1+S}{\nu_1+\nu_2} T_1.\end{aligned}$$

Therefore, as $s \to 0$, $\phi_2^+(S)T_2$ converges to $T_1/(\nu_1+\nu_2)$.

As s gets large, the denominator term on the right hand side of (A.55) goes to one, while the numerator goes to zero in the limit. Therefore, $\lim_{s\to\infty}\hat{\lambda}=\nu_2$, which implies that $\phi_2^+(S)\,T_2 \to T_2/\nu_2$ as s gets large.

A.8 Moments and cdf of the distribution of V^*

Derivation of the expected value of V^*, associated with inference about $J\sigma_\alpha^2$, is given below, along with its limiting behavior. Following Weerahandi (1994), an expression for the cdf of V^* is also derived.

A.8.1 Expectation of V^*

Given the observed sums of squares t_1 and t_2, the random quantity

$$V^*(t_1,t_2)=\left(\frac{t_2}{U_2}-\frac{t_1}{U_1}\right)\Big|W\le s$$

is used as a basis for inference about $J\sigma_\alpha^2$. For each (t_1,t_2) pair, a confidence distribution of V^* is obtainable from the joint PDP density (2.5), from which interval estimates of various types can be computed. Over all (t_1,t_2) pairs, lower and upper confidence functions for estimating $J\sigma_\alpha^2$ can be constructed, based on the family of densities of V^*. Additionally, one can find the mean function $E(V^*(T_1,T_2))$ by evaluating $E(V^*)$ at each (t_1,t_2) pair. This is the present task.

Theorem A.40. *Given an arbitrary (t_1,t_2) pair, the expected value of V^* is*

$$E(V^*|t_1,t_2)=\frac{t_2}{\nu_2-2}\,\frac{I_z(\phi_2-1,\phi_1)}{I_z(\phi_2,\phi_1)}-\frac{t_1}{\nu_1-2}\,\frac{I_z(\phi_2,\phi_1-1)}{I_z(\phi_2,\phi_1)},$$

where $\nu_1>\nu_2>2$.

Proof. Starting from the transformed joint density of (V_1, V_2) given by (2.30) in section 2.4.7, make the bivariate transformation

$$W_1 = \frac{t_1}{V_1} \qquad W_2 = \frac{t_2}{V_1 + V_2}, \tag{A.56}$$

whose inverse transform is

$$V_1 = \frac{t_1}{W_1} \qquad V_2 = \frac{t_2}{W_2} - \frac{t_1}{W_1}, \tag{A.57}$$

with Jacobian $t_1 t_2/(W_1 W_2)^2$. Since $V_2 > 0$,

$$\frac{t_2}{W_2} - \frac{t_1}{W_1} \geq 0 \quad \Longrightarrow \quad W_2 \leq sW_1,$$

so the transformation (A.56) affects the integration limits. Then,

$$E(V^*|t_1, t_2) = \frac{1}{t_1 t_2 K} \int_0^\infty \int_0^{sw_1} \left(\frac{t_2}{w_2} - \frac{t_1}{w_1}\right) w_1^{\phi_1+1} w_2^{\phi_2+1} \left(\frac{t_1 t_2}{w_1^2 w_2^2}\right) e^{-(w_1+w_2)}\, dw_2\, dw_1,$$

where $K = \Gamma(\phi_1)\Gamma(\phi_2)2^{\phi_1+\phi_2}\, I_z(\phi_2, \phi_1)$. Expanding terms,

$$\begin{aligned}
&= K^{-1}\left[\int_0^\infty w_1^{\phi_1-1} e^{-w_1/2}\left[t_2 \int_0^{sw_1} w_2^{(\phi_2-1)-1} e^{-w_2/2}\, dw_2\right] dw_1\right. \\
&\qquad \left. - t_1 \int_0^\infty w_1^{(\phi_1-1)-1} e^{-w_1/2}\left[\int_0^{sw_1} w_2^{\phi_2-1} e^{-w_2/2}\, dw_2\right] dw_1\right] \\
&= K^{-1}\left[t_2\, 2^{\phi_2-1}\Gamma(\phi_2-1) \int_0^\infty w_1^{\phi_1-1} e^{-w_1/2}\, G_{\nu_2-2}(sw_1)\, dw_1\right. \\
&\qquad \left. - t_1 2^{\phi_2}\Gamma(\phi_2) \int_0^\infty w_1^{(\phi_1-1)-1} e^{-w_1/2}\, G_{\nu_2}(sw_1)\, dw_1\right] \\
&= K_2^{-1}\left[t_2\Gamma(\phi_2-1)\Gamma(\phi_1) I_z(\phi_2-1, \phi_1) - t_1\Gamma(\phi_2)\Gamma(\phi_1-1) I_z(\phi_2, \phi_1-1)\right],
\end{aligned}$$

by Lemma A.24, where $K_2 = K/2^{\phi_1+\phi_2-1}$. Expanding the constant K_2,

$$= \frac{1}{2}\,\frac{t_2\Gamma(\phi_2-1)\Gamma(\phi_1) I_z(\phi_2-1, \phi_1) - t_1\Gamma(\phi_2)\Gamma(\phi_1-1) I_z(\phi_2, \phi_1-1)}{\Gamma(\phi_1)\Gamma(\phi_2) I_z(\phi_2, \phi_1)},$$

so that

$$\begin{aligned}
E(V^*|t_1, t_2) &= \frac{t_2}{2}\left[\frac{\Gamma(\phi_2-1)}{\Gamma(\phi_2)}\,\frac{I_z(\phi_2-1, \phi_1)}{I_z(\phi_2, \phi_1)} - \frac{1}{s}\,\frac{\Gamma(\phi_1-1)}{\Gamma(\phi_1)}\,\frac{I_z(\phi_2, \phi_1-1)}{I_z(\phi_2, \phi_1)}\right] \\
&= \frac{t_2}{\nu_2-2}\,\frac{I_z(\phi_2-1, \phi_1)}{I_z(\phi_2, \phi_1)} - \frac{t_1}{\nu_1-2}\,\frac{I_z(\phi_2, \phi_1-1)}{I_z(\phi_2, \phi_1)}.
\end{aligned}$$

□

A.8.2 Limiting values of $E(V^*)$

Since V^* is a function of both t_1 and t_2, the limiting value of the mean $E(V^*)$ depends on how t_1 and t_2 behave in the limit. There are four cases to consider.

Case 1: t_2 fixed, $s \to \infty$.
In this situation, we are concerned about the term in brackets in (2.32). As $s \to \infty$ is equivalent to $z \to 1$, we find that the Beta cdf's all go to one in the limit and $1/s$ goes to zero, so that

$$\lim_{s\to\infty} E(V^*) = \frac{t_2}{\nu_2 - 2}.$$

Case 2: t_1 fixed, $s \to \infty$.
If t_1 is fixed and $s \to \infty$, it follows that $t_2 \to \infty$, which means that by (2.32), $E(V^*) \to \infty$.

In other words, as $s \to \infty$, it matters which of t_1 or t_2 is fixed if we want to ensure that the limiting mean is finite as s gets large. The same remarks apply as $s \to 0$, as the following pair of cases show.

Case 3: t_1 fixed, $s \to 0$.

$$\lim_{s\to 0} E(V^*) = \frac{t_1}{2} \lim_{s\to 0} \left[\frac{s\, I_z(\phi_2 - 1, \phi_1)}{(\phi_2 - 1)\, I_z(\phi_2, \phi_1)} - \frac{I_z(\phi_2, \phi_1 - 1)}{(\phi_1 - 1)\, I_z(\phi_2, \phi_1)} \right]$$

We need to be concerned with the limit of both terms within brackets. Focusing on the latter term first, we have

$$\begin{aligned}
(\phi_1 - 1)\frac{I_z(\phi_2, \phi_1)}{I_z(\phi_2, \phi_1 - 1)} &= (\phi_1 - 1)\left(1 + \frac{\delta_z(\phi_2, \phi_1 - 1)}{\phi_1 - 1}\right) \\
&= (\phi_1 - 1) + \delta_z(\phi_2, \phi_1 - 1) \\
&\to \phi_1 + \phi_2 - 1 \quad \text{as } s \to 0.
\end{aligned}$$

With regard to the first term in brackets, make the substitution $s = z/(1 - z)$, so that we have the following result.

Lemma A.41.

$$\lim_{z\to 0} \frac{z\, I_z(\phi_2-1,\phi_1)}{(\phi_2-1)(1-z)\, I_z(\phi_2,\phi_1)} = \frac{2\nu_2}{(\nu_2-2)(\nu_1+\nu_2-2)}.$$

Proof. By Fact A.4,

$$\begin{aligned}\frac{z\, I_z(\phi_2-1,\phi_1)}{(\phi_2-1)(1-z)\, I_z(\phi_2,\phi_1)} &= \frac{I_z(\phi_2,\phi_1)-(1-z)\, I_z(\phi_2,\phi_1-1)}{(\phi_2-1)(1-z)\, I_z(\phi_2,\phi_1)} \\ &= \frac{1}{\phi_2-1}\left[\frac{1}{1-z} - \frac{I_z(\phi_2,\phi_1-1)}{I_z(\phi_2,\phi_1)}\right].\end{aligned}$$

By Theorem A.2,

$$\begin{aligned} &= \frac{1}{\phi_2-1}\left[\frac{1}{1-z} - \left(1+\frac{\delta_z(\phi_2,\phi_1-1)}{\phi_1-1}\right)^{-1}\right] \\ &= \frac{1}{\phi_2-1}\left[\frac{1}{1-z} - \frac{\phi_1-1}{(\phi_1-1)+\delta_z(\phi_2,\phi_1-1)}\right] \\ &\to \frac{1}{\phi_2-1}\left[1-\frac{\nu_1-2}{\nu_1+\nu_2-2}\right] \quad \text{as } z\to 0 \\ &= \frac{2\,\nu_2}{(\nu_2-2)(\nu_1+\nu_2-2)}.\end{aligned}$$

□

Therefore, the limiting value of $E(V^*)$ is

$$\begin{aligned}\lim_{s\to 0} E(V^*) &= \frac{t_1}{2}\left[\frac{\nu_2}{(\phi_2-1)(\nu_1+\nu_2-2)} - \frac{1}{\phi_1+\phi_2-1}\right] \\ &= \frac{t_1}{\nu_1+\nu_2-2}\left[\frac{\nu_2}{\nu_2-2}-1\right] \\ &= \frac{2t_1}{(\nu_2-2)(\nu_1+\nu_2-2)}.\end{aligned}$$

Case 4: t_2 fixed, $s\to 0$.

The limiting mean of V^* can be written as

$$\begin{aligned}\lim_{s\to 0} E(V^*) &= \frac{t_2}{2}\lim_{s\to 0}\left[\frac{I_z(\phi_2-1,\phi_1)}{(\phi_2-1)I_z(\phi_2,\phi_1)} - \frac{1}{s}\,\frac{I_z(\phi_2,\phi_1-1)}{(\phi_1-1)\, I_z(\phi_2,\phi_1)}\right] \\ &= \frac{t_2}{2}\lim_{s\to 0}\frac{1}{s}\left[\frac{s\, I_z(\phi_2-1,\phi_1)}{(\phi_2-1)\, I_z(\phi_2,\phi_1)} - \frac{I_z(\phi_2,\phi_1-1)}{(\phi_1-1)\, I_z(\phi_2,\phi_1)}\right].\end{aligned}$$

From Case 3, we know that the term in brackets is finite, as is $t_2/2$. This leaves one with $\lim_{s\to 0} s^{-1} = \infty$, so the limiting value of $E(V^*)$ in this case is infinite.

Consequently, to ensure that $E(V^*)$ tends to a finite limit, we must have either t_1 fixed as $t_2 \to 0$, or t_2 fixed as $t_1 \to 0$.

A.8.3 cdf of V^*.

Following Weerahandi (1994), one can obtain an integral expression for the cdf of V^* for each (t_1, t_2) pair. Let $c > 0$ be a point in the domain of $V^*(t_1, t_2)$; then

$$\begin{aligned} P(V^* \le c) &= P\Big(\frac{t_2}{U_2} - \frac{t_1}{U_1} \le c \mid W \le s\Big) \\ &= P\Big(\frac{t_2}{U_2} \le \frac{t_1}{U_1} + c \mid W \le s\Big). \end{aligned}$$

Multiplying through by U_1,

$$\begin{aligned} &= P\Big(\frac{t_2}{W} \le t_1 + cU_1 \mid W \le s\Big) \\ &= P\Big(\frac{1}{W} \le \frac{1}{s} + \frac{cU_1}{t_2} \mid W \le s\Big). \end{aligned}$$

Multiplying through by $1 + W$,

$$= P\Big(\frac{1+W}{W} \le \frac{1+W}{s} + \frac{cU_1(1+W)}{t_2} \mid W \le s\Big).$$

Now, $U_1(1+W) = U_1 + U_2$, so that

$$= P\Big(\frac{1+W}{W} \le \frac{1+W}{s} + \frac{c(U_1+U_2)}{t_2} \mid W \le s\Big).$$

Rearranging terms to isolate $U_1 + U_2$,

$$= P\Big(U_1 + U_2 \ge \frac{t_2}{c}\Big(\frac{1+W}{W} - \frac{1+W}{s}\Big) \mid W \le s\Big).$$

Now, $V = W/(1+W) \sim \text{Beta}(\phi_2, \phi_1)$, so that $1 - V = (1+W)^{-1}$. Then, the conditioning event $W \leq s$ can be reexpressed as $V \leq z$, where $z = s/(1+s)$. Then,

$$\begin{aligned}
&= P\Big(U_1 + U_2 \geq \frac{t_2}{c}\left(\frac{1}{V} - \frac{1}{s(1-V)}\right) \mid V \leq z\Big) \\
&= P\Big(U_1 + U_2 \geq \frac{1}{c}\left(\frac{t_2}{V} - \frac{t_1}{1-V}\right) \mid V \leq z\Big) \\
&= 1 - P\Big(U_1 + U_2 \leq \frac{1}{c}\left(\frac{t_2}{V} - \frac{t_1}{1-V}\right) \mid V \leq z\Big) \\
&= 1 - \frac{\int_0^z G_{\nu_1+\nu_2}\left[\frac{1}{c}\left(\frac{t_2}{v} - \frac{t_1}{1-v}\right)\right] g_V(v)\, dv}{I_z(\phi_2, \phi_1)}.
\end{aligned} \tag{A.58}$$

To find the probability γ between two points ψ_1 and ψ_2, we solve

$$\begin{aligned}
\gamma &= P(\psi_1 \leq \frac{t_2}{U_2} - \frac{t_1}{U_1} \leq \psi_2 \mid W \leq s) \\
&= P\Big(\frac{\psi_1 U_1}{t_2} \leq \frac{1}{W} - \frac{1}{s} \leq \frac{\psi_2 U_1}{t_2} \mid W \leq s\Big).
\end{aligned}$$

Let $Y = W^{-1} - s^{-1}$; then,

$$\begin{aligned}
&= P\Big(\frac{\psi_1}{t_2 Y} \leq \frac{1}{U_1} \leq \frac{\psi_2}{t_2 Y} \mid W \leq s\Big) \\
&= P\Big(\frac{t_2 Y}{\psi_2} \leq U_1 \leq \frac{t_2 Y}{\psi_1} \mid W \leq s\Big).
\end{aligned}$$

Multiplying through by $1 + W$,

$$\begin{aligned}
&= P\Big(\frac{t_2 Y(1+W)}{\psi_2} \leq U_1(1+W) \leq \frac{t_2 Y(1+W)}{\psi_1} \mid W \leq s\Big) \\
&= P\Big(\frac{t_2}{\psi_2}\left(\frac{1+W}{W} - \frac{1+W}{s}\right) \leq U_1 + U_2 \leq \frac{t_2}{\psi_1}\left(\frac{1+W}{W} - \frac{1+W}{s}\right) \mid W \leq s\Big).
\end{aligned}$$

Letting $R = W/(1+W) \sim \text{Beta}(\phi_2, \phi_1)$ with observed value $z = s/(1+s)$,

$$\begin{aligned}
&= P\Big(\frac{1}{\psi_2}\left(\frac{t_2}{R} - \frac{t_1}{1-R}\right) \leq U_1 + U_2 \leq \frac{1}{\psi_1}\left(\frac{t_2}{R} - \frac{t_1}{1-R}\right) \mid R \leq z\Big) \\
&= \frac{\int_0^z \left[G_{\nu_1+\nu_2}\left(\frac{1}{\psi_1}\left(\frac{t_2}{r} - \frac{t_1}{1-r}\right)\right) - G_{\nu_1+\nu_2}\left(\frac{1}{\psi_2}\left(\frac{t_2}{r} - \frac{t_1}{1-r}\right)\right)\right] g_R(r)\, dr}{I_z(\phi_2, \phi_1)},
\end{aligned} \tag{A.59}$$

by the independence of R and $U_1 + U_2$.

Appendix B

Complements: three-stage nested model

This appendix contains derivations of some of the distribution theory results cited in Chapter 3.

B.1 Normalizing constant

The normalizing constant of the joint density is the denominator term in (3.3), which can be expressed as

$$P(W_1 \leq s_1, W_2 \leq s_2) = \int_0^\infty \int_0^{s_1 u_1} \int_0^{s_2 u_2} g_{\nu_1}(u_1)\, g_{\nu_2}(u_2)\, g_{\nu_3}(u_3)\, du_3\, du_2\, du_1, \qquad \text{(B.1)}$$

for each set of realized sums of squares (t_1, t_2, t_3).

This probability can be obtained in several different ways, depending on one's choice of transformations and the direction one chooses to integrate. The following regions are inferentially equivalent:

$$\{U_1 > 0, U_2 \leq s_1 U_1, U_3 \leq s_2 U_2\}$$

$$\{U_1 \geq U_2/s_1, U_2 > 0, U_3 \leq s_2 U_2\}$$

$$\{U_1 \geq U_2/s_1, U_2 \geq U_3/s_2, U_3 > 0\}.$$

Hence, if we let

$$t = g_{\nu_1}(u_1)\, g_{\nu_2}(u_2)\, g_{\nu_3}(u_3)\, du_3\, du_2\, du_1,$$

it follows that

$$\int_0^\infty \int_0^{s_1 u_1} \int_0^{s_2 u_2} t = \int_{u_2/s_1}^\infty \int_0^\infty \int_0^{s_2 u_2} t = \int_{u_2/s_1}^\infty \int_{u_3/s_2}^\infty \int_0^\infty t.$$

To this end, we will look at two derivations of this probability, as it becomes relevant in the determination of moments of marginal PDP densities. The direct method is to transform $(U_1, U_2, U_3) \mapsto (U_1, W_1, W_2)$ and integrate out U_1.

Theorem B.1. *Given* $s_1, s_2 > 0$,

$$P(W_1 \le s_1, W_2 \le s_2) = \int_0^{s_1} \int_0^{s_2} \frac{\Gamma(\phi_1 + \phi_2 + \phi_3)}{\Gamma(\phi_1)\Gamma(\phi_2)\Gamma(\phi_3)} \frac{v_2^{\phi_2+\phi_3-1}\, v_3^{\phi_3-1}}{(1 + v_2 + v_2 v_3)^{\phi_1+\phi_2+\phi_3}}\, dv_3\, dv_2. \tag{B.2}$$

Proof. Given $s_1, s_2 > 0$, begin with (B.1) and make the transformation

$$\begin{aligned} V_1 &= U_1 & & U_1 = V_1 \\ V_2 &= U_2/U_1 & \implies\quad & U_2 = V_1\, V_2 \\ V_3 &= U_3/U_2 & & U_3 = V_1\, V_2\, V_3 \end{aligned}$$

with Jacobian $V_1^2 V_2$, so that the integral in the transformed scale is

$$\int_0^\infty \int_0^{s_1} \int_0^{s_2} \frac{v_1^{\phi_1+\phi_2+\phi_3-1}\, v_2^{\phi_2+\phi_3-1}\, v_3^{\phi_3-1}\, \exp\{-v_1(1 + v_2 + v_2 v_3)/2\}\, dv_3\, dv_2\, dv_1}{\Gamma(\phi_1)\Gamma(\phi_2)\Gamma(\phi_3)\, 2^{\phi_1+\phi_2+\phi_3}}.$$

To integrate out V_1, exchange the order of integration, yielding

$$= \int_0^{s_1} \int_0^{s_2} \frac{v_2^{\phi_2+\phi_3-1}\, v_3^{\phi_3-1}}{\Gamma(\phi_1)\Gamma(\phi_2)\Gamma(\phi_3)} \int_0^\infty \frac{v_1^{\phi_1+\phi_2+\phi_3-1}\, \exp\{-v_1(1 + v_2 + v_2 v_3)/2\}}{2^{\phi_1+\phi_2+\phi_3}}\, dv_1\, dv_3\, dv_2.$$

Under the innermost integral, substitute $x = v_1(1 + v_2 + v_2 v_3)$, which implies that

$$v_1 = \frac{x}{1 + v_2 + v_2 v_3} \qquad \implies \qquad dv_1 = \frac{dx}{1 + v_2 + v_2 v_3}.$$

Then, after some rearrangement of terms,

$$= \int_0^{s_1}\int_0^{s_2} \frac{v_2^{\phi_2+\phi_3-1}\, v_3^{\phi_3-1}\,(1+v_2+v_2v_3)^{-(\phi_1+\phi_2+\phi_3)}}{\Gamma(\phi_1)\Gamma(\phi_2)\Gamma(\phi_3)} \int_0^{\infty} \frac{x^{\phi_1+\phi_2+\phi_3-1}\, e^{-x/2}}{2^{\phi_1+\phi_2+\phi_3}}\, dx\, dv_3\, dv_2.$$

The innermost integral evaluates to $\Gamma(\phi_1+\phi_2+\phi_3)$, so that

$$P(W_1 \le s_1, W_2 \le s_2) = \int_0^{s_1}\int_0^{s_2} \frac{\Gamma(\phi_1+\phi_2+\phi_3)}{\Gamma(\phi_1)\Gamma(\phi_2)\Gamma(\phi_3)} \frac{v_2^{\phi_2+\phi_3-1}\, v_3^{\phi_3-1}}{(1+v_2+v_2v_3)^{\phi_1+\phi_2+\phi_3}}\, dv_3\, dv_2.$$

□

The other approach to deriving the normalizing constant (B.1) is to recognize it as a Dirichlet cdf, generalizing the Beta cdf of Chapter 2. This is the preferred method, as higher moments can be found by following similar lines of proof.

Theorem B.2. *Given $s_1, s_2 > 0$, the cdf of a Dirichlet(ϕ_3, ϕ_2, ϕ_1) random variable is given by*

$$I_{z_1,z_2}(\phi_3,\phi_2,\phi_1) = \int_0^{z_1}\int_0^{z_2} \frac{\Gamma(\phi_1+\phi_2+\phi_3)}{\Gamma(\phi_1)\Gamma(\phi_2)\Gamma(\phi_3)}\, v_2^{\phi_2-1}\, v_3^{\phi_3-1}\,(1-v_2-v_3)^{\phi_1-1}\, dv_3\, dv_2. \tag{B.3}$$

Proof. Given $s_1, s_2 > 0$, begin with (B.1) and make the Dirichlet transformation

$$\begin{aligned} V_1 &= U_1+U_2+U_3 & & U_1 = V_1(1-V_2-V_3) \\ V_2 &= \frac{U_2}{U_1+U_2+U_3} & \Longrightarrow\quad & U_2 = V_1V_2 \\ V_3 &= \frac{U_3}{U_1+U_2+U_3} & & U_3 = V_1V_3, \end{aligned} \tag{B.4}$$

with Jacobian V_1^2. The upper integration limits are transformed from (∞, s_1u_1, s_2u_2) to (∞, z_1, z_2), where

$$z_1 = \frac{s_1}{1+s_1+s_1s_2}, \qquad z_2 = \frac{s_1s_2}{1+s_1+s_1s_2}; \tag{B.5}$$

hence, (B.1) becomes

$$\int_0^{\infty}\int_0^{z_2}\int_0^{z_1} \frac{v_1^{\phi_1+\phi_2+\phi_3-1}\, v_2^{\phi_2-1}\, v_3^{\phi_3-1}\,(1-v_2-v_3)^{\phi_1-1}\, e^{-v_1/2}}{\Gamma(\phi_1)\Gamma(\phi_2)\Gamma(\phi_3)\, 2^{\phi_1+\phi_2+\phi_3}}\, dv_3\, dv_2\, dv_1,$$

in the transformed scale. To eliminate V_1, exchange the order of integration to obtain

$$= \int_0^{z_2}\int_0^{z_1} \frac{v_2^{\phi_2-1}\, v_3^{\phi_3-1}\,(1-v_2-v_3)^{\phi_1-1}}{\Gamma(\phi_1)\Gamma(\phi_2)\Gamma(\phi_3)} \left[\int_0^{\infty} \frac{v_1^{\phi_1+\phi_2+\phi_3-1}\, e^{-v_1/2}}{2^{\phi_1+\phi_2+\phi_3}}\, dv_1\right] dv_3\, dv_2$$
$$= \int_0^{z_2}\int_0^{z_1} \frac{\Gamma(\phi_1+\phi_2+\phi_3)}{\Gamma(\phi_1)\Gamma(\phi_2)\Gamma(\phi_3)}\, v_3^{\phi_3-1}\, v_2^{\phi_2-1}\,(1-v_2-v_3)^{\phi_1-1}\, dv_3\, dv_2.$$

The integrand is a Dirichlet(ϕ_3, ϕ_2, ϕ_1) density. The notation

$$I_{z_1,z_2}(\phi_3,\phi_2,\phi_1) = \int_0^{z_1}\int_0^{z_2} \frac{\Gamma(\phi_1+\phi_2+\phi_3)}{\Gamma(\phi_1)\Gamma(\phi_2)\Gamma(\phi_3)}\, v_2^{\phi_2-1}\, v_3^{\phi_3-1}\,(1-v_2-v_3)^{\phi_1-1}\, dv_3\, dv_2$$

signifies that the subscript represents the upper limits of v_2 and v_3, in that order, and the parenthesized term represents the parameter values associated with v_3, v_2, and $1-v_2-v_3$, in that order. This is the form of the normalizing constant adopted in Chapter 3, given by (3.6). □

B.2 Preservation of total df: general results

In the balanced one-way random model, it was shown that the sum of the conditional means preserved the total degrees of freedom; i.e.,

$$E(U_1^*) + E(U_2^*) = \nu_1 + \nu_2 = E(U_1) + E(U_2),$$

where $U_i^* = U_i | U_2 \le sU_1$, $i = 1, 2$. We begin this section by showing that the above result holds more generally.

Henceforth through this Appendix, we will use $C(s_1, s_2)$ to denote the conditioning event; i.e., $C(s_1, s_2) = \{U_2 \le s_1 U_1, U_3 \le s_2 U_2\}$. In general, the argument will be dropped and C will be used to denote the event.

Before presenting a more general argument, we first show that the total degrees of freedom are preserved by PDP conditioning in the balanced three-stage nested random effects model.

Lemma B.3. *Under the three-stage nested model* (3.1),

$$E(U_1 + U_2 + U_3|C) = \nu_1 + \nu_2 + \nu_3$$

for any $s_1, s_2 > 0$.

Proof. Given $s_1, s_2 > 0$,

$$E(U_1 + U_2 + U_3|C) = \int_0^\infty \int_0^{s_1 u_1} \int_0^{s_2 u_2} (u_1 + u_2 + u_3) \frac{g_{\nu_1}(u_1)\, g_{\nu_2}(u_2)\, g_{\nu_3}(u_3)}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)}\, du_3\, du_2\, du_1.$$

Make the standard Dirichlet transformation (B.4) with Jacobian v_1^2. In the transformed scale, the integral becomes

$$\begin{aligned}
&= \int_0^\infty \int_0^{z_1} \int_0^{z_2} v_1 \frac{[v_1(1 - v_2 - v_3)]^{\phi_1 - 1} (v_1 v_2)^{\phi_2 - 1} (v_1 v_3)^{\phi_3 - 1}\, v_1^2\, e^{-v_1/2}}{\Gamma(\phi_1)\, \Gamma(\phi_2)\, \Gamma(\phi_3)\, 2^{\phi_T}\, I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)}\, dv_3\, dv_2\, dv_1 \\
&= \int_0^{z_1} \int_0^{z_2} \frac{(1 - v_2 - v_3)^{\phi_1 - 1}\, v_2^{\phi_2 - 1}\, v_3^{\phi_3 - 1}}{\Gamma(\phi_1)\, \Gamma(\phi_2)\, \Gamma(\phi_3)\, I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)} \left[\int_0^\infty \frac{v_1^{\phi_T + 1 - 1}\, e^{-v_1/2}}{2^{\phi_T}}\, dv_1\right] dv_3\, dv_2 \\
&= 2\, \frac{\Gamma(\phi_T + 1)}{\Gamma(\phi_T)} \int_0^{z_1} \int_0^{z_2} \frac{\Gamma(\phi_T)}{\Gamma(\phi_1)\, \Gamma(\phi_2)\, \Gamma(\phi_3)} \frac{v_3^{\phi_3 - 1}\, v_2^{\phi_2 - 1}\, (1 - v_2 - v_3)^{\phi_1 - 1}}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)}\, dv_3\, dv_2 \\
&= \nu_1 + \nu_2 + \nu_3,
\end{aligned}$$

where $\phi_T = \phi_1 + \phi_2 + \phi_3$. This establishes the result. □

To generalize this property to any random effects model utilizing PDP conditioning, we start with the following well-known distributional result, stated without proof:

Theorem B.4. *Let* $U_1, \ldots, U_k$ *denote independently distributed* $\chi^2_{\nu_i}$ *random variables,* $i = 1, 2, \ldots, k$. *Under the general Dirichlet transformation*

$$V_1 = \sum U_i, \quad V_j = U_j / V_1, \qquad j = 2, 3, \ldots, k,$$

$V_1 \sim \chi^2_{\nu_T}$ *(where* $\nu_T = \sum \nu_i$*) is distributed independently of the joint distribution of* $(V_2, V_3, \ldots, V_k)$*, which follows a Dirichlet distribution with parameter vector* $(\phi_k, \ldots, \phi_2)$.

Consequently, the distribution of V_1 conditional on $V_2, \ldots, V_k$ is still $\chi^2_{\nu_T}$, i. e., it is unaffected by conditioning on $V_2, \ldots, V_k$, or any 1–1 function of $V_2, \ldots, V_k$. In particular, the moments of V_1 are unaffected by conditioning events involving (functions of) $V_2, \ldots, V_k$, so the total degrees of freedom under a balanced random effects ANOVA model involving PDP conditioning are not affected by such conditioning.

B.3 Distributions of the W_i^*

In this section, marginal densities of the $W_j^* = W_j|C$ $(j = 1, 2, 3)$ are derived, where $W_1 = U_2/U_1$, $W_2 = U_3/U_2$ and

$$W_3 = \frac{U_3}{U_1} = \frac{U_3}{U_2}\frac{U_2}{U_1} = W_1 W_2.$$

In the following, a pair (s_1, s_2), $s_1, s_2 > 0$ is given, and the PDP densities to be derived below obtain for any such pair. Conditional on the event $C(s_1, s_2)$, all three pivotal ratios are correlated.

Theorem B.5. *For each fixed (s_1, s_2) pair, the joint PDP density of the pivotal ratios $W_1, W_2 | W_1 \le s_1, W_2 \le s_2$ is given by*

$$g^*(w_1, w_2|s_1, s_2) = \frac{\Gamma(\phi_1+\phi_2+\phi_3)}{\Gamma(\phi_1)\Gamma(\phi_2)\Gamma(\phi_3)} \frac{w_1^{\phi_2+\phi_3-1} w_2^{\phi_3-1}}{(1+w_1+w_1w_2)^{\phi_1+\phi_2+\phi_3}} \frac{1}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)}, \tag{B.6}$$

defined over the region $(0 < w_1 \le s_1, 0 < w_2 \le s_2)$.

Proof. Let s_1 and s_2 be given. Starting from the joint PDP density (3.3) of (U_1, U_2, U_3), make the transformation

$$\begin{aligned} V_1 &= U_1 & U_1 &= V_1 \\ V_2 &= U_2/U_1 & U_2 &= V_1 V_2 \\ V_3 &= U_3/U_2 & U_3 &= V_1 V_2 V_3, \end{aligned}$$

with Jacobian $V_1^2 V_2$. The joint PDP density of the V's is then

$$g^*(v_1, v_2, v_3 | s_1, s_2) = \frac{v_1^{\phi_1+\phi_2+\phi_3-1} v_2^{\phi_2+\phi_3-1} v_3^{\phi_3-1} \exp\{-v_1(1+v_2+v_2v_3)/2\}}{\Gamma(\phi_1)\Gamma(\phi_2)\Gamma(\phi_3)\, 2^{\phi_1+\phi_2+\phi_3} I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)}.$$

The joint density of interest is obtained by integrating the above density over V_1, i. e.,

$$g^*(v_2, v_3 | s_1, s_2) = \frac{v_2^{\phi_2+\phi_3-1} v_3^{\phi_3-1}}{\Gamma(\phi_1)\Gamma(\phi_2)\Gamma(\phi_3)} \int_0^\infty \frac{v_1^{\phi_1+\phi_2+\phi_3-1} \exp\{-v_1(1+v_2+v_2v_3)/2\}}{2^{\phi_1+\phi_2+\phi_3} I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)} dv_1.$$

Let $y = v_1(1+v_2+v_2v_3)$, so that

$$v_1 = \frac{y}{1+v_2+v_2v_3} \qquad \Longrightarrow \qquad dv_1 = \frac{dy}{1+v_2+v_2v_3};$$

with this substitution,

$$= \frac{v_2^{\phi_2+\phi_3-1} v_3^{\phi_3-1} (1+v_2+v_2v_3)^{-(\phi_1+\phi_2+\phi_3)}}{\Gamma(\phi_1)\Gamma(\phi_2)\Gamma(\phi_3)\, I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)} \int_0^\infty \frac{y^{\phi_1+\phi_2+\phi_3-1}\, e^{-y/2}}{2^{\phi_1+\phi_2+\phi_3}} dy$$

Identifying v_2 with w_1 and v_3 with w_2, the result follows. Note that the joint density of $W_1, W_2 | C$ is doubly right truncated. □

B.3.1 Marginal densities of the W_i^*.

Marginal density of $W_1|C$. There are two ways to obtain the marginal density of $W_1^* = W_1|C$: integrating over the joint density (B.6) of $W_1, W_2|C$, or from the joint density (3.3) of the U's. Both methods yield the same result, so we will integrate over (B.6).

Theorem B.6. *For any fixed $s_1, s_2 > 0$, the marginal density of $W_1^*(s_1, s_2)$ is*

$$g^*(w_1 | s_1, s_2) = \frac{\Gamma(\phi_1+\phi_2)}{\Gamma(\phi_1)\Gamma(\phi_2)} \frac{w_1^{\phi_2-1}}{(1+w_1)^{\phi_1+\phi_2}} \frac{I_{t^*(w_1)}(\phi_3, \phi_1+\phi_2)}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)},$$

defined on the region $0 < w_1 \leq s_1$, where

$$t^*(w_1) = \frac{s_1 s_2}{1+w_1+s_1 s_2} \tag{B.7}$$

Proof. Starting with (B.6), make the transformation

$$\begin{aligned} X_1 &= W_1 & \qquad W_1 &= X_1 \\ X_2 &= W_1 W_2 & \qquad W_2 &= X_2/X_1. \end{aligned} \tag{B.8}$$

The transformed joint density is then

$$g^*(x_1, x_2|s_1, s_2) = \frac{\Gamma(\phi_1+\phi_2+\phi_3)}{\Gamma(\phi_1)\Gamma(\phi_2)\Gamma(\phi_3)} \frac{x_1^{\phi_2-1} x_2^{\phi_3-1}}{(1+x_1+x_2)^{\phi_1+\phi_2+\phi_3}} \frac{1}{I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)}. \tag{B.9}$$

Integrating over X_2, the marginal of X_1 is

$$g^*(x_1|s_1, s_2) = \frac{\Gamma(\phi_1+\phi_2+\phi_3)}{\Gamma(\phi_1)\Gamma(\phi_2)\Gamma(\phi_3)} \frac{x_1^{\phi_2-1}}{I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)} \int_0^{s_1 s_2} \frac{x_2^{\phi_3-1}}{(1+x_1+x_2)^{\phi_1+\phi_2+\phi_3}}\, dx_2.$$

Multiply both numerator and denominator by $(1+x_1)^{\phi_1+\phi_2}$; then, the right hand side becomes

$$\frac{\Gamma(\phi_1+\phi_2+\phi_3)}{\Gamma(\phi_1)\Gamma(\phi_2)\Gamma(\phi_3)} \frac{x_1^{\phi_2-1}(1+x_1)^{-(\phi_1+\phi_2)}}{I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)} \int_0^{s_1 s_2} \frac{x_2^{\phi_3-1}(1+x_1)^{\phi_1+\phi_2}}{(1+x_1+x_2)^{\phi_1+\phi_2+\phi_3}}\, dx_2.$$

Make the standard Beta transformation, with $v = x_2/(1+x_1+x_2)$; the integration limits transform to

$$t^* = \frac{s_1 s_2}{1+x_1+s_1 s_2},$$

a function of x_1, so that

$$g^*(x_1|s_1, s_2) = \frac{\Gamma(\phi_1+\phi_2+\phi_3)}{\Gamma(\phi_1)\Gamma(\phi_2)\Gamma(\phi_3)} \frac{x_1^{\phi_2-1}(1+x_1)^{-(\phi_1+\phi_2)}}{I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)} \int_0^{t^*} v^{\phi_3-1}(1-v)^{\phi_1+\phi_2-1}\, dv.$$

The integral evaluates to

$$\frac{\Gamma(\phi_3)\Gamma(\phi_1+\phi_2)}{\Gamma(\phi_1+\phi_2+\phi_3)} I_{t^*}(\phi_3, \phi_1+\phi_2);$$

recalling that $x_1 = w_1$, the marginal density of W_1^* is

$$g^*(w_1|s_1, s_2) = \frac{\Gamma(\phi_1+\phi_2)}{\Gamma(\phi_1)\Gamma(\phi_2)} \frac{w_1^{\phi_2-1}}{(1+w_1)^{\phi_1+\phi_2}} \frac{I_{t^*}(\phi_3,\phi_1+\phi_2)}{I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)},$$

completing the proof. □

Marginal density of $W_2^* = W_2|C$. Integrating (B.6) over W_1, the marginal PDP density of W_2^* is obtained as follows:

Theorem B.7. *Given $s_1, s_2 > 0$, the marginal density of $W_2^*(s_1, s_2) = W_2|C(s_1, s_2)$ is given by*

$$g^*(w_2|s_1, s_2) = \frac{\Gamma(\phi_2+\phi_3)}{\Gamma(\phi_2)\Gamma(\phi_3)} \frac{w_2^{\phi_3-1}}{(1+w_2)^{\phi_2+\phi_3)}} \frac{I_{r^*(w_2)}(\phi_2+\phi_3, \phi_1)}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)}, \tag{B.10}$$

defined on $0 < w_2 \leq s_2$, where

$$r^*(w_2) = \frac{s_1\,(1+w_2)}{1+s_1\,(1+w_2)} \tag{B.11}$$

Proof. Given $s_1, s_2 > 0$, integrate the joint density (B.6) over w_1, which produces

$$g^*(w_2|s_1, s_2) = \frac{\Gamma(\phi_1+\phi_2+\phi_3)}{\Gamma(\phi_1)\Gamma(\phi_2)\Gamma(\phi_3)} \frac{w_2^{\phi_3-1}}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)} \int_0^{s_1} \frac{w_1^{\phi_2+\phi_3-1}}{(1+w_1+w_1w_2)^{\phi_1+\phi_2+\phi_3}}\, dw_1.$$

Let $x = w_1(1+w_2)$, which implies $w_1 = x/(1+w_2)$ and $dw_1 = dx/(1+w_2)$. Making these substitutions into the above equation, we get

$$g^*(w_2|s_1, s_2) = \frac{\Gamma(\phi_1+\phi_2+\phi_3)}{\Gamma(\phi_1)\Gamma(\phi_2)\Gamma(\phi_3)} \frac{w_2^{\phi_3-1}(1+w_2)^{-(\phi_2+\phi_3)}}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)} \int_0^{s_1(1+w_2)} \frac{x^{\phi_2+\phi_3-1}}{(1+x)^{\phi_1+\phi_2+\phi_3}}\, dx.$$

Now, let $v = x/(1+x)$ be the standard Beta transformation, with $dv = dx/(1+x)$; the integration limit becomes

$$r^*(w_2) = \frac{s_1\,(1+w_2)}{1+s_1\,(1+w_2)}$$

and the integral evaluates to

$$\frac{\Gamma(\phi_2+\phi_3)\Gamma(\phi_1)}{\Gamma(\phi_1+\phi_2+\phi_3)} I_{r^*}(\phi_2+\phi_3, \phi_1);$$

the marginal density of W_2^* is then

$$g^*(w_2|s_1, s_2) = \frac{\Gamma(\phi_2+\phi_3)}{\Gamma(\phi_2)\Gamma(\phi_3)} \frac{w_2^{\phi_3-1}}{(1+w_2)^{\phi_2+\phi_3)}} \frac{I_{r^*}(\phi_2+\phi_3, \phi_1)}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)},$$

defined on $0 < w_2 \leq s_2$. This completes the proof. □

Marginal density of $W_3^* = W_1W_2|C$. It is also useful to consider the PDP density of $W_3 = W_1W_2$. We have the following result concerning its marginal density:

Theorem B.8. *Given $s_1, s_2 > 0$, the marginal density of $W_3^*(s_1, s_2) = W_3|C(s_1, s_2)$ is given by*

$$g^*(w_3|s_1, s_2) = \frac{\Gamma(\phi_1+\phi_3)}{\Gamma(\phi_1)\Gamma(\phi_3)} \frac{w_3^{\phi_3-1}}{(1+w_3)^{\phi_1+\phi_3}} \frac{I_{p^*(w_3)}(\phi_2, \phi_1+\phi_3)}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)},$$

defined on the interval $0 < w_3 \leq s_1 s_2$, where

$$p^*(w_3) = \frac{s_1}{1+s_1+w_3}. \tag{B.12}$$

Proof. Starting from the joint density (B.6), make the bivariate transformation (B.8), which results in (B.9). In this case, the marginal PDP density of $X_2 = W_1W_2$ is

$$\begin{aligned} g^*(x_2|s_1, s_2) &= \frac{\Gamma(\phi_1+\phi_2+\phi_3)}{\Gamma(\phi_1)\Gamma(\phi_2)\Gamma(\phi_3)} \frac{x_2^{\phi_3-1}}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)} \int_0^{s_1} \frac{x_1^{\phi_2-1}}{(1+x_1+x_2)^{\phi_1+\phi_2+\phi_3}}\, dx_1 \\ &= \frac{\Gamma(\phi_1+\phi_2+\phi_3)}{\Gamma(\phi_1)\Gamma(\phi_2)\Gamma(\phi_3)} \frac{x_2^{\phi_3-1}(1+x_2)^{-(\phi_1+\phi_3)}}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)} \int_0^{s_1} \frac{x_1^{\phi_2-1}(1+x_2)^{\phi_1+\phi_3}}{(1+x_1+x_2)^{\phi_1+\phi_2+\phi_3}}\, dx_1. \end{aligned}$$

Now, make the transformation

$$v = \frac{x_1}{1+x_1+x_2} \qquad \Longrightarrow \qquad dv = \frac{(1+x_2)\, dx_1}{(1+x_1+x_2)^2};$$

This is a standard Beta transformation, which transforms the integration limit to

$$p^*(x_2) = \frac{s_1}{1+s_1+x_2};$$

the integral then evaluates to

$$\frac{\Gamma(\phi_2)\Gamma(\phi_1+\phi_3)}{\Gamma(\phi_1+\phi_2+\phi_3)} I_{p^*}(\phi_2, \phi_1+\phi_3),$$

so the marginal density of $W_3^* = W_1W_2|C$ is given by

$$g^*(w_3|s_1, s_2) = \frac{\Gamma(\phi_1+\phi_3)}{\Gamma(\phi_1)\Gamma(\phi_3)} \frac{w_3^{\phi_3-1}}{(1+w_3)^{\phi_1+\phi_3}} \frac{I_{p^*}(\phi_2, \phi_1+\phi_3)}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)},$$

where x_2 is replaced by w_3 in (B.12). This density is defined on $0 < w_3 \leq s_1 s_2$, and the proof is complete. □

B.4 Moment generating functions of U_i^*

mgf of U_1^*. For each fixed (s_1, s_2) pair, the moment generating function of U_1 is obtained as follows.

Theorem B.9. *Given realized sums of squares ratios (s_1, s_2), the mgf of U_1^* is*

$$M_{U_1^*}(t|s_1,s_2) = (1-2t)^{-\phi_1} \frac{I_{z_1',z_2'}(\phi_3,\phi_2,\phi_1)}{I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)}, \tag{B.13}$$

where

$$z_1 = \frac{s_1}{1+s_1+s_1s_2} \qquad z_2 = \frac{s_1s_2}{1+s_1+s_1s_2}$$

$$z_1' = \frac{s_1}{(1-2t)+s_1+s_1s_2} \qquad z_2' = \frac{s_1s_2}{(1-2t)+s_1+s_1s_2}.$$

Proof. By definition,

$$\begin{aligned}
M_{U_1^*}(t|s_1,s_2) &= \int_0^\infty \int_0^{s_1u_1} \int_0^{s_2u_2} \frac{e^{tu_1}\, g_{\nu_1}(u_1)\, g_{\nu_2}(u_2)\, g_{\nu_3}(u_3)\, du_3\, du_2\, du_1}{I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)} \\
&= \int_0^\infty \int_0^{s_1u_1} \int_0^{s_2u_2} \frac{e^{tu_1} \prod_{i=1}^3 u_i^{\phi_i-1} e^{-(u_1+u_2+u_3)/2}}{\Gamma(\phi_1)\Gamma(\phi_2)\Gamma(\phi_3)\, 2^{\phi_1+\phi_2+\phi_3}\, I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)}\, du_3\, du_2\, du_1 \\
&= \int_0^\infty \int_0^{s_1u_1} \int_0^{s_2u_2} \frac{\prod_{i=1}^3 u_i^{\phi_i-1} e^{-(u_1(1-2t)+u_2+u_3)/2}}{\Gamma(\phi_1)\Gamma(\phi_2)\Gamma(\phi_3)\, 2^{\phi_1+\phi_2+\phi_3}\, I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)}\, du_3\, du_2\, du_1.
\end{aligned}$$

Let

$$\begin{aligned}
V_1 &= (1-2t)U_1 + U_2 + U_3 & & U_1 = V_1(1-V_2-V_3)(1-2t)^{-1} \\
V_2 &= \frac{U_2}{(1-2t)U_1+U_2+U_3} & \Longrightarrow \quad & U_2 = V_1V_2 \\
V_3 &= \frac{U_3}{(1-2t)U_1+U_2+U_3} & & U_3 = V_1V_3,
\end{aligned}$$

with Jacobian $V_1^2/(1-2t)$. This transformation changes the upper integration limits from (∞, s_1u_1, s_2u_2) to (∞, z_1', z_2'). In the transformed scale, this integral becomes

$$(1-2t)^{-\phi_1} \int_0^\infty \int_0^{z_1'} \int_0^{z_2'} \frac{v_1^{\phi_1+\phi_2+\phi_3-1} v_2^{\phi_2-1} v_3^{\phi_3-1} (1-v_2-v_3)^{\phi_1-1}\, e^{-v_1/2}\, dv_3\, dv_2\, dv_1}{\Gamma(\phi_1)\Gamma(\phi_2)\Gamma(\phi_3)\, 2^{\phi_1+\phi_2+\phi_3}\, I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)};$$

Exchanging the order of integration to eliminate V_1,

$$= (1-2t)^{-\phi_1} \int_0^{z_1'} \int_0^{z_2'} \frac{\Gamma(\phi_1+\phi_2+\phi_3)\, v_2^{\phi_2-1} v_3^{\phi_3-1} (1-v_2-v_3)^{\phi_1-1}\, dv_3\, dv_2}{\Gamma(\phi_1)\Gamma(\phi_2)\Gamma(\phi_3)\, I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)},$$

so that

$$M_{U_1}^*(t|s_1,s_2) = (1-2t)^{-\phi_1} \frac{I_{z_1',z_2'}(\phi_3,\phi_2,\phi_1)}{I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)},$$

as required. □

Similar lines of proof apply to derivation of mgf's for U_2^* and U_3^* for each fixed (s_1, s_2) pair, producing

$$M_{U_2}^*(t|s_1,s_2) = (1-2t)^{-\phi_2} \frac{I_{z_1^*,z_2^*}(\phi_3,\phi_2,\phi_1)}{I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)}, \tag{B.14}$$

where

$$z_1^* = \frac{1}{1+(1-2t)s_1+s_1s_2} \quad \text{and} \quad z_2^* = \frac{s_1s_2}{1+(1-2t)s_1+s_1s_2},$$

and

$$M_{U_3}^*(t|s_1,s_2) = (1-2t)^{-\phi_3} \frac{I_{z_1^+,z_2^+}(\phi_3,\phi_2,\phi_1)}{I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)}, \tag{B.15}$$

where

$$z_1^+ = \frac{1}{1+s_1+(1-2t)s_1s_2} \quad \text{and} \quad z_2^* = \frac{s_1}{1+s_1+(1-2t)s_1s_2}.$$

B.5 Moments of the $U_i|C$ and $W_i|C$

In this section, two primary theorems are introduced: one involving the sum of higher order moments, and the other involving the product of higher order moments in balanced random three-stage nested models. All of the necessary first and second moments, product moments, variances and covariances among the $U_i|C$ and $W_i|C$ are special cases of these results.

B.5.1 Moments of $U_i|C$

Theorem B.10. *Given $s_1, s_2 > 0$,*

$$E(U_1^{k_1} + U_2^{k_2} + U_3^{k_3}|C) = \frac{2^{k_1}\,\Gamma(\phi_1 + k_1)}{\Gamma(\phi_1} \frac{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1 + k_1)}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)} + \frac{2^{k_2}\,\Gamma(\phi_2 + k_2)}{\Gamma(\phi_2} \frac{I_{z_1,z_2}(\phi_3, \phi_2 + k_2, \phi_1)}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)} + \frac{2^{k_3}\,\Gamma(\phi_3 + k_3)}{\Gamma(\phi_3} \frac{I_{z_1,z_2}(\phi_3 + k_3, \phi_2, \phi_1)}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)}, \tag{B.16}$$

where $C = \{U_2 \le s_1 U_1, U_3 \le s_2 U_2\}$ is the conditioning event.

Remark B.1. The k_i's can assume any nonnegative value, and can also assume negative values so long as $k_i > -\phi_i$, $i = 1, 2, 3$.

Proof. It suffices to prove one of these, as they all have the same general form. We choose to find $E(U_1^{k_1}|C)$. Now,

$$E(U_1^{k_1}|C) = \int_0^\infty \int_0^{s_1 u_1} \int_0^{s_2 u_2} u_1^{k_1} \frac{g_{\nu_1}(u_1)\, g_{\nu_2}(u_2)\, g_{\nu_3}(u_3)}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)}\, du_3\, du_2\, du_1.$$

By Theorem A.22,

$$= A_1 \int_0^\infty \int_0^{s_1 u_1} \int_0^{s_2 u_2} \frac{u_1^{\phi_1+k_1-1}\, u_2^{\phi_2-1}\, u_3^{\phi_3-1}\, e^{-u_T/2}\, du_3\, du_2\, du_1}{\Gamma(\phi_1 + k_1)\, \Gamma(\phi_2)\, \Gamma(\phi_3)\, 2^{\phi_T+k_1}\, I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)},$$

where

$$A_1 = \frac{2^{k_1}\,\Gamma(\phi_1 + k_1)}{\Gamma(\phi_1)}, \qquad u_T = u_1 + u_2 + u_3, \quad \text{and} \quad \phi_T = \phi_1 + \phi_2 + \phi_3.$$

Next, apply the Dirichlet transformation (B.4); the integral expression then becomes

$$= A_1 \int_0^{z_1} \int_0^{z_2} \frac{v_3^{\phi_3-1}\, v_2^{\phi_2-1}\, (1 - v_2 - v_3)^{\phi_1+k_1-1}}{\Gamma(\phi_1 + k_1)\, \Gamma(\phi_2)\, \Gamma(\phi_3)\, I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)} \left[\int_0^\infty \frac{v_1^{\phi_T+k_1-1}\, e^{v_1/2}}{2^{\phi_T+k_1}}\, dv_1 \right] dv_3\, dv_2$$
$$= A_1 \int_0^{z_1} \int_0^{z_2} \frac{\Gamma(\phi_T + k_1)}{\Gamma(\phi_1 + k_1)\, \Gamma(\phi_2)\, \Gamma(\phi_3)} \frac{v_3^{\phi_3-1}\, v_2^{\phi_2-1}\, (1 - v_2 - v_3)^{\phi_1+k_1-1}}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)}\, dv_3\, dv_2$$
$$= \frac{2^{k_1}\,\Gamma(\phi_1 + k_1)}{\Gamma(\phi_1)} \frac{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1 + k_1)}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)},$$

by Theorem B.2.

By a similar line of argument, one can show that

$$E(U_2^{k_2}|C) = \frac{2^{k_2}\,\Gamma(\phi_2+k_2)}{\Gamma(\phi_2)}\,\frac{I_{z_1,z_2}(\phi_3,\phi_2+k_2,\phi_1)}{I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)},$$
$$E(U_3^{k_3}|C) = \frac{2^{k_3}\,\Gamma(\phi_3+k_3)}{\Gamma(\phi_3)}\,\frac{I_{z_1,z_2}(\phi_3+k_3,\phi_2,\phi_1)}{I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)}.$$

The result follows by addition. □

Theorem B.11. *For any fixed $s_1, s_2 > 0$,*

$$E(U_1^{k_1}\,U_2^{k_2}\,U_3^{k_3}|C) = \Big(\prod_{i=1}^{3} 2^{k_i}\,\frac{\Gamma(\phi_i+k_1)}{\Gamma(\phi_i)}\Big)\,\frac{I_{z_1,z_2}(\phi_3+k_3,\phi_2+k_2,\phi_1+k_1)}{I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)}, \qquad \text{(B.17)}$$

which holds for any integer $k_i > -\phi_i$, $i = 1, 2, 3$.

Proof. Given $s_1, s_2 > 0$,

$$E(U_1^{k_1}\,U_2^{k_2}\,U_3^{k_3}|C) = \int_0^\infty\int_0^{s_1u_1}\int_0^{s_2u_2} u_1^{k_1}\,u_2^{k_2}\,u_3^{k_3}\,\frac{\prod_{i=1}^{3} g_{\nu_i}(u_i)}{I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)}\,du_3\,du_2\,du_1.$$

By Theorem A.22,

$$= B_1 \int_0^\infty\int_0^{s_1u_1}\int_0^{s_2u_2} \frac{\prod_{i=1}^{3} g_{\nu_i+2k_i}(u_i)}{I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)}\,du_3\,du_2\,du_1,$$

where

$$B_1 = \prod_{i=1}^{3} 2^{k_i}\,\frac{\Gamma(\phi_i+k_1)}{\Gamma(\phi_i)}.$$

Apply the Dirichlet transformation (B.4); after integrating out v_1, what remains is

$$= B_1 \int_0^{z_1}\int_0^{z_2} \frac{\Gamma(\phi_T+k_1+k_2+k_3)}{\prod_{i=1}^{3}\Gamma(\phi_i+k_i)}\,\frac{v_3^{\phi_3+k_3-1}\,v_2^{\phi_2+k_2-1}\,(1-v_2-v_3)^{\phi_1+k_1-1}}{I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)}\,dv_3\,dv_2$$
$$= B_1\,\frac{I_{z_1,z_2}(\phi_3+k_3,\phi_2+k_2,\phi_1+k_1)}{I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)},$$

completing the proof. □

It is convenient to label the exponents as the triplet (k_3, k_2, k_1) corresponding to (U_3, U_2, U_1), in that order. For example, $E(U_1^2|C)$ would correspond to the triplet $(0, 0, 2)$, while the product moment $E(U_1U_2|C)$ would correspond to $(0, 1, 1)$. Negative moments can also be handled; for example, $W_1 = U_2/U_1$ would correspond to $(0, 1, -1)$. Nearly all of the (product) moments of the $U_i|C$ and $W_i|C$ can be ascertained from the above result.

B.5.2 Moments of $W_i|C$

Define $W_1 = U_2/U_1$, $W_2 = U_3/U_2$ and $W_3 = W_1W_2 = U_3/U_1$.

Corollary B.11.1. *For any fixed $s_1, s_2 > 0$,*

$$E(W_1|C) = \frac{\nu_2}{\nu_1 - 2} \frac{I_{z_1,z_2}(\phi_3, \phi_2 + 1, \phi_1 - 1)}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)}, \qquad \nu_1 > 2 \tag{B.18}$$

$$E(W_2|C) = \frac{\nu_3}{\nu_2 - 2} \frac{I_{z_1,z_2}(\phi_3 + 1, \phi_2 - 1, \phi_1)}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)}, \qquad \nu_2 > 2 \tag{B.19}$$

$$E(W_3|C) = \frac{\nu_3}{\nu_1 - 2} \frac{I_{z_1,z_2}(\phi_3 + 1, \phi_2, \phi_1 - 1)}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)}, \qquad \nu_1 > 2. \tag{B.20}$$

The proof follows by taking $(k_3, k_2, k_1) = (0, 1, -1)$, $(1, -1, 0)$ and $(1, 0, -1)$, respectively, and applying Theorem B.11.

Corollary B.11.2. *For any fixed $s_1, s_2 > 0$, the k^{th} moments of $W_i|C$, $i = 1, 2, 3$ are*

$$E(W_1^k|C) = 2^k \frac{\Gamma(\phi_2 + k)}{\Gamma(\phi_2)} 2^{-k} \frac{\Gamma(\phi_1 - k)}{\Gamma(\phi_1)} \frac{I_{z_1,z_2}(\phi_3, \phi_2 + k, \phi_1 - k)}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)}, \qquad \nu_1 > 2k \tag{B.21}$$

$$E(W_2^k|C) = 2^k \frac{\Gamma(\phi_3 + k)}{\Gamma(\phi_3)} 2^{-k} \frac{\Gamma(\phi_2 - k)}{\Gamma(\phi_2)} \frac{I_{z_1,z_2}(\phi_3 + k, \phi_2 - k, \phi_1)}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)}, \qquad \nu_2 > 2k \tag{B.22}$$

$$E(W_3^k|C) = 2^k \frac{\Gamma(\phi_3 + k)}{\Gamma(\phi_3)} 2^{-k} \frac{\Gamma(\phi_1 - k)}{\Gamma(\phi_1)} \frac{I_{z_1,z_2}(\phi_3 + k, \phi_2, \phi_1 - k)}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)}, \qquad \nu_1 > 2k. \tag{B.23}$$

The proof follows from Theorem B.11 with $(k_3, k_2, k_1) = (0, k, -k)$, $(k, -k, 0)$ and $(k, 0, -k)$, respectively.

Corollary B.11.3. *For any fixed $s_1, s_2 > 0$, the variances of $W_i|C$, $i = 1, 2, 3$ are*

$$\begin{aligned}\operatorname{Var}(W_1|C) = &\frac{\nu_2(\nu_2+2)}{(\nu_1-2)(\nu_1-4)}\frac{I_{z_1,z_2}(\phi_3,\phi_2+2,\phi_1-2)}{I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)}\\ &-\left(\frac{\nu_2}{\nu_1-2}\right)^2\frac{I^2_{z_1,z_2}(\phi_3,\phi_2+1,\phi_1-1)}{I^2_{z_1,z_2}(\phi_3,\phi_2,\phi_1)}, \qquad \nu_1 > 4,\end{aligned} \tag{B.24}$$

$$\begin{aligned}\operatorname{Var}(W_2|C) = &\frac{\nu_3(\nu_3+2)}{(\nu_2-2)(\nu_2-4)}\frac{I_{z_1,z_2}(\phi_3+2,\phi_2-2,\phi_1)}{I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)}\\ &-\left(\frac{\nu_3}{\nu_2-2}\right)^2\frac{I^2_{z_1,z_2}(\phi_3+1,\phi_2-1,\phi_1)}{I^2_{z_1,z_2}(\phi_3,\phi_2,\phi_1)} \qquad \nu_2 > 4,\end{aligned} \tag{B.25}$$

$$\begin{aligned}\operatorname{Var}(W_3|C) = &\frac{\nu_3(\nu_3+2)}{(\nu_1-2)(\nu_1-4)}\frac{I_{z_1,z_2}(\phi_3+2,\phi_2,\phi_1-2)}{I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)}\\ &-\left(\frac{\nu_3}{\nu_1-2}\right)^2\frac{I^2_{z_1,z_2}(\phi_3+1,\phi_2,\phi_1-1)}{I^2_{z_1,z_2}(\phi_3,\phi_2,\phi_1)} \qquad \nu_1 > 4.\end{aligned} \tag{B.26}$$

The proof follows from Corollary B.11.2 and the definition of variance.

To obtain covariances among pairs of W's given C, we first require a result on product moments of the $W_iW_j|C$.

Corollary B.11.4. *For any fixed $s_1, s_2 > 0$,*

$$E(W_1W_2|C) = E(W_3|C) = \frac{\nu_3}{\nu_1-2}\frac{I_{z_1,z_2}(\phi_3+1,\phi_2,\phi_1-1)}{I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)}, \qquad \nu_1 > 2 \tag{B.27}$$

$$E(W_1W_3|C) = E\left(\frac{U_2U_3}{U_1^2}\Big|C\right) = \frac{\nu_2\nu_3\, I_{z_1,z_2}(\phi_3+1,\phi_2+1,\phi_1-2)}{(\nu_1-2)(\nu_1-4)\, I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)}, \quad \nu_1 > 4 \tag{B.28}$$

$$E(W_2W_3|C) = E\left(\frac{U_3^2}{U_1U_2}\Big|C\right) = \frac{\nu_3(\nu_3+2)}{(\nu_1-2)(\nu_2-2)}\frac{I_{z_1,z_2}(\phi_3+2,\phi_2-1,\phi_1-1)}{I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)}, \tag{B.29}$$

where $\nu_1 > 2, \nu_2 > 2$ for (B.29).

The proof follows from Theorem B.11 with $(k_3, k_2, k_1) = (1, 0, -1)$, $(1, 1, -2)$ and $(2, -1, -1)$, respectively.

Corollary B.11.5. *For any fixed* $s_1, s_2 > 0$,

$$\begin{aligned} \operatorname{cov}(W_1, W_2|C) = & \frac{\nu_3}{\nu_1 - 2} \frac{I_{z_1,z_2}(\phi_3 + 1, \phi_2, \phi_1 - 1)}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)} \\ & - \frac{\nu_2 \nu_3 \; I_{z_1,z_2}(\phi_3, \phi_2 + 1, \phi_1 - 1) \, I_{z_1,z_2}(\phi_3 + 1, \phi_2 - 1, \phi_1)}{(\nu_1 - 2)(\nu_2 - 2) \; I^2_{z_1,z_2}(\phi_3, \phi_2, \phi_1)}; \end{aligned} \tag{B.30}$$

$$\begin{aligned} \operatorname{cov}(W_1, W_3|C) = & \frac{\nu_2 \nu_3}{(\nu_1 - 2)(\nu_1 - 4)} \frac{I_{z_1,z_2}(\phi_3 + 1, \phi_2 + 1, \phi_1 - 1)}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)} \\ & - \frac{\nu_2 \nu_3}{(\nu_1 - 2)^2} \frac{I_{z_1,z_2}(\phi_3, \phi_2 + 1, \phi_1 - 1) \, I_{z_1,z_2}(\phi_3 + 1, \phi_1, \phi_1 - 1)}{I^2_{z_1,z_2}(\phi_3, \phi_2, \phi_1)}; \end{aligned} \tag{B.31}$$

$$\begin{aligned} \operatorname{cov}(W_2, W_3|C) = & \frac{\nu_3(\nu_3 + 2)}{(\nu_1 - 2)(\nu_2 - 2)} \frac{I_{z_1,z_2}(\phi_3 + 2, \phi_2 - 1, \phi_1 - 1)}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)} \\ & - \frac{\nu_3^2 \; I_{z_1,z_2}(\phi_3 + 1, \phi_2, \phi_1 - 1) \, I_{z_1,z_2}(\phi_3 + 1, \phi_2 - 1, \phi_1)}{(\nu_1 - 2)(\nu_2 - 2) \; I^2_{z_1,z_2}(\phi_3, \phi_2, \phi_1)}. \end{aligned} \tag{B.32}$$

The proof follows from Corollaries B.11.4 and B.11.1, along with the definition of covariance.

B.6 Distributions for $K\sigma_\beta^2$ and $JK\sigma_\alpha^2$

Given $t_1, t_2, t_3 > 0$, define

$$V_1^*(t_1, t_2, t_3) = \frac{t_2}{U_2} - \frac{t_1}{U_1} \Big| W_1 \le s_1, W_2 \le s_2,$$

used for inference about $K\sigma_\beta^2$, and

$$V_2^*(t_1, t_2, t_3) = \frac{t_3}{U_3} - \frac{t_2}{U_2} \Big| W_1 \le s_1, W_2 \le s_2,$$

used for inference about $JK\sigma_\alpha^2$. We are interested in the densities and moments of the V_i^*, $i = 1, 2$.

B.6.1 Results concerning V_1^*.

Starting from the joint density (3.3) of the U's, make the transformation

$$V_1 = \frac{t_2}{U_2} - \frac{t_1}{U_1} \qquad U_1 = \frac{t_1}{V_2}$$

$$V_2 = \frac{t_1}{U_1} \qquad U_2 = \frac{t_2}{V_1 + V_2}$$

$$V_3 = U_3 \qquad U_3 = V_3,$$

with Jacobian $t_1 t_2[v_1(v_1+v_2)]^{-2}$. In the coordinate system of the $U's$, the domain is $\{U_1 > 0, U_2 \le s_1 U_1, U_3 \le s_2 U_2\}$; in the transformed V scale, the domain becomes $\{V_1 > 0, V_2 > 0, V_3 \le s_2 t_2/(V_1+V_2)\}$. The joint density of $V_1, V_2, V_3|C$ is then

$$g^*(v_1, v_2, v_3|\boldsymbol{t}) = \frac{\left(\frac{t_1}{v_2}\right)^{\phi_1+1}\left(\frac{t_2}{v_1+v_2}\right)^{\phi_2+1} v_3^{\phi_3-1} \exp\left\{-\frac{1}{2}\left(\frac{t_2}{v_2} + \frac{t_2}{v_1+v_2} + v_3\right)\right\}}{t_1 t_2\,\Gamma(\phi_1)\Gamma(\phi_2)\Gamma(\phi_3) 2^{\phi_1+\phi_2+\phi_3} I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)},$$

where $\boldsymbol{t} = (t_1, t_2, t_3)$. The marginal density of V_1^* is obtained by integrating over V_2 and V_3; the integral expression is

$$\int_0^\infty \int_0^{\frac{s_2 t_2}{v_1+v_2}} \frac{\left(\frac{t_1}{v_2}\right)^{\phi_1+1}\left(\frac{t_2}{v_1+v_2}\right)^{\phi_2+1} v_3^{\phi_3-1} \exp\left\{-\frac{1}{2}\left(\frac{t_1}{v_2} + \frac{t_2}{v_1+v_2} + v_3\right)\right\}}{t_1 t_2\,\Gamma(\phi_1)\Gamma(\phi_2)\Gamma(\phi_3)\, 2^{\phi_1+\phi_2+\phi_3} I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)}\, dv_3\, dv_2.$$

Evaluating the inner integral, the density of V_1^* is

$$g^*(v_1|\boldsymbol{t}) = \int_0^\infty \frac{\left(\frac{t_1}{v_2}\right)^{\phi_1+1}\left(\frac{t_2}{v_1+v_2}\right)^{\phi_2+1} \exp\left\{-\frac{1}{2}\left(\frac{t_1}{v_2} + \frac{t_2}{v_1+v_2}\right)\right\} G_{\nu_3}\left(\frac{s_2 t_2}{v_1+v_2}\right)}{t_1 t_2\,\Gamma(\phi_1)\Gamma(\phi_2)\, 2^{\phi_1+\phi_2} I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)}\, dv_2.$$

As was the case in the one-way random model, the marginal density of V_1^* does not admit a closed form solution.

To find the moments of V_1^*, it is convenient to go back to the (U_1, U_2, U_3) scale.

Using (3.17), we have

$$\begin{aligned}
E(V_1^*|\boldsymbol{t}) &= \int_0^\infty \int_0^{s_1u_1} \int_0^{s_2u_2} \left(\frac{t_2}{u_2} - \frac{t_1}{u_1}\right) \frac{g_{\nu_1}(u_1)\, g_{\nu_2}(u_2)\, g_{\nu_3}(u_3)}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)}\, du_3\, du_2\, du_1 \\
&= t_2 \int_0^\infty \int_0^{s_1u_1} \int_0^{s_2u_2} \frac{\Gamma(\phi_2-1)}{\Gamma(\phi_2)} \frac{g_{\nu_1}(u_1) g_{\nu_2-2}(u_2) g_{\nu_3}(u_3)}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)}\, du_3\, du_2\, du_1 \\
&\quad - t_1 \int_0^\infty \int_0^{s_1u_1} \int_0^{s_2u_2} \frac{\Gamma(\phi_1-1)}{\Gamma(\phi_1)} \frac{g_{\nu_1-2}(u_1) g_{\nu_2}(u_2) g_{\nu_3}(u_3)}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)}\, du_3\, du_2\, du_1 \\
&= \frac{t_2}{2}\left[\frac{\Gamma(\phi_2-1)}{\Gamma(\phi_2)} \frac{I_{z_1,z_2}(\phi_3, \phi_2-1, \phi_1)}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)} - \frac{1}{s_1} \frac{\Gamma(\phi_1-1)}{\Gamma(\phi_1)} \frac{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1-1)}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)}\right].
\end{aligned}$$

Since

$$E(V_1^{*2}|\boldsymbol{t}) = E\left[\left(\frac{t_2}{U_2} - \frac{t_1}{U_1}\right)^2 \middle| C\right],$$

expansion of the integral leads to

$$\begin{aligned}
E(V_1^{*2}|\boldsymbol{t}) = \left(\frac{t_2}{2}\right)^2 \Bigg[&\frac{\Gamma(\phi_2-2)}{\Gamma(\phi_2)} \frac{I_{z_1,z_2}(\phi_3, \phi_2-2, \phi_1)}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)} \\
&- \frac{2}{s_1} \frac{\Gamma(\phi_1-1)\Gamma(\phi_2-1)}{\Gamma(\phi_1)\Gamma(\phi_2)} \frac{I_{z_1,z_2}(\phi_3, \phi_2-1, \phi_1-1)}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)} \\
&+ \frac{1}{s_1^2} \frac{\Gamma(\phi_1-2)}{\Gamma(\phi_1)} \frac{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1-2)}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)}. \Bigg]
\end{aligned}$$

Compare $E(V_1^*|\boldsymbol{t})$ and $E(V_1^{*2}|\boldsymbol{t})$ with (2.32) and (2.34), respectively.

We can extend formula (2.35) for the r^{th} moment of V^* in Chapter 2 to obtain the r^{th} moment of V_1^*, given by

$$\left(\frac{t_2}{2}\right)^r \sum_{k=0}^r \binom{r}{k} \left(-\frac{1}{s_1}\right)^k \frac{\Gamma(\phi_1-k)\,\Gamma(\phi_2-r+k)}{\Gamma(\phi_1)\,\Gamma(\phi_2)} \frac{I_{z_1,z_2}(\phi_3, \phi_2-r+k, \phi_1-k)}{I_{z_1,z_2}(\phi_3, \phi_2, \phi_1)},$$

which we will denote as $E(V_1^{*r})$.

B.6.2 Results concerning V_2^*.

Starting from the joint density (3.3) of the U's, make the transformation

$$\begin{aligned} W_1 &= U_1 & U_1 &= W_1 \\ W_2 &= \frac{t_2}{U_2} & U_2 &= \frac{t_2}{W_2} \\ W_3 &= \frac{t_3}{U_3} - \frac{t_2}{U_2} & U_3 &= \frac{t_3}{W_2 + W_3}, \end{aligned}$$

with Jacobian $|J| = t_2 t_3 [W_2^2 (W_2 + W_3)^2]^{-1}$. The joint density of $(W_1, W_2, W_3)|C$ is then transformed to $g^*(w_1, w_2, w_3|s_1, s_2)$, given by

$$\frac{w_1^{\phi_1 - 1}\left(\frac{t_2}{w_2}\right)^{\phi_2+1}\left(\frac{t_3}{w_2+w_3}\right)^{\phi_3+1}\exp\left\{-\frac{1}{2}\left(w_1 + \frac{t_2}{w_2} + \frac{t_3}{w_2+w_3}\right)\right\}}{t_2 t_3\,\Gamma(\phi_1)\Gamma(\phi_2)\Gamma(\phi_3)\,2^{\phi_1+\phi_2+\phi_3}\,I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)},$$

defined over the region $w_1 \geq t_1/w_2, w_2 > 0, w_3 > 0$. The marginal density of V_2^* is identified with that of $W_3|C$, given by

$$g^*(w_3|s_1, s_2) = \int_{t_1/w_2}^{\infty}\int_0^{\infty} g^*(w_1, w_2, w_3|s_1, s_2)\, dw_2\, dw_1;$$

exchanging the order of integration, the marginal density works out to

$$\int_0^{\infty} \frac{\left(\frac{t_2}{w_2}\right)^{\phi_2+1}\left(\frac{t_3}{w_2+w_3}\right)^{\phi_3+1}\exp\left\{-\frac{1}{2}\left(\frac{t_2}{w_2} + \frac{t_3}{w_2+w_3}\right)\right\}[1 - G_{\nu_1}(t_1/w_2)]}{t_2 t_3\,\Gamma(\phi_2)\Gamma(\phi_3)\,2^{\phi_2+\phi_3}\,I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)}\, dw_2.$$

As above, to find the moments of $V_2^* = W_3|C$, it is convenient to return to the (U_1, U_2, U_3) coordinate system. The mean of V_2^* is

$$\begin{aligned} E(V_2^*|\boldsymbol{t}) &= \int_0^{\infty}\int_0^{s_1u_1}\int_0^{s_2u_2}\left(\frac{t_3}{u_3} - \frac{t_2}{u_2}\right)\frac{g_{\nu_1}(u_1)\,g_{\nu_2}(u_2)\,g_{\nu_3}(u_3)}{I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)}\, du_3\, du_2\, du_1 \\ &= t_3 \int_0^{\infty}\int_0^{s_1u_1}\int_0^{s_2u_2}\frac{\Gamma(\phi_3 - 1)}{\Gamma(\phi_3)}\frac{g_{\nu_1}(u_1) g_{\nu_2}(u_2) g_{\nu_3 - 2}(u_3)}{I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)}\, du_3\, du_2\, du_1 \\ &\quad - t_2 \int_0^{\infty}\int_0^{s_1u_1}\int_0^{s_2u_2}\frac{\Gamma(\phi_2 - 1)}{\Gamma(\phi_2)}\frac{g_{\nu_1}(u_1) g_{\nu_2 - 2}(u_2) g_{\nu_3}(u_3)}{I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)}\, du_3\, du_2\, du_1, \end{aligned}$$

so that

$$E(V_2^*|\boldsymbol{t}) = \frac{t_3}{2}\Big[\frac{\Gamma(\phi_3-1)}{\Gamma(\phi_3)}\,\frac{I_{z_1,z_2}(\phi_3-1,\phi_2,\phi_1)}{I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)} - \frac{1}{s_2}\,\frac{\Gamma(\phi_2-1)}{\Gamma(\phi_2)}\,\frac{I_{z_1,z_2}(\phi_3,\phi_2-1,\phi_1)}{I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)}\Big].$$

In general, the r^{th} moment of V_2^* is:

$$\Big(\frac{t_3}{2}\Big)^r \sum_{k=0}^{r} \binom{r}{k} \Big(-\frac{1}{s_2}\Big)^k \frac{\Gamma(\phi_2-k)\,\Gamma(\phi_3-r+k)}{\Gamma(\phi_2)\,\Gamma(\phi_3)}\,\frac{I_{z_1,z_2}(\phi_3-r+k,\phi_2-k,\phi_1)}{I_{z_1,z_2}(\phi_3,\phi_2,\phi_1)}.$$

Appendix C

References

Abramowitz, M. and I. Stegun (1964). *A Handbook of Mathematical Functions.* New York: Dover.

Barnard, G. (1963). Some logical aspects of the fiducial argument. *Journal of the Royal Statistical Society, Series B*, **25**, 111–114.

Barnard, G. A. (1995). Pivotal models and the fiducial argument. *International Statistical Review*, **63**, 309–323.

Barnett, V. (1982). *Comparative Statistical Inference, Second Edition.* New York: Wiley.

Becker, N. and I. Gordon (1983). On Cox's criterion for discriminating between alternative ancillary statistics. *International Statistical Review*, **51**, 89–92.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis, Second Edition.* New York: Springer-Verlag.

Berger, J. O. (1988). An alternative: the estimated confidence approach. Discussion of Casella (1988).

Berger, J. O. (1990). Comment on Maatta and Casella (1990). *Statistical Science*, **5**, 102–103.

Berger, J. O. and R. Wolpert (1984). *The Likelihood Principle.* Hayward, CA: IMS Lecture Notes–Monograph Series, vol. 6.

Boardman, T. (1974). Confidence intervals for variance components—a comparative Monte Carlo study. *Biometrics*, **30**, 251–262.

Bondar, J. V. (1977). A conditional confidence principle. *Annals of Statistics*, **5**, 881–891.

Box, G. E. P., and G. C. Tiao (1973). *Bayesian Inference in Statistical Analysis.* Reading, MA: Addison-Wesley.

Brewster, J. F. (1972). *On the admissibility of scale and quantile estimators.* Unpublished Ph.D Thesis, University of British Columbia.

Brewster, J. F. and J. V. Zidek (1974). Improving on equivariant estimators. *Annals of Statistics*, **2**, 21–38.

Brown, L. D. (1967). The conditional level of Student's test. *Annals of Mathematical Statistics*, **38**, 1068–1071.

Brown, L. D. (1968). Inadmissibility of the usual estimators of scale parameters in problems with unknown location and scale parameters. *Annals of Mathematical Statistics*, **39**, 29–48.

Brown, L. D. (1978). A contribution to Kiefer's theory of conditional confidence procedures. *Annals of Statistics*, **6**, 59–71.

Brown, L. D. (1990). Discussion of Maatta and Casella (1990). *Statistical Science*, **5**, 103–106.

Brownie, C. and J. Kiefer (1977). The ideas of conditional confidence in the simplest setting. *Communications in Statistics*, **A6(8)**, 691–751.

Buehler, R. J. (1959). Some validity criteria for statistical inferences. *Annals of Mathematical Statistics*, **30**, 845–863.

Buehler, R. J. and A. P. Fedderson (1963). Note on a conditional property of Student's *t*. *Annals of Mathematical Statistics*, **34**, 1098–1100.

Casella, G. (1987). Conditionally acceptable recentered set estimators. *Annals of Statistics*, **15**, 1363–1371.

Casella, G. (1988). Conditionally acceptable frequentist solutions. In *Statistical Decision Theory and Related Topics IV* (S. S. Gupta and J. O. Berger, eds). New York: Springer, 73–118 (with discussion).

Casella, G. (1992). Conditional inference from confidence sets. In *Current Issues in Statistical Inference: Essays in Honor of D. Basu*, M. Ghosh and P. K. Pathak, eds. Hayward, CA: IMS Lecture Notes–Monograph Series, vol. 17, 1–12.

Casella, G. and R. L. Berger (1990). *Statistical Inference.* Pacific Grove, CA: Wadsworth.

Cohen, A. (1972). Improved confidence intervals for the variance of a normal distribution. *Journal of the American Statistical Association*, **67**, 382–387.

Cox, D. R. (1971). The choice between ancillary statistics. *Journal of the Royal Statistical Society, Series B*, **33**, 251–255.

Cox, D. R. and D. V. Hinkley (1974). *Theoretical Statistics.* London: Chapman and Hall.

Das, K. (1992). Improved estimation of the ratio of variance components for a balanced one-way random effects model. *Statistics and Probability Letters*, **13**, 99–108.

Dawid, A P., M. Stone and J. V. Zidek (1973). Marginalization paradoxes in Bayesian and structural inference (with discussion). *Journal of the Royal Statistical Society, Series B*, **35**, 189–233.

Donner, A. (1986). A review of inference procedures for the intraclass correlation coefficient in the one-way random effects model. *International Statistical Review*, **54**, 67–82.

Fisher, R. A. (1934). Two new properties of mathematical likelihood. *Proceedings of the Royal Society of London, A*, **144**, 285–307.

Fisher, R. A. (1935). The fiducial argument in statistical inference. *Annals of Eugenics*, **4**, 391–398.

Fisher, R. A. (1956). *Statistical Methods and Scientific Inference*, 3rd edition. New York: Hafner.

Fraser, D. A. S. (1968). *The Structure of Inference.* New York: Wiley.

Fraser, D. A. S. (1979). *Inference and Linear Models.* New York: McGraw-Hill.

Fraser, D. A. S. (1983). Structural inference. In *Encyclopedia of Statistical Science*, N. L. Johnson and S. Kotz, eds., vol. 4, 105–114.

Fraser, D. A. S. (1996). Comment on "Pivotal inference and the fiducial argument" (Barnard, 1995). *International Statistical Review*, 64, 231–235.

Girschick, M. A. and L. J. Savage (1951). Bayes and minimax estimates for quadratic loss functions. *Proceedings from the Second Berkeley Symposium in Mathematical Statistics and Probability*, 53–73. University of California Press.

Goutis, C. (1989). *Improved set estimation of a normal variance with generalizations.* Ph.D. dissertation, Cornell U.

Goutis, C. and G. Casella (1991). Improved invariant confidence intervals for a normal variance. *Annals of Statistics*, **19**, 2015–2031.

Goutis, C. and G. Casella (1995). Frequentist post-data inference. *International Statistical Review*, **63**, 325–344.

Graybill, F. A. (1976). *Theory and Application of the Linear Model.* North Scituate, MA: Duxbury.

Graybill, F. A. and R. K. Burdick (1992). *Confidence Intervals on Variance Components.* New York: Marcel Dekker.

Healy, M. J. R. (1963). Fiducial limits for a variance component. *Journal of the Royal Statistical Society, Series B*, **25**, 128–130.

Hodges, J. L., Jr. and E. L. Lehmann (1951). Some applications of the Cramèr-Rao inequality. *Proceedings from the Second Berkeley Symposium in Mathematical Statistics and Probability*, 13–22. University of California Press.

James, W. and C. Stein (1961). Estimation with quadratic loss. *Proceedings from the Fourth Berkeley Symposium in Mathematical Statistics and Probability*, 13–22. University of California Press.

Kiefer, J. (1975). Conditional confidence approach in multidecision problems. In *Proceedings of the Fourth Dayton Multivariate Conference* (P. R. Krishnaiah, ed.). Amsterdam: North Holland, 143–158.

Kiefer, J. (1976). Admissibility of conditional confidence procedures. *Annals of Statistics*, **4**, 836–865.

Kiefer, J. (1977). Conditional confidence statements and confidence estimators (with discussion). *Journal of the American Statistical Association*, **72**, 789–827.

Kiefer, J. (1983). Conditional inference. In *Encyclopedia of Statistical Science*, vol. 4, N. L. Johnson and S. Kotz, eds.

Kubokawa, T. (1994). A unified approach to improving equivariant estimators. *Annals of Statistics*, **22**, 290–299.

Kubokawa, T. (1997). The Stein phenomenon in simultaneous estimation: a review. *Discussion Paper Series 97-F-14, Faculty of Economics, University of Tokyo.*

Kubokawa, T. (1998). Shrinkage and modification techniques in estimation of variance and the related problems: a review. *Discussion Paper Series CIRJE-F-6, Faculty of Economics, University of Tokyo.*

Kubokawa, T., A. K. Md. E. Saleh and Y. Konno (1998). Bayes, minimax and nonnegative estimators of variance components under Kullback-Leibler loss. *Discussion Paper Series CIRJE-F-19, Faculty of Economics, University of Tokyo.*

Lawless, J. (1982). *Statistical Models and Methods for Lifetime Data.* New York: Wiley.

Lehmann, E. (1986). *Testing Statistical Hypotheses, 2^{nd} Edition.* New York: Wiley.

Lehmann, E. and G. Casella (1998). *Theory of Point Estimation, Second Edition.* New York: Springer-Verlag.

Loh, W. (1986). Improved estimators for ratios of variance components. *Journal of the American Statistical Association*, **81**, 699–702.

Maatta, J. M. and G. Casella (1987). Conditional properties of interval estimators of the normal variance. *Annals of Statistics*, **15**, 1372–1388.

Maatta, J. M. and G. Casella (1990). Developments in decision-theoretic variance estimation. *Statistical Science*, **5**, 90–120 (with discussion).

Mayo, D. G. (1981). In defense of the Neyman-Pearson theory of confidence intervals. *Philosophy of Science*, **48**, 269–280.

Meng, X. (1994). Posterior predictive p-values. *Annals of Statistics*, **22**, 1142–1160.

Nelder, J. A. (1977). A reformulation of linear models. *Journal of the Royal Statistical Society, Series A*, **140**, 48–77 (with discussion).

Olshen, R. A. (1973). The conditional level of the F test. *Journal of the American Statistical Association*, **68**, 692–698.

Proskin, H. M. (1985). *An admissibility theorem with applications to the estimation of the variance of the normal distribution.* Ph.D. dissertation, Rutgers U.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications, Second Edition.* New York: Wiley.

Robinson, G. K. (1979a). Conditional properties of statistical procedures. *Annals of Statistics*, **7**, 742–755.

Robinson, G. K. (1979b). Conditional properties of statistical procedures for location and scale parameters. *Annals of Statistics*, **7**, 756–771.

Robinson, J. (1965). The distribution of a general quadratic form in normal variates. *Australian Journal of Statistics*, **7**, 110–113. (Corrigenda in **8**, p. 180).

Savage, J. (1962). Subjective probability and statistical practice. In *The Foundations of Statistical Inference: A discussion opened by Professor L. J. Savage at a meeting of the Joint Statistics Seminar, Birkbeck and Imperial Colleges, University of London.* London: Methuen.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, **2**, 110–114.

Searle, S. R., G. Casella and R. McCulloch (1992). *Variance Components.* New York: Wiley.

Seidenfeld, T. (1979). *Philosophical Problems of Statistical Inference.* Dordrecht: Reidel.

Seidenfeld, T. (1981). On after-trial properties of best Neyman-Pearson confidence intervals: Discussion of Mayo (1981). *Philosophy of Science*, **48**, 281–291.

Shinozaki, N. (1995). Some modifications of improving estimators of a normal variance. *Annals of the Institute of Statistical Mathematics*, **47**, 273–286.

Shorrock, G. (1990). Improved confidence intervals for a normal variance. *Annals of Statistics*, **18**, 972-980.

Smith, D. W. and L. W. Murray (1984). An alternative to Eisenhart's model II and mixed model in the case of negative variance estimates. *Journal of the American Statistical Association*, **79**, 145–151.

Strawderman, W. E. (1974). Minimax estimation of powers of the variance of a normal population under squared error loss. *Annals of Statistics*, **2**, 190–198.

Strawderman, W. E. (1990). Discussion of Maatta and Casella (1990), pp. 117–118.

Stein, C. (1964). Inadmissibility of the usual estimator for the variance of a normal distribution with an unknown mean. *Annals of the Institute of Statistical Mathematics*, **16**, 155–160.

Tate, R. F. and G. W. Klett (1959). Optimal confidence intervals for the variance of the normal distribution. *Journal of the American Statistical Association*, **54**, 674–682.

Tukey, J. W. (1951). Components in regression. *Biometrics*, **7**, 33–69.

Venables, W. and A. T. James (1978). Interval estimates for variance components. *Canadian Journal of Statistics*, *6*, 103–111.

Wang, C. M. (1990). On the lower bound of confidence coefficients for a confidence interval on variance components. *Biometrics*, **46**, 187–192.

Weerahandi, S. (1994). *Exact Statistical Methods in Data Analysis.* New York: Springer-Verlag.

Welsh, A. H. (1996). *Aspects of Statistical Inference.* New York: Wiley.

Williams, J. S. (1962). A confidence interval for variance components. *Biometrika*, **49**, 278–281.

Wild, C. J. (1981). Interval estimates for variance components. *Canadian Journal of Statistics*, *9*, 195–201.

Wolpert, R. L. (1988). Discussion of Casella (1988). In *Statistical Decision Theory and Related Topics IV* (S. S. Gupta and J. O. Berger, eds). New York: Springer, 104–110.