THE UNIVERSITY OF MANITOBA

LIBRARY

AUTHOR ..WEISS, Günter Max Theodor..

TITLE ....A LARGE DEVIATION STUDY OF CONSISTENT ESTIMATION OF A TRANSLATION

....INVARIANT LOCATION PARAMETER.......................................

....

THESIS ...Ph.D., 1983...

I, the undersigned, agree to refrain from producing, or reproducing, the above-named work, or any part thereof, in any material form, without the written consent of the author:

THE UNIVERSITY OF MANITOBA

A LARGE DEVIATION STUDY OF CONSISTENT ESTIMATION OF A TRANSLATION INVARIANT LOCATION PARAMETER

by

GÜNTER MAX THEODOR WEISS

A THESIS

SUBMITTED TO THE FACULTY OF

GRADUATE STUDIES IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Department of Statistics

Winnipeg, Manitoba

February, 1983

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a Ph.D. thesis entitled: A Large Deviation Study of Consistent Estimation of a Translation Invariant Location Parameter

submitted by Günter Max Theodor Weiss

in partial fulfilment of the requirements for the Ph.D. degree.

James C. Fu
Advisor

J. Brewster

A. Rukhin

E. Shim

Dr. A. Rukhin
External Examiner
Dept. of Statistics
Purdue University
West Lafayette, In. U.S.A.

Date of oral examination: March 29th, 1983.

The student has satisfactorily completed and passed the Ph.D. oral examination.

James C. Fu
Advisor

J. Brewster

A. Rukhin

E. Shim

G. Williams
Chairman of Ph.D. Oral*

(*The signature of the Chairman does not necessarily signify that the Chairman has read the complete thesis.)

A LARGE DEVIATION STUDY

OF CONSISTENT ESTIMATION

OF A TRANSLATION INVARIANT

LOCATION PARAMETER

BY

GÜNTER MAX THEODOR WEISS

A thesis submitted to the Faculty of Graduate Studies of the University of Manitoba in partial fulfillment of the requirements of the degree of

DOCTOR OF PHILOSOPHY

ACKNOWLEDGMENTS

I wish to thank Dr. Kathleen Subrahmaniam (Kocherlakota) for introducing me to the Wonderful World of Statistics, Dr. John F. Brewster for the many hours that he has spent both as my supervisor in my Masters program and as a member of my PhD committee; most of all, I wish to thank Dr. James C. Fu for suggesting this topic, for the many enlightening discussions we have had, and for his careful guidance, advice, and supervision of my research.

CONTENTS

ABSTRACT

R. A. Fisher (1922) defined the concept of the efficiency of estimators of a parameter $\theta \in \Theta \subset R$ in terms of limiting Normal distributions. As this definition entails several deficiencies, one remedy was to compare the behaviours of the estimators in the tails of their distributions since any reasonable estimator should be well-behaved near the true value of θ (eg: the Central Limit Theorem). This has led to the establishment of the Theory of the Probability of Large Deviations.

For consistent estimator-sequences $\{T_n\} = T$, the tail probability

$$a_n(T_n,\theta,\varepsilon) = P_\theta[\,|T_n - \theta| \geq \varepsilon\,],$$

tends to zero as n tends to infinity. Basu (1956) proposed that the rate at which $a_n \to 0$ be used as a measure of the asymptotic behaviour of the estimator. Since this rate is typically exponential in nature, Bahadur (1971) suggested that we compute the exponential rate of convergence:

$$b(T,\theta,\varepsilon) = \lim_{n\to\infty} -\frac{1}{n} \log a_n(T_n,\theta,\varepsilon).$$

Fu (1971) showed that for any consistent estimator-sequence, the exponential rate is bounded above by

$$B(\theta,\varepsilon) = \inf_{\theta^*}\{E_{\theta^*}\log \frac{f(X|\theta^*)}{f(X|\theta)} : |\theta-\theta^*| > \varepsilon\},$$

which is called the Bahadur bound. A theorem by Chernoff (1952) as modified by Bahadur (1960) is often used to compute the exponential rate in particular cases.

The Method of Maximum Likelihood Estimation has been shown to possess many desirable properties. Indeed, the Maximum Likelihood Estimator (mle) does have a limiting Normal distribution and is Fisher-efficient. Also, Rao (1961, 1962, 1963) shows that the mle, $\hat{\theta}$, is generally favoured among efficient estimators in that $\hat{\theta}$ has minimum loss of sample information on θ and is thus second-order efficient. This property is measured in terms of a parameter γ^2, where γ was named the statistical curvature by Efron (1975).

In this thesis, we consider a translation invariant location parameter $\theta \in R^1$. We propose a competitor to the mle in the form of a Probability Ratio Estimator (pre), which is defined in §2.1. We show that, subject to certain regularity conditions on the underlying distributions, the pre is also second-order efficient.

Also, for symmetric densities, the pre is shown to 'dominate' the mle in the sense that the exponential rate of the pre is optimal for the class of translation invariant estimators of which the mle is a member. Some examples are also provided to support the contention that the pre also dominates when the symmetry is lacking.

Then, as the local behaviour of the exponential rates as $\varepsilon \to 0$ is of especial interest, we felt that some insight could be provided by the Taylor Series expansions of the exponential rates. This follows the work of Fu (1982) who obtained expansions up to the fourth-degree, which is sufficient for second-order efficiency comparisons. Since both the mle and the pre are second-order efficient, we were required to find expansions up to the sixth-degree in order to obtain theoretically meaningful differences. Obtaining these expansions led us to define several constants, in particular, a parameter, λ^3, which appears to be a cubic extension of the quadratic parameter γ^2, the squared-curvature. We then considered several practical cases, including the Logistic and Hyperbolic Secant distributions, which numerically demonstrate that, although the mle is second-order effient, it is unlikely that it is third-order efficient (a concept that is as yet undefined). We also consider cases where some of the regularity conditions do not hold. In the semi-regular case of a mixture of Normals, we obtain, through direct computation, results similar to those in the Logistic case. For the non-regular Double Exponential distribution, the pre still seems to dominate the mle (in the sense of exponential rate), although here neither estimator-sequence appears to be second-order efficient (likely due to the non-regularity). An appendix is provided which gives details of proofs and computations, and provides some useful reference tables and equations.

CHAPTER ONE
INTRODUCTION

1.1 Large Sample Theory and Large Deviations

Until fairly recently most statistical estimation and inference was concerned with small deviations about the mean. The study of large samples has led to the Central Limit Theorem (CLT) which basically states that for a sample of independent observations on a population with mean $\theta<\infty$, the sampling distribution of the sample mean will tend to a Normal Distribution, in the neighbourhood of θ, as the sample size becomes large. In this way, Fisher (1922) defines the *CRITERION OF EFFICIENCY* as being satisfied by all statistics (estimating the same quantity) which, when derived from large samples tend to a Normal distribution with the least possible variance. Some major limitations are apparent:

1. How large is large? That is, how large a sample size is required before we can apply these large sample results (such as CLT). This is especially important in the tail regions of the probability distribution, away from θ.

2. How relevant are asymptotic results when dealing with finite sample sizes? How efficient is an estimator derived from a finite sample? Or, in another sense, how can we choose between estimators which satisfy Fisher's Criterion of Efficiency when we have finite samples?

3. What do we do in the case of non-Normal limiting distributions? That is, how can we define the efficiency of estimators which do not tend to Normality?

A remedy, in the small sample case, has been to consider *ROBUST PROCEDURES*. Here, one deals with tests and estimators which are not greatly affected by small changes in the underlying distribution [see Huber (1972, 1977)]. One unfortunate tendency of these procedures is that many of these methods will tend to discard the observations in the tails of the distribution ("the outliers"). A recent paper by Stigler (1977) suggests, however, that in many real data situations, the non-robust sample mean seems to perform almost as well as the best robust estimator (of those considered), and much better than most of the other robust estimators, in particular, the median, which does quite poorly overall. This may be partly due to the fact that many of the robust procedures were designed to handle hypothetical situations which rarely arise in the real world (e.g. the Cauchy and Double Exponential Distributions).

Another approach to remedy these limitations of Fisher's definition is to examine the behaviour of the estimators in the tails of the distribution. This has led to the foundation of Large Deviation Theory.

Consider a sample $S = \{X_i\}$ of n observations with mean θ and variance $\sigma^2 < \infty$. By the Central Limit Theorem, $\bar{X}_n = \Sigma X_i/n$ will tend to $N(\theta, \sigma^2/n)$ as $n \to \infty$. Hence, at a deviation of $\varepsilon/\sqrt{n}$ from the mean, we will get a fixed distribution as $n \to \infty$. If, however, we take some fixed point, $\theta + \varepsilon$, the tail of the distribution will get smaller and smaller as the sample size increases [see Figure 1.1]. Suppose, we have several "efficient" estimators with distributions tending to $N(\theta, \sigma^2/n)$. Then we would wish to select the estimator with the fastest rate of convergence to this limiting Normal distribution. One way of measuring the rate of this convergence is to measure the rate at which the tail probability is tending to zero. This rate is suggested by Basu (1956) as a measure of the asymptotic accuracy of the statistics. He calls this rate the <u>concentration</u>. Large Deviation Theory is then largely concerned with measuring the rate of

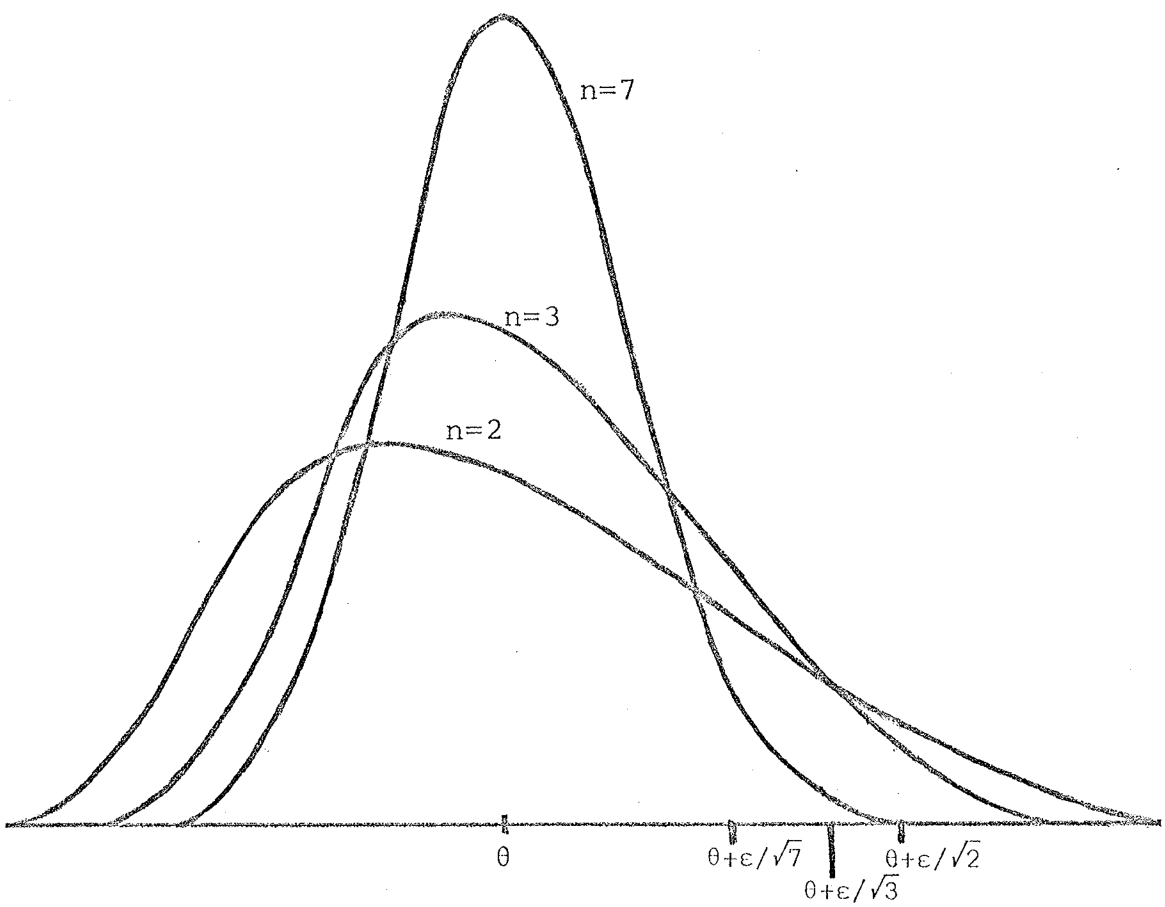

FIGURE 1.1: Convergence to Normality in the small deviation.

convergence to zero of the tail probability of a statistic.

In general, if T_n is any point estimate of θ, we have, for $a_n > 0$,

$$A_n = P(T_n - \theta \geq a_n). \tag{1.1}$$

Then, we define:

1. If $a_n = O(\sqrt{n})$ then a_n is an ordinary or small deviation. [The notation $O(r_n)$ means $O(r_n)/r_n$ is bounded and $o(r_n)$ means $o(r_n)/r_n \to 0$ as $n\to\infty$. See Chernoff (1956).]

2. If $a_n = O(\sqrt{\log n/n})$, then a_n is a moderate deviation, a concept introduced by Rubin & Sethuraman (1965).

3. If $a_n = O(1)$, then a_n is a large deviation.

For an introduction to the theory of large and moderate deviations see Sethuraman (1970). Some more advanced theory is provided by Chernoff (1952) and Bahadur (1971).

1.2 Review of the Literature

The first major result in Large Deviation Theory was an extension, by Harald Cramér, in 1938, of the Central Limit Theorem [see Chernoff (1956) or Sethuraman (1970)], which may be stated as:

Theorem 1.1 [Cramér, 1938] *Let* $\{X_i\}$ *be a random sample from a distribution with mean* θ *and variance* σ^2*, and with* $Ee^{tX}<\infty$ *in some neighbourhood of* $t=0$*. If* $a_n>0$ *and* $a_n = o(1)$*, then*

$$P((\bar{X}_n - \theta)/\sigma \le -a_n) = \Phi(-a_n\sqrt{n})e^{-na_n^3\lambda(-a_n)}[1+O(a_n)] \qquad (1.2)$$

where $\lambda(t)$ *is a power series in* t *which involves the cumulants of* X *and converges for small* t *(called the* Cramér Series*).*

Chernoff (1956) obtains the result for the important special case when $a_n = a = -(c - \theta)/\sigma$,

Theorem 1.2 [Chernoff, 1956] *If* $Ee^{tX}<\infty$ *for* t *in some neighbourhood of zero, and* $c<\theta$*, then*

$$P(\bar{X}_n \le c) = \rho^n/\sqrt{n}(b + O(1/n)) \qquad (1.3)$$

where $b>0$ *and* $\rho = \inf_t Ee^{t(X - c)}$.

Chernoff (1952, 1956) also obtains the rates of convergence of the error probabilities in testing a simple hypothesis H_0 versus a simple alternative H_1, with a test of the form:

$$\text{"Reject } H_0 \text{ if } \bar{Y}_n \ge k_n\text{"} \qquad (1.4)$$

where $\bar{Y}_n = \Sigma Y_i/n$ is the mean of n independent observations on a random variable Y where $\mu_0 = E(Y|H_0) < E(Y|H_1) = \mu_1$. For example, the likelihood ratio test is of this form, where $Y = \log [f_1(X)/f_0(X)]$. The error probabilities are then given by:

$$\alpha_n = P(\bar{Y}_n \geq k_n | H_0), \tag{1.5}$$

$$\beta_n = P(\bar{Y}_n < k_n | H_1). \tag{1.6}$$

Chernoff then notes that the selection of k_n to minimize the risk with respect to some arbitrary convex loss function is equivalent to the selection of k_n to minimize the linear combination, $\beta_n + \lambda\alpha_n$, for some $\lambda > 0$, and he hence obtains the result:

$$\lim_{n\to\infty} [\inf_{k_n} (\beta_n + \lambda\alpha_n)]^{n^{-1}} = \rho, \tag{1.7}$$

where ρ is independent of λ, and

$$\rho = \inf_{\mu_0 \leq a \leq \mu_1} \rho(a), \tag{1.8}$$

$$\rho(a) = \max[m_0(a), m_1(a)], \tag{1.9}$$

$$m_i(a) = \inf_t E[e^{t(Y - a)} | H_i], \; i=0,1. \tag{1.10}$$

Chernoff then suggests that $-\log \rho$ be considered as a kind of information measure for a test of hypothesis based on the sums of observations, and hence defines the relative efficiency of two tests with indices ρ and ρ^* respectively, as

$$e = \frac{\log \rho^*}{\log \rho}. \tag{1.11}$$

Kallenberg (1982) investigates some of the properties of "Chernoff's Efficiency". Then, in particular, for the likelihood-ratio test,

$$\rho_{LR} = \inf_{0<t<1} \int_{-\infty}^{\infty} (f_1(x))^t (f_0(x))^{1-t}\, dx, \qquad (1.12)$$

where f_0, f_1 are the densities of X under H_0 and H_1 respectively. "Chernoff's Information", $-\log \rho_{LR}$, is sub-additive in that:

(1) Information on m replications of the same random variable is m times the information on one single observation.

(2) Information on the sum of independent observations of several random variables is less than or equal to the sums of the corresponding information.

Chernoff (1956) also considers the test for which k_n is chosen to minimize β_n for a fixed $\alpha_n = \alpha_0$ and he shows that in this case

$$\lim_{n\to\infty} [\inf_{k_n} \beta_n]^{n^{-1}} = \rho^* = m_1(\mu_0), \qquad (1.13)$$

and, for the likelihood-ratio test, the corresponding measure of information is

$$-\log \rho^*_{LR} = \int_{-\infty}^{\infty} f_0(x) \log\left[\frac{f_0(x)}{f_1(x)}\right] dx, \qquad (1.14)$$

which is the information measure for the discrimination between H_0 and H_1, as defined by Kullback & Leibler (1951). The Kullback-Leibler information numbers are additive and play an important rôle in large deviation theory. An interesting discussion on the physical interpretation of Kullback-Leibler information numbers is given in Hoeffding (1979, §8.12).

In this thesis, we will be concerned mainly with large deviation theory applied to the problem of estimating a one-dimensional parameter $\theta\in\Theta\equiv\mathbb{R}^1$. In general, a large deviation set is any set whose closure does not contain θ. In this case, the large deviation sets of the greatest interest are the tail sets, $\mathcal{D}_\varepsilon(\theta)$, where

$$\mathcal{D}_\varepsilon(\theta) = \{x\in\mathbb{R}^1 | x \le \theta - \varepsilon,\ x \ge \theta + \varepsilon\}. \tag{1.15}$$

The tail probability of an estimator $T = \{T_n(X_1,\ldots,X_n)\}$ of θ is then given by $a_n(T_n,\theta,\varepsilon)$,

$$a_n(T_n,\theta,\varepsilon) = P_\theta(T_n(X_1,\ldots,X_n) \in \mathcal{D}_\varepsilon(\theta)). \tag{1.16}$$

Then, for large samples, we need only be concerned with estimator-sequences which are consistent for θ, that is, $T_n(X_1,\ldots,X_n) \xrightarrow{p} \theta$ as $n\to\infty$, and thus, for which $a_n(T_n,\theta,\varepsilon)$ tends to zero. As mentioned, the rate at which this quantity tends to zero has been proposed by Basu (1956) as measure of the asymptotic performance of an estimator.

But since in typical cases the tails of the probability distributions are exponential in nature,

$$a_n(T_n,\theta,\varepsilon) = e^{-n[b(T,\theta,\varepsilon) + o(1)]}. \qquad (1.17)$$

Thus, Bahadur (1967, 1971) suggests that we instead compute the exponential rate of convergence,

$$b(T,\theta,\varepsilon) = \lim_{n\to\infty} -\frac{1}{n} \log a_n(T_n,\theta,\varepsilon), \qquad (1.18)$$

which is often easier to compute and allows for easier comparisons to be made. Sievers (1978) introduced the inaccuracy function

$$A_n(T_n,\theta,\varepsilon) = \max[P_\theta(T_n \le \theta-\varepsilon),\ P_\theta(T_n \ge \theta+\varepsilon)], \qquad (1.19)$$

which he adapted from a kind of minimax procedure of Huber (1967, 1972). Correspondingly, he defines the inaccuracy rate as

$$B(T,\theta,\varepsilon) = \lim_{n\to\infty} -\frac{1}{n} \log A_n(T_n,\theta,\varepsilon). \qquad (1.20)$$

But, as both rates depend solely on the rate of convergence of the larger of the two tails, we have $b(T,\theta,\varepsilon) = B(T,\theta,\varepsilon)$. We can thus determine the exponential rate of convergence of the tail probability to zero by either (1.18) or (1.20), and hence we will refer to either as the (exponential) rate. One final note, that although

$b(T,\theta,\varepsilon) = B(T,\theta,\varepsilon)$ it is usually not true that the exact rates are equal ($a_n(T_n,\theta,\varepsilon) \neq A_n(T_n,\theta,\varepsilon)$), and also if two estimators T^1 and T^2 have the same exponential rate, this does not necessarily mean that they will also have the same exact rates ($a_n(T^1,\theta,\varepsilon) \neq a_n(T^2,\theta,\varepsilon)$).

Evaluating the rate $b(T,\theta,\varepsilon)$ is generally not an easy matter. If the estimator $T_n = T_n(X_1, X_2, \ldots, X_n)$ is defined by a sum, ΣY_i, where $Y = g(X)$ is any transformation of X then Chernoff's Theorem (Theorem 1.2) can be used to obtain the exponential rate. Bahadur (1960, 1971) adapted the procedure of Chernoff to the problem of estimation, and using exponential centering [see Feller (1969)] proved the following general result:

Theorem 1.3 [Bahadur-Chernoff] *If* $\{Y_i\}$ *are independent replicates of a random variable* Y*, with finite moment-generating function,* $Ee^{tY}<\infty$*, in some neighbourhood of* $t=0$*, and* $P(Y>0)>0$*, then*

$$\lim_{n\to\infty} -\frac{1}{n} \log P_\theta[\Sigma Y_i \geq 0] = -\log \rho \qquad (1.21)$$

where $\rho = \inf_{t\geq 0} E_\theta e^{tY}$.

This result can thus be applied to compute the rate whenever T_n is defined by means of a convolution, such as:

1. Sample Mean ($g(X) = X/n$),

2. Maximum Likelihood Estimator (mle)

   $[g(X) = \partial/\partial\theta \log f(X|\theta)]$,

3. Probability Ratio Estimators (pre)

   $[g(X) = \log [f_1(X)/f_0(X)]]$,

4. Huber M-Estimators (maximum likelihood-type estimators).

Other methods to compute rates, under special sets of conditions, have been given, by Sievers (1969) and Book (1975) using the moment-generating function of the estimator; by Killeen, Hettmansperger and Sievers (1972) using densities; by Fu (1971, 1975) using monotonic log-likelihood ratios; by Bahadur (1971) using bounds whose ratio tend to one. For the discrete (multinomial) case a method to obtain the exponential rate was given by Sanov (1957). Also, Efron & Truax (1968) give methods for computing large, moderate and small deviations for the exponential family of distributions.

If we are unable to directly compute the exponential rate, then it would at least be desirable to be able to bound it. Bahadur (1960, 1967, 1971), in various special cases, obtains a bound $B(\theta,\varepsilon)$ which is generally called the Bahadur bound, and is defined as:

$$B(\theta,\varepsilon) = \inf_{\theta^*}\{K(\theta^*,\theta): |\theta - \theta^*| > \varepsilon\}, \qquad (1.22)$$

where $K(\theta^*, \theta)$ is the Kullback-Leibler information number, in this case given by

$$K(\theta^*, \theta) = E_{\theta^*} \log \frac{f(x|\theta^*)}{f(x|\theta)} . \tag{1.23}$$

Fu (1971) showed directly that for any consistent estimator T_n of θ, and without any conditions on $f(x|\theta)$, that

$$\mathit{lim\ sup}_{n\to\infty} -\frac{1}{n} \log a_n(T_n, \theta, \varepsilon) \leq B(\theta, \varepsilon). \tag{1.24}$$

For testing composite hypotheses, H_0: $\theta \epsilon \Theta_0$, Bahadur (1965, 1971) defines the level, $L_n(S)$, of the test statistic, $T_n(S)$, as the probability of obtaining as large or a larger value of T_n as that observed. Under the alternative hypothesis, H_a: $\theta \not\in \Theta_0$, we would have $L_n(S) \to 0$ almost surely. Hence the rate at which the level tends to zero,

$$\mathit{lim}_{n\to\infty} -\frac{1}{n} \log L_n(S) = \tfrac{1}{2}\, c(\theta), \tag{1.25}$$

is a measure of the asymptotic efficiency of T_n for θ. Here, $c(\theta)$ is called the exact slope. Bahadur (1967), under certain restrictions, and Raghavachari (1970) and Bahadur (1971), without any restrictions, obtain a result similar to (1.24) for the level:

$$\mathit{lim\ sup}_{n\to\infty} -\frac{1}{n} \log L_n(S) \leq \mathit{inf}_{\theta_0} \{K(\theta, \theta_0) | \theta_0 \epsilon \Theta_0\}. \tag{1.26}$$

Bahadur (1967) also shows that this upper bound on the level is attained by the likelihood ratio statistic. From (1.25), the Bahadur efficiency is defined as

$$e_{12}(\theta) = c_1(\theta)/c_2(\theta), \tag{1.27}$$

which is equal to limit of Pitman efficiency,

$$lim_{\varepsilon\to 0} \frac{N_2(\varepsilon)}{N_1(\varepsilon)} = e_{12}(\theta). \tag{1.28}$$

Here, $N_i(\varepsilon)$ is the sample size required to make $T_n^{(i)}$ significant at level ε [Bahadur (1971)]. Fu (1975), Wieand (1976) and Grooneboom & Oosterhoff (1977) have also studied the relationship between these efficiencies.

Bahadur (1960, 1967) and Fu (1971, 1973) have shown that

$$limsup_{\varepsilon\to 0}\ limsup_{n\to\infty} -\frac{1}{n\varepsilon^2} \log \alpha_n(T_n,\theta,\varepsilon) \le \tfrac{1}{2}I(\theta). \tag{1.29}$$

with the equality attained for the mle (and others). Here, $I(\theta) = \mu_{200}$ is the Fisher information number. Bahadur's (1971) approach is fundamentally different from the others in that he attempts to extend on Fisher's Criterion of Efficiency by defining the effective standard deviation, $\tau_\theta^2(\varepsilon)$, as

$$P_\theta(|T_n - \theta| \ge \varepsilon) = P(|N(0,1)| \ge \varepsilon/\tau_\theta(\varepsilon)). \tag{1.30}$$

With this approach, it is not possible to obtain any measures of efficiency beyond the second-order.

Due to the inadequacy of Fisher's definition, Rao (1961) defines the criterion of first-order efficiency as being satisfied by all estimators T_n, for which the Fisher information per unit observation contained in the statistic, $I(T_n;\theta)$, tends to the Fisher information in the sample, $I(\theta)$, for large n. This is equivalent to the requirement that the equality in (1.29) be achieved, that is, that

$$\ell im_{\varepsilon\to 0}\ \ell im_{n\to\infty}\ -\frac{1}{n\varepsilon^2}\log \alpha_n(T_n,\theta,\varepsilon) = \tfrac{1}{2}\, I(\theta)\,, \qquad (1.31)$$

which is in turn equivalent to the rate equation,

$$\ell im_{\varepsilon\to 0}\ [B(\theta,\varepsilon) - b(T,\theta,\varepsilon)]/\varepsilon^2 = 0\,. \qquad (1.32)$$

This criterion of first-order efficiency is met by the mle as well as a large class of other "reasonable" estimators of θ. Rao (1961, 1962, 1963) also defines the criterion of second-order efficiency as being attained by all the first-order efficient estimators for which the total asymptotic loss in information of the sample, E_2, is a minimum, where

$$E_2(T) = \ell im_{n\to\infty}\ |nI(\theta) - nI(T_n;\theta)|\,. \qquad (1.33)$$

Again, subject to certain regularity conditions, the mle is second-order efficient [Rao (1961)]. This minimum of $E_2(T)$ is equal to $\gamma_\theta^2 \cdot I(\theta)$ where γ_θ^2 is a parameter which was first infered by Fisher and first evaluated by Rao (1961, 1962, 1963) as

$$\gamma_\theta^2 = \frac{\mu_{400} - 2\mu_{210} + \mu_{020}}{I^2(\theta)} - 1 - \frac{(\mu_{300} - \mu_{110})^2}{I^3(\theta)}, \qquad (1.33)$$

where the μ_{ijk} are the Fisher moments, defined as

$$\mu_{ijk..} = E_\theta[d_1^i d_2^j d_3^k \ldots], \qquad (1.34)$$

and

$$d_i = \{\partial^i/\partial\theta^i \ f(x|\theta)\}/\ f(x|\theta). \qquad (1.35)$$

Efron (1975, 1978, 1982a, 1982b), Amari (1982a, 1982b), Madsen (1979) and Kass (1982) investigate the geometry of the statistical estimation and inference problem in order to make statistical (and geometrical) interpretations of the nature of parameter γ_θ, which Efron has named the "statistical curvature" by showing that the definition co-incides with the regular mathematical curvature in the case of "curved exponential" families. Efron (1975, 1982b) goes on to infer that γ_θ^2 is a measure of the departure from exponentiality since $\gamma_\theta^2 = 0$ if and only if the underlying distribution is from the exponential family [see also Madsen (1979)].

Fu (1982) has shown that, subject to certain regularity conditions, where θ is the location parameter of a symmetric log-concave density $f(x|\theta)$, and T_n is a first-order efficient translation invariant estimator,

$$\lim_{\varepsilon\to 0} [B(\theta,\varepsilon) - b(T,\theta,\varepsilon)]/\varepsilon^4 \geq I^2(\theta)\gamma_\theta^2/8, \quad (1.36)$$

with equality for the mle. We believe that symmetry is not required, and that possibly many of the other restrictions may also be removed so that it may be possible to define Rao's second-order efficiency by an equation similar to (1.36). To this end, and in a manner similar to Hodges & Lehmann (1970) and Kallenberg (1982) let us define a measure of the difference between the Bahadur bound and the exponential rate of the statistic as the Bahadur Divergence. An estimator will be said to be *BAHADUR DIVERGENT IN THE* k^{TH}*-DEGREE* if there exists a finite constant $D_k>0$ such that:

$$\lim_{\varepsilon\to 0} D_r(\varepsilon) = \begin{cases} 0 & r < k \\ D_k & r = k \\ \infty & r > k \end{cases} \quad (1.37)$$

where

$$D_r(\varepsilon) = [B(\theta,\varepsilon) - b(T,\theta,\varepsilon)]/\varepsilon^r. \quad (1.38)$$

We will call $D_r(\varepsilon)$ the (Bahadur) Divergence Function and D_k the Bahadur Divergence.

Ghosh & Subramanyam (1974) apply the concept of second-order efficiency to multi-parameter family of exponentials, and Ghosh, Sinha & Wieand (1980) investigate the second-order efficiency property of the mle with respect to any arbitrary bounded bowl-shaped loss (increases from zero as $|\theta-a|$ increases).

Pfanzagl & Wefelmeyer (1978) attempt to show that the mle is "third-order" optimal in the multivariate case, although, as the authors themselves note, their definition of third-order efficiency would only correspond to Rao's second-order, explaining that:

> *For properties related to the second term of the asymptotic expansion of the covariance matrix of an estimator-sequence, C.R.Rao coined the term "second-order efficiency". As the asymptotic expansion of the covariance matrix proceeds in powers of* n^{-1} *whereas the asymptotic expansion of the distribution of a standardized estimator-sequence proceeds in powers of* $n^{-\frac{1}{2}}$. *Hence in our set-up it seems more natural to use the term "third-order efficiency" for what corresponds to C.R.Rao's "second-order efficiency".*

We are more inclined to Rao's terminology. Firstly, what would be the second-order efficiency by Pfanzagl and would it be possible to construct an estimator-sequence which is second-order (Pfanzagl) efficient but not third-order (Pfanzagl) efficient? Also, in terms of the asymptotic expansion in ε of the rates, we believe that the first-order criterion corresponds to the second-degree term in the expansion, while Rao's second-order (Pfanzagl's third-order) corresponds to the fourth-degree term. Hence Pfanzagl's second-order would correspond to the third-degree term of the expansion. However all the terms of odd degree are very largely influenced by the skewness of the underlying distribution and will generally vanish under symmetry. Hence only the even terms are used in determining the efficiency. Indeed we will show that, although in the general case the mle is second-order (Rao) efficient, it is not third-order efficient as it will diverge from the optimal in the sixth-degree term of the rate expansion under symmetry, and from the fifth-degree term under asymmetry [and hence is third- or fourth-order efficient by Pfanzagl, depending on whether or not the underlying distribution is symmetric].

Kester (1981) shows that, under fairly weak regularity conditions, the mle is optimal for exponential families with convex parameter spaces.

Ever since R. A. Fisher introduced the Method of Maximum Likelihood in 1912 [see Fisher (1922)], and had shown it to be consistent, asymptotically efficient, and asymptotically Normal, there have been many attempts to show the method to be "optimal" and in many instances the criterion of optimality implied that the method be maximum likelihood. In time, some statisticians began to doubt the universal superiority accorded to the mle, in particular, Kallianpur & Rao (1955), Bahadur (1958), and Berkson (1980). As we will show, for the simple problem of estimating the location parameter of a one-dimensional translation invariant distribution, the mle may be second-order efficient, but is not third-order efficient (in general) and certainly not optimal. We will show that an optimal estimator among the class of translation invariant estimators may exist in the form of probability ratio estimator, and that the mle may be no better than second-order efficient within this class. The optimality of probability ratio (likelihood ratio) estimators have been previously investigated, under other circumstances by Bahadur (1965) and Krafft & Plachky (1970).

Other important contributions to large deviation theory, in areas not directly relating to this study, include Hoeffding (1965), Groeneboom, Oosterhoff & Ruymgaart (1979), Bahadur & Zabell (1979), Fu (1980), Bahadur, Gupta & Zabell (1980).

1.3 Formulation of the General Problem

Consider the situation in which we wish to estimate the location parameter, $\theta \epsilon R^1$, of an absolutely continuous translation invariant distribution, $F(x|\theta)$, with corresponding density function $f(x|\theta)$. Let $s = \{x_1, x_2, \ldots, x_n\}$ be a sample of n independent observations from $F(x|\theta)$, and C the class of all (consistent) estimators of θ. Then for $\varepsilon>0$, we define the tail probability of the estimator as

$$a_n(T_n,\theta,\varepsilon) = P_\theta(|T_n - \theta| \geq \varepsilon), \tag{1.39}$$

which tends to zero as $n\to\infty$ for all consistent estimator-sequences $\{T_n(s)\} \epsilon C$. To measure the asymptotic behaviour of $T = \{T_n(s)\}$, we compute the exponential rate,

$$b(T,\theta,\varepsilon) = \ell im_{n\to\infty} -\frac{1}{n} \log a_n(T_n,\theta,\varepsilon), \tag{1.40}$$

which is identically equal to the inaccuracy rate,

$$\ell im_{n\to\infty} -\frac{1}{n} \log max[P_\theta(T_n \geq \theta+\varepsilon),\ P_\theta(T_n \leq \theta-\varepsilon)]. \tag{1.41}$$

Then, for all consistent estimator-sequences, $\{T_n\} \epsilon C$,

$$b(T,\theta,\varepsilon) \le B(\theta,\varepsilon) = inf_{\theta^*}\{K(\theta^*,\ \theta) : |\theta-\theta^*|>\varepsilon\}.\quad(1.42)$$

In certain cases, we can construct an estimator-sequence, $T^* = \{T_n^*(s)\} \in C$, which will attain the bound, $b(T^*,\theta,\varepsilon) = B(\theta,\varepsilon)$. In that case, T^* is (asymptotically) exponentially optimal for θ. Note that it may be possible to construct a non-consistent (i.e.: asymptotically biased) estimator-sequence $\tilde{T} = \{\tilde{T}_n(s)\} \notin C$, for which

$$\lim_{n\to\infty} -\frac{1}{n}\log P_\theta[|\tilde{T}_n - \theta| \ge \varepsilon] > B(\theta,\varepsilon), \quad (1.43)$$

exceeds the Bahadur bound. However if such a sequence is then corrected for asymptotic bias,

$$\tilde{T}_n^*(s) = \tilde{T}_n(s) - \lim_{n\to\infty}[E_\theta\tilde{T}_n(s)-\theta], \quad (1.44)$$

then the resultant consistent estimator-sequence, $\tilde{T}^* = \{\tilde{T}_n^*(s)\} \in C$, will satisfy (1.42), $b(\tilde{T}^*,\theta,\varepsilon) \le B(\theta,\varepsilon)$. An example of this will be given in §3.1.

In most instances, however, there does not appear to be any consistent estimator-sequence whose exponential rate will attain the Bahadur bound. In such a case there may not be an optimal estimator, or the Bahadur bound may be too large. We may then consider a smaller class of estimators for which an optimal estimator-sequence does exist. In the case of estimating the location parameter of a translation invariant distribution, there is, in general no optimal estimator-sequence [only for the exponential family has optimality been demonstrated by the mle, in Kester (1981)]. Let us then restrict ourselves to considering the class of estimators which are translation invariant and consistent for θ, $T \subset C$. Sievers (1978) obtains an exponentially optimal sequence of Huber M-Estimates and hence obtains the optimal rate for T for various underlying distributions. However, his result is only valid for symmetric distributions (of which all of his numerical examples consisted), since under asymmetry his estimator-sequence is not consistent for θ. In Chapter 2 we obtain the exponentially optimal rate for the class T of translation invariant estimators and obtain the local expansion of this rate about $\varepsilon=0$ up to the sixth-degree, which we compare with the sixth-degree expansion of the rate of the mle and the Bahadur bound. We thereby show that in particular cases,

$$b(\hat{\theta},\theta,\varepsilon) < b_T(\theta,\varepsilon) < B(\theta,\varepsilon), \qquad (1.45)$$

and that the differences between these expansions, up to the sixth-degree, depend on four parameters which we will call the rate coefficients: γ_θ^2, δ_θ, λ_θ^3, ν_θ^3.

CHAPTER TWO

EXPONENTIAL OPTIMALITY FOR THE CLASS OF TRANSLATION INVARIANT CONSISTENT ESTIMATORS

2.1 Probability Ratio Estimators

We will require that the underlying distribution be sufficiently smooth and well-behaved, hence we define the following regularity conditions.

Definition 2.1: A distribution, $F(x|\theta)$, with density function $f(x|\theta)$, will be said to be a *REGULAR DISTRIBUTION* if the following conditions are satisfied:

[R1] The density function, $f(x|\theta)$, is absolutely continuous for all $x\epsilon R$, $\theta\epsilon R$.

[R2] The distribution is translation invariant for θ: $f(x+a|\theta+a) = f(x|\theta)$, for all $x\epsilon R$, $\theta\epsilon R$, and each constant a.

[R3] For each fixed $\varepsilon>0$, the probability ratio, $c(x|\theta,\varepsilon) = f(x|\theta + \varepsilon) / f(x|\theta - \varepsilon)$, is strictly increasing in x (or decreasing in θ, by the translation invariance).

[R4] The moment-generating function of the log-probability ratio, $E_{\theta^*}\exp\{t\cdot\log c(x|\theta,\varepsilon)\}$ exists and is finite for some t in the neighbourhood of 0, and $|\theta - \theta^*| \leq \varepsilon$.

[R5] The density function is strictly log-concave.

Note that [R3] follows directly from [R5]. The stronger condition [R5] is required for Lemma 2.4 & Theorem 2.2 which give the rate of the mle, but is not required elsewhere. A distribution which satisfies [R1] through [R4] but not [R5] will thus be called semi-regular. Examples of regular distributions include the Normal and the Logistic.

Then, for an underlying regular distribution, we would like to obtain an estimator-sequence, $T = \{T_n(s)\} \in \mathcal{T}$, which is consistent for θ, translation invariant, and is exponentially optimal in the sense that if $T^* = \{T_n^*(s)\} \in \mathcal{T}$ is any other translation invariant estimator-sequence that is consistent for θ,

$$b(T^*,\theta,\varepsilon) \leq b(T,\theta,\varepsilon). \qquad (2.1)$$

Therefore it follows that the exponentially optimal rate for the class $\mathcal{T}$ of translation invariant estimators consistent for θ is given by the rate of T, namely

$$b_{\mathcal{T}}(\theta,\varepsilon) = b(T,\theta,\varepsilon). \qquad (2.2)$$

Sievers (1978) gives a theorem in which he identifies such an estimator-sequence by taking a sequence of Huber (1967) M-estimates. Sievers' (1978) estimator-sequence, however, is only consistent under symmetric distributions (a property shared by all of his illustrative examples). We will apply the Huber (1967) procedure to obtain a consistent estimator-sequence which we shall call the Probability Ratio Estimator (pre).

To construct the probability ratio estimator, we consider the following artificial "Test of Hypothesis":

$$\begin{aligned} H_0 &: F_0(x|\theta) = F(x|\theta - \varepsilon) \\ H_1 &: F_1(x|\theta) = F(x|\theta + \varepsilon), \end{aligned} \tag{2.3}$$

with probability ratio,

$$c(x|\theta,\varepsilon) = \frac{f(x|\theta + \varepsilon)}{f(x|\theta - \varepsilon)} = \frac{f_0(x|\theta)}{f_1(x|\theta)}. \tag{2.4}$$

Then, for a sample $s = \{x_1, x_2, \ldots, x_n\}$ of n independent observations from the population $F(x|\theta)$, we define the likelihood ratio for this test as $c_n(s|\theta,\varepsilon) = \Pi_i\, c(x_i|\theta,\varepsilon)$, and the test function:

$$\phi_n(k_n) = \phi_n(k_n;s|\theta,\varepsilon) = \begin{cases} 1 & c_n(s|\theta,\varepsilon) \geq k_n \\ 0 & c_n(s|\theta,\varepsilon) < k_n \end{cases} \tag{2.5}$$

for each fixed k_n (the critical value).

Now, we are ready to define our probability ratio estimator, in reference to the test of hypothesis that we established. (We use the term probability ratio estimator rather than likelihood ratio estimator since the 'test of hypothesis' was artificially constructed in order to define our estimator and is not an actual test that we would ever wish to perform; in fact, we are apparantly 'testing' $F(x|\theta - \varepsilon)$ against $F(x|\theta + \varepsilon)$ when we are actually sampling from $F(x|\theta)$.) Then, as θ lies midway between H_0 and H_1, a 'logical' estimate for θ, corresponding to the test $\phi_n(k_n)$, would be to take the value of θ at the boundary between H_0 and H_1, that is, the critical value of the test, namely,

$$T_n(k_n;s|\varepsilon) = \{\theta | c_n(s|\theta,\varepsilon) = k_n\}. \tag{2.6}$$

Since $c_n(s|\theta,\varepsilon)$ strictly decreases in θ by [R3], then $T_n(k_n;s|\varepsilon)$ is well-defined for any fixed k_n. How should we then select the k_n? Since we require an estimator-sequence to be consistent for θ, we will require that the $\{k_n\}$ are such that $T_n(k_n;s|\varepsilon) \xrightarrow{p} \theta$.

It is well known [Chernoff (1954)] that $-2 \log c_n$ tends to a chi-square distribution with n degrees of freedom, hence $(1/n)\log c_n(s|\theta,\varepsilon)$ degenerates to its mean:

$$\frac{1}{n}\log c_n(x|\theta,\varepsilon) = \frac{1}{n}\sum_i \log c(x_i|\theta,\varepsilon) \to E_\theta \log c(X|\theta,\varepsilon). \qquad (2.7)$$

Hence the estimator $T_n(k_n;s|\varepsilon)$ in (2.6) will only be consistent if the k_n are such that

$$\frac{1}{n}\log k_n \to \lim_{n\to\infty}\frac{1}{n}\log c_n(s|\theta,\varepsilon) = E_\theta \log c(X|\theta,\varepsilon). \quad (2.8)$$

Therefore, define the probability ratio estimator, $\tilde{\theta} = \{\tilde{\theta}_n\}$ as (see Figure 2.1):

$$\tilde{\theta}_n = \tilde{\theta}_n(s|\varepsilon) = \{\theta|\frac{1}{n}\log c_n(s|\theta,\varepsilon)=\mu(\varepsilon)\}, \qquad (2.9)$$

where

$$\mu(\varepsilon) = E_\theta \log c(X|\theta,\varepsilon) = K(\theta,\theta-\varepsilon)-K(\theta,\theta+\varepsilon). \quad (2.10)$$

This estimator-sequence will be consistent and translation invariant [Fu (1983)]. For symmetric regular distributions, Sievers (1978) and Fu (1983) have shown $\tilde{\theta}$ to be optimal in exponential rate for the class of translation invariant consistent estimators of θ, T.

Sievers (1978) also gives an expression for the exponential rate of convergence of $\mathcal{D}_\varepsilon(\theta)$ to zero for his estimator-sequence. However if the underlying distribution is not symmetric, then this exponential

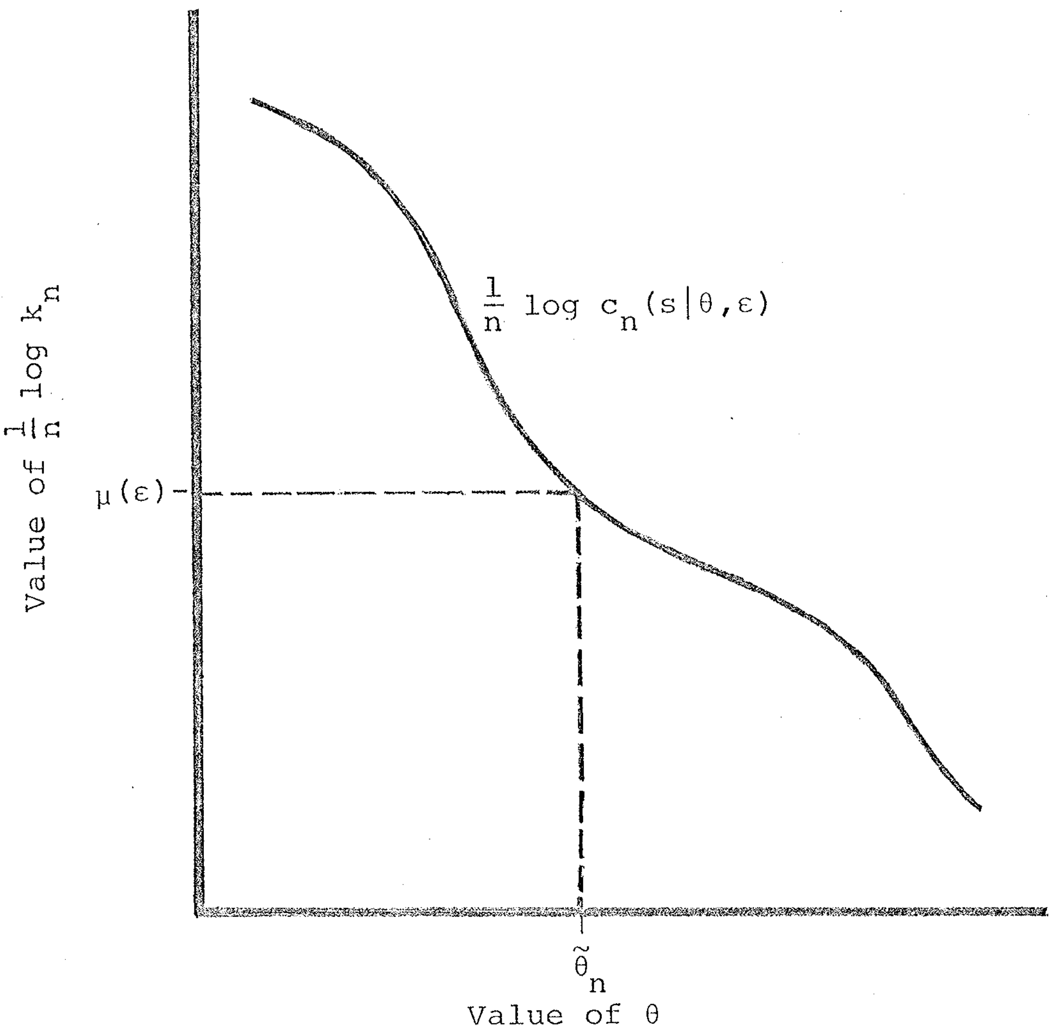

FIGURE 2.1: Probability Ratio Estimator.

rate may exceed the Bahadur bound as the estimator-sequence is not consistent [see example in §3.1]. The exponential rate of $\tilde{\theta}$ in the general case is given by the following theorem:

Theorem 2.1: *If* $F(x|\theta)$ *is a (semi-) regular distribution, the probability ratio estimator* $\tilde{\theta}$, *defined in* (2.9) *has the exponential rate of convergence of the tail probability to zero given by:*

$$b(\tilde{\theta},\theta,\varepsilon) = -\log \max\{\rho_0, \rho_1\} \tag{2.11}$$

where

$$\begin{aligned} \rho_0 &= \inf_{0<t<1} e^{-t\mu(\varepsilon)} m_0(t), \\ \rho_1 &= \inf_{0<t<1} e^{t\mu(\varepsilon)} m_1(-t), \end{aligned} \tag{2.12}$$

and where $\mu(\varepsilon)$ *is defined by* (2.10) *and* $m_i(t)$ *is the moment-generating function of the log-probability ratio with respect to* $F_i(x|\theta)$:

$$m_i(t) = E_{F_i} e^{t\cdot\log c(x|\theta,\varepsilon)}, \quad i=0,1. \tag{2.13}$$

VProof:

Let us define,

$$Y = Y(x|\theta,\varepsilon) = \log c(x|\theta,\varepsilon), \tag{2.14}$$

then, the sample mean is

$$\bar{Y}_n = \bar{Y}_n(s|\theta,\varepsilon) = \frac{1}{n}\log c_n(s|\theta,\varepsilon). \tag{2.15}$$

By definition, the exponential rate is

$$b(\tilde{\theta},\theta,\varepsilon) = \lim_{n\to\infty} -\frac{1}{n}\log \max\{\alpha_n, \beta_n\}, \tag{2.16}$$

where

$$\begin{aligned} \alpha_n &= P_\theta(\tilde{\theta}_n \geq \theta + \varepsilon), \\ \beta_n &= P_\theta(\tilde{\theta}_n \leq \theta - \varepsilon). \end{aligned} \tag{2.17}$$

Then, by the strict monotonicity of $\bar{Y}_n$ (by [R3]),

$$\begin{aligned} \alpha_n &= P_\theta(\tilde{\theta}_n \geq \theta + \varepsilon) \\ &= P_{F_0}(\tilde{\theta}_n \geq \theta) \\ &= P_{F_0}(\bar{Y}_n \geq \mu(\varepsilon)), \end{aligned} \tag{2.18}$$

and so, by conditions [R3] and [R4], we can apply the Bahadur-Chernoff Theorem (Theorem 1.3) to obtain:

$$\begin{aligned} -\frac{1}{n}\log \alpha_n &\to -\log \inf_{t\geq 0} E_{F_0} e^{t[Y - \mu(\varepsilon)]} \\ &= -\log \inf_{t\geq 0} e^{-t\mu(\varepsilon)}\, m_0(t), \end{aligned} \tag{2.19}$$

where $m_0(t)$ is the Chernoff Information number of (1.12),

$$m_0(t) = E_{F_0} e^{tY} = E_{F_0} \exp[t \cdot \log \frac{f(X|\theta + \varepsilon)}{f(X|\theta - \varepsilon)}]$$
$$= \int_{-\infty}^{\infty} f^{1-t}(x|\theta - \varepsilon) f^{t}(x|\theta + \varepsilon)dx. \tag{2.20}$$

Then, as $\partial^2/\partial t^2\{m_0(t)\} > 0$, $m_0(t)$ is convex and continuous and as $m_0(0) = m_0(1) = 1$, the infimum of $m_0(t)$ will occur for $0<t<1$ and hence the infimum in (2.19) will also occur on $0<t<1$. Thus we obtain

$$-\frac{1}{n} \log \alpha_n \to - \log \inf_{0<t<1} e^{-t\mu(\varepsilon)} m_0(t). \tag{2.21}$$

Similarly for β_n, we obtain

$$-\frac{1}{n} \log \beta_n \to - \log \inf_{t\geq 0} E_{F_1} e^{t[\mu(\varepsilon) - Y]}$$
$$= - \log \inf_{0<t<1} e^{t\mu(\varepsilon)} m_1(-t), \tag{2.22}$$

where

$$m_1(t) = E_{F_1} e^{tY} = E_{F_1} \exp[t \cdot \log \frac{f(X|\theta + \varepsilon)}{f(X|\theta - \varepsilon)}]$$
$$= \int_{-\infty}^{\infty} f^{1+t}(x|\theta + \varepsilon) f^{-t}(x|\theta - \varepsilon)dx. \tag{2.23}$$

Then, combining these results we get

$$b(\tilde{\theta},\theta,\varepsilon) = - \log \max\{\rho_0, \rho_1\}, \tag{2.24}$$

where ρ_0 and ρ_1 are as defined in (2.12). Δ

Remark 1: As $m_1(t) = m_0(t+1)$, we get $\rho_1 = \rho_0 e^{\mu(\varepsilon)}$, and so we need only compute one of ρ_0 and ρ_1.

Remark 2: Under symmetry, $\tilde{\theta}$ is the Sievers' (1978) estimator-sequence and $\rho_0 = \rho_1 = m_0(\frac{1}{2})$ is the Chernoff Information number, the exponential rate expression obtained by Sievers (1978).

Remark 3: Although $\tilde{\theta}$ has optimal exponential rate for the class of translation invariant estimators under symmetric distributions [Sievers(1978), Fu(1983)], we believe that it may be optimal for T in the general case, although we have not yet been able to verify this.

Remark 4: The exponential rates of convergence of the tail probability to zero for the mle and the pre can both be determined from a theorem by Rubin & Rukhin (1982) in which they find the exponential rate for a general M-estimator: $\sum_i w(x_i, \theta) = 0$, where $w(x, \theta)$ is a strictly monotonic function. This monotonicity condition is the condition [R3] for the pre and [R5] for the mle.

2.2 Local Expansion of Exponential Rates about $\varepsilon=0$

One of the questions we would like to answer is: How nearly optimal is the mle? Also, how does the pre compare with the mle? Then, since we are mostly concerned with the behaviour of an estimator when ε is small, let us examine the Taylor series expansions of the Bahadur bound and of the exponential rates of the mle and pre, in the neighbourhood of $\varepsilon=0$. Fu (1982) obtained expansions of the Bahadur bound and the rate of the mle up to degree four, which is sufficient to make second-order efficiency comparisons. The mle has been shown to hold a favoured position among (first-order) efficient estimators in that it minimizes the loss of sample information on θ, and is thereby "second-order" efficient [Efron (1975)]. However, as we shall see later, the pre is also second-order efficient. This is not surprising as it can be shown [Fu (1983)] that the mle is a limit of probability ratio estimators*, $\tilde{\theta}_n(s|\varepsilon)$ as $\varepsilon\to 0$. In fact, there are many estimators which are second-order efficient. The question is then, does the mle also hold a favoured position among these second-order efficient estimators? (i.e.: Is it "third-order" efficient?) To investigate this we require that

---

* *Actually, the entire class of Probability Ratio-type Estimators, as defined by* (2.6), *collapses to the mle.*

the Taylor series expansions be up to degree six, since as mentioned, the odd degree terms vanish under symmetry. These higher degree terms are important to differentiate between the mle and pre and any other second-order efficient estimators, and they may possibly relate to the concept of third-order efficiency in this case. We will show that, in general, using the criteria of exponential rates, the mle is not optimal. We will show by the expansions that we obtain, and by illustrative examples, that for certain distributions, and each $\varepsilon>0$, the mle can be dominated in exponential rate of convergence of the probability of the tail set $\mathcal{D}_{\varepsilon}(\theta)$ by some pre, $\tilde{\theta}_n(s|\varepsilon)$. This means that the mle may not be "third-order" efficient in the sense of exponential rates, although it may be regarded as locally "third-order" efficient in the neighbourhood of $\varepsilon=0$.

As the probability ratio estimator is defined in terms of H_0 and H_1, as is its' exponential rate, we have re-written the Bahadur bound and the exponential rate of the mle to correspond, which will make for better and more meaningful comparisons. We have thus shifted the expectations with respect to $F(x|\theta)$ as in Fu (1982) to expectations with respect to $F_0(x|\theta)$ and $F_1(x|\theta)$. This shifting is mathematically valid only for translation invariant location parameter. This means that these Taylor series expansions will be valid only for the translation invariant location parameter (in fact, $\tilde{\theta}$ is defined only in this instance). This also means that any conclusions we reach concerning third-order efficiencies, will relate only to a translation invariant location parameter. This should not be too distressing in that it seems likely that the higher order efficiencies will depend on the nature of the parameter of estimation. The sixth-degree expansions for the mle are extremely difficult to obtain. Without the shifting of the expectations the procedure is much more protracted and becomes practically intractable.

Due to the shifting of the expectations, the series expressions that we obtain in this case may result in terms whose coefficients, when expressed in terms of Fisher moments may differ from those of Fu (1982), however, they will be mathematically equivalent (under translation invariance). In fact, under translation invariance, there are certain relationships that exist between Fisher moments of the same degree as given by the following lemma.

Lemma 2.1: *If* $F(x|\theta)$ *is a regular distribution with continuous derivative-ratios,*

$$d_i = \{\partial^i/\partial\theta^i\ f(x|\theta)\}/\ f(x|\theta),\ i=1,2,3,.. \quad (2.25)$$

and finite Fisher moments,

$$\mu_{ijk..} = E_\theta[d_1^i d_2^j d_3^k \ldots] \quad (2.26)$$

of degree $i+2j+3k+....$, *then by the translation invariance property, certain relationships exist between moments of the same degree:*

Degree 3

(A) $\mu_{300} - 2\mu_{110} = 0$,

Degree 4

(B) $2\mu_{400} - 3\mu_{210} = 0$,

(C) $\mu_{210} - \mu_{020} - \mu_{101} = 0$,

Degree 5

(D) $3\mu_{500} - 4\mu_{310} = 0$,

(E) $2\mu_{310} - 2\mu_{120} - \mu_{201} = 0$,

(F) $\mu_{201} - \mu_{011} - \mu_{1001} = 0$, (2.27)

(G) $\mu_{120} - 2\mu_{011} = 0$,

Degree 6

(H) $4\mu_{600} - 6\mu_{410} = 0$,

(I) $3\mu_{410} - 3\mu_{220} - \mu_{301} = 0$,

(J) $2\mu_{220} - 2\mu_{111} - \mu_{030} = 0$,

$$\left|\begin{array}{ll}(K) & 2\mu_{301} - 2\mu_{111} - \mu_{2001} = 0,\\ (L) & \mu_{111} - \mu_{002} - \mu_{0101} = 0,\\ (M) & \mu_{2001} - \mu_{0101} - \mu_{10001} = 0.\end{array}\right.$$

∇*Proof:*

It follows directly from integration by parts, for $p \geq 1$ and $p+q>2$, that

$$E_\theta[d_1^p d_\nu^q] = \int_{-\infty}^{\infty} \{f^{(1)}\}^p \{f^{(\nu)}\}^q / f^{p+q-1}\, dx$$

$$= -\left.\frac{\{f^{(1)}\}^{p-1}\{f^{(\nu)}\}^q}{(p+q-2)f^{p+q-2}}\right|_{-\infty}^{\infty}$$

$$+ \frac{1}{p+q-2}\int_{-\infty}^{\infty}\left\{\frac{(p-1)\{f^{(1)}\}^{p-2} f^{(2)}\{f^{(\nu)}\}^q}{f^{p+q-2}}\right. \tag{2.28}$$

$$\left. + \frac{q\{f^{(1)}\}^{p-1}\{f^{(\nu)}\}^{q-1} f^{(\nu+1)}}{f^{p+q-2}}\right\} dx$$

$$= 0 + \frac{(p-1)E_\theta[d_1^{p-2} d_2 d_\nu^q] + qE_\theta[d_1^{p-1} d_\nu^{q-1} d_{\nu+1}]}{p+q-2}.$$

Here, the initial term must tend to zero under translation invariance. Thus, for example, for $\nu=1$, $(p,q)=(3,0)$ (or $(2,1)$ or $(1,2)$) we obtain (2.27*A*), and so on. Δ

In order that the Taylor series expansions be valid, we will require some additional regularity conditions, in addition to [R1] through [R5]:

[C1] The density function, $f(x|\theta)$, must be at least six times differentiable, with each such derivative, $f^{(i)}(x|\theta) = \partial^i/\partial\theta^i\, f(x|\theta)$, being absolutely continuous and bounded for all x, θ.

[C2] The log-density function,

$$\ell \equiv \ell(x|\theta) = \log f(x|\theta), \tag{2.29}$$

and the score function,

$$\dot{\ell} \equiv \ell^{(1)} = \frac{\partial \ell}{\partial \theta}, \tag{2.30}$$

are both continuous in x for each θ.

[C3] There exists a constant $u(\theta)$ such that

$$\begin{aligned} P_\theta(\dot{\ell}(x|\theta + \varepsilon) < 0) > 0, \\ P_\theta(\dot{\ell}(x|\theta - \varepsilon) > 0) > 0, \end{aligned} \tag{2.31}$$

for all $0<\varepsilon<u(\theta)$.

[C4] There exists a second constant $v(\theta)$, for each $u(\theta)$, such that the mgf,

$$\psi(t,\theta,\varepsilon) = E_\theta e^{t\dot{\ell}(x|\theta + \varepsilon)}, \tag{2.32}$$

exists and is finite for all $|t| < v(\theta)$ and $|\varepsilon| < u(\theta)$.

The Fisher moments are related to the moments of the score function, as shown by the following lemma:

Lemma 2.2: *Let* $f(x|\theta)$ *be any continuous distribution satisfying conditions* [C1] *through* [C4]. *With the Fisher moments defined in Lemma 2.1, and the score function given by* (2.30), *we similarly define the score moments as*

$$\eta_{ijk..} = E_\theta[h_1^i h_2^j h_3^k \ldots], \tag{2.33}$$

where

$$h_i = \ell^{(i)} - E_\theta \ell^{(i)}, \quad i=1,2,3,\ldots. \tag{2.34}$$

Then, certain relationships exist between the $\mu_{ijk\ldots}$ *and the* $\eta_{ijk\ldots}$:

(A) $E_\theta \ell^{(1)} = \mu_{100} = 0,$

(B) $E_\theta \ell^{(2)} = -\mu_{200},$

(C) $E_\theta \ell^{(3)} = 2\mu_{300} - 3\mu_{110},$

(D) $E_\theta \ell^{(4)} = -6\mu_{400} + 12\mu_{210} - 4\mu_{101} - 3\mu_{020},$

(E) $E_\theta \ell^{(5)} = 24\mu_{500} - 60\mu_{310} + 20\mu_{201} + 30\mu_{120} - 10\mu_{011} - 5\mu_{1001},$

(F) $E_\theta \ell^{(6)} = -120\mu_{600} + 360\mu_{410} - 120\mu_{301} - 270\mu_{220} + 120\mu_{111}$
$+ 30\mu_{030} - 10\mu_{002} + 30\mu_{2001} - 15\mu_{0101} - 6\mu_{10001},$

(G) $\eta_{200} = V_\theta \ell^{(1)} = \mu_{200} \equiv I(\theta),$ (2.35)

(H) $\eta_{020} = V_\theta \ell^{(2)} = \mu_{400} - 2\mu_{210} + \mu_{020} - I^2(\theta),$

(I) $\eta_{002} = V_\theta \ell^{(3)} = 4\mu_{600} - 12\mu_{410} + 4\mu_{301} + 9\mu_{220}$
$- 6\mu_{111} + \mu_{002} - (2\mu_{300} - 3\mu_{110})^2,$

$$
\begin{cases}
(J)\ \eta_{110} = Cov_\theta[\ell^{(1)},\ell^{(2)}] = -\mu_{300} + \mu_{110} \equiv J(\theta), \\
(K)\ \eta_{101} = Cov_\theta[\ell^{(1)},\ell^{(3)}] = 2\mu_{400} - 3\mu_{210} + \mu_{101}, \\
(L)\ \eta_{1001} = Cov_\theta[\ell^{(1)},\ell^{(4)}] = -6\mu_{500} + 12\mu_{310} - 4\mu_{201} - 3\mu_{120} + \mu_{1001}, \\
(M)\ \eta_{011} = Cov_\theta[\ell^{(2)},\ell^{(3)}] = -2\mu_{500} + 5\mu_{310} - \mu_{201} - 3\mu_{120} \\
\qquad\qquad + \mu_{011} + I(\theta)(2\mu_{300} - 3\mu_{110}), \\
(N)\ \eta_{0101} = Cov_\theta[\ell^{(2)},\ell^{(4)}] = 6\mu_{600} - 18\mu_{410} + 4\mu_{301} - 4\mu_{111} - \mu_{2001} + \\
\qquad 15\mu_{220} - 3\mu_{030} + \mu_{0101} - I(\theta)(6\mu_{400} - 12\mu_{210} + 4\mu_{101} + 3\mu_{020}), \\
(O)\ \eta_{300} = E_\theta\{\ell^{(1)}\}^3 = \mu_{300}, \\
(P)\ \eta_{210} = E_\theta\{\ell^{(1)}\}^2\{\ell^{(2)} + I(\theta)\} = -\mu_{400} + \mu_{210} + I^2(\theta), \\
(Q)\ \eta_{120} = E_\theta\{\ell^{(1)}\}\{\ell^{(2)} + I(\theta)\}^2 = \mu_{500} - 2\mu_{310} + \mu_{120} + 2I(\theta)J(\theta), \\
(R)\ \eta_{030} = E_\theta\{\ell^{(2)} + I(\theta)\}^3 = -\mu_{600} + 3\mu_{410} - 3\mu_{220} + \mu_{030} + 3\eta_{020}I(\theta) + I^3(\theta).
\end{cases}
$$

∇*Proof:* By expansion and direct application of the definitions. Δ

Here, the principal variance is $\eta_{200} = \mu_{200} = I(\theta)$, the Fisher information number. An important and frequently occuring term is the principal covariance, which we will denote by $J(\theta) = \eta_{110} = \mu_{110} - \mu_{300}$. We may also be interested in the normalized form $\omega_\theta^2 = J^2(\theta)/I^3(\theta)$.

Before we state the Taylor series expansions, it would be useful to define a number of quantities that will be encountered in our work. Since these definitions only require existence of the moments, we need not confine ourselves to regular distributions, although their application in the non-regular case may differ.

Definition 2.2: For each underlying distribution $F(x|\theta)$ satisfying conditions [C1] to [C4], and finite moments $\mu_{ijk..}$ and $\eta_{ijk..}$, with h_i as defined in (2.34), and with $I(\theta) = \eta_{200} = \mu_{200}$, $J(\theta) = \eta_{110} = \mu_{110} - \mu_{300}$, and $\omega_\theta^2 = J^2(\theta)/I^3(\theta)$, we define the following constants:

$$\gamma_\theta^2 = I^{-3}(\theta)\begin{vmatrix}\eta_{200} & \eta_{110}\\ \eta_{110} & \eta_{020}\end{vmatrix} = I^{-4}(\theta)E_\theta\begin{vmatrix}I(\theta) & h_1\\ J(\theta) & h_2\end{vmatrix}^2$$

$$= (\mu_{400} - 2\mu_{210} + \mu_{020})/I^2(\theta) - 1 - \omega_\theta^2, \quad (2.36)$$

$$\delta_\theta = \frac{1}{I^2(\theta)J(\theta)}\begin{vmatrix}\eta_{200} & \eta_{101}\\ \eta_{110} & \eta_{011}\end{vmatrix} \qquad (J(\theta)\neq 0)$$

$$= \frac{1}{I^3(\theta)J(\theta)} E_\theta\begin{vmatrix}I(\theta) & h_1\\ J(\theta) & h_2\end{vmatrix}\cdot\begin{vmatrix}I(\theta) & h_1\\ J(\theta) & h_3\end{vmatrix} \quad (2.37)$$

$$= -(2\mu_{500} - 5\mu_{310} + \mu_{201} + 3\mu_{120} - \mu_{011})/I(\theta)J(\theta)$$

$$-(2\mu_{400} - 3\mu_{210} + \mu_{101})/I^2(\theta) + (2\mu_{300} - 3\mu_{110})/J(\theta),$$

$$\nu_\theta^3 = I^{-4}(\theta)\begin{vmatrix}\eta_{200} & \eta_{101}\\ \eta_{101} & \eta_{002}\end{vmatrix} + 3\omega_\theta^2\delta_\theta \qquad (2.38)$$

$$= (4\mu_{600} - 12\mu_{410} + 4\mu_{301} + 9\mu_{220} - 6\mu_{111} + \mu_{002})/I^3(\theta)$$

$$-\left\{\frac{2\mu_{400} - 3\mu_{210} + \mu_{101}}{I^2(\theta)}\right\}^2 - \omega_\theta^2\left\{\frac{2\mu_{300} - 3\mu_{110}}{J(\theta)}\right\}^2 + 3\omega_\theta^2\delta_\theta,$$

$$\lambda_\theta^3 = I^{-6}(\theta)\, E_\theta \begin{vmatrix} I(\theta) & h_1 \\ J(\theta) & h_2 \end{vmatrix}^3 +$$

$$I^{-5}(\theta) \begin{pmatrix} 2J(\theta) \\ I(\theta) \end{pmatrix}' \begin{pmatrix} \eta_{1010} & -\eta_{0110} \\ -\eta_{1001} & \eta_{0101} \end{pmatrix} \begin{pmatrix} J(\theta) \\ I(\theta) \end{pmatrix} \tag{2.39}$$

$$= (5\mu_{600}-15\mu_{410}+4\mu_{301}+12\mu_{220}-4\mu_{111}-2\mu_{030}-\mu_{2001}+\mu_{0101})/I^3(\theta)$$

$$-(3\mu_{400}-6\mu_{210}+4\mu_{101})/I^2(\theta) - 2 - \omega_\theta^2\{(\mu_{300}-3\mu_{110})/J(\theta)$$

$$-(7\mu_{500}-16\mu_{310}+6\mu_{201}+6\mu_{120}-2\mu_{011}-\mu_{1001})/I(\theta)J(\theta)$$

$$-(\mu_{400}-3\mu_{210}+2\mu_{101})/I^2(\theta)\} - \omega_\theta^4\{\mu_{300}/J(\theta)\}.$$

Now we can proceed with the Taylor series expansions. First the Bahadur bound:

Lemma 2.3: *If $F(x|\theta)$ is a regular distribution, and the additional condition* [C1] *holds, then a local expansion about $\varepsilon=0$ for the Bahadur bound for the consistent estimators of a translation invariant location parameter, θ, is given by*

$$B(\theta,\varepsilon) = \min_{\pm\varepsilon}\Big[\sum_{i=1}^{k} b_i \varepsilon^i/i! + o(\varepsilon^k)\Big], \tag{2.40}$$

where the coefficients b_i are given by

$$b_i = -E_\theta \ell^{(i)}, \quad i=1,2,3,\ldots. \tag{2.41}$$

Note: This result is valid for translation invariant location parameter only.

∇*Proof*

By the regularity conditions, the Kullback-Leibler information, $K(\theta^*,\theta)$, is continuous and convex. Hence,

$$B(\theta,\varepsilon) = \min\{K(\theta+\varepsilon,\theta),\ K(\theta-\varepsilon,\theta)\}. \qquad (2.42)$$

Under translation invariance,

$$K(\theta-\varepsilon,\theta) = K(\theta,\theta+\varepsilon) = E_\theta \log \frac{f(x|\theta)}{f(x|\theta+\varepsilon)}. \qquad (2.43)$$

Then we expand $\log f(x|\theta+\varepsilon)$ about $\varepsilon=0$, so that

$$\log f(x|\theta+\varepsilon) = \sum_{i=0}^{k} \ell^{(i)} \varepsilon^i/i! + o(\varepsilon^k) \qquad (2.44)$$

exists for some $k>0$. Condition [C1] is sufficient to guarantee the validity of this expansion to at least the sixth degree ($k=6$). By taking the expectations, we obtain the result

$$K(\theta-\varepsilon,\theta) = -\sum_{i=2}^{k} E_\theta \ell^{(i)} \varepsilon^i/i! + o(\varepsilon^k). \qquad (2.45)$$

Then since $K(\theta+\varepsilon,\theta) = K(\theta-(-\varepsilon),\theta)$, the lemma follows.Δ

The method for obtaining the expansions of the rates of the mle and pre is quite lengthy and tedious, and so, these results are given in sections 2.3 and 2.4 respectively.

2.3 Exponential Rate of the mle

The exponential rate for the mle can be obtained by the Bahadur-Chernoff Theorem under certain regularity conditions. The result is well known and so is stated as the following lemma without proof.

Lemma 2.4: *If* $F(x|\theta)$ *is a regular distribution and conditions* [C2] *to* [C4] *hold, and* $\hat{\theta}$ *is the maximum likelihood estimator, which is the only solution of the likelihood equation:*

$$\sum_{i=1}^{n} \partial/\partial\theta \log f(x_i|\hat{\theta}) = 0, \tag{2.46}$$

then the exponential rate of the mle is given by

$$b(\hat{\theta},\theta,\varepsilon) = -\log \max\{\rho_0^*, \rho_1^*\}, \tag{2.47}$$

where

$$\begin{aligned} \rho_0^* &= \inf_{t\geq 0} \rho_0^*(t), \\ \rho_1^* &= \inf_{t\leq 0} \rho_1^*(t), \end{aligned} \tag{2.48}$$

and

$$\rho_i^*(t) = E_{F_i} e^{t\dot{\ell}}, \quad i=0,1. \tag{2.49}$$

Using a method similar to Fu (1982), we will obtain a local Taylor series expansion of $b(\hat{\theta},\theta,\varepsilon)$ up to degree six as stated in the following theorem.

<u>Theorem 2.2:</u> *Under the conditions of Lemma 2.4, and with the additional condition* [C1], *a local expansion in the neighbourhood of* $\varepsilon=0$ *of the exponential rate of convergence to zero of the tail probability of the mle, as stated in Lemma 2.4, is given by*

$$\begin{aligned} b(\hat{\theta},\theta,\varepsilon) = \min_{\pm\varepsilon}\Big\{ & b_2\varepsilon^2/2! + b_3\varepsilon^3/3! \\ & + [b_4 - 3I^2(\theta)\gamma_\theta^2]\varepsilon^4/4! \\ & + [b_5 - 10I(\theta)J(\theta)\delta_\theta]\varepsilon^5/5! \qquad (2.50) \\ & + [b_6 - 15I^3(\theta)\lambda_\theta^3 \\ & \quad - 10I^3(\theta)\nu_\theta^3]\varepsilon^6/6! + o(\varepsilon^6)\Big\} \end{aligned}$$

Note: This expansion is again valid only for a translation invariant location parameter.

∇*Proof:*

Under the regularity conditions, it follows from Lemma 1 in Hoeffding (1965) that there exists a unique $t=\tau_\theta^*$ which minimizes $\rho_0^*(t)$.

In order to obtain a Taylor series expansion about $\varepsilon=0$ of the exponential rate of the mle as given by Lemma 2.4, we must expand

$$\rho_0^* = \inf_{t\geq 0} \rho_0^*(t). \tag{2.51}$$

Fu (1973, 1982) has shown that the unique t of Hoeffding (1965) is of the form

$$t = \tau_\theta^*(\varepsilon) = \varepsilon + o(\varepsilon). \tag{2.52}$$

Hence, let us first expand $\rho_0^*(t)$ into a double Taylor series in t and ε, as

$$\begin{aligned} \rho_0^*(t) &= E_{F_0} e^{t\cdot\partial/\partial\theta \log f(x|\theta)} \\ &= E_\theta\left\{\sum_{m=0}^{\infty} (d_m(-\varepsilon)^m/m!)e^{td_1}\right\} \qquad (2.53) \\ &= \sum_{m=0}^{\infty}\ \sum_{n=0}^{\infty} \frac{t^n}{n!}\left\{E_\theta(d_1 d_m)\frac{(\;\varepsilon)^m}{m!}\right\}. \end{aligned}$$

The d_i are the derivative ratios defined in (2.25). By the definition of the Fisher moments (2.26) we get

$$\rho_0^*(t) = 1 + t\{-I\varepsilon+\mu_{110}\varepsilon^2/2!-\mu_{101}\varepsilon^3/3!+\mu_{1001}\varepsilon^4/4!-\mu_{10001}\varepsilon^5/5!\}$$
$$+ \frac{t^2}{2!}\{I-\mu_{300}\varepsilon+\mu_{210}\varepsilon^2/2!-\mu_{201}\varepsilon^3/3!+\mu_{2001}\varepsilon^4/4!\}$$
$$+ \frac{t^3}{3!}\{\mu_{300}-\mu_{400}\varepsilon+\mu_{310}\varepsilon^2/2!-\mu_{301}\varepsilon^3/3!\} \qquad (2.54)$$
$$+ \frac{t^4}{4!}\{\mu_{400}-\mu_{500}\varepsilon+\mu_{410}\varepsilon^2/2!\} + \frac{t^5}{5!}\{\mu_{500}-\mu_{600}\varepsilon\}$$
$$+ \frac{t^6}{6!}\{\mu_{600}\} + o(\varepsilon^6).$$

where $I = I(\theta) = \mu_{200}$ is the Fisher information. As $\tau_\theta^*(\varepsilon)$ is of order ε, let us write

$$t_0^*=\tau_\theta^*(\varepsilon)=\varepsilon+A\varepsilon^2/2!+B\varepsilon^3/3!+C\varepsilon^4/4!+D\varepsilon^5/5!+o(\varepsilon^5). \qquad (2.55)$$

We insert (2.55) into (2.54) to obtain $\rho_0^* = \rho_0^*(t_0^*)$,

$$\rho_0^* =1+\{(-2I\varepsilon^2/2!+3\mu_{110}\varepsilon^3/3!-4\mu_{101}\varepsilon^4/4!+5\mu_{1001}\varepsilon^5/5!-6\mu_{10001}\varepsilon^6/6!)$$
$$+(-3AI\varepsilon^3/3!+6A\mu_{110}\varepsilon^4/4!-10A\mu_{101}\varepsilon^5/5!+15A\mu_{1001}\varepsilon^6/6!)$$
$$+(-4BI\varepsilon^4/4!+10B\mu_{110}\varepsilon^5/5!-20B\mu_{101}\varepsilon^6/6!)$$
$$+(-5CI\varepsilon^5/5!+15C\mu_{110}\varepsilon^6/6!) + (-6DI\varepsilon^6/6!)\}$$
$$+ \{(I\varepsilon^2/2!-3\mu_{300}\varepsilon^3/3!+6\mu_{210}\varepsilon^4/4!-10\mu_{201}\varepsilon^5/5!+15\mu_{2001}\varepsilon^6/6!)$$
$$+(3AI\varepsilon^3/3!-12A\mu_{300}\varepsilon^4/4!+30A\mu_{210}\varepsilon^5/5!-60A\mu_{201}\varepsilon^6/6!)$$
$$+(3A^2I\varepsilon^4/4!-15A^2\mu_{300}\varepsilon^5/5!+45A^2\mu_{210}\varepsilon^6/6!) + (4BI\varepsilon^4/4!$$
$$-20B\mu_{300}\varepsilon^5/5!+60B\mu_{210}\varepsilon^6/6!) + (5CI\varepsilon^5/5!-30C\mu_{300}\varepsilon^6/6!)$$
$$+(10ABI\varepsilon^5/5!-60AB\mu_{300}\varepsilon^6/6!) + (6DI\varepsilon^6/6!) + (15ACI\varepsilon^6/6!)$$
$$+(10B^2I\varepsilon^6/6!)\} + \{(\mu_{300}\varepsilon^3/3!-4\mu_{400}\varepsilon^4/4!+10\mu_{310}\varepsilon^5/5!-20\mu_{301}\varepsilon^6/6!)$$
$$+(6A\mu_{300}\varepsilon^4/4!-30A\mu_{400}\varepsilon^5/5!+90A\mu_{310}\varepsilon^6/6!) + (10B\mu_{300}\varepsilon^5/5!$$
$$-60B\mu_{400}\varepsilon^6/6!) + (15A^2\mu_{300}\varepsilon^5/5!-90A^2\mu_{400}\varepsilon^6/6!)$$
$$+(60AB\mu_{300}\varepsilon^6/6!) + (15A^3\mu_{300}\varepsilon^6/6!) + (15C\mu_{300}\varepsilon^6/6!)\}$$

(CONT→)

$$+\ \{(\mu_{400}\varepsilon^4/4!-5\mu_{500}\varepsilon^5/5!+15\mu_{600}\varepsilon^6/6!) + (10A\mu_{400}\varepsilon^5/5!$$
$$-60A\mu_{500}\varepsilon^6/6!) + (20B\mu_{400}\varepsilon^6/6!) + (45A^2\mu_{400}\varepsilon^6/6!)\}$$
$$+\ \{(\mu_{500}\varepsilon^5/5!-6\mu_{600}\varepsilon^6/6!) + (15A\mu_{500}\varepsilon^6/6!)\}$$
$$+\ \{(\mu_{600}\varepsilon^6/6!)\} + o(\varepsilon^6). \qquad (2.56)$$

Simplifying, we get

$$\begin{aligned}\rho_0^* = 1 &- I\varepsilon^2/2! - (2\mu_{300}-3\mu_{110})\varepsilon^3/3! - \{(3\mu_{400}-6\mu_{210}+4\mu_{101})\\ &-3A^2I+6A(\mu_{300}-\mu_{110})\}\varepsilon^4/4! - \{(4\mu_{500}-10\mu_{310}+10\mu_{201}-5\mu_{1001})\\ &+10A(2\mu_{400}-3\mu_{210}+\mu_{101})-10ABI+10B(\mu_{300}-\mu_{110})\}\varepsilon^5/5! \qquad (2.57)\\ &- \{(5\mu_{600}-15\mu_{410}+20\mu_{301}-15\mu_{2001}+6\mu_{10001})-15A^3\mu_{300}\\ &+45A^2(\mu_{400}-\mu_{210})+15A(3\mu_{500}-6\mu_{310}+4\mu_{201}-\mu_{1001})-15ACI\\ &-10B^2I+20B(2\mu_{400}-3\mu_{210}+\mu_{101})+15C(\mu_{300}-\mu_{110})\}\varepsilon^6/6! + o(\varepsilon^6).\end{aligned}$$

Similarly, if we differentiate (2.54) with respect to t and then insert $t=t_0^*$,

$$\begin{aligned}\dot{\rho}_0^*(t_0^*) &= \{-I\varepsilon+\mu_{110}\varepsilon^2/2!-\mu_{101}\varepsilon^3/3!\} + \{(I\varepsilon-2\mu_{300}\varepsilon^2/2!+3\mu_{210}\varepsilon^3/3!\\ &\quad+(AI\varepsilon^2/2!-3A\mu_{300}\varepsilon^3/3!)+(BI\varepsilon^3/3!)\} + \{(\mu_{300}\varepsilon^2/2!-3\mu_{400}\varepsilon^3/3!)\\ &\quad+(3A\mu_{300}\varepsilon^3/3!)\} + \{(\mu_{400}\varepsilon^3/3!)\} + o(\varepsilon^3) \qquad (2.58)\\ &= (AI+\mu_{110}-\mu_{300})\varepsilon^2/2! + (BI-2\mu_{400}+3\mu_{210}-\mu_{101})\varepsilon^3/3! + o(\varepsilon^3).\end{aligned}$$

Hence, setting each coefficient to zero, we have

$$A = (\mu_{300} - \mu_{110})/I(\theta) = -J(\theta)/I(\theta), \qquad (2.59)$$
$$B = (2\mu_{400} - 3\mu_{210} + \mu_{101})/I(\theta) = \eta_{101}/I(\theta). \qquad (2.60)$$

The value of C is not needed as the terms containing C will cancel when we insert (2.59) and (2.60) into (2.57).

$$\begin{aligned}\rho_0^* = 1 &- I\varepsilon^2/2! - (2\mu_{300}-3\mu_{110})\varepsilon^3/3! - \{(3\mu_{400}-6\mu_{210}+4\mu_{101})\\
&+3J^2/I\}\varepsilon^4/4! - \{(4\mu_{500}-10\mu_{310}+10\mu_{201}-5\mu_{1001})\\
&-10J(2\mu_{400}-3\mu_{210}+\mu_{101})/I\}\varepsilon^5/5! - \{(5\mu_{600}-15\mu_{410}+20\mu_{301}\\
&-15\mu_{2001}+6\mu_{10001})+15J^3\mu_{300}/I^3+45J^2(\mu_{400}-\mu_{210})/I^2\\
&-15J(3\mu_{500}-6\mu_{310}+4\mu_{201}-\mu_{1001})/I+10(2\mu_{400}-3\mu_{210}+\mu_{101})^2/I\}\varepsilon^6/6!\\
&+ o(\varepsilon^6). \qquad (2.61)\end{aligned}$$

We expand (2.61) into a logarithmic series,

$$\begin{aligned}-\log\rho_0^* = I(\theta)\varepsilon^2/2! &+ (2\mu_{300}-3\mu_{110})\varepsilon^3/3! + \{(3\mu_{400}-6\mu_{210}+4\mu_{101})\\
&+3J^2/I+3I^2\}\varepsilon^4/4! + \{(4\mu_{500}-10\mu_{310}+10\mu_{201}-5\mu_{1001})\\
&-10J(2\mu_{400}-3\mu_{210}+\mu_{101})/I+10I(2\mu_{300}-3\mu_{110})\}\varepsilon^5/5!\\
&+ \{(5\mu_{600}-15\mu_{410}+20\mu_{301}-15\mu_{2001}+6\mu_{10001})+15J^3\mu_{300}/I^3\\
&+15J^2(3\mu_{400}-3\mu_{210})/I^2+10(2\mu_{400}-3\mu_{210}+\mu_{101})^2/I \qquad (2.62)\\
&-15J(3\mu_{500}-6\mu_{310}+4\mu_{201}-\mu_{1001})/I+10(2\mu_{300}-3\mu_{110})^2\\
&+15I(3\mu_{400}-6\mu_{210}+4\mu_{101})+30I^3+45J^2\}\varepsilon^6/6! + o(\varepsilon^6).\end{aligned}$$

We write this in terms of the coefficients of the Bahadur bound, $b_i = -E_\theta \ell^{(i)}$, which are given in Lemma 2.2,

$$\begin{aligned}-\log\rho_0^* = b_2\varepsilon^2/2! &- b_3\varepsilon^3/3! + \{b_4-3I^2[(\mu_{400}-2\mu_{210}+\mu_{020})/I^2-1\\
&-J^2/I^3]\}\varepsilon^4/4! - \{b_5+10IJ[(2\mu_{500}-5\mu_{310}+\mu_{201}+3\mu_{120}-\mu_{011})/IJ\\
&+(2\mu_{400}-3\mu_{210}+\mu_{101})/I^2-(2\mu_{300}-3\mu_{110})/J]\}\varepsilon^5/5!\\
&+ \{b_6-15I^3[(5\mu_{600}-15\mu_{410}+4\mu_{301}+12\mu_{220}-4\mu_{111}-2\mu_{030}-\mu_{2001}\\
&+\mu_{0101})/I^3-(3\mu_{400}-6\mu_{210}+4\mu_{101})/I^2-2-J(\mu_{300}-3\mu_{110})/I^3\end{aligned}$$

(CONT.)

$$+J(7\mu_{500}-16\mu_{310}+6\mu_{201}+6\mu_{120}-2\mu_{011}-\mu_{1001})/I^4$$

$$+J^2(\mu_{400}-3\mu_{210}+2\mu_{101})/I^5-J^3\mu_{300}/I^6]$$

$$-10I^3[(4\mu_{600}-12\mu_{410}+4\mu_{301}+9\mu_{220}-6\mu_{111}+\mu_{002})/I^3$$

$$-(2\mu_{400}-3\mu_{210}+\mu_{101})^2/I^4-(2\mu_{300}-3\mu_{110})^2/I^3$$

$$-3J(2\mu_{500}-5\mu_{310}+\mu_{201}+3\mu_{120}-\mu_{011})/I^4+3J(2\mu_{300}-3\mu_{110})/I^3$$

$$-3J^2(2\mu_{400}-3\mu_{210}+\mu_{101})/I^5]]\varepsilon^6/6! + o(\varepsilon^6). \qquad (2.63)$$

By Definition 2.2, we get the final result,

$$-\log \rho_0^* = b_2\varepsilon^2/2! - b_3\varepsilon^3/3! + \{b_4 - 3I^2(\theta)\gamma_\theta^2\}\varepsilon^4/4!$$
$$-\{b_5 - 10I(\theta)J(\theta)\delta_\theta\}\varepsilon^5/5! + \{b_6 - 15I^3(\theta)\lambda_\theta^3$$
$$- 10I^3(\theta)\nu_\theta^3\}\varepsilon^6/6! + o(\varepsilon^6). \qquad (2.64)$$

Then, to obtain $-\log \rho_1$ we can follow the same procedure, or more simply, use the complementation principle in that reversing H_0 and H_1 is equivalent to replacing ε with $-\varepsilon$. Hence,

$$-\log \rho_1^* = b_2\varepsilon^2/2! + b_3\varepsilon^3/3! + \{b_4 - 3I^2(\theta)\gamma_\theta^2\}\varepsilon^4/4!$$
$$+\{b_5 - 10I(\theta)J(\theta)\delta_\theta\}\varepsilon^5/5! + \{b_6 - 15I^3(\theta)\lambda_\theta^3$$
$$- 10I^3(\theta)\nu_\theta^3\}\varepsilon^6/6! + o(\varepsilon^6). \qquad (2.65)$$

Finally, we combine (2.64) and (2.65), the exponential rate of the mle is the larger of the two tail probabilities, hence the smaller of the two expressions above. This completes the proof.Δ

2.4 Exponential Rate of the pre

To expand the exponential rate of the pre, as given by Theorem 2.1, we use a similar procedure as was used for the mle.

Theorem 2.3: *If* $F(x|\theta)$ *is a regular translation invariant distribution for the location parameter* θ *and the additional condition* [C1] *holds, then the exponential rate of convergence of the tail probability to zero for the pre, as given in Theorem 2.1, has a local expansion in the neighbourhood of* $\varepsilon=0$ *given by*

$$b(\tilde{\theta},\theta,\varepsilon) = \min_{\pm\varepsilon}\{b_2\varepsilon^2/2! + b_3\varepsilon^3/3! + [b_4 - 3I^2(\theta)\gamma_\theta^2]\varepsilon^4/4! + b_5\varepsilon^5/5! + [b_6 - 15I^3(\theta)\lambda_\theta^3]\varepsilon^6/6! + o(\varepsilon^6)\}. \tag{2.66}$$

∇*Proof:*

In this case, the functions to be minimized in (2.12) are strictly convex functions (under the regularity conditions) with the infima in $0<t<1$. Again, by Hoeffding (1965), there is a unique τ_θ such that the minimum value is attained at τ_θ. Under symmetry, $\tau_\theta=\frac{1}{2}$ by Sievers (1978), hence in general: $\tau_\theta=\tau_\theta(\varepsilon)=\frac{1}{2}+o(\varepsilon)$.

Referring back to Theorem 2.1, we proceed as in the last section, minimizing

$$\rho_0(t) = E_{F_0} e^{t(Y - \mu)} = E_{F_0} e^{sW}, \tag{2.67}$$

where $s=2t$ and $W=\frac{1}{2}(Y - \mu)$. We expand Y as a Taylor series in ε, to obtain

$$\begin{aligned}
Y &= \log f(x|\theta + \varepsilon)/f(x|\theta - \varepsilon) \\
&= \log f(x|\theta + \varepsilon) - \log f(x|\theta - \varepsilon) \\
&= \log\left\{ \sum_{i=0}^{\infty} \frac{\partial^i f(x|\theta)}{\partial\theta^i} \frac{\varepsilon^i}{i!}\right\} - \log\left\{ \sum_{i=0}^{\infty} \frac{\partial^i f(x|\theta)}{\partial\theta^i} \frac{(-\varepsilon)^i}{i!}\right\} \\
&= \log\left\{ f(x|\theta) \sum_{i=0}^{\infty} d_i \frac{\varepsilon^i}{i!}\right\} - \log\left\{ f(x|\theta) \sum_{i=0}^{\infty} d_i \frac{(-\varepsilon)^i}{i!}\right\} \\
&= \{d_1\varepsilon + (d_2-d_1^2)\varepsilon^2/2! + (d_3-3d_1d_2+2d_1^3)\varepsilon^3/3! \\
&\quad + (d_4-4d_1d_3-3d_2^2+12d_1^2d_2-6d_1^4)\varepsilon^4/4! + (d_5 \\
&\quad -5d_1d_4-10d_2d_3+30d_1d_2^2+20d_1^2d_3-60d_1^3d_2 \\
&\quad +24d_1^5)\varepsilon^5/5! + (d_6-6d_1d_5-15d_2d_4+30d_1^2d_4 \\
&\quad -10d_3^2+30d_2^3+120d_1d_2d_3-270d_1^2d_2^2-120d_1^3d_3 \\
&\quad +360d_1^4d_2-120d_1^6)\varepsilon^6/6!\} - \{-d_1\varepsilon + (d_2 \\
&\quad -d_1^2)\varepsilon^2/2! - (d_3-3d_1d_2+2d_1^3)\varepsilon^3/3! + (d_4 \\
&\quad -4d_1d_3-3d_2^2+12d_1^2d_2-6d_1^4)\varepsilon^4/4! - (d_5-5d_1d_4 \\
&\quad -10d_2d_3+30d_1d_2^2+20d_1^2d_3-60d_1^3d_2+24d_1^5)\varepsilon^5/5! \\
&\quad + (d_6-6d_1d_5-15d_2d_4+30d_1^2d_4-10d_3^2+30d_2^3+120d_1d_2d_3 \\
&\quad -270d_1^2d_2^2-120d_1^3d_3+360d_1^4d_2-120d_1^6)\varepsilon^6/6!\} + o(\varepsilon^6)
\end{aligned}$$

$$= 2d_1\varepsilon + 2(d_3-3d_1d_2+2d_1^3)\varepsilon^3/3! + 2(d_5-5d_1d_4-10d_2d_3$$
$$+30d_1d_2^2+20d_1^2d_3-60d_1^4d_2+24d_1^6)\varepsilon^5/5! + o(\varepsilon^6). \qquad (2.68)$$

Taking the expectations, and using the definition of Fisher moments (2.26), we obtain $\mu = \mu(\varepsilon) = E_\theta Y$,

$$\mu = 2(2\mu_{300}-3\mu_{110})\varepsilon^3/3! + 2(24\mu_{500}-60\mu_{310}+20\mu_{201}+30\mu_{120}$$
$$-10\mu_{011}-5\mu_{1001})\varepsilon^5/5! + o(\varepsilon^6) \qquad (2.69)$$
$$=-2b_3\varepsilon^3/3! - 2b_5\varepsilon^5/5! + o(\varepsilon^6).$$

Hence,

$$W = d_1\varepsilon + (d_3-3d_1d_2+2d_1^3+b_3)\varepsilon^3/3! + (d_5-5d_1d_4-10d_2d_3$$
$$+30d_1d_2^2+20d_1^2d_3-60d_1^4d_2+24d_1^6+b_5)\varepsilon^5/5! + o(\varepsilon^6). \qquad (2.70)$$

Returning to (2.67) and noting that minimization with respect to t is equivalent to minimization with respect to s, we expand $\rho_0^*(s)$ into a double Taylor series, in s and ε,

$$\rho_0(s) = E_{F_0} e^{sW}$$

$$= E_\theta\left\{\left(\sum_{m=0}^{\infty} d_m \frac{(-\varepsilon)^m}{m!}\right)\cdot\left(\sum_{n=0}^{\infty} \frac{(sW)^n}{n!}\right)\right\} \qquad (2.71)$$

$$= \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} \frac{s^n}{n!}\left\{E_\theta(d_m W^n)\frac{(-\varepsilon)^m}{m!}\right\}.$$

Inserting (2.70) into (2.71) and expanding up to the sixth degree in ε, we obtain

$$
\begin{aligned}
\rho_0(s) = 1 &+ sE_\theta\{-[2d_1^2\varepsilon^2/2!+4(d_1d_3-3d_1^2d_2+2d_1^4+d_1b_3)\varepsilon^4/4! \\
&\quad +6(d_1d_5-5d_1^2d_4-10d_1d_2d_3+30d_1^2d_2^2+20d_1^3d_3-60d_1^4d_2 \\
&\quad +24d_1^6+d_1b_5)\varepsilon^6/6!] + [3d_1d_2\varepsilon^3/3!+10(d_2d_3-3d_1d_2^2 \\
&\quad +2d_1^3d_2+d_2b_3)\varepsilon^5/5!] - [4d_1d_3\varepsilon^4/4!+20(d_3^2-3d_1d_2d_3 \\
&\quad +2d_1^3d_3+d_3b_3)\varepsilon^6/6!] + [5d_1d_4\varepsilon^5/5!] - [6d_1d_5\varepsilon^6/6!]\} \\
&+ s^2E_\theta\{[d_1^2\varepsilon^2/2!+4(d_1d_3-3d_1^2d_2+2d_1^4+d_1b_3)\varepsilon^4/4! \\
&\quad +(6d_1d_5-30d_1^2d_4-120d_1d_2d_3+270d_1^2d_2^2+160d_1^3d_3+10d_3^2 \\
&\quad -480d_1^4d_2+184d_1^6+10b_3^2+20b_3(d_3-3d_1d_2+2d_1^3)+6d_1b_5)\frac{\varepsilon^6}{6!}] \\
&\quad - [3d_1^3\varepsilon^3/3!+20(d_1^2d_3-3d_1^3d_2+2d_1^5+d_1^2b_3)\varepsilon^5/5!] \\
&\quad + [6d_1^2d_2\varepsilon^4/4!+60(d_1d_2d_3-3d_1^2d_2^2+2d_1^4d_2+d_1d_2b_3)\frac{\varepsilon^6}{6!}] \\
&\quad - [-10d_1^2d_3\varepsilon^5/5!] + [15d_1^2d_4\varepsilon^6/6!]\} \\
&+ s^3E_\theta\{[d_1^3\varepsilon^3/3!+10(d_1^2d_3-3d_1^3d_2+2d_1^5+d_1^2b_3)\varepsilon^5/5!] \\
&\quad - [4d_1^4\varepsilon^4/4!+60(d_1^3d_3-3d_1^4d_2+2d_1^6+d_1^3b_3)\varepsilon^6/6!] \\
&\quad + [10d_1^3d_2\varepsilon^5/5!] - [20d_1^3d_3\varepsilon^6/6!]\} \\
&+ s^4E_\theta\{[d_1^4\varepsilon^4/4!+20(d_1^3d_3-3d_1^4d_2+2d_1^6+d_1b_3)\varepsilon^6/6!] \\
&\quad - [5d_1^5\varepsilon^5/5!] + [15d_1^4d_2\varepsilon^6/6!]\} \\
&+ s^5E_\theta\{[d_1^5\varepsilon^5/5!] - [6d_1^6\varepsilon^6/6!]\} \\
&+ s^6E_\theta\{[d_1^6\varepsilon^6/6!]\} + o(\varepsilon^6). \qquad (2.72)
\end{aligned}
$$

Taking the expectations, using the definition of the Fisher moments (2.26), and collecting terms, we get the following:

$$\rho_0(s) = 1 + s[-2I\varepsilon^2/2!+3\mu_{110}\varepsilon^3/3!-4(2\mu_{400}-3\mu_{210}+2\mu_{101})\varepsilon^4/4!$$
$$+5(4\mu_{310}-6\mu_{120}+\mu_{1001}+2\mu_{011})\varepsilon^5/5!-(144\mu_{600}-360\mu_{410}$$
$$+160\mu_{301}+180\mu_{220}-120\mu_{111}-30\mu_{2001}+20\mu_{002}+12\mu_{10001})\varepsilon^6/6!]$$
$$+ s^2[I\varepsilon^2/2!-3\mu_{300}\varepsilon^3/3!+2(4\mu_{400}-3\mu_{210}+2\mu_{101})\varepsilon^4/4!-10(4\mu_{500}$$
$$-6\mu_{310}+3\mu_{201}+2Ib_3)\varepsilon^5/5!+(184\mu_{600}-360\mu_{410}+160\mu_{301}+90\mu_{220}$$
$$-60\mu_{111}-15\mu_{2001}+10\mu_{002}+6\mu_{10001}-10b_3^2+60\mu_{110}b_3)\varepsilon^6/6!]$$
$$+ s^3[\mu_{300}\varepsilon^3/3!-4\mu_{400}\varepsilon^4/4!+10(2\mu_{500}-2\mu_{310}+\mu_{201}+Ib_3)\varepsilon^5/5!$$
$$-20(6\mu_{600}-9\mu_{410}+4\mu_{301}+3\mu_{300}b_3)\varepsilon^6/6!]$$
$$+ s^4[\mu_{400}\varepsilon^4/4!-5\mu_{500}\varepsilon^5/5!+5(8\mu_{600}-9\mu_{410}+4\mu_{301}+4\mu_{300}b_3)\frac{\varepsilon^6}{6!}]$$
$$+ s^5[\mu_{500}\varepsilon^5/5!-6\mu_{600}\varepsilon^6/6!] + s^6[\mu_{600}\varepsilon^6/6!] + o(\varepsilon^6). \quad (2.73)$$

Since the minimum value of $\rho_0(t)$ is attained at $t=\frac{1}{2}$ under symmetry, we would expect that the minimum is near $\frac{1}{2}$ in the general non-symmetric case, hence let,

$$t_0=\tfrac{1}{2}s_0=\tfrac{1}{2}+A\varepsilon+B\varepsilon^2/2!+C\varepsilon^3/3!+D\varepsilon^4/4!+E\varepsilon^5/5!+o(\varepsilon^5). \quad (2.74)$$

We insert (2.74) into (2.73), which after collecting terms, gives the following:

$$\rho_0=\rho_0(s_0) = 1 + [-I\varepsilon^2/2!-(2\mu_{300}-3\mu_{110})\varepsilon^3/3!-(3\mu_{400}-6\mu_{210}+4\mu_{101})\frac{\varepsilon^4}{4!}$$
$$-(24\mu_{500}-60\mu_{310}+20\mu_{201}+30\mu_{120}-10\mu_{011}-5\mu_{1001}+10Ib_3)\frac{\varepsilon^5}{5!}$$
$$-(45\mu_{600}-135\mu_{410}+60\mu_{301}+90\mu_{220}-60\mu_{111}-15\mu_{2001}+10\mu_{002}$$
$$+6\mu_{10001}-10b_3^2)\varepsilon^6/6!] - [24A(\mu_{300}-\mu_{110})\varepsilon^4/4!+60A(7\mu_{500}$$
$$-16\mu_{310}+6\mu_{201}+6\mu_{120}-2\mu_{011}-\mu_{1001}+2Ib_3)\varepsilon^6/6!] + [48A^2I\frac{\varepsilon^4}{4!}$$
$$-240A^2(\mu_{400}-3\mu_{210}+2\mu_{101})\varepsilon^6/6!] + [960A^3\mu_{300}\varepsilon^6/6!]$$
$$+ [240ABI\varepsilon^5/5!] - [60B(\mu_{300}-\mu_{110})\varepsilon^5/5!] + [360B^2I\varepsilon^6/6!]$$
$$+ [480ACI\varepsilon^6/6!] - [120C(\mu_{300}-\mu_{110})\varepsilon^6/6!] + o(\varepsilon^6). \quad (2.75)$$

Similarly, differentiating (2.73) with respect to s and inserting (2.74), and collecting terms, we get

$$\dot{\rho}_0(s_0) = [12AI+3\mu_{110}-3\mu_{300}]\varepsilon^3/3!+[24BI]\varepsilon^4/4!+o(\varepsilon^4). \quad (2.76)$$

Hence,

$$A = (\mu_{300} - \mu_{110})/4I(\theta) = -J(\theta)/4I(\theta), \quad (2.77)$$

$$B = 0. \quad (2.78)$$

Once again, the terms involving C in (2.75) will cancel when we insert the A and B above. This gives

$$\begin{aligned}\rho_0 = 1 &- I\varepsilon^2/2! - (2\mu_{300}-3\mu_{110})\varepsilon^3/3! - [(3\mu_{400}-6\mu_{210}+4\mu_{101})\\ &+3J^2/I]\varepsilon^4/4! - [(24\mu_{500}-60\mu_{310}+20\mu_{201}+30\mu_{120}-10\mu_{011}\\ &-5\mu_{1001})+10Ib_3]\varepsilon^5/5! - [(45\mu_{600}-135\mu_{410}+60\mu_{301}+90\mu_{220}\\ &-60\mu_{111}-15\mu_{2001}+10\mu_{002}-6\mu_{10001})+10b_3(5\mu_{300}-6\mu_{110})\\ &+15J^3\mu_{300}/I^3-15J^2(\mu_{400}-3\mu_{210}+2\mu_{101})/I^2-15J(7\mu_{500}-16\mu_{310}\\ &+6\mu_{201}+6\mu_{120}-2\mu_{011}-\mu_{1001})/I]\varepsilon^6/6! + o(\varepsilon^6). \quad (2.79)\end{aligned}$$

We expand (2.79) in a logarithmic series to obtain

$$\begin{aligned}-\log\ \rho_0 = &\ I\varepsilon^2/2! + (2\mu_{300}-3\mu_{110})\varepsilon^3/3! + [(3\mu_{400}-6\mu_{210}+4\mu_{101})\\ &+3J^2/I+3I^2]\varepsilon^4/4! + (24\mu_{500}-60\mu_{310}+20\mu_{201}+30\mu_{120}\\ &-10\mu_{011}-5\mu_{1001})\varepsilon^5/5! + [(45\mu_{600}-135\mu_{410}+60\mu_{301}\\ &+90\mu_{220}-60\mu_{111}-15\mu_{2001}+10\mu_{002}+6\mu_{10001})+15J(\mu_{300}\\ &-3\mu_{110})+30I^3+15I(3\mu_{400}-6\mu_{210}+4\mu_{101})-15J^3\mu_{300}/I^3\\ &-15J(7\mu_{500}-16\mu_{310}+6\mu_{201}+6\mu_{120}-2\mu_{011}-\mu_{1001})/I\\ &-15J^2(\mu_{400}-3\mu_{210}+2\mu_{101})/I^2]\varepsilon^6/6! + o(\varepsilon^6). \quad (2.80)\end{aligned}$$

In terms of the coefficients of the Bahadur bound, $b_i = -E_\theta \ell^{(i)}$, equation (2.80) becomes

$$
\begin{aligned}
-\log \rho_0 = {} & b_2\varepsilon^2/2! - b_3\varepsilon^3/3! + [b_4 - 3(\mu_{400} - 2\mu_{210} + \mu_{020}) \\
& + 3J^2/I + 3I^2]\varepsilon^4/4! - b_5\varepsilon^5/5! + [b_6 - 15(5\mu_{600} \\
& - 15\mu_{410} + 4\mu_{301} + 12\mu_{220} - 4\mu_{111} - 2\mu_{030} - \mu_{2001} + \mu_{0101}) \\
& + 15J(\mu_{300} - 3\mu_{110}) + 30I^3 + 15I(3\mu_{400} - 6\mu_{210} + 4\mu_{101}) \\
& + 15J^3\mu_{300}/I^3 - 15J(7\mu_{500} - 16\mu_{310} + 6\mu_{201} + 6\mu_{120} - 2\mu_{011} \\
& - \mu_{1001})/I - 15J^2(\mu_{400} - 3\mu_{210} + 2\mu_{101})/I^2]\frac{\varepsilon^6}{6!} + o(\varepsilon^6). \quad (2.81)
\end{aligned}
$$

By Definition 2.2, we then obtain

$$
\begin{aligned}
-\log \rho_0 = {} & b_2\varepsilon^2/2! - b_3\varepsilon^3/3! + [b_4 - 3I^2(\theta)\gamma_\theta^2]\varepsilon^4/4! \\
& - b_5\varepsilon^5/5! + [b_6 - 15I^3(\theta)\lambda_\theta^3]\varepsilon^6/6! + o(\varepsilon^6).
\end{aligned} \quad (2.82)
$$

And, again, we can obtain $-\log \rho_1$ by the same procedure or by replacing ε with $-\varepsilon$ in (2.82). This gives

$$
\begin{aligned}
-\log \rho_1 = {} & b_2\varepsilon^2/2! + b_3\varepsilon^3/3! + [b_4 - 3I^2(\theta)\gamma_\theta^2]\varepsilon^4/4! \\
& + b_5\varepsilon^5/5! + [b_6 - 15I^3(\theta)\lambda_\theta^3]\varepsilon^6/6! + o(\varepsilon^6).
\end{aligned} \quad (2.83)
$$

The theorem follows. Δ

Thus we have seen that the differences between the exponential rates of the pre and mle with the Bahadur bound depend on the rate coefficients, γ_θ^2, δ_θ, λ_θ^3, and ν_θ^3, as defined in Definition 2.2. The rate of the pre differs from the Bahadur bound only in terms which depend on γ_θ^2 and λ_θ^3, while the rate of the mle differs from the Bahadur bound in terms of all four parameters. The parameters δ_θ and ν_θ^3, which describe the divergence of the mle from the pre, are of particular interest since both these estimator-sequences do exist. On the other hand, it has not been established as to what conditions are neccessary in order that the Bahadur bound be attainable. In the cases where the bound cannot be attained, it should be possible to improve upon this bound with an even sharper bound. We have seen that the mle is optimal and attains the Bahadur bound for the exponential location family, when $\gamma_\theta^2=0$. Also, Madsen (1979) has shown that $\gamma_\theta^2=0$ only for the exponential family. Theorems 2.2 and 2.3 seem to suggest that the Bahadur bound can be attained only if $\gamma_\theta^2=0$ (i.e.: exponential family). Thus we are led to believe that, except for the exponential family, the Bahadur bound is too large. We therefore make the following conjectures.

Conjecture 1: *The optimal exponential rate of convergence of the tail probability of any consistent estimator of a translation invariant location parameter θ should under non-exponential families be less than the Bahadur bound, with a fifth-degree expansion given by:*

$$b_C(\theta,\varepsilon) = \min_{\pm\varepsilon}\{b_2\varepsilon^2/2! + b_3\varepsilon^3/3! + (b_4 - 3I(\theta)\gamma_\theta^2)\varepsilon^4/4! + b_5\varepsilon^5/5! + o(\varepsilon^5)\}, \tag{2.84}$$

where C is the class of all consistent estimators of θ.

This conjecture follows from the fact that $B(\theta,\varepsilon) \geq b_C(\theta,\varepsilon) \geq b(\tilde{\theta},\theta,\varepsilon)$ and by the second-order efficiency (1.36), which means that the fourth-degree term's coefficient can be no larger than that given in (2.84).

Conjecture 2: *The optimal exponential rate of convergence of the tail probability of any consistent estimator of a translation invariant location parameter θ should equal to the optimal exponential rate for the class of translation invariant estimators:* $b_C(\theta,\varepsilon) = b_T(\theta,\varepsilon)$.

This conjecture follows from the apparent paradox that would ensue if the class of optimal estimators is not translation invariant. If this were the case, then the exponential rates would not be translation invariant with θ. This would mean that the optimal class of estimators would be indexed by θ. To avoid such an occurrence, we would require the optimal class to have the same exponential rate for all T, and hence be translation invariant.

Lastly, we conjecture that the pre is optimal for T and hence for C as well (by Conjecture 2), in the general case (beyond symmetry) of a regular translation invariant population.

Conjecture 3: The optimal exponential rate of convergence to zero of the tail probability of any consistent estimator of a translation invariant location parameter, θ, should be given by the exponential rate of the pre, as stated in Theorem 2.1: $b_C(\theta,\varepsilon) = b_T(\theta,\varepsilon) = b(\tilde{\theta},\theta,\varepsilon)$.

This conjecture follows from Conjecture 2 and Remark 3 following Theorem 2.1. Also, by Conjecture 1, the pre and $b_C(\theta,\varepsilon)$ would have at least fifth-degree contact since the expansion of the pre co-incides with the expansion of the Bahadur bound in the odd degree terms.

In the next chapter we will look at various applications of these results under various regular distributions and observe the behaviour of the rate coefficients in order to assess their properties and statistical significance. In Chapter 4 we will examine some non-regular distributions in which the pre is shown to dominate the mle, even though the theorems may not apply and the Taylor series expansions may not be valid. In these cases, we can still compute the rate coefficients, although they are not measures of the differences of the Taylor series expansions. This may shed some further light on the statistical interpretation of these quantities, which we will discuss in Chapter 5.

CHAPTER THREE

Translation Invariance and Scale Invariance

3.1 Log-Transformation of Scale Invariant Exponentials

In this chapter we would like to examine the properties of the rate coefficients, γ_θ^2, δ_θ, λ_θ^3, ν_θ^3, and investigate some applications. A good place to begin is with the exponential family which is generally well-behaved and tractable. If we consider the general scale invariant distribution on σ,

$$g(y|\sigma) = \frac{cpm^q}{\Gamma(q)} \frac{y^{pq-1} e^{-m(y/\sigma)^p}}{\sigma^{pq}}, \quad \sigma>0, \tag{3.1}$$

with $p \geq 1$; $q>0$; $m>0$; and if $c=1$ then $y>0$, while if $c=\frac{1}{2}$ and p is an even integer then Y is defined here on all of R^1. The family defined by (3.1) for suitable choices of (m,p,q) includes: the Normal ($m=\frac{1}{2}$, $p=2$, $q=\frac{1}{2}$, $c=\frac{1}{2}$); the Gamma ($m=1$, $p=1$, $q=\alpha$, $c=1$); and the Weibull ($m=1$, $c=1$). Then, under a logarithmic transformation, $X = \log |Y|$ and with $\theta = \log \sigma$, we obtain the translation invariant family, defined by the density

$$f(x|\theta) = \frac{p\ m^q}{\Gamma(q)} \exp\left\{pq(x - \theta) - me^{p(x - \theta)}\right\}, \tag{3.2}$$

for $-\infty<x<\infty$ and $-\infty<\theta<\infty$. The graph of the log-transform of the Normal is given in Figure 3.1. This distribution is log-concave and satisfies all regularity

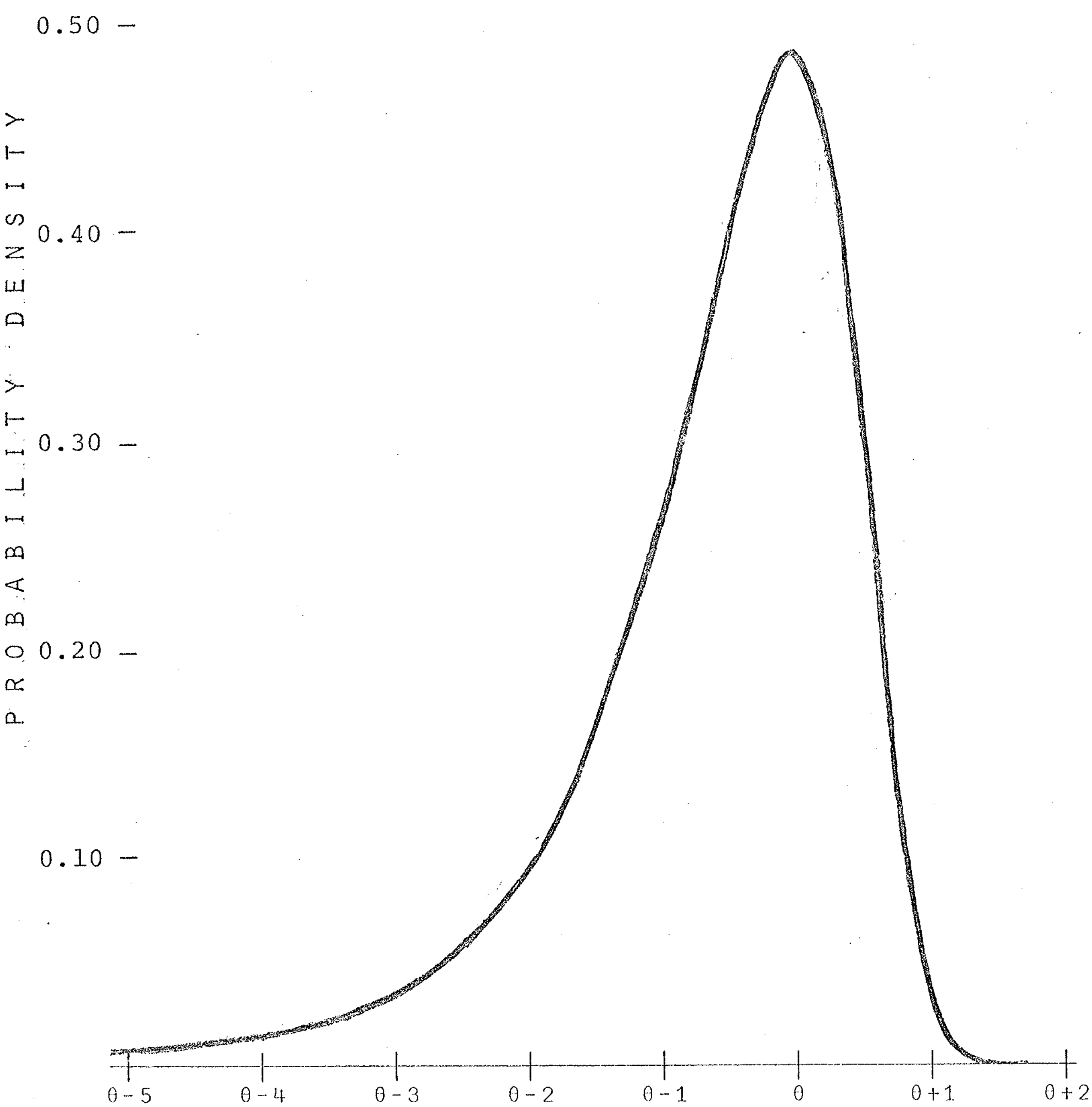

FIGURE 3.1: Graph of the density function

$$f(x|\theta) = \sqrt{2/\pi}\ \exp\{(x - \theta) - \tfrac{1}{2}\, e^{2(x - \theta)}\}.$$

conditions. For the density function in (3.2), the probability ratio for the pre is given by,

$$\begin{aligned} c(x|\theta) &= f(x|\theta + \varepsilon) \,/\, f(x|\theta - \varepsilon) \\ &= \exp[-2pq\varepsilon + 2me^{p(x-\theta)}\sinh(p\varepsilon)], \end{aligned} \tag{3.3}$$

and

$$c_n(s|\theta) = \exp[-2npq\varepsilon + 2nqe^{p(\hat{\theta}_n-\theta)}\sinh(p\varepsilon)]. \tag{3.4}$$

Here $\hat{\theta}_n = \hat{\theta}_n(s)$ is the mle, which is

$$\hat{\theta}_n(s) = \frac{1}{p}\log\left[\frac{m}{nq}\sum_{i=1}^{n}\exp(px_i)\right]. \tag{3.5}$$

The Kullback-Leibler information number is

$$\begin{aligned} K(\theta^*, \theta) &= E_{\theta^*}\log f(x|\theta^*)/f(x|\theta) \\ &= pq(\theta-\theta^*) - q[1 - e^{-p(\theta-\theta^*)}]. \end{aligned} \tag{3.6}$$

Thus, the Bahadur bound for consistent estimates of θ,

$$\begin{aligned} B(\theta,\varepsilon) &= \inf_{\theta^*}\{K(\theta^*, \theta) : |\theta-\theta^*| > \varepsilon\} \\ &= pq\varepsilon - q(1 - e^{-p\varepsilon}). \end{aligned} \tag{3.7}$$

Then, since $F(x|\theta)$ is a regular distribution, we can apply Theorems 2.1 to obtain the exponential rate of the probability ratio estimator. First, $\mu = E_\theta Y$, which is given by (2.10), is computed as

$$\begin{aligned}\mu &= K(\theta,\ \theta - \varepsilon) - K(\theta,\ \theta + \varepsilon)\\ &= [-pq\varepsilon - q(1-e^{p\varepsilon})] - [pq\varepsilon - q(1-e^{-p\varepsilon})] \qquad (3.8)\\ &= -2pq\varepsilon + 2q\sinh(p\varepsilon).\end{aligned}$$

The Chernoff Information is given by

$$\begin{aligned}m_0(t) &= E_\theta e^{tY}\\ &= \frac{p\ m^q}{\Gamma(q)}\, e^{pq(1-2t)\varepsilon}\\ &\quad \cdot\int_{-\infty}^{\infty} \exp\,[pq(x-\theta) + m[(1-t)e^{p\varepsilon} + te^{-p\varepsilon}]e^{p(x-\theta)}]\,dx\\ &= \frac{p\ m^q}{\Gamma(q)}\, e^{pq(1-2t)\varepsilon}\left(\frac{\Gamma(q)}{p\,(m[(1-t)e^{p\varepsilon}+te^{-p\varepsilon}])^q}\right)\\ &= \left(\frac{e^{p(1-2t)\varepsilon}}{(1-t)e^{p\varepsilon} + te^{-p\varepsilon}}\right)^q. \qquad (3.9)\end{aligned}$$

We then obtain

$$\begin{aligned}\rho_0(t) &= e^{-t\mu}\ m_0(t) \qquad (3.10)\\ &= [\,e^{p\varepsilon - 2t\sinh(p\varepsilon)}/[(1-t)e^{p\varepsilon}+te^{-p\varepsilon}]]^q.\end{aligned}$$

This quantity attains its minimum when $t = \tau_\theta(\varepsilon)$ is the solution of $\partial/\partial t\ \rho_0(t) = 0$, which gives

$$\tau_\theta(\varepsilon) = e^{\frac{1}{2}p\varepsilon}\ \sinh(\tfrac{1}{2}p\varepsilon)/\sinh(p\varepsilon). \qquad (3.11)$$

Hence

$$\rho_0 = \inf_{0<t<1} \rho_0(t) = \rho_0(\tau_\theta(\varepsilon))$$

$$= \exp[pq\varepsilon - 2qe^{p\varepsilon} \sinh(p\varepsilon)] \tag{3.12}$$

$$= \exp[pq\varepsilon + q(1 - e^{p\varepsilon})].$$

Then ρ_1 can be obtained from (3.12) by replacing ε with $-\varepsilon$, or from Remark 1 of Theorem 2.1,

$$\rho_1 = \exp[-pq\varepsilon + q(1 - e^{-p\varepsilon})], \tag{3.13}$$

and so

$$\begin{aligned} b(\tilde{\theta},\theta,\varepsilon) &= -\log max\{\rho_0, \rho_1\} \\ &= pq\varepsilon - q(1 - e^{-p\varepsilon}) = B(\theta,\varepsilon). \end{aligned} \tag{3.14}$$

Therefore the pre is optimal for the class of all consistent estimators of θ. The rate of the mle, as given in Lemma 2.4, is computed from

$$\rho_0^*(t) = E_{F_0} \exp\{t \cdot \frac{\partial}{\partial\theta} \log f(x|\theta)\}$$

$$= \frac{p\ m^q}{\Gamma(q)} e^{-pq(t - \theta)}$$

$$\cdot \int_{-\infty}^{\infty} \exp [pq(x-\theta) - me^{p\varepsilon}(1-pte^{-p\varepsilon})e^{p(x-\theta)}]dx$$

$$= \frac{p\ m^q}{\Gamma(q)} e^{-pq(t-\theta)} \left(\frac{\Gamma(q)}{p[me^{p\varepsilon}(1-pte^{-p\varepsilon})]^q}\right)$$

$$= [e^{-pt}/(1 - pte^{-p\varepsilon})]^q. \tag{3.15}$$

The infimum is attained for

$$\tau_\theta(\varepsilon) = (e^{p\varepsilon} - 1)/p, \tag{3.16}$$

which gives

$$\begin{aligned} \rho_0 &= \inf_{t\geq 0} \rho_0(t) = \rho_0(\tau_\theta(\varepsilon)) \\ &= \exp[pq\varepsilon + q(1 - e^{p\varepsilon})]. \end{aligned} \tag{3.17}$$

And again, we obtain ρ_1 by replacing ε with $-\varepsilon$,

$$\rho_1 = \exp[-pq\varepsilon + q(1 - e^{-p\varepsilon})]. \tag{3.18}$$

Hence, the rate of the mle is given by

$$\begin{aligned} b(\hat{\theta},\theta,\varepsilon) &= -\log \max\{\rho_0, \rho_1\} \\ &= pq\varepsilon - q(1 - e^{-p\varepsilon}) \\ &= b(\tilde{\theta},\theta,\varepsilon) = B(\theta,\varepsilon). \end{aligned} \tag{3.19}$$

Thus, the mle is also optimal for the class of all consistent estimators of θ, as was proven by Kester (1981). In fact, if we examine the estimator-sequence of the probability ratio estimator, $\tilde{\theta}_n = \tilde{\theta}_n(s)$,

$$\begin{aligned} \tilde{\theta}_n &= \{\theta | c_n(s|\theta) = k_n\} \\ &= \hat{\theta}_n(s) - \frac{1}{p} \log\left\{\frac{(\log k_n)/n + 2pq\varepsilon}{2q\sinh(p\varepsilon)}\right\}, \end{aligned} \tag{3.20}$$

where

$$\frac{1}{n} \log k_n \to \mu = -2pq\varepsilon + 2q\sinh(p\varepsilon). \tag{3.21}$$

Therefore, $\tilde{\theta}_n = \hat{\theta}_n + o(1)$, that is, each pre estimator-sequence is also asymptotically mle.

Since both the mle and the pre are optimal, the rate coefficients, γ_θ^2, δ_0, ν_θ^3, λ_θ^3, are all zero for this regular exponential family.

Sievers (1978) shows that the estimator-sequence with the fastest exponential rate is a probability ratio-type estimator where the $k_n = k_n^*$ are chosen such that $\alpha_n(k_n^*) = \beta_n(k_n^*)$. Such an estimator-sequence has exponential rate

$$\begin{aligned} b(\theta^*,\theta,\varepsilon) &= \lim_{n\to\infty} -\frac{1}{n} \log P_\theta(|\theta_n^* - \theta| \geq \varepsilon) \\ &= \inf_{0<t<1} m_0(t) \qquad (3.22) \\ &= q[p\varepsilon \coth(p\varepsilon) - 1 - \log(p\varepsilon/\sinh(p\varepsilon))], \end{aligned}$$

where the infimum occurs for $t = 1/(1+e^{-2p\varepsilon}) - 1/2p\varepsilon$. It can be shown that $b(\theta^*,\theta,\varepsilon) > B(\theta,\varepsilon)$. Hence as probability ratio-type estimators are related to the mle, in this case, through (3.20), we can equate the two rate equations of the mle to solve for this limit of the k_n^*, to get

$$\frac{1}{n} \log k_n^* \to 0 \neq \mu. \qquad (3.22a)$$

Therefore, from (3.20) we obtain

$$\theta_n^* \rightarrow \theta - \frac{1}{p} \log\left(\frac{p\varepsilon}{\sinh(p\varepsilon)}\right). \tag{3.22b}$$

If we then correct θ_n^* for asymptotic bias, the corrected estimator-sequence will again be asymptotically mle and therefore has exponential rate equal to the exponential rate of the mle and the Bahadur bound.

3.2 Duality of Translation and Scale Invariance

The transformation of the last section is illustrative of a more general duality that exists between families that are scale invariant and those which are translation invariant. Efron & Truax (1968) have made use of this duality to obtain some results for scale invariant families by making use of the nice linear properties of the translation invariant family of exponentials [and which formed the basis of Efron's (1975) theories concerning the relationship of the statistical curvature to the mathematical curvature].

Let h(t) be a non-negative function which will integrate to unity over the positive reals. Then, we can define two random variables in terms of h(t):

1. X which is translation invariant for θ, and has density function

$$f(x|\theta) = e^{x-\theta}\ h(e^{x-\theta}). \tag{3.23}$$

2. Y which is scale invariant for σ, and has density function

$$g(y|\sigma) = (1/\sigma)h(y/\sigma). \tag{3.24}$$

Then, X and Y are implicity related, through h(t), and will be called Location-Scale Duals (through this transformation, we can explicitly write X = log Y and θ = log σ). Let us define the derivative ratios as

$$\xi_i = \{\partial^i/\partial\theta^i\ f(x|\theta)\}/\ f(x|\theta), \tag{3.25}$$

and

$$\zeta_i = \{\partial^i/\partial\sigma^i\ g(y|\sigma)\}/\ g(y|\sigma). \tag{3.26}$$

But since $e^{x-\theta} = y/\sigma$ through the duality, we get

$$d_i = \left.\frac{\frac{d^i}{dt^i}\ h(t)}{h(t)}\right|_{t=e^{x-\theta}} = \left.\frac{\frac{d^i}{dt^i}\ h(t)}{h(t)}\right|_{t=y/\sigma}. \tag{3.27}$$

Then we can express both ξ_i and ζ_i in terms of the d_i and hence in terms of each other, as given in the following lemma.

Lemma 3.1: *Under Location-Scale Duality with the derivative ratios ζ_i and ξ_i as defined in (3.25) and (3.26), the following relationships exist:*

$$
\begin{aligned}
&(A)\quad \zeta_1 = \xi_1/\sigma,\\
&(B)\quad \zeta_2 = (\xi_2 - \xi_1)/\sigma^2,\\
&(C)\quad \zeta_3 = (\xi_3 - 3\xi_2 + 2\xi_1)/\sigma^3,\\
&(D)\quad \zeta_4 = (\xi_4 - 6\xi_3 + 11\xi_2 - 6\xi_1)/\sigma^4,\\
&(E)\quad \zeta_5 = (\xi_5 - 10\xi_4 + 35\xi_3 - 50\xi_2 + 24\xi_1)/\sigma^5.
\end{aligned}
\tag{3.28}
$$

Proof:

We differentiate (3.23) to obtain expressions for ζ_i in terms of the d_i defined in (3.27), where $t=e^{x-\theta}$,

$$
\begin{aligned}
&(A)\quad \zeta_1 = -(1+td_1),\\
&(B)\quad \zeta_2 = (1+3td_1+t^2d_2),\\
&(C)\quad \zeta_3 = -(1+7td_1+6t^2d_2+t^3d_3),\\
&(D)\quad \zeta_4 = (1+15td_1+25t^2d_2+10t^3d_3+t^4d_4),\\
&(E)\quad \zeta_5 = -(1+31td_1+90t^2d_2+65t^3d_3+15t^4d_4+t^5d_5).
\end{aligned}
\tag{3.29}
$$

Similarly, we differentiate (3.24) to obtain expressions for ξ_i in terms of the d_i defined in (3.27), where $t=y/\sigma$,

$$
\begin{aligned}
&(A)\quad \xi_1 = -(1+td_1)/\sigma,\\
&(B)\quad \xi_2 = (2+4td_1+t^2d_2)/\sigma^2,
\end{aligned}
$$

$$\text{(C)}\quad \xi_3 = -(\;6+18td_1+\;9t^2d_2+\quad t^3d_3)/\sigma^3, \quad (3.30)$$

$$\text{(D)}\quad \xi_4 = \;(\;24+\;96td_1+\;72t^2d_2+\;16t^3d_3+\quad t^4d_4)/\sigma^4,$$

$$\text{(E)}\quad \xi_5 = -(120+600td_1+600t^2d_2+200t^3d_3+25t^4d_4+t^5d_5)/\sigma^5.$$

Then, as $e^{x-\theta} = t = y/\sigma$, we can implicity solve (3.29) for the d_i and insert these into (3.30) to obtain the required results.Δ

The rate coefficients γ_θ^2, δ_θ, λ_θ^3 and ν_θ^3, defined in (2.36) to (2.39), were developed as measures of the differences between the asymptotic expansions of the pre and mle with the Bahadur bound in the case of a translation invariant location parameter from a regular distribution, however they may still have some statistical significance in general for any underlying distribution for which these constants are finite. Efron (1975) has, in a sense, done this for γ by defining this parameter as the statistical curvature and attempting to give it certain statistical interpretations. We would like to be able to do the same for the other rate coefficients, and also to make an extension of Efron's interpretations concerning the nature of the statistical curvature. Hence, it may be useful to investigate the various relationships that must exist between the Location-Scale Duals: how do γ_θ^2, δ_θ, λ_θ^3, ν_θ^3 relate to γ_σ^2, δ_σ, λ_σ^3, ν_σ^3 ?

Lemma 3.2: *Under the Location-Scale Duality, as defined by (3.23) and (3.24), the following relationships exist between the rate coefficients of σ and the rate coefficients of θ,*

$$
\begin{aligned}
&(A)\quad \gamma_\sigma^2 = \gamma_\theta^2,\\
&(B)\quad \delta_\sigma = [J(\theta)\delta_\theta - 3I(\theta)\gamma_\theta^2]/[J(\theta)-I(\theta)],\\
&(C)\quad \nu_\sigma^3 = \nu_\theta^3 + [18I(\theta)\gamma_\theta^2 - 3J(\theta)(3\gamma_\theta^2+3\delta_\theta)]/I^2(\theta),\\
&(D)\quad \lambda_\sigma^3 = \lambda_\theta^3 + [\ 5I(\theta)\gamma_\theta^2 + 2J(\theta)(3\gamma_\theta^2 - 2\delta_\theta)]/I^2(\theta).
\end{aligned}
\tag{3.31}
$$

∇*Proof:*

By applying Lemma 3.1, we obtain the relationships between the Fisher moments of σ and θ,

$$
\begin{aligned}
&(A)\quad \mu_{200}(\sigma) = E\zeta_1^2 = E\xi_1^2/\sigma^2 = \mu_{200}(\theta)/\sigma^2,\\
&(B)\quad \mu_{300}(\sigma) = E\zeta_1^3 = E\xi_1^3/\sigma^3 = \mu_{300}(\theta)/\sigma^3,\\
&(C)\quad \mu_{110}(\sigma) = E\zeta_1\zeta_2 = E(\xi_1\xi_2-\xi_1^2)/\sigma^3 = [\mu_{110}(\theta)-\mu_{200}(\theta)]/\sigma^3,\\
&(D)\quad \mu_{400}(\sigma) = E\zeta_1^4 = E\xi_1^4/\sigma^4 = \mu_{400}(\theta)/\sigma^4,\\
&(E)\quad \mu_{210}(\sigma) = E\zeta_1^2\zeta_2 = E(\xi_1^2\xi_2-\xi_1^3)/\sigma^4 = [\mu_{210}(\theta)-\mu_{300}(\theta)]/\sigma^4,\\
&(F)\quad \mu_{020}(\sigma) = E\zeta_2^2 = E(\xi_2^2-2\xi_1\xi_2+\xi_2^2)/\sigma^4 = [\mu_{020}(\theta)-2\mu_{110}(\theta)+\mu_{200}(\theta)]/\sigma^4,\\
&(G)\quad \mu_{101}(\sigma) = E\zeta_1\zeta_3 = E(\xi_1\xi_3-3\xi_1\xi_2+2\xi_1^2)/\sigma^4 = [\mu_{101}(\theta)-3\mu_{110}(\theta)+2\mu_{200}(\theta)]/\sigma^4,\\
&(H)\quad \mu_{500}(\sigma) = E\zeta_1^5 = E\xi_1^5/\sigma^5 = \mu_{500}(\theta)/\sigma^5,\\
&(I)\quad \mu_{310}(\sigma) = E\zeta_1^3\zeta_2 = E(\xi_1^3\xi_2-\xi_1^4)/\sigma^5 = [\mu_{310}(\theta)-\mu_{400}(\theta)]/\sigma^5,\\
&(J)\quad \mu_{201}(\sigma) = E\zeta_1^2\zeta_3 = E(\xi_1^2\xi_3-3\xi_1^2\xi_2+2\xi_1^3)/\sigma^5 = [\mu_{201}(\theta)-3\mu_{210}(\theta)+2\mu_{300}(\theta)]/\sigma^5,\\
&(K)\quad \mu_{120}(\sigma) = E\zeta_1\zeta_2^2 = E(\xi_1\xi_2^2-2\xi_1^2\xi_2+\xi_1^3)/\sigma^5 = [\mu_{120}(\theta)-2\mu_{210}(\theta)+\mu_{300}(\theta)]/\sigma^5,\\
&(L)\quad \mu_{011}(\sigma) = E\zeta_2\zeta_3 = E(\xi_2\xi_3-3\xi_2^2-\xi_1\xi_3+5\xi_1\xi_2-2\xi_2^2)/\sigma^5\\
&\qquad\qquad = [\mu_{011}(\theta)-3\mu_{020}(\theta)-\mu_{101}(\theta)+5\mu_{110}(\theta)-2\mu_{200}(\theta)]/\sigma^5,
\end{aligned}
$$

$$\text{(M)}\quad \mu_{1001}(\sigma) = E\zeta_1\zeta_4 = E(\xi_1\xi_4-6\xi_1\xi_3+11\xi_1\xi_2-6\xi_2^2)/\sigma^5$$
$$= [\mu_{1001}(\theta)-6\mu_{101}(\theta)+11\mu_{110}(\theta)-6\mu_{200}(\theta)]/\sigma^5,$$

$$\text{(N)}\quad \mu_{600}(\sigma) = E\zeta_1^6 = E\xi_1^6/\sigma^6 = \mu_{600}(\theta)/\sigma^6, \qquad (3.32)$$

$$\text{(O)}\quad \mu_{410}(\sigma) = E\zeta_1^4\zeta_2 = E(\xi_1^4\xi_2-\xi_1^5)/\sigma^6 = [\mu_{410}(\theta)-\mu_{500}(\theta)]/\sigma^6,$$

$$\text{(P)}\quad \mu_{301}(\sigma) = E\zeta_1^3\zeta_3 = E(\xi_1^3\xi_3-3\xi_1^3\xi_2+2\xi_1^4)/\sigma^6 = [\mu_{301}(\theta)-3\mu_{310}(\theta)+2\mu_{400}(\theta)]/\sigma^6,$$

$$\text{(Q)}\quad \mu_{220}(\sigma) = E\zeta_1^2\zeta_2^2 = E(\xi_1^2\xi_2^2-2\xi_1^2\xi_2+\xi_1^4)/\sigma^6 = [\mu_{220}(\theta)-2\mu_{310}(\theta)+\mu_{400}(\theta)]/\sigma^6,$$

$$\text{(R)}\quad \mu_{111}(\sigma) = E\zeta_1\zeta_2\zeta_3 = E(\xi_1\xi_2\xi_3-3\xi_1\xi_2^2-\xi_1^2\xi_3+5\xi_1^2\xi_2-2\xi_1^3)/\sigma^6$$
$$= [\mu_{111}(\theta)-3\mu_{120}(\theta)-\mu_{201}(\theta)+5\mu_{210}(\theta)-2\mu_{300}(\theta)]/\sigma^6,$$

$$\text{(S)}\quad \mu_{030}(\sigma) = E\zeta_2^3 = E(\xi_2^3-3\xi_1\xi_2^2+3\xi_1^2\xi_2-\xi_1^3)/\sigma^6$$
$$= [\mu_{030}(\theta)-3\mu_{120}(\theta)+3\mu_{210}(\theta)-\mu_{300}(\theta)]/\sigma^6,$$

$$\text{(T)}\quad \mu_{002}(\sigma) = E\zeta_3^2 = E(\xi_3^2+9\xi_2^2+4\xi_1^2-6\xi_2\xi_3-12\xi_1\xi_2+4\xi_1\xi_3)/\sigma^6$$
$$= [\mu_{002}(\theta)+9\mu_{020}(\theta)+4\mu_{200}(\theta)-6\mu_{011}(\theta)-12\mu_{110}(\theta)+4\mu_{101}(\theta)]/\sigma^6,$$

$$\text{(U)}\quad \mu_{2001}(\sigma) = E\zeta_1^2\zeta_4 = E(\xi_1^2\xi_4-6\xi_1^2\xi_3+11\xi_1^2\xi_2-6\xi_1^3)/\sigma^6$$
$$=[\mu_{2001}(\theta)-6\mu_{201}(\theta)+11\mu_{210}(\theta)-6\mu_{300}(\theta)]/\sigma^6,$$

$$\text{(V)}\quad \mu_{0101}(\sigma) = E\zeta_2\zeta_4 = E(\xi_2\xi_4-6\xi_2\xi_3+11\xi_2^2-\xi_1\xi_4+6\xi_1\xi_3-17\xi_1\xi_2+6\xi_1^2)/\sigma^6$$
$$= [\mu_{0101}(\theta)-6\mu_{011}(\theta)+11\mu_{02}(\theta)-\mu_{1001}(\theta)+6\mu_{101}(\theta)-17\mu_{110}(\theta)+6\mu_{200}(\theta)]/\sigma^6,$$

$$\text{(W)}\quad \mu_{10001}(\sigma) = E\zeta_1\zeta_5 = E(\xi_1\xi_5-10\xi_1\xi_4+35\xi_1\xi_3-50\xi_1\xi_2+24\xi_1^2)/\sigma^6$$
$$= [\mu_{10001}(\theta)-10\mu_{1001}(\theta)+35\mu_{101}(\theta)-50\mu_{110}(\theta)+24\mu_{200}(\theta)]/\sigma^6.$$

Then we also have

$$I(\sigma) = \mu_{200}(\sigma) = \mu_{200}(\theta)/\sigma^2 = I(\theta)/\sigma^2, \qquad (3.33)$$

and

$$J(\sigma) = \mu_{110}(\sigma) = \mu_{300}(\sigma) \qquad (3.34)$$
$$= [\mu_{110}(\theta)-\mu_{200}(\theta)]/\sigma^3-\mu_{300}(\theta)/\sigma^3 = [J(\theta)-I(\theta)]/\sigma^3.$$

The results of the lemma can thus be obtained by inserting these equations in (3.32) into the definitions, (2.36) to (2.39).Δ

Some brief comments concerning Lemma 3.2 can be stated. First, we see that the statistical curvature is invariant under the logarithmic transformation [as established by Efron (1975)]. Also, $\omega_\theta^2, \gamma_\theta^2, \delta_\theta, \lambda_\theta^3, \nu_\theta^3$ are all translation invariant and all of the rate coefficients (for θ and σ) are scale invariant since each has been normalized and is unitless. Hence the rate coefficients for θ are invariant under any linear transformation. Furthermore, we can write γ_θ^2, from the definition, as

$$\begin{aligned} \gamma_\theta^2 &= (V_\theta \dot{\ell} \cdot V_\theta \ddot{\ell} - Cov_\theta^2(\dot{\ell}, \ddot{\ell}))/I^3(\theta) \\ &= [1 - \{Corr(\dot{\ell}, \ddot{\ell})\}^2] CV^2(\ddot{\ell}), \end{aligned} \tag{3.35}$$

where the coefficient of variation for $\ddot{\ell}$ is

$$CV(\ddot{\ell}) = \sqrt{V_\theta \ddot{\ell}} \,/ -I(\theta) = -(\eta_{020})^{\frac{1}{2}}/\eta_{010}\,. \tag{3.36}$$

Let us now consider the statistical curvature. It follows directly from (3.35) that γ_θ^2 is invariant under linear transformations and also that γ_θ^2 must be non-negative and equal to zero if and only if $\ddot{\ell}$ is a constant, or linear in $\dot{\ell}$ (i.e. $\ddot{\ell} = a\dot{\ell} + b$, where

a and b are constant in X). This 'linearity' in X is satisfied if and only if the underlying distribution is from the exponential family,

$$f(x|\theta) = \exp[A(\theta)B(x) + C(x) + D(\theta)]. \qquad (3.37)$$

In a similar manner, we have defined $J(\theta)$ as the covariance between $\dot{\ell}$ and $\ddot{\ell}$ and hence $J(\sigma) = 0$ if $\ddot{\ell}$ is constant in Y, which means that the underlying distribution is from the exponential scale family [as in (3.37), but with $A(\theta)=\theta$]. Then from (3.34), it follows that $J(\theta) = I(\theta)$ for the corresponding translation dual, hence $J(\theta)$ is a kind of measure of 'scale-information' in this instance. Also, since $J(\theta)$ is defined in terms of the odd Fisher moments, $J(\theta)=0$ for any symmetric translation family and, again by (3.34), $J(\sigma) = -I(\sigma)/\sigma$ for the scale dual of this family. Let us now examine the behaviour of these various parameters under a variety of regular distributions. First we consider the logistic distribution, an example of non-optimality in which the mle is shown to be no better than second-order efficient in the sense of Rao.

3.3 Logistic Distribution

To illustrate a case of non-optimality, let us consider the logistic distribution,

$$f(x|\theta) = \tfrac{1}{4}\mathrm{sech}^2[\tfrac{x-\theta}{2}] = \frac{e^{-(x-\theta)}}{[1+e^{-(x-\theta)}]^2}, \qquad (3.38)$$

$$-\infty<x<\infty, \quad -\infty<\theta<\infty .$$

This distribution is not in the exponential family, but it does have a scale dual of the form,

$$g(y|\sigma) = \sigma/(\sigma + y)^2, \qquad y>0, \ \sigma>0. \qquad (3.39)$$

In this case the Bahadur bound and the rate of the mle are well known:

$$B(\theta,\varepsilon) = \varepsilon - 2 + 2\varepsilon/(e^{\varepsilon} - 1), \qquad (3.40)$$

$$b(\hat{\theta},\theta,\varepsilon) = -\log \inf_{t\geq 0} e^{-t}\int_{-\infty}^{\infty} \exp[2t/(1+e^{-x+\theta+\varepsilon})]dF(x|\theta). \ (3.41)$$

The mle $\hat{\theta}_n$ is the unique solution of

$$n = \sum_{i=1}^{n} 2/\left\{1 + e^{-(x_i + \hat{\theta}_n)}\right\}. \qquad (3.42)$$

By symmetry, $t=\tau_\theta(\varepsilon)=\tfrac{1}{2}$ and $\mu=0$ and so the rate of pre, which is the optimal rate for the class of translation invariant estimators is given by

$$
\begin{aligned}
b_T(\theta,\varepsilon) &= -\log \int_{-\infty}^{\infty} \sqrt{f(x|\theta-\varepsilon)f(x|\theta+\varepsilon)}\,dx \\
&= -\log \tfrac{1}{4}\int_{-\infty}^{\infty} \operatorname{sech}\left(\frac{x-\theta+\varepsilon}{2}\right)\operatorname{sech}\left(\frac{x-\theta-\varepsilon}{2}\right)dx \\
&= -\log \tfrac{1}{4}\int_{-\infty}^{\infty} \left\{\frac{\tanh\left(\frac{x-\theta+\varepsilon}{2}\right)-\tanh\left(\frac{x-\theta-\varepsilon}{2}\right)}{\sinh\varepsilon}\right\}dx \\
&= -\log \left.\frac{2\log\left[\cosh\left(\frac{x-\theta+\varepsilon}{2}\right)/\cosh\left(\frac{x-\theta-\varepsilon}{2}\right)\right]}{4\sinh\varepsilon}\right|_{-\infty}^{\infty} \\
&= -\log[(\varepsilon-(-\varepsilon))/2\sinh\varepsilon] \\
&= \log[\sinh\varepsilon/\varepsilon], \qquad (3.43)
\end{aligned}
$$

by making use of (4.5.45) and (4.5.79) of Abramowitz & Stegun (1964). Direct comparison of these rates is not possible, particularly as the rate of the mle (3.41) is not known in closed form and can only be obtained by numerical integration, as has been done by Sievers (1978). Even graphs of these rates show very little difference between the three curves, which may not be too unusual when we consider that the logistic is very nearly Normal, except in the tails, and hence we might expect 'near'-Optimality [Sievers' graph of the ratio of the rate of the mle to the optimal rate is almost constant at one].

Hence in order to make any meaningful comparisons, especially when ε is small, we should examine the Taylor series expansions, which we can obtain from Lemma 2.3 and Theorems 2.2 and 2.3, which give the following series:

$$B(\theta,\varepsilon) = \varepsilon^2/6 - \varepsilon^4/360 + \varepsilon^6/15120 + o(\varepsilon^6), \tag{3.44}$$

$$b_T(\theta,\varepsilon) = \varepsilon^2/6 - \varepsilon^4/180 + \varepsilon^6/2835 + o(\varepsilon^6), \tag{3.45}$$

$$b(\hat{\theta},\theta,\varepsilon) = \varepsilon^2/6 - \varepsilon^4/180 + 31\varepsilon^6/113400 + o(\varepsilon^6). \tag{3.46}$$

The expansions for the Bahadur bound and the optimal translation invariant rate can also be obtained by direct expansion of (3.40) and (3.43). Hence, from the above, we have the strict inequality for ε near 0,

$$b(\hat{\theta},\theta,\varepsilon) < b_T(\theta,\varepsilon) < B(\theta,\varepsilon). \tag{3.47}$$

Thus the mle is second-order efficient but not third-order efficient since it is dominated by the pre in the sixth-degree term. Also, the pre is optimal among the class of translation invariant estimators consistent for θ, but it may not be optimal among the class of all consistent estimators since its rate does not attain the Bahadur bound. Indeed, from the second-order efficiency of the mle and pre, it would follow that there can be no estimator-sequence which will attain the bound since the fourth-degree term can be no smaller than that of the mle and pre, and hence the Bahadur bound is too large in this case. Thus the probability ratio estimators may be optimal for the class of all consistent estimators as we have not been able to find any estimator-sequence with a larger exponential rate.

The rate coefficients for the logistic distribution and its scale dual are given in Table 3.1 following.

TABLE 3.1: Rate Coefficients

Co-efficient	ω^2	γ^2	δ	ν^3	λ^3
Logistic	0	1/5	0*	27/175	-13/35
Scale Dual	3	1/5	3/5	1917/175	92/35

Here the Fisher information number is $I(\theta) = 1/3$. Since the logistic is symmetric, the coefficient δ_θ is not defined and is thus given a value of 0 (*).

In this case, we also decided to compute the exponential rates and the Bahadur bound directly from the definitions by numerical integration and optimization, as is done for the mixture of Normals in Chapter Four. The computer programs in §6.2 were modified for the logistic distribution and simplified somewhat (due to the symmetry). We then numerically computed the rates for a variety of values of ε and produced the graph in Figure 3.2. We also computed the Bahadur Divergence Function of degree 2,3,4,5, and 6, for the pre, mle, and the mle relative to the pre, which is in this case optimal for the class of translation invariant consistent estimators, hence we label "ML/T".

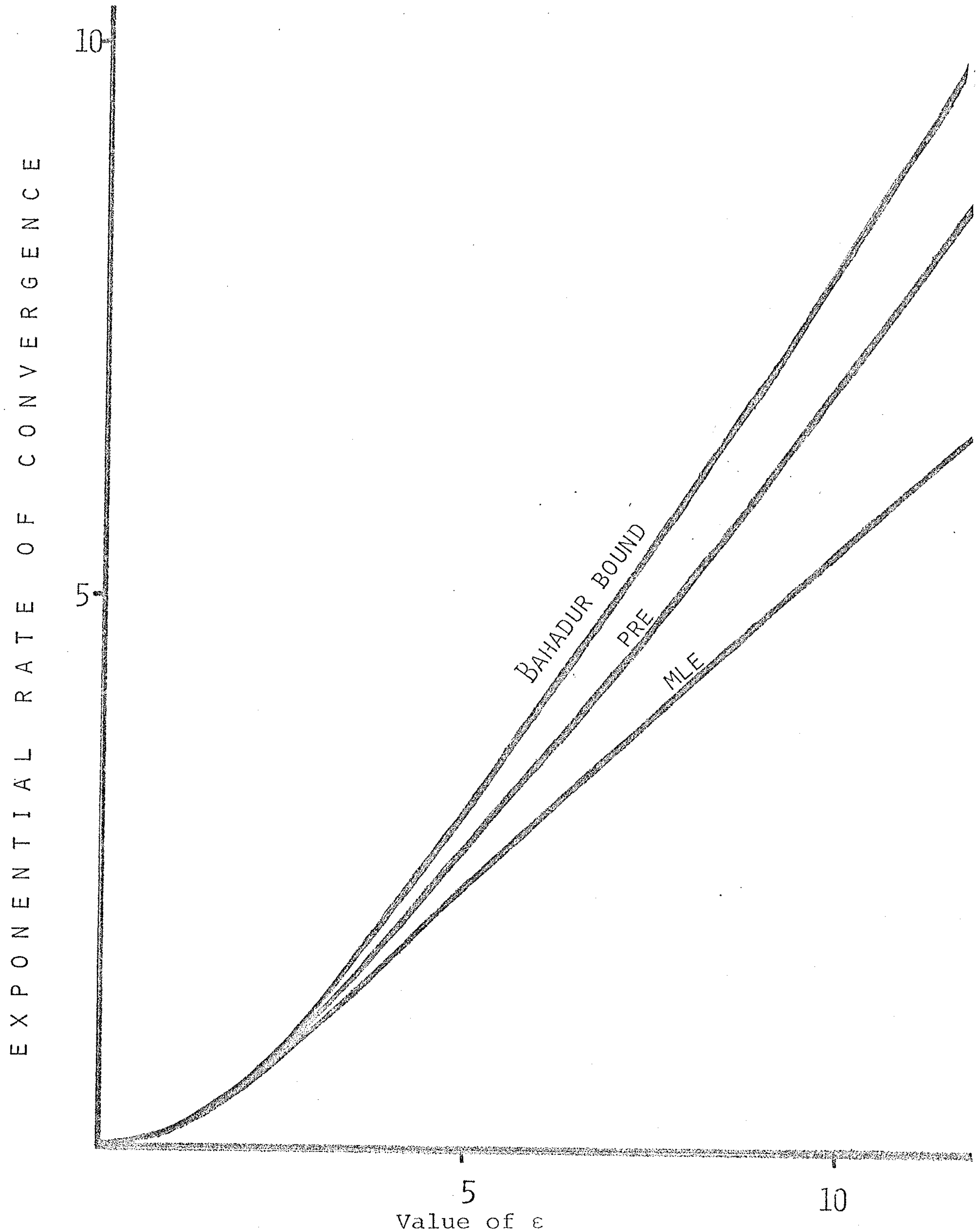

FIGURE 3.2: Exponential Rates of Convergence of pre and mle and Bahadur bound for the Logistic.

The Divergence Functions are graphed in Figures 3.3 through 3.7. For each graph a logarithmic spacing is used for ε in order to spread out the graphs and bring out detail. Figure 3.3 has the Divergence Functions of degree 2, which shows that all three graphs tend to zero as ε tends to zero, which is characteristic of all first-order efficient estimators. Also we note that the mle has a much higher graph than the pre, which would indicate that the pre appears to be more efficient, especially for ε in the range of about 1 to 100. The Divergence Functions of degree 3 (Figure 3.4) have similar graphs except that the difference between the mle and pre is less pronounced. This is to be expected since for the symmetric logistic, the odd-degree terms in any expansion will vanish, and hence play no role in determining efficiencies. Since both the mle and pre are second-order efficient, their fourth-degree Divergence Functions should tend to a constant value, $D_4 = I^2(\theta)\gamma_\theta^2/8 = 0.002778$, which can be seen in Figure 3.5. In this case, the divergence function of the mle relative to the pre tends to zero as ε tends to zero, which indicates that the mle has fourth-degree contact in exponential rate, near zero, with the class of translation invariant consistent estimators for θ.

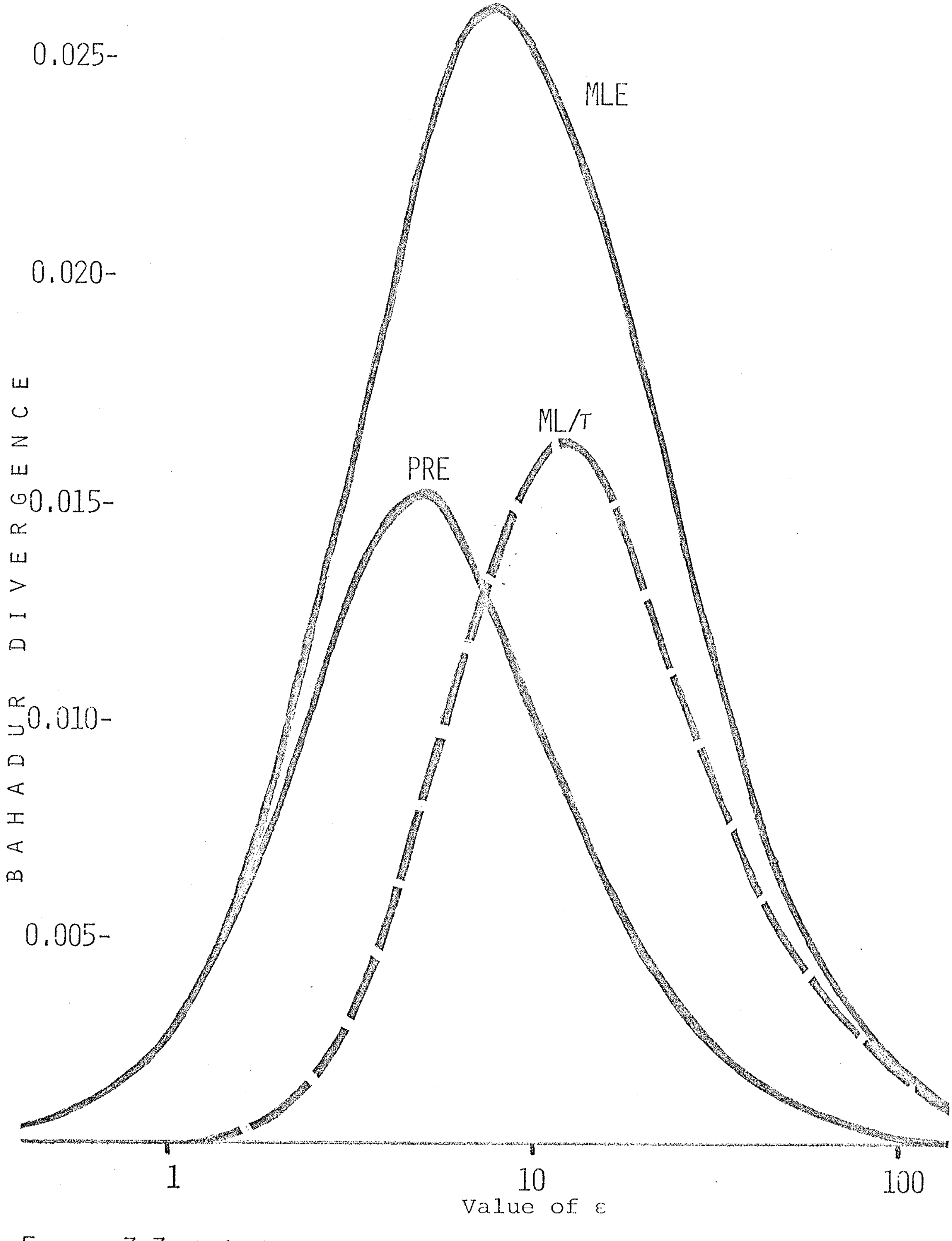

FIGURE 3.3: Bahadur Divergence of degree 2 for Logistic.

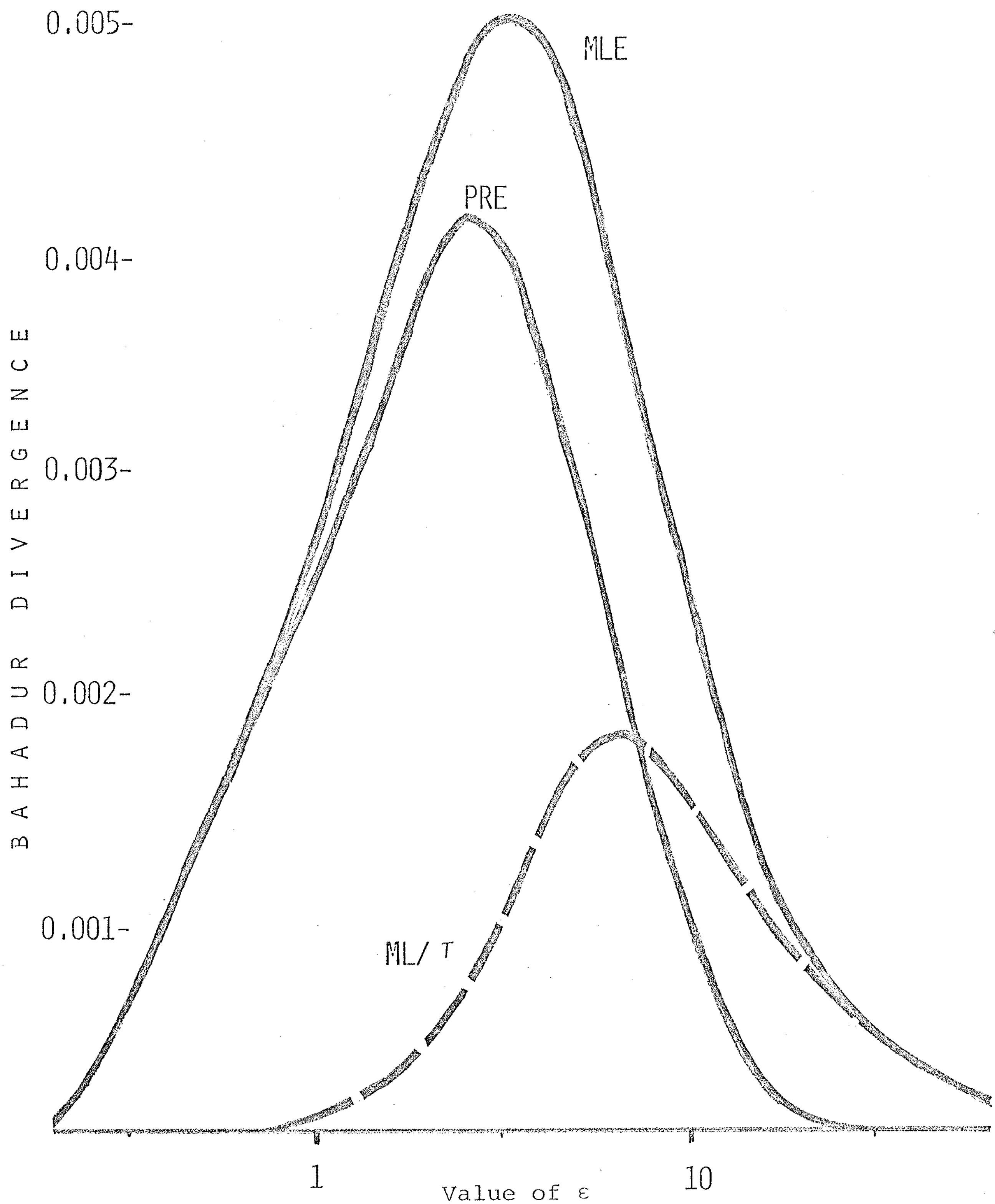

FIGURE 3.4: Bahadur Divergence of degree 3 for Logistic.

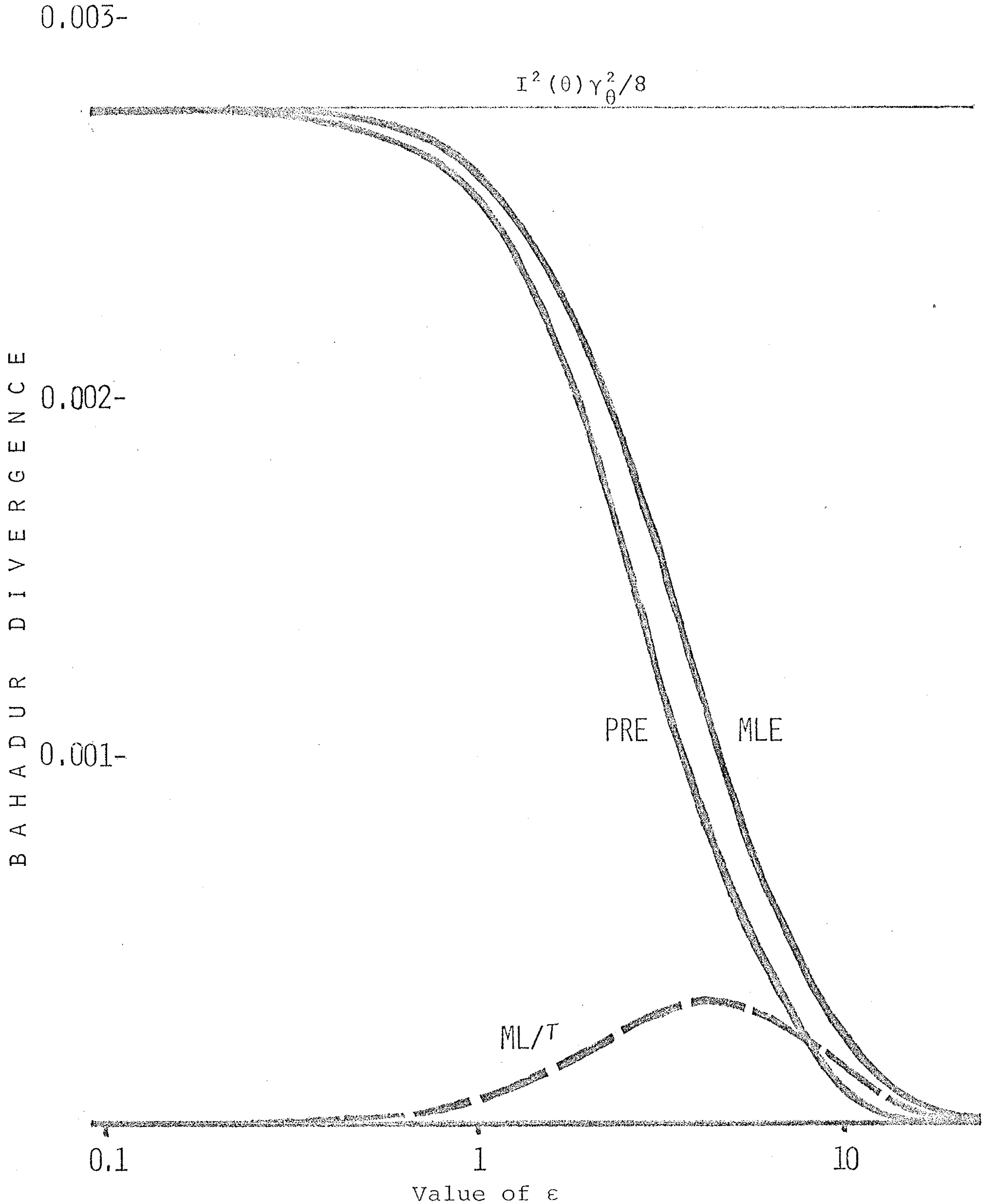

FIGURE 3.5: Bahadur Divergence of degree 4 for Logistic.

The Divergence Functions of degree 5 in Figure 3.6, are similar but with the graphs for the mle and pre now tending to ∞, as would be expected from (1.37). Here we see that the mle also has fifth degree contact with T. However in Figure 3.7, we see that the mle is divergent in degree 6 from the class of translation invariant consistent estimators, where $D_6(ML/T) = I^3(\theta)\nu_0^3/12 = 7.9365(10)^{-5}$. In this way it is felt that the mle would be at least third-order efficient amoung the class of translation invariant consistent estimators of θ(in this instance), although possibly not among all estimators. As we stated, third order efficiency is as yet undefined, although we infer that any definition consistent with Rao (1961) should somehow relate to the Bahadur Divergence of degree 6, just as the first- and second-order efficiencies relate to the Bahadur Divergence of degrees 2 and 4 respectively. Again, the odd-degree Divergence Functions are discounted as they will tend to be largely influenced by the skewness of the underlying distribution, and are thus of little use in determining the relative efficiencies.

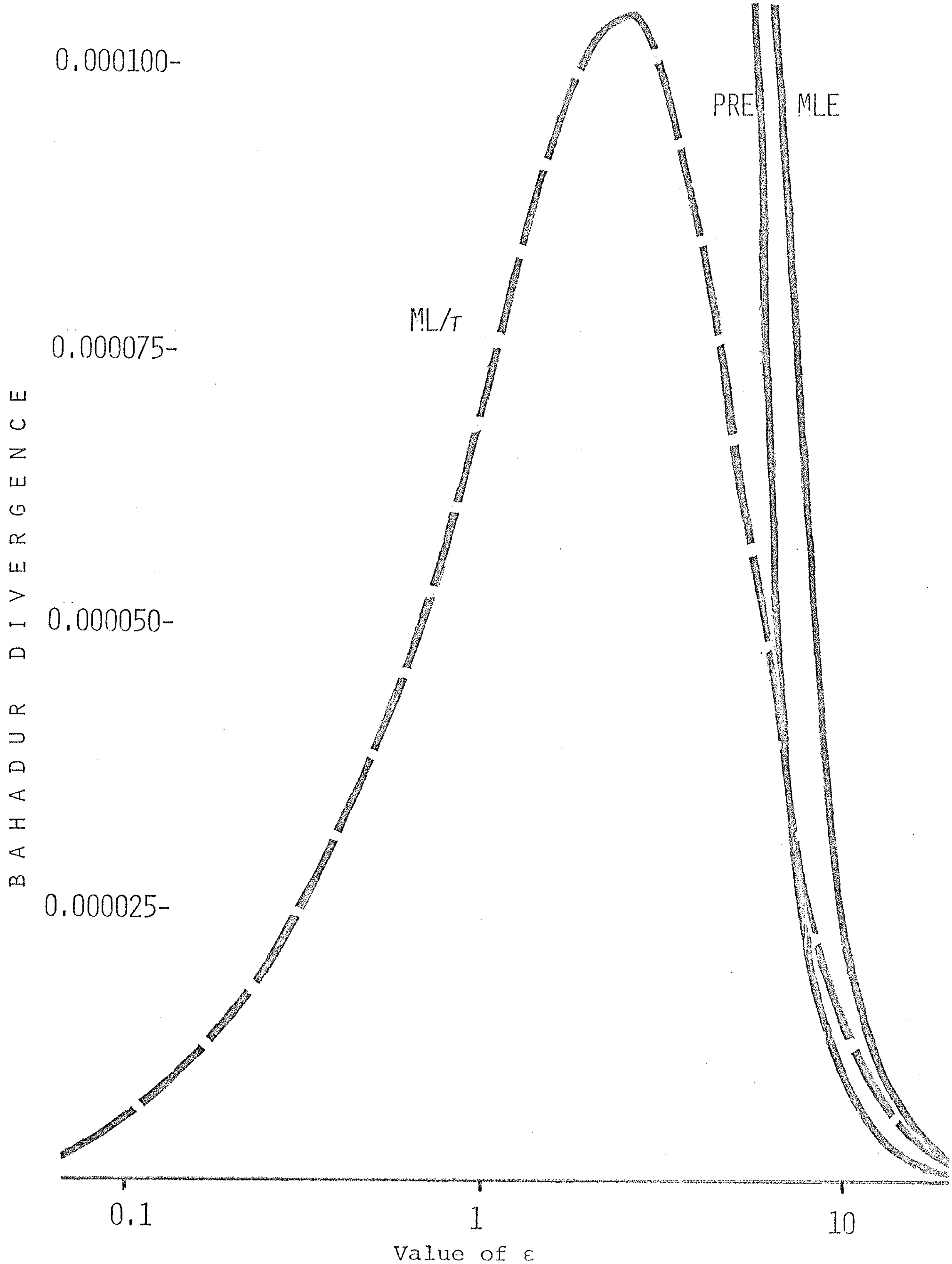

FIGURE 3.6: Bahadur Divergence of degree 5 for Logistic.

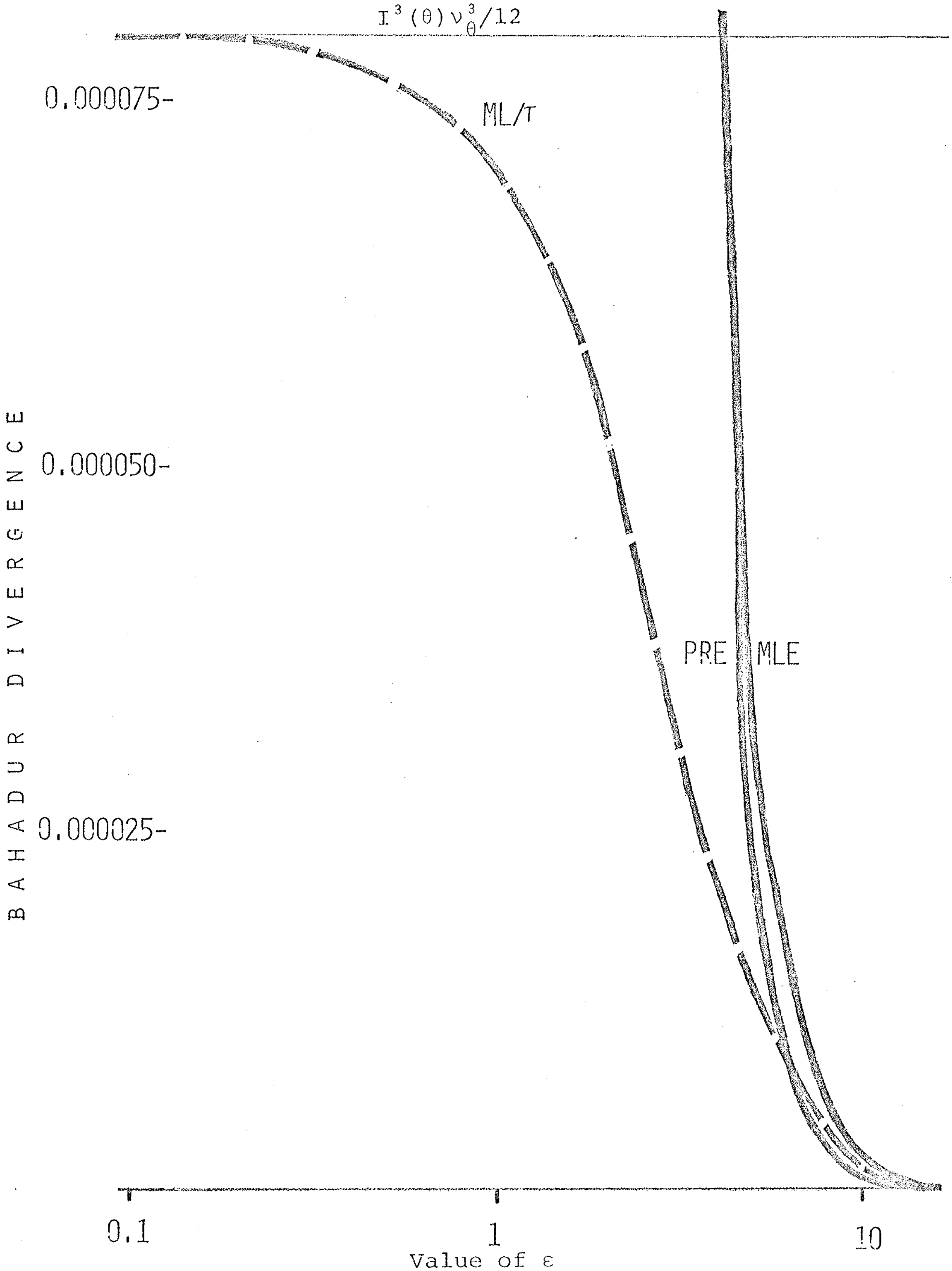

FIGURE 3.7: Bahadur Divergence of degree 6 for Logistic.

3.4 Other Regular Distributions

3.4.1 General Exponential Family

Consider the general exponential family of (3.2), for which $I(\theta)=qp^2$ and $J(\theta)=-qp^3=-pI(\theta)$. Here, p is a pure scale parameter and q is a pure shape parameter. However, the normalized coefficient $\omega_\theta^2=J^2(\theta)/I^3(\theta)=1/q$ is a pure shape measure since it depends only on q. Indeed, the coefficient of skewness for this population is $\beta_1(\theta)=\ddot{\psi}^2(q)/\dot{\psi}^3(q)$, which tends to $1/q=\omega_\theta^2$ as q becomes large (where $\psi(x)$ is the digamma function: $\psi(x)=\partial/\partial x \log \Gamma(x)$).

3.4.2 Hyperbolic Secant Distribution

To make use of the Location-Scale Duality to find suitably smooth regular distributions, we tried the Hyperbolic Secant Distribution, where

$$h(t) = e^{-\frac{1}{2}t} \operatorname{sech} \tfrac{1}{2}t \;/\; \log 4 , \tag{3.48}$$

and, hence, we get the regular distribution:

$$f(x|\theta) = \exp(x-\theta-e^{x-\theta})/\left\{\log 2\cdot[1+e^{-e^{x-\theta}}]\right\}. \tag{3.49}$$

as graphed in Figure 3.8. The exponential rates can not be computed directly in this case, but they can be evaluated for various ε by numerical integration. However, the differences between the rates appear to be no larger than the integration bias, and so, the

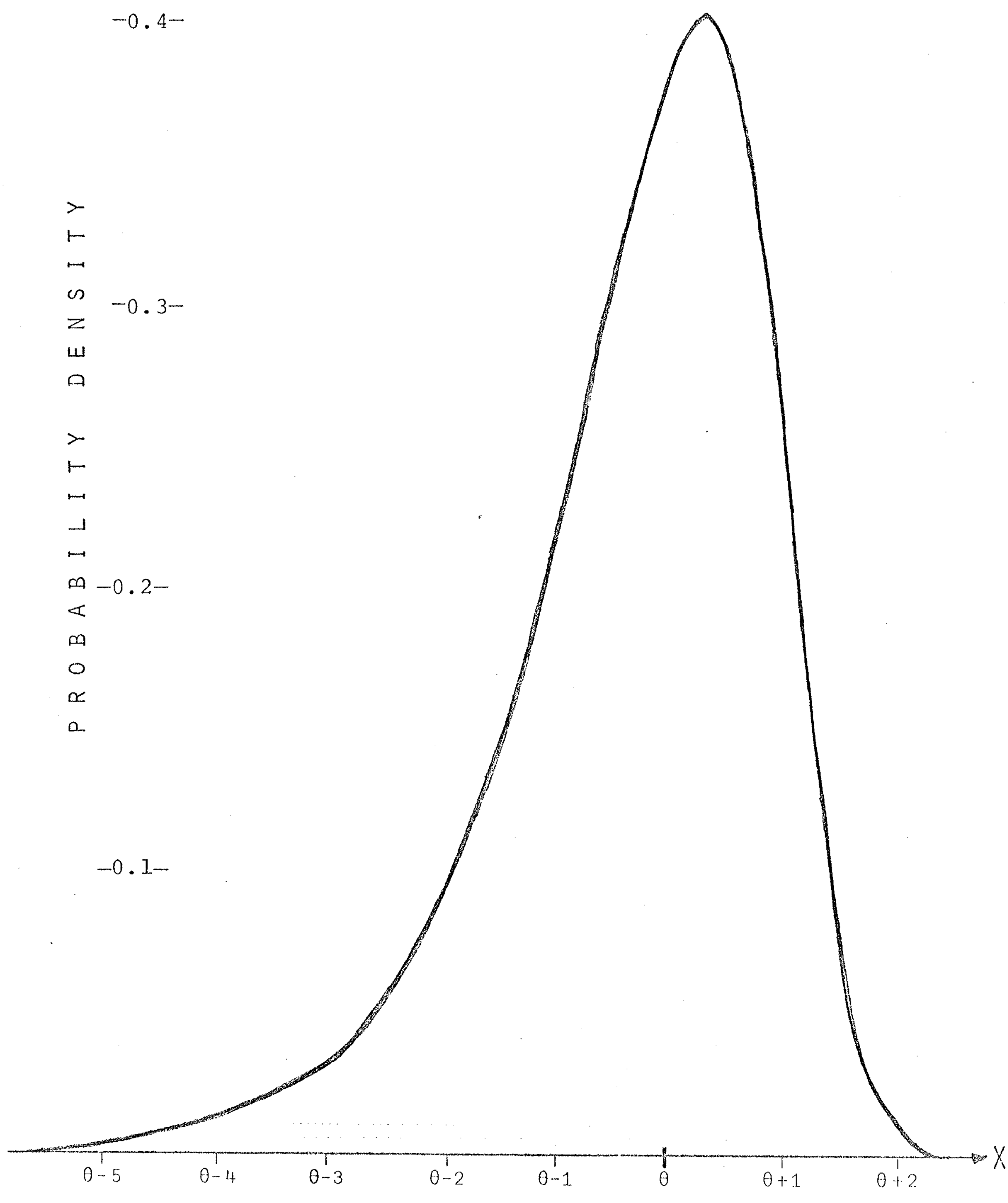

FIGURE 3.8: $$f(x|\theta) = \frac{\exp\{(x - \theta) - e^{(x - \theta)}\}}{\{\log 2\cdot(1 + e^{-e^{(x - \theta)}})\}}$$

rates are almost all the same (i.e. the mle and pre are almost optimal). This near-optimality is mainly due to the extremely small statistical curvature,

$$\begin{aligned}\gamma_\theta^2 &= [6\kappa_1\kappa_2+4\kappa_1\kappa_3-12\kappa_1^2(\kappa_1+1)-3\kappa_2^2]/12\kappa_1^3 \\ &= .009228601,\end{aligned} \tag{3.50}$$

where $\kappa_i = i!\eta(i+1)/\log 2$, and $\eta(j)$ is related to the Reimann Zeta Function [see pages 807-811 of Abramowitz & Stegun (1964)]. Computation of the Fisher moments is given in Appendix 6.1. Here the Fisher information is $I(\theta) = \kappa_1 = 1.18656911$ and $J(\theta) = \frac{1}{2}\kappa_2 = 1.3006511$ and so $\omega_\theta^2 = 4.0504431$.

We also generated 1000 samples of various sizes in order to estimate the rates of the mle and pre. Again, the two rates were almost identical in every case, and a plot of this common value of the rate of the mle and pre, suitably normalized by the value of ε, is given in Figure 3.9. As can be seen, very large sample sizes are required before the rates are within range of their 'theoretical' limits.

As direct computation of the exponential rates is not possible and the numerical integration and the Monte Carlo study suggest that the rates are all nearly equal, due to the extremely small statistical curvature,

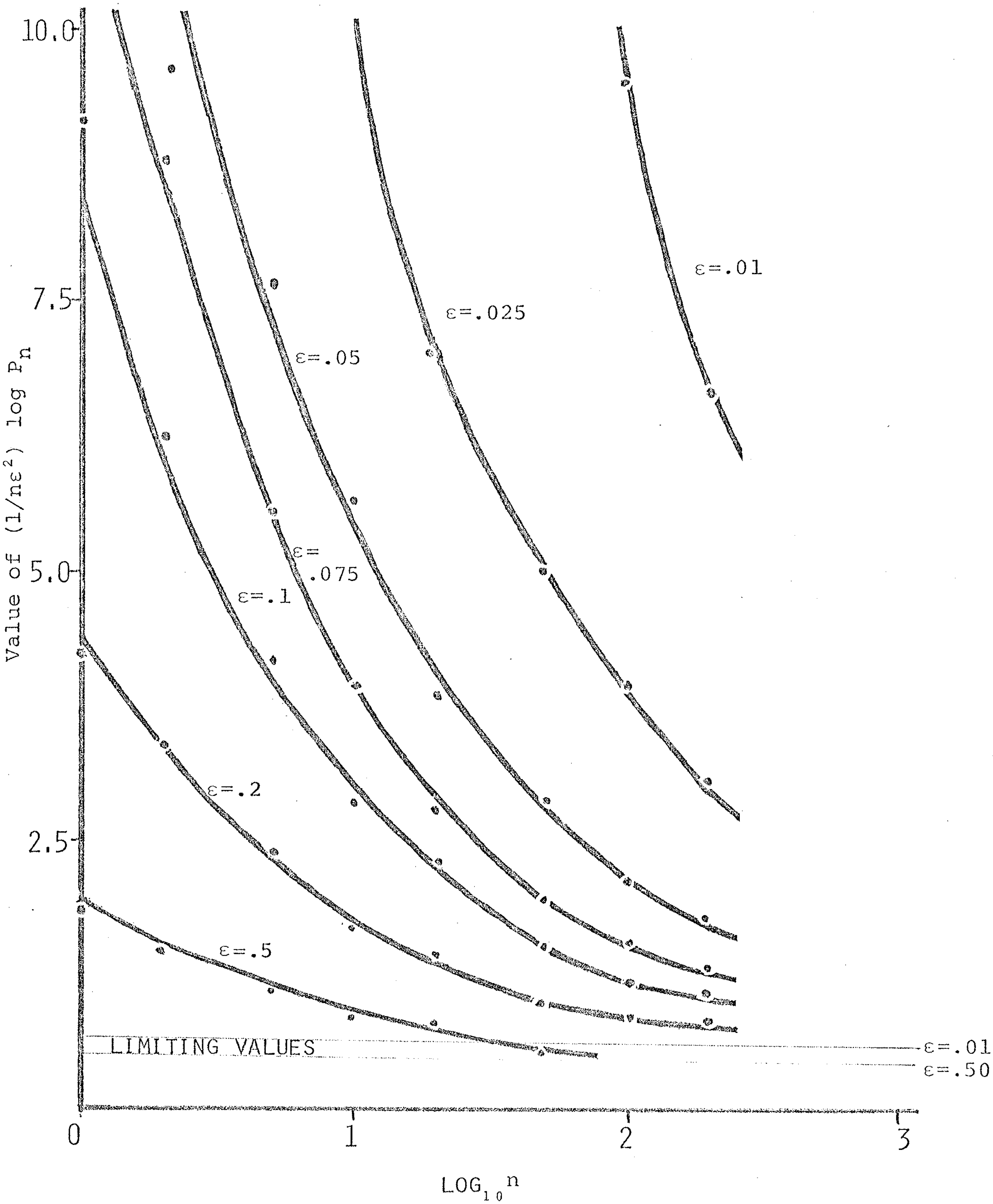

FIGURE 3.9: Estimated Rates of Convergence based on the generation of 1000 samples from the distribution in Figure 3.8.

then this leaves us with only the methods of §2.2-§2.4 to investigate the relationships between the rates. In appendix 6.1.4 we compute the Fisher moments as given in the following table:

TABLE 3.3: Fisher moments of the Hyperbolic Secant Distribution

$\mu_{200} = \kappa_1 = 1.18657$	$\mu_{500} = (10\kappa_3+3\kappa_4)/3 = 60.98493$
$\mu_{300} = -\kappa_2 = -2.60130$	$\mu_{310} = (10\kappa_3+3\kappa_4)/4 = 45.73870$
$\mu_{110} = -\frac{1}{2}\kappa_2 = -1.30065$	$\mu_{201} = (12\kappa_2+6\kappa_3+\kappa_4)/6 = 19.01017$
$\mu_{400} = (3\kappa_2+2\kappa_3)/2 = 12.09963$	$\mu_{120} = (-3\kappa_2+6\kappa_3+2\kappa_4)/3 = 36.23361$
$\mu_{210} = (3\kappa_2+2\kappa_3)/3 = 8.06642$	$\mu_{011} = (-3\kappa_2+6\kappa_3+2\kappa_4)/6 = 18.11680$
$\mu_{101} = \kappa_1 = 1.18657$	$\mu_{1001} = (15\kappa_2-\kappa_4)/6 = 0.89337$
$\mu_{020} = (-3\kappa_1+3\kappa_2+2\kappa_3)/3 = 6.87983$	

Then we obtain the fifth order rate coefficient,

$$\delta_\theta = -(6\kappa_1\kappa_2 + 2\kappa_3 + \kappa_4)/6\kappa_1\kappa_2 = -3.09841, \quad (3.51)$$

and the Bahadur bound coefficients, $b_i = -E_\theta \ell^{(i)}$,

$$\begin{aligned}
&(A)\ b_2 = \mu_{200} = \kappa_1 = 1.18657,\\
&(B)\ b_3 = -2\mu_{300}+3\mu_{110} = \tfrac{1}{2}\kappa_2 = 1.30065,\\
&(C)\ b_4 = 6\mu_{400}-12\mu_{210}+4\mu_{101}+3\mu_{020} = \kappa_1 = 1.18657,\\
&(D)\ b_5 = -24\mu_{500}+60\mu_{310}-20\mu_{201}-30\mu_{120}+10\mu_{011}+5\mu_{1001}\\
&\qquad = -(15\kappa_2 - \kappa_4)/6 = -0.89337.
\end{aligned} \quad (3.52)$$

Then since b_3 is positive, the 'minimums' in the rate expressions occur for $-\varepsilon$, hence,

$$B(\theta,\varepsilon) = 0.59328\varepsilon^2-0.21678\varepsilon^3+0.04944\varepsilon^4+0.00744\varepsilon^5+o(\varepsilon^5), \quad (3.53)$$

$$b(\tilde{\theta},\theta,\varepsilon) = 0.59328\varepsilon^2-0.21678\varepsilon^3+0.04782\varepsilon^4+0.00744\varepsilon^5+o(\varepsilon^5), \quad (3.54)$$

$$b(\hat{\theta},\theta,\varepsilon) = 0.59328\varepsilon^2-0.21678\varepsilon^3+0.04782\varepsilon^4-0.39104\varepsilon^5+o(\varepsilon^5). \quad (3.55)$$

Hence we have, strictly,

$$B(\theta,\varepsilon) > b(\tilde{\theta},\theta,\varepsilon) > b(\hat{\theta},\theta,\varepsilon). \quad (3.56)$$

For a numerical comparison, Table 3.4 gives the values of the rates, as computed by these expansions, for selected (small) values of ε, with some indication of the order of the accuracy.

TABLE 3.4: Numerical Comparison of Rates

ε	$B(\theta,\varepsilon)$	$b(\tilde{\theta},\theta,\varepsilon)$	$b(\hat{\theta},\theta,\varepsilon)$	Error
0.50	$1.24545(10)^{-1}$	$1.24444(10)^{-1}$	$1.11991(10)^{-1}$	10^{-2}
0.20	$2.20784(10)^{-2}$	$2.20759(10)^{-2}$	$2.19483(10)^{-2}$	10^{-4}
0.10	$5.72104(10)^{-3}$	$5.72088(10)^{-3}$	$5.71689(10)^{-3}$	10^{-6}
0.05	$1.45641(10)^{-3}$	$1.45640(10)^{-3}$	$1.45628(10)^{-3}$	10^{-7}
0.02	$2.35586(10)^{-4}$	$2.35585(10)^{-4}$	$2.35584(10)^{-4}$	10^{-9}
0.01	$5.91117(10)^{-5}$	$5.91117(10)^{-5}$	$5.91117(10)^{-5}$	10^{-11}

3.4.3 One parameter Normal

Consider the Normal family where the standard deviation is some power of the mean: $N(\theta,\theta^{k})$, where

$$\ell=-\tfrac{1}{2}\log(2\pi\theta^{k})-x^{2}/2\theta^{k}+x/\theta^{k-1}-\tfrac{1}{2}/\theta^{k-2}. \tag{3.57}$$

Then, this distribution is in the exponential family only when $k=0$ (location, symmetric) or $k=1$ (scale). For other values of k the distribution is not of the exponential family and θ is neither a pure location nor a pure scale parameter. Then

$$I(\theta) = 1/\theta^{k} + k^{2}/2\theta^{2}, \tag{3.58}$$

$$J(\theta) = -2k/\theta^{k+1} - k^{2}(k+1)/2\theta^{3}, \tag{3.59}$$

and,

$$\gamma_{\theta}^{2} = k^{2}(k-1)^{2}/[2I^{3}(\theta)\theta^{k+4}]. \tag{3.60}$$

Hence, as expected $\gamma_{\theta}^{2}=0$ only for the two exponential family members, when $k=0,1$. We also note that $J(\theta)=0$ only for the symmetric location family ($k=0$), while for $k=1$, the exponential scale family member, we get $J(\theta) = -2I(\theta)/\theta$. The multiplier '2' likely arises from the fact that the scale of X is in quadratic units.

The exponential scale distribution, $N(\theta,\theta)$, is not scale invariant,

$$h(x|\theta) = \frac{1}{\sqrt{2\pi\theta}} e^{-x^2/2\theta + x-\theta/2}, \qquad (3.61)$$

and has location dual, for $z = \log x$ and $\mu = \frac{1}{2}\log\theta$,

$$g(z|\mu) = \frac{1}{\sqrt{2\pi}} \exp[(z-\mu)-\tfrac{1}{2}(e^{z-\mu}-e^{\mu})^2], \qquad (3.62)$$

which is not translation invariant. Here, $I(\mu) = e^{2\mu} + \frac{1}{2} = -J(\mu)$ and $\omega_\mu^2 = 1/I(\mu)$. All of the rate coefficients, for both θ and μ, are zero for this non-regular distribution. We might also examine the scale invariant distribution for $k=2$, $N(\theta,\theta^2)$,

$$h(x|\theta) = \frac{1}{\sqrt{2\pi}\,\theta} e^{-x^2/2\theta^2+x/\theta-\frac{1}{2}}, \qquad (3.63)$$

which, however is not exponential family, but does have a location dual which is a regular distribution,

$$g(y|\mu) = \frac{1}{\sqrt{2\pi}} \exp[(y-\mu)-\tfrac{1}{2}(e^{y-\mu}-1)^2]. \qquad (3.64)$$

The rate coefficients for this regular distribution and its scale dual (3.63) are give in the table below.

TABLE 3.5: Rate Coefficients

Coefficient	ω^2	γ^2	δ	ν^3	λ^3
θ (Scale)	100/27	2/27	2/15	64/27	-40/729
μ (Location)	49/27	2/27	2/21	20/27	-94/729

CHAPTER FOUR
EXAMPLES UNDER NON-REGULAR DISTRIBUTIONS

4.1 Mixtures of Normals

In our search for distributions which are sufficiently 'smooth', but which are non-exponential and not symmetric, we decided on Normal mixtures. Unfortunately, it seems that all such mixtures will satisfy all of the regularity conditions except [R5], which is needed only for the rate of the mle. As it turns out, the rate of the mle cannot be obtained from Lemma 2.4, which we show in §4.1.2.

4.1.1 Symmetry about θ

Consider the mixture of the two Normals with common mean θ, but with scale σ and $r\sigma$ respectively,

$$\begin{aligned} g(x|\theta) &= N(\theta, \sigma^2), \\ h(x|\theta) &= N(\theta, r^2\sigma^2), \end{aligned} \tag{4.1}$$

and with mixture coefficient p, we have $MN_p(0, r)$,

$$f(x|\theta) = pg(x|\theta) + (1 - p)h(x|\theta). \tag{4.2}$$

Then, if we denote the maximum likelihood estimators of θ with respect to g, h and f as $\hat{\theta}_g$, $\hat{\theta}_h$ and $\hat{\theta}(=\hat{\theta}_f)$ respectively; then, we have, in this case,

$$\hat{\theta}_f = \hat{\theta}_g = \hat{\theta}_h = \bar{X}_n = \sum_{i=1}^{n} X_i/n. \tag{4.3}$$

Then, the moment-generating function of X is,

$$m_X(t) = pe^{\theta t + \frac{1}{2}\sigma^2 t^2} + (1-p)e^{\theta t + \frac{1}{2}r^2\sigma^2 t^2}. \tag{4.4}$$

Hence, the moment-generating function for $\bar{X}$ is

$$\begin{aligned} m_{\bar{X}}(t) &= \left[pe^{\theta t/n + \sigma^2 t^2/2n^2} + (1-p)e^{\theta t/n + r^2\sigma^2 t^2/2n^2}\right]^n \qquad (4.5) \\ &= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} e^{\theta t + [k+(n-k)r^2]\sigma^2 t^2/2n^2}. \end{aligned}$$

And so the distribution of the mle is Normal mixture,

$$\hat{\theta} \sim \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} N\left(\theta, \frac{k+(n-k)r^2}{n^2}\sigma^2\right). \tag{4.6}$$

Thus, we can obtain the tail probability directly,

$$P_n = P(|\hat{\theta}-\theta| \geq \varepsilon) = 2\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \Phi\left(\frac{-n\varepsilon}{\sqrt{k+(n-k)r^2}\,\sigma}\right). \tag{4.7}$$

We have applied (4.7) directly to approximate the exponential rate of the mle by computing $-\frac{1}{n}\log P_n$ for n=2,4,8,16,....,etc. until we obtain a convergent sequence. No approximations were made to (4.7), with the Normal tails computed by a partial fraction method [see Weiss (1981)]. Also, as these sequences were converging from above, a lower 'bound' for the rate was obtained by extrapolating the differences by a geometric series, which gave a new sequence which converges from below, and gives some idea of precision.

The computer program is given in appendix 6.2. As a numerical example, consider $p=\frac{1}{2}$ and $r=2$, which is graphed in Figure 4.1. For $\varepsilon=\frac{1}{2}$, for example, the convergent sequence, obtained by successive doubling of n is given below in Table 4.1. We will denote the 'estimate' of the exponential rate of the mle as $b_n(\hat{\theta}) = -\frac{1}{n} \log P_n$, and the geometric extrapolation by $b_n^*(\hat{\theta})$.

TABLE 4.1: Exponential Rates of Convergence

n	P_n	$b_n(\hat{\theta})$	$b_n^*(\hat{\theta})$
2	$6.28154(10)^{-1}$	.232485	—
4	$5.13255(10)^{-1}$	.166745	—
8	$3.64937(10)^{-1}$	.126004	.059602
16	$2.04146(10)^{-1}$	.099308	.048568
32	$7.37330(10)^{-2}$	.081478	.045628
64	$1.16081(10)^{-2}$	.069626	.046122
128	$3.62931(10)^{-4}$	.061885	.047313
256	$4.65276(10)^{-7}$	.056956	.048311
512	$1.03312(10)^{-12}$	.053903	.048939
1024	$7.02078(10)^{-24}$	.052064	.049274
2048	$4.52212(10)^{-46}$	.050981	.049435
4096	$2.63421(10)^{-90}$	.050357	.049508
8192	$1.25910(10)^{-178}$	.050003	.049540
16384	$4.06407(10)^{-355}$	.049806	.049554
32768	$5.97900(10)^{-708}$	.049696	.049560
65536	$1.82932(10)^{-1413}$	.049636	.049563

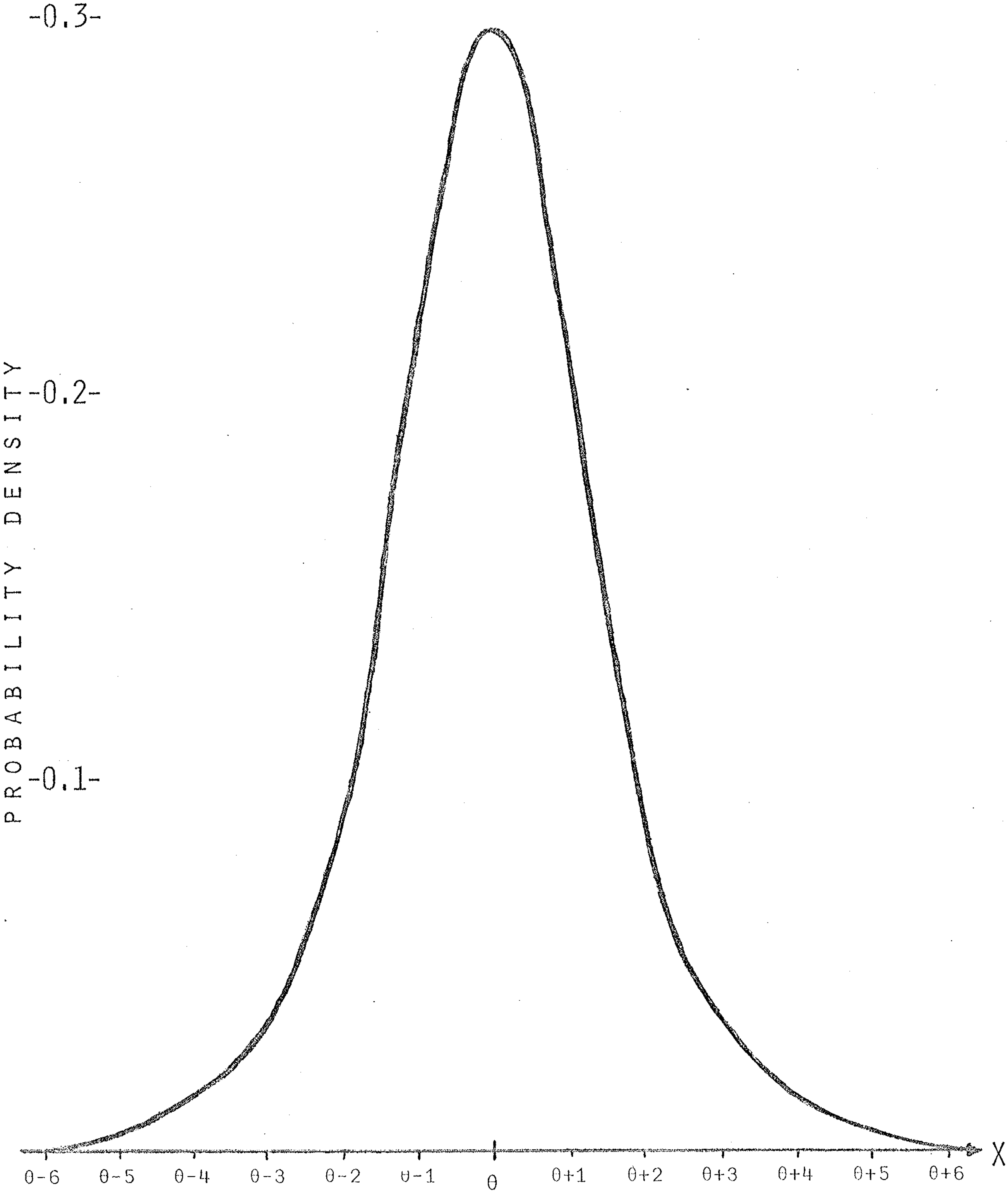

FIGURE 4.1: Probability density function of the mixture

$MN_{\frac{1}{2}}(0,2) = \frac{1}{2}N(0,1) + \frac{1}{2}N(0,2^2)$.

We see that $b_n(\hat{\theta})$ seems to be converging to some value between 0.049563 and 0.049636. Hence the exponential rate for the mle is approximately 0.0496, compared with a value of 0.057233 for the Bahadur bound and 0.056623 for the probability ratio estimator. Both values have been obtained through numerical integration of the rates as given by (2.42) and Theorem 2.1 (computer program for this also given in Appendix 6.2). A summary table of the rates for selected values of ε is given below in Table 4.2, and plotted in Figure 4.2.

TABLE 4.2: Comparison of Rates for $MN_{\frac{1}{2}}(0,2)$

ε	$b_n(\hat{\theta})$	$b_n^*(\hat{\theta})$	$b(\hat{\theta},\theta,\varepsilon)$	$b(\tilde{\theta},\theta,\varepsilon)$	$B(\theta,\varepsilon)$
0.1	.0026708	.0018609	.002	.002306	.002307
0.2	.0089653	.0078254	.008	.009202	.009218
0.5	.049636	.049563	.0496	.056623	.057233
1.0	.193821	.193665	.1937	.215247	.223677
2.0	.724378	.723205	.723	.750053	.822717
3.0	1.51370	1.51334	1.5134	1.52199	1.65945
4.0	2.52304	2.52267	2.5227	2.52603	2.66282
5.0	3.73898	3.73762	3.738	3.74127	3.84726
10.	13.19453	13.19304	13.193	13.19307	13.25945

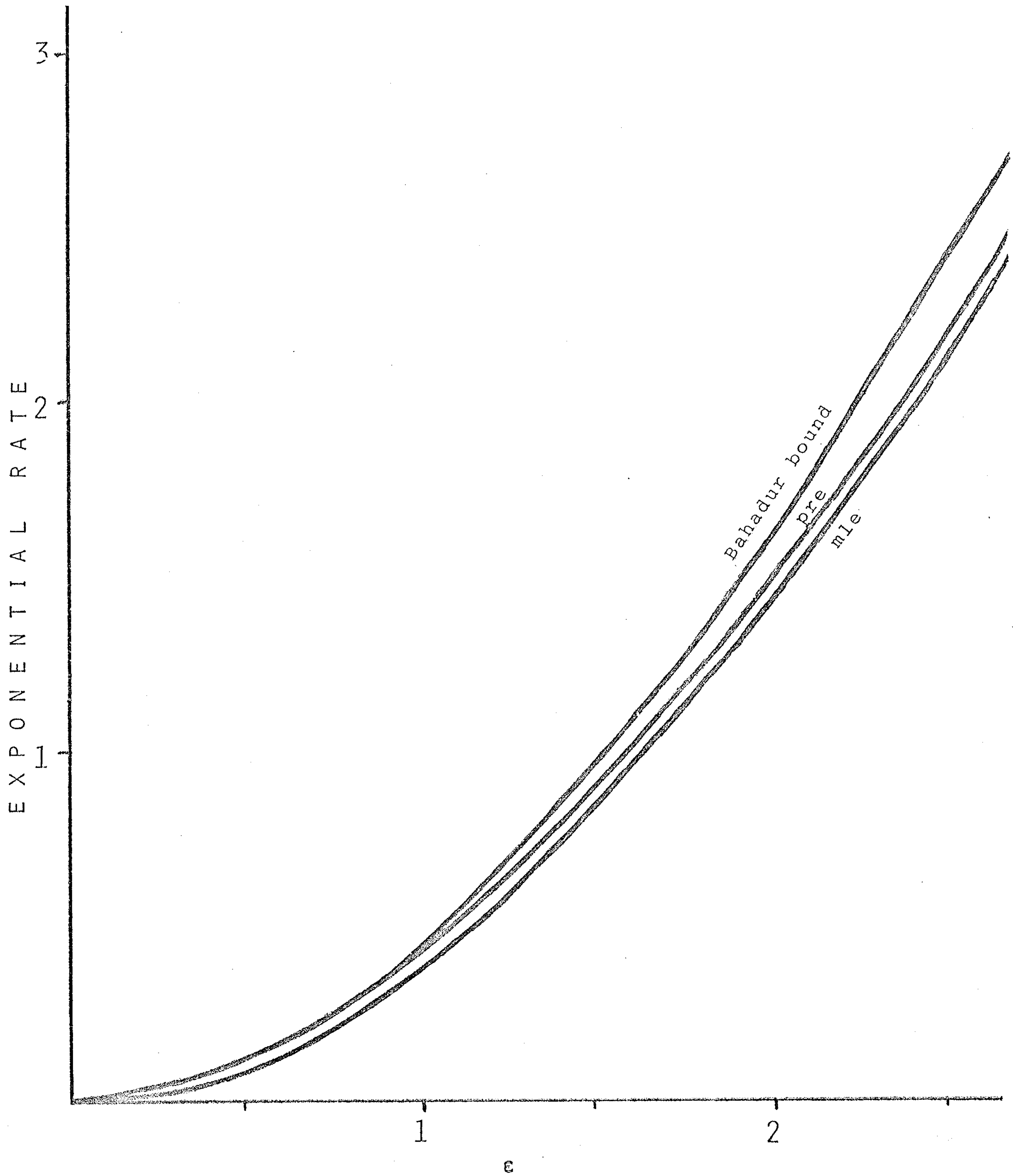

FIGURE 4.2: Approximate Exponential Rates for $MN_{\frac{1}{2}}(0,2)$.

We note that for large values of ε, the rate of the mle and pre are closer to each other than to the Bahadur bound. This is likely because both estimators are second-order efficient and not 'optimal' (in the sense that they do not attain the Bahadur bound). On the other hand, for small values of ε, the rate of the pre is closer to the Bahadur bound than to the rate of the mle, which would seem to indicate that the mle is not optimal within the class of translation invariant estimators (for which the pre is optimal).

4.1.2 Non-symmetry

Now let us consider a mixture in which the mean of the second distribution is shifted by a distance a,

$$\begin{aligned} g(x|\theta) &= N(\theta\ , \ \sigma^2)\ , \\ h(x|\theta) &= N(\theta+a,\ r^2\sigma^2)\ , \end{aligned} \tag{4.8}$$

and so the mixture, which we denote $MN_p(a,\ r)$ is

$$f(x|\theta) = pg(x|\theta) + (1-p)h(x|\theta). \tag{4.9}$$

Once again, we will denote the mle's of g, h and f as $\hat{\theta}_g$, $\hat{\theta}_h$ and $\hat{\theta}(=\hat{\theta}_f)$ respectively, where in this case

$$\hat{\theta}_g = \bar{X} = \textstyle\sum_i X_i/n\ , \tag{4.10}$$

$$\hat{\theta}_h = \bar{X} - a\ , \tag{4.11}$$

and,

$$\hat{\theta}_g > \hat{\theta}_f > \hat{\theta}_h . \tag{4.12}$$

Let us consider the unbiased sample mean estimator,

$$\theta^* = \bar{X} - (1-p)a . \tag{4.13}$$

Then, since the mle from each Normal is related to the sample mean, the mle for this mixture should, at least, be asymptotically equivalent, that is

$$\hat{\theta} - \theta^* \xrightarrow{p} 0. \tag{4.14}$$

so that, proceeding as in §4.1.1, the distribution of θ^* and hence of $\hat{\theta}_f$ is approximately

$$\hat{\theta} \sim \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} N\left(\theta + (p-\frac{k}{n})a,\ \frac{k+(n-k)r^2}{n^2}\sigma^2\right). \tag{4.15}$$

Thus, we can use (4.15) to compute the exponential rate of the mle, as before. As a numerical example, the mixture with $p=\frac{1}{2}$, $a=2$, and $r=2$, which is graphed in Figure 4.2, has the exponential rates as given in Table 4.3, which follows.

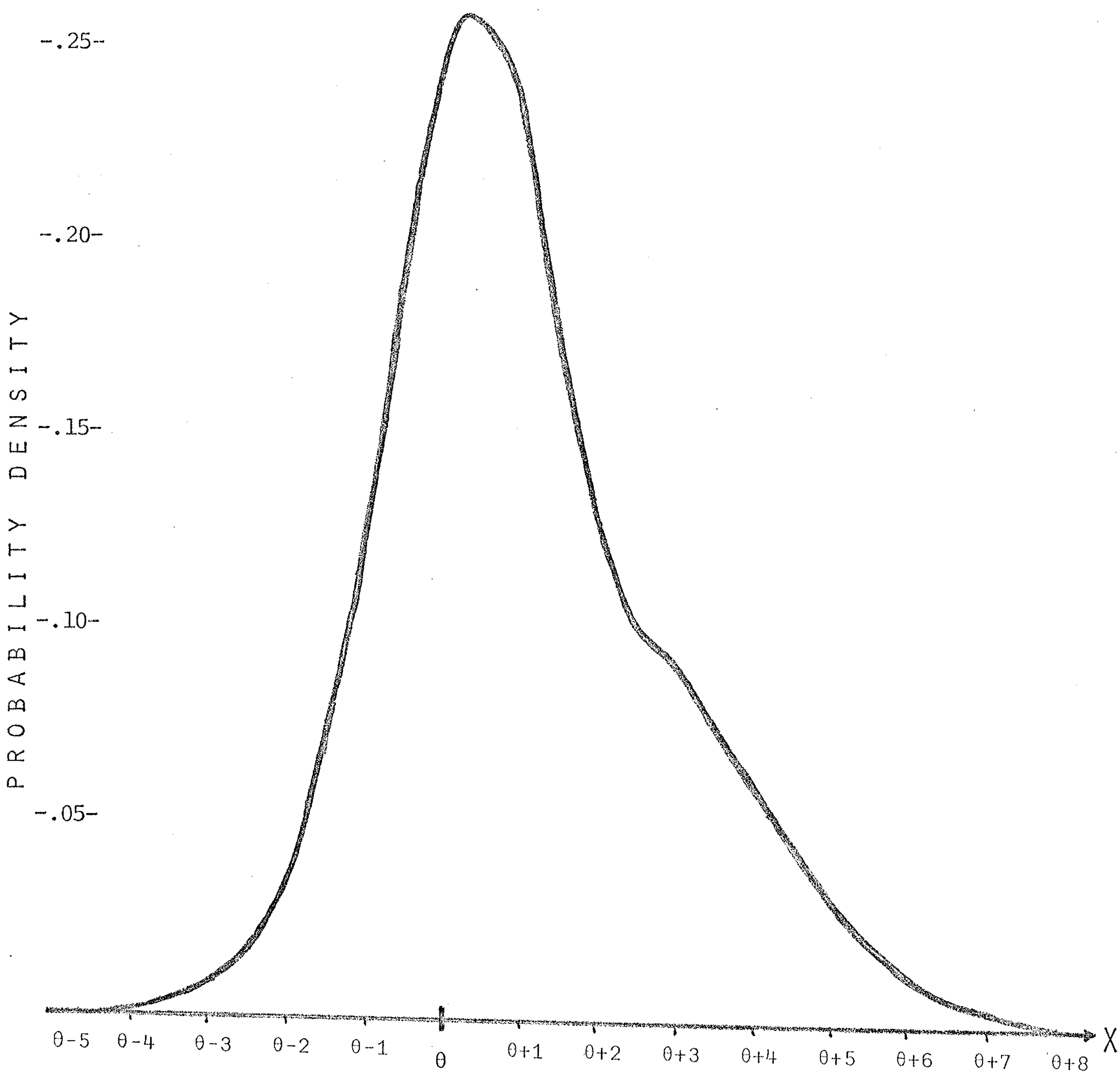

FIGURE 4.3 : Probability density function of mixture of Normals, $MN_{\frac{1}{2}}(2,2) = \frac{1}{2}N(0,1) + \frac{1}{2}N(2,2^2)$.

TABLE 4.3: Exponential Rates for $MN_{\frac{1}{2}}(2,2)$

ε	$b(\hat{\theta},\theta,\varepsilon)$	$b(\tilde{\theta},\theta,\varepsilon)$	$B(\theta,\varepsilon)$	Lemma 2.4
0.1	.0015*	.0020553	.0020564	.0020533
0.2	.006*	.0081082	.0081260	.0081092
0.5	.034	.048135	.048798	.048207
1.0	.129	.17183	.18080	.17322
2.0	.476	.53181	.60725	.53861
3.0	1.014	1.02068	1.16307	.98006
10.0	10.818	10.81901	10.87824	9.45403

(* Very approximate, convergence is very slow)

The last column of Table 4.3 is the value of the exponential rate for the mle as given by the Bahadur-Chernoff Theorem in Lemma 2.4. We note that these values are quite different from the rates computed by (4.15), and that for $\varepsilon<3$, they exceed the rate of of the pre. Similar results were obtained under symmetry, when the pre is optimal for the class of translation invariant estimators to which they both belong. Hence Lemma 2.4 does not appear to hold.

Note also that the rate of the mle as computed from (4.15) by summing terms, may be subject to greater error when the probabilities are small, which together with the fact that the convergence is very slow for small ε, makes the table entries $b(\hat{\theta},\theta,\varepsilon)$ very approximate indeed (and not really comparable with the other columns, for small ε).

4.2 Double Exponential Distribution

The Double Exponential Distribution,

$$f(x|\theta) = \tfrac{1}{2}e^{-|x-\theta|}, \qquad \begin{array}{l} -\infty<x<\infty, \\ -\infty<\theta<\infty, \end{array} \tag{4.16}$$

does not satisfy all of the regularity conditions of Chapter 2, and, in particular, the mle is not the unique solution of the likelihood equation. The mle, $\hat{\theta}_n = med\{x_1, x_2, \ldots, x_n\}$, has exponential rate, determined by Bahadur (1971, Example 6.1),

$$b(\hat{\theta},\theta,\varepsilon) = -\tfrac{1}{2}\log e^{-\varepsilon}(2 - e^{-\varepsilon}). \tag{4.17}$$

The Bahadur bound is easily determined and well known,

$$B(\theta,\varepsilon) = e^{-\varepsilon} + \varepsilon - 1. \tag{4.18}$$

The distribution is suitably smooth so that Theorem 2.1 will still hold, so that the exponential rate of the probability ratio estimator, which is also the optimal rate for the class of translation invariant estimators, consistent for θ, is given by Theorem 2.1, as obtained by Sievers (1978),

$$b_T(\theta,\varepsilon) = b(\tilde{\theta},\theta,\varepsilon) = -\log(1 + \varepsilon). \tag{4.19}$$

By symmetry, all of the Fisher moment of odd order are zero, but since [C1] to [C4] fail to hold, the Taylor series expansions of §2.2-§2.4 are not valid.

However, each of the above rate expressions can be expanded directly into Taylor series,

$$b(\hat{\theta},\theta,\varepsilon) = \varepsilon^2/2! - 1\varepsilon^3/3! + 1\varepsilon^4/4! + o(\varepsilon^4), \quad (4.20)$$

$$b_T(\theta,\varepsilon) = \varepsilon^2/2! - 2\varepsilon^3/3! + 6\varepsilon^4/4! + o(\varepsilon^4), \quad (4.21)$$

$$B(\theta,\varepsilon) = \varepsilon^2/2! - 3\varepsilon^3/3! + 13\varepsilon^4/4! + o(\varepsilon^4). \quad (4.22)$$

Therefore, for small ε,

$$b(\hat{\theta},\theta,\varepsilon) < b_T(\theta,\varepsilon) < B(\theta,\varepsilon), \quad (4.23)$$

as in the case of the logistic distribution. One interesting point is that although the distribution is symmetric and the odd-order Fisher moments all vanish, the odd-degree terms of the above Taylor series <u>do not</u> vanish. This means that for this non-regular distribution, the mle is not second-order efficient, and possibly, the pre is also not second-order efficient as we do not know what the maximum attainable rate will be in this case, assumming again that the Bahadur bound is too large and that there is no estimator-sequence whose exponential rate will achieve this bound.

CHAPTER FIVE
CONCLUSIONS AND EXTENSIONS

5.1 Statistical Curvature

Efron (1975, 1978), Madsen (1979), Amari (1982a, 1982b), Kass (1982) and others have investigated the geometry of the statistical distributions and demonstrated the relationship of statistical curvature to the mathematical curvature. Efron (1975) and Madsen (1979) have shown that the exponential family is statistically 'linear' and has curvature zero, $\gamma_\theta^2 = 0$. Consequently γ_θ^2 measures the extent to which score function departs from linearity, which is also a measure of how much the underlying distribution departs from the exponential family. Efron (1975) goes on to infer that γ_θ^2 also measures the loss in information on the mle $\hat{\theta}$, since $\hat{\theta}$ is optimal only for the exponential family. However the mle is just one member of a large class of estimators for which this 'loss of information' is measured by γ_θ^2. Indeed among all 'efficient' (by Fisher) estimators, the definition by Rao (1961) of second-order efficiency (which led to the definition of γ_θ^2) was satisfied by a large class of estimators which includes the mle, and in the case of translation invariance, the pre as well. However, attributes such as 'loss of information', 'departure from exponentiallity' or 'efficiency' cannot be adequately characterized by a single measure. Thus, γ_θ^2 is only the first of a sequence of such measures. We will wish to define other such measures, measures which will relate to the higher order efficiences. With this consideration in mind, we defined the higher order rate coefficients, δ_θ, λ_θ^3, and ν_θ^3. In particular, λ_θ^3 appears to be a kind of third-order extension of the squared curvature.

Let us return to the specific case of translation invariance under a regular distribution. Both the mle and the pre are second-order efficient, which is shown by the fact that their exponential rates co-incide up to the fourth-degree term, i.e. the term involving γ_θ^2. However, in general for regular distributions, the pre but not the mle dominates and may be optimal for the class of translation invariant estimators of which both are members. This is demonstrated in the case of the Logistic distribution and the Hyperbolic Secant distribution (and numerically for the Normal mixtures) in that

$$b(\tilde{\theta},\theta,\varepsilon) > b(\hat{\theta},\theta,\varepsilon), \tag{5.1}$$

strictly. This suggests that the prominant place accorded to the mle may be unjustified. It is optimal for exponential family, but only for the exponential family. Also, the mle has been singled out by Rao (1961, etc) and others, by showing that the mle has a favored place among (first-order) efficient estimators in that it has the minimum loss of sample information and is hence, by this definition, second-order efficient. However, we have shown by (5.1) that the mle is <u>not</u> favored among the second-order efficient estimators in that the mle is clearly dominated by the pre in these instances. Indeed, the Taylor series expansions of §2.2-§2.4, and with the actual application in the case of the Hyperbolic Secant distribution, demonstrate that if we define third-order efficiency in terms of the exponential rates of convergence, then among all of the second-order efficient estimators, the mle cannot be third-order efficient. Also, as the pre has only been shown to be optimal for the class of translation invariant estimators, and since $B(\theta,\varepsilon) > b(\tilde{\theta},\theta,\varepsilon)$, and differ in the fourth- and sixth-degree terms, it follows that the pre may not be third-order efficient either. If we refer to the Conjecture 3 in Chapter Two, however, we do believe that, in this case, the pre is third-order efficient, and in fact, optimal for all consistent estimators, not just the translation invariant class under symmetric distributions.

Hence, in this one instance, we have shown the mle to be non-optimal and not even third-order efficient, and we believe that the pre may in this case be third-order efficient and optimal. On the other hand, the mle is optimal for the exponential family. In general, it may be possible that there is no one estimator which is universally best. Intuitively, we feel that the class of second-order efficient estimators is partially ordered with no dominant estimator-sequence.

5.2 Significance of the Rate Coefficients

The special properties of the statistical curvature, γ_θ^2, by which it remains unaffected by transformation, allow us to interpret γ_θ^2 the same way regardless of the nature of the parameter θ, whether it is a location parameter, a scale parameter, shape parameter, or whatever. Indeed, we have seen that γ_θ^2 characterizes the efficiency of all second-order efficient estimators. We would like to be able to define other parameters which would characterize the higher order efficiencies as well. This may not be as easily accomplished, as is evidenced by Lemma 3.2 which relates the rate coefficients of Location-scale Duals. This lemma suggests that any measure characterizing the efficiency beyond the second-order may depend on the nature of the parameter θ. This would undoubtably increase the complexity of the situation, but should not really be too surprising. Basically, γ_θ^2, being the first rate coefficient, is a kind of 'linear' or 'steady-state' component of the efficiency, measuring the amount of departure or curvature from the 'linear' exponential family for which $\gamma_\theta^2=0$. Going beyond this 'linearity', the interpretation and significance of the parameters we define could depend on the nature of θ.

In Chapter Two we defined several parameters with specific application to the problem of a translation invariant location parameter θ of a regular distribution. The divergence of the mle from the pre is characterized by δ_θ and ν_θ^3, while the difference of the pre from the Bahadur bound depends on γ_θ^2 and λ_θ^3 (see Definition 2.2). By our conjecture of the optimality of the pre, and by the definition of Rao, it appears that γ_θ^2 and λ_θ^3 are in a sense 'corrections' to the Bahadur bound which appears to be too large. Refering to the definitions (2.36) and (2.39),

$$\gamma_\theta^2 = I^{-4}(\theta)E_\theta\begin{vmatrix} I(\theta) & h_1 \\ J(\theta) & h_2 \end{vmatrix}^2, \qquad (5.2)$$

$$\lambda_\theta^3 = I^{-6}(\theta) E_\theta \begin{vmatrix} I(\theta) & h_1 \\ J(\theta) & h_2 \end{vmatrix}^3 \tag{5.3}$$

$$+ I^{-5}(\theta) \begin{pmatrix} 2J(\theta) \\ I(\theta) \end{pmatrix}' \begin{pmatrix} \eta_{1010} & -\eta_{0110} \\ -\eta_{1001} & \eta_{0101} \end{pmatrix} \begin{pmatrix} J(\theta) \\ I(\theta) \end{pmatrix},$$

we see that λ_θ^3 has two parts:

1. A cubic generalization of the quadratic in γ_θ^2, which involves only the first two derivatives of the log-density function, $\ell = \log f(x|\theta)$.

2. A 'correction' term involving the covariances between the vector of the first two derivatives $(\dot{\ell}, \ddot{\ell})$, and the vector of the next two derivatives $(\ell^{(3)}, \ell^{(4)})$.

Judging by this definition and its application, we may give a similar interpretation to λ_θ^3 as γ_θ^2, in that γ_θ^2 is a kind of quadratic distance measure, while λ_θ^3 is cubic distance. Indeed, by our conjecture, γ_θ^2 and λ_θ^3 are the quadatic and cubic corrections to the Bahadur bound corresponding to the second- and third-order efficiencies.

It would also seem that δ_θ and ν_θ^3, the parameters defined on the mle, may not have as wide an interpretation, however we note that their definitions bear some striking similarities to the previous. For instance, (2.37), the definition of δ_θ,

$$\delta_\theta = \frac{1}{I^3(\theta)J(\theta)} E_\theta \begin{vmatrix} I(\theta) & h_1 \\ J(\theta) & h_2 \end{vmatrix} \begin{vmatrix} I(\theta) & h_1 \\ J(\theta) & h_3 \end{vmatrix}, \qquad (5.4)$$

is similar to that of γ_θ^2 and the first term of λ_θ^3, except that it involves the mean of the cross-product of $\begin{vmatrix} I(\theta) & h_1 \\ J(\theta) & h_2 \end{vmatrix}$ which is a function of $(\dot{\ell}, \ddot{\ell})$ and a similar determinant, in this case a function of $(\dot{\ell}, \ell^{(3)})$. This kind of a cross-product is due to the fact that δ_θ corresponds to the fifth-degree term in the difference of the exponential rates, whereas γ_θ^2 and λ_θ^3 correspond to even degree terms. By the alternate definition of δ_θ in terms of the covariances of the score function,

$$\delta_\theta = \frac{1}{I^2(\theta)J(\theta)} \begin{vmatrix} \eta_{200} & \eta_{101} \\ \eta_{110} & \eta_{011} \end{vmatrix}. \qquad (5.5)$$

The matrix here is similar to γ_θ^2 and the matrix in the second term of λ_θ^3, but involves the covariances of $(\dot{\ell}, \ddot{\ell})\times(\dot{\ell}, \ell^{(3)})$ rather than $(\dot{\ell}, \ddot{\ell})\times(\dot{\ell}, \ddot{\ell})$ or $(\dot{\ell}, \ddot{\ell})\times(\ell^{(3)}, \ell^{(4)})$. Similarly, in terms of covariances,

$$\gamma_\theta^2 = I^{-3}(\theta) \begin{vmatrix} \eta_{200} & \eta_{110} \\ \eta_{110} & \eta_{020} \end{vmatrix}, \qquad (5.6)$$

and,

$$\nu_\theta^3 = I^{-4}(\theta)\begin{vmatrix} \eta_{200} & \eta_{101} \\ \eta_{101} & \eta_{002} \end{vmatrix} + 3\omega_\theta^2\delta_\theta. \tag{5.7}$$

Again, ν_θ^3 contains two parts:

1. A covariance matrix of $(\dot{\ell}, \ell^{(3)})\times(\dot{\ell}, \ell^{(3)})$, similar to the cross-product matrices in γ_θ^2, δ_θ and λ_θ^3.

2. A 'correction' term involving δ_θ.

Similarities of definition aside, these terms are somewhat more difficult to interpret (for general distributions). How do we interpret their values in the non-regular case, or even the semi-regular case, as with the mixtures of Normals, when the Taylor expansions of the mle are not valid and hence δ_θ and ν_θ^3 are in this case not measures of the divergence of the mle from the pre? Indeed, as we vary the parameters of the Normal mixtures, we are able to vary δ_θ from positive through negative values, while the actual fifth-degree divergence of the mle relative to the pre remains positive.

5.3 Conclusions

The major conclusions to be drawn from this investigation are the following:

1. The Maximum Likelihood Estimator is second-order efficient but optimal only for the exponential family. In the case of translation invariant location parameter of a regular distribution, the mle is shown to be not even third-order efficient.

2. For translation invariant location parameter of a regular distribution, the Probability Ratio Estimator may be optimal for the class of translation invariant estimators, which includes the mle; and the pre may, in fact, be optimal for the class of all consistent estimators.

3. The differences of the exponential rates depend on coefficients γ_θ^3, δ_θ, λ_θ^3 and ν_θ^3, of which λ_θ^3 appears to be a third-order extension of the squared statistical curvature, γ_θ^2.

4. For non-exponential families, the Bahadur bound appears to be too large in that it does not appear possible to construct a consistent estimator-sequence whose exponential rate will attain this bound. The 'corrections' necessary to make the bound attainable depend on a sequence of parameters, which we believe begin with γ_θ^2, λ_θ^3, ξ_θ^4,....; where ξ_θ^4 is as yet undefined.

5.4 Extensions

There are still some aspects of our investigation that require some additional probing. As well, other extensions to our work are suggested. The following list of problems, of varying complexity, are among those inspired by this investigation:

(A) Similar studies under other family restrictions, such as scale invariance.

(B) To prove (or disprove), in general, Point #3 of §5.3 and prove (or disprove) that λ_θ^3 is the third-order extension of the statistical curvature, regardless of the nature of θ.

(C) Further investigate the nature and the properties of λ_θ^3.

(D) Along the same lines as Problem (A), to define third-order efficiency and its relationship to λ_θ^3.

(E) Refering to Point #4 of §5.3, to define ξ_θ^4 and conduct investigations similar to those proposed for λ_θ^3 above.

(F) To prove (or disprove), in general, the assertion in Point #4 of §5.3, and to obtain the corrected upper bound, $B^*(\theta,\varepsilon) < B(\theta,\varepsilon)$.

(G) To show that $B^*(\theta,\varepsilon)$ is attainable and to construct such an optimal estimator-sequence.

(H) Failing (F) and (G), to construct optimal estimator-sequences in restricted cases, such as:

(i) Show that $B^*(\theta,\varepsilon) = b_T(\theta,\varepsilon)$ as has been conjectured.

(ii) Obtain $b_S(\sigma,\varepsilon)$ and show that $B^*(\sigma,\varepsilon) = b_S(\sigma,\varepsilon)$ where S is the class of scale invariant estimators consistent for the pure scale parameter σ.

Some of these problems may not be soluble and indeed further investigation may make some obsolete and create new problems, as is often the case with research. However, it seems appropriate to conclude this investigation at this point with the above list of possible further investigations.

APPENDICES

6.1 Computation of Fisher moments

In this appendix we give the details involving the compution of the Fisher moments of the distributions in Chapter 3: the Logistic distribution, the general exponential location family, the one-parameter Normal distribution, and the Hyperbolic Secant distribution.

6.1.1 Logistic Distribution (§3.3)

Let us reparametrize, to make integration easier,

$$g = F(x|\theta) = [1 + e^{-(x - \theta)}]^{-1}. \tag{6.1}$$

Then, the derivatives, in terms of g are

$$\begin{aligned}
&(A)\quad f(x|\theta) = -\dot{g} = -(g^2 - g),\\
&(B)\quad f^{(1)}(x|\theta) = -(2g - 1)\dot{g} = -(2g^3 - 3g^2 + g),\\
&(C)\quad f^{(2)}(x|\theta) = -(6g^2 - 6g + 1)\dot{g}\\
&\qquad\qquad = -(6g^4 - 12g^3 + 7g^2 - g),\\
&(D)\quad f^{(3)}(x|\theta) = -(24g^3 - 36g^2 + 14g - 1)\dot{g}\\
&\qquad\qquad = -(24g^5 - 60g^4 + 50g^3 - 15g^2 + g),\\
&(E)\quad f^{(4)}(x|\theta) = -(120g^4 - 240g^3 + 150g^2 - 30g + 1)\dot{g}\\
&\qquad\qquad = -(120g^6 - 360g^5 + 390g^4 - 180g^3 + 31g^2 - g),\\
&(F)\quad f^{(5)}(x|\theta) = -(720g^5 - 1800g^4 + 1560g^3 - 540g^2 + 62g - 1)\dot{g}.
\end{aligned} \tag{6.2}$$

and, the integral of a polynomial in g can be evaluated from the following:

$$\int_{-\infty}^{\infty} g^k \, dF(x|\theta) = \int_0^1 g^k \, dg = \left.\frac{g^{k+1}}{k+1}\right|_0^1 = \frac{1}{k+1}. \tag{6.3}$$

The derivative ratios are then,

$$
\begin{aligned}
&\text{(A)}\quad d_1 = 2g - 1,\\
&\text{(B)}\quad d_2 = 6g^2 - 6g + 1,\\
&\text{(C)}\quad d_3 = 24g^3 - 36g^2 + 14g - 1, \qquad\qquad (6.4)\\
&\text{(D)}\quad d_4 = 120g^4 - 240g^3 + 150g^2 + 30g - 1,\\
&\text{(E)}\quad d_5 = 720g^5 - 1800g^4 + 1560g^3 - 540g^2 + 62g - 1.
\end{aligned}
$$

And so, by symmetry all the odd order moments are zero, thus,

$$
\begin{aligned}
&\text{(A)}\quad \mu_{200} = E_\theta(4g^2-4g+1) = 4/3 - 4/2 + 1 = 1/3,\\
&\text{(B)}\quad \mu_{400} = E_\theta(16g^4-32g^3+24g^2-8g+1)\\
&\qquad = 16/5-32/4+24/3-8/2+1 = 1/5,\\
&\text{(C)}\quad \mu_{210} = E_\theta(24g^4-48g^3+34g^2-10g+1)\\
&\qquad = 24/5-48/4+34/3-10/2+1 = 2/15,\\
&\text{(D)}\quad \mu_{101} = E_\theta(48g^4-96g^3+64g^2-16g+1)\\
&\qquad = 48/5-96/4+64/3-16/2+1 = -1/15,\\
&\text{(E)}\quad \mu_{020} = E_\theta(36g^4-72g^3+48g^2-12g+1) \qquad\qquad (6.5)\\
&\qquad = 36/5-72/4+48/3-12/2+1 = 1/5,\\
&\text{(F)}\quad \mu_{600} = E_\theta(64g^6-192g^5+240g^4-160g^3+60g^2-12g+1)\\
&\qquad = 64/7-192/6+240/5-160/4+60/3-12/2+1 = 1/7,\\
&\text{(G)}\quad \mu_{410} = E_\theta(96g^6-288g^5+352g^4-224g^3+78g^2-14g+1)\\
&\qquad = 96/7-288/6+352/5-224/4+78/3-14/2+1 = 4/35,\\
&\text{(H)}\quad \mu_{301} = E_\theta(192g^6-576g^5+688g^4-416g^3+132g^2-20g+1)\\
&\qquad = 192/7-576/6+688/5-416/4+132/3-20/2+1 = 1/35,
\end{aligned}
$$

(I) $\mu_{220} = E_\theta(144g^6-432g^5+516g^4-312g^3+100g^2-16g+1)$

$= 144/7-432/6+516/5-312/4+100/3-16/2+1 = 11/105,$

(J) $\mu_{111} = E_\theta(288g^6-864g^5+1008g^4-576g^3+166g^2-22g+1)$

$= 288/7-864/6+1008/5-576/4+166/3-22/2+1 = 8/105,$

(K) $\mu_{030} = E_\theta(216g^6-648g^5+756g^4-432g^3+126g^2-18g+1)$

$= 216/7-648/6+756/5-432/4+126/3-18/2+1 = 2/35,$

(L) $\mu_{002} = E_\theta(576g^6-1728g^5+1968g^4-1056g^3+268g^2-28g+1)$

$= 576/7-1728/6+1968/5-1056/4+268/3-28/2+1 = 23/105,$

(M) $\mu_{2001} = E_\theta(480g^6-1440g^5+1680g^4-960g^3+274g^2-34g+1)$

$= 480/7-1440/6+1680/5-960/4+274/3-34/2+1 = -2/21,$

(N) $\mu_{0101} = E_\theta(720g^6-2160g^5+2460g^4-1320g^3+336g^2-36g+1)$

$= 720/7-2160/6+2460/5-1320/4+336/3-36/2+1 = -1/7,$

(O) $\mu_{10001} = E_\theta(1440g^6-4320g^5+4920g^4-2640g^3+664g^2-64g+1)$

$= 1440/7-4320/6+4920/5-2640/4+664/3-64/2+1 = 1/21.$

Similarly, for the scale dual of the logistic (3.39),

$$h = 1 - F(y|\sigma) = \sigma/(\sigma + y), \tag{6.6}$$

and so, differentiating by σ, with $\dot{h} = (h - h^2)/\sigma$,

(A) $f(y|\sigma) = h^2/\sigma,$

(B) $f^{(1)}(y|\sigma) = 2h\dot{h}/\sigma - h^2/\sigma^2 = (h^2 - 2h^3)/\sigma^2,$

(C) $f^{(2)}(y|\sigma) = (2h - 6h^2)\dot{h}/\sigma^2 - 2(h^2 - 2h^3)/\sigma^3$

$= (6h^4 - 4h^3)/\sigma^3,$ (6.7)

(D) $f^{(3)}(y|\sigma) = (24h^3 - 12h^2)\dot{h}/\sigma^3 - 3(6h^4 - 4h^3)/\sigma^4$

$= (18h^4 - 24h^5)/\sigma^4,$

$$\text{(E)}\quad f^{(4)}(y|\sigma) = (72h^3 - 120h^4)\dot{h}/\sigma^4 - 4(18h^4 - 24h^5)/\sigma^5$$
$$= (120h^6 - 96h^5)/\sigma^5,$$
$$\text{(F)}\quad f^{(5)}(y|\theta) = (720h^5 - 480h^4)\dot{h}/\sigma^5 - 5(120h^6 - 96h^5)/\sigma^6$$
$$= (600h^6 - 720h^7)/\sigma^6.$$

and, again, the integral of any polynomial in h can be determined from the following:

$$\int_{-\infty}^{\infty} h^k \, dF(y|\sigma) = \int_0^1 h^k \, dh = \left.\frac{h^{k+1}}{k+1}\right|_0^1 = \frac{1}{k+1}. \quad (6.8)$$

Then, the derivative ratios are

$$\begin{aligned}
&\text{(A)}\quad d_1 = (1 - 2h)/\sigma, \\
&\text{(B)}\quad d_2 = (6h^2 - 4h)/\sigma^2, \\
&\text{(C)}\quad d_3 = (18h^2 - 24h^3)/\sigma^3, \\
&\text{(D)}\quad d_4 = (120h^4 - 96h^3)/\sigma^4, \\
&\text{(E)}\quad d_5 = (600h^4 - 720h^5)/\sigma^5.
\end{aligned} \quad (6.9)$$

And hence the Fisher moments of the scale dual of the logistic are as follows:

$$\text{(A)}\quad \mu_{200} = E_\sigma(1-4h+4h^2)/\sigma^2 = (1-4/2+4/3)/\sigma^2 = 1/3\sigma^2,$$
$$\text{(B)}\quad \mu_{300} = E_\sigma(1-6h+12h^2-8h^3)/\sigma^3 = (1-6/2+12/3-8/4)/\sigma^3 = 0,$$
$$\text{(C)}\quad \mu_{110} = E_\sigma(-4h+14h^2-12h^3)/\sigma^3 = (-4/2+14/3-12/4)/\sigma^3 = -\frac{1}{3}\sigma^{-3},$$
$$\text{(D)}\quad \mu_{400} = E_\sigma(1-8h+24h^2-32h^3+16h^4)/\sigma^4$$
$$= (1-8/2+24/3-32/4+16/5)/\sigma^4 = 1/5\sigma^4,$$

(E) $\mu_{210} = E_\sigma(-4h^1+22h^2-40h^3+24h^4)/\sigma^4$

$= (-4/2+22/3-40/4+24/5)/\sigma^4 = 2/15\sigma^4$,

(F) $\mu_{020} = E_\sigma(16h^2-48h^3+36h^4)/\sigma^4$

$= (16/3-48/4+36/5)/\sigma^4 = 8/15\sigma^4$,

(G) $\mu_{101} = E_\sigma(18h^2-60h^3+48h^4)/\sigma^4$

$= (18/3-60/4+48/5)/\sigma^4 = 3/5\sigma^4$,

(H) $\mu_{500} = E_\sigma(1-10h+40h^2-80h^3+80h^4-32h^5)/\sigma^5$

$= (1-10/2+40/3-80/4+80/5-32/6)/\sigma^5 = 0$,

(I) $\mu_{310} = E_\sigma(-4h+30h^2-84h^3+104h^4-48h^5)/\sigma^5$

$= (-4/2+30/3-84/4+104/5-48/6)/\sigma^5 = -1/5\sigma^5$,

(J) $\mu_{201} = E_\sigma(18h^2-96h^3+168h^4-96h^5)/\sigma^5$

$= (18/3-96/4+168/5-96/6)/\sigma^5 = -2/5\sigma^5$,

(K) $\mu_{120} = E_\sigma(16h^2-80h^3+132h^4-72h^5)/\sigma^5$

$= (16/3-80/4+132/5-72/6)/\sigma^5 = -4/15\sigma^5$,

(L) $\mu_{1001} = E_\sigma(-96h^3+312h^4-240h^5)/\sigma^5$ (6.10)

$= (-96/4+312/5-240/6)/\sigma^5 = -8/5\sigma^5$,

(M) $\mu_{011} = E_\sigma(-72h^3+204h^4-144h^5)/\sigma^5$

$= (-72/4+204/5-144/6)/\sigma^5 = -6/5\sigma^5$,

(N) $\mu_{600} = E_\sigma(1-12h+60h^2-160h^3+240h^4-192h^5+64h^6)/\sigma^6$

$= (1-12/2+60/3-160/4+240/5-192/6+64/7)/\sigma^6 = 1/7\sigma^6$,

(O) $\mu_{410} = E_\sigma(-4h+38h^2-144h^3+272h^4-256h^5+96h^6)/\sigma^6$

$= (-4/2+38/3-144/4+272/5-256/6+96/7)/\sigma^6 = 4/35\sigma^6$,

(P) $\mu_{301} = E_\sigma(18h^2-132h^3+360h^4-432h^5+192h^6)/\sigma^6$

$= (18/3-132/4+360/5-432/6+192/7)/\sigma^6 = 3/7\sigma^6$,

$$\text{(Q)}\quad \mu_{220} = E_\sigma(16h^2-112h^3+292h^4-336h^5+144h^6)/\sigma^6$$
$$= (16/3-112/4+292/5-336/6+144/7)/\sigma^6 = 32/105\sigma^6,$$
$$\text{(R)}\quad \mu_{111} = E_\sigma(-72h^3+348h^4-552h^5+288h^6)/\sigma^6$$
$$= (-72/4+348/5-552/6+288/7)/\sigma^6 = 26/35\sigma^6,$$
$$\text{(S)}\quad \mu_{030} = E_\sigma(-64h^3+288h^4-432h^5+216h^6)/\sigma^6$$
$$= (-64/4+288/5-432/6+216/7)/\sigma^6 = 16/35\sigma^6,$$
$$\text{(T)}\quad \mu_{002} = E_\sigma(324h^4-864h^5+576h^6)/\sigma^6$$
$$= (324/5-864/6+576/7)/\sigma^6 = 108/35\sigma^6,$$
$$\text{(U)}\quad \mu_{0101} = E_\sigma(384h^4-1056h^5+720h^6)/\sigma^6$$
$$= (384/5-1056/6+720/7)/\sigma^6 = 128/35\sigma^6,$$
$$\text{(V)}\quad \mu_{2001} = E_\sigma(-96h^3+504h^4-864h^5+480h^6)/\sigma^6$$
$$= (-96/4+504/5-864/6+480/7)/\sigma^6 = 48/35\sigma^6,$$
$$\text{(W)}\quad \mu_{10001} = E_\sigma(600h^4-1920h^5+1440h^6)/\sigma^6$$
$$= (600/5-1920/6+1440/7)/\sigma^6 = 40/7\sigma^6.$$

We could have obtained the above results from (3.32) in Lemma 3.2, and the Fisher moments of the original distribution (6.5). Note, however, that this scale dual is not translation invariant and hence Lemma 2.1 does not apply to the above moments (6.10), although it does apply to the moments of the original translation invariant logistic distribution (6.5). Note also that the rate coefficients as given in Table 3.1 can be computed directly from the above moments (6.10), or by applying Lemma 3.2.

6.1.2 General Exponential Family (§3.4.1)

Let us reparametrize as $t=me^{p(x-\theta)}$, then

$$f(x|\theta) = \frac{p\ m^q}{\Gamma(q)} e^{pq(x-\theta)-me^{p(x-\theta)}} = pt^q e^{-t}/\Gamma(q). \qquad (6.11)$$

And so diffentiating by θ, with $\dot{t}=-pt$, we get

$$\begin{aligned} \text{(A)}\quad \dot{f}(x|\theta) &= p(qt^{q-1} - t^q)\dot{t}\ e^{-t}/\Gamma(q) \\ &= p^2 t^q (t - q)e^{-t}/\Gamma(q), \end{aligned} \qquad (6.12)$$

$$\begin{aligned} \text{(B)}\quad \ddot{f}(x|\theta) &= p^2[(q+1)t^q - q^2 t^{q-1} - t^q(t - q)]\dot{t}\ e^{-t}/\Gamma(q) \\ &= p^3 t^q [t^2 - (2q+1)t + q^2]e^{-t}/\Gamma(q). \end{aligned}$$

Then, integrating a polynomial in t,

$$\begin{aligned} \int_{-\infty}^{\infty} t^k\ dF(x|\theta) &= \int_{-\infty}^{\infty} t^k\ pt^q e^{-t}/\Gamma(q)\ dx \\ &= \int_{0}^{\infty} t^k\ t^{q-1}e^{-t}/\Gamma(q)\ dt \qquad (6.13) \\ &= \frac{\Gamma(q+k)}{\Gamma(q)} = q(q+1)(q+2)\ldots(q+k-1). \end{aligned}$$

Hence, the derivative ratios are

$$\begin{aligned} \text{(A)}\quad d_1 &= (t - q)p, \\ \text{(B)}\quad d_2 &= (t^2 - (2q+1)t + q^2)p^2. \end{aligned} \qquad (6.14)$$

And so, the Fisher information is

$$\begin{aligned} I(\theta) &= p^2 E_\theta(t^2 - 2qt + q^2) \\ &= p^2[q(q+1) - 2q^2 + q^2] = qp^2. \end{aligned} \qquad (6.15)$$

Then, the third-order Fisher moments are

$$
\begin{aligned}
&(\mathrm{A})\quad \mu_{300} = p^3 E_\theta\; t^3-3qt^2+3q^2t-q^3 \\
&\qquad\qquad = p^3[q(q+1)(q+2)-3q^2(q+1)+3q^3-q^3] = 2qp^3\,, \\
&(\mathrm{B})\quad \mu_{110} = p^3 E_\theta\; t^3-(3q+1)t^2+q(3q+1)t-q^3 \\
&\qquad\qquad = p^3[q(q+1)(q+2)-(3q+1)q(q+1)+(3q+1)q^2-q^3] = qp^3.
\end{aligned}
\tag{6.16}
$$

Therefore, the 'scale-information' is

$$
\begin{aligned}
J(\theta) &= \mu_{110} - \mu_{300} = qp^3 - 2qp^3 = -qp^3\,, \\
\omega_\theta^2 &= J^2(\theta)/I^3(\theta) = q^2p^6/q^3p^6 = 1/q\,.
\end{aligned}
\tag{6.17}
$$

The higher order Fisher moments are not required, since for this underlying exponential family the other rate coefficients will all be zero.

Now to obtain the coefficient of skewness, we require the central moments μ_2 and μ_3. To obtain these we first find the moments about the location,

$$
\begin{aligned}
\xi_i &= E_\theta(X-\theta)^i \\
&= \int_{-\infty}^{\infty} (x-\theta)^i\, \frac{p\;m^q}{\Gamma(q)} e^{pq(x-\theta)-me^{p(x-\theta)}}\, dx \\
&= \int_0^{\infty} \left(\frac{1}{p}\log z\right)^i \frac{m^q}{\Gamma(q)}\, z^{q-1}e^{-mz}\, dz \\
&= \frac{p^{-i}m^q}{\Gamma(q)} \int_0^{\infty} (\log z)^i z^{q-1} e^{-mz}\, dz \\
&= \frac{p^{-i}m^q}{\Gamma(q)}\, \frac{\partial^i}{\partial q^i}\{\Gamma(q)/m^q\}\,.
\end{aligned}
\tag{6.18}
$$

Hence,

$$\xi_1 = \frac{p^{-1} m^q}{\Gamma(q)} \frac{\partial}{\partial q} (\Gamma(q)/m^q)$$

$$= \frac{p^{-1} m^q}{\Gamma(q)} (\dot{\Gamma}(q)/m^q - \Gamma(q)/m^q \log m)$$

$$= \frac{1}{p} (\dot{\Gamma}(q)/\Gamma(q) - \log m) \tag{6.19}$$

$$= \frac{1}{p} (\psi(q) - \log m),$$

where $\psi(x)$ is the digamma function,

$$\psi(x) = \frac{\partial}{\partial x} \log \Gamma(x) = \dot{\Gamma}(x)/\Gamma(x). \tag{6.20}$$

Also,

$$\xi_2 = \frac{p^{-2} m^q}{\Gamma(q)} \frac{\partial^2}{\partial q^2} (\Gamma(q)/m^q)$$

$$= \frac{p^{-2} m^q}{\Gamma(q)} (\ddot{\Gamma}(q)/m^q - 2\dot{\Gamma}(q)/m^q \cdot \log m + \Gamma(q)/m^q \cdot \log^2 m)$$

$$= p^{-2} (\ddot{\Gamma}(q)/\Gamma(q) - 2\dot{\Gamma}(q)/\Gamma(q) \log m + \{\log m\}^2) \tag{6.21}$$

$$= p^{-2} (\dot{\psi}(q) + \psi^2(q) - 2\psi(q) \log m + \{\log m\}^2),$$

and so, the population variance is

$$\mu_2 = \xi_2 - \xi_1^2$$

$$= \dot{\psi}(q)/p^2. \tag{6.22}$$

Similarly,

$$\xi_3 = \frac{p^{-3}m^q}{\Gamma(q)}\,\frac{\partial^3}{\partial q^3}\,(\Gamma(q)/m^q)$$

$$= \frac{p^{-3}m^q}{\Gamma(q)}\left\{\frac{\Gamma(q)}{m^q} - 3\frac{\ddot{\Gamma}(q)}{m^q}\log\, m + 3\frac{\dot{\Gamma}(q)}{m^q}\log^2 m - \frac{\Gamma(q)}{m^q}\log^3 m\right\}$$

$$= p^{-3}\left\{\frac{\Gamma(q)}{\Gamma(q)} - 3\frac{\ddot{\Gamma}(q)}{\Gamma(q)}\log\, m + 3\frac{\dot{\Gamma}(q)}{\Gamma(q)}\log^2 m - \log^3 m\right\} \qquad (6.23)$$

$$= p^{-3}(\ddot{\psi}(q) + 3\dot{\psi}(q)\psi(q) + \psi^3(q) - 3[\dot{\psi}(q) + \psi^2(q)]\log\, m$$

$$+ 3\psi(q)\log^2 m - \log^3 m).$$

Then, the third central moment is

$$\mu_3 = \xi_3 - 3\xi_1\xi_2 + 2\xi_1^2 = \ddot{\psi}(q)/p^3. \qquad (6.24)$$

Therefore, the coefficient of skewness is

$$\beta_1 = \mu_3^2/\mu_2^3 = \ddot{\psi}^2(q)/\dot{\psi}^3(q). \qquad (6.25)$$

Then from page 260 of Abramowitz & Stegun (1964),

$$\dot{\psi}(x) \sim 1/x + o(1/x),$$
$$\ddot{\psi}(x) \sim -1/x^2 + o(1/x^2), \qquad (6.26)$$

for large x, and so, as $q\to\infty$,

$$\beta_1 \sim (-1/q^2)^2 \div (1/q)^3 = 1/q. \qquad (6.27)$$

6.1.3 Hyperbolic Secant Distribution (§3.4.2)

In this case, to make the computations easier let us use a double re-parametrization,

$$g = e^{x - \theta}, \qquad h = 1/(1 + e^{-g}), \tag{6.28}$$

and the derivatives, with respect to θ,

$$\dot{g} = -e^{x - \theta} = -g, \qquad \dot{h} = \dot{g}e^{-g}/(1 + e^{-g})^2 = g(h^2 - h). \tag{6.29}$$

Therefore,

$$\begin{aligned} f(x|\theta) &= e^{(x-\theta)-e^{x-\theta}}/[\log 2\cdot(1 + e^{-e^{x-\theta}})] \\ &= cge^{-g}/(1 + e^{-g}) = cg(1 - h), \end{aligned} \tag{6.30}$$

where $c = 1/\log 2$. Then, the derivatives in θ,

$$\begin{aligned} \text{(A)}\quad f^{(1)}(x|\theta) &= c[\dot{g}(1 - h) - g\dot{h}] \\ &= c[-g(1 - h) + g^2(h - h^2)], \end{aligned} \tag{6.31}$$

$$\begin{aligned} \text{(B)}\quad f^{(2)}(x|\theta) &= c[-\dot{g}\{(1-h)-2g(h-h^2)\}+\{g+g^2(1-2h)\}\dot{h}] \\ &= c[g(1-h)-3g^2(h-h^2)-g^3(h-3h^2+2h^3)], \end{aligned}$$

$$\begin{aligned} \text{(C)}\quad f^{(3)}(x|\theta) &= c[\dot{g}\{(1-h)-6g(h-h^2)-3g^2(h-3h^2+2h^3)\} \\ &\quad -\{(g+3g^2(1-2h)+g^3(1-6h+6h^2)\}\dot{h}] \\ &= c[-g(1-h)+7g^2(h-h^2)+6g^3(h-3h^2+2h^3) \\ &\quad +g^4(h-7h^2+12h^3-6h^4)], \end{aligned}$$

(D) $f^{(4)}(x|\theta) = c[-\dot{g}\{(1-h)-14g(h-h^2)-18g^2(h-3h^2$
$+2h^3)-4g^3(h-7h^2+12h^3-6h^4)\}$
$+\{g+7g^2(1-2h)+6g^3(1-6h+6h^2)$
$+g^4(1-14h+36h^2-24h^3)\}\dot{h}]$
$= c[g(1-h)-15g^2(h-h^2)-25g^3(h-3h^2+2h^3)$
$-10g^4(h-7h^2+12h^3-6h^4)-g^5(h-15h^2$
$+50h^3-60h^4+24h^5)]$,

(E) $f^{(5)}(x|\theta) = c[\dot{g}\{(1-h)-30g(h-h^2)-75g^2(h-3h^2+2h^3)$
$-40g^3(h-7h^2+12h^3-6h^4)-5g^4(h-15h^2$
$+50h^3-60h^4+24h^5)\} - \{g+15g^2(1-2h)$
$+25g^3(1-6h+6h^2)+10g^4(1-14h+36h^2$
$-24h^3)+g^5(1-30h+150h^2-240h^3+120h^4)\}\dot{h}]$
$= c[-g(1-h)+31g^2(h-h^2)+90g^3(h-3h^2+2h^3)$
$+65g^4(h-7h^2+12h^3-6h^4)+15g^5(h-15h^2$
$+50h^3-60h^4+24h^5)+g^6(h-31h^2+180h^3$
$-390h^4-360h^5)]$.

And hence the derivative ratios are

(A) $d_1 = -[1 - gh]$,

(B) $d_2 = [1 - 3gh - g^2h + 2g^2h^2]$,

(C) $d_3 = -[1-7gh-6g^2h+12g^2h^2-g^3h+6g^3h^2-6g^3h^3]$,(6.32)

(D) $d_4 = [1-15gh-25g^2h+50g^2h^2-10g^3h+60g^3h^2-60g^3h^3$
$-g^4h+14g^4h^2-36g^4h^3+24g^4h^4]$,

(E) $d_5 = -[1-31gh-90g^2h+180g^2h^2-65g^3h+390g^3h^2+390g^3h^3$
$-15g^4h+210g^4h^2-540g^4h^3+360g^4h^4-g^5h+30g^5h^2$
$-150g^5h^3+240g^5h^4-120g^5h^5]$.

Then, we also need the mean of $g^m h^n$ for $n \le m$,

$$
\begin{aligned}
\xi_{mn} &= E_\theta g^m h^n = \int_{-\infty}^{\infty} g^m h^n \, dF(x|\theta) \\
&= \int_{-\infty}^{\infty} c g^{m+1} h^n (1 - h) \, dx \\
&= \int_0^{\infty} c g^m h^n (1 - h) \, dg \\
&= \int_0^{\infty} c g^m e^{-g} dg/(1 + e^{-g})^n \\
&= \int_0^{\infty} c g^m e^{ng} dg/(1 + e^{g})^n \qquad (6.33) \\
&= c\left\{ -\frac{1}{n} g^m e^{(n-1)g} (1 + e^g)^{-n} \Big|_0^{\infty} \right. \\
&\qquad \left. + \frac{1}{n}\int_0^{\infty} [m g^{m-1} + (n-1) g^m] e^{(n-1)g} dg/(1 + e^g)^n \right\} \\
&= \frac{1}{n}[m\xi_{m-1,n-1} + (n-1)\xi_{m,n-1}] .
\end{aligned}
$$

So, from (23.2.8) of Abramowitz & Stegun (1964), p807,

$$
\begin{aligned}
\kappa_m &= \xi_{m0} = \int_0^{\infty} c g^m dg/(1 + e^g) \\
&= c\zeta(m+1)[1+2^{-m}]\Gamma(m+1) \qquad (6.34) \\
&= m!\eta(m+1)/\log 2 ,
\end{aligned}
$$

where $\zeta(x)$ is the Reimann Zeta function and $\eta(x)$ is the sum of the alternating series corresponding to $\zeta(x)$. We can then define each ξ_{mn} in terms of κ_i as in the following table.

TABLE 6.1: ξ_{mn} in terms of κ_i

$$\xi_{11} = \kappa_0 = 1 \qquad \xi_{21} = 2\kappa_1 \qquad \xi_{31} = 3\kappa_2$$

$$\xi_{41} = 4\kappa_3 \qquad \xi_{51} = 5\kappa_4 \qquad \xi_{61} = 6\kappa_5$$

$$\xi_{22} = (1 + \kappa_1) \qquad \xi_{32} = (3\kappa_1 + \tfrac{3}{2}\kappa_2) \qquad \xi_{42} = (6\kappa_2 + 2\kappa_3)$$

$$\xi_{52} = (10\kappa_3 + \tfrac{5}{2}\kappa_4) \qquad \xi_{62} = (15\kappa_4 + 3\kappa_5) \qquad \xi_{33} = (1 + 3\kappa_1 + \kappa_2)$$

$$\xi_{43} = (4\kappa_1+6\kappa_2+\tfrac{4}{3}\kappa_3) \qquad \xi_{53} = (10\kappa_2+10\kappa_3+\tfrac{5}{3}\kappa_4) \qquad \xi_{63} = (20\kappa_3+15\kappa_4+2\kappa_5)$$

$$\xi_{44} = (1+6\kappa_1+\tfrac{11}{2}\kappa_2+\kappa_3) \qquad \xi_{54} = (5\kappa_1+15\kappa_2+\tfrac{55}{6}\kappa_3+\tfrac{3}{2}\kappa_5)$$

$$\xi_{64} = (15\kappa_2+30\kappa_3+\tfrac{55}{4}\kappa_4+\tfrac{3}{2}\kappa_5) \qquad \xi_{55} = (1+10\kappa_1+\tfrac{35}{2}\kappa_2+\tfrac{25}{3}\kappa_3+\kappa_4)$$

$$\xi_{65} = (6\kappa_1+30\kappa_2+35\kappa_3+\tfrac{25}{2}\kappa_4+\tfrac{6}{5}\kappa_5) \qquad \xi_{66} = (1+15\kappa_1+\tfrac{85}{2}\kappa_2+\tfrac{75}{2}\kappa_3+\tfrac{137}{12}\kappa_4+\kappa_5)$$

Then from (6.32), and using Table 6.1 to simplify,

$$\begin{aligned}
\text{(A)}\quad \mu_{200} &= E_\theta d_1^2 = E_\theta(1 - gh)^2 \\
&= E_\theta(1 - 2gh + g^2h^2) \\
&= 1 - 2\xi_{11} + \xi_{22} = 1 - 2 + (1 + \kappa_1) = \kappa_1, \\
\text{(B)}\quad \mu_{300} &= E_\theta d_1^3 = -E_\theta(1 - gh)^3 \\
&= -E_\theta(1 - 3gh + 3g^2h^2 - g^3h^3) \\
&= 1 - 3\xi_{11} + 3\xi_{22} - \xi_{33} = 1-3+3(1+\kappa_1)-(1+3\kappa_1+\kappa_2) = -\kappa_2, \\
\text{(C)}\quad \mu_{110} &= E_\theta d_1 d_2 = -E_\theta(1-gh)(1-3gh-g^2h+2g^2h^2) \\
&= -E_\theta(1-4gh-g^2h+5g^2h^2+g^3h^2-2g^3h^3) \\
&= 1 - 4\xi_{11} - \xi_{21} + 5\xi_{22} + \xi_{32} - 2\xi_{33} \\
&= 1-4-2\kappa_1+5(1+\kappa_1)+(3\kappa_1+\tfrac{3}{2}\kappa_2)-2(1+3\kappa_1+\kappa_2) = -\tfrac{1}{2}\kappa_2, \\
\text{(D)}\quad \mu_{400} &= E_\theta d_1^4 = E_\theta(1 - gh)^4 \\
&= 1 - 4\xi_{11} + 6\xi_{22} - 4\xi_{33} + \xi_{44} \\
&= 1-4+6(1+\kappa_1)-4(1+3\kappa_1+\kappa_2)+(1+6\kappa_1+\tfrac{11}{2}\kappa_2+\kappa_3) = \tfrac{3}{2}\kappa_2 + \kappa_3,
\end{aligned}$$

$$(E)\quad \mu_{210} = 1-5\xi_{11}-\xi_{21}+9\xi_{22}+2\xi_{32}-7\xi_{33}-\xi_{43}+2\xi_{44}$$

$$= 1-5-2\kappa_1+9(1+\kappa_1)+2(3\kappa_1+\tfrac{3}{2}\kappa_2)-7(1+3\kappa_1+\kappa_2)$$

$$-(4\kappa_1+6\kappa_2+\tfrac{4}{3}\kappa_3)+2(1+6\kappa_1+\tfrac{11}{2}\kappa_2+\kappa_3) = \kappa_2 + \tfrac{2}{3}\kappa_3,$$

$$(F)\quad \mu_{101} = 1-8\xi_{11}-6\xi_{21}-\xi_{31}+19\xi_{22}+12\xi_{32}+\xi_{42}-18\xi_{33}-6\xi_{43}+6\xi_{44}$$

$$= 1-8-12\kappa_1-3\kappa_2+19(1+\kappa_1)+12(3\kappa_1+\tfrac{3}{2}\kappa_2)+(6\kappa_2+2\kappa_3)$$

$$-18(1+3\kappa_1+\kappa_2)-6(4\kappa_1+6\kappa_2+\tfrac{4}{3}\kappa_3)+6(1+6\kappa_1+\tfrac{11}{2}\kappa_2+\kappa_3) = \kappa_1,$$

$$(G)\quad \mu_{020} = 1-6\xi_{11}-2\xi_{21}+13\xi_{22}+6\xi_{32}+\xi_{42}-12\xi_{33}-4\xi_{43}+4\xi_{44}$$

$$= 1-6-4\kappa_1+13(1+\kappa_1)+6(3\kappa_1+\tfrac{3}{2}\kappa_2)+(6\kappa_2+2\kappa_3)-12(1+3\kappa_1$$

$$+\kappa_2)-4(4\kappa_1+6\kappa_2+\tfrac{4}{3}\kappa_3)+4(1+6\kappa_1+\tfrac{11}{2}\kappa_2+\kappa_3) = -\kappa_1+\kappa_2+\tfrac{2}{3}\kappa_3,$$

$$(H)\quad \mu_{500} = -1+5\xi_{11}-10\xi_{22}+10\xi_{33}-5\xi_{44}+\xi_{55}$$

$$= -1+5-10(1+\kappa_1)+10(1+3\kappa_1+\kappa_2)-5(1+6\kappa_1+\tfrac{11}{2}\kappa_2+\kappa_3)$$

$$+(1+10\kappa_1+\tfrac{35}{2}\kappa_2+\tfrac{25}{3}\kappa_3+\kappa_4) = \tfrac{10}{3}\kappa_3 + \kappa_4, \tag{6.35}$$

$$(I)\quad \mu_{310} = -1+6\xi_{11}+\xi_{21}-14\xi_{22}-3\xi_{32}+16\xi_{33}+3\xi_{43}-9\xi_{44}-\xi_{54}+2\xi_{55}$$

$$= -1+6+2\kappa_1-14(1+\kappa_1)-3(3\kappa_1+\tfrac{3}{2}\kappa_2)+16(1+3\kappa_1+\kappa_2)+3(4\kappa_1$$

$$+6\kappa_2+\tfrac{4}{3}\kappa_3)-9(1+6\kappa_1+\tfrac{11}{2}\kappa_2+\kappa_3)-(5\kappa_1+15\kappa_2+\tfrac{55}{6}\kappa_3+\tfrac{5}{4}\kappa_4)$$

$$+2(1+10\kappa_1+\tfrac{35}{2}\kappa_2+\tfrac{25}{3}\kappa_3+\kappa_4) = \tfrac{5}{2}\kappa_3 + \tfrac{3}{4}\kappa_4,$$

$$(J)\quad \mu_{120} = -1+7\xi_{11}+2\xi_{21}-19\xi_{22}-8\xi_{32}+25\xi_{33}-\xi_{42}+10\xi_{43}-16\xi_{44}$$

$$+\xi_{53}-4\xi_{54}+4\xi_{55} = -1+7+4\kappa_1-19(1+\kappa_1)-8(3\kappa_1+\tfrac{3}{2}\kappa_2)$$

$$+25(1+3\kappa_1+\kappa_2)-(6\kappa_2+2\kappa_3)+10(4\kappa_1+6\kappa_2+\tfrac{4}{3}\kappa_3)-16(1+6\kappa_1$$

$$+\tfrac{11}{2}\kappa_2+\kappa_3)+(10\kappa_2+10\kappa_3+\tfrac{5}{3}\kappa_4)-4(5\kappa_1+15\kappa_2+\tfrac{55}{6}\kappa_3+\tfrac{5}{4}\kappa_4)$$

$$+4(1+10\kappa_1+\tfrac{35}{2}\kappa_2+\tfrac{25}{3}\kappa_3+\kappa_4) = -\kappa_2 + 2\kappa_3 + \tfrac{2}{3}\kappa_4,$$

$$\text{(K)}\quad \mu_{201} = -1+9\xi_{11}+6\xi_{21}-27\xi_{22}+\xi_{31}-18\xi_{32}+37\xi_{33}-2\xi_{42}$$
$$+18\xi_{43}-24\xi_{44}+\xi_{53}-6\xi_{54}+6\xi_{55}$$
$$= -1+9+12\kappa_1-27(1+\kappa_1)+3\kappa_2-18(3\kappa_1+\tfrac{3}{2}\kappa_2)+37(1+3\kappa_1$$
$$+\kappa_2)-2(6\kappa_2+2\kappa_3)+18(4\kappa_1+6\kappa_2+\tfrac{4}{3}\kappa_3)-24(1+6\kappa_1$$
$$+\tfrac{11}{2}\kappa_2+\kappa_3)+(10\kappa_2+10\kappa_3+\tfrac{5}{3}\kappa_4)-6(5\kappa_1+15\kappa_2+\tfrac{55}{6}\kappa_3$$
$$+\tfrac{5}{4}\kappa_4)+6(1+10\kappa_1+\tfrac{35}{2}\kappa_2+\tfrac{25}{3}\kappa_3+\kappa_4) = 2\kappa_2 + \kappa_3 + \tfrac{1}{6}\kappa_4 ,$$

$$\text{(L)}\quad \mu_{011} = -1+10\xi_{11}+7\xi_{21}-35\xi_{22}+\xi_{31}-31\xi_{32}+56\xi_{33}-9\xi_{42}$$
$$+42\xi_{43}-42\xi_{44}-\xi_{52}+8\xi_{53}-18\xi_{54}+12\xi_{55}$$
$$= -1+10+14\kappa_1-35(1+\kappa_1)+3\kappa_2-31(3\kappa_1+\tfrac{3}{2}\kappa_2)+56(1+3\kappa_1$$
$$+\kappa_2)-9(6\kappa_2+2\kappa_3)+42(4\kappa_1+6\kappa_2+\tfrac{4}{3}\kappa_3)-42(1+6\kappa_1$$
$$+\tfrac{11}{2}\kappa_2+\kappa_3)-(10\kappa_3+\tfrac{5}{2}\kappa_4)+8(10\kappa_2+10\kappa_3+\tfrac{4}{3}\kappa_4)-18(5\kappa_1$$
$$+15\kappa_2+\tfrac{55}{6}\kappa_3+\tfrac{5}{4}\kappa_4)+12(1+10\kappa_1+\tfrac{35}{2}\kappa_2+\tfrac{25}{3}\kappa_3+\kappa_4)$$
$$= -\tfrac{1}{2}\kappa_2 + \kappa_3 + \tfrac{1}{3}\kappa_4 ,$$

$$\text{(M)}\quad \mu_{1001} = -1+16\xi_{11}+25\xi_{21}-65\xi_{22}+10\xi_{31}-85\xi_{32}+110\xi_{33}+\xi_{41}$$
$$-24\xi_{42}+96\xi_{43}-84\xi_{44}-\xi_{52}+14\xi_{53}-36\xi_{54}+24\xi_{55}$$
$$= -1+16+50\kappa_1-65(1+\kappa_1)+30\kappa_2-85(3\kappa_1+\tfrac{3}{2}\kappa_2)+110(1+3\kappa_1$$
$$+\kappa_2)+4\kappa_3-24(6\kappa_2+2\kappa_3)+96(4\kappa_1+6\kappa_2+\tfrac{4}{3}\kappa_3)-84(1+6\kappa_1$$
$$+\tfrac{11}{2}\kappa_2+\kappa_3)-(10\kappa_3+\tfrac{5}{2}\kappa_4)+14(10\kappa_2+10\kappa_3+\tfrac{5}{3}\kappa_4)-36(5\kappa_1$$
$$+15\kappa_2+\tfrac{55}{6}\kappa_3+\tfrac{5}{4}\kappa_4)+24(1+10\kappa_1+\tfrac{35}{2}\kappa_2+\tfrac{25}{3}\kappa_3+\kappa_4)$$
$$= \tfrac{5}{2}\kappa_2 - \tfrac{1}{6}\kappa_4 .$$

6.1.4 One parameter Normal (§3.4.3)

To compute the rate co-efficients for the one parameter Normal, $N(\theta,\theta^k)$, it seems to be easier to compute the log-density moments, η_{ijk}, rather than the Fisher moments (which can then always be obtained through Lemma 2.2). First we can make use of the well known moments of the $N(\mu,\sigma^2)$,

$$
\begin{aligned}
&(A)\quad EX = \mu = \theta,\\
&(B)\quad EX^2 = \mu^2 + \sigma^2 = \theta^2 + \theta^k,\\
&(C)\quad EX^3 = \mu^3 + 3\mu\sigma^2 = \theta^3 + 3\theta^{k+1},\\
&(D)\quad EX^4 = \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4 = \theta^4 + 6\theta^{k+2} + 3\theta^{2k}.
\end{aligned}
\tag{6.36}
$$

Therefore the log-density and its derivatives in θ are

$$
\begin{aligned}
&(A)\quad \ell = -\tfrac{1}{2}\log(2\pi\theta^k) - x^2/2\theta^k + x/\theta^{k-1} - \tfrac{1}{2}/\theta^{k-2},\\
&(B)\quad \dot{\ell} = -k/2\theta + kx^2/2\theta^{k+1} - (k-1)x/\theta^k + (k-2)/2\theta^{k-1},\\
&(C)\quad \ddot{\ell} = k/2\theta^2 - k(k+1)x^2/2\theta^{k+2} + k(k-1)x/\theta^{k+1} - (k-2)(k-1)/2\theta^k.
\end{aligned}
\tag{6.37}
$$

Hence, the Fisher information is

$$
\begin{aligned}
\mu_{200} &= E\dot{\ell}^2 = -E\ddot{\ell}\\
&= -[k/2\theta^2 - k(k+1)/2\theta^2 - k(k+1)/2\theta^k + 2k(k-1)/2\theta^k\\
&\qquad -(k-2)(k-1)/2\theta^k]\\
&= k^2/2\theta^2 + 1/\theta^k.
\end{aligned}
\tag{6.38}
$$

And, the 'scale-information' is

$$\begin{aligned}
J(\theta) &= Cov_\theta(\dot{\ell}, \ddot{\ell}) = E_\theta \dot{\ell}\ddot{\ell} \\
&= -k^2(k+1)/4\theta^{2k+3}\, EX^4 + k(k-1)(2k+1)/2\theta^{2k+2}\, EX^3 \\
&- k(3k^2-6k+2)/2\theta^{2k+1}\, EX^2 + k^2(k+2)/4\theta^{k+3}\, EX^2 \\
&+ (k-1)(k-2)(2k+1)/2\theta^{2k}\, EX - (k-1)(k-2)^2/4\theta^{2k-1} \\
&+ k^2(k-2)/4\theta^{k+1} - k^2/4\theta^3 - k(k-1)/2\theta^{k+2}\, EX \\
&= -2k/\theta^{k+1} - k^2(k+1)/2\theta^3. \qquad (6.39)
\end{aligned}$$

Hence for the exponential family members, $J(\theta)=0$ for $k=0$ (exponential translation family) and $J(\theta)=-2I(\theta)/\theta$ for $k=1$ (exponential scale family). Also,

$$E_\theta \ddot{\ell}^2 = 1/\theta^{2k} + 5k^2/\theta^{k+2} + (3k^4+4k^3+2k^2)/4\theta^4, \quad (6.40)$$

and so

$$\eta_{020} = E_\theta \ddot{\ell}^2 - \mu_{200}^2 = 4k^2/\theta^{k+2} + k^2(k+1)^2/2\theta^4. \quad (6.41)$$

Hence, the statistical curvature is

$$\begin{aligned}
\gamma_\theta^2 &= [I(\theta)\eta_{200} - J^2(\theta)]/I^3(\theta) \\
&= k^2(k-1)^2/[2I^3(\theta)\theta^{k+4}], \qquad (6.42)
\end{aligned}$$

which is zero only if $k=0$ or $k=1$, the only two values of k for which the distribution is of the exponential family.

6.2 Computer programs in BASIC

All computations were done on a Commodore PET 2001-32K micro-computer with algorithms written in BASIC. Since only computations and no simulation was involved, the 9-significant digit precision of the micro-computer was considered sufficient. Some of the programs were also tested on a mainframe computer (in FORTRAN) and yielded output in double precision which was identical in the 9-significant digits of the micro-computer. As there was no printer for the micro-computer, all output and program listings had to be reported manually. The major programs are thus listed below.

6.2.1 Direct Computation of the Exponential Rates

The following BASIC program computes the Bahadur bound and the exponential rates of the pre and mle directly from the definitions (2.42), Theorem 2.1 and Lemma 2.4. The program as presented considers the mixtures of Normals of Chapter 5, but can easily be modified for any other underlying distribution by changing the defining subroutine at line 140, and modifying other lines as necessary (such as 5 to 15).

```
1 DIM TM(5), TP(5), T(5), ML(5), PR(5), M(22),
   BB(6), CC(6), DD(6)
2 DATA .25, .45, .5, .55, .75
5 INPUT "P,A,R";P,A,R: Q=1-P: L2=SQR(2*π): R2=R*R:
   LR=LOG(R2)
10 DEF FNS(Z)=-Z*Z/2:LQ=LOG(Q/R): LP=LOG(P)
15 DEF FNP(Z)=INT(1000000*Z+.5)/1000000: DEF FNC(Z)=
   INT(ABS(.5-SGN(Z)))
17 X0=INT(A-9.3*R)-1: IF X0>-10 THEN X0=-10
18 X1=INT(A+9.3*R)+1: IF X1< 10 THEN X1= 10
20 INPUT "DX";DX: C=DX/L2: INPUT "MOMENTS";IT:
   IF IT=1 THEN GOSUB 200
25 INPUT "E,PRECISION";E,I: SR=10↑(-I)
30 FOR I=1 TO 5: READ TP(I): TM(I)=2*E*TP(I): NEXT:
   INPUT "DP,DM,R0";DP,DM,R0
32 INPUT "T";IT: IF IT=1 THEN FOR I=1 TO 5:
   INPUT "PR,ML";TP(I),TM(I): NEXT
35 K0=0: K1=0: RESTORE: FOR I=1 TO 5: ML(I)=0:
   PR(I)=0: NEXT
40 FOR X=X0 TO X1 STEP DX: GOSUB 140: FX=EXP(L):
   K0=K0+FX*(L-L0): K1=K1+FX*(L-L1)
45 FOR I=1 TO 5: ML(I)=ML(I)+EXP(L0+TM(I)*D1+DM)
50 PR(I)=PR(I)+EXP(L0-TP(I)*(R0-L1+L0)+DP): NEXT I,DX
55 K0=C*K0: K1=C*K1: RT=K0-K1: PRINT "K-L =" K0,K1:
   PRINT "RT =" RT: IF A=0 OR R=1 THEN RT=0
```

```
60 FOR I=1 TO 5: ML(I)=C*ML(I): PR(I)=C*PR(I)*
    EXP(TP(I)*(R0-RT)): NEXT
65 D=DP: PRINT "PRE": FOR I=1 TO 5: M(I)=PR(I):
    T(I)=TP(I): PRINT T(I) TAB(8); D-LOG(M(I)): NEXT
70 DEF FNE(Z)=EXP(L0-Z*(RT-L1+L0)+D): GOSUB 100
75 D=DM: PRINT "MLE": FOR I=1 TO 5: M(I)=ML(I):
    T(I)=TM(I): PRINT T(I) TAB(8); D-LOG(M(I)): NEXT
80 DEF FNE(Z)=EXP(L0+Z*D1+D): GOSUB 100
98 GO TO 25
99 END
```

With the subroutines to compute the Normal mixtures as follows:

```
140 Z=X: GOSUB 160: L=LN: D1=X*EXP(LA-L)+(X-A)*
     EXP(LB-L)/R2
145 Z=X+E: GOSUB 160: L0=LN: Z=X-E: GOSUB 160:
     L1=LN: RETURN
160 LA=FNS(Z)+LP: LB=FNS((Z-A)/R)+LQ: IF LA-LB>20
     THEN LN=LA: RETURN
165 LN=LB+LOG(1+EXP(LA-LB)): RETURN
170 LA=-1000: S=FNC(ZX): IF ZX<>0 THEN LA=LG+LOG(ABS(ZX))
175 LB=-900: SY=FNC(ZY): IF ZY<>0 THEN LB=LH+LOG(ABS(ZY))
180 IF LA-LB>20 THEN LN=LA: RETURN
185 IF S=SY THEN GOSUB 165: RETURN
190 IF LA>LB THEN LN=LA+LOG(1-EXP(LB-LA)): RETURN
195 S=SY: LN=LB+LOG(1-EXP(LA-LB)): RETURN
```

Subroutine to evaluate the infimum of the $\rho(t)$ in the rate expressions of the pre and mle, using a trisection method to obtain the optimum value of a convex or concave curve for a closed interval:

```
100 IF M(3)< M(2) THEN T(1)=T(2): M(1)=M(2)
105 IF M(3)< M(4) THEN T(5)=T(4): M(5)=M(4)
110 IF M(3)<=M(2) THEN T(5)=T(3): M(5)=M(3):
     T(3)=T(2): M(3)=M(2)
115 IF M(3)<=M(4) THEN T(1)=T(3): M(1)=M(3):
     T(3)=T(4): M(3)=M(4)
120 T(2)=.2*T(1)+.8*T(3): T(4)=.8*T(3)+.2*T(5)
122 DT=ABS(LOG(M(3))/LOG(M(1)))+ABS(LOG(M(3))/LOG(M(5)))
124 PRINT D-LOG(M(3)) DT: IF DT<SR THEN RETURN
125 M(2)=0: M(4)=0: FOR X=X0 TO X1 STEP DX: GOSUB 140
130 M(2)=M(2)+FNE(T(2)): M(4)=M(4)+FNE(T(4)): NEXT
135 M(2)=C*M(2): M(4)=C*M(4): GO TO 100
```

6.2.2 Computation of Fisher moments

Here we give the subroutine to compute the Fisher moments up to the sixth-degree, which are then used to obtain the coefficients of the Taylor series expansions of the Bahadur bound and the rates of the pre and mle, as given in §2.2-§2.4. This program is given for the Normal mixtures and requires the subroutines beginning at line 160 in §6.2.1

```
200 FOR X=X0 TO X1 STEP DX: Z=X: GOSUB 160:
     LG=LA-LN: LH=LB-LN: L=LN
205 PRINT FNP(X) TAB(12) FNP(EXP(LN)/L2) TAB(25) LN
210 X2=X*X: XA=(X-A)/R2: A2=XA*(X-A)
215 ZX=X: ZY=XA: GOSUB 170: D1=LN: G1=S: LH=LH-LR
218 ZX=X2-1: ZY=A2-1: GOSUB 170: D2=LN: G2=S
220 ZX=X*(X2-3): ZY=XA*(A2-3): GOSUB 170: D3=LN:
     G3=S: LH=LH-LR
222 ZX=X2*(X2-6)+3: ZY=A2*(A2-6)+3: GOSUB 170: D4=LN: G4=S
225 ZX=X*(X2*(X2-10)+15): ZY=xa*(A2*(A2-10)+15): GOSUB 170
227 M(22)=M(22)+(-1)↑(G1+S)*EXP(L+D1+LN)
230 J=0: FOR S=2 TO 6: FOR I4=0 TO S/5: S3=S-4*I4
235 FOR I3=0 TO S3/2-1: S2=S3-3*I3
240 FOR I2=0 to S2/2: I1=S2-2*I2: IF I2=S*S/4 GO TO 250
245 M(J)=M(J)+(-1)↑(I1*G1+I2*G2+I3*G3+I4*G4)*
     EXP(L+I1*D1+I2*D2+I3*D3+I4*D4): J=J+1
250 NEXT I2, I3, I4, S, X
```

```
252 FOR I=0 TO 22: M(I)=C*M(I): IF ABS(M(I))<1E-9
     THEN M(I)=0
253 NEXT: I=M(0): J=M(2)-M(1): PRINT "I ="I"J ="J
255 I2=I*I: I3=I*I2: K=J*J/I3:K1=2*M(3)-3*M(4)+M(6):
     K2=3*M(3)-6*M(4)+4*M(6)
260 K3=M(3)-2*M(4)+M(5): K4=K2-K1-M(6)
265 L1=2*M(7)-5*M(8)+3*M(9)+M(10)-M(11)
270 L2=7*M(7)-16*M(8)+6*M(9)+6*M(10)-2*M(11)-M(12)
275 L3=4*M(7)-10*M(8)+10*M(10)-5*M(12)
280 M1=5*M(13)-15*M(14)+12*M(15)-2*M(16)+4*M(17)
     -4*M(18)-M(20)+M(21)
285 M2=4*M(13)-12*M(14)+9*M(15)+4*M(17)-6*M(18)+M(19)
290 M3=5*M(13)-15*M(14)+20*M(17)-15*M(20)+6*M(22)
295 B3=-3*J-M(1): K0=K1/I2: G=K3/I2-1-K: D4=I2*G/8
300 IF J<>0 THEN J1=B3/J: J2=3+J1: J3=(K4/I+L2/J)/I
     -J1-J2: D=L1/I/J-J1+K0: D=-D
305 L=(M1/I-K2)/I2-2+K*(J3+K*J2): D3=I3*L/48
310 N=M2/I3-K0*K0-K*(J1*J1-3*D): D2=I3*N/72: D1=I*J*D/12
315 BB(2)=I/2: CC(2)=BB(2): DD(2)=BB(2): BB(3)=B3/6:
     CC(3)=BB(3): DD(3)=BB(3)
320 BB(4)=K2/24+K3/8: CC(4)=BB(4)-D4: DD(4)=CC(4)
325 BB(5)=(L1+L3/10)/12: CC(5)=BB(5): DD(5)=BB(5)-D1
330 BB(6)=M1/48+(M2+M3/10)/72: CC(6)=BB(6)-D3:
     DD(6)=CC(6)-D2
```

6.2.3 Exponential Rate of mle

For the mixure of Normals, the rate of the mle cannot be obtained from Lemma 2.4. Below is given the program for computing the rate directly, as described in §4.1:

```
5 INPUT "P,A,R";P,AA,R: NC=.5*LOG(2*π): R2=R*R:
   LP=LOG(P): LQ=LOG(1-P): LR=LP-LQ
10 DIM T(50): LT=LOG(10): INPUT "E,N";E,N: D0=1
15 DV=R2-1: AN=AA/N: MN=P*AA: LC=N*LQ: V=N*R2:
   FOR K=0 TO N: SE=N/SQR(V)
20 Z=SE*(E+MN): GOSUB 400: L0=L
25 Z=SE*(E-MN): GOSUB 400: L1=L: IF L1>L0 THEN
   L=L0:L0=L1:L1=L
30 IF K=0 THEN P0=LC+L0: P1=LC+L1: GO TO 40
35 P0=P0+LOG(1+EXP(LC+L0-P0))
38 P1=P1+LOG(1+EXP(LC+L1-P1)): IF K=N GO TO 45
40 MN=MN-AN: V=V-DV: LC=LC+LR+LOG((N-K)/(K+1))
45 NEXT: PRINT "N ="N: PRINT "P0 ="EXP(P0) TAB(20)
   "P1 ="EXP(P1)
46 IF P0<P1 THEN PN=P0: P0=P1: P1=PN
47 PN=P0+LOG(1+EXP(P1-P0))
48 IF PN>-75 THEN PRINT "PN ="EXP(PN): GO TO 50
49 EN=INT(PN/LT): PRINT "PN ="EXP(PN-EN*LT)"E"EN
50 AN=PN/N: T(I)=AN: I=I+1: D=DA-AN: RN:D/D0: D0=D
```

This last part of the mainline computes a lower bound by geometric extrapolation. In the computation, as we keep doubling n, we obtain a sequence which seems to converge from above (eventually). Hence, the sequence of consecutive differences is converging to zero from above. If r_n is the ratio of the n^{th} pair of consecutive differences, we can extrapolate the series of differences to infinity by approximating the unknown series for $i>n$ by a geometric series with ratio r_n. Our lower bound is then the last term plus the sum of the series.

```
50 AN=PN/N: T(I)=AN: I=I+1: D=DA-AN: RN=D/D0: D0=D: DA=AN
55 D=RN*D: AN=AN-D
60 IF D/AN>1E-10 GO TO 55
65 PRINT "GEOMETRIC EXTRAPOLATION ="AN; TAB(30)
    "RN ="RN
70 PRINT "UPPER BOUND ="DA
80 N=2*N: GO TO 15
99 END
```

To compute the Standard Normal probabilities, a method using expansions of the Normal probability integral into continued fractions was used [see Weiss (1981) for description of method]. Two continued fraction expansions are needed: one which converges rapidly for large deviates and one for small deviates. [26.2.14 & 26.2.15 of Abramowitz & Stegun (1964)].

```
400 IF Z=0 THEN L=-LOG(2): RETURN

402 T=Z: IF Z<0 THEN T=-Z

405 IF<T 2.75 THEN GOSUB 200: GO TO 415

410 GOSUB 100

415 IF Z<0 THEN L=LOG(1-EXP(L)): RETURN


100 A1=0: B1=1: A=1: B=T: J=1

105 A0=A1: B0=B1: A1=A: B1=B: L=A/B

110 A=T*A1+J*A0: B=T*B1+J*B0: J=J+1: IF B>1E30 GO TO 120

115 IF L*B/A<>1 GO TO 105

120 L=LOG(A/B)-NC-T*T/2: RETURN


200 A1=0: B1=1: A=T: B=1: J=1: S=-1: X=T*T

205 A0=A1: B0=B1: A1=A: B1=B: L=A/B: M=2*J+1

210 A=M*A1+S*J*X*A0: B=M*B1+S*J*X*B0: J=J+1: S=-S:
      IF ABS(B)>1E30 GO TO 220

215 IF L*B/A<>1 GO TO 205

220 L=LOG(.5-A/B*EXP(-NC-X/2)): RETURN
```

Notes: Returns log $\Phi(x)$. NC = .5*log(2*π) is defined in the mainline (line 5).

REFERENCES

[1] Abramowitz, M., & Stegun, I. A. (1964) *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables.* National Bureau of Standards, Government Printing Office. (Paperback by Dover, 1965)

[2] Amari, S. (1982a) Geometrical theory of asymptotic ancillarity and conditional inference. *Biometrika* 69: 1-17.

[3] Amari, S. (1982b) Differential Geometry of Curved Exponential Families. *Annals of Statistics* 10: 357-385.

[4] Bahadur, R. R. (1958) Examples of Inconsistency of Maximum Likelihood Estimates. *Sankhyā* 20: 207-210.

[5] Bahadur, R. R. (1960) On the Asymptotic Efficiency of Tests and Estimates. *Sankhyā* 22: 229-252.

[6] Bahadur, R. R. (1965) An Optimal Property of the Likelihood Ratio Statistic. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* 1: 13-26.

[7] Bahadur, R. R. (1967) Rates of Convergence of Estimates and Test Statistics. *Annals of Mathematical Statistics* 38: 303-324.

[8] Bahadur, R. R. (1971) *Some Limit Theorems in Statistics*. Society of Industrial and Applied Mathematics. Philadelphia.

[9] Bahadur, R. R., & Zabell, S. L. (1979) Large deviations of the sample mean in general vector spaces. *Annals of Probability* 7: 587-621.

[10] Bahadur, R. R., Gupta, J. C., & Zabell, S. L. (1980) Large Deviations, Tests, and Estimates. In *Asymptotic Theory of Statistical Tests and Estimation*, ed. by I.M.Chakravarti. Academic Press, NY.

[11] Basu, D. (1956) The Concept of Asymptotic Efficiency. *Sankhyā* 17: 193-196.

[12] Berkson, J. J. (1980) Minimum Chi-square, not Maximum Likelihood! (with discussants). *Annals of Statistics* 8: 457-488.

[13] Book, S. A. (1975) Convergence rates for a class of large deviation probabilities. *Annals of Probability* 3: 516-525.

[14] Chernoff, H. (1952) A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Annals of Mathematical Statistics* 23: 493-507.

[15] Chernoff, H. (1954) On the distribution of the Likelihood Ratio. *Annals of Mathematical Statistics* 25: 573-578.

[16] Chernoff, H. (1956) Large-sample theory: Parametric case. *Annals of Mathematical Statistics* 27: 1-22.

[17] Efron, B. E., & Truax, D. (1968) Large deviations in exponential families. *Annals of Mathematical Statistics* 39: 1402-1424.

[18] Efron, B. E. (1975) Defining the curvature of a statistical problem (with applications to second-order efficiency), (with discussants). *Annals of Statistics* 3: 1189-1242.

[19] Efron, B. E. (1978) The geometry of exponential families. *Annals of Statistics* 6: 362-376.

[20] Efron, B. E. (1982a) Transformation Theory: How Normal is a family of distributions? *Annals of Statistics* 10: 323-339.

[21] Efron, B. E. (1982b) Maximum Likelihood and Decision Theory. *Annals of Statistics* 10: 340-355.

[22] Feller, W. (1969) Limit Theorems for Probabilities of Large Deviations. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandter Gebiete* 14: 1-20.

[23] Fisher, R. A. (1922) On the Mathematical Foundations of Theoretical Statistics. *Philosophical Transactions of the Royal Society of London, Series A* 222: 309-368.

[24] Fu, J. C. (1971) *On some large-sample methods in statistical point estimation*. Unpublished doctoral dissertation, Johns Hopkins University. Baltimore.

[25] Fu, J. C. (1973) On a theorem of Bahadur on the rate of convergence of point estimators. *Annals of Statistics* 1: 745-749.

[26] Fu, J. C. (1975) The rate of convergence of consistent point estimators. *Annals of Statistics* 3: 234-240.

[27] Fu, J. C. (1980) Large Deviations, Bahadur Efficiency, and the rate of convergence of linear function of order statistics. *Bulletin of the Institute of Mathematics Academia Sinica* 8: 15-37.

[28] Fu, J. C. (1981) Large-sample point estimation: A Large Deviation Theory approach. *Annals of Statistics* 10: 762-771.

[29] Fu, J. C. (1983) Large sample point estimation for the location parameter: Neyman-Pearson Lemma or Maximum Likelihood Estimator. (To appear)

[30] Ghosh, J. K., & Subramanyam, K. (1974) Second-order Efficiency of Maximum Likelihood Estimators. *Sankhyā, series A* 36: 325-358.

[31] Ghosh, J. K., Sinha, B. K., & Wieand, H. (1980) Second-order efficiency of the MLE with respect to any bounded bowl-shaped loss function. *Annals of Statistics* 8: 506-521.

[32] Groeneboom, P., & Oosterhoff, J. (1977) Bahadur efficiency and the probabilities of large deviations. *Statistica Neerlandica* 31: 1-24.

[33] Groeneboom, P., Oosterhoff, J., & Ruymgaart, F. H. (1979) Large Deviation Theorem for Empirical Probability Measures. *Annals of Probability* 7: 553-586.

[34] Hodges, J. L., & Lehmann, E. L. (1970) Deficiency. *Annals of Mathematical Statistics* 41: 783-801.

[35] Hoeffding, W. (1965) On Probabilities of Large Deviations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* 1: 203-219.

[36] Hoeffding, W. (1979) *Theory of Estimation and Hypothesis Testing*. Lecture notes for a course given at the University of North Carolina. Chapel Hill. (Issued as a technical report.)

[37] Huber, P. J. (1967) Robust Confidence Limits. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandter Gebiete* 10: 269-278.

[38] Huber, P. J. (1972) Robust Statistics: A review. *Annals of Mathematical Statistics* 43: 1041-1067.

[39] Huber, P. J. (1977) *Robust Statistical Procedures*. Society for Industrial and Applied Mathematics. Philadelphia.

[40] Kallenberg, W. C. (1982) Chernoff Efficiency and Deficiency. *Annals of Statistics* 10: 583-594.

[41] Kallianpur, G., & Rao, C. R. (1955) On Fisher's Lower Bound to Asymptotic Variance of a Consistent Estimator. *Sankhyā* 15: 331-342.

[42] Kass, R. E. (1982) Canonical Parameterization on Zero Parameter-effects Curvature. Technical report, Department of Statistics, Carnegie-Mellon University. Pittsburgh.

[43] Kester, A. (1981) Large Deviation Optimality of MLE's in Exponential Families. Technical Report #160, Mathematics Department, Amsterdam Free University. Amsterdam.

[44] Killeen, T., Hettmansperger, T., & Sievers, G. L. (1972) An elementary theorem on the probability of large deviations. *Annals of Mathematical Statistics* 43: 181-192.

[45] Krafft, O., & Plachky, D. (1970) Bounds for the power of Likelihood Ratio Tests and their asymptotic properties. *Annals of Mathematical Statistics* 41: 1646-1654.

[46] Kullback, S., & Leibler, R. A. (1951) On Information and Sufficiency. *Annals of Mathematical Statistics* 22: 79-86.

[47] Madsen, L. T. (1979) The geometry of the statistical model—a generalization of curvature. Research Report #79-1, Statistical Research Unit, Danish Medical Research Council. Copenhagen.

[48] Pfanzagl, J., & Wefelmeyer (1978) A Third-order Optimum Property of the Maximum Likelihood Estimator. *Journal of Multivariate Analysis* 8: 1-29.

[49] Raghavachari, M. (1970) On a theorem of Bahadur on the rate of convergence of test statistics. *Annals of Mathematical Statistics* 41: 1659-1699.

[50] Rao, C. R. (1961) Asymptotic Efficiency and Limiting Information. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* 1: 531-545.

[51] Rao, C. R. (1962) Efficient Estimates and Optimum Inference Procedures in Large Samples. *Journal of the Royal Statistical Society, series B* 24: 46-72.

[52] Rao, C. R. (1963) Criteria of Estimation in Large Samples. *Sankhyā* 25: 189-206.

[53] Rubin, H., & Sethuraman, J. (1965) Bayes Risk Efficiency. *Sankhyā* 27: 325-346.

[54] Rubin, H., & Rukhin, A. (1982) Convergence Rates of Large Deviations Probabilities for Point Estimators. (To appear in *Statistics and Probability Letters*).

[55] Sanov, I. N. (1957) On the Probability of Large Deviations of Random Variables. *Selected Translations in Mathematical Statistics and Probability* 1: 213-244.

[56] Sethuraman, J. (1970) Probabilities of Deviations. Chapter 34 of *Essays in Probability and Statistics*, edited by R. C. Bose, et al., University of North Carolina Press. Chapel Hill.

[57] Sievers, G. L. (1969) On the probability of large deviations and exact slopes. *Annals of Mathematical Statistics* 40: 1908-1921.

[58] Sievers, G. L. (1978) Estimates of Location: A Large Deviation Comparison. *Annals of Statistics* 6: 610-618.

[59] Stigler, S. M. (1977) Do robust estimators work with *real* data? *Annals of Statistics* 5: 1055-1098.

[60] Weiss, G. M. T. (1981) Review of *Using Continued Fractions for Micro-computing Standard Distributions* by J. H. Klotz. Statistical Association of Manitoba Newsletter. Winnipeg.

[61] Wieand, H. (1976) A condition under which Pitman and Bahadur Efficiencies coincide. *Annals of Statistics* 4: 1003-1011.

VITAE

Günter Max Theodor Weiss was born in Walsum-am-Niederrhein on the 16th of June in 1953, the son of Heinz Herbert Otto Weiss and Elfriede Gertrude *neé* Lempke. In 1956 the family emigrated to Canada, later becoming Canadian citizens. Günter and his family settled in Winnipeg, where he attended Elmwood High School and graduated in 1971. In the fall of 1971, Günter was admitted to the University of Manitoba. He received his BSc(Hons) with first class honours, in May, 1975. Günter was a sessional lecturer in statistics from 1976-1981, and received his MSc in Statistics in May, 1978. In 1979, he was admitted to the PhD program in Statistics, and the following year was awarded a University of Manitoba Fellowship, which he held until the completion of his program in 1983. He is currently employed as a sessional lecturer in Statistics at the University of Manitoba.