Bayesian Analysis of Binary and Count Data in Two-arm Trials

by

CYNTHIA KPEKPENA

A Thesis submitted to
the Faculty of Graduate Studies
In Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE

Department of STATISTICS
University of Manitoba
Winnipeg, Manitoba

Abstract

Binary and count data naturally arise in clinical trials in health sciences. We consider a Bayesian analysis of binary and count data arising from two-arm clinical trials for testing hypotheses of equivalence. For each type of data, we discuss the development of likelihood, the prior and the posterior distributions of parameters of interest. For binary data, we also examine the suitability of a normal approximation to the posterior distribution obtained via a Taylor series expansion.

When the posterior distribution is complex and high-dimensional, the Bayesian inference is carried out using Markov Chain Monte Carlo (MCMC) methods. We also discuss a meta-analysis approach for data arising from two-arm trials with multiple studies. We assign a Dirichlet process prior for the study effects parameters for accounting heterogeneity among multiple studies. We illustrate the methods using actual data arising from several health studies.

Acknowledgment Page

Dedication Page

I dedicate this research to the memory of my late
father and to my mother.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1   Binary Data in Two-arm Trials

An arm is a standard term for describing clinical trial and it represents
a treatment group or a set of subjects. A single-arm study involves only
one treatment where as the normal two-arm study compares a drug with
a placebo or drug A with drug B. A binary outcome is an outcome whose
unit can take on only two possible states "0" and "1". Health studies
outcomes such as the morbidity and mortality studies are often binary in
nature.

As an example, consider a clinical trial, where a pharmaceutical com-
pany wants to test a new drug against a currently existing drug. The clin-
ical trial end point is the binary success or failure of the treatment. This
success/failure response variable could be heart disease (Yes/No), patient
condition (Good/Critical), how often patient feel depressed (Never/Often)
etc.

The natural distribution for modeling these types of binary data is the binomial distribution. The binomial is a discrete probability distribution that summarizes the likelihood that a random variable will take one of two independent values under a given set of parameters and assumptions. It is assumed that there are only two outcomes (denoted 'success' or 'failure') and a fixed number of trials (n). The trials are independent with a constant probability of success.

The probability mass function for the binomial random variable is given as:

$$f(x;p) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{for} \quad x = 0, 1, \ldots, n, \quad p \in (0,1).$$

The mean and variance for the binomial random variable are $E(X) = np$ and $Var(X) = np(1-p)$ respectively.

1.2 Count Data in Two-arm Trials

Count data refers to the occurrence of observations that can take only the non-negative integer values $\{0, 1, 2, 3, ...\}$, and these integers arise from counting rather than ranking (data composed of counts of the number of events occurring within a specific observation period). When data are not dominated by zeros, it is reasonable to assume such count data as continuous and fit the usual linear models. However, real world count variables such as the number of accidents on a particular spot on a highway, the number of fish in a pond etc. are bound to be characterised by excessive zero values, often called zero-inflated.

In clinical trials, observations are sometimes in the form of counts, for example, in an anti-viral therapeutic vaccine efficacy study, subjects are assessed every day for viral shedding during the study follow-up period and the number of seizures in epileptic patients during a follow-up period. In these instances, only counts of the number with the attribute of interest is taken but not the number without the attributes. The natural distribution for modeling these type of count data is Poisson distribution. This is a discrete distribution used to model the count of a specified event in a given time interval. The assumptions underlying the Poisson distribution are that:

- The number of events in disjoint intervals are independent of each other

- The probability distribution of the number of events counted in any time interval only depends on the length of the interval

- Events cannot be simultaneous

The probability mass function of the Poisson random variable is

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} \quad \text{for} \quad x = 0, 1, \ldots, \lambda > 0.$$

The expection of the Poisson random variable is $E(X) = \lambda$ and the variance is $\text{Var}(X) = \lambda$.

1.3 Hypothesis Testing in Two-arm Trials

The main objective of a clinical trial is to determine whether there is a significant difference between active treatment (new drug) and reference treatment (current drug). Tests of significance has generally been argued not to be enough. That is, if the p-value for a test of significance leads to the non-rejection of the null hypothesis, it is not a proof that the null hypothesis holds. In other words, lack of significance does not imply the two treatments are equivalent. The clinician may want to test hypothesis of a relevant difference or a hypothesis stating one treatment is not lower in standard than the another. To establish the credibility of the null hypothesis, post hoc tests of treatment means have to be conducted. These post hoc test could be formulated in terms of a null hypotheses of equivalence against an alternative hypothesis that states that there is a sufficient difference between the two drugs.

Equivalence testing is widely used when a choice is to be made between a drug (or a treatment) and an alternative. The term equivalence in the statistical sense is used to mean a weak pattern displayed by the data under study regarding the underlying population distribution. Equivalence tests are designed to show the non-existence of a relevant difference between two treatments. It is known that the Fisher's one sided exact test is the same as the test for equivalence in the frequentist approach [26]. This testing procedure is similar to the classical two sided test procedure but involves an equivalence zone determined by equivalence margin (δ) explained in section 1.4.

Noninferiority test on the other hand are designed to show that a new treatment does not fall short in efficacy by some clinically acceptable amount when compared to some existing treatment. The objective is to establish that the new treatment is no worse than the standard already existing. This means the new treatment measures up to the stated standard (not lower in standard than the current drug usually by a margin). Noninferiority test are formulated by placing an upper limit on the difference in treatment means [19].

For example, multiple injections that used to characterise polio vaccinations usually resulted in side effects. An alternative could be a vaccine that combines all the active ingredients of the individual vaccines. Then, it will have to be investigated that the mixture vaccine is as effective as each of the individual vaccines. In another instance, the innovator of a drug with a patent right may come up with a different formulation of the drug with the same ingredients in the innovated drug. At this time the drug is about to be out for competition, other manufacturers may claim their product perform equally well as the innovated drug. The manufacturers different formulation of the drug together with the other products constitute alternatives to the innovated drug. Each of these alternatives require the proof of equivalence of average bioavailabilities(ABE). The concept of bioavailability refers to the rate and extent by which the drug is available at its site of action [19].

1.4 The Equivalence Margin

The equivalence margin (δ), which represents a margin of clinical indifference, is usually estimated from previous studies and as such is also based primarily on clinical criteria as well as statistical principle. It is influenced by statistical principle but largely dependent on the interest of the experimenter and research questions clinicians wish to answer. As such, the statistical method employed together with the design of the study must be in such a manner that the margin of difference is not too restrictive to capture the bounds of the research question. This is usually chosen to be a value less than the least expected disparity between the new treatment and a placebo. For a test of equivalence of two binomial proportions, the equivalence margin is discussed in [26]. When the goal is to establish that one treatment is not equivalent to the other, the equivalence margin has been presented as a fraction f of the lower limit of a confidence interval for the difference in treatment means, but the choice of f is a matter of clinical judgment and also overall benefit-cost and benefit-risk assessment [14]. The frequentist approach to equivalence testing is the two one-sided test (TOST) procedure. By the TOST, equivalence is established at the α significance level if a $(1-2\alpha)\times100\%$ confidence interval for the difference in treatment means $\mu_i - \mu_j$ is contained within the interval $(-\delta, \ \delta)$ where δ is the equivalence margin. For a generic drug (G) and an Active Comparator (A), if Δ is the population treatment group difference ($\Delta = A - G$), $d^\star$ is a threshold of clinical meaningfulness and δ the non-inferiority margin,

G is clinically superior to A if $\Delta > d^\star$ and A is clinically superior to G if $\Delta < -d^\star$. G is inferior to A if $A - G < \delta$ and A is non-inferior to G if $A - G > -\delta$ [7].

1.5 Bayesian Model Ingredients

1.5.1 The Prior

The Statistical inferential procedure is similar to an inversion method where the "cause" (parameters) are extracted from the "effects" (data) [25]. The parameter represents a true state of nature whose value is usually unknown and cannot be observed directly. In the usual classical paradigm, the parameter of interest θ is assumed to be fixed (some constant value) where as in the Bayesian paradigm the parameter is assumed to vary (random in nature). For instance in estimating the recovery rate of a patient, it is natural to assume the rate varies depending on several other factors. This implies θ is a random variable and therefore has a distribution $\pi(\theta)$, called the prior. If the distribution of θ depends on another parameter τ, then the prior is $\pi(\theta|\tau)$, where the parameter τ is called a hyperparameter.

The prior distribution of θ reflects previous knowledge about the parameter θ. The prior could be noninformative or subjective. An informative prior gives a numerical information specific to the problem under consideration. Prior distributions that are uniform with the intention of bringing out the information from the likelihood in probabilistic terms are noninformative. For example, for the variance parameter σ^2 of a normal

distribution for data in which the variability is low, a prior distribution proportional to the inverse of σ^2 is appropriate. This distribution summarizes available prior information in the form of an appropriately chosen probability distribution or mass function. As another example, the probability of success (p) in Bernoulli trials lies between 0 and 1 and therefore an appropriate prior will be a density whose support lies in the range [0, 1], for instance the Beta distribution or the Uniform(0, 1) distribution [25]. Prior distributions that do not provide contradicting information but are capable of suppressing inaccurate deductions not reflected by the likelihood are weakly informative prior. A subjective prior is the Statistician's best judgment about the uncertain parameters in a problem expressed in scientific terms [9].

Conjugate Priors: If the posterior distribution (explained in section 1.5.3) p are in the same family as the prior probability distribution p, the prior and posterior are called conjugate distributions, and the prior is called a conjugate prior for the likelihood. Conjugate priors lead to posterior distributions that belong to the same family as the prior and are analytically tractable.

1.5.2 The Likelihood

The idea of likelihood denotes that, there is some data (observed responses) for which we want to make statements (generalise) about some unknown characteristics. Making inference about the parameter θ requires a probability model. That is a description of values of the parameter that are most

possible in parametric form considering the observed data. Some values of the parameter θ are more likely to produce the data than others are and will be advisable to make inference about those values and the likelihood can be thought of as a means of measuring the relative plausibility of various values of θ by comparing their likelihood ratios [10].

Suppose a parametric model $f(x; \theta)$ is being considered, which is the probability density function with respect to a suitable measure for a random variable X. If the parameter is assumed to be k-dimensional and the data are assumed to be n-dimensional, sometimes representing a sequence of independent identically distributed random variables: $X = (X_1, ... X_n)$, then the likelihood function represented by $L(\theta)$ [22] is given by

$$L(\theta) = L(\theta; x) = \prod_{i=1}^{n} f(x_i; \theta).$$

From the frequentist perspective, the parameter θ is assumed to be some fixed value and data x is assumed to be one realisation of the random variable X. Inference about θ involves calculating relevant summary statistic (about θ without loss of substantial information) which can be used to test hypothesis [12]. "Although the use of likelihood as a plausibility scale is sometimes of interest, probability statements are usually preferred in applications. The most direct way to obtain these is by combining the likelihood with a prior probability function for θ to obtain a posterior probability function" [22].

1.5.3   The Posterior Distribution

The posterior distribution portrays the present state of affairs concerning the unknown parameters. It is the updated state of the prior knowledge by the observed data including missing, latent, and unobserved potential data. The posterior distribution has its source from the Bayes Theorem which states that for two events A and B, the conditional probability A given B is defined as

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$$

Let $X_1, X_2, \ldots X_n$ be a random sample from $f(x|\theta)$ and $\pi(\theta)$ be the prior of θ. The conditional distribution of θ given x, denoted by $\pi(\theta|x)$ is called the posterior distribution of θ. Based on the Bayes Theorem, the posterior distribution is

$$\pi(\theta|x) = \frac{L(x|\theta)\pi(\theta)}{\displaystyle\int L(\theta|x)\pi(\theta)d\theta}. \tag{1.1}$$

The denominator term in 1.1 is known as the normalizing constant.

1.6   Meta-analysis in Clinical Trials

Meta-analysis includes the systematic methods which use statistical techniques for combining results from several independent studies and the aim is to get a consistent estimation of the global effect of an intervention or treatment [6]. A meta-analysis combines in a single conclusion the results

of different studies conducted on the same topic and with the same methods [11]. The most prominent area in which meta-analysis is being used is genetics and health research. When it comes to health issues, everyone is interested in what works and what does not [27] and, meta-analysis, when well designed and appropriately performed, is a great tool that helps in understanding the results of interventions in medicine. The updating of clinical topics through the publication of medical reviews and guidelines shows the need for clinicians to practice evidence-based medicine. Evidence-based medicine has introduced well-defined rules for the critical evaluation of medical data. The use of meta-analysis has a prominent role in the validation and interpretation of the results of clinical studies. In other words, if a well designed and well conducted meta-analysis has shown that drug A is more effective than drug B, we can assume that this information is correct and there would be no need for further investigation on this issue"[11].

In medicine, the effect size is called treatment effect but is simply called effect size in other fields such as the Arts. The term effect size is appropriate when the index is used to quantify the relationship between two variables or a difference between two groups (for instance comparing the performance of girls and boys on a subject) whilst treatment effect is appropriate only for an index used to measure the impact of a deliberate intervention, for example the impact of a new malaria drug [2].

The first step is the statement of the research problem in definite terms. The question or the hypothesis of interest guides the researcher on which

studies to choose and also the kind of data that justifies the inclusion of a study in the meta-analysis. Upon stating the problem, the researcher can start with the search for the relevant studies on the topic. This is done through journals, electronic databases and references on articles. The researcher needs to locate studies that have not been published as well to avoid inclusion of only studies that are statistically significant since inclusion of only studies which conclude the treatment improves for instance patient's condition will cause the result of the meta-analysis to be shifted towards significance.

It is believed that studies that are not statistically significant are not published in most cases [6]. When the manufacturer of a drug gives funding to a reseacher to conduct research on the effectiveness of a drug in a given geographical area, if the results conclude that there is no treatment effect, it is likely that only results from other researchers or other geographical locations that are significant will be published. This points to the issue of bias in publication of research articles. Inclusion of the non-published results in the meta-analysis may cause the conclusion drawn from the meta-analysis to change .

Publication bias arises either because there is an already existing assertion and it will be easier publishing results that validate the opinion or authors may consider their results redundant because findings from various studies follow the same trend and people want something new that has been discovered. The author may not be interested in publishing a research that does not produce positive results and the editorial policy of

the journal in which the paper must be published may also be a potential source of bias. Publication bias can be detected by making a funnel plot. This is a plot of effect size (using risk ratios or odds ratios) against the size of each study. If there is no bias in the publication on a topic, then the plot is an inverted funnel. Departure from this pattern indicates the presence of publication bias. The funnel plot, however, is only a graphical tool. The Klein's procedure provides a test on the dependability of the meta-analysis with regard to publication bias. The Klein's procedure is an answer to the question "assuming publication bias is present , how many studies are needed to change the conclusion of the meta-analysis from statistical significance to no treatment effect"[11]. Bias could also result from the search procedure, it is known that the rate at which an expert can identify the relevant studies is between 32% and 80% and this rate is obviously lower for inexperienced users [11]. Access to all the relevant studies depends on the ability of the researcher to search the Internet or other sources to recover all studies on the topic. In addition, if the criteria for inclusion of studies in the meta-analysis is not clearly defined at the start of the research and also if the selection criteria is such that important studies are neglected , the results of the meta-analysis will be biased as well.

A correct systematic review on a topic requires collection and analysis of all published data and not only those which are more interesting, relevant, or easily available - the available literature must be completely covered. The methods used in meta-analysis limit the bias and help improve the reliability (precision) and validates the conclusion made. "In clinical trails

and cohort studies, meta-analysis gives an indication of more events in the groups observed (that is meta-analysis gives an indication of variables that are not of immediate concern). In the absence of meta-analysis, these events of interest and promising leads will be overlooked and researchers will spend time and resources to find solutions to that which had already been addressed elsewhere"[27].

Despite the difficulty that may sometimes be encountered in locating studies to be included in meta-analysis, we have access to information from many studies with less effort and hassle when the search procedure is successful. Money and energy are saved compared to what would have been required in survey planning and data collection and a considerable amount of time is saved as well. Single studies rarely provide answers to clinical questions. Meta-analysis of multiple studies establishes whether the results of different studies on an issue are consistent and can be generalized across populations, settings and treatment variations, or whether findings vary by particular subsets. By pooling studies together by way of weighting, sample size is increased with greater power and it is expected that the estimates from a meta-analysis would be more precise compared to that from single studies. Randomized control trials are presumed to be the best in most cases but findings from different studies based on the randomized controlled design do not necessarily produce similar results [21]. For a treatment, some studies may report the benefits of the treatment while others report its hazards.

1.6.1 Odds Ratios

The effect size of a disease or an intervention drug is usually computed by ratios such as the risk ratio. The Odds ratio is one of the several statistics that is becoming increasingly important in clinical research and decision making. It is particularly useful because as a treatment effect, it gives clear and direct information to clinicians about which treatment approach has the best odds of benefiting the patient. The odds ratio (OR) can be said to be the ratio of two odds and may sometimes provide information on the strength of the relationship between two variables[15]. The odds ratio of a disease (say lung cancer) is the odds of cancer in the exposed group divided by the odds of the cancer in the unexposed group. The odds ratio is usually computed in case control studies - this is where individuals with condition of interest are being compared with similar subjects without conditions (the controls). For example, suppose

- tt is the number of subjects exposed (smoke) and have experienced condition (lung cancer)

- tc is the number of subjects who have experienced condition (lung cancer) in the control group(non-smokers)

- qt is the number of subjects exposed (smoke) but don't have lung cancer

- qc is number of subjects in the control group who does not have lung cancer

Then the odds of lung cancer in the exposed group is $\dfrac{tt}{qt}$. The odds of cancer in the control group is $\dfrac{tc}{qc}$. Then odds ratio of having cancer is $\dfrac{tt}{qt} \Big/ \dfrac{tc}{qc}$.

When the odds ratio is less than 1, the risk is less likely in the exposed group and if it is greater than 1, the risk is more likely in the exposed group. An odds ratio of 0.75 means that the outcome of interest is 25% less likely in the exposed group. An odds ratio 1 indicates no difference and is called the null value. Examples of the odds ratio are: the Likelihood Ratio Chi-Square, Fishers Exact Probability test and the Pearson Chi-Square.

In Meta-analysis, individual studies will have respective odds ratios calculated $(OR_1, OR_2, \dots)$, then the combined odds ratio can be calculated by different methods:

Mantel-Haenszel method: Let the approximated variance from each study be V_i and associated weights $W_i = \frac{1}{V_i}$. Then by the Mantel-Haenszel [8] method, the combined odds ratio is

$$OR_{MH} = \frac{(OR_1 * W_1) + (OR_2 * W_2) + \cdots + (OR_k * W_k)}{W_1 + W_2 + \cdots + W_k} \qquad (1.2)$$

The chi-square test statistic under the Mantel-Haenszel method is given as

$$Q = \sum_i^k W_i (\ln OR_i - \ln OR_{MH}).$$

The Peto method: The Peto method gives confidence interval that covers the combined odds ratio. Suppose V_i is the variance corresponding

to study i . For each study, the expected frequency(E_i) of each cell is obtained. Then the natural logarithm of the odds ratio of the ith study is

$$L_n\text{OR}_i = \frac{\text{sum of (observed - expected)}}{\text{sum of the variances}} \quad \text{and } OR_i = exp(L_n \ OR_i).$$

The $(1-\alpha)$ % confidence interval for the pooled odds ratio is

$$exp\left(OR_i \pm \frac{Z_{\frac{\alpha}{2}}}{\sqrt{\sum_i^k V_i}}\right).$$

The chi-square test Statistic when odds ratios are calculated by the Peto method is

$$Q = \sum \left[w_i * (O_i - E_i)^2\right] - \frac{\sum(O_i - E_i)^2}{\sum V_i}.$$

1.7 Organization of the Thesis

The motivation for this thesis is based on the fact that for a given disease, there is likely to be many other substitute drugs or new drugs that can be used to treat the patients. But these drugs may not all be at the same cost, some may possibly have adverse side effects and the method of application could be complex for others. On grounds of these information, we do equivalence testing to see if two different drugs can be regarded as equivalent in terms of the their treatment effect. A meta-analysis would answer the question of whether on a large scale or in the long run the drug will be beneficial.

The remaining section of this thesis is organized as follows. In Chapter 2, the inferential procedures for binary and count data are discussed.

Chapter 3 presents the statistical models and the analytic procedures in Meta-analysis as well as a review of the Dirichlet process. In Chapter 4, data on counts of the number of people experiencing myocardial infarction from the use of drugs with an active ingredient "rosiglitazone" is analyzed by testing hypothesis about the binomial proportions as well as multiple determination of treatment effects through Meta-analysis. A count data model is then considered. Chapter 5 presents a discussion of the results and conclusions.

As future work, we will be interested in exploring Network meta-analysis and the methods involved. This is a meta-analysis in which multiple treatments are compared in multivariate analysis.

Chapter 2

Statistical Models

2.1 Statistical Inference for Binary Data

Let X_t be the number of individuals with positive exposure out of a total
of n_t patients in treatment group with proportion P_t. Accordingly, let X_c
denote the number of individuals with positive exposure out of a total n_c
in the control group with proportion P_c. Then

$$X_t \sim \text{Bin}(n_t, P_t) \quad \text{and}$$

$$X_c \sim \text{Bin}(n_c, P_c).$$

The priors on the parameters, P_t and P_c are given by

$$P_t \sim \text{Beta}(\alpha, \beta) \quad \text{and}$$

$$P_c \sim \text{Beta}(\epsilon, \eta).$$

The posterior distribution of P_t is given by:

$$\pi(P_t|X_t) \propto L(X_t|P_t)\pi(P_t)$$

$$\propto \binom{n_t}{x_t} P_t^{x_t}(1-P_t)^{n_t-x_t} \frac{1}{B(\alpha,\beta)} P_t^{\alpha-1}(1-P_t)^{\beta-1}$$

$$\propto \binom{n_t}{x_t} \frac{1}{B(\alpha,\beta)} P_t^{x_t+\alpha-1}(1-P_t)^{n_t+\beta-x_t-1}$$

$$\propto \text{Beta}(x_t+\alpha, \ n_t+\beta-x_t)$$

Similarly, the posterior distribution of P_c is

$$\pi(P_c|X_c) \propto L(X_c|P_c)\pi(P_c)$$

$$\propto \binom{n_c}{x_c} P_c^{x_c}(1-P_c)^{n_c-x_c} \frac{1}{B(\epsilon,\eta)} P_c^{\epsilon-1}(1-P_c)^{\eta-1}$$

$$\propto \binom{n_c}{x_c} \frac{1}{B(\epsilon,\eta)} P_c^{x_c+\alpha-1}(1-P_c)^{n_c+\eta-x_c-1}$$

$$\propto \text{Beta}(x_c+\epsilon, \ n_c+\eta-x_c)$$

For Bayesian inference about treatment effect, a test is required to determine whether the posterior probability of treatment proportions P_t and P_c lies within the bounds of the equivalence margin or not. There is therefore, the need to sample from the posterior distribution of $P_t - P_c$. The marginal posteriors of P_t and P_c are Beta distributions and therefore $\pi(P_t-P_t|X_t, X_c)$ is not in an analytically tractable form. So, $P_{1t}, P_{2t}, \ldots P_{nt}$ are generated from $\pi(P_t|X_t)$ and independently $P_{1c}, P_{2c}, \ldots P_{nc}$ generated from $\pi(P_c|X_c)$ because λ_t and λ_c are independent. Then $P_{1t} - P_{1c}, \ P_{2t} - P_{2c}, \ \ldots, P_{nt}-P_{nc}$ can be treated as a random sample from $\pi(P_t-P_c|X_t, X_c)$.

2.2 Normal Approximation to the Beta Posterior Distribution

Our posterior distributions of P_t, P_c are Beta distributions. A normal approximation to posteriors can be obtained using a Taylor series expansion of the Beta distribution. We derive this approximation as follows: Let the best estimate of P, P_0 be the value of P for which the posterior is at it's maximum. That is,

$$\frac{d\pi(P|x)}{dp}|_{P_0} = 0 \quad \text{and}$$

$$\frac{d^2\pi(P|x)}{dP^2}|_{P_0} < 0$$

The Taylor series expansion of a function $f(x)$ at $X = x_0$ is

$$f(x) = \sum_{m=0}^{\infty} \frac{f^m(x_0)}{m!}(x - x_0)^m$$

Let the log of the posterior distribution be

$$L(P) = \log(\pi(P|X)).$$

By applying a Taylor series expansion to $L(P)$ at P_0 with first three terms,

$$L(P) = L(P_0) + \frac{dL(P)}{dP}|_{P_0}(P - P_0) + 1/2\frac{d^2L(P)}{dP^2}|_{P_0}(P - P_0)^2 + \dots$$

$$= \text{constant} + 1/2\frac{d^2L(P)}{dP^2}|_{P_0}(P - P_0)^2 + \dots$$

By taking the exponential of $L(P)$,

$$\pi(P|X) \propto K \exp^{\frac{1}{2}\frac{d^2L(P)}{dP^2}}|_{P_0}(P - P_0)^2$$

where K is a normalising constant.

Let $\mu = P_0$ and $\sigma = \frac{1}{\left[\frac{-d^2 L(P)}{dP^2}|_{P_0}\right]^{1/2}}$. This gives

$$\pi(P|X) \approx \mathcal{N}(\mu, \sigma).$$

$$\pi(P_t|X_t) \sim \text{Beta}(x_t + \alpha, n_t + \beta - x_t)$$

$$\sim P_t^{x_t + \alpha - 1}(1 - P_t)^{n_t + \beta - x_t - 1}$$

$$\implies L(P) = k + (x_t + \alpha - 1)\log P_t + (n_t + \beta - x_t - 1)\log P_t$$

$$\frac{dL(P_t)}{dP_t} = \frac{(x_t + \alpha - 1)}{P_t} - \frac{(n_t + \beta - x_t - 1)}{1 - P_t} = 0$$

$$\implies (1 - P_t)(x_t + \alpha - 1) - P_t(n_t + \beta - x_t - 1) = 0$$

$$\alpha - 1 + x_t + 2P_t - \alpha P_t - n_t P_t - \beta P_t = 0 \quad \text{and}$$

$$2P_t + x_t + \alpha - 1 - \alpha P_t - n_t P_t - \beta P_t = 0$$

$$P_0 = \frac{1 - \alpha - x_t}{2 - \alpha - n_t - \beta}$$

$$\frac{dL(P_t)}{dP_t} = (x_t + \alpha - 1)P_t^{-1} - (n_t + \beta - x_t - 1)(1 - P_t)^{-1}$$

$$\frac{d^2}{dP_t^2}(\pi(P_t|X_t)) = -(x_t + \alpha - 1)P_t^{-2} - [-(-1)(1 - P_t)^{-2}(n_t + \beta - x_t - 1)]$$

$$= \frac{-(x_t + \alpha - 1)}{P_t^2} - \frac{-(n_t + \beta - x_t - 1)}{(1 - P_t)^2}$$

$$1 - P_0 = 1 - \frac{1 - \alpha - x_t}{2 - \alpha - n_t - \beta}$$

$$= \frac{2 - \alpha - n_t - \beta - 1 + x_t + \alpha}{2 - \alpha - n_t - \beta}$$

$$= \frac{1 - n_t - \beta + x_t}{2 - \alpha - n_t - \beta}$$

$$\frac{d^2}{dP_t^2}(\pi(P_t|X_t))|_{P_0} = \frac{(1 - x_t - \alpha)}{\left[\frac{1 - x_t - \alpha}{2 - \alpha - n_t - \beta}\right]^2} + \frac{(1 - n_t - \beta + x_t)}{\left[\frac{1 - n_t - \beta + x_t}{2 - \alpha - n_t - \beta}\right]^2}$$

$$= (1 - x_t - \alpha)\left[\frac{(2 - \alpha - n_t - \beta)^2}{(1 - x_t - \alpha)^2}\right] + (1 - n_t - \beta + x_t)\left[\frac{(2 - \alpha - n_t - \beta)^2}{(1 - n_t - \beta + x_t)^2}\right]$$

$$= \frac{(2 - \alpha - n_t - \beta)^2}{1 - x_t - \alpha} + \frac{(2 - \alpha - n_t - \beta)^2}{1 - n_t - \beta + x_t}$$

$$= (2 - \alpha - n_t - \beta)^2\left[\frac{1 - n_t - \beta + x_t + 1 - x_t - \alpha}{(1 - x_t - \alpha)(1 - n_t - \beta + x_t)}\right]$$

$$= \frac{(2 - \alpha - n_t - \beta)^3}{(1 - x_t - \alpha)(1 - n_t - \beta + x_t)}$$

$$\sigma = \frac{1}{\left[-\frac{d^2}{dP_t^2}(\pi(P_t|X_t))|_{P_0}\right]^{\frac{1}{2}}}$$

$$= \frac{1}{\left[\frac{-(2 - \alpha - n_t - \beta)^3}{(1 - x_t - \alpha)(1 - n_t - \beta + x_t)}\right]^{\frac{1}{2}}}$$

Table 2.1 provides some approximations based on this development. We investigate these approximations in Figures 2.1, 2.2 and 2.3. It is clear that this approximation starts to work well for values of the posterior parameters

from $x + \alpha = 10$ and $n + \beta - x = 10$. However, the approximation is not suitable when Beta posterior parameters are less than 10.

Table 2.1:   Normal Approximation to the Beta Distribution

Exact Distribution	Approximation
Beta(2, 1)	$N(1, \infty)$
Beta(1, 2)	$N(0, \infty)$
Beta(10, 10)	$N(0.5000, 8.4853)$
Beta(5, 1)	$N(1, \infty)$
Beta(1, 5)	$N(0, \infty)$
Beta(2, 2)	$N(0.5000, 2.8284)$
Beta(3, 3)	$N(0.5000, 4.0)$
Beta(2, 4)	$N(0.2500, 4.6188)$
Beta(4, 4)	$N(0.5000, 4.8990)$
Beta(5, 5)	$N(1, 5.6569)$
Beta(30, 20)	$N(0.6042, 14.1673)$
Beta(20, 30)	$N(0.3958, 14.1673)$
Beta(50, 20)	$N(0.7206, 18.3776)$
Beta(20, 50)	$N(0.2794, 18.3776)$

2.3   Statistical Inference for Count Data

Modelling count data is common in clinical trials. When the outcome can take any value $\{0, 1, \dots\}$, one can model these outcomes using a Poisson distribution. The Poisson distribution with parameter λ has the probability mass function

$$P(X|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad \lambda > 0, \quad k = 0, 1, \dots.$$

Classical inference involves obtaining the maximum likelihood estimator of the parameter λ and making statements about it. For reasons of overdispersion, there is the need to investigate whether the data actually follows a Poisson distribution. This is done by a chi-square test.

Let X_t and X_c be the number of counts in the treatment and control groups which are assumed to follow Poisson distributions with probability mass functions $P(\lambda_t)$ and $P(\lambda_c)$. For a Bayesian inference, the parameters λ_t and λ_c are assigned a prior distribution for which the posterior distributions given the observed data are found. The prior distributions $\pi(\lambda_t)$ and $\pi(\lambda_c)$ are both Gamma. The posterior distributions of λ_t and λ_c are derived below:

$$\pi(\lambda_t|X_t) \propto \prod_{i=1}^{n_t} P(X_t|\lambda_t)\pi(\lambda_t)$$

$$\propto \frac{e^{-n_t\lambda_t}\lambda_t^{\sum x_{it}}}{\prod_{i=1}^{n} x_{it}!} \frac{\lambda_t^{\alpha_t-1}\beta_t^{\alpha_t}e^{-\lambda_t\beta_t}}{\Gamma(\alpha_t)}$$

$$\propto \lambda_t^{(\sum x_{it}+\alpha_t-1)}\beta_t^{\alpha_t}e^{-(n_t+\beta_t)\lambda_t}$$

$$\propto \lambda_t^{(\sum x_{it}+\alpha_t-1)}\beta_t^{\alpha_t}e^{-(n_t+\beta_t)\lambda_t}$$

$$\propto Gamma\left(\sum x_{it}+\alpha_t,\ \beta_t+n_t\right).$$

Hence the posterior distribution of λ_t is $\ \mathrm{Gamma}(\sum x_{it}+\alpha_t,\ \ \beta_t+n_t)$. Similarly,

$$\pi(\lambda_c|X_c) \propto \prod_{i=1}^{n_c} P(X_c|\lambda_c)\pi(\lambda_c)$$

$$\propto \frac{e^{-n_c\lambda_c}\lambda_c^{\sum x_{ic}}}{\prod_{i=1}^{n} x_{ic}!} \frac{\lambda_c^{\alpha_c-1}\beta_c^{\alpha_c}e^{-\lambda_c\beta_c}}{\Gamma(\alpha_c)}$$

$$\propto \lambda_c^{(\sum x_{ic}+\alpha_c-1)}\beta_c^{\alpha_c}e^{-(n_c+\beta_c)\lambda_c}$$

$$\propto Gamma(\sum x_{ic}+\alpha_c,\ \beta_c+n_c).$$

25

Hence the posterior distribution of λ_c is Gamma$(\sum x_{ic} + \alpha_c,\ \beta_c + n_c)$. To test the hypothesis of equivalence of the treatment mean λ_t and the control mean λ_c, we require the posterior distribution of $\lambda_t - \lambda_c$ $(\pi(\lambda_t - \lambda_c|X_t, X_c))$ which is not in analytically tractable form. If the marginal posterior distributions of λ_t and λ_c happened to be Normal, then $\pi(\lambda_t - \lambda_c|X_t, X_c)$ will be Normal too. However the marginal posteriors are Gamma and we don't know the form of $\pi(\lambda_t - \lambda_c|X_t, X_c)$. Therefore, $\lambda_t^1, \lambda_t^2, \ldots, \lambda_t^N$ are generated from the marginal posterior distribution of λ_t and another set of values $\lambda_c^1, \lambda_c^2, \lambda_c^3 \ldots, \lambda_c^N$ are independently generated from the marginal posterior distribution of λ_c. Subsequently, generating from $\pi(\lambda_t - \lambda_c|X_t, X_c)$ is the same as taking the differences $\lambda_t^1 - \lambda_c^1,\ \ \lambda_t^2 - \lambda_c^2,\ \ \ldots,\ \ \lambda_t^N - \lambda_c^N$.

2.4   Estimating Missing Data in Arms

Missing data is easily handled in Bayesian inference by treating them as another set of parameters. We estimate the missing values conditioning on the observed data. For example, let $X_1, \ldots X_n$ be a binary random sample from Ber(P) in an arm and suppose that X_m is missing. Let $P \sim Beta(\alpha, \beta)$ and $Y = \sum_{i \neq m}^{n} X_i$. Then the likelihood of the observed data is

$$L(X_{\text{obs}}|P) = \binom{n-1}{y} P^y (1-P)^{n-1-y}.$$

The posterior of P based on the complete data $X = (Y,\ X_m)$ is

$$\pi(P|X) \propto P^{y+x_m}(1-P)^{n-y-x_m} \frac{1}{B(\alpha, \beta)} P^{\alpha-1}(1-P)^{\beta-1}.$$

The full conditionals of P and X_m are

$$\pi(P|y, x_m) \sim \text{Beta}(y + x_m + \alpha, n - y - x_m + \beta)$$

$$\pi(x_m|y, P) \sim \text{Ber}(P).$$

It is easy to generate from these full conditionals in R so P and x_m can be estimated using Gibbs sampling.
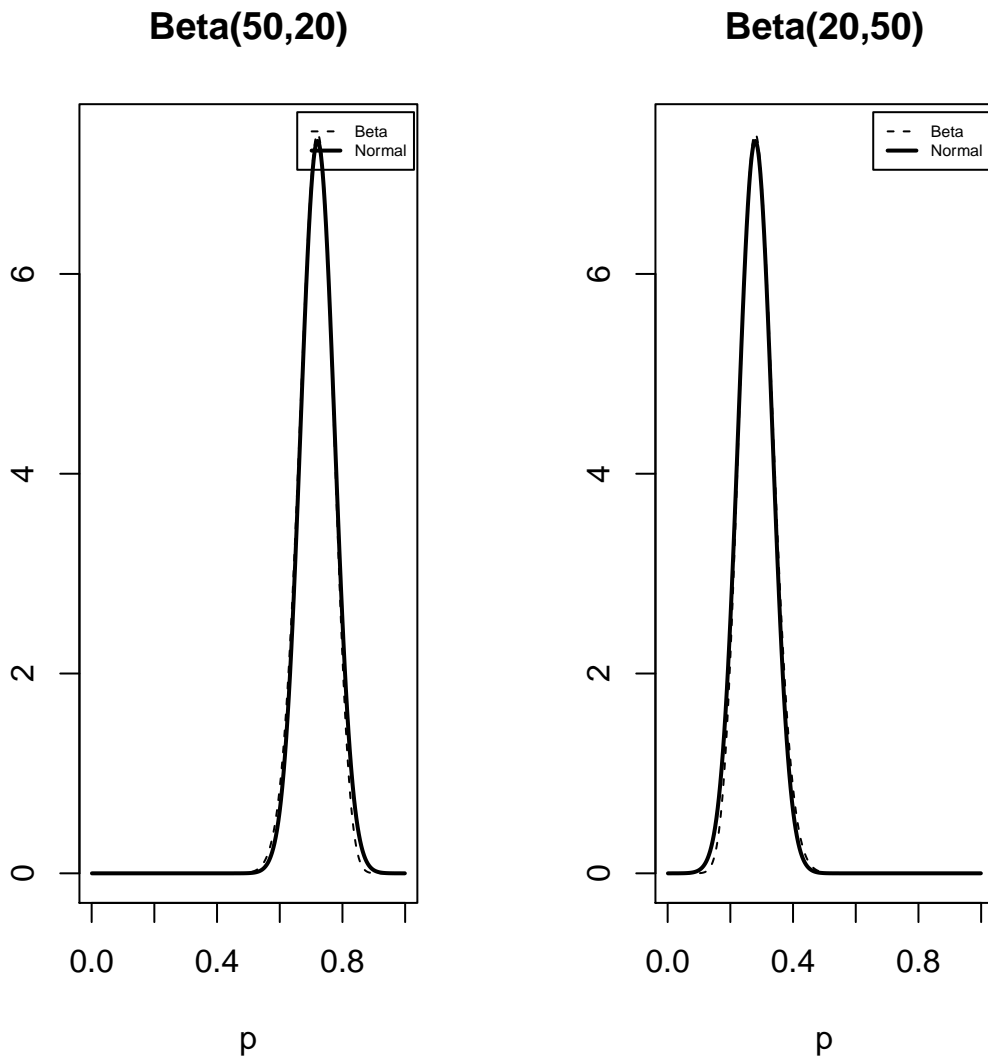
Figure 2.1: The normal approximations for Beta$(50, 20)$ and Beta$(20, 50)$
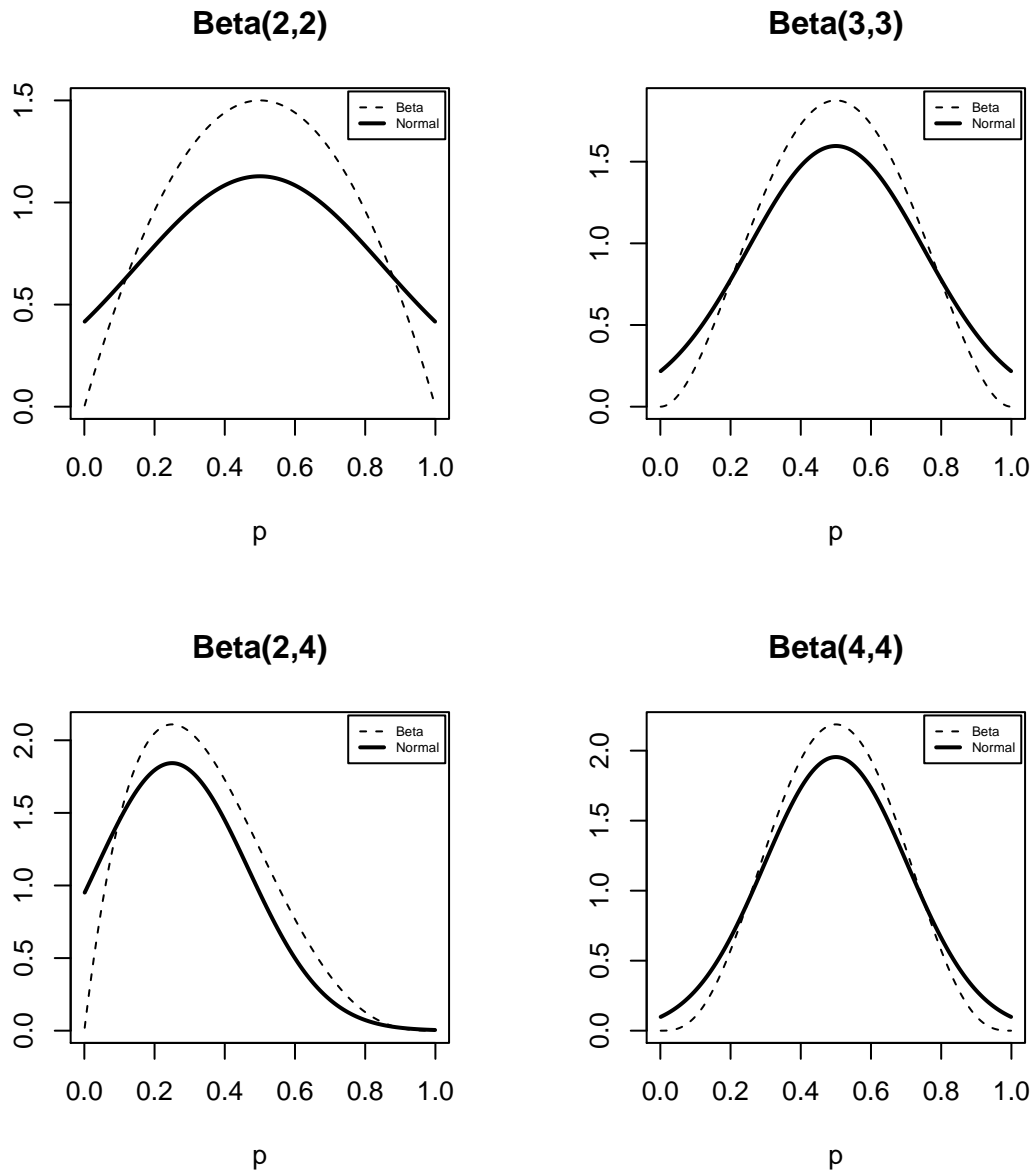
Figure 2.2: The normal approximations of Beta$(2, 2)$, Beta$(3, 3)$, Beta$(2,4)$ and Beta$(4, 4)$
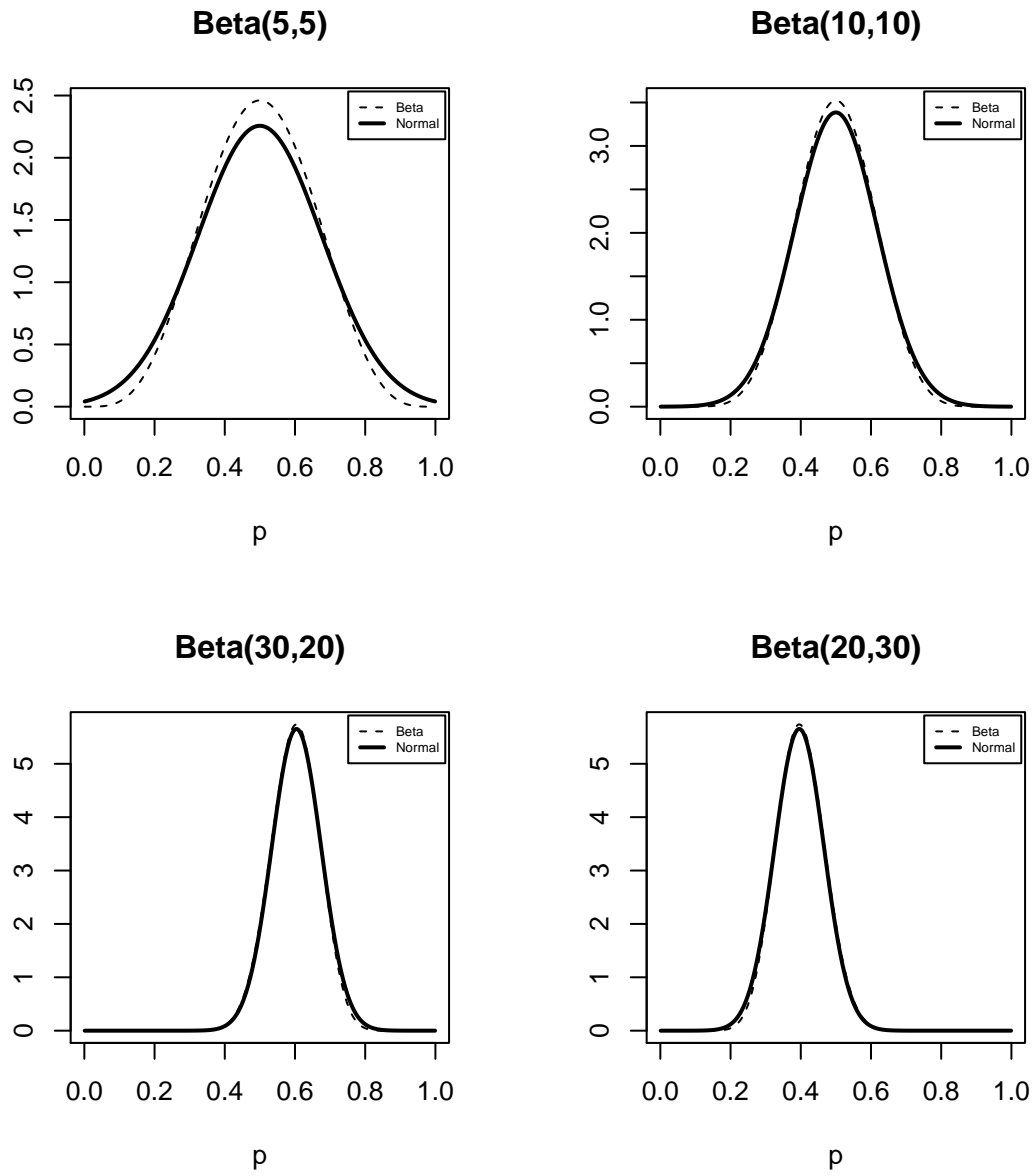
Figure 2.3:   The Normal approximations Beta$(5,5)$,   Beta$(10,10)$, Beta$(30,20)$ and Beta$(20,30)$

Chapter 3

The Meta-analysis Procedure with Multiple Studies

3.1 Fixed Effects and Random Effects Model

The assumption underlying the combined effect (true population) across studies determines whether the model can be classified as Fixed Effects Model (FEM) or Random Effects model (REM) [6].

3.1.1 Fixed Effects Model

The fixed effect model (FEM) is constructed under the assumption that individual study effect sizes can be regarded as estimates of some common effect size (true population effect size) as a whole. That is, estimates can be regarded as coming from the same distribution and the factors that influence effect size are the same [2]. The individual studies in a FEM are believed to be practically alike. It is therefore not possible to generalize conclusions beyond the domain of the studies involved since populations

may differ from the common distribution from which the effect sizes are drawn. Under the assumption that the true effect mean is constant in each study, the observed effect size of the individual studies nevertheless may deviate from the studies true effect mean (this is assumed mainly to be due to sampling error) and this constitutes the within study variance. The true effect size of a study is the effect size in the underlying distribution and is usually unknown. To justify the use of the fixed effect model, there is the need to determine that statistical diversity (heterogeneity) is non-existant among the different studies. Since the FEM is predicated on the assumption that the studies share a common effect, the test of heterogeneity establishes whether the population parameter is constant or not. When the test of heterogeneity is significant(that is we conclude the true effect varies between studies), then the FEM will not be appropriate. The chi-squared test of heterogeneity is one common test used to determine whether the studies in the meta-analysis deal with the same parameter or not. The test of the null hypothesis that all studies share a common effect size is done by comparing the p-value of the Statistic Q (which has a chi -square distribution with degree of freedom $df = k - 1$, where k is the number of studies) with a stated level of significance. The statistic Q is given as

$$Q = \sum_{i=1}^{k} W_i (Y_i - M)^2 \quad \text{where}$$

W_i is the weight (or precision) of the ith study and is calculated as the inverse of the variance of the ith study

Y_i is the ith study effect size

M is an estimate of the true effect size and

k is the total number of studies

Another measure of heterogeneity is I^2, which reflects the proportion of total variability (in effect size) that is real (for instance not due to chance or measurement error). This is calculated as [11]

$$I^2 = \left(\frac{Q - df}{Q}\right) * 100\%.$$

I^2 could be viewed as the ratio of actual heterogeneity to total variability. I^2 is a way of quantifying heterogeneity with values of 25%, 50% and 75% regarded as low, moderate and high respectively. However, I^2 value near zero does not necessarily indicate effects are clustered within a narrow range; the observed effects could be dispersed over a wide range in studies with a lot of error [2]. When the condition for FEM is fulfilled, the combined effect size is the weighted average of individual study effects. The weights corresponding to each study is calculated as $W_i = 1/V_{Y_i}$ where V_{Y_i} is the within-study variance for the ith study . If we let μ represent the combined effect then,

$$\hat{u} = \frac{\sum_{i=1}^{k} W_i Y_i}{\sum_{i=1}^{k} W_i}$$

where Y_i is the ith study effect size.

3.1.2 Random Effects Model

In a REM, the population effect size is assumed to vary from study to study. The studies included in a given meta-analysis may be regarded as being sampled from a universe of possible effects or some parent population [1]. If each study is assumed to have come from a different population, then the estimates of the effect sizes are expected to differ. If it was feasible to perform an infinite number of studies from the different conceivable distributions, "then the effect sizes for the studies will be distributed about some average. The observed effect sizes of trials actually performed are assumed to be a random sample from the effect sizes of the different populations of distributions and the REM is appropriate in this instance"[2]. In most experiments, there may be other variables that influence the response variable but may not be of direct interest. These variables are referred to as covariates. For instance, in an experiment to determine the impact of smoking on lung cancer, other factors such as duration of smoking, family record of lung cancer can have an effect on the outcome. These covariates will definitely vary from study to study and therefore cause variations in the effect size across studies. This introduces randomness in the analysis and the random effects model is appropriate.

If y_i is the estimate of the true effect size μ_i corresponding to the ith study, α_i the random effect of the ith study and the variance of the ith study is $\sigma_i^2 (> 0)$, then the random effects model is given as

$$y_i = \mu + \alpha_i + e_i, \quad i = 1, \ldots, k \tag{3.1}$$

where the study effects α_i are assumed to be different but related. The variation between α_i are assumed to be equal τ^2. The random study effects α_i and the random error term e_i are assumed to be distributed as follows.

$$e_i \overset{i.i.d}{\sim} N(0, \sigma_i^2)$$

$$\alpha_i \overset{i.i.d}{\sim} N(0, \tau^2), \quad i = 1, \ldots, k \tag{3.2}$$

where $N(\theta, \eta^2)$ is a normal random variable with mean θ and variance η^2. The combined effect size in the REM is calculated as the weighted average of individual effect sizes where the weights w_i are inversely related to the ith study variance. Let the variance of the ith study be $V_{Y_i}^\star$, and this has two components. $V_{Y_i}^\star$ is the sum of the within study variance (σ_i^2) and the between study variance. Assuming T^2 is an estimate of the between study variance (τ^2), then

$$V_{Y_i}^\star = \sigma_i^2 + T^2$$

The Dersimonian and Laird method gives the frequentist estimates of the overall mean effect μ and the estimate of the between study variation. The Dersimonian and Laird estimate of the variation between studies is [11]

$$\hat{\tau}_{DL}^2 = \max\left(0, \ \frac{Q - (k-1)}{\sum_{i=1}^{k} W_i - \left(\frac{\sum_{i=1}^{k} W_i^2}{\sum_{i=1}^{k} W_i}\right)}\right)$$

where k is the number of studies, $W_i = \frac{1}{\sigma_i^2}$ and $Q = \sum_{i=1}^{k}\left(y_i - \left(\sum_{i=1}^{k} W_i y_i / \sum_{i=1}^{k} W_i\right)\right)^2$.

When the normality assumption holds, a uniformly minimun-variance unbiased (UMVUE) of μ is given as the weighted average. That is

$$\hat{\mu} = \frac{\sum_{i=1}^{k} w_i^\star y_i}{\sum_{i=1}^{k} w_i^\star} \quad \text{and the variance of the UMVUE is}$$

$$\text{Var}(\hat{\mu}) = \sigma_\mu^2 = \frac{1}{\sum_{i=1}^{k} w_i^\star} \quad \text{where } w_i^\star = \frac{1}{\tau^2 + \sigma_i^2}.$$

The ith study weight estimate $\hat{w}_i^\star = \frac{1}{\hat{\tau}_{DL}^2 + \sigma_i^2}$ and the estimate of μ is given as

$$\hat{\mu}_{DL} = \frac{\sum_{i=1}^{k} \hat{w}_i^\star y_i}{\sum_{i=1}^{k} \hat{w}_i^\star}.$$

In the Bayesian paradigm, parameters are assumed to be random. On the assumption that the study effects $\alpha_1, \alpha_2, \ldots, \alpha_k$ are unknown and random, then the full likelihood function is given as [16]

$$L(\mu, \alpha_1, \alpha_2, \ldots, \alpha_k, |y_1, y_2, \ldots, y_k, \sigma_i^2, \ldots, \sigma_k^2) \propto \prod_{i=1}^{k} \left\{ \frac{1}{(\sigma_i^2)^{\frac{1}{2}}} exp \left(-\frac{(y_i - (\alpha_i + \mu)}{2\sigma_i^2}) \right) \right\}$$

Suppose the prior distributions for μ, $(\alpha_1, \alpha_2, \ldots, \alpha_k)$, and τ^2 are given as

$$\pi(\mu) \propto c, -\infty \le \mu \le \infty$$

$$\alpha_1, \ldots, \alpha_k \overset{iid}{\sim} N(0, \tau^2)$$

$$\tau^2 \sim IG(\eta, \lambda)$$

The conditional posterior probability density functions (p.d.f) of $\mu, (\alpha_1, \alpha_2, \ldots, \alpha_k)$ and τ^2 are given as

$$\mu | \text{rest} \sim N(\mu^\star, \sigma_{\mu^\star}^2) \quad \text{where} \quad \mu^\star = \frac{\sum_{i=1}^{k} w_i(y_i - \alpha_i)}{\sum_{i=1}^{k} w_i}, \sigma_{\mu^\star}^2 = \left(\sum_{i=1}^{k} w_i \right)^{-1}, \quad w_i = \frac{1}{\sigma_i^2}$$

$$\alpha_i | \text{rest} \overset{iid}{\sim} N(\alpha_i^\star, \sigma_{\alpha_i^\star}^2), \alpha_i^\star = \frac{\sigma_i^2 \tau_i^2 (y_i - \alpha_i)}{\tau^2 + \sigma_i^2}, \quad \sigma_{\alpha_i^\star}^2 = \frac{\tau^2 \sigma_i^2}{\tau^2 + \sigma_i^2}, \quad i = 1, \ldots, k;$$

$$\tau^2 | \text{rest} \sim IG(\eta^\star, \lambda^\star), \eta^\star = \eta + \frac{k}{2}, \lambda^\star = \lambda + \frac{1}{2} \sum_{i=1}^{k} \alpha_i^2$$

where conditioning on "rest" implies the other parameters that are not of immediate interest [16]. Note that the model in 3.1 can be reparameterized as follows:

$$y_i = \mu_i + e_i \quad \text{where} \quad e_i \sim N(0, \sigma_i^2). \tag{3.3}$$

Then,

$$Y_i | \mu_i, \sigma_i^2 \sim N(\mu_i, \sigma_i^2)$$

$$\mu_i | \mu, \tau^2 \sim N(\mu, \tau^2)$$

$$\mu | \mu_0, \sigma_0^2 \sim N(\mu_0, \ \sigma_0^2)$$

$$\tau^2 | \eta, \lambda \sim IG(\eta, \ \lambda)$$

We derive the full conditional distributions of this model in the next section.

3.2 Deriving Full Conditional Distributions of Model Parameters in Random Effects Meta-analysis

The full conditional distributions of the parameters conditional on all other parameters are found from the distributions that has information about the parameter of interest. The conditional posterior distribution of μ_i is proportional to the product of the distribution of y_i conditional on μ_i, σ_i^2 and the prior distribution on μ_i.

That is,

$$p(\mu_i|\text{others}) \propto p(Y_i|\mu_i, \sigma_i^2)p(\mu_i|\mu, \tau^2)$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma_i^2}}\right)\exp\left\{\frac{-(y_i - \mu_i)^2}{2\sigma_i^2}\right\}\left(\frac{1}{\sqrt{2\pi\tau^2}}\right)\exp\left\{\frac{-(\mu_i - \mu)^2}{2\tau^2}\right\}$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma_i^2}}\right)\left(\frac{1}{\sqrt{2\pi\tau^2}}\right)\exp\left\{(-\frac{1}{2\sigma_i^2\tau^2})\left[(y_i - \mu_i)^2\tau^2 + \sigma_i^2(\mu_i - \mu)^2\right]\right\}$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma_i^2}}\right)\left(\frac{1}{\sqrt{2\pi\tau^2}}\right) \times$$

$$\exp\left\{(-\frac{1}{2\sigma_i^2\tau^2})[\tau^2(y_i^2 - 2\mu_i y_i + \mu_i^2) + \sigma_i^2(\mu_i^2 - 2\mu\mu_i + \mu^2)]\right\}$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma_i^2}}\right)\left(\frac{1}{\sqrt{2\pi\tau^2}}\right)\exp\left\{(-\frac{1}{2\sigma_i^2\tau^2})\left[(\tau^2 + \sigma_i^2)\mu_i^2 - 2\mu_i(\tau^2 y_i + \mu\sigma_i^2) + \tau^2 y_i^2 + \mu^2\sigma_i^2\right]\right\}$$

$$= \frac{1}{2\pi\sigma_i\tau}\exp\left\{(-\frac{\tau^2 + \sigma_i^2}{2\sigma_i^2\tau^2})\left[\mu_i^2 - 2\mu_i\frac{(\tau^2 y_i + \mu\sigma_i^2)}{\tau^2 + \sigma_i^2} + \frac{\tau^2 y_i^2 + \mu^2\sigma_i^2}{\tau^2 + \sigma_i^2}\right]\right\}$$

Now, consider the exponential term as a quadratic in μ_i below:

$$\mu_i^2 - 2\mu_i \frac{(\tau^2 y_i + \mu\sigma_i^2)}{\tau^2 + \sigma_i^2} + \frac{\tau^2 y_i^2 + \mu^2\sigma_i^2}{\tau^2 + \sigma_i^2}$$

Completing the squares gives

$$\mu_i^2 - 2\mu_i \frac{(\tau^2 y_i + \mu\sigma_i^2)}{\tau^2 + \sigma_i^2} + \left[\frac{(\tau^2 y_i + \mu\sigma_i^2)}{\tau^2 + \sigma_i^2}\right]^2 + \frac{\tau^2 y_i^2 + \mu^2\sigma_i^2}{\tau^2 + \sigma_i^2} - \left[\frac{(\tau^2 y_i + \mu\sigma_i^2)}{\tau^2 + \sigma_i^2}\right]^2$$

$$= \left(\mu_i - \frac{(\tau^2 y_i + \mu\sigma_i^2)}{\tau^2 + \sigma_i^2}\right)^2 + \frac{\tau^2 y_i^2 + \mu^2\sigma_i^2}{\tau^2 + \sigma_i^2} - \left[\frac{(\tau^2 y_i + \mu\sigma_i^2)}{\tau^2 + \sigma_i^2}\right]^2$$

Hence

$$p(\mu_i|rest) \propto \exp\left\{\left(-\frac{\tau^2 + \sigma_i^2}{2\sigma_i^2\tau^2}\right)\left[\mu_i - \frac{(\tau^2 y_i + \mu\sigma_i^2)}{\tau^2 + \sigma_i^2}\right]^2\right\}.$$

Therefore the posterior distribution of μ_i given all the others is

$$\mathcal{N}\left(\frac{\tau^2 y_i + \mu\sigma_i^2}{\tau^2 + \sigma_i^2}, \quad \frac{\sigma_i^2\tau^2}{\tau^2 + \sigma_i^2}\right).$$

The posterior distribution of μ conditional on all the other parameters is derived as follows:

$$p(\mu|rest) \propto \left(\prod_{i=1}^{k} p(\mu_i|\mu, \tau^2) \right) p(\mu)$$

$$\propto \exp \left\{ \frac{-1}{2\tau^2} \sum_{i=1}^{k} (\mu_i - \mu)^2 - \frac{1}{2\sigma_0^2} (\mu - \mu_0)^2 \right\}$$

$$= \exp \left\{ \frac{-1}{2\tau^2} \left[\sum_{i=1}^{k} \mu_i - 2\mu \sum_{i=1}^{k} \mu_i + k\mu^2 \right] - \frac{1}{2\sigma_0^2} [\mu^2 - 2\mu\mu_0 + \mu_0^2] \right\}$$

$$\propto \exp \left\{ -1/2 \left[\mu^2 \left(\frac{k}{\tau^2} + \frac{1}{\sigma_0^2} \right) - 2\mu \left(\frac{\sum_{i=1}^{k} \mu_i}{\tau^2} + \frac{\mu_0}{\sigma_0^2} \right) \right] \right\}$$

$$\propto \exp \left\{ \frac{-1}{2} \left(\frac{k}{\tau^2} + \frac{1}{\sigma_0^2} \right) \left[\mu^2 - 2\mu \frac{\left(\frac{\sum_{i=1}^{k} \mu_i}{\tau^2} + \frac{\mu_0}{\sigma_0^2} \right)}{\frac{k}{\tau^2} + \frac{1}{\sigma_0^2}} \right] \right\}$$

Adding $\left[\dfrac{\left(\frac{\sum_{i=1}^{k} \mu_i}{\tau^2} + \frac{\mu_0}{\sigma_0^2} \right)}{\frac{k}{\tau^2} + \frac{1}{\sigma_0^2}} \right]^2$ gives

$$p(\mu|\text{others}) \propto \exp \left\{ \frac{-1}{2} \left(\frac{k}{\tau^2} + \frac{1}{\sigma_0^2} \right) \left[\mu - \frac{\frac{\sum_{i=1}^{k} \mu_i}{\tau^2} + \frac{\mu_0}{\sigma_0^2}}{\frac{k}{\tau^2} + \frac{1}{\sigma_0^2}} \right]^2 \right\}$$

Hence the posterior distribution of μ given all other parameters is

$$\mathcal{N} \left(\frac{\sigma_0^2 \sum \mu_i + \tau^2 \mu_0}{k\sigma_0^2 + \tau^2}, \frac{\tau^2 \sigma_0^2}{k\sigma_0^2 + \tau^2} \right)$$

40

The posterior distribution of τ^2 is proportional to the product of μ_i conditional on μ, τ^2 and τ^2 conditional on η, λ. That is,

$$p(\tau^2|\text{rest}) \propto \left(\prod_{i=1}^{k} p(\mu_i|\mu,\tau^2)\right) p(\tau^2|\eta,\lambda)$$

$$= \left(\prod_{i=1}^{k} \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{\left(\frac{-1}{2\tau^2}\right)\sum(\mu_i-\mu)\right\}\right)\left(\frac{1}{\tau^2}\right)^{\eta+1}\exp\left(\frac{-\lambda}{\tau^2}\right)$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)\left(\frac{1}{\tau^2}\right)^{(\frac{k}{2}+\eta+1)}\exp\left\{\frac{-(\sum(\mu_i-\mu)^2-2\lambda)}{2\tau^2}\right\}$$

$$\therefore p(\tau^2|\text{rest}) \propto \left(\frac{1}{\tau^2}\right)^{(\frac{k}{2}+\eta+1)}\exp\left\{\frac{-\left(\sum(\mu_i-\mu)^2+2\lambda\right)}{2\tau^2}\right\}$$

Hence the conditional distribution for τ^2 is $IG\left(\frac{k}{2}+\eta, \frac{\sum(\mu_i-\mu)^2+2\lambda}{2}\right)$.

3.3   Markov Chain Monte Carlo (MCMC) Methods

Gibbs Sampling: In the Bayesian paradigm, inference is based on the posterior distribution of θ given the observed data y, where θ is a vector of the parameters of interest. The posterior distribution $p(\theta|y) \propto p(y|\theta)p(\theta)$ can be represented as $f(\theta)$ for fixed y which is the nonnormalised posterior density [17].

Gibbs sampling is a simulation technique employed to sample from the nonnormalised posterior density in order to make inference in the Bayesian framework. The Gibbs sampling procedure is based on the Markov chain

monte carlo methods via full conditional distributions of parameters. Markov chain monte carlo (MCMC) methods are a class of algorithms for sampling from probability distributions based on constructing a Markov chain that has the desired distribution(the posterior density) as its equilibrium distribution. A Markov chain denotes a sequence of random variables $\theta^1, \theta^2, \ldots$, for which, for any t, the distribution of θ^t given all previous θ's depends only on the most recent value, θ^{t-1} [9]. "In the applications of Markov chain simulation, several independent sequences of simulation draws are created; each sequence, θ^t, $t = 1, 2, 3, \ldots$ is produced by starting at some point θ^0 and then, for each t, drawing θ^t from it's full conditional distribution" [9].

Practical problems present situations in which it is not possible to sample directly from the posterior distribution $p(\theta|y)$ and as such MCMC sampling only approximates the target distribution. Sampling is carried out in a manner in which at the long-run the distribution of the sample coincides with the target distribution, in particular, it is anticipated that at each iteration the distribution gets closer to the posterior $P(\theta|y)$ and the quality of the sample improves as a function of the number of steps.

The Metropolis Algorithm : When the full conditionals of parameters are not in closed form, one can use Metropolis sampling. This algorithm is derived from the process of a random walk and is based on an acceptance/rejection rule to converge to the intended posterior distribution. The procedure involved in the algorithm is as follows [9].

Step 1 : Draw a starting point θ^0, for which $p(\theta^0|y) > 0$, from a starting distribution $P_0(\theta)$. The starting distribution is mostly based on an approximation.

Step 2   For iteration $t = 1, 2, \ldots$:

(a)   sample a proposal $\theta^\star$ from a jump distribution (or proposal distribution) at time t, $J_t(\theta^\star|\theta^{t-1})$. The jump distribution must be symmetric, satisfying the condition $J_t(\theta_a|\theta_b) = J_t(\theta_b|\theta_a)$ for all θ_a, θ_b , and t.   (b) Calculate the ratio of the densities ,

$$ r = \frac{p(\theta^\star|y)}{p(\theta^{t-1}|y)}. $$

(c) Set

$$ \theta^t = \begin{cases} \theta^\star & \text{with probability } \min(r, 1). \\ \theta^{t-1} & \text{otherwise} \end{cases} $$

$\theta^t = \theta^{t-1}$ implies the jump is not accepted and the process must be repeated (iteration in the algorithm).

The Metropolist Hastings algorithm proceeds similarly as the Metropolist algorithm except that the jumping distribution is not required to be symmetric and the ratio is modified as follows

$$ r = \frac{p(\theta^\star|y)/J_t(\theta^\star|\theta^{t-1})}{p(\theta^{t-1}|y)/J_t(\theta^{t-1}|\theta^\star)}. \tag{3.4} $$

The common application of MCMC–based algorithms involves numerically calculating multi-dimensional integrals. Inferencial methods emanating directly from the posterior is based on obtaining marginal distributions.

43

In these instances , integration is also required to find marginal expectations and distribution of functions of subsets of the parameter θ. "The difficulty in obtaining marginal distributions from a nonnormalised joint density lies in integration. Suppose, for example, that θ is a $p \times 1$ vector and $f(\theta)$ is a nonnormalised joint density for θ with respect to Lebesgue measure. Normalising f entails calculating $\int f(\theta)d\theta$. To marginalise, say for θ_i, requires $h(\theta_i) = \int f(\theta)d\theta_{(i)} / \int f(\theta)d\theta$, where $\theta_{(i)}$ denotes all components of θ except θ_i. When p is large, such integration is analytically infeasible [8].

The challenge of using MCMC methods lies in determining the mixing time of the Markov chain. The mixing time of a Markov chain is the time until the Markov chain is "close" to its steady state distribution. Essentially, the experimenter needs to address the question of how large must t be until the time-t distribution is approximately π, where π is the posterior distribution. The variation distance mixing time, is defined as the smallest t such that

$$|P(Y_t \in A) - \pi(A)| \leq \frac{1}{4} \qquad (3.5)$$

for all subsets A of states and all initial states.

3.4 Bayesian Model Selection Criteria- The Bayes Factor

The bayes factor is used to decide between two contesting discrete set of hypothesis of interest. "The statistician (or scientist) is required to choose one particular hypothesis out of the two available and there must be a zero-one loss on that decision" [13]. The Bayes factor denotes the ratio of the marginal likelihood under one model to the marginal likelihood under a second model. If the two hypothesis are represented as H_0 and H_1 with priors $p(H_0)$ and $p(H_1)$, the ratio of the posterior probabilities is given as :

$$\frac{p(H_1|y)}{p(H_0|y)} = \frac{p(H_1)}{p(H_0)} * \text{Bayes factor}(H_1, H_0) \quad \text{where}$$

$$B = \text{Bayes factor}(H_1, H_0) = \frac{p(y|H_1)}{p(y|H_0)} = \frac{\int p(\theta_1|H_1)p(y|\theta_1, H_1)d\theta_1}{\int p(\theta_0|H_0)p(y|\theta_0, H_0)d\theta_0}$$

$$= \frac{P(H_1|y)/P(H_1)}{P(H_0|y)/P(H_0)}.$$

Table 3.1 gives an interpretation of the Bayes Factor based on the Jeffreys criteria for model selection [13].

Table 3.1 shows how the Bayes factor is used to choose between two hypothesis. For values of the Bayes factor between 1 and 3, the evidence against H_0 (the equivalence hypothesis) is not worth more than a bare mention. For values of the Bayes factor between 3 and 10, the evidence for H_1 is substantial.

Table 3.1:   Table showing decision rule using Bayes Factor

Bayes Factor(B)	Strength of Evidence
$B \leq 0.1$	Strong against
$0.1 < B \leq (1/3)$	Substantial against
$(1/3) < B < 1$	Barely worth mentioning against
$1 \leq B < 3$	Barely worth mentioning for
$3 \leq B < 10$	Substantial for
$10 \leq B < \infty$	Strong for

Note that the Bayes factor is only defined when the marginal density of y under each model is proper. The goal when using Bayes factors is to choose a single model H_i or average over a discrete set using their posterior distributions, $p(H_i|y)$.

3.5   The Dirichlet Process

A Dirichlet process (DP) is a distribution over probability distributions [20]. Assume that G is a probability distribution over a measurable space Θ, then a DP is a probability distribution over all the distributions of the subsets of Θ. The Dirichlet process is specified by the pair (M, H) for which H is the base distribution and $M > 0$ is a concentration parameter. Two major methods of constructing a DP are discussed below [20]:

Stick-breaking construction: Suppose that an infinite sequence of "weights"

$\{\pi_k\}_{k=1}^{\infty}$ are generated such that

$$\beta_k \sim \text{Beta}(1, M)$$

$$\pi_k = \beta_k \prod_{l=1}^{k-1}(1 - \beta_l)$$

Consider the discrete random probability distribution:

$$G(\theta) = \sum_{k=1}^{\infty} \pi_k \delta_{(\theta = \zeta_k)} \quad \text{where} \quad \zeta_k \overset{iid}{\sim} H \text{ and } \delta \text{ is an indicator function.}$$

Then $G \sim \text{DP}(M, H)$.

Polya urn scheme: Suppose that colored balls are drawn from an urn G and let θ_i represent the color of the ith ball drawn from the urn. Suppose that for each ball drawn, it is replaced and another ball of the same color is added. As more balls of the given color are drawn, it becomes more likely to draw balls of the given color at subsequent draws. To add diversity, a ball is occasionally drawn from a different urn H, replaced and a ball of the same color added to the original urn G. If $G \sim DP(M, H)$ and $\theta_1, ..., \theta_N \sim G$, then as the draw continues indefinitely G_N converges to a random discrete distribution which is a $\text{DP}(M, H)$ [24].

It is observed that the normality assumption on μ_i is too restrictive when the heterogeneity among studies is quiet appreciable and that this assumption can be relaxed using a Dirichlet process. Muthukumarana & Tiwari [16] considers a hierarchical Dirichlet Process formulation for α_i of

the model 3.1 based on

$$\alpha_i | G \overset{iid}{\sim} G, \ i = 1, \ldots, k$$

$$G \sim DP(M_1, \ H_1), \quad M_1 \text{ fixed}$$

$$H_1 \sim N(0, \tau^2)$$

$$\tau^2 \sim IG(\eta, \ \lambda).$$

We consider a Dirichlet Process formulation for μ_i in our Random effects model 3.3 as follows.

$$\mu_i | F \sim F$$

$$F \sim \text{DP}(M_2, \ H_2)$$

$$H_2 \sim N(\mu, \ \tau^2)$$

$$\mu \sim N(\mu_0, d\tau^2)$$

$$1/\tau^2 \sim G(a, \ b).$$

where M_2, μ_0 and d are known.

Note that the above formulations of the Dirichlet Process are known as the Ordinary and Conditional Dirichlet Processes respectively.

Chapter 4

Data Analysis

4.1 Example 1

The data used in this sectionn provides information on diabetes patients, 42 diabetes treatments, and possible heart condition or death resulting from the use of rosiglitazone (a treatment for diabetes). This data is attached as part of the appendix. For each of the 42 treatments, a test of equivalence is done to ascertain whether the treatment proportion is equivalent to the control proportion. This example is based on the Statistical inferential procedure for binary data discussed in Section 2.1. For each arm, the number of patients who had myocardial infarction out of a total n_t as a result of using the diabetes treatment is considered to be the number of successes in n_t binomial trials. Similarly, the number of cases in the control group is treated as a binomial outcome independent of the treatment group. The equivalence margin δ is chosen to be as small as possible such that if the absolute value of the difference in the control and treatment proportions is less than δ, we can say that the two proportions are equivalent. For

example, we assume a practically meaningful equivalence margin $\delta = 0.01$. The hypothesis for a test of equivalence of treatment number 20 and it's control group is as follows:

$$H_0 : |P_{t20} - P_{c20}| \leq \delta$$

$$H_1 : |P_{t20} - P_{c20}| > \delta.$$

To evaluate how the Beta posterior is sensitive to the Beta prior assumptions, a plot of the likelihood, prior and posterior distribution is examined for some of the treatments. The plots for four of the treatments with their respective controls beside them are presented in Figures 4.1 and 4.2. Each of these graphs depicts a pattern in which either the posterior distribution looks like the likelihood distribution or the posterior seems to be a blend of the likelihood and the prior. This implies values generated from this posterior will reflect the state of the data because data is supposed to have come from the likelihood.

The equivalence test is carried out using the Bayes factor. Tables 4.1 and 4.2 give the results of the equivalence test. The first column D_i is the ith drug (treatment). Columns 2 and 3 are the treatment proportion (x_t/n_t) and control proportion (x_c/n_c) respectively. Columns 4 $(P(H_0|X))$ and 5 $(P_A(H_0|X))$ are the posterior probabilities that $H_0 : |P_{ti} - P_{ci}| \leq \delta$ is true under the Beta posterior distributions and under the normal approximation to the Beta posterior respectively. Column 6 (B) is the Bayes Factor for exact posterior and B_A is the Bayes Factor based on the normal approximation. The Bayes Factors are calculated on the assumption that

50

Di	Pt_i	Pc_i	$P(H_0|X)$	$P_A(H_0|X)$	B	B_A
D_1	0.0019	0.0000	0.7607	0.6757	0.3146	0.4788
D_2	0.0017	0.0016	0.7190	0.6842	0.3908	0.4615
D_3	0.0003	0.0018	0.4861	0.1857	1.0572	4.3826
D_4	0.0000	0.0037	0.3428	0.2548	1.9171	2.9244
D_5	0.0013	0.0000	0.5916	0.3325	0.6903	2.0079
D_6	0.0000	0.0111	0.1924	0.0402	4.1975	25.8938
D_7	0.0032	0.0032	0.4401	0.6145	1.2722	0.6272
D_8	0.0280	0.0082	0.1806	0.7738	4.5370	0.2924
D_9	0.0007	0.0000	0.8897	0.9711	0.1240	0.0297
D_{10}	0.0010	0.0000	0.6673	0.3861	0.4986	1.5898
D_{11}	0.0000	0.0009	0.8083	0.6501	0.2372	0.5382
D_{12}	0.0011	0.0000	0.6860	0.7311	0.4577	0.3677
D_{13}	0.0026	0.0011	0.7079	0.9717	0.4126	0.0291
D_{14}	0.0016	0.0000	0.5808	0.7937	0.7218	0.2604
D_{15}	0.0017	0.0017	0.6950	0.6392	0.4388	0.5646
D_{16}	0.0016	0.0038	0.4136	0.3957	1.4180	1.5710
D_{17}	0.0039	0.0097	0.3147	0.0381	2.1776	26.9081
D_{18}	0.0037	0.0000	0.5330	0.4928	0.876	1.0210
D_{19}	0.0110	0.0027	0.3362	0.5785	1.9744	0.7285
D_{20}	0.0000	0.0000	0.5145	0.6454	0.9436	0.5495
D_{21}	0.0000	0.0033	0.4455	0.0949	1.2447	9.5432

H_0 and H_1 are equally likely, that is, $P(H_0) = P(H_1) = 0.5$. For drug number six labelled as 49653/085, the Bayes Factor for the exact posterior is 4.1975 where as that of the normal approximation is 25.8938. Both Bayes Factors are above 1 which imples H_1 is more likely to be true and H_1 is the hypothesis that the treatment proportion is not equivalent to the control proportion. Where as the evidence for H_1 is substantial based on the exact posterior distribution, there is a strong evidence for H_1 based on the normal approximation.

We now consider a missing data analysis in an arm. As an example,

Table 4.2: Posterior Probabilities and Bayes Factor (Continuation of Table 4.1)

Di	Pt_i	Pc_i	$P(H_0|X)$	$P_A(H_0|X)$	B	B_A
D_{22}	0.0053	0.000	0.5829	0.5570	0.7156	0.7953
D_{23}	0.0051	0.0048	0.2397	0.3435	3.1719	1.9111
D_{24}	0.0256	0.000	0.1638	0.3695	5.1050	1.7476
D_{25}	0.0000	0.0072	0.5181	0.1164	0.9301	7.5725
D_{26}	0.0172	0.0270	0.3149	0.0217	2.1756	45.0188
D_{27}	0.0068	0.0000	0.5122	0.9982	0 .9552	0.0018
D_{28}	0.0043	0.0000	0.8242	0.844	0.2133	0.1818
D_{29}	0.0112	0.0000	0.3482	0.3491	1.8719	1.8641
D_{30}	0.0060	0.0000	0.5609	0.9130	0.7828	0.0953
D_{31}	0.0172	0.0270	0.3178	0.8182	2.1466	0.2222
D_{32}	0.0009	0.0000	0.9483	0.7935	0.0545	0.2602
D_{33}	0.0000	0.0000	0.9135	0.7935	0.0946	0.7134
D_{34}	0.0049	0.0108	0.5297	0.4546	0.8879	1.1996
D_{35}	0.0035	0.0000	0.0033	0.5398	0.2544	0.8528
D_{36}	0.0032	0.0000	0.7441	0.7334	0.3039	0.3635
D_{37}	0.0032	0.0000	0.7164	0.4792	0.3958	1.0868
D_{38}	0.0000	0.0000	0.6692	0.5836	0.4943	0.7134
D_{39}	0.0023	0.0000	0.5644	0.4546	0.7718	1.996
D_{40}	0.0025	0.0000	0.6196	0.5814	0.6137	0.7198
D_{41}	0.0057	0.0034	0.9997	0.9998	0.0003	0.0003
D_{42}	0.0185	0.0142	0.8822	0.9995	0.1335	0.0005

suppose an observation was missing in the treatment labelled 49653/234 with three cases out of a sample of size 111. We estimate this missing value using Gibbs sampling derived in section 2.4. R code for the Gibbs sampling is given in Appendix.

Figures are based on 20000 MCMC simulations. According to Figure 4.3, it is likely that x_m is 0. The trace plot in Figure 4.4 shows that mixing is good enough and there are no large spikes in the autocorrelation plot after lag 0. This is an indication of convergence of the Markov Chain.

4.2 Example 2

We now consider a dataset relating to the number of deaths arising from lung cancer as a consequence of smoking. This is a survey carried out by Princeton University and the data is attached as part of the appendix. It can also be accessed at http://data.princeton.edu/wws509/datasets/smoking.dat. The dataset present two classes of smokers named "heavy" and "light" smokers. The light smokers comprise the non-smokers and what has been classified as cigarPipeOnly. The 'heavy' smokers are those who smoke cigarrette and cigarrettePlus (probably large packets of ciggarrete in addition to cigar). Equivalence testing is done to determine if the average number of deaths resulting from light smoking is different from the average number of deaths arising from heavy smoking. The equivalence hypothesis is given by

$$H_0 : |\lambda_h - \lambda_l| < \delta$$

$$H_1 : |\lambda_h - \lambda_l| > \delta$$

where λ_h is the average number of lung cancer deaths resulting from heavy smoking and λ_l is the average number of people who died from light smoking. We assume an equivalence margin of $\delta = 0.01$. The data are assumed to come from Poisson distributions and gamma priors are imposed on λ's. The distributions of Heavy and Light smokers are shown in Figure 4.5. The joint posterior distribution of $(\lambda_t, \ \lambda_c)$ is shown in Figure 4.6. To do the equivalence test, the posterior probabilities of H_0 and H_1 are calculated and the higher probability is more likely. From section 2.2,

53

the posterior distributions of λ_t's are $\text{Gamma}(\sum x_{it} + \alpha_t, \quad \beta_t + n_t)$ and $\text{Gamma}(\sum x_{ic} + \alpha_c, \quad \beta_c + n_c)$ respectively. To test the equivalence hypothesis, a function is written in R to count the number of Monte Carlo estimates that falls within the margin specified in the null hypothesis. The posterior probability that H_0 is true is 0 for an equivalent margin of 0.01 which implies it is certain that the average number of deaths from heavy smoking is not equivalent to the average number of deaths from light smoking. For an equivalence margin of 2, the posterior probability that H_0 is true is still less likely with a probability of 0.0437.

4.3 Example 3

We now re-analyse the data in example 1 in terms of a meta-analysis. It has been observed that 65% of deaths in diabetes patients are from cardiovascular causes [18]. It is therefore of importance to investigate the effect of rosiglitazone on heart conditions. Out of a total of 116 studies available, 42 of the studies satisfied the pre-determined conditions for a meta-analysis. The 42 trials comprise 15565 diabetes patients who were put on rosiglitazone(treatment group) and 12282 diabetes patients assigned to medication that does not contain rosiglitazone(control group). The average age of patients in the 42 trials is approximately 52 years. The interest is on myocardial infarction and death from rosiglitazone as a treatment for diabetes. Since the follow-up periods below treatments are similar for all trials, the use of odds ratio as treatment effect is valid. Most of the responses from the treatment are zero. Out of the 42 trials, only 13 treatment effects have

been estimated by the Mantel-Haenszel method. Consequently, the odds ratio calculated by the Mantel-Haenszel method has values designated as 0 or ∞. For instance, treatments labelled SB-712753/002, AVA100193 has a lower 95% limit as C.I as undefined and upper 95% C.I limit as infinity. The values of all the estimated odds ratios fall within the 95% confidence interval. This implies that even in cases where myocardial infarction is more likely in the treatment group, the occurance of the events (myocardial infarction) are not significant. The estimate of the combined odds ratio by the Mantel -Haenszel method is 1.39 with a 95% confidence interval of $(1.01, 1.91)$. That is myocardial infarction is 39% more likely in the diabetes patients treated with rosiglitazone compared to diabetes patients not treated with rosiglitazone. The Dersimonian and Laird method gives the summary odds ratio to be 1.25 and an estimate of the between study variance to be 0.

It is clear that treatment effects are not estimable in this case. The authors provided a remedy by pooling some of the studies. That is by combining treatments in order to have values for each cell to be able to estimate treatment effect. This in turn gave estimates for treatment effects. The chi-square test for heterogeneity is found to be 6.61 from the Mantel–Haenszel method with a high p-value of 0.8825 which seeks to justify the FEM where the studies as a group is assumed to have some common effect size which can be found by combining the studies. Nevertheless, this approach is not the best since the high p-value only indicates statistical non–significance and not practical significance. The merged cells represent

different treatments and as such combining them may not be meaningful. Moreover, the study has been carried out at different centers representing different populations with different characteristics and as such some amount of variability is expected between the studies.

The literature suggest that when responses are mostly zeros, each cell be adjusted by a value that is small in magnitude. In particular, adding a value of 0.5 to all the cells [11]. This approach has been adapted in the current study and the odds ratios re–estimated. The odds ratios of this modification is shown in Table 4.3 and 4.4 . The summary odds ratio for the modified data is 1.2, that is rosiglitazone is 20% more likely to cause cardiovascular effects and death. A 95% confidence interval is (0.91, 1.6). The DerSimonian–Laird method estimate of the summary odds ratio is 1.21 which does not vary so much from the Mantel-Haenszel estimate. The value of the chi-square test statistic is 17.88 with a p-value of 0.9994. The chi-square test statistic only assesses whether observed differences in treatment across studies are compatible with chance.

Generally, if confidence intervals for the results of individual studies (depicted graphically using horizontal lines) are non overlapping, this indicates the presence of heterogeneity. A look at the forest plot of the data in figure 4.7 shows the horizontal lines do not overlap. Figures 4.7 and 4.8 are the plots of the confidence intervals associated with the treatments. Each study is represented by a horizontal line. However, studies having zero events in both groups will not have lines representing them. The lines represent the length of the confidence interval for each study. The line

Table 4.3: The estimates of odds ratios by the Mantel–Haenszel method after adding 0.5 to each response

Treatment	OR	lower 95%	upper 95%
49653/011	2.36	0.11	49.33
49653/020	0.88	0.12	6.72
49653/024	0.24	0.02	2.30
49653/093	0.17	0.01	4.17
49653/094	1.50	0.06	37.19
100684	0.36	0.01	9.00
49653/143	3.55	0.14	88.01
49653/211	2.35	0.51	10.72
49653/284	3.02	0.12	74.46
712753/008	1.43	0.06	35.29
AMM100264	0.34	0.01	8.41
BRL49653C/185	1.26	0.06	26.44
BRL49653/334	1.68	0.22	12.80
BRL49653/347	2.55	0.12	53.25
49653/015	0.83	0.11	6.36
49653/079	0.52	0.05	5.05
49653/080	0.56	0.07	4.36
49653/082	2.54	0.12	53.42
49653/085	2.39	0.35	16.39
49653/095	0.49	0.01	24.81
49653/097	0.33	0.01	8.06

for each study has a box located on it and middle of the box represents the magnitude of the treatment effect for the corresponding study. The area of the box represent the weight assigned to each study. The diamond is the combined treatment effect. Hence there is some inherent heterogeneity and a random effects model is fit to the data in this thesis. Even though adding 0.5 to each cell enabled us to calculate odds ratios, it is still not the best approach. In this study, this data is re-analysed by fitting a semi–parametric random effects model described in Chapter 3.

Forest plot of observed treatment effects and 95% confidence intervals

	OR	(lower 95%	upper 95%)
		Table 4.4: Continuation of 4.3	
49653/125	0.33	0.01	8.10
49653/127	3.17	0.13	79.37
49653/128	3.00	0.12	76.03
49653/134	0.10	0.00	2.04
49653/135	0.68	0.13	3.50
49653/136	2.92	0.12	72.23
49653/145	3.16	0.13	77.89
49653/147	3.00	0.12	74.66
49653/162	3.09	0.12	76.39
49653/234	0.68	0.13	3.50
49653/330	0.96	0.04	23.74
49653/331	0.46	0.01	23.23
49653/137	0.54	0.07	4.13
SB-712753/002	2.93	0.12	72.15
SB-712753/003	3.23	0.13	79.55
SB-712753/007	1.47	0.06	36.38
SB-712753/009	0.99	0.02	50.08
49653/132	0.76	0.03	18.77
AVA100193	0.94	0.04	23.32
DREAM	1.63	0.73	3.67
ADOPT	1.32	0.81	2.15

for rosiglitazone study. The horizontal lines represent the length of the confidence interval. The center of each box represent the magnitude of the study effect and the area of the box is the weight assigned to each study.

The funnel plot in figure 4.9 shows the actual responses of effect sizes where as figure 4.10 represents the funnel plot after adjusting the responses (by adding 0.5 to the treatment and control cases). Both shapes do not deviate so much from the pattern of a funnel turned upside down. This shows that publication bias may not be a problem with the rosiglitazone dataset.

In the Bayesian setting, when the posterior probability of the data given a specific model is the highest, then that model is the preferred model. "However, it is difficult to calculate the two marginal likelihoods m_h^c and m_h^o exactly, or very difficult to evaluate accurately even when feasible [4]. But, it is possible to estimate their ratio (the Bayes factor) m_h^c/m_h^o for all h from a single Markov chain, run under model $\mathcal{M}_{h_1}^o$, where h_1 is some prespecified value of the hyperparameter $h_1 = (M_1, d_1)$, M is the precision parameter and d is vector of starting values for the hyperparameters. $\mathcal{M}_c$ and $\mathcal{M}_o$ are respectively the Conditional Dirichlet and the Ordinary Dirichlet model and m_h^c and m_h^o are the respective marginals" [5]. Figure 4.11 shows the plot of Bayes factors for choosing between the mixtures of Conditional Dirichlet model and the Ordinary Dirichlet model. The plot shows that the ratio m_h^c/m_h^o is always greater than 1 and the Conditional Dirichlet model is preferred for the rosiglitazone dataset.

We now investigate the choice of M, precision parameters of DP. We consider $M = 1$ and $M = 10$. The posterior distributions of μ (mu) and τ (tau) are displayed in Figure 4.12. The posterior distributions of the mean look similar for values of the concentration parameter equal 1 and 10. For $M = 10$, the responses seem to be clustered around 0 and the tails of the distribution for $M = 10$ are flatter . However, the distribution of τ is skewed to the right. The initial values and hyper parameters for the Gibbs estimation is in table 4.5.

The parameters of the model are estimated by Gibbs sampling algorithm implemented in R. The R code for the Gibbs sampling is attached

Table 4.5: Initial Values for Gibbs sampling

μ_0	τ_0^2	μ	d	a	b
0	1	0	0.001	1	2

as part of the appendix. The estimates of study effects (μ_i) are given in Table 4.6.

Table 4.6: The estimates of posterior treatments and standard deviations

Parameter	Estimate	S.d	Parameter	Estimate	S.d
τ^2	0.74	0.2794073	μ_{21}	-0.78	0.7748052
μ	0.71	0.4142608	μ_{22}	-0.78	0.7449046
μ_1	- 0.73	0.8151172	μ_{23}	-0.71	0.8175963
μ_2	-0.63	0.7358434	μ_{24}	-0.75	0.8185433
μ_3	1.2	0.4914719	μ_{25}	-1.6	0.4702446
μ_4	-1.1	0.6541732	μ_{26}	-0.57	0.5950003
μ_5	-0.74	0.8379231	μ_{27}	-0.74	0.8104195
μ_6	-0.77	0.7702218	μ_{28}	-0.74	0.8192542
μ_7	-0.70	0.8191522	μ_{29}	-0.74	0.8091605
μ_8	-0.61	0.7714286	μ_{30}	-0.73	0.8372302
μ_9	-0.74	0.7989194	μ_{31}	-0.58	0.6008941
μ_{10}	-0.75	0.8194161	μ_{32}	-0.72	0.8096048
μ_{11}	-0.77	0.7766973	μ_{33}	-0.76	0.8134446
μ_{12}	-0.74	0.8008516	μ_{34}	-0.69	0.6464147
μ_{13}	-0.67	0.7848413	μ_{35}	-0.72	0.8196591
μ_{14}	-0.73	0.8148389	μ_{36}	-0.75	0.8109735
μ_{15}	-0.65	0.7427869	μ_{37}	-0.71	0.8172117
μ_{16}	-0.72	0.6874473	μ_{38}	-0.72	0.8127885
μ_{17}	-0.67	0.6670385	μ_{39}	-0.74	0.8024024
μ_{18}	-0.74	0.8121317	μ_{40}	-0.73	0.8114458
μ_{19}	-0.69	0.7916802	μ_{41}	-0.189	0.5258747
μ_{20}	-0.74	0.8151257	μ_{42}	0.01	0.3224310

4.4   A Simulation Study

In this simulation study, each study has been simulated by means of a binomial random variable in which the number of cases in the treatment group and the control group are generated as independent binomial random variables. That is, for the arm labeled 49653/011 for which there are 375 total number of patients in the treatment group with 2 cases, this is regarded as 2 'successes' out of a total of 375 trials with 'success probability' $p = 2/375$. In order to determine how the model performs, a typical approach is the examination of estimates of the model to see if they make sense [9]. As an example, we generate twenty binomial successes using the rbinom random generator. We assume $n = 200$ in each case and fix the p at 0.7. This setting is similar to administering a treatment in twenty hospitals with 200 patients in each hospital. Fixing p at 0.7 generates number of cases that do not vary so much from each other. This is confirmed in the non significance of the chi-square test for heterogeneity. Another set of twenty 'number of cases' is generated from the binomial distribution but this time we induce heterogeneity. This is done by varying the success probability of each trial. For instance rbinom(1, 200, 0.86), rbinom(1, 200, 0.10), rbinom(1, 200, 0.55) . . .

Interest is in comparing the posterior treatment means of the heterogeneous studies with the studies that are not heterogeneous. Table 4.7 compares the posterior treatment means of 20 studies with heterogeneity to the treatment means of 20 other studies in which there is no heterogeneity. Column 1 is the posterior treatment means of the non–heterogeneous

(μ_i) studies where as $\mu_i^\star$ in column 2 posterior treatments of the heterogeneous studies. Treatment means in column 1 (μ_i) are mostly 0.68 or just slightly below or above it. On the other hand, all the treatment means in column 2 ($\mu_i^\star$) differ from each other significantly. If the responses are similar, the treatment effects are supposed to be an estimate of a common treatment mean, hence the model can be regarded as good.

Table 4.7: Estimates of treatment means for twenty studies with 200 observations within each study

Study	μ_i	$\mu_i^\star$
1	0.68	0.092
2	0.67	0.39
3	0.67	0.80
4	0.68	0.69
5	0.68	0.76
6	0.68	-2.8
7	0.68	-1.4
8	0.68	0.35
9	0.68	-0.11
10	0.70	0.53
11	0.67	0.69
12	0.70	-0.054
13	0.68	0.72
14	0.67	0.39
15	0.68	0.41
16	0.69	0.79
17	0.67	0.81
18	0.69	-0.94
19	0.68	-1.6
20	0.69	0.81

The estimates considered in this model are the posterior treatment means and the respective standard deviations. Estimates have been obtained in the different cases including small number of studies(k) involving

small number of patients(n), large number of studies(k) involving small number of patients(n), large k with large n and where both k and n are small. Table 4.8 presents results for the case where there are small number of studies$(k = 5)$ with large number of patients$(n = 200)$. Column 2 of Table 4.8 labelled μ_i gives the treatment mean of five studies in which there is no heterogeneity – the p–value for the chi–square test of heterogeneity is 0.35 with the associated posterior standard deviation in column 3 labelled σ_i. Columns 3 and 4 give the estimates of five different studies in which the studies differ from each other significantly (with a p–value for the chi–square test of heterogeneity as 0). The estimate of the posterior standard deviation for the five studies with heterogeneity is slightly lower than the posterior standard deviation of the five studies that are similar. This is possibly due to the fact that the semi–parametric model fitted to the data is a random effects model and therefore gives more precise estimates when there is some heterogeneity among the studies. The case where there are a smaller number of studies (k) with large number of patients appears to be a practical situation but a more realistic scenario could be experiments on a chronic disease which is characterised by a few patients (n) and possibly a small number of studies (k) .

Table 4.8: μ_i and σ_i are estimates of treatment mean and posterior standard deviation from five studies that are similar where as $\mu_i^\star$ and $\sigma_i^\star$ are estimates of five that studies that are heterogeneous

	μ_i	σ_i	$\mu_i^\star$	$\sigma_i^\star$
μ	1.2	7.83	3.6	3.87
τ	0.92	12.6	4.7	4.26
1	1.01	0.92	0.46	0.31
2	1.01	0.92	0.38	0.26
3	1.01	0.92	7.70	0.59
4	1.02	0.86	0.40	0.20
5	1.01	0.88	0.38	0.21
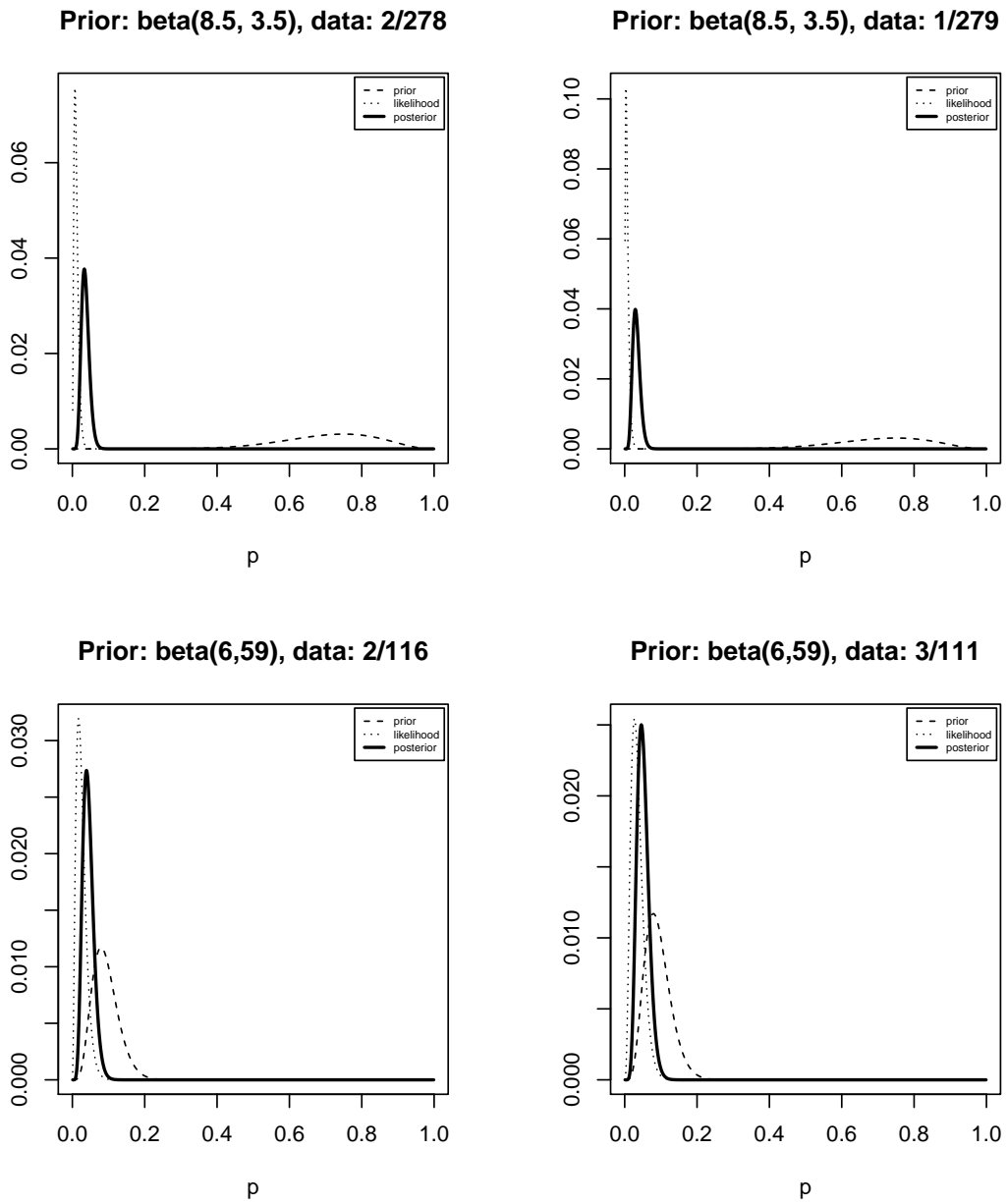
Figure 4.1: Graph showing the distributions of the Prior, Likelihood and Posterior for treatment $BRL49653/334$ and $49653/135$ with the respective controls at the right hand side
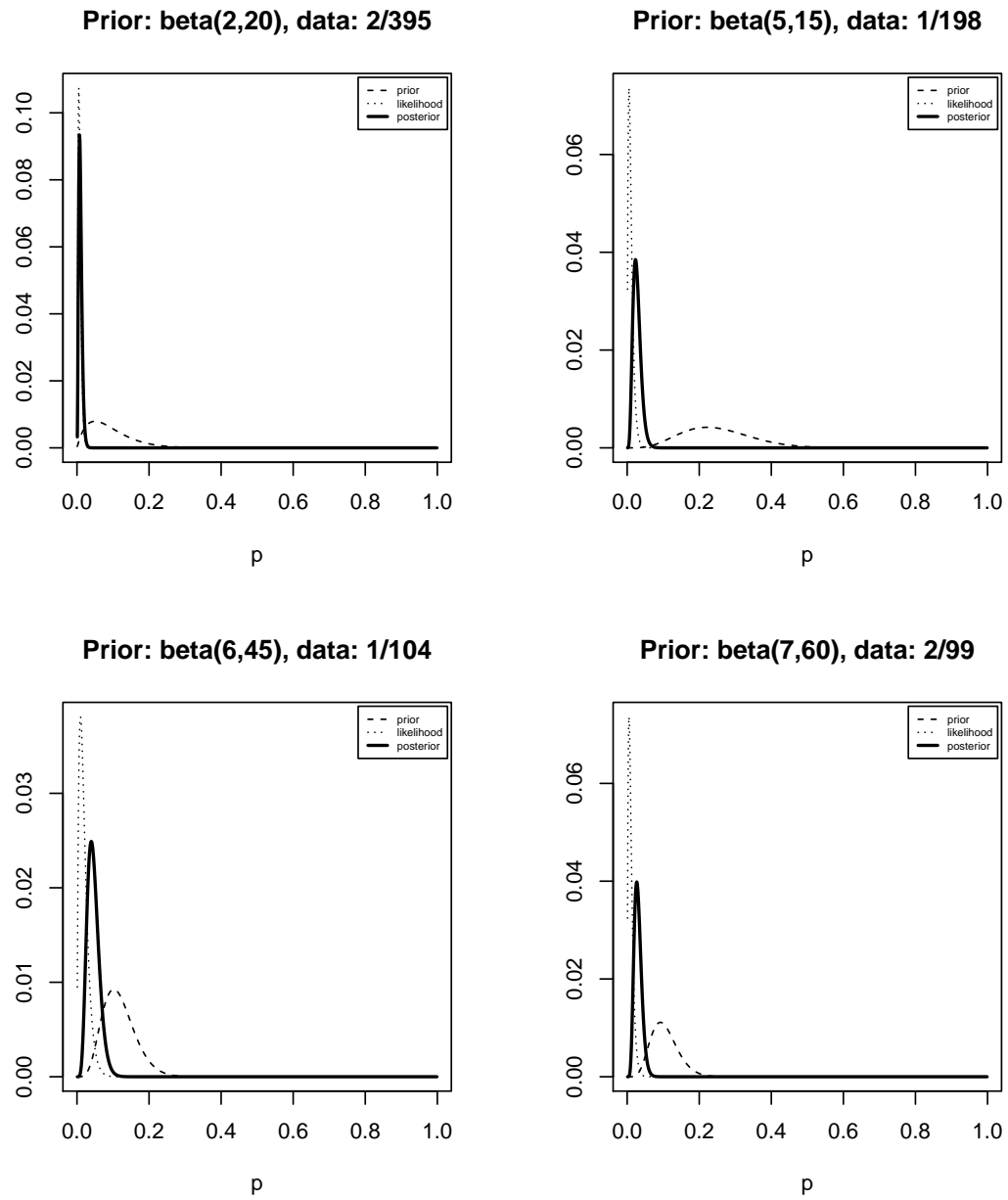
Figure 4.2: Densities of the Prior, Likelihood and Posterior for the arms 49653/015 and 49653/080 and their controls at the right

Posteriors of parameters using MCMC
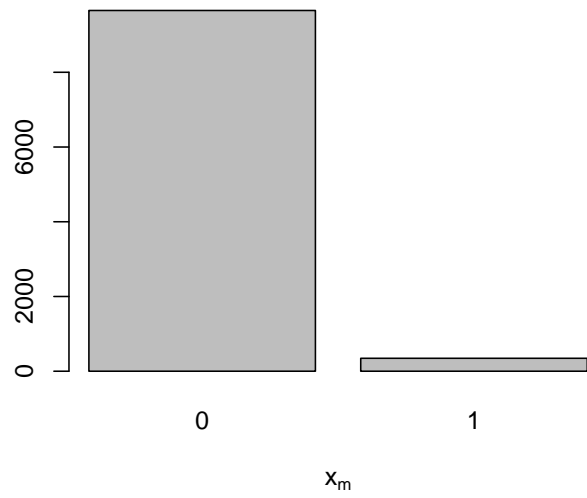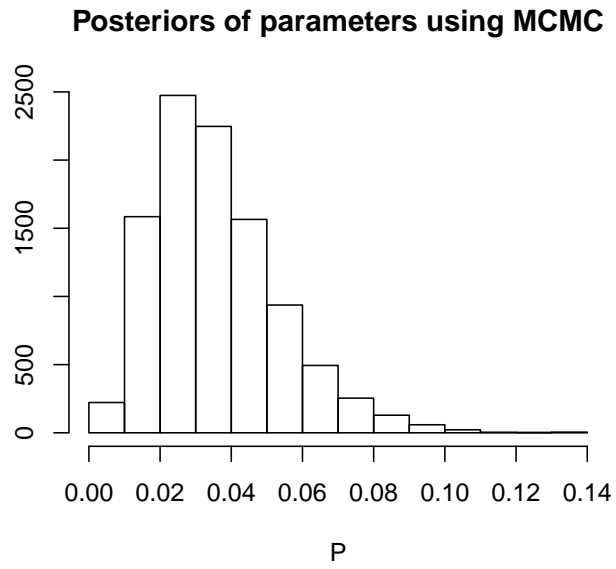
Figure 4.3: The distribution of x_m shows it is more likely to be 0
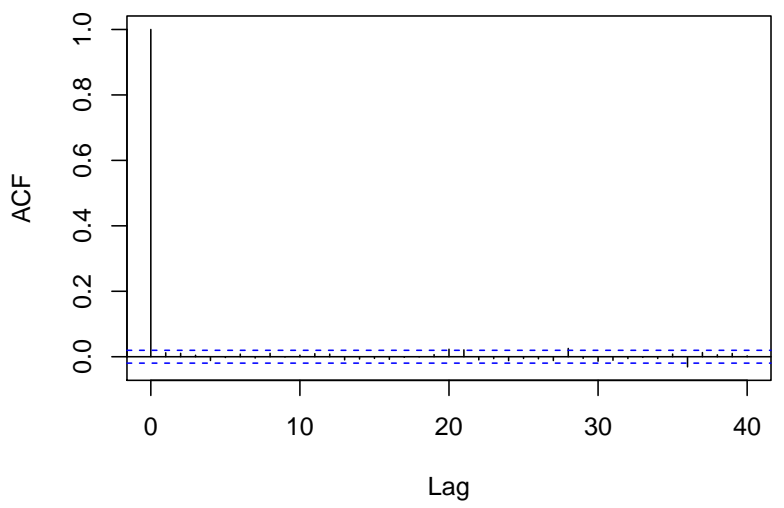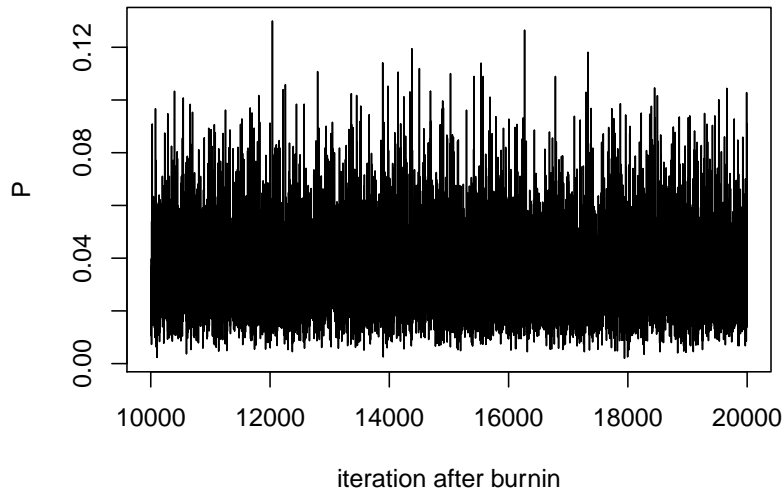
Figure 4.4: There is no discernible pattern in the trace plot and no large spikes after lag 0 in the autocorrelation plot
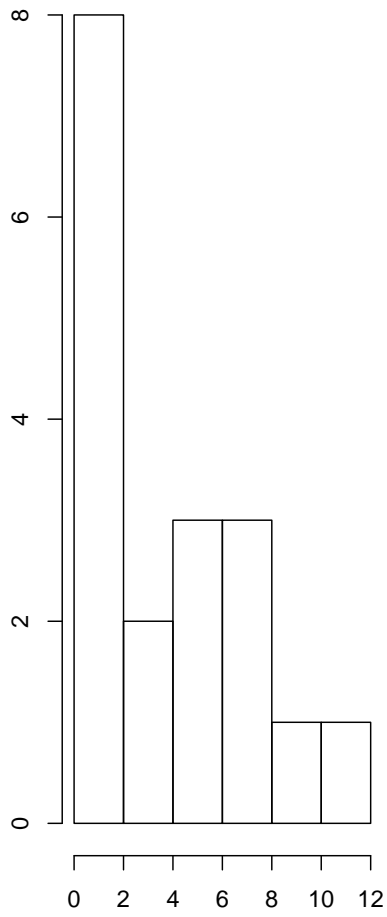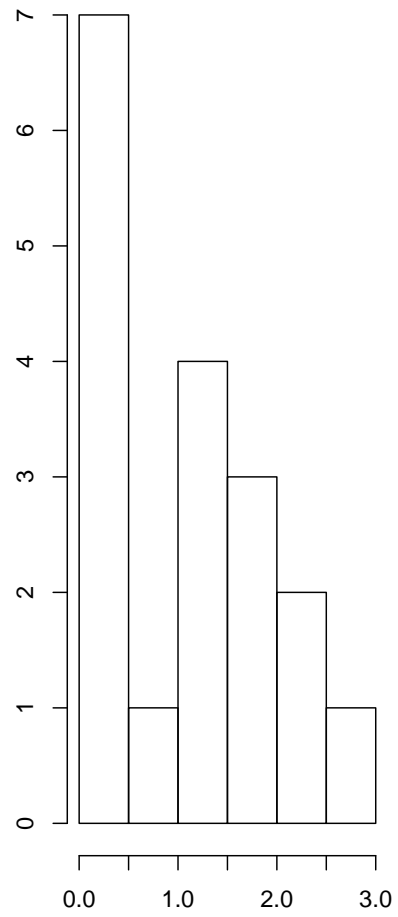
Figure 4.5: Histogram showing the distributions of Heavy and Light smok-
ers.
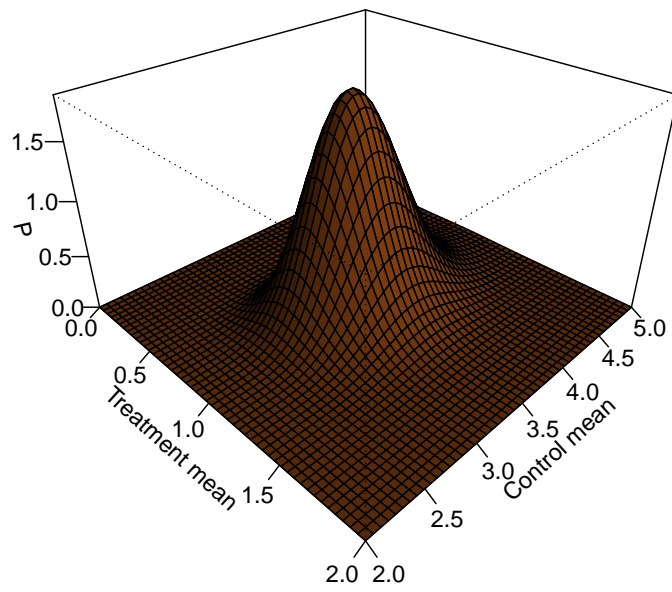
Figure 4.6: The joint distribution of the Treatment mean (λ_t) and Control mean (λ_c)
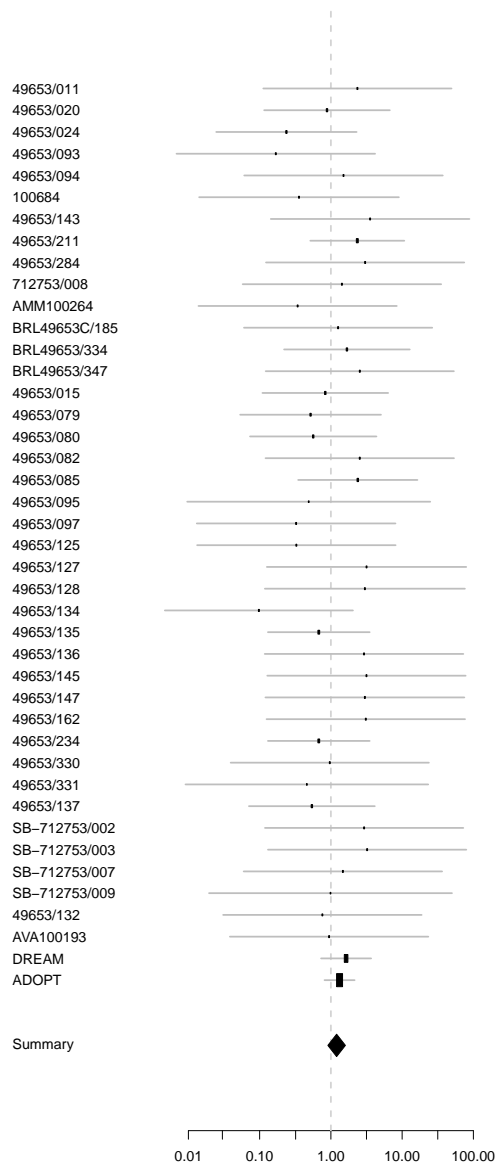
Figure 4.7: Forest plot of data after adjusting responses by addition of 0.5
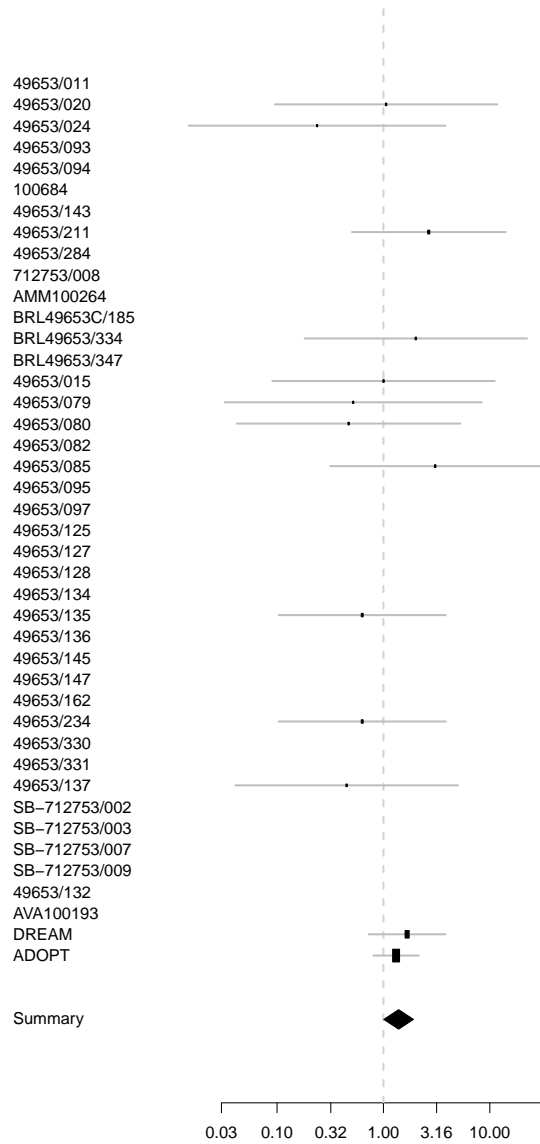
Figure 4.8: Forest plot of observed treatment effects and 95% confidence intervals for rosiglitazone study
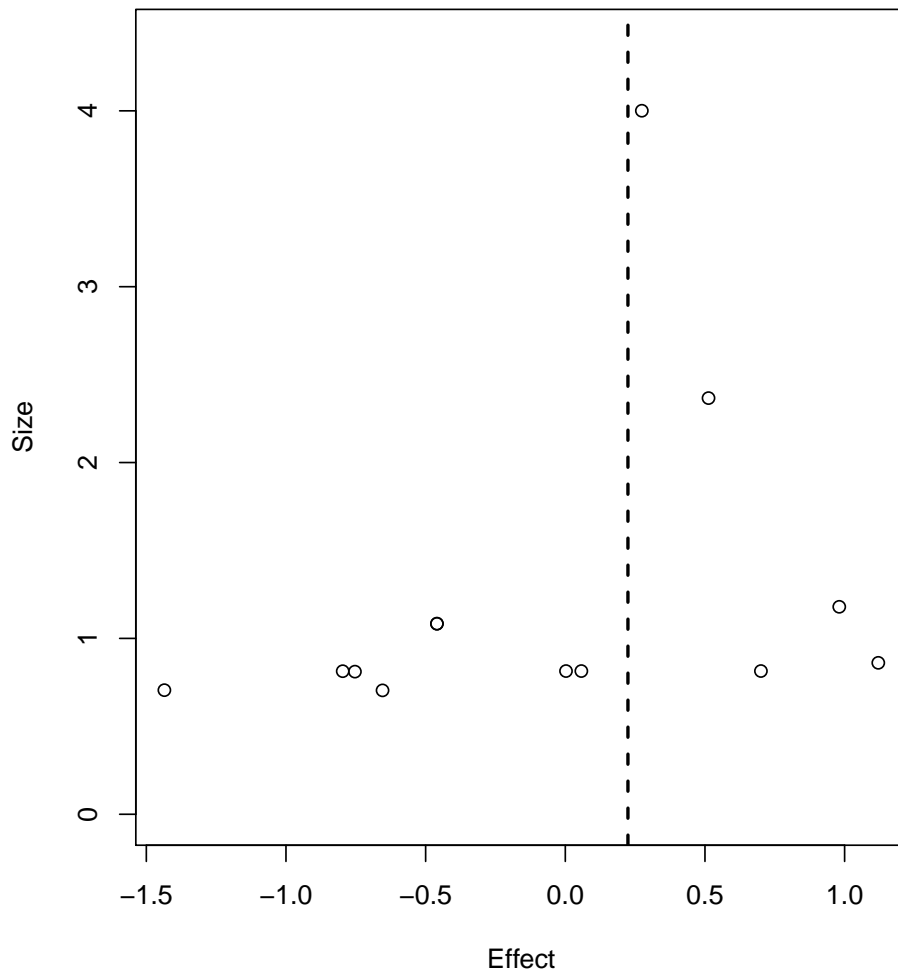
Figure 4.9: Funnel plot of rosiglitazone data
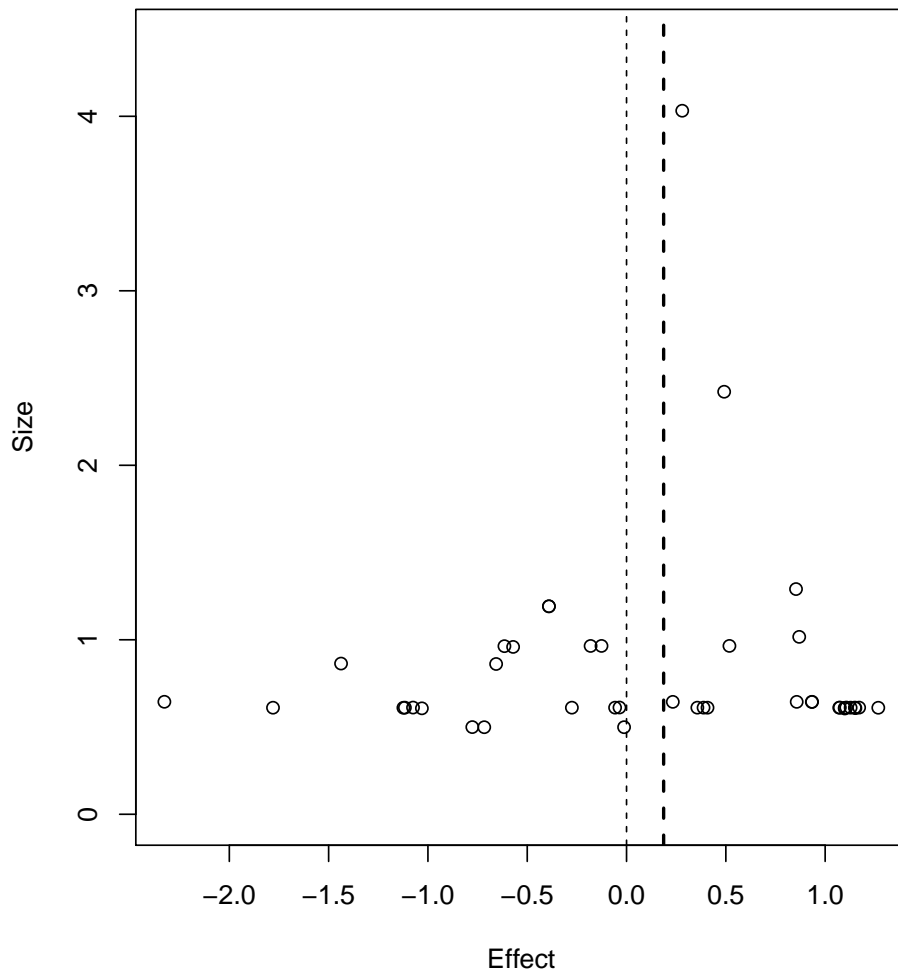
Figure 4.10: Funnel plot of rosiglitazone data after adjustment
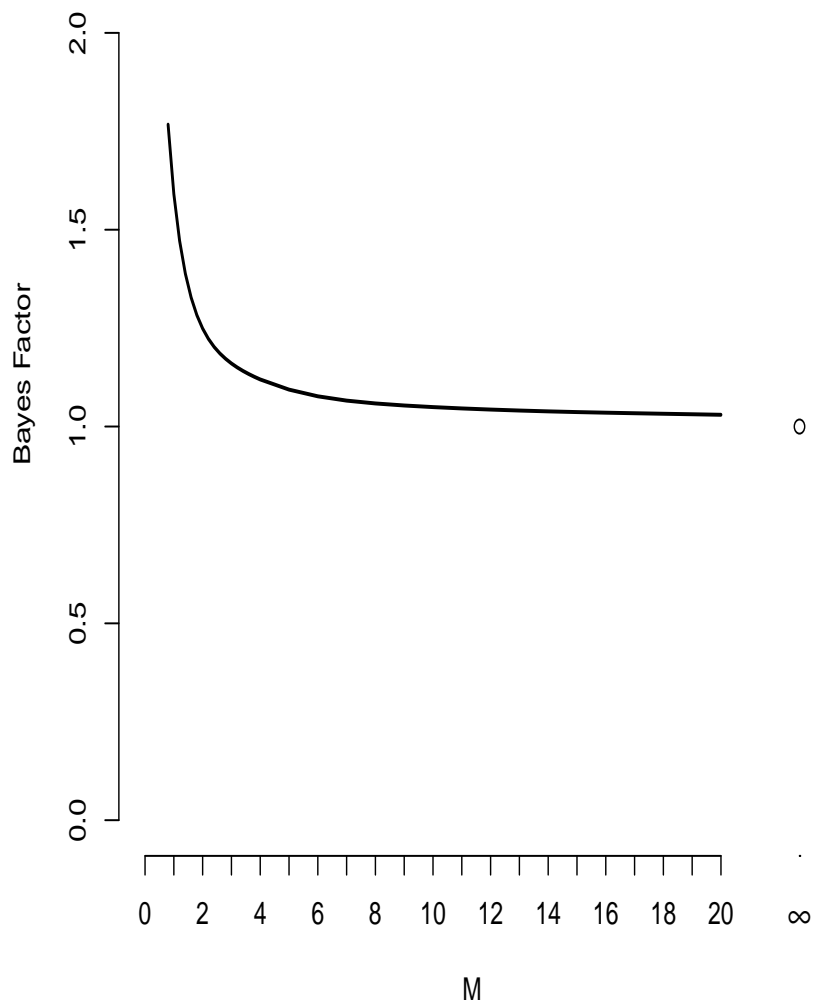
Figure 4.11: Graph of Bayes Factor for choosing between the Ordinary and Conditional Dirichlet models
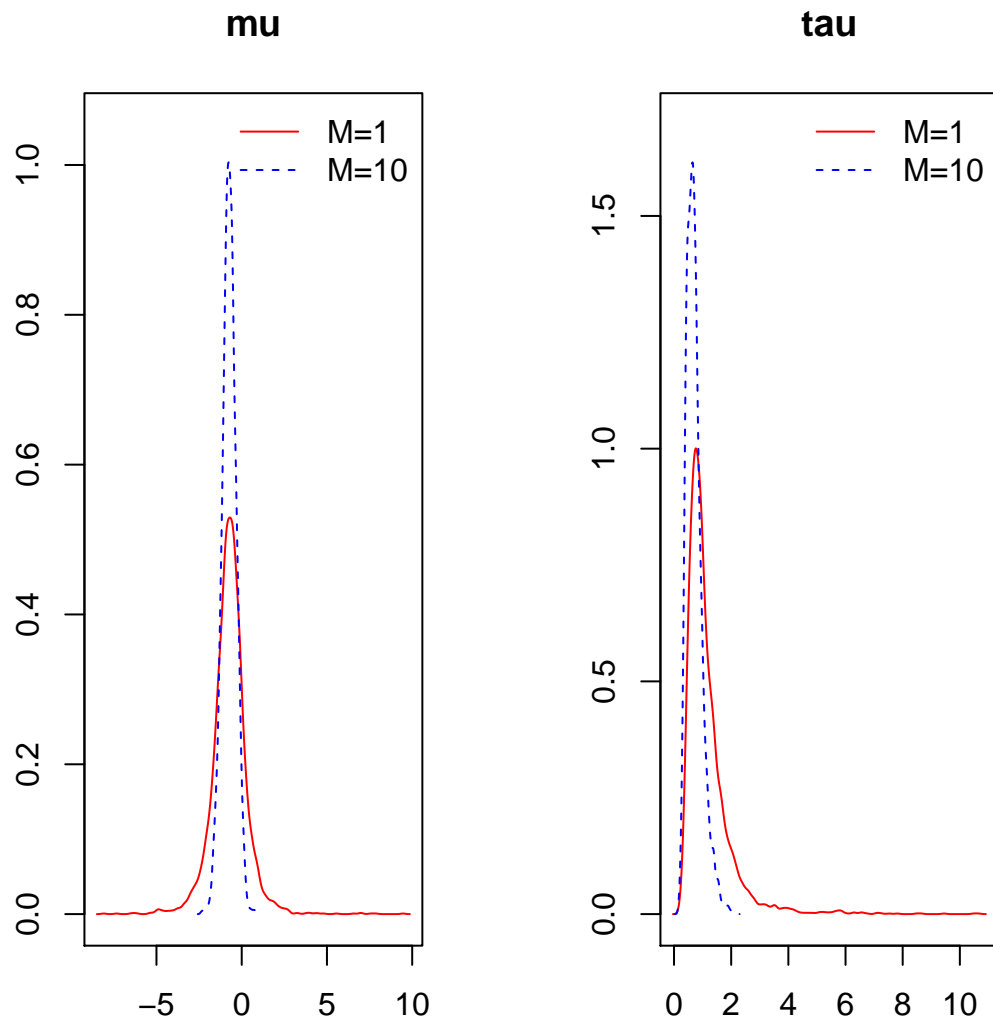
Figure 4.12: The posterior distributions of μ and τ for M equals "1" and "10"

Chapter 5

Conclusion

We have considered a Bayesian analysis of binary and count data in clinical trials. For each type of data, Bayesian formulation was considered for testing hypothesis of equivalence. We observe that normal approximation to the beta posterior can used for moderately large sample sizes.

We also considered a meta analysis approach for data arising from multiple studies. In our example, the primary aim of the Meta-analysis of different studies on the impact of the treatment of interest (rosiglitazone) on myocardial infarction is to determine the overall effect. The individual studies used in the Meta-analysis reported different effects of rosiglitazone – some of which are positive and others negative.

The Bayes factor has been used to choose between the ordinary Dirichlet process and the conditional Dirichlet process as priors and based on the data, the conditional Dirichlet process is chosen. From the estimates obtained, the posterior probability that the overall relative risk is less than 1 is .83 which means that the use of rosiglitazone as a treatment for diabetes actually reduces the risk of myocardial infarction.

A clinical equivalent test procedure has been employed to test for the equivalence of treatment means. The estimates of the posterior means obtained from the Semiparametric model has been used to do an equivalent test. The test concludes that all treatment means are not the same and therefore fitting a random effects model to the data is appropriate.

The conclusion of the Meta–analysis varies from the conclusion from Maximum likelihood method called the Dersimonian–Laird method. Where as the Meta–analysis concludes that rosiglitazone reduces the myocardial infarction, Dersimonian–Laird method gives the summary odds ratio to be 1.21 which means rosiglitazone increases the risk of myocardial infarction by 21%.

We would like to pursue some future work along the methods discussed in the thesis. W are interested in enhancing the method to accommodate extra covariates into the model as well as when there are multiple treatments in one arm. The incorporation of covariates makes the Bayes Factor inappropriate for model selection. We would like to examine the other model selection criterions in place of Bayes Factor.

How a model fits data can be summarized numerically by the weighted mean square error given as $T(y, \theta) = \frac{1}{n} \sum_{i=1}^{n} (y_i - E(y_i|\theta))^2 / var(y_i)$. Another measure which is proportional to the mean square of the model is the deviance given as

$$D(y, \theta) = -2 \log p(y|\theta) \tag{5.1}$$

The disparity between data and the model fitted can be assessed by any measure of discrepancy but the deviance is a standard measure. For a

measure of the disparity that depends only on data y and independent of θ, the quantity $D_{\hat{\theta}}(y) = D(y, \hat{\theta}(y))$ can be used. A point estimate of θ for instance the median can be used in the above formula.

The above disparity can be averaged as follows:

$$D_{\text{avg}}(y) = E(D(y, \theta)|y) \tag{5.2}$$

An estimate of the average in 5.2 is obtained using posterior simulations θ_l and this estimate is given as :

$$\hat{D}_{avg}(y) = \frac{1}{L} \sum_{l=1}^{L} D(y, \theta^l)$$

"The expected deviance — computed by averaging out the deviance over the sampling distribution $f(y)$ — equals 2 times the Kullback-Leibler information, up to a fixed constant , $\int f(y) \log f(y) dy$ which does not depend on θ . In the limit of large sample sizes, the model with the lowest Kullback-Leibler information — and thus , the lowest expected deviance will have the highest posterior probability " [9] The difference between the estimated posterior mean deviance and the deviance at $\hat{\theta}$ is used as a measure of the effective number of parameters that should be in the model. This is represented as :

$$p_D^{(1)} = \hat{D}_{\text{avg}}(y) - D_{\hat{\theta}}(y) \tag{5.3}$$

A relative measure of model complexity is calculated as half the posterior variance of the deviance which is estimated from the posterior simulations

and given by the formula:

$$p_D^{(2)} = \frac{1}{2} \frac{1}{L-1} \sum_{l=1}^{L} (D(y, \theta^l) - \hat{D}_{\text{avg}}(y))^2$$

In hierarchical models, the effective number of parameters is greatly influenced by the variance of the group-level parameters. Another approach to measuring the disparity between data and the fitted model is by estimating the error anticipated when the model is applied to future data for instance the expected mean squared predictive error, $D_{\text{avg}}^{\text{pred}}(y) = E[\frac{1}{n} \sum_{i=1}^{n} (y_i - E(y_i|y))^2]$, where the expectation averages over the posterior predictive distribution of replicated data y^{rep}. The expected deviance for replicated data can be computed as

$$D_{\text{avg}}^{\text{pred}} = E\left[D(y^{\text{rep}}, \hat{\theta}(y))\right]$$

where $D(y^{\text{rep}}, \theta) = -2 \log p(y^{\text{rep}}|\theta)$, and $\hat{\theta}$ a parameter estimate such as the mean. The expected predictive deviance $D_{\text{avg}}^{\text{pred}}$ is usually greater than the expected deviance $\hat{D}_{\text{avg}}$ since the predictive data y^{rep} are being compared to a model estimated from data y. The expected predictive deviance $D_{\text{avg}}^{\text{pred}}$ has been recommended as a yardstick of model fit when the aim is to pick a model with best out-of-sample predictive power [9]. An estimate for the expected predictive deviance is called the deviance information criterion (DIC):

$$DIC = \hat{D}_{\text{avg}}^{\text{pred}}(y) = 2\hat{D}_{\text{avg}} - D_{\hat{\theta}}(y)$$

The Akaike Information Criterion is based on the Kullback–Leibler (K-L) information . The K–L information is a measure (a distance in an heuristic sense) between conceptual reality, f and approximating model, g, and is defined for continuous functions as the integral

$$I(f,g) = \int f(x) \log_e \left(\frac{f(x)}{g(x|\theta)} \right) dx$$

where f and g are n–dimensional probability distributions, $l(f,g)$ represent a measure of the information lost in approximating the real model f by g[3].

The goal here is to look for an approximating model that loses as little information as possible which is equivalent to minimising $l(f,g)$ over the set of models of interest. The link between K–L information and maximum likelihood estimation which makes it possible to bring estimation and model selection under one framework is called optimization. The estimator of the expected relative K–L information is based on the maximised log–likelihood function. The derivation is an asymptotic result (for large samples) and relies on the K–L information as an averaged entropy and this lead to Akaike's information criterion (AIC) given as

$$\text{AIC} = n \log_e(L(\hat{\theta}|\text{data})) + 2K$$

where $\log_e(L(\hat{\theta}|\text{data}))$ is the value of the maximised log-likelihood over the unknown parameters (θ), given the data and the model, and K is the number of estimable parameters in that approximating model. In a linear

model with normally distributed errors for all models under consideration, the AIC is stated as:

$$\text{AIC} = n \log(\hat{\theta}) + 2K$$

where $\hat{\sigma}^2 = \frac{\sum \hat{\epsilon}^2}{n}$. The model with the smallest AIC is comparatively better than all others and is the one selected." The AIC is asymptotically efficient but not consistent and can be used to compare non-nested models. A substantial advantage in using information-theoretic criteria is that they are valid for nonnested models. Of course, traditional likelihood ratio tests are defined only for nested models, and this represents another substantial limitation in the use of hypothesis testing in model selection " [3].

Table 5.1: Table showing empirical support for AIC

AIC_i - AIC_{min}	Level of Empirical Support for Model i
$0 - 2$	Substantial
$4 - 7$	Considerably Less
≥ 10	Essentially None

From Table 5.1, small values of AIC between 0 and 2 provides substantial evidence in support of the model under consideration. Large values of AIC gives considerably less evidence in support of the model. The BIC as well as the AIC is a classical way of estimating the dimension of a model . By the maximum likelihood principle, the model for which $\log M_j(X_1, \ldots, X_n) - \frac{1}{2}k_j \log n$ is the largest should be chosen [23] . In choosing among different models, the likelihood function for each model is maximized to get a Maximum Likelihood Estimate (MLE) of the form $M_j(X_1, \ldots, X_n)$ and k_j is the dimension of the jth model. This result has

been validated by as a large sample version of the Bayes procedure.

Chapter 6

Appendix

```
######################################################################
To install and load packages required to estimate odds by
 the  Mantel-Haenszel method
######################################################################

install.packages("HSAUR2")

library("HSAUR2")

install.packages("rmeta")

library("rmeta")

######################################################################
 R code to estimate odd ratios by the  Mantel-Haenszel method
######################################################################
a <- read.table("C:\\Users\\Cynthia\\Desktop\\Thesis\\data.txt",
header=TRUE)

aOR <- meta.MH(a[["tt"]], a[["tc"]],
               a[["qt"]], a[["qc"]],
```

```
                        names = rownames(a))

summary(aOR)

O <- summary(aOR)


##########################################################################
 R code to make a Forest Plot of the Rosiglitazone data
 by Mantel-Haenszel method
##########################################################################
pdf('forestplot_A.pdf',width=7,height=13)

plot(aOR, ylab = "",cex.lab=0.05)

dev.off()

getwd()



##########################################################################
 R code to estimate Odds Ratios for the modified data
##########################################################################
a1 <- read.table("C:\\Users\\Cynthia\\Desktop\\Thesis\\data1.txt",

header=TRUE)

aO1R <- meta.MH(a1[["tt"]], a1[["tc"]],

              a1[["qt"]], a1[["qc"]],

              names = rownames(a1))

summary(aO1R)


a1DSL <- meta.DSL(a1[["tt"]], a1[["tc"]],

              a1[["qt"]], a1[["qc"]],

              names = rownames(a1))
```

```r
print(a1DSL)

 pdf('forestplotmodified.pdf',width=7,height=15)

plot(aO1R, ylab = "",cex.lab=0.05)

dev.off()

getwd()


pdf('funnelplot_B.pdf',width=7,height=7)

funnelplot(a1DSL$logs, a1DSL$selogs,

           summ = a1DSL$logDSL, xlim = c(-1.7, 1.7))

abline(v = 0, lty = 2)

dev.off()

getwd()

#######################################################################
 Bayesian analysis
 To install package required for the Bayesian Semi-parametric model
#######################################################################
install.packages("bspmma")

library("bspmma")

Ba <- read.table("C:\\Users\\Cynthia\\Desktop\\Thesis\\data2.txt",

header=TRUE)

Ba.new <- as.matrix(Ba)

attach(Ba)

## R code to change data to the log of odd ratios and standard errors
```

```
Bam <- data.frame(OR, lower, upper)

se <- (upper -lower)/3.92

OR1 <- log(OR)


#######################################################################
 R code to compute and make a plot of Bayes factors
#######################################################################
Ba <- read.table("C:\\Users\\Cynthia\\Desktop\\Thesis\\Bayesdata.txt",

header=TRUE)

rosiglitazone.data <- as.matrix(Ba)

chain1.list <- bf1(rosiglitazone.data)

cc          <-  bf2(chain1.list)

chain2.list <- bf1(rosiglitazone.data, seed=2)

rosiglitazone.bfc <- bf.c(to=20, cc=cc, mat.list=chain2.list)

draw.bf(rosiglitazone.bfc)


#######################################################################
 R code to compute Bayes for choosing between Conditional
 and Ordinary Dirichlet Models
#######################################################################
load("rosiglitazone-rdat-2lists-1000")

rosiglitazone.bfco <- bf.c.o(to=20, cc=cc, mat.list=chain2.list)

draw.bf(rosiglitazone.bfco)


#######################################################################
R code to generate MCMC chians, plot autocorrelation,
```

```
obtain posterior descriptives and graph of mu and tau

##################################################################

install.packages("bspmma")

library("bspmma")


Alt <- read.table("Altered.txt",header=FALSE)

rosiglitazone <- as.matrix(Alt)


set.seed(1)

Alt.c5 <- dirichlet.c(rosiglitazone, ncycles = 4000, M =1,

d=c(.1,.1, 0, 1000))

set.seed(1)

Alt.c6 <- dirichlet.c(rosiglitazone , ncycles = 4000, M =10,

d=c(.1,.1, 0, 1000))


pdf('Autocorrelation3.pdf',width=7,height=7)

Alt.coda <- mcmc(Alt.c5$chain)

autocorr.plot(Alt.coda[, 15:19])

dev.off()


## R code to make Graphs of mu and tau

Alt.c5c6 <- list("1" =Alt.c5$chain, "10" = Alt.c6$chain)

pdf('Graph3.pdf',width=6,height=6)

draw.post(Alt.c5c6, burnin = 100)

dev.off()
```

```
describe.post(Alt.c5c6, burnin = 100)

data3<-capture.output(describe.post(Alt.c5c6, burnin = 100))

cat(data3,file="estimate3.txt",sep="\n",append=TRUE)


chain1.list <- bf1(rosiglitazone, ncycles = 5000, burnin = 1000)

cc          <-  bf2(chain1.list)

chain2.list <- bf1(rosiglitazone, seed=2, ncycles = 5000, burnin = 1000)

rosiglitazone.bfco <- bf.c.o(from =0.8, incr = 0.2, to = 20, cc = cc,

mat.list = chain2.list)

pdf('BayesModel.pdf',width=6,height=6)

draw.bf(rosiglitazone.bfco)

dev.off

getwd()


sd(Alt.c6$chain)

sigma10_i <- capture.output(sd(Alt.c6$chain))

cat(sigma10_i,file="standarddeviation.txt",sep="\n",append=TRUE)


rosiglitazone.bfc <- bf.c(df=-99, from = 0.8, incr = 0.2, to = 20, cc =cc,

 mat.list = chain2.list)

pdf('BayesM.pdf',width=6,height=6)

draw.bf(rosiglitazone.bfc)

dev.off()

getwd()


rosiglitazone.bfc$y[9]/rosiglitazone.bfc$yinfinity

value <- capture.output(rosiglitazone.bfc$y[9]/rosiglitazone.bfc$yinfinity)
```

```
cat(value,file="Bayesfactor.txt",sep="\n",append=TRUE)

set.seed(1)
Alt.c7 <- dirichlet.o(rosiglitazone, ncycles = 4000, M =1,
d=c(.1,.1, 0, 1000))
Alt.c7<-matrix(Alt.c7)
set.seed(1)
Alt.c8 <- dirichlet.o(rosiglitazone , ncycles = 4000, M =10,
d=c(.1,.1, 0, 1000) )
Alt.c8<-matrix(Alt.c8)

Alt.c7c8 <- list("1"=Alt.c7$chain, "10"=Alt.c8$chain)
Alt.c7
pdf('Grapho.pdf',width=6,height=6)
draw.post(Alt.c7c8, burnin = 100)
dev.off()

describe.post(Alt.c7c8, burnin = 100)
colnames(Alt.c7c8) <-c(Alt.c7,Alt.c8)

rosiglitazone.bfco <- bf.c.o(from = 0.8, incr = 0.2, to = 20,
cc = cc, mat.list = chain2.list)
pdf('BayesMo.pdf',width=6,height=6)
draw.bf(rosiglitazone.bfco)

#####################################################################
Simulation Study
#####################################################################
```

```
qt  <- c(rbinom(5, 200, 0.7))

qc <- c(rbinom(5, 200, 0.3))

tt <- rep(200, 5)

tc <- rep(200, 5)

Sdata  <-  cbind(tt, qt, tc, qc,  deparse.level = 1)

Sdata

Sdata0 <- capture.output(Sdata)

cat(Sdata0, file="Sdata.txt",sep="\n",append=TRUE)


Sdata1 <- read.table("C:\\Users\\Cynthia\\Desktop\\Thesis\\Sdata.txt",

header=TRUE)


Sdata1OR  <- meta.MH(Sdata1[["tt"]], Sdata1[["tc"]],

                     Sdata1[["qt"]], Sdata1[["qc"]],

                     names = rownames(Sdata1))

summary(Sdata1OR)

SMH <- capture.output(summary(Sdata1OR))

cat(SMH, file="S_MH.txt",sep="\n",append=TRUE)


nh5 <- read.table("C:\\Users\\Cynthia\\Desktop\\Thesis\\S.txt",

header=TRUE)

attach(nh5)

Sdata2 <- data.frame(OR, lower, upper)

se <- (upper -lower)/3.92

OR1 <- log(OR)

Sdata.new  <- cbind(se, OR1, deparse.level = 1)
```

```r
Simulation <- capture.output(Sdata.new)

cat(Simulation, file="Simulated_D.txt",sep="\n",append=TRUE)


Sbinom <- read.table("Simulated_D.txt",header=TRUE)

Sbinom1 <- as.matrix(Sbinom)

set.seed(1)

Alt.c1 <- dirichlet.c(Sbinom1, ncycles = 4000, M =1,d=c(.1,.1, 0, 1000))

set.seed(1)

Alt.c2 <- dirichlet.c(Sbinom1 , ncycles = 4000, M =10,d=c(.1,.1, 0, 1000))

Alt.c1c2 <- list("1"=Alt.c1$chain, "10"=Alt.c2$chain)

describe.post(Alt.c1c2, burnin = 100)

Mean <- capture.output(describe.post(Alt.c1c2, burnin = 100))

cat(Mean, file="Smeans.txt",sep="\n",append=TRUE)

deviation <- capture.output(sd(Alt.c1$chain))

cat(deviation, file="Smeans.txt",sep="\n",append=TRUE)


qt  <- c(rbinrbinom(1, 200, 0.45),(1, 200, 0.7), rbinom(1, 200, 0.01),

 rbinom(1, 200, 0.9),   rbinrbinom(1, 200, 0.65), rbinom(1, 200, 0.2))

qc <- c(rbinom(5, 200, 0.3))

tt <- rep(200, 5)

tc <- rep(200, 5)

SdataH1  <-  cbind(tt, qt, tc, qc,  deparse.level = 1)

H1_D <- capture.output(SdataH1)

cat(H1_D, file="Simulated_H1.txt",sep="\n",append=TRUE)


SH1 <- read.table("C:\\Users\\Cynthia\\Desktop\\Thesis\\Simulated_H1.txt",
```

```
header=TRUE)

SH1OR  <- meta.MH(SH1[["tt"]], SH1[["tc"]],
                  SH1[["qt"]], SH1[["qc"]],
                  names = rownames(SH1))
summary(SH1OR)
SMH1 <- capture.output(summary(SH1OR))
cat(SMH1, file="S_MH1.txt",sep="\n",append=TRUE)

SD1 <- read.table("C:\\Users\\Cynthia\\Desktop\\Thesis\\S1.txt",header=TRUE)
attach(SD1)

Sdata3 <- data.frame(OR1, lower1, upper1)
se1 <- (upper1 -lower1)/3.92
OR2 <- log(OR1)
SdataH.new  <- cbind(se1, OR2, deparse.level = 1)
SimH <- capture.output(SdataH.new)
cat(SimH, file="SH2.txt",sep="\n",append=TRUE)

Sbinom1 <- read.table("SH2.txt",header=TRUE)
Sbinom2 <- as.matrix(Sbinom1)
set.seed(1)
Alt.c2 <- dirichlet.c(Sbinom2, ncycles = 4000, M =1,d=c(.1,.1, 0, 1000))
set.seed(1)
Alt.c3 <- dirichlet.c(Sbinom2 , ncycles = 4000, M =10,d=c(.1,.1, 0, 1000))
Alt.c2c3 <- list("1"=Alt.c2$chain, "10"=Alt.c3$chain)
describe.post(Alt.c2c3, burnin = 100)
```

```
Mean <- capture.output(describe.post(Alt.c2c3, burnin = 100))

cat(Mean, file="Smeans.txt",sep="\n",append=TRUE)

deviation1 <- capture.output(sd(Alt.c2$chain))

cat(deviation1, file="Smeans.txt",sep="\n",append=TRUE)


qt <- c(rbinom(1, 200, 0.7), rbinom(1, 200, 0.65), rbinom(1, 200, 0.02),
 rbinom(1, 200, 0.09),  rbinom(1, 200, 0.86), rbinom(1, 200, 0.01),
 rbinom(1, 200, 0.19), rbinom(1, 200, 0.35),  rbinom(1, 200, 0.49),
rbinom(1, 200, 0.80), rbinom(1, 200, 0.11), rbinom(1, 200, 0.55),
 rbinom(1, 200, 0.79), rbinom(1, 200, 0.27), rbinom(1, 200, 0.38),
 rbinom(1, 200, 0.43),rbinom(1, 200, 0.46), rbinom(1, 200, 0.22),
 rbinom(1, 200, 0.29), rbinom(1, 200, 0.63))

qc <- c(rbinom(20, 200, 0.3))

tt <- rep(200, 20)

tc <- rep(200, 20)

H_s  <-  cbind(tt, qt, tc, qc,  deparse.level = 1)

S_0 <- capture.output(H_s)

cat(S_0, file="Sdata20.txt",sep="\n",append=TRUE)


SH20 <- read.table("C:\\Users\\Cynthia\\Desktop\\Thesis\\Sdata20.txt",
header=TRUE)


SH20_mh  <- meta.MH(SH20[["tt"]], SH20[["tc"]],
                  SH20[["qt"]], SH20[["qc"]],
                  names = rownames(SH20))

summary(SH20_mh )
```

```
SMH20<- capture.output(summary(SH20_mh))

cat(SMH20, file="S_MH20.txt",sep="\n",append=TRUE)


Shbin <- read.table("C:\\Users\\Cynthia\\Desktop\\Thesis\\S_20.txt",

header=TRUE)

attach(Shbin)


Shbin2 <- data.frame(OR20, lower20, upper20)

se20 <- (upper20 -lower20)/3.92

OR_20 <- log(OR20)

Shbin3  <- cbind(OR_20, se20, deparse.level = 1)

Shbin4 <- capture.output(Shbin3)

cat(Shbin4, file="Shbin20.txt",sep="\n",append=TRUE)


Sbinom20 <- read.table("Shbin20.txt",header=TRUE)

Sbinom_20 <- as.matrix(Sbinom20)

set.seed(1)

Alt.c20 <- dirichlet.c(Sbinom_20, ncycles = 4000, M =1,

d=c(.1,.1, 0, 1000))

set.seed(1)

Alt.c21 <- dirichlet.c(Sbinom_20 , ncycles = 4000, M =10,

d=c(.1,.1, 0, 1000))


Alt.c20c21 <- list("1"=Alt.c20$chain, "10"=Alt.c21$chain)

describe.post(Alt.c20c21, burnin = 100)

Mean <- capture.output(describe.post(Alt.c20c21, burnin = 100))

cat(Mean, file="Smeans20.txt",sep="\n",append=TRUE)
```

```
sd(Alt.c21$chain)

sd20 <- capture.output(sd(Alt.c21$chain))

cat(sd20, file="Smeans20.txt",sep="\n",append=TRUE)


qt <- c(rbinom(1, 200, 0.7), rbinom(1, 200, 0.65), rbinom(1, 200, 0.02),
 rbinom(1, 200, 0.09), rbinom(1, 200, 0.86), rbinom(1, 200, 0.01),
 rbinom(1, 200, 0.19), rbinom(1, 200, 0.35), rbinom(1, 200, 0.49),
 rbinom(1, 200, 0.80), rbinom(1, 200, 0.11), rbinom(1, 200, 0.55),
 rbinom(1, 200, 0.79), rbinom(1, 200, 0.27), rbinom(1, 200, 0.38),
rbinom(1, 200, 0.43), rbinom(1, 200, 0.46), rbinom(1, 200, 0.22),
 rbinom(1, 200, 0.29), rbinom(1, 200, 0.63), binom(1, 200, 0.7),
 rbinom(1, 200, 0.65), rbinom(1, 200, 0.31), rbinom(1, 200, 0.09),
 rbinom(1, 200, 0.86), rbinom(1, 200, 0.53), rbinom(1, 200, 0.32),
 rbinom(1, 200, 0.35), rbinom(1, 200, 0.49), rbinom(1, 200, 0.10),
rbinom(1, 200, 0.7), rbinom(1, 200, 0.01), rbinom(1, 200, 0.52),
 rbinom(1, 200, 0.45), rbinom(1, 200, 0.2), rbinom(1, 200, 0.12),
  rbinom(1, 200, 0.06), rbinom(1, 200, 0.36),
 rbinom(1, 200, 0.44), rbinom(1, 200, 0.34))
qc <- c(rbinom(40, 200, 0.3))
tt <- rep(200, 40)
tc <- rep(200, 40)
H_s40  <-  cbind(tt, qt, tc, qc,  deparse.level = 1)
S_40 <- capture.output(H_s40)
cat(S_40, file="Sdata40.txt",sep="\n",append=TRUE)


SH40 <- read.table("C:\\Users\\Cynthia\\Desktop\\Thesis\\Sdata40.txt",
```

```
header=TRUE)

SH40_mh  <- meta.MH(SH40[["tt"]], SH40[["tc"]],
                    SH40[["qt"]], SH40[["qc"]],
                    names = rownames(SH40))
summary(SH40_mh )
SMH40<- capture.output(summary(SH40_mh))
cat(SMH40, file="S_MH40.txt",sep="\n",append=TRUE)

Sh40 <- read.table("C:\\Users\\Cynthia\\Desktop\\Thesis\\d40.txt",
header=TRUE)
attach(Sh40)

Shbin40 <- data.frame(OR40, lower40, upper40)
se40 <- (upper40 -lower40)/3.92
OR_40 <- log(OR40)
Shbin5  <- cbind(OR_40, se40, deparse.level = 1)
Shbin_40 <- capture.output(Shbin5)
cat(Shbin_40, file="Shb40.txt",sep="\n",append=TRUE)

Sbinom40 <- read.table("Shb40.txt",header=TRUE)
Sbinom_40 <- as.matrix(Sbinom40)
set.seed(1)
Alt.c40 <- dirichlet.c(Sbinom_40, ncycles = 4000, M =1,d=c(.1,.1, 0, 1000))
set.seed(1)
Alt.c41 <- dirichlet.c(Sbinom_40, ncycles = 4000, M =10,d=c(.1,.1, 0, 1000))
```

```r
Alt.c40c41 <- list("1"=Alt.c40$chain, "10"=Alt.c41$chain)

describe.post(Alt.c40c41, burnin = 100)

Mean <- capture.output(describe.post(Alt.c40c41, burnin = 100))

cat(Mean, file="Smeans40.txt",sep="\n",append=TRUE)

sd(Alt.c41$chain)

sd40 <- capture.output(sd(Alt.c21$chain))

cat(sd40, file="Smeans40.txt",sep="\n",append=TRUE)


Alt <- read.table("Altered.txt",header=FALSE)

y <- as.matrix(Alt)


################################################################
R code to Estimate parameters of the model by Gibbs Sampling
################################################################
mu0=0; sigma0=10000; eta=c=.001; lambda=d=.001; tau2=1; sigma2=1; mmu=0

n=nrow(y)

for(i in 1:20000){

  mui= rnorm(n,
             mean=(((tau2*(y[,1]+y[,2]))+sigma2*mmu)/(2*tau2+sigma2)),
             sd=sqrt((tau2*sigma2)/(2*tau2+sigma2)))

  mu =rnorm(1,
            mean=(tau2*mu0+sigma0*sum(mui))/((tau2+n*sigma0)),
            sd=sqrt((tau2*sigma0)/((tau2+n*sigma0))))

  phi=rgamma(1, shape=(n/2+eta), rate=2/(sum((mui -mu)^2)+2*lambda))
```

```
     mu0  = mu
   mmu   = mui
   tau2  = 1/phi
sigma0   = sigma0

  if(i%%10==0 | i==1)
  {print(c(i,mui[1],mu,tau2,sigma0))
   write(c(i,mui[1],mu,tau2,sigma0),
       file="c:\\result.out",append=T,ncol=5)}

}

xt <- c(2,2,1,0,1,0,1,5,1,1,0,2,2,2,2,1,1,2,3,0,0,0,1,
       1,0,2,1,1,1,1,0,1,0,1,1,1,1,0,1,1,15,27)
xc <- c(0,1,1,1,0,1,0,2,0,0,1,0,1,0,1,1,2,0,1,0,1,
       1,0,0,2,3,0,0,0,0,0,0,0,2,0,0,0,0,0,0,9,41)
nt <- c(357,391,774,213,232,43,121,110,382,284,294,563,
       278,418,395,203,104,212,138,196,122,175,
       56,39,561,116,148,231,89,168,116,1172,
       706,204,288,254,314,162,442,394,2635,1456)
nc <-c(176,207,185,109,116,47,142,114,384,135,302,142,279,
       212,198,106,99,107,139,96,120,173,58,38,276,111,
       143,242,88,172,111,377,325,185,280,272,154,160,112,
       124,1634,1895)

p1 <- xt/nt
p2 <- xc/nc
```

```r
alpha<-2

beta<-5

pc <- qbeta(p2, xc + alpha, nc+beta-xc )

pt <- qbeta(p1, xt + alpha, nt+beta-xt )

pt

pc

Pct <- capture.output(pt)

cat(Pct, file="Proportions.txt",sep="\n",append=TRUE)

Pct1 <- capture.output(pc)

cat(Pct1, file="Proportions.txt",sep="\n",append=TRUE)


#####################################################################
R code to calculate posterior probabilities for Equivalence test
#####################################################################
count = 0

H0_prob <-function(xc, xt, alpha, beta, nc, nt){

  for(i in 1:10000){

     pc[i] <- rbeta(1, xc + alpha, nc+beta-xc)

    pt[i] <- rbeta(1, xt + alpha, nt+beta-xt)

                  D[i] <- pt[i]-pc[i]

  count = ifelse(D[i] < 0.01 & D[i]>-0.01, count+1, count)

  }

  return(count)

}

R <- Probability(41,27,2,5,2895,1456)
```

R

```
################################################################
 R code to plot the prior, likelihood and posterior for thr Beta-binomial
################################################################
beta_binom<-function(n,y,a=1,b=1,main=""){
  #likelihood: y|p~binom(n,p)
  #prior: p~beta(a,b)
  #posterior: p|y~beta(a+y,n-y+b)

  p<-seq(0.001,0.999,0.001)
  prior<-dbeta(p,a,b)
  if(n>0){likelihood<-dbinom(rep(y,length(p)),n,p)}
  if(n>0){posterior<-dbeta(p,a+y,n-y+b)}

  #standardize!
  prior<-prior/sum(prior)
  if(n>0){likelihood<-likelihood/sum(likelihood)}
  if(n>0){posterior<-posterior/sum(posterior)}

  ylim<-c(0,max(prior))
  if(n>0){ylim<-c(0,max(c(prior,likelihood,posterior)))}

  plot(p,prior,type="l",lty=2,xlab="p",ylab="",main=main,ylim=ylim)
  if(n>0){lines(p,likelihood,lty=3)}
  if(n>0){lines(p,posterior,lty=1,lwd=2)}
  legend("topright",c("prior","likelihood","posterior"),
```

```
          lty=c(2,3,1),lwd=c(1,1,2),inset=0.01,cex=.5)

}

##

pdf('Plot1n1.pdf',width=7,height=8)

par(mfrow=c(2,2))

beta_binom(278,2,8.5,3.5,main="Prior: beta(8.5, 3.5), data: 2/278")

beta_binom(279,1,8.5,3.5,main="Prior: beta(8.5, 3.5), data: 1/279")

beta_binom(116,2,6,59,main="Prior: beta(6,59), data: 2/116")

beta_binom(111,3,6,59,main="Prior: beta(6,59), data: 3/111")

dev.off()

getwd()


## 49653/015n49653/080

pdf('Plot2.pdf',width=7,height=8)

par(mfrow=c(2,2))

beta_binom(395,2,2,20,main="Prior: beta(2,20), data: 2/395")

beta_binom(198,1,5,15,main="Prior: beta(5,15), data: 1/198")

beta_binom(102,1,6,45,main="Prior: beta(6,45), data: 1/104")

beta_binom(198,1,7,60,main="Prior: beta(7,60), data: 2/99")

dev.off()

getwd()


## 49653/211n49653/011

pdf('Plot3.pdf',width=7,height=7)

par(mfrow=c(2,2))
```

```
beta_binom(375,1,5,2,main="Prior: beta(2,5), data: 5/110")

beta_binom(176,0,2,5,main="Prior: beta(2,5), data: 2/114")

beta_binom(375,2,4,75,main="Prior: beta(4,75), data: 2/375")

beta_binom(176,0,4,75,main="Prior: beta(4,75), data: 0/176")

dev.off()

getwd()


############################################################
 R code to calculate posterior probabilities for Poisson model
############################################################
alpha=1

beta=1

xc= 6021

xt=5101

count = 0

probability <-function(xc, xt, alpha, beta){

  lambdac <- vector(length = 10000)

  lambdat <- vector(length = 10000)

  D       <- vector(length =10000)

  for(i in 1:10000){

    lambdac[i] <- rgamma(1, xc + alpha, beta +19)

    lambdat[i] <- rgamma(1, xt + alpha, beta+19)

    D[i] <- lambdat[i]-lambdac[i]

    count = ifelse(D[i] < 0.01 & D[i]>-0.01, count+1, count)

  }

  return(count)
```

```
}
R <- probability(5,10,1,2)

 R
R < probability

###########################################################
R code to plot the prior and posterior for Poisson likelihood and
Gamma prior
###########################################################

p_gamma <- function(y,a,b,main=""){
  #likelihood: y|lambda~Poisson(lambda)
  #prior: lambda~gamma(lambda)
  #posterior: lambda|y~gamma(a+y,n+b)
  a=2
  b=1
  n=19
lambda <- c(seq(1,10, length.out=1000))
  y <- c(0.18,0.22,0.19,0.55,1.17,1.70,1.79,
         1.20,1.20,0.02,0.04,0.03,0.38,1.13,
         1.73,2.12,2.43,2.53)
  prior <- dgamma(lambda,a,b)
  likelihood<- (lambda^(sum(y))*exp(-n*lambda))/prod(factorial(y))
         #likelihood <- exp(sum(y)*log(lambda) -n*lambda-sum(log(y)))

y1=sum(y)
```

```r
      posterior<-dgamma(lambda,a+y1,n+b)
  #loglikelihood <- log(loglikelihood1)
  # Standardize
#loglikelihood <- loglikelihood/sum(loglikelihood)
 # posterior <-posterior/sum(posterior)
  #ylim<-c(0,max(prior))
  ylim<-c(0,max(c(prior,likelihood,posterior)))

  plot(lambda,prior,type="l",lty=2,xlab="lambda",ylab="")
      #lines(lambda,likelihood,lty=3)
      lines(lambda,posterior,lty=1,lwd=2)
      legend("topright",c("prior","likelihood","posterior"),
            lty=c(2,3,1),lwd=c(1,1,2),inset=0.01,cex=.5)
}


Light<- c(0.18,0.22,0.19,0.55,1.17,1.70,1.79,
                1.20,1.20,0.02,0.04,0.03,0.38,1.13,
                1.73,2.12,2.43,2.53)
Heavy <- c(1.49,1.69,1.93,5.73,10.01,9.01,6.13,3.37,
                1.89,1.24,1.40,1.87,5.14,7.78,6.89,4.32,
                2.14,0.63)
## Plot of Prior and Posterior distributions
pdf('distrns.pdf',width=7,height=8)
par(mfrow=c(2,1))
p_gamma(Heavysmoke,2,1,main="beta(2,1)")
```

```
  p_gamma(Lightsmoke,3,2,main="beta(3,2)")

dev.off()

getwd()


## Histogram of Smoking datasets

pdf('histmoker1.pdf',width=7,height=8)

par(mfrow=c(1,2))

hist(Heavy,sub="Deaths of Heavy Smokers",main="",xlab="", ylab="")

hist(Light,sub="Deaths of Light Smokers",main="",xlab="",ylab="")

dev.off()

getwd()


pdf('trial1.pdf',width=7,height=8)

p_gamma(y<-c(14,6,8,15,18,24,52,53,127,252,364,491,638,655,712,652,527,493),1,1,be

dev.off()

getwd()


####################################################################
 R  code to compute the posterior probabability for testing
the equivalence of two Poisson rates
####################################################################
xl <- c(0.18,0.22,0.19,0.55,1.17,1.70,1.79,
        1.20,1.20,0.02,0.04,0.03,0.38,1.13,
        1.73,2.12,2.43,2.53)
xh<- c(1.49,1.69,1.93,5.73,10.01,9.01,6.13,3.37,
        1.89,1.24,1.40,1.87,5.14,7.78,6.89,4.32,
        2.14,0.63)
```

```r
xh <- rpois(30,1)

xl <- rpois(30,1.2)


xh <- rpois(30,1.5)

xl <-c(rpois(5,5),rpois(5,20),rpois(5,1),rpois(5,89),rpois(5,3),

rpois(5,200))

sum(xh)

sum(xl)

nh=30

nl=30

lambdah <- dgamma(xh,2,1)

lambdal <-dgamma(xl,2,1)

count = 0

posterior<-function(xh, xl, alpha, beta, nh, nl){

  lambdac <- vector(length = 100)

  lambdat <- vector(length = 100)

  D       <- vector(length =100)

  for(i in 1:100){

 lambdah[i] <- rgamma(1, 38+alpha, nh+beta)

    lambdal[i] <- rgamma(1, 1563+alpha, nl+beta)

    D[i] <- lambdah[i]-lambdal[i]

    count = ifelse(D[i] < 200 & D[i]>-200, count+1, count)

  }

  return(count)

}
```

```
posterior(xh,xl,2,1,30,30)

###################################################################
R code for aa plot of Normal approximation to Beta distribution
###################################################################
beta_approx <- function(alpha,beta){
 #a+xt = alpha
 #nt+b-xt = beta
 #a+b+nt= alpha+beta
 S=alpha+beta
  P_0 =(1-alpha)/(2-S)
sigma <- sqrt(-(2-S)^3/((1-beta)*(1-alpha)))
N <- c(P_0,sigma)
 return(N)
}

beta_approx1<-function(alpha=1,beta=1,main=""){
  S=alpha+beta
  P_0 =(1-alpha)/(2-S)
    sigma <-1/sqrt(-(2-S)^3/((1-beta)*(1-alpha)))
  p<-seq(0.001,0.999,0.001)
  Beta<-dbeta(p,alpha,beta)
  #T <- qnorm(p,P_0,sigma)
  if(n>0){Normal <-dnorm(p,P_0,sigma)}

  #standardize!
```

```r
 #Beta<-Beta/sum(Beta)
 #if(n>0){Normal<-Normal/sum(Normal)}

 ylim<-c(0,max(Beta))
 if(n>0){ylim<-c(0,max(c(Beta,Normal)))}

 plot(p,Beta,type="l",lty=2,xlab="p",ylab="",main=main,ylim=ylim)
 if(n>0){lines(p,Normal,lty=1,lwd=2)}
 legend("topright",c("Beta","Normal"),
        lty=c(2,1),lwd=c(1,2),inset=0.01,cex=.5)
}


pdf('Napprox.pdf',width=7,height=8)
par(mfrow=c(5,2))
beta_approx1(2,2,"Beta(2,2)")
beta_approx1(3,3,"Beta(3,3)")
beta_approx1(2,4,"Beta(2,4)")
beta_approx1(4,4,"Beta(4,4)")
beta_approx1(5,5,"Beta(5,5)")
beta_approx1(10,10,"Beta(10,10)")
beta_approx1(30,20,"Beta(30,20)")
beta_approx1(20,30,"Beta(20,30)")
beta_approx1(50,20,"Beta(50,20)")
beta_approx1(20,50,"Beta(20,50)")
dev.off()
```

```
getwd()

pdf('Napprox21.pdf',width=6,height=6.5)

par(mfrow=c(2,2))

beta_approx1(2,2,"Beta(2,2)")

beta_approx1(3,3,"Beta(3,3)")

beta_approx1(2,4,"Beta(2,4)")

beta_approx1(4,4,"Beta(4,4)")

dev.off()

getwd()

pdf('Napprox31.pdf',width=6,height=6.5)

par(mfrow=c(2,2))

beta_approx1(5,5,"Beta(5,5)")

beta_approx1(10,10,"Beta(10,10)")

beta_approx1(30,20,"Beta(30,20)")

beta_approx1(20,30,"Beta(20,30)")

dev.off()

getwd()

pdf('Napprox41.pdf',width=6,height=6)

par(mfrow=c(1,2))

beta_approx1(50,20,"Beta(50,20)")

beta_approx1(20,50,"Beta(20,50)")

dev.off()

getwd()
```

```
######################################################################
A 3D plot of the joint posterior of lambdat and lambdac
######################################################################
xl <- c(0.18,0.22,0.19,0.55,1.17,1.70,1.79,
        1.20,1.20,0.02,0.04,0.03,0.38,1.13,
        1.73,2.12,2.43,2.53)
xh<- c(1.49,1.69,1.93,5.73,10.01,9.01,6.13,3.37,
       1.89,1.24,1.40,1.87,5.14,7.78,6.89,4.32,
       2.14,0.63)
sum(xl)=18
sum(xh)=72.66

## plot of exact posterior
Pgpost <- function(lambdat,lambdac,alphat=2,alphac=2,betat=3
,betac=3,nt=18,nc=18)
{
P =  dgamma(lambdat,18+alphat,betat+nt)*dgamma(lambdac,
72.66+alphac,betac+nc)
}
xc =seq(2,5, length = 50)
xt  =seq(0,2, length = 50)
P = outer(xt,xc,Pgpost)
pdf('ppgamma1.pdf',width=6,height=8)
persp(xt,xc,P,theta = 45,phi=30,expand = 0.6,ltheta = 120, shade = 0.7,
 ticktype = "detailed",xlab="Treatment mean",
```

```r
ylab="Control mean", col="saddlebrown")
dev.off()
getwd()

Pgbeta <- function(Pt,Pc,alpha=2,beta=3,eta=2,epsilon=3)
{
  xt <- c(2,2,1,0,1,0,1,5,1,1,0,2,2,2,2,1,1,2,3,0,0,0,1,
          1,0,2,1,1,1,1,0,1,0,1,1,1,1,0,1,1,15,27)
  xc <- c(0,1,1,1,0,1,0,2,0,0,1,0,1,0,1,1,2,0,1,0,1,
          1,0,0,2,3,0,0,0,0,0,0,0,2,0,0,0,0,0,0,9,41)
  nt <- c(357,391,774,213,232,43,121,110,382,284,294,563,
          278,418,395,203,104,212,138,196,122,175,
          56,39,561,116,148,231,89,168,116,1172,
          706,204,288,254,314,162,442,394,2635,1456)
  nc <-c(176,207,185,109,116,47,142,114,384,135,302,142,279,
          212,198,106,99,107,139,96,120,173,58,38,276,111,
          143,242,88,172,111,377,325,185,280,272,154,160,112,
          124,1634,1895)

  Pt <- xt/nt
  Pc <- xc/nc

  Posterior =  dbeta(Pt,xt+alpha,beta+nt-xt)*dbeta(Pc,xc+epsilon,eta+nc-xc)
}
Pt <-seq(0, 1, length.out=1
Pc<- seq(0,1,length.out=42)
Posterior = outer(Pt,Pc,Pgbeta)
```

```r
pdf('ppbeta.pdf',width=7,height=8)
persp(xt,xc,Posterior,theta = 45,phi=30,expand = 0.6,ltheta = 120,
 shade = 0.7, ticktype = "detailed",xlab="Pt",ylab="Pc",
col="saddlebrown")
dev.off()
getwd()


#######################################################################
R code to estimate missing data in arm
#######################################################################
m = 20000  # no of mcmc
burnin = 10000 # burn-in length

# initial values
P= 0.5
xm = 1

# matrix for mcmc
Px = matrix(0, m , 2)

##data
y = 3
n = 111

Px[1,] = c(P, xm)

### generating the mcmc

for( i in 2:m){
```

```r
  P = Px[i-1,1]

  Px[i,2] = rbinom(1,1,P)

  Px[i,1] = rbeta(1, y+ Px[i,2]+1, n - y - Px[i,2] +1)

}

## get data after burn-in

b = burnin + 1

data = Px[b:m,]

### trace plots and acf for assessing convergence

burnin = b:m

index2 = 1:m

pdf('tracenacf1.pdf',width=5.5,height=8.5)

par(mfrow = c(2,1))

plot(burnin,data[,1],type ="l",xlab="iteration after burnin", ylab = "P")

acf(data[,1],main="")

dev.off()

getwd()

## Posterior summaries using MCMC after Burn-in

colMeans(data)

xm.freq = table(data[,2])

pdf('Histogrammcmc.pdf',width=4.8,height=8.5)

par(mfrow = c(2,1))

hist(data[,1],main = paste("Posteriors of parameters using MCMC"),
```

```
xlab = "P",ylab="")

barplot(xm.freq,xlab=expression(x[m]))

dev.off()

getwd()


Nap <- function(alpha=2,beta=3,xt,nt,xc,nc){

  mu1= (1-alpha-xt)/(2-alpha-nt-beta)

  sigma1= 1/sqrt(-(2-alpha-nt-beta)^3/((1-xt-alpha)*(1-nt-beta+xt)))

  mu2  = (1-alpha-xc)/(2-alpha-nc-beta)

sigma2 = 1/sqrt((-(2-alpha-nc-beta)^3/((1-xc-alpha)*(1-nc-beta+xc))))

 mu    = mu1 - mu2

sigma = sqrt(sigma1^2 + sigma2^2)

H0 = pnorm(-0.01,mu,sigma)

  H = 1-2*H0

  B   = (1-H)/H

  set = c(H,B)

return(set)

}
```

	nt	xt	nc	xc
49653/011	375	2	176	0
49653/020	391	2	207	1
49653/024	774	1	185	1
49653/093	213	0	109	1
49653/094	232	1	116	0
100684	43	0	47	1
49653/143	121	1	142	0
49653/211	110	5	114	2
49653/284	382	1	384	0
712753/008	284	1	135	0
AMM100264	294	0	302	1
BRL49653C/185	563	2	142	0
BRL49653/334	278	2	279	1
BRL49653/347	418	2	212	0
49653/015	395	2	198	1
49653/079	203	1	106	1
49653/080	104	1	99	2
49653/082	212	2	107	0
49653/085	138	3	139	1
49653/095	196	0	96	0
49653/097	122	0	120	1
49653/125	175	0	173	1
49653/127	56	1	58	0
49653/128	39	1	38	0
49653/134	561	0	276	2

49653/135	116	2	111	3
49653/136	148	1	143	0
49653/145	231	1	242	0
49653/147	89	1	88	0
49653/162	168	1	172	0
49653/234	116	2	111	3
49653/330	1172	1	377	0
49653/331	706	0	325	0
49653/137	204	1	185	2
SB-712753/002	288	1	280	0
SB-712753/003	254	1	272	0
SB-712753/007	314	1	154	0
SB-712753/009	162	0	160	0
49653/132	442	1	112	0
AVA100193	394	1	124	0
DREAM	2635	15	2634	9
ADOPT	1456	27	2895	41

Bibliography

[1] Betsy Jane Becker. Multivariate meta-analysis: Contributions of ingram olkin. *Statistical Science*, 22:401– 406, 2007. (Cited on page 34.)

[2] Michael Borenstein, Larry Hedges, and Hannah Rothstein. *Introduction to Meta-Analysis*. Wiley, 2009. (Cited on pages 11, 31, 33 and 34.)

[3] Kenneth Burnham and David Anderson. Kullback–leibler information as a basis for strong inference in ecological studies. *Wildlife Research*, 28:111–119, 2001. (Cited on pages 81 and 82.)

[4] Debora Burr and Hanani Doss. A bayesian semiparametric model for random effects meta–analysis. *Journal of American Statistical Association*, 100:242–251, 2005. (Cited on page 59.)

[5] Deborah Burr. An r package for bayesian semiparametric models for meta analysis. *Journal of Statistical Software*, 50, 2012. (Cited on page 59.)

[6] Francois Delahaye, Gilles Landrivon, and Cyrille Colin. Meta-analysis. *Health Policy*, 19:185–196, 1991. (Cited on pages 10, 12 and 31.)

[7] Kaul Diamond. Good enough : a primer on the analysis and interpretation of noninferiority trials. *Annals of Internal Medicine*, 145:62 – 69, 2006. (Cited on page 7.)

[8] Alan Gelfand. Gibbs sampling. *Journal of American Statistical Association*, 95, 2000. (Cited on pages 16 and 44.)

[9] Andrew Gelman, John Carlin, and Hal Stern. *Bayesian Data Analysis*. Chapman & Hall, 2004. (Cited on pages 8, 42, 61, 79 and 80.)

[10] Jeff Gill. *Bayesian Methods A social and Behavioural Sciences Approach*. Chapman & Chap/CRC, 2008. (Cited on page 9.)

[11] Leandro Giocchino. *Meta-analysis in Medical Research: The Handbook for the Understanding and Practice of Meta-analysis*. EBSCO, 2005. (Cited on pages 11, 13, 33, 35 and 56.)

[12] David V Hinkley. Likelihood. *The Canadian Journal of Statistics*, 8:151–163, 1980. (Cited on page 9.)

[13] Joseph KADANE and Nicole LAZAR. Methods and criteria for model selection. *Journal of the American Association*, 99, 2004. (Cited on page 45.)

[14] Mark Mamalo, Rui Wu, and Ram Tiwari. Bayesian approach to non-inferiority trials for proportions. *Journal of Biopharmaceutical Statistics*, 21:902–919, 2011. (Cited on page 6.)

119

[15] Mary MCHugh. Odds ratios and interpretation. *Biochemia Medica*, 19:120–6, 2009. (Cited on page 15.)

[16] Saman Muthukumarana and Ram C Tiwari. Meta-analysis using dirichlet process. *SAGE*, 10:1–14, 2012. (Cited on pages 36, 37 and 47.)

[17] McCullagh & Nelder. *Generalized Linear Models*. Chapman & Hall/CRC, 1989. (Cited on page 41.)

[18] Steven Nissen and Kathy Wolski. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *The New England Journal of Medicine*, 356:2457–2471, 2007. (Cited on page 54.)

[19] Jordi Ocana, Pilar Sanchez, and Alex Sanchez. On equivalence and bioequivalence testing. *SORT*, 32:151– 171, 2008. (Cited on page 5.)

[20] Emin Orhan. *Dirichlet Processes*. PhD thesis, Rochester University, 2012. (Cited on page 46.)

[21] Mehesh Patel. An introduction to meta–analysis. *Health Policy*, 11:79–85, 1988. (Cited on page 14.)

[22] N Reid. Likelihood. *Journal of the American Statistical Association*, 95:13335–1340, 2000. (Cited on page 9.)

[23] Gideon Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6:461 – 464, 1978. (Cited on page 82.)

[24] Yee Whye Teh, Michael Jordan, and Matthew Beal. Hierarchical dirichlet processes. *Journal of the Royal Statistical Society*, 61:487–527, 1999. (Cited on page 47.)

[25] Steve Wang. *Principles of Statistical Inference: Likelihood and the BayesianParadigm*, chapter 18, pages 1–18. Chapman & Hall, 1998. (Cited on pages 7 and 8.)

[26] Stefan Wellek. *Testing Statistical Hypothesis of Equivalence and Non-inferiority*. CRC Press, 2010. (Cited on pages 4 and 6.)

[27] S Zodpey. Meta-analysis in medicine. *RESEARCH METHDOLOGY*, 69:416–420, 2003. (Cited on pages 11 and 14.)