Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

The CWI is a research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a nonprofit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

CWI Tract 18

Some large deviation results in statistics

A.D.M. Kester

Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

1980 Mathematics Subject Classification:
Primary: 62F10, 62F12, 62F03, 62F05
Secondary: 60F10, 60F05
ISBN 90 6196 289 7

Printed in the Netherlands

FOREWORD

The research leading to this treatise was performed at the Wiskundig Seminarium and at the Vakgroep Medische Statistiek of the Vrije Universiteit, Amsterdam.

Professor Dr. J. Oosterhoff suggested the subject of the second chapter, Dr. W.C.M. Kallenberg both posed the problem of the third chapter and provided the key idea for its major theorem. I would like to thank both also for their unfailing scrutiny of the manuscript.

Professors R.R. Bahadur, J.C. Fu and A.L. Rukhin kindly sent me preprints of their papers.

I am indebted to the CWI for the opportunity to publish this work in its present form.

April 1985,
Arnold Kester.

CONTENTS

CHAPTER I

INTRODUCTION

1. GENERAL INTRODUCTION AND SUMMARY

This treatise concerns two topics, both involving large deviations. In Chapter II, probabilities of gross errors (large deviations) of point estimates are considered and in Chapter III we determine the Bahadur efficiency and deficiency of classical two-sample conditional tests in exponential families. The present chapter contains an introduction to both subjects, a summary of the results and some more technical preliminaries. We start with an outline of the estimation problem.

Let $X_1, X_2, \ldots$ be a sequence of independent random variables with a common probability distribution P_θ on $\mathbb{R}^k$, where P_θ is a member of a parametric family $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ (in the initial two sections of Chapter II a somewhat greater generality is allowed). For each $n = 1,2,\ldots$ let an estimate $T_n = T_n(X_1,\ldots,X_n)$ of $g(\theta)$ be given, where g is a map of Θ into $\mathbb{R}^d$. Examples of g's of interest are $g(\theta) = E_\theta X_1$, $g(\theta) = \mathrm{var}_\theta\, X_1$ or $g(\theta) = \inf\{x : P_\theta((-\infty,x]) \geq \frac{1}{2}\}$ when $k = 1$, in case $\theta = (\theta^{(1)},\ldots,\theta^{(d)}) \in \mathbb{R}^d$ one might take $g(\theta) = \theta$ or $g(\theta) = \theta^{(i)}$.

The quality of T_n is often measured by its (normalized) expected quadratic loss

$$(1.1) \qquad n \cdot E_\theta \|T_n - g(\theta)\|^2$$

or, since (1.1) may be hard to obtain, by the variance (or covariance matrix) of the limit distribution of $n^{\frac{1}{2}}(T_n - g(\theta))$. This variance is then compared with the Cramér-Rao bound. Note that the limit distribution of $n^{\frac{1}{2}}(T_n - g(\theta))$ allows a sensible approximation of the probability of errors

$$(1.2) \qquad P_\theta(\|T_n - g(\theta)\| > \varepsilon_n)$$

when $\varepsilon_n = O(n^{-\frac{1}{2}})$ as $n \to \infty$. We aim to take, however, the probability of gross errors

$$(1.3) \qquad P_\theta(\|T_n - g(\theta)\| > \varepsilon)$$

for *fixed* values of ε as a basis for comparison of different estimates of $g(\theta)$, a point of view first taken by Basu (1956), Bahadur (1960 b) and Huber (1968). Just like the expected quadratic loss (1.1), the probability of gross errors (1.3) - also called the inaccuracy function - can often not be evaluated explicitly and one considers an asymptotic expression instead.

For consistent (sequences of) estimates $\{T_n\}$, the inaccuracy function tends to zero as $n \to \infty$, usually exponentially fast (when $\varepsilon > 0$ is fixed); we take the exponential rate of convergence

$$(1.4) \qquad -\limsup_{n\to\infty} n^{-1} \log P_\theta(\|T_n - g(\theta)\| > \varepsilon),$$

coined *inaccuracy rate* by Sievers (1978), as a yardstick to compare estimates of $g(\theta)$: the larger the inaccuracy rate the better the estimate.

In Section II.1 a bound $b(\varepsilon,\theta)$ on the inaccuracy rate of consistent estimates originally due to Bahadur (1960 b) is discussed. Bahadur's bound $b(\varepsilon,\theta)$ plays a similar role for the inaccuracy rate as the Cramér-Rao bound for the variance of the limit distribution. An obvious target of this study is to determine (consistent) estimates which attain Bahadur's bound, if they exist. Since inaccuracy rates (1.4) and bounds $b(\varepsilon,\theta)$ are often hard to evaluate and small ε's are particularly important, most authors (Bahadur (1960 b, 1967, 1971, 1983), Fu (1973, 1975, 1982), Perng (1978)) concentrate on the behaviour of the inaccuracy rate and $b(\varepsilon,\theta)$ as $\varepsilon \to 0$. In sufficiently smooth one-parameter families with $g(\theta) = \theta$ it holds that

$$b(\varepsilon,\theta) = \tfrac{1}{2}\varepsilon^2 i_\theta + o(\varepsilon^2) \quad \text{as } \varepsilon \to 0,$$

where i_θ denotes the Fisher information. Bahadur (1960 b, 1967) proved that the inaccuracy rate of the maximum likelihood estimate (MLE) equals $b(\varepsilon,\theta)$ to first order as $\varepsilon \to 0$. This result was extended to certain maximum probability estimates by Fu (1973) and, in k-parameter families, to regular best asymptotic normal estimates by Perng (1978). Perng's result includes

MLE's in k-parameter exponential families. In contrast, we shall mainly study inaccuracy rates for fixed $\varepsilon > 0$.

In Section II.2 a simple but important sufficient condition (II.2.2) for an estimate to attain Bahadur's bound is seen to fail in families which are not "*exponentially convex*", see Section 2a and Lemma II.2.2, giving a partial explanation for the elusiveness of estimates which attain the bound. In Example II.2.2 Bahadur's bound is proved to be unattainable.

Section II.3 treats estimation in *exponential families*. Exponential convexity is here equivalent to a convexity condition on the parameter space, and indeed, for exponentially convex exponential families the MLE generally attains Bahadur's bound when ε is not too large, extending Perng's (1978) result for infinitesimal ε's to the present fixed-ε case. *Curved exponential families* are considered next as an important subclass of the non-exponentially convex exponential families. Here the nice result of the convex case is not available, though examples are given where the bound is attained at a single θ or for a single $\varepsilon > 0$. Therefore we slightly change tack and investigate only estimates which satisfy a natural linearity restriction. In that class of estimates the MLE turns out to have the best inaccuracy rate for each sufficiently small, but fixed ε.

In the final section of Chapter II we look at *shift families* on the real line having Lebesgue densities. There are only a few essentially different exponentially convex shift families (examples II.4.1-3); in these families the MLE attains Bahadur's bound. For other shift families the bound is not often attained for all θ's simultaneously, but examples are given (examples II.4.4 and II.4.11) where a consistent estimate attains the bound for each $\varepsilon > 0$ and at a single θ. In shift families it is natural to restrict attention to (translation) equivariant estimates, an estimate T_n being equivariant when $T_n(X_1+c,\dots,X_n+c) = T_n(X_1,\dots,X_n) + c$ for each $c \in \mathbb{R}$. Sievers (1978) noted that when $p(x-\varepsilon) / p(x+\varepsilon)$ is nondecreasing, where p is the density of P_0, an equivariant estimate T_n^{ε} exists which minimizes for fixed n

$$P_\theta(|T_n - \theta| > \varepsilon)$$

over the class of equivariant estimates, cf. also Huber (1968).

Thus, the inaccuracy rate of equivariant estimates is bounded by the inaccuracy rate of $\{T_n^{\varepsilon}\}$ (Sievers (1978), Thm. 2.1), when p satisfies the condition above.

For a wider class of shift families we derive a bound which coincides with Sievers' bound when p satisfies his condition. When p is sufficiently smooth and ε is small enough, an M-estimate is constructed which attains this bound. In constrast to Sievers we employ a typical large deviation approach. Furthermore, we prove these estimates to be essentially unique in the class of M-estimates and we give a necessary condition for Gâteaux-differentiable estimates to attain the bound. An example is given where a Gâteaux-differentiable L-estimate (a trimmed mean) indeed attains the bound in the double exponential shift family. Remarkably, Sievers' bound can be *higher* than Bahadur's bound. The reason is that Sievers' bound concerns *equivariant* estimates, which are not necessarily *consistent* (examples II.4.2-3). When p is symmetric and sufficiently regular though, Sievers' estimate $\{T_n^{\varepsilon}\}$ is consistent and hence Sievers' bound is not larger than Bahadur's.

In the last part of Section II.4 we show by an example that the approach of Sievers cannot easily be generalized to location-scale families.

In Chapter III we consider, in a full one-parameter exponential family $\{P_\theta : \theta \in \Theta^*\}$, the problem of testing the hypothesis $\theta_1 = \theta_2$ against the alternative $\theta_1 > \theta_2$ on the basis of samples $X_1,\ldots,X_m$ from P_{θ_1} and $Y_1,\ldots,Y_n$ from P_{θ_2}. Without loss of generality we assume that the family is in its canonical form, cf. Section 2a, implying that $\Sigma_{i=1}^m X_i$ and $\Sigma_{i=1}^n Y_i$ are sufficient for θ_1 and θ_2.

The uniformly most powerful unbiased (UMPU) test for the above testing problem is the - possibly randomized - conditional test δ_c defined by means of the conditional distribution of ΣX_i given $\Sigma X_i + \Sigma Y_i$, cf. Lehmann (1959).

EXAMPLE 1.1. Let X and Y have binomial distributions with parameters (m,p_1) and (n,p_2), respectively (X and Y are sums of independent Bernoulli random variables). Fisher's exact test, randomized so that the conditional size equals α, is UMPU for $p_1 = p_2$ against $p_1 > p_2$.

The main purpose of this chapter is to determine how much the power of the conditional test falls short of the envelope power, in terms of *Bahadur efficiency* and *deficiency*. The envelope power is determined by most powerful (MP) tests against simple alternatives, hence the difference in power describes how much we lose by not knowing the nuisance parameter. We shall now briefly introduce Bahadur efficiency and deficiency for the two-sample case; more general introductions can be found in Bahadur (1971) and in Groeneboom and Oosterhoff (1981). The efficiency concept was originated by Bahadur (1960 a).

Let $\{m_N\}$ and $\{n_N\}$ be nondecreasing sequences of integers such that $m_N + n_N = N$ for each N, and define $N_+(\alpha,\beta,\theta_1,\theta_2)$ to be the smallest number N such that the MP test of $H = \{(\theta,\theta) : \theta \in \Theta^*\}$ against $\{(\theta_1,\theta_2)\}$ of size α has power at least β in (θ_1,θ_2). Let $N_c(\alpha,\beta,\theta_1,\theta_2)$ be similarly defined for the conditional test.

The *Bahadur efficiency* of the conditional test versus the MP test is defined as

$$e^B = \lim_{\alpha\to 0} N_+(\alpha,\beta,\theta_1,\theta_2) \,/\, N_c(\alpha,\beta,\theta_1,\theta_2),$$

keeping the other parameters fixed (the more classical Pitman efficiency may be defined as

$$e^P = \lim_{(\theta_1,\theta_2)\to(\theta,\theta)} N_+(\alpha,\beta,\theta_1,\theta_2) \,/\, N_c(\alpha,\beta,\theta_1,\theta_2),$$

keeping α and β fixed).

The Bahadur efficiency turns out to equal 1 in our testing problem, hence it is useful to consider the speed of convergence by looking at the difference of the required sample sizes $N_c - N_+$, named *deficiency* by Hodges and Lehmann (1970).

Bahadur deficiency of order $\mathcal{O}(D(N))$ is defined by

$$\frac{N_c - N_+}{D(N_+)} = \mathcal{O}(1) \qquad \text{as } \alpha \to 0,$$

where β, θ_1 and θ_2 are kept fixed. Bahadur deficiency of order $o(D(N))$ is similarly defined. Deficiency of order $\mathcal{O}(1)$ is also called bounded deficiency. Note that $e^B = 1$ is equivalent to Bahadur deficiency of order $o(N)$.

Bahadur (1965,1971) proved, under general conditions in a one-sample situation, that the likelihood ratio (LR) test has Bahadur efficiency 1 with respect to the MP test. Kallenberg (1978,1981) proved the LR test to have Bahadur deficiency (with respect to the MP test) of order $\mathcal{O}(\log N)$ in k-parameter exponential families, and $\mathcal{O}(1)$ in one-parameter exponential families. In the normal location-scale family however, the t-test was found to have bounded deficiency, too (Kallenberg (1981)). Note that the t-test is actually UMPU. The present situation is somewhat similar in that $\theta_1 - \theta_2 = 0$ is to be tested against $\theta_1 - \theta_2 > 0$, where θ_2 is a nuisance parameter like the variance in the t-test.

We shall prove, for one-parameter exponential families with either a lattice distribution or a Lebesgue density satisfying some regularity conditions, that *the Bahadur deficiency of the conditional test with respect to the MP test is bounded.* Furthermore, an explicit upper bound for the asymptotic deficiency will be obtained for families with a density. Boundedness of the deficiency implies, of course, that the *Bahadur efficiency equals* 1.

For the binomial distribution of example 1.1, the theoretical results are complemented by computer calculations. For some selected values of (p_1,p_2), the power β and the sample size N_+, an upper bound on the deficiency $N_c - N_+$ has been determined. This upper bound turns out to be low and remarkably constant as a function of N_+.

Michel (1979) investigated the asymptotic power of the (one sample) conditional test in a k+1-parameter exponential family, where the hypothesis concerns the first parameter only, the other k being nuisance parameters. For local alternatives he proved that this UMPU test has the same power up to $\mathcal{O}(N^{-1})$ as a test which is MP in a larger class of not necessarily unbiased tests. Using the methods of the present study one may prove in Michel's testing situation, but for fixed alternatives and $\alpha \to 0$, that the conditional test has the same power up to $\mathcal{O}(N_+^{-\frac{1}{2}})$ as the MP test.

Albers (1974) determined the *Pitman deficiency* of the conditional test with sample ratio $m_N / N = \frac{1}{2}$ versus the same test with the Pitman-optimal ratio. He proved that the Pitman-optimal ratio tends to $\frac{1}{2}$ and that the deficiency is *bounded*. Thus, for local alternatives the ratio $m_N / N = \frac{1}{2}$ is almost optimal. For fixed alternatives, however, the situation is different.

The Bahadur-optimal sample ratio converges to a number ν_0 which depends on the alternative and is in general not equal to $\frac{1}{2}$. Moreover, as we shall prove in Section III.7, the *Bahadur efficiency* of the conditional test with ratio $\frac{1}{2}$ versus the conditional test with the Bahadur-optimal ratio is in general *smaller than one*, in contrast to Albers' result.

We conclude this section with some remarks on the notation. In references to a relation, theorem, etc., within a chapter, the chapter number is omitted. Probability measure(s) is frequently abbreviated to pm(s), P-almost surely is denoted [P]; almost everywhere will refer to Lebesgue measure and is abbreviated to a.e. Furthermore, convergence in distribution and in probability under a pm P are denoted $\xrightarrow{\mathcal{D}_P}$ and $\xrightarrow{P}$, and finally, the end (or omission) of a proof will be indicated by □.

2. GENERAL PRELIMINARIES

2a. Exponential families

A d-dimensional random vector Y is distributed according to a k-parameter exponential family when its densities with respect to a σ-finite measure ν on $\mathbb{R}^d$ are of the form

$$dP_\theta^Y(y) = c(\theta) \exp\left\{\sum_{j=1}^{k} Q_j(\theta)T_j(y)\right\} h(y)d\nu(y), \tag{2.1}$$

where c(θ) is the norming constant and the functions T_j, $j = 1,\ldots,k$ and h are assumed to be ν-measurable.

EXAMPLE 2.1. The binomial distribution has density with respect to counting measure

$$\binom{n}{y} p^y (1-p)^{n-y} = (1-p)^n \exp\left\{y \log \frac{p}{1-p}\right\}\binom{n}{y}.$$

EXAMPLE 2.2. The normal $N(\mu,\sigma^2)$ distribution has Lebesgue density

$$\{2\pi\sigma^2\}^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}(y-\mu)^2/\sigma^2\}$$

$$= \{2\pi\sigma^2\}^{-\frac{1}{2}} \exp\left\{-\frac{\mu^2}{2\sigma^2}\right\} \exp\left\{\frac{\mu}{\sigma^2}y - \frac{1}{2\sigma^2}y^2\right\}.$$

Since P_θ^Y depends on θ only through the functions Q_j, a more natural parametrization has the form (absorbing the factor h into ν)

$$dP_\theta^Y(y) = c(\theta) \exp\left\{\sum_1^k \theta_j T_j(y)\right\} d\nu(y), \qquad \theta \in \Theta \subset \mathbb{R}^k. \tag{2.2}$$

Since the vector $T = (T_1(Y),\ldots,T_k(Y))'$ is sufficient for θ, all statistical inference is based on T, whence we may as well consider the densities of T itself, which can be written as

$$(2.3) \qquad dP_\theta^T(t) = c(\theta) \exp \{\Sigma\, \theta_j t_j\} d\nu^T(t), \quad \theta \in \Theta,$$

the (σ-finite) measure ν^T being induced by the map T, cf. Witting (1966), p. 57.

Applying vector notation and changing the name of the statistic to X we obtain the representation which shall be employed in the sequel:

$$(2.4) \qquad dP_\theta^X(x) = \exp \{\theta' x - \psi(\theta)\} d\mu(x), \quad \theta \in \Theta,$$

where

$$(2.5) \qquad \psi(\theta) = \log \int e^{\theta' x} d\mu(x)$$

and where Θ is a subset of the *full parameter space*

$$\Theta^* = \{\theta \in \mathbb{R}^k : \int e^{\theta' x} d\mu(x) < \infty\}.$$

As is well known, Θ^* is convex. Throughout this thesis we will assume that Θ^* has a non-empty interior with respect to the Euclidean topology on $\mathbb{R}^k$, denoted int Θ^*, and that the measure μ is not supported on a flat of dimension lower than k. Barndorff-Nielsen (1978) has called such a representation of the family $\{P_\theta : \theta \in \Theta^*\}$ "full canonical" and the statistic X "minimal canonical". As noted by Berk (1972), these assumptions do not restrict generality since they can be met by transforming and / or reparametrizing to lower dimensional subspaces. The assumptions also imply that ψ is strictly convex on Θ^* and hence continuous on int Θ^*. When $\theta_0 \in$ int Θ^*,

$$dP_\theta^X(x) = \exp \{(\theta - \theta_0)' x - \psi(\theta) + \psi(\theta_0)\} dP_{\theta_0}^X(x)$$

holds for each $\theta \in \Theta^*$, hence it is not a restriction to assume that μ is a probability measure and that $0 \in$ int Θ^*.

<u>EXAMPLE 2.1. (continued)</u>. The binomial distribution has density

$$dP_\theta^X(x) = \exp \left\{\theta x + n \log \frac{2}{1+e^\theta}\right\} dP_0^X(x)$$

where P_0 is the binomial $B(n,\frac{1}{2})$ distribution and $\theta = \log (p/(1-p))$.

EXAMPLE 2.2. (continued). For the normal $N(\mu,\sigma^2)$ distribution, the density of the minimal canonical statistic $(X_1,X_2) = (Y,Y^2)$ can be written as

$$dP_\theta^X(x) = \exp \left\{\theta_1 x_1 + \theta_2 x_2 - \left[\tfrac{1}{4}\theta_1^2 / (\tfrac{1}{2} - \theta_2) - \tfrac{1}{2} \log (1 - 2\theta_2)\right]\right\} dP_0^X(x),$$

where P_0^X is the distribution of X for $\mu = 0$, $\sigma^2 = 1$. The relation of (μ,σ^2) and (θ_1,θ_2) is given by $\theta_1 = \mu/\sigma^2$, $\theta_2 = \frac{1}{2} - 1/(2\sigma^2)$. Note that $\Theta^* = \{(\theta_1,\theta_2) : -\infty < \theta_1 < \infty, -\infty < \theta_2 < \frac{1}{2}\}$.

Let

$$\Theta^1 = \{\theta \in \Theta^* : E_\theta||X|| < \infty\},$$

where $||\cdot||$ denotes Euclidean norm, then int $\Theta^* \subset \Theta^1$ (Berk (1972)). On Θ^1 we define

$$\lambda(\theta) = E_\theta X.$$

The map λ is 1-1 on Θ^1 (Berk (1972)); its inverse λ^{-1} is defined on $\Lambda = \lambda(\Theta^1)$. For $\theta \in$ int Θ^* we have $\lambda(\theta) = \text{grad}\, \psi(\theta)$ and the covariance matrix Σ_θ is equal to the matrix of second order derivatives of ψ.

Furthermore, since μ is not supported on a flat, Σ_θ is positive definite on int Θ^*. Since moments of all orders exist on int Θ^* (Lehmann (1959)), each moment $E_\theta||X||^j$ is uniformly bounded for $\theta \in A$ when A is a compact subset of int Θ^*.

Finally, when $(X_1,X_2,\ldots,X_n)$ is a random sample from P_θ, and $0 \in$ int Θ^*, the distribution of $\overline{X}_n = \frac{1}{n} \Sigma_{i=1}^n X_i$ is given by

$$dP_\theta^{(n)}(x) = \exp \{n\theta' x - n\psi(\theta)\} d\mu^{(n)}(x),$$

where $P_\theta^{(n)}$ is the distribution of $\overline{X}_n$ and $\mu^{(n)} = P_0^{(n)}$.

For one-parameter exponential families, we list some additional properties. Since Θ^* is convex, it is a possibly infinite interval in $\mathbb{R}$. Kallenberg (1978), Lemma 2.2.1, proved that ψ is continuous on Θ^* and that $\lambda(\theta) = E_\theta X$ is properly defined and continuous on Θ^*, considered as a mapping into the extended real line. Moreover, λ is increasing on Θ^*; its inverse λ^{-1} is increasing and continuous on Λ. The second derivative of ψ is denoted σ^2.

To conclude this subsection we introduce a special one-parameter exponential family. Let P^* be the class of all pms on a set X, let $P,Q \in P^*$ and let μ be a σ-finite measure dominating both P and Q. Furthermore, let p and q be the μ-densities of P and Q and define $Y = \{x \in X : p(x)\cdot q(x) > 0\}$.

DEFINITION 2.1. *The exponential family between* P *and* Q, $P^{P,Q} = \{R_\alpha : \alpha \in [0,1]\}$ is defined by its μ-densities $\{r_\alpha\}$,

$$r_\alpha(x) = \exp\{\alpha \log \frac{q(x)}{p(x)} - \psi^{P,Q}(\alpha)\}p(x)1_Y(x), \quad \alpha \in [0,1].$$

Note that $P^{P,Q}$ is empty when $Y = \emptyset$, $P^{P,Q}$ contains more than one pm unless $q(x)/p(x)$ is a constant on Y and that the family is independent of the measure μ. Furthermore, $r_0(x) = p(x)1_Y(x)/P(Y)$ and $r_1(x) = q(x)1_Y(x)/Q(Y)$, thus when $P(Y) = 1$ we have $P = R_0$.

Exponential families as in Definition 2.1 were earlier used by Brown (1971). Note that the dominating measure $p(x)1_Y(x)\mu(x)$ of $\{R_\alpha\}$ is not necessarily a probability measure.

When P_η, P_θ are members of an exponential family $\{P_\theta : \theta \in \Theta\}$, the family between P_η and P_θ is a linear subfamily of $\{P_\theta : \theta \in \Theta^*\}$:

$$P^{P_\eta,P_\theta} = \{P_{(1-\alpha)\eta+\alpha\theta} : \alpha \in [0,1]\}.$$

Note that $P^{P_\eta,P_\theta} \subset \{P_\theta : \theta \in \Theta\}$ for all $\eta,\theta \in \Theta$ iff Θ is convex. More generally, we shall call a set P of pms *exponentially convex* when $P,Q \in P$ implies $P^{P,Q} \subset P$.

2b. Kullback-Leibler information

Let P^* be the class of all pms on a set X. For $P,Q \in P^*$ the Kullback-Leibler information $K(P,Q)$ is defined as

$$K(P,Q) = E_P \log \frac{dP}{dQ} \text{ when } P \ll Q, \ \infty \text{ otherwise.}$$

Some basic properties are $0 \le K(P,Q) \le \infty$, $K(P,Q) = 0 \Leftrightarrow dP/dQ = 1\ [P]$, $K(\alpha P_1 + (1-\alpha)P_0,Q) \le \alpha K(P_1,Q) + (1-\alpha)K(P_0,Q)$. The last inequality implies convexity of the sets

(2.6) $\Gamma^*(b,Q) = \{P \in \mathcal{P}^* : K(P,Q) \le b\}, \quad Q \in \mathcal{P}^*, \quad 0 \le b \le \infty.$

In spite of these distance-like properties Kullback-Leibler information is not a metric (it does not satisfy the triangle inequality and is not symmetric, for instance). Note that the set $\Gamma^*(b,Q)$ is not exponentially convex. We give an example:

EXAMPLE 2.3. Let x_1, x_2, x_3 be distinct points and define $Q\{x_i\} = \frac{1}{3}$, $i = 1,2,3$; $P_0\{x_i\} = \frac{1}{2}$, $i = 1,2$ and $P_1\{x_i\} = \frac{1}{2}$, $i = 2,3$. $\mathcal{P}^{P_0,P_1}$ contains one pm R with $K(R,Q) = \log 3$ whereas $K(P_0,Q) = K(P_1,Q) = \log \frac{3}{2}$.

A useful identity concerning Kullback-Leibler information is

(2.7) $$K(R,P) - K(R,Q) = \int \log \frac{dQ}{dP}\, dR,$$

which holds when both K's are finite and $Q \ll P$ [R] (Csiszār (1975)). The following lemma was also proved by Csiszār ((1975), Lemma 2.1).

LEMMA 2.1 (Csiszār). *When* $K(P,R)$ *and* $K(R,Q)$ *are finite, then* (i) *and* (ii) *below are equivalent.*

(i) $K(\alpha P + (1-\alpha)Q) \ge K(R,Q)$ *for each* $\alpha \in [0,1]$.

(ii) $\int \log \frac{dR}{dQ}\, dP \ge K(R,Q)$.

□

A quantity related to Kullback-Leibler information is $M(P,Q)$, defined for P and $Q \in \mathcal{P}^*$ by

(2.8) $$M(P,Q) = \inf \{\max [K(R,P), K(R,Q)] : R \in \mathcal{P}^*\}.$$

$M(P,Q)$ occurs in the theory of Chernoff efficiency of tests - it is the logarithm of the Chernoff index of the most powerful test of P against Q -, cf. Chernoff (1952), Kallenberg (1982) and, more implicitly, Brown (1971). One might say that $M(P,Q)$ is the Kullback-Leibler information of the "middle" of $\mathcal{P}^{P,Q}$ with respect to P and Q. A lemma supports this view:

LEMMA 2.2. *When it is finite, the infimum in* (2.8) *is attained for a unique pm* $R_{\tilde{\alpha}} \in \mathcal{P}^{P,Q}$. *Without restrictions we have, writing* $p = dP/d\mu$ *and* $q = dQ/d\mu$,

(2.9) $$M(P,Q) = -\log \inf_{0<\alpha<1} \int q^{\alpha} p^{1-\alpha}\, d\mu.$$

The proof is deferred to Section 2d.

Often we shall not be concerned with $\mathcal{P}^*$ itself, but with a subset $\mathcal{P}$ of $\mathcal{P}^*$. Therefore we define for $P,Q \in \mathcal{P}$

$$M_{\mathcal{P}}(P,Q) = \inf \{\max [K(R,P),K(R,Q)] : R \in \mathcal{P}\}. \tag{2.10}$$

Note that $R_{\tilde{\alpha}} \in \mathcal{P}$ implies $M_{\mathcal{P}}(P,Q) = M(P,Q)$. The converse also often holds:

<u>LEMMA 2.3</u>. *If* $\mathcal{P}$ *is closed in the metric of total variation and* $P,Q \in \mathcal{P}$ *then* $M_{\mathcal{P}}(P,Q) = M(P,Q) < \infty$ *implies that the infimum in* (2.10) *is attained in* $\mathcal{P}$, *i.e.* $R_{\tilde{\alpha}} \in \mathcal{P}$.

The proof is again in 2d.

In exponential families (2.4) we have

$$K(\eta,\theta) = K(P_\eta,P_\theta) = (\eta-\theta)'\lambda(\eta) - \psi(\eta) + \psi(\theta) \tag{2.11}$$

for all $\eta \in \Theta^1$, $\theta \in \Theta^*$. In one-parameter exponential families, (2.11) holds for all $\eta,\theta \in \Theta^*$.

<u>EXAMPLE 2.4</u>. Let $\{P_\theta : \theta \in \mathbb{R}^k\}$ be the multivariate normal family with mean vector θ and identity covariance matrix I_k, then we have $dP_\theta(x) = \exp\{\theta'x - \frac{1}{2}\|\theta\|^2\}dP_0(x)$, hence $K(\eta,\theta) = \frac{1}{2}\|\eta-\theta\|^2$.

In other exponential families the relation between Kullback-Leibler information and Euclidean distance is not so nice, but there is still a connection, as was proved by Kallenberg (1981), Lemma 3.1.a:

<u>LEMMA 2.4. (Kallenberg)</u>. *If* $\{P_\theta : \theta \in \Theta^*\}$ *is an exponential family and* C *a compact subset of* int Θ^*, *then* $\|\lambda(\eta) - \lambda(\theta)\| / \|\eta-\theta\|$, $K(\eta,\theta) / \|\eta-\theta\|^2$ *and* $K(\eta,\theta)/((\eta-\theta)'(\lambda(\eta) - \lambda(\theta)))$ *are uniformly bounded away from zero and infinity, for* $\eta,\theta \in C$, $\eta \neq \theta$. □

In exponential families, the equivalent of (2.7) is, for $\eta,\theta \in \Theta^1$ and $\xi \in \Theta^*$,

$$K(\theta,\xi) - K(\theta,\eta) = (\eta-\xi)'\lambda(\theta) - \psi(\eta) + \psi(\xi)$$

or, sometimes more conveniently,

$$K(\theta,\xi) - K(\theta,\eta) = (\eta-\xi)'(\lambda(\theta) - \lambda(\eta)) + K(\eta,\xi). \tag{2.12}$$

As an analogue to (2.6) we define for $\theta \in \Theta^*$

(2.13) $\Gamma(b,\theta) = \{\eta \in \Theta^* : K(\eta,\theta) \le b\}$.

A sort of Kullback-Leibler "distance" $K(\theta)$ of the boundary of Θ^* to θ is defined for interior points of Θ^* by

(2.14) $K(\theta) = \sup \{a : \exists \text{ compact } C_a \subset \operatorname{int} \Theta^* \text{ with } \Gamma(a,\theta) \subset C_a\}$.

Note that by continuity of $K(\cdot,\theta)$ on int Θ^* for $\theta \in$ int Θ^*, $\Gamma(a,\theta)$ is closed when $a < K(\theta)$.

To see that $K(\theta)$ is positive on int Θ^*, take a compact Euclidean ball $\{\eta : ||\eta-\theta|| \le b\} \subset$ int Θ^*. $K(\cdot,\theta)$ attains its infimum over the compact surface of this ball in $\tilde{\eta}$, say. It follows that $K(\theta) \ge K(\tilde{\eta},\theta) > 0$.

A lemma of Kallenberg (personal communication) gives conditions ensuring that $K(\theta) = \infty$ on int Θ^*.

LEMMA 2.5. (Kallenberg). *If* Θ^* *is open and*

(2.15) $B = \{x \in \mathbb{R}^k : \sup_{\theta\in\Theta^*} (\theta'x - \psi(\theta)) < \infty\}$ *is open*

then $K(\theta) = \infty$ *for each* $\theta \in \Theta^*$.

We reproduce Kallenberg's proof here with his permission.

PROOF of Lemma 2.5. (Kallenberg). Let $\theta_0 \in \Theta^*$. It is sufficient to prove, for each $a > 0$ that $\Gamma_a = \{\theta \in \Theta^* : K(\theta,\theta_0) \le a\}$ is compact. Define

$$B_a = \{x : \sup_{\theta\in\Theta^*} [(\theta-\theta_0)'x - \psi(\theta) + \psi(\theta_0)] \le a\}$$

then B_a is closed (its complement is open), and bounded since

$$(\delta x / ||x||)'x - \psi(\theta_0 + \delta x / ||x||) + \psi(\theta_0) \to \infty$$

as $||x|| \to \infty$ when δ is small enough, hence B_a is compact. Note that $\Lambda = \lambda(\Theta^*)$ is open. It shall be proved that $\Gamma_a = \lambda^{-1}(B_a)$ and that $B_a \subset \Lambda$, implying compactness of Γ_a by continuity of λ^{-1} on Λ.

First, let C be the closure of the convex hull of the support of μ, then, by Theorems 9.1 (ii)* and 9.2 of Barndorff-Nielsen (1978), int $C \subset B \subset C$ and $\lambda(\Theta^*)$ = int C, which, since B is open, implies

(2.16) $\quad B = \text{int } C = \lambda(\Theta^*)$.

Observe, that (2.16) implies $B_a \subset B = \Lambda$. It remains to show

$$\lambda^{-1}(B_a) = \Gamma_a$$

but this follows readily from

$$\sup_{\theta\in\Theta^*} \{(\theta-\theta_0)'\lambda(\tilde{\theta}) - \psi(\theta) + \psi(\theta_0)\} = K(\tilde{\theta},\theta_0)$$

for all $\theta_0 \subset \Theta^*$, and the 1-1 relation between $B = \Lambda$ and Θ^*. □

REMARK 2.1. The condition that Θ^* is open is frequently satisfied. Barndorff-Nielsen (1978) calls families with open Θ^* *regular*. Furthermore, by (2.16) the conditions imply that the likelihood $\ell(\theta;x) = \theta'x - \psi(\theta)$ attains its supremum (in $\lambda^{-1}(x)$) when $x \in B$. Since $B = \lambda(\Theta^*)$ is open we have $P_\theta(\overline{X}_n \in B) \to 1$ as $n \to \infty$ for each $\theta \in \Theta^*$ so that the probability that the maximum likelihood estimate $\lambda^{-1}(\overline{X}_n)$ exists tends to 1 as $n \to \infty$.

2c. A large deviation lemma

Here we present a lemma on large deviations in one-parameter exponential families. Though the lemma follows readily from Chernoff's theorem (Chernoff (1952)), we include a proof here since this exemplifies the proof of a large deviation theorem in its simplest form.

LEMMA 2.6. *Let* $\{P_\theta : \theta \in \Theta\}$ *be a one-parameter exponential family. When* $\theta \in \text{int } \Theta^*$ *and* $a \in \text{int } \Lambda$ *with* $\lambda(\theta) < a$ *then*

$$\lim_{n\to\infty} n^{-1} \log P_\theta(\overline{X}_n \geq a) = -K(\lambda^{-1}(a),\theta).$$

PROOF. Writing $\eta = \lambda^{-1}(a)$ we have $\eta > \theta$ and

$$\begin{aligned} (2.17) \qquad P_\theta(\overline{X}_n \geq a) &= \int_{[a,\infty)} dP_\theta^{(n)}(x) \\ &= \int_{[a,\infty)} \exp\{-n[(\eta-\theta)x - \psi(\eta) + \psi(\theta)]\} dP_\eta^{(n)}(x) \\ &= \exp\{-nK(\eta,\theta)\} \int_{[a,\infty)} \exp\{-n(\eta-\theta)(x-a)\} dP_\eta^{(n)}(x). \end{aligned}$$

The last integral is bounded since x-a is nonnegative. Furthermore, it is larger than

$$(2.18)\qquad \int_{[a,a+n^{-\frac{1}{2}})} \exp\{-n^{\frac{1}{2}}(\eta-\theta)\}dP_{\eta}^{(n)}(x)$$

$$= \exp\{-n^{\frac{1}{2}}(\eta-\theta)\}P_{\eta}(a \le \overline{X}_n < a+n^{-\frac{1}{2}}).$$

Since $n^{\frac{1}{2}}(\overline{X}_n - a)$ is asymptotically normal with mean zero under P_{η}, the probability in (2.18) tends to a positive limit and the lemma is proved. □

REMARK 2.2. The technique of the proof above is called *exponential centering* (the pm $P_{\eta}^{(n)}$ is determined so that the integration region is a central part of the distribution) and is not only a standard tool in proofs of large deviation theorems, but also - then known as the saddle point method - used to obtain better approximations of distribution functions for small values of n, cf. Daniels (1954), Barndorff-Nielsen and Cox (1979).

Exponential centering is particularly easy in exponential families since the centered pm - called conjugate or associate - is a member of the same full exponential family. The large deviation theorems of sections III.4 and 5 will be proved using exponential centering, with a more accurate evaluation of the integral.

2d. Proofs

PROOF of Lemma 2.2. First assume $\mu(pq > 0) = 0$, then $M(P,Q) = \infty$ and (2.9) holds trivially. In case $\mu(pq > 0) > 0$, both $K(R_0,P)$ and $K(R_1,Q)$ are finite. For each R with $R << R_0$ we have

$$(2.19)\qquad K(R,P) = \int \log\frac{dR}{dP}\,dR = \int \left\{\log\frac{dR}{dR_0} + \log\frac{dR_0}{dP}\right\} dR$$

$$= K(R,R_0) + K(R_0,P),$$

where the last equality holds true since dR_0/dP is a constant $[R_0]$. Similarly,

$$(2.20)\qquad K(R,Q) = K(R,R_1) + K(R_1,Q).$$

Since (2.19) and (2.20) hold for $R = R_{\alpha}$ and since $\alpha \mapsto K(R_{\alpha},R_i)$ is continuous and monotone for $i = 0,1$, an $\tilde{\alpha}$ exists which uniquely minimizes

$$(2.21)\qquad \max\,[K(R_{\alpha},P),K(R_{\alpha},Q)]$$

over $\alpha \in [0,1]$. Note that this minimum is finite, implying $M(P,Q) < \infty$.

Now let $R \in \mathcal{P}^*$ with $K(R,P) < \infty$ and $K(R,Q) < \infty$ then $R << R_0$ and we shall prove that

$$(2.22) \qquad \max [K(R,P),K(R,Q)] \geq \max [K(R_{\tilde{\alpha}},P),K(R_{\tilde{\alpha}},Q)]$$

with equality iff $R = R_{\tilde{\alpha}}$.

First assume $\tilde{\alpha} = 0$, then $K(R_0,Q) \leq K(R_0,P)$ and hence (2.19) implies (2.22) with equality iff $R = R_0$.

In case $0 < \tilde{\alpha} < 1$ we have $K(\tilde{R},Q) = K(\tilde{R},P)$ where $\tilde{R} = R_{\tilde{\alpha}}$, implying by (2.7) that

$$(2.23) \qquad \int \log \frac{dQ}{dP} d\tilde{R} = 0.$$

By definition 2.1 we have on $\mathcal{Y} = \{x : p(x)q(x) > 0\}$

$$(2.24) \qquad \log \frac{d\tilde{R}}{dP} = \tilde{\alpha} \log \frac{dQ}{dP} - \psi^{P,Q}(\tilde{\alpha}),$$

hence integration with respect to $\tilde{R}$ yields

$$(2.25) \qquad K(\tilde{R},P) = -\psi^{P,Q}(\tilde{\alpha}).$$

Assume without loss of generality that $K(R,Q) \leq K(R,P)$ then (2.7) implies

$$\int \log \frac{dQ}{dP} dR \geq 0.$$

In view of (2.25), integration with respect to R of (2.24) now yields

$$\int \log \frac{d\tilde{R}}{dP} dR \geq K(\tilde{R},P)$$

and, using (2.7),

$$(2.26) \qquad K(R,P) \geq K(R,\tilde{R}) + K(\tilde{R},P).$$

The case $\tilde{\alpha} = 1$ is analogous to that of $\tilde{\alpha} = 0$.

To prove (2.9), observe that by Definition 2.1 we have

$$\psi^{P,Q}(\alpha) = \log \int q^{\alpha} p^{1-\alpha} 1_{\{pq>0\}} d\mu, \qquad \alpha \in [0,1]$$

and, since $\log \frac{dQ}{dP}$ is the sufficient statistic in the family $\mathcal{P}^{P,Q}$,

$$(2.27) \qquad E_{\alpha} \log \frac{dQ}{dP} = \int \log \frac{dQ}{dP} dR_{\alpha} = \frac{d}{d\alpha} \psi^{P,Q}(\alpha).$$

For $\tilde{\alpha} \in (0,1)$, (2.9) now follows from (2.23). When $\tilde{\alpha} = 0$ (say) we have $K(R_\alpha,Q) \le K(R_\alpha,P)$ implying by (2.7) that (2.27) is nonnegative for each $\alpha \in (0,1)$ and hence (2.9) follows by continuity of $\psi^{P,Q}$. □

PROOF of Lemma 2.3. Let $\{R_n\} \subset \mathcal{P}$ satisfy, as $n \to \infty$,

(2.28) $$\max\,[K(R_n,P),K(R_n,Q)] \to M(P,Q).$$

Assume $\tilde{\alpha} = 0$, then $M(P,Q) = K(R_0,P)$, hence (2.19) implies $K(R_n,P) = K(R_n,R_0) + M(P,Q)$. It follows now from (2.28) that

$$\lim_{n\to\infty} K(R_n,R_0) = 0.$$

By the corollary to Theorem 2.4.2 in Pinsker (1975) this implies that $\{R_n\}$ tends to R_0 in total variation. Now let $0 < \tilde{\alpha} < 1$ and assume R_n satisfies

(2.29) $$K(R_n,Q) \le K(R_n,P).$$

By the previous proof, (2.26) then holds for $R = R_n$, which can be rewritten as

$$K(R_n,\tilde{R}) \le K(R_n,P) - M(P,Q).$$

Together with the analogous result for the R_n's which do not satisfy (2.29) we have

$$K(R_n,\tilde{R}) \le \max\,[K(R_n,P),K(R_n,Q)] - M(P,Q),$$

implying convergence to zero of the left hand side and $R_n \to \tilde{R}$ in total variation. □

CHAPTER II

LARGE DEVIATIONS OF ESTIMATES

1. INTRODUCTION

Let $\mathcal{X}$ be a topological space and $\mathcal{B}$ its Borel σ-algebra, let $\mathcal{P}^*$ be the class of all probability measures on $\mathcal{B}$ and let $\mathcal{P}$ be a subclass of $\mathcal{P}^*$. Furthermore, let $g : \mathcal{P} \to \mathbb{R}^d$ be a map.

Mostly, $\mathcal{P}$ will be a parametric family $\{P_\theta : \theta \in \Theta\}$ with $\Theta \subset \mathbb{R}^k$, but we also want to be able to treat other $\mathcal{P}$'s such as the class of pms on $\mathbb{R}$ which have symmetric Lebesgue densities. In that case, g(P) could be the centre of symmetry of P.

Let the sequence $\{X_1, X_2, \ldots\}$ with values in $\mathcal{X}^\infty$ be distributed according to the product measure P^∞ on $\mathcal{B}^\infty$ for some $P \in \mathcal{P}$ and let $\{T_n\}_{n=1}^\infty$ be a sequence of estimates of g(P) such that $T_n = T_n(X_1, \ldots, X_n)$.
We shall only consider estimates that take values in $g(\mathcal{P})$.

The quality of an estimate T_n is usually measured by its (normalized) expected quadratic loss

$$(1.1) \qquad v_n(T_n, P) = \int_{\mathcal{X}^n} n \cdot \|T_n - g(P)\|^2 dP^n,$$

or, since $v_n(T_n, P)$ can be hard to obtain, by the variance of the limit distribution of $n^{\frac{1}{2}}(T_n - g(P))$.

In this chapter we concentrate on the inaccuracy function

$$(1.2) \qquad \alpha_n(\varepsilon, P, T_n) = P^n(\|T_n - g(P)\| > \varepsilon)$$

as a criterion to judge the quality of T_n, a point of view taken by Basu (1956), Bahadur (1960 b), Huber (1968) and others. Just like the expected quadratic loss (1.1), the inaccuracy function (1.2) can usually not be evaluated explicitly and an asymptotic expression is taken instead. For consistent (sequences of) estimates $\{T_n\}$ the inaccuracy function tends to zero as $n \to \infty$, typically exponentially fast when $\varepsilon > 0$ is fixed.

EXAMPLE 1.1. Let $X_1, X_2, \dots$ be i.i.d. normal with mean θ and unit variance, let $g(\theta) = \theta$ and $T_n = \overline{X}_n$, then $P_\theta(|\overline{X}_n - \theta| > \varepsilon) = \exp\{-n(\varepsilon^2/2) + o(n)\}$ as $n \to \infty$.

Note that $g(P_\theta)$ is abbreviated to $g(\theta)$ and that the exponent on P_θ is suppressed. We will continue to do this unless it causes ambiguities.

EXAMPLE 1.2. Let $X_1, X_2, \dots$ be i.i.d. uniform $(0,\theta)$, $g(\theta) = \theta$ and $T_n = X_{n:n}$, the largest order statistic. We have $P_\theta(X_{n:n} < \theta-\varepsilon) = \exp\{-n \log(\theta/(\theta-\varepsilon))\}$. This example is a rare occasion where the inaccuracy function is simple and explicit.

The exponential rate of convergence ($\varepsilon^2/2$ and $\log(\theta/(\theta-\varepsilon))$ in the examples above) was coined *inaccuracy rate* by Sievers (1978) in a shift family context. We give a somewhat modified definition.

DEFINITION 1.1. The inaccuracy rate $e(\varepsilon, P, \{T_n\})$ of the sequence $\{T_n\}$ of estimates of $g(P)$ is defined as

$$e(\varepsilon, P, \{T_n\}) = -\limsup_{n\to\infty} n^{-1} \log P(\|T_n - g(P)\| > \varepsilon). \tag{1.3}$$

When $\mathcal{P}$ is a parametric family $\{P_\theta : \theta \in \Theta\}$, the inaccuracy rate will be written as $e(\varepsilon, \theta, \{T_n\})$. Note that large values of e mean that estimation errors are rarely larger than ε.

In contrast to the variance of the asymptotic distribution (to be called asymptotic variance hereafter), the inaccuracy rate is determined by the non-local or large deviation-behaviour of the estimate $\{T_n\}$. The same contrast is found in testing problems between Pitman efficiency and Bahadur efficiency. The local criteria - asymptotic variance and Pitman efficiency - may be called classical and are extensively studied. There have also been, in the last two decades, many publications on Bahadur efficiency, see Kallenberg (1981) for references. In estimation theory, however, the inaccuracy rate approach is relatively unexplored, especially for fixed values of ε, though it was already proposed by Basu (1956) and Bahadur (1960 b).

Our aim is to find estimates that are best with respect to the inaccuracy rate; to avoid trivial "estimates" such as $T_n \equiv c$, which has inaccuracy rate infinity when $\|c - g(P)\| \le \varepsilon$ and zero otherwise, we shall only consider *consistent* estimates. To know whether an estimate indeed attains the maximal value of the inaccuracy rate (in the class of consistent estimates) it is useful to have an upper bound for this criterion. Such a bound was found for consistent estimates by Bahadur (1960 b, 1971) and recently generalized by Bahadur *et al* (1980) to non-i.i.d. frameworks. Bahadur's bound on the inaccuracy rate can be seen as an equivalent of the Cramér-Rao bound on the asymptotic variance.

The main tool for the derivation of Bahadur's bound is Stein's lemma (see Chernoff (1956)), which is an application of the fundamental lemma of Neyman and Pearson. We restate the inequality-part of Stein's lemma here as

LEMMA 1.1. *If* $P, Q \in P^*$ *and* $\{A_n\}_{n=1}^{\infty}$ *is a sequence of subsets* $A_n \subset X^n$, *then*

$$\liminf_{n\to\infty} Q^n((X_1, X_2, \ldots, X_n) \in A_n) > 0$$

implies

$$\liminf_{n\to\infty} n^{-1} \log P^n((X_1, X_2, \ldots, X_n) \in A_n) \ge -K(Q,P).$$

□

For applications of this lemma where a limsup is needed we have a simple corollary:

COROLLARY 1.2. *Under the conditions of Lemma* 1.1,

$$\limsup_{n\to\infty} Q\{A_n\} > 0 \Rightarrow \limsup_{n\to\infty} n^{-1} \log P\{A_n\} \ge -K(Q,P),$$

where $\{A_n\}$ *abbreviates* $((X_1, \ldots, X_n) \in A_n)$.

PROOF. Take a subsequence $\{n_k\}$ with $\lim_{n\to\infty} Q\{A_{n_k}\} = \limsup_{n\to\infty} Q\{A_n\}$. Define sets A_n' as

$$A_n' = \{(x_1, \ldots, x_n) : (x_1, \ldots, x_{n_k}) \in A_{n_k}\}, \qquad n_k \le n < n_{k+1}$$

then $\lim_{n\to\infty} Q\{A_n'\} = \limsup_{n\to\infty} Q\{A_n\} > 0$. The corollary now follows from Lemma 1.1 and

$$\limsup_{n\to\infty} n^{-1} \log P\{A_n\} \ge \liminf_{n\to\infty} n^{-1} \log P\{A_n'\}.$$

□

The upper bound on the inaccuracy rate for consistent estimates is now easily derived: When $\{T_n\}$ is consistent we have for each $Q \in \mathcal{P}$ with

(1.4) $$\|g(Q) - g(P)\| > \varepsilon,$$

that $\liminf_{n\to\infty} Q(\|T_n - g(P)\| > \varepsilon) > 0$, implying by Lemma 1.1

(1.5) $$\liminf_{n\to\infty} n^{-1} \log P(\|T_n - g(P)\| > \varepsilon) \geq -K(Q,P).$$

Since (1.5) holds for each $Q \in \mathcal{P}$ satisfying (1.4), the following lemma is proved:

LEMMA 1.3. *If* $\{T_n\}$ *is a consistent estimate of* $g(P)$ *for each* $P \in \mathcal{P}$, *then*

(1.6) $$-\liminf_{n\to\infty} n^{-1} \log P(\|T_n - g(P)\| > \varepsilon) \leq b(\varepsilon,P),$$

where $b(\varepsilon,P)$ *is Bahadur's bound*

(1.7) $$b(\varepsilon,P) = \inf\{K(Q,P) : Q \in \mathcal{P}, \|g(Q) - g(P)\| > \varepsilon\}.$$

□

When $\mathcal{P}$ is a parametric family $\{P_\theta : \theta \in \Theta\}$, Bahadur's bound is denoted $b(\varepsilon,\theta)$. Note that (1.6) is stronger than the formal statement of the bound

(1.8) $$e(\varepsilon,P,\{T_n\}) \leq b(\varepsilon,P).$$

In view of (1.8) we shall call a sequence of estimates $\{T_n\}$ of $g(P)$ *inaccuracy rate optimal* when

(1.9) $$e(\varepsilon,P,\{T_n\}) = b(\varepsilon,P)$$

for all $\varepsilon > 0$, $P \in \mathcal{P}$. Since this optimality is often unattainable, we say that $\{T_n\}$ is inaccuracy rate optimal for ε_0 at P_0 when (1.9) holds for $\varepsilon = \varepsilon_0$ and $P = P_0$.

EXAMPLE 1.1. (continued). We have $K(\eta,\theta) = \frac{1}{2}|\eta-\theta|^2$ implying $b(\varepsilon,\theta) = \frac{1}{2}\varepsilon^2$. It follows that $\overline{X}_n$ is an inaccuracy rate optimal estimate of θ.

EXAMPLE 1.2. (continued). $K(\eta,\theta) = \log(\theta/\eta)$ when $0 < \eta \leq \theta$ and ∞ otherwise, hence $X_{n:n}$ is inaccuracy rate optimal.

EXAMPLE 1.3. Let $\mathcal{X} = \mathbb{R}$ and $\mathcal{B}$ the σ-algebra of Borel sets on $\mathbb{R}$. Let $\mathcal{P}$ be the class of pms having a positive Lebesgue density on $\mathbb{R}$ and let g map each pm onto its median. It has been noted in Bahadur *et al* (1980) that in this case the sample median $X_{[(n+1)/2]:n}$ is inaccuracy rate optimal. Now this is hardly useful, since the class P is so large that there are very few essentially different consistent estimates indeed.

As remarked in Bahadur *et al* (1980), Example 1.3 parallels a result of Pfanzagl (1976) concerning the optimality of the sample median with respect to the asymptotic variance. We shall extend this parallel, thereby providing an answer to a question in Bahadur *et al* (1980), Example 4.2.

EXAMPLE 1.4. Let X, B and g as in Example 1.3 and let P_S be the class of *symmetric* positive Lebesgue densities. There are many consistent estimates of the median (e.g. "robust" symmetric estimates), but, as will be proved in Section 2, none of them attains the bound $b(\varepsilon,P)$ for each $P \in P_S$, not even for any fixed $\varepsilon > 0$ (see Example 2.2).

This example parallels the known fact (for references, see Pfanzagl (1976)) that the sample median is not optimal with respect to the asymptotic variance in P_S.

In general the determination of $e(\varepsilon,P,\{T_n\})$ can be quite a problem, even in parametric families. This is one of the reasons that many authors (Bahadur (1960 b,1967,1971), Fu (1973,1975), Perng (1978)) mainly study the behaviour of $e(\varepsilon,P,\{T_n\})$ and $b(\varepsilon,P)$ as $\varepsilon \to 0$. In regular one-parameter cases (with $g(\theta) \equiv \theta$) they find

$$(1.10) \qquad b(\varepsilon,\theta) = \tfrac{1}{2}\varepsilon^2 i_\theta + o(\varepsilon^2) \quad \text{as } \varepsilon \to 0,$$

where i_θ denotes the Fisher information, and moreover, that

$$(1.11) \qquad \lim_{\varepsilon\to 0} e(\varepsilon,\theta,\{T_n\}) \,/\, b(\varepsilon,\theta) = 1$$

holds for the MLE of θ. The optimality property (1.11) has been called "local asymptotic optimality" by Bahadur (1980), in earlier papers it has been named "efficiency in the sense of asymptotic effective variances" (Bahadur (1960 b,1971)) and "asymptotic efficiency in Bahadur's sense" (Fu (1973,1975)). In view of our earlier definition following (1.8), we shall refer to (1.11) as *local inaccuracy rate optimality*.

In Section 2, a sufficient condition for inaccuracy rate optimality due to Bahadur will be presented, together with an explanation why that condition cannot hold in situations where $\mathcal{P}$ is not exponentially convex.

In Section 3, exponential families $\{P_\theta : \theta \in \Theta\}$ are treated. These families are exponentially convex iff Θ is convex and in that case, the MLE is usually inaccuracy rate optimal. As an important subclass of the non-convex exponential families, curved exponential families are examined next. Under a linearity restriction on the class of estimates, it turns out that inaccuracy rates are either equal to the inaccuracy rate of the MLE for each ε in an interval $(0,\varepsilon_0)$ or strictly lower than the inaccuracy rate of the MLE for each ε in an interval $(0,\varepsilon_0)$.

In the final section of this chapter we look at shift families on the real line. A logical restriction is here to translation equivariant estimates. Results of Sievers (1978) will be extended and his "optimal estimates" will be proved to be essentially unique.

2. A NATURAL CONDITION FOR OPTIMALITY

Suppose $\{\tilde{P}_n\}$ is a consistent estimate of P and g is continuous, then $\{T_n\} = \{g(\tilde{P}_n)\}$ is a consistent estimate of g(P). Note that maximum likelihood estimates often have this structure. Now regard $\{K(\tilde{P}_n,P_0)\}$ as an estimate of $K(P,P_0)$. If $K(\cdot,P_0)$ is continuous this estimate is consistent, hence (1.7) yields

$$(2.1) \qquad -\liminf_{n\to\infty} n^{-1} \log P_0\ (K(\tilde{P}_n,P_0) > b) \le \inf\{K(Q,P_0) : Q \in \mathcal{P},\ K(Q,P_0) > b\}.$$

Note that the right hand side of (2.1) can be larger than b. The simple but useful Proposition 2 of Bahadur (1980) (also in Section 2 in Bahadur (1983)) states that optimality with respect to the inaccuracy rate of $\{K(\tilde{P}_n,P_0)\}$ as an estimate of $K(P,P_0)$ yields optimality of $g(\tilde{P}_n)$ as an estimate of g(P). We give a slightly refined version:

<u>PROPOSITION 2.1. (Bahadur)</u>. *If* g *is continuous and* $\{\tilde{P}_n\}$ *is a consistent estimate of* P *such that for each* $b < b_0 = b_0(P_0)$

$$(2.2) \qquad \limsup_{n\to\infty} n^{-1} \log P_0\ (K(\tilde{P}_n,P_0) \ge b) \le -b$$

then $e(\varepsilon,P_0,\{g(\tilde{P}_n)\}) = b(\varepsilon,P_0)$ *for each* $\varepsilon > 0$ *with* $b(\varepsilon,P_0) < b_0$, *and hence*

$g(\tilde{P}_n)$ *is an inaccuracy rate optimal estimate of* $g(P)$, *for each* $\varepsilon > 0$ *with* $b(\varepsilon,P_0) < b_0$, *at* P_0. □

Several remarks should be made concerning this proposition:

REMARK 2.1. Continuity of g and consistency of $\tilde{P}_n$ can only be defined when $\mathcal{P}$ is equipped with a topology, e.g. the topology of total variation. When $\mathcal{P}$ is a parametric family $\{P_\theta : \theta \in \Theta\}$, consistency and continuity are usually defined with respect to the topology on Θ. Equivalence of both definitions depends on the bi-continuity of the map $\theta \mapsto P_\theta$.

REMARK 2.2. Since optimality with respect to the inaccuracy rate of $\{K(\tilde{P}_n,P_0)\}$ as an estimate of $K(P,P_0)$ implies optimality of $\{g(\tilde{P}_n)\}$ as an estimate of $g(P)$ for each continuous g, Kullback-Leibler information is a sort of "canonical distance" when dealing with large deviations of estimates. To further support this, note that when $\tilde{\theta}_n$ estimates θ, the probability $P_\theta(K(\tilde{\theta}_n,\theta) \geq b)$ does not depend on the parametrization, whereas $P_\theta(\|g(\tilde{\theta}_n) - g(\theta)\| > \varepsilon)$ generally does.

REMARK 2.3. The condition on ε in Proposition 2.1 holds for each $b_0 > 0$ when ε is small enough, provided g is not a constant on any "open Kullback-Leibler ball" $\{Q : K(Q,P_0) < \delta\}$. To see this, take Q with $K(Q,P_0) < b_0$ and $g(Q) \neq g(P_0)$, then let $\varepsilon < \|g(Q) - g(P_0)\|$.

We give a simple example of an application of Proposition 2.1.

EXAMPLE 2.1. Let $\{P_\theta : \theta \in \mathbb{R}^k\}$ denote a multivariate normal family with fixed covariance Σ and mean vector θ. Since $K(\eta,\theta) = \frac{1}{2}(\eta-\theta)'\Sigma^{-1}(\eta-\theta)$ we have, taking $\overline{X}_n$ as an estimate of θ,

$$P_\theta(K(\overline{X}_n,\theta) \geq b) = P(\tfrac{1}{2}\chi_k^2 \geq nb) = \int_{2nb}^{\infty} \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1}e^{-x/2}dx.$$

Since for each positive δ, when x is large enough

$$1 \leq x^{k/2-1} < e^{\delta x},$$

we have as $n \to \infty$,

$$n^{-1} \log P_\theta(K(\overline{X}_n,\theta) \geq b) \to -b. \tag{2.3}$$

Thus (2.2) holds and $g(\overline{X}_n)$ is an inaccuracy rate optimal estimate of $g(\theta)$

for each continuous g.

A theorem of Efron and Truax (1968) generalizes (2.3) to exponential families. This will be treated in Section 3. Another application of Proposition 2.1 is found in Example 4.2.

Bahadur (1980), Proposition 3, gives a local version of Proposition 2.1, too: if

$$\text{(2.4)} \qquad \limsup_{b\to 0}\ \limsup_{n\to\infty}\ (nb)^{-1}\log P(K(\tilde{P}_n,P)\geq b)\leq -1,$$

then $g(\tilde{P}_n)$ is locally inaccuracy rate optimal, cf. (1.11). Condition (2.4) is related to the existence of estimates which are optimal with respect to asymptotic variances: In sufficiently regular parametric families $\{P_\theta : \theta\in\Theta\}$ with $\Theta\subset\mathbb{R}^k$, it holds that as $\eta\to\theta$,

$$K(\eta,\theta) = \tfrac{1}{2}(\eta-\theta)'I_\theta(\eta-\theta) + o(\|\eta-\theta\|^2)$$

where I_θ is Fisher's information matrix. If an estimate $\tilde{\theta}_n$ of θ exists which is optimal with respect to asymtotic variances, i.e.

$$n^{\frac{1}{2}}(\tilde{\theta}_n-\theta)'I^{\frac{1}{2}} \xrightarrow{\mathcal{D}} N(0,I_k),$$

then $nK(\tilde{\theta}_n,\theta)\to\frac{1}{2}\chi^2_k$ in distribution. When this convergence is stronger so that

$$\text{(2.5)} \qquad n^{-1}\log P_\theta(K(\tilde{\theta}_n,\theta)\geq b)-n^{-1}\log P(\tfrac{1}{2}\chi^2_k\geq nb)$$

tends to zero uniformly for $b>0$, then (2.2) follows by calculation of the tail probability of the χ^2_k distribution. The uniform convergence of (2.5) is a stringent condition, but note that for (2.4) to hold the convergence is only needed for "infinitesimal" values of b and this may be expected to occur more frequently. Indeed (2.4) has been proved (implicitly) for the MLE in various situations, cf. Fu (1973,1975), Perng (1978) and, in certain Markov chains, Bahadur (1983).

As mentioned before, condition (2.2) is less frequently satisfied. We shall now give a lemma indicating that (2.2) cannot hold for each $P\in\mathcal{P}$ when $\mathcal{P}$ is not exponentially convex, cf. Section I.2.

<u>LEMMA 2.2</u>. *If* $P_1,P_2\in\mathcal{P}$ *such that*

(2.6) $\qquad M(P_1,P_2) < M_P(P_1,P_2) \qquad (\leq \infty)$

cf. (I.2.9), (I.2.10), *and* $\tilde{P}_n$ *is an estimate of* P *then condition* (2.2) *fails for each* b *with* $M(P_1,P_2) < b < M_P(P_1,P_2)$, *at least at one of the pms* P_1,P_2.

The proof is preceded by some remarks.

REMARK 2.4. When P is not exponentially convex and P is closed in the total variation topology, there are $P_1,P_2 \in P$ such that (2.6) holds, see Lemma 2.3. If moreover P is connected, such P's usually exist with $M(P_1,P_2)$ arbitrarily small, thereby refuting (2.2) for arbitrarily small values of b. Note that the class P in Example 1.3 is exponentially convex, whereas P_S of Example 1.4 is not. It should also be noted that the family of uniform $(0,\theta)$ distributions, cf. Example 1.2, is exponentially convex, $P^{\eta,\theta}$ consisting of one pm with parameter min (η,θ).

REMARK 2.5. Condition (2.2) is *not necessary* for an estimate $g(\tilde{P}_n)$ to be inaccuracy rate optimal. Thus, the occurence of (2.6) does not exclude the existence of an optimal estimate but merely indicates a reason for the possible elusiveness of such an estimate.

PROOF of Lemma 2.2. Let $M(P_1,P_2) < b < M_P(P_1,P_2)$ and choose $Q \in P^*$ such that

(2.7) $\qquad \max\,\{K(Q,P_1),K(Q,P_2)\} < b.$

Since $\tilde{P}_n$ takes values in P only we have $Q(\max\,\{K(\tilde{P}_n,P_1),K(\tilde{P}_n,P_2)\} > b) = 1$ for each n, implying either for i = 1 or for i = 2

$$\limsup_{n\to\infty} Q(K(P_n,P_i) > b) > 0.$$

Corollary 1.2 together with (2.7) completes the proof. □

Now we give an example where not only (2.6) holds, but indeed Bahadur's bound is proved to be unattainable. Moreover, we shall sketch a modification showing that local inaccuracy rate optimality, cf. (1.11), cannot hold either.

EXAMPLE 2.2. (continued from Example 1.4). We shall construct, for $\varepsilon = 1$, a pair P_1,P_2 of pms with symmetric densities such that their medians $g(P_1)$

and $g(P_2)$ satisfy

$$(2.8) \qquad |g(P_1) - g(P_2)| = 2\varepsilon$$

and such that

$$(2.9) \qquad M(P_1,P_2) < \min\{b(\varepsilon,P_1), b(\varepsilon,P_2)\},$$

implying not only (2.6), but also the existence of a pm $Q \in \mathcal{P}^*$ such that

$$(2.10) \qquad K(Q,P_i) < b(\varepsilon,P_i), \quad i = 1,2.$$

Since for any nondegenerate estimate $\{T_n\}$ of $g(P)$

$$Q(|T_n - g(P_1)| > \varepsilon \text{ or } |T_n - g(P_2)| > \varepsilon) > 0,$$

we have that for $i = 1$ or for $i = 2$

$$\limsup_{n\to\infty} Q(|T_n - g(P_i)| > \varepsilon) > 0$$

and hence by Corollary 1.2 and (2.10),

$$-\limsup_{n\to\infty} n^{-1} \log P_i(|T_n - g(P_i)| > \varepsilon) \le K(Q,P_i) < b(\varepsilon,P_i)$$

for $i = 1$ or for $i = 2$, proving Bahadur's bound to be unattainable.

Define P_1 and P_2 as follows: The symmetric density $p(x)$ is given by

$$p(x) = \begin{cases} \frac{1}{6} & \text{for } |x| < 1, \\ \frac{1}{6} \cdot 2^{-k} & \text{for } 2k-1 \le |x| < 2k+1, \quad k \ge 1. \end{cases}$$

Let a positive function f be periodic with period 4, symmetric around -1 and symmetric around $+1$, such that

$$(2.11) \qquad \int_{2k}^{2k+2} f(x)\,dx = 2,$$

but not symmetric around 0. For instance, f can be chosen as in fig. 2.1.

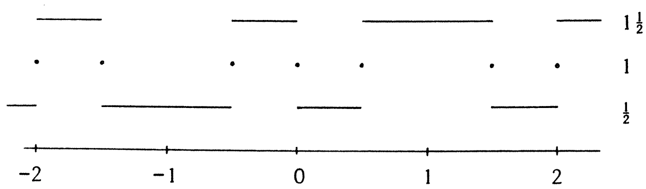

fig. 2.1. A possible f.

Now define P_1 and P_2 by their densities p_1 and p_2 as

$$p_1(x) = f(x) \cdot p(x+1),$$
$$p_2(x) = f(x) \cdot p(x-1).$$

Note that P_1 and P_2 are symmetric around -1 and 1 respectively, implying (2.8).

We determine $M(P_1,P_2)$. By Lemma I.2.2 and (I.2.25) we have $M(P_1,P_2) = K(\tilde{R},P_1) = -\psi^{P_1,P_2}(\tilde{\alpha})$ (Note that indeed $0 < \tilde{\alpha} < 1$ since $R_0 = P_1$ and $R_1 = P_2$). Furthermore, by (2.11)

$$(2.12) \qquad -\psi^{P_1,P_2}(\tilde{\alpha}) = -\log \int \exp\left\{\tilde{\alpha} \log \frac{p_2(x)}{p_1(x)}\right\} p_1(x)dx$$
$$= -\log \int \exp\left\{\tilde{\alpha} \log \frac{p(x-1)}{p(x+1)}\right\} f(x)p(x+1)dx$$

is independent of f since $p(x-1)$ and $p(x+1)$ are constant on $(2k,2k+2)$ for each $k \in \mathbb{Z}$. It follows that $\tilde{\alpha} = \frac{1}{2}$, yielding

$$(2.13) \qquad M(P_1,P_2) = -\log \int \{p(x+1)p(x-1)\}^{\frac{1}{2}}dx = \tfrac{1}{2} \log \tfrac{9}{8}.$$

The value of $b(\varepsilon,P_1)$ is found for $\varepsilon = 1$ as

$$\inf\left\{\int_{-\infty}^{\infty} \log \frac{r(x)}{p_1(x)} r(x)dx\colon r \text{ a density with } r(x) = r(-x)\right\}$$
$$= \inf\left\{\int_0^{\infty}\left[\log \frac{r(x)}{p_1(x)} + \log \frac{r(x)}{p_1(-x)}\right]r(x)dx\colon r(x) \geq 0, \int_0^{\infty} r(x)dx = \tfrac{1}{2}\right\}$$
$$= \inf\left\{\int_0^{\infty} \log \frac{r^2(x)}{p(x+1)p(x-1)f(x)f(-x)} r(x)dx\colon \ldots\right\}.$$

The infimum is attained when r is proportional to $\{p(x+1)p(x-1)f(x)f(-x)\}^{\frac{1}{2}}$ on $(0,\infty)$, hence

$$(2.14) \qquad b(\varepsilon,P_1) = -\log 2 \int_0^{\infty} \{p(x+1)p(x-1)f(x)f(-x)\}^{\frac{1}{2}}dx$$
$$= -\log 2 \sum_{k=0}^{\infty} \left[\tfrac{1}{6}\, 2^{-k-\frac{1}{2}} \int_{2k}^{2k+2} \{f(x)f(-x)\}^{\frac{1}{2}}dx\right]$$
$$= -\log \left(2 \cdot \tfrac{1}{6} \cdot 2^{\frac{1}{2}} \int_0^2 \{f(x)f(-x)\}^{\frac{1}{2}}dx\right),$$

where the periodicity of f is used. Note, that the last member of (2.14) can be written as

$$(2.15) \qquad \tfrac{1}{2} \log \tfrac{9}{8} - \log \tfrac{1}{2} \int_0^2 \{f(x)f(-x)\}^{\frac{1}{2}}dx,$$

which is larger than $M(P_1,P_2)$ (see (2.13)) by the conditions on f and Schwarz' inequality, thus (2.9) holds since $b(\varepsilon,P_2) = b(\varepsilon,P_1)$.

It follows moreover from (2.13) - (2.15), that the difference $b(\varepsilon,P_1) - M(P_1,P_2)$ can be made arbitrarily large by choosing f appropriately. When f is chosen as in fig. 2.1, this difference is $\frac{1}{2}\log\frac{4}{3}$.

Now we sketch how the example can be modified for arbitrary $\varepsilon > 0$: Take

$$p(x) = \begin{cases} c & \text{for } |x| < \varepsilon \\ c\rho^k & \text{for } (2k-1)\varepsilon \le |x| < (2k+1)\varepsilon \end{cases}$$

where $0 < \rho < 1$ can be chosen dependent on ε such that $p(x) \to \frac{1}{2}\exp\{-|x|\}$ as $\varepsilon \to 0$ ($\rho = \exp\{-2\varepsilon\}$), and let f be periodic with period 4ε, symmetric about $-\varepsilon$ and $+\varepsilon$, etc. Define $p_1(x) = f(x)p(x+\varepsilon)$, and $p_2(x) = f(x)p(x-\varepsilon)$, then the difference of $b(\varepsilon,P_1)$ and $M(P_1,P_2)$ is again $-\log\frac{1}{2\varepsilon}\int_0^{2\varepsilon}\{f(x)f(-x)\}^{\frac{1}{2}}dx$, implying (2.9) for arbitrary $\varepsilon > 0$.

Finally we consider $\varepsilon \to 0$. Let $\rho = \exp\{-2\varepsilon\}$, then, by analogy to (2.13) we obtain

$$M(P_1,P_2) = \varepsilon + \log\tfrac{1}{2}(1+e^{-2\varepsilon})$$

which is of order ε^2 as $\varepsilon \to 0$.

The difference $b(\varepsilon,P_1) - M(P_1,P_2)$ can be made to have order ε by choosing f such that $\frac{1}{2\varepsilon}\int_0^{2\varepsilon}\{f(x)f(-x)\}^{\frac{1}{2}}dx = \{1-\varepsilon\}^{\frac{1}{2}}$ (this can be achieved with an f similar to fig. 2.1 which satisfies $|f(x)-1| \le \sqrt{\varepsilon}$). It follows that local inaccuracy rate optimality, cf. (1.11), cannot hold, at least not uniformly on any total variation neighbourhood of the double exponential in the family of symmetric densities, for estimates of the median. Note that for the double exponential itself, $b(\varepsilon,P)$ is of order ε^2.

This section is concluded with a technical lemma.

LEMMA 2.3. *If $\mathcal{P}$ is closed in total variation and $\mathcal{P}$ is not exponentially convex, then there are $P_1,P_2 \in \mathcal{P}$ such that* (2.6) *holds.*

PROOF. The non-convexity implies the existence of P and $Q \in \mathcal{P}$ such that $\mathcal{P}^{P,Q} - \mathcal{P}$ is nonempty. Let $\mathcal{P}^{P,Q} = \{R_\alpha : \alpha \in [0,1]\}$ then $\{\alpha : R_\alpha \in \mathcal{P}\}$ is closed, since $K(\cdot,\cdot)$ is continuous on $\{R_\alpha\}$ and $K(R_n,R_\alpha) \to 0$ implies convergence in total variation of R_n to R_α, cf. Pinsker (1975), p. 20. Now let α_1,α_2 be the endpoints of the (a) largest open (relative to [0,1]) interval U with $\{R_\alpha : \alpha \in U\} \cap P = \emptyset$. Define

$$P_1 = \begin{cases} R_{\alpha_1} & \text{when } \alpha_1 > 0 \\ R_0 & \text{when } \alpha_1 = 0 \text{ and } R_0 \in P \\ P & \text{when } \alpha_1 = 0 \text{ and } R_0 \notin P \end{cases}$$

and P_2 similarly for α_2, R_1 and Q. Now (2.6) holds for P_1, P_2 in view of Lemma I.2.3. □

3. EXPONENTIAL FAMILIES

3a. Introduction

In this section let $\{P_\theta : \theta \in \Theta^*\}$ be a full k-parameter exponential family, given by its densities with respect to a σ-finite measure μ on $\mathbb{R}^k$

$$dP_\theta(x) = \exp\{\theta' x - \psi(\theta)\} d\mu(x), \quad x \in \mathbb{R}^k, \quad \theta \in \Theta^*$$

as discussed in Section I.2. We shall consider here a subfamily $P = \{P_\theta : \theta \in \Theta\}$, where $\Theta \subset \Theta^*$. Since $\overline{X}_n$ is sufficient for θ, we allow only statistics T_n which are functions of $\overline{X}_n$: $T_n = t_n(\overline{X}_n)$. When $\overline{X}_n \in \Lambda$, T_n can equivalently be written as a function of $\lambda^{-1}(\overline{X}_n)$, since λ is one-to-one on Θ^1. To further simplify the notation we introduce the statistic

$$\hat{\theta}_n^* = \begin{cases} \lambda^{-1}(\overline{X}_n) & \text{when } \overline{X}_n \in \Lambda \\ \theta_\infty & \text{otherwise} \end{cases}$$

where it is convenient to think of θ_∞ as a point outside Θ^1. Note that when $\overline{X}_n \in \text{int}\,\Lambda$, $\hat{\theta}_n^*$ is the MLE of θ when $\Theta = \Theta^*$, cf. Barndorff-Nielsen (1978), Theorem 9.13. We arrive at the representation

$$T_n = \tau_n(\hat{\theta}_n^*) \text{ when } \overline{X}_n \in \Lambda.$$

Here τ_n is a map from Θ^1 into $g(\Theta)$ when estimates of $g(\theta)$ are considered.

As mentioned briefly in Example 2.1, a theorem of Efron and Truax ((1968), Thm. 6) proves condition (2.1) for the MLE $\hat{\theta}_n^*$, when Θ is the full parameter space Θ^*:

THEOREM 3.1. (Efron and Truax). *If* $b < K(\theta)$ (cf. (I.2.14)) *then, as* $n \to \infty$,

$$P_\theta(K(\hat{\theta}_n^*, \theta) > b) = (nb)^{\frac{1}{2}(k-1)} \exp\{-nb + O(1)\}.$$

□

By continuity the theorem also holds for events $K(\hat{\theta}_n^*,\theta) \geq b$. A full-length proof of a generalized version of this theorem was given by Kallenberg (1978). Lemma I.2.5 gives simple conditions ensuring $K(\theta) = \infty$ for each $\theta \in \text{int } \Theta^*$; without any condition, $K(\theta) > 0$ on int Θ^*.

Note that Theorem 3.1 together with Proposition 2.1 solves the optimality problem (see Lemma 1.3) with respect to the inaccuracy rate in full exponential families for continuous g, $T_n = g(\hat{\theta}_n^*)$ being the optimal estimate. This result will be extended to families where Θ is a *convex* subset of Θ^* in Theorem 3.5. These exponential families (with convex Θ) are exponentially convex, cf. the last alinea of Section I.2a; we shall abbreviate the phrase "exponentially convex exponential family" to "convex exponential family".

For non-convex exponential families, the general result of the convex case does not hold: When Θ is closed (and non-convex), condition (2.2) of Proposition 2.1 fails by Lemma 2.2, for at least one pair θ,b, see also Remark 2.4. An example of a non-convex exponential family will be given, however, where an estimate $\{T_n\}$ of $g(\theta)$ exists such that $e(\varepsilon,\theta,\{T_n\}) = b(\varepsilon,\theta)$ for a fixed $\varepsilon > 0$ (Example 3.7).

Curved exponential families form an important subclass of the non-convex case. We shall in Theorem 3.8 prove that estimates which satisfy a linearity condition (these estimates are called linear M-estimates (LME's), cf. Definition 3.1) and are optimal with respect to asymptotic variances, all have the same inaccuracy rate as the MLE, which also satisfies these conditions. The explanation is that optimality with respect to the asymptotic variance implies "convergence" to the MLE, in the class of LME's. In Theorem 3.10 we prove that LME's which do not converge to the MLE indeed have a lower inaccuracy rate, when ε is small enough. Thus, in the class of LME's the MLE can not be improved upon with respect to the inaccuracy rate.

Since the MLE plays a prominent role in this section, we shall introduce this estimate more fully. A detailed account of the consistency properties of the MLE can be found in Berk (1972). As mentioned before, the likelihood of a sample $X_1,\dots,X_n$ is maximized over Θ^* in the point $\hat{\theta}_n^* = \lambda^{-1}(\overline{X}_n)$ when $\overline{X}_n \in \Lambda$. A condition like $b < K(\theta)$ of Theorem 3.1 will be imposed (implicitly or explicitly) in theorems and lemmas of this section to ensure that the event $\overline{X}_n \notin \Lambda$ occurs with negligible probability.

Maximizing the likelihood over a subset Θ of Θ^* is equivalent to minimizing the Kullback-Leibler information $K(\hat{\theta}_n^*,\cdot)$ over Θ (when $\overline{X}_n \in \Lambda$), cf. Efron (1978). If it exists the unique point $\hat{\theta}(\eta)$ minimizing $K(\eta,\cdot)$ over Θ is called the *Kullback-Leibler projection of* η *on* Θ. Thus, when $\hat{\theta}_n^* = \eta \in \Theta^1$ and $\hat{\theta}(\eta)$ exists, $\hat{\theta}(\hat{\theta}_n^*)$ is the MLE of θ on Θ. We shall use the notation

$$\hat{\theta}_n = \hat{\theta}(\hat{\theta}_n^*)$$

and we shall assume that $\hat{\theta}_n$ can be completed to a measurable function. In our theorems and lemma's, care will be taken to ensure that $\hat{\theta}$ exists on the relevant parts of Θ^1. A sufficient condition for the existence of $\hat{\theta}$ at η is given in Lemma 3.12.

When $\hat{\theta}_n$ exists, i.e. $\hat{\theta}_n$ maximizes the likelihood of $\overline{X}_n$ over Θ, the MLE of $g(\theta)$ equals $g(\hat{\theta}_n)$. Berk (1972), Thm. 3.1, proved that the estimate $\hat{\theta}_n$ eventually exists and is consistent for $\theta \in \Theta \cap \text{int}\,\Theta^*$ when Θ is locally compact and locally convex, thus for continuous g, the MLE $g(\hat{\theta}_n)$ is consistent under those conditions.

In part b of this section we present the main results and some examples. The proofs and the more technical lemmas are given in part c.

We conclude this part with a useful lemma on large deviations which is sufficient to prove the assertions made in the examples of subsection 3b. For $A \subset \mathbb{R}^k$, $\delta > 0$ define, abbreviating $\inf\{\|y-x\| : x \in A\}$ to $\|y-A\|$,

$$(A)_\delta = \{y \in \mathbb{R}^k : \|y-A\| \le \delta\} .$$

The *Hausdorff distance* $d_H(A,B)$ of sets $A,B \in \mathbb{R}^k$ is defined as

$$d_H(A,B) = \inf\{\delta : A \subset (B)_\delta \text{ and } A^c \subset (B^c)_\delta\},$$

where A^c denotes the complement $\mathbb{R}^k \setminus A$ of A.

<u>LEMMA 3.2</u>. *Let* $E,E_1,E_2,\ldots \subset \Theta^*$ *and* $\theta_0 \in \text{int}\,\Theta^*$. *If*

(3.1) $$K(\text{cl}\,E,\theta_0) = K(\text{int}\,E,\theta_0) < K(\theta_0)$$

and, for some $\tilde{b}$ *with* $K(E,\theta_0) < \tilde{b} < K(\theta_0)$,

(3.2) $$d_H(E_n \cap \Gamma(\tilde{b},\theta_0), E \cap \Gamma(\tilde{b},\theta_0)) \to 0 \textit{ as } n \to \infty,$$

where Γ *is defined in* (I.2.13), *then*

$$\lim_{n\to\infty} n^{-1} \log P_{\theta_0}(\hat{\theta}_n^* \in E_n) = -K(E,\theta_0). \tag{3.3}$$

The proof is given in part c of this section.

REMARK 3.1. The equality part of condition (3.1) is similar to condition (3.4) of Theorem 3.1 in Groeneboom *et al* (1979). We give examples to illustrate this condition:

EXAMPLE 3.1. (i) Let $\{P_\theta : \theta \in \mathbb{R}\}$ be the normal $N(\theta,1)$ shift family and let $E_n = E = [1,\infty) \cup \{\frac{1}{2}\}$. Since $K(\eta,\theta) = \frac{1}{2}(\eta-\theta)^2$ we have

$$K(E,0) = K(\mathrm{cl}\, E,0) = \tfrac{1}{8}\,,\ K(\mathrm{int}\ E,0) = \tfrac{1}{2}$$

and, by Lemma I.2.6 and $P(\overline{X}_n = \frac{1}{2}) = 0$,

$$\lim_{n\to\infty} -n^{-1} \log P_0(\overline{X}_n \in E_n) = \tfrac{1}{2}.$$

(ii) Take $\{P_\theta : \theta \in \mathbb{R}\}$ and E as in (i). Let $E_n = [1,\infty) \cup [\frac{1}{2},\frac{1}{2}+1/\sqrt{n})$. By an inspection of the proof of Lemma I.2.6 it is clear that

$$\lim_{n\to\infty} -n^{-1} \log P_0(\overline{X}_n \in E_n) = \tfrac{1}{8}\,,$$

thus (3.3) may or may not hold when (3.1) fails.
(iii) Condition (3.1) is not redundant when E_n and E are replaced by int E_n and int E in condition (3.2):

Take for $\{P_\theta\}$ the family of binomial (2,p) distributions, see Example I.2.1. We have $p = \lambda(\theta) = e^\theta/(1+e^\theta)$. Let $E_n = \lambda^{-1}([\frac{3}{4},1) \cup \{\frac{1}{2}\})$, $E = \lambda^{-1}([\frac{3}{4},1))$, then (3.2) holds for int E_n and int E. Now $n\overline{X}_n$ has a binomial $(2n,\frac{1}{2})$ distribution when $\theta = 0$, hence

$$\lim_{n\to\infty} -n^{-1} \log P_0(\hat{\theta}_n^* \in E_n) \le \lim_{n\to\infty} -n^{-1} \log P_0(\overline{X}_n = \tfrac{1}{2}) =$$

$$\lim_{n\to\infty} -n^{-1} \log \left[\binom{2n}{n} 2^{-2n}\right] = 0,$$

whereas $K(E,0) = \frac{3}{4} \log 3 + \log \frac{1}{2}$.

3b. Results and examples

We begin with a lemma that establishes the inaccuracy rate of an estimate as the Kullback-Leibler information of a certain set with respect to θ. As an important corollary we find sufficient conditions for estimates to have the same inaccuracy rate. The proofs are given in 3c.

LEMMA 3.3. *Let* $T_n = \tau_n(\hat{\theta}_n^*)$ *be a consistent estimate of* $g(\theta)$, *let* $\theta \in \Theta$ *and* $\varepsilon > 0$. *If* $b(\varepsilon,\theta) < K(\theta)$ (cf. (1.7), (I.2.14)) *and there exist a continuous function* $\tau : \Theta^1 \to g(\Theta)$ *and a constant* $\tilde{b}$ *with* $b(\varepsilon,\theta) < \tilde{b} < K(\theta)$ *such that*

$$(3.4) \qquad \lim_{n\to\infty} \sup \{\|\tau_n(\eta) - \tau(\eta)\| \ : \ \eta \in \Gamma(\tilde{b},\theta)\} = 0$$

then

$$\text{(i)} \qquad \liminf_{n\to\infty} n^{-1} \log P_\theta(\|T_n - g(\theta)\| > \varepsilon) \geq -K(E(\varepsilon,\theta,\tau),\theta)$$

$$\text{(ii)} \qquad \limsup_{n\to\infty} n^{-1} \log P_\theta(\|T_n - g(\theta)\| > \varepsilon) \leq - \lim_{\xi\uparrow\varepsilon} K(E(\xi,\theta,\tau),\theta)$$

$$\text{(iii)} \qquad e(\varepsilon,\theta,\{\tau(\hat{\theta}_n^*)\}) = K(E(\varepsilon,\theta,\tau),\theta),$$

where

$$E(\varepsilon,\theta,\tau) = \{\eta \in \Theta^1 : \|\tau(\eta) - g(\theta)\| > \varepsilon\}.$$

COROLLARY 3.4. *If the conditions of Lemma* 3.3 *hold and* $K(E(\cdot,\theta,\tau),\theta)$ *is left continuous at* ε, *then*

$$e(\varepsilon,\theta,\{T_n\}) = e(\varepsilon,\theta,\{\tau(\hat{\theta}_n^*)\}).$$

□

REMARK 3.2. When $\Theta = \Theta^*$ and g is continuous, taking $\tau = g$ yields $K(E(\varepsilon,\theta,\tau),\theta) = b(\varepsilon,\theta)$ when $b(\varepsilon,\theta) < K(\theta)$. Consequently, $g(\hat{\theta}_n^*)$ is optimal with respect to the inaccuracy rate when $b(\varepsilon,\theta) < K(\theta)$. As mentioned in the introduction to this section, this also follows from Theorem 3.1 and Proposition 2.1.

REMARK 3.3. It should be noted that condition (3.4) does not prescribe the *speed* of convergence. The uniformity, however, is essential. We give an example:

EXAMPLE 3.2. Let $\{P_\theta : \theta \in \mathbb{R}\}$ be the normal shift family $N(\theta,1)$, let $g(\theta) = \theta$, $\tau(\theta) = \theta$. Define

$$\tau_n(\theta) = \begin{cases} \theta & \text{when } \theta \le 0 \text{ or } \theta \ge 1/n \\ 1+\theta & \text{when } 0 < \theta < 1/n \end{cases}$$

then $\tau_n(\overline{X}_n)$ is a consistent estimate of θ and τ_n converges pointwise to τ, but

$$P_0(|\tau_n(\overline{X}_n)| > 1) > P_0(0 < \overline{X}_n < n^{-1}) > cn^{-\frac{1}{2}}$$

for some positive c, implying $e(1,0,\{\tau_n(\overline{X}_n)\}) = 0$, contrasting with $e(1,0,\overline{X}_n) = \frac{1}{2}$.

Next we give an example where equality holds in assertions (i) and (ii) of Lemma 3.3, but the values are different, and where the estimate is optimal with respect to asymptotic variances but not inaccuracy rate optimal for each ε.

EXAMPLE 3.3. Let $\{P_\theta : \theta \in \mathbb{R}^2\}$ be the family of bivariate normal pms with mean $\theta = (\theta_1,\theta_2)$ and covariance I_2. Define $\Theta = \{(\theta_1,0) : \theta_1 \in \mathbb{R}\}$, let $g(\theta) = \theta_1$ and $\theta_0 = (0,0)$.

Define the map $\tau : \mathbb{R}^2 \to g(\Theta)$ as

$$\tau(\eta_1,\eta_2) = \begin{cases} \eta_1 & \text{if } \eta_2 < 0 \text{ or } \eta_1 \ge 1 \\ 1 & \text{if } \eta_2 \ge \sqrt{1-\eta_1} \text{ and } \eta_1 < 1 \\ \eta_1+\eta_2^2 & \text{if } 0 \le \eta_2 < \sqrt{1-\eta_1} \text{ and } \eta_1 < 1, \end{cases}$$

see fig. 3.1.

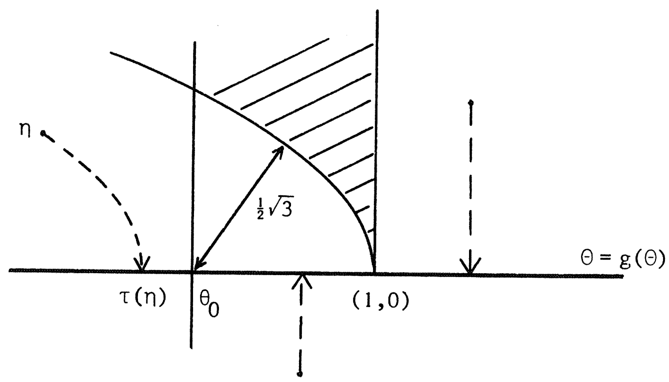

fig. 3.1. For simplicity, $g(\Theta)$ is drawn superimposed on Θ. τ maps the shaded region on the point (1,0). In the region between Θ and the parabole the inverse images under τ are halfparaboles.

Since $K(\eta,\theta) = \frac{1}{2}\|\eta-\theta\|^2$, we have $K(E(1,\theta_0,\tau),\theta_0) = \frac{1}{2}$ and $\lim_{\varepsilon\uparrow 1} K(E(\varepsilon,\theta_0,\tau),\theta_0) = \frac{3}{8}$.

Now define $\{T_n\} = \{\tau_n(\hat{\theta}_n^*)\}$ by

$$\tau_n(\eta) = \begin{cases} \tau(\eta) & \text{when n is even} \\ \tau(\eta) + 1/n & \text{when n is odd} \end{cases}$$

then (i) and (ii) of Lemma 3.3 hold with equality for $\theta = \theta_0$ and $\varepsilon = 1$. This also shows that $\{T_n\}$ is not inaccuracy rate optimal for $\theta = \theta_0$ and $\varepsilon = 1$, $\overline{X}_{1,n}$ being optimal with $e(\varepsilon,\theta,\{\overline{X}_{1,n}\}) = \frac{1}{2}\varepsilon^2$ for all θ,ε. To prove optimality with respect to asymptotic variances, it is sufficient to consider $\tau(\hat{\theta}_n^*)$. Since $\tau(\hat{\theta}_n^*) = \overline{X}_{1,n} + O(\overline{X}_{2,n}^2)$ when $\theta \in \Theta$, $\sqrt{n}(\tau(\hat{\theta}_n^*) - \theta_1)$ is asymptotically normal $N(0,1)$.

Example 3.3 shows that optimality with respect to asymptotic variances, which is a local property, leaves a lot of freedom for the behaviour of the estimate in values of $\hat{\theta}_n^*$ remote from Θ. It is often the behaviour in these non-local values which determines the large deviation properties of the estimate. Another clear illustration of this phenomenon can be found in Example 3.5.

We now treat exponential families with a convex parameter space ("convex exponential families"). Here the possible existence of an estimate of θ which satisfies condition (2.2) of Proposition 2.1 is not "threatened", since convex exponential families are exponentially convex. Indeed we shall prove, under weak conditions on θ,ε and Θ, that (2.2) holds for the MLE $\hat{\theta}_n$ on Θ, implying inaccuracy rate optimality of $\hat{g}_n = g(\hat{\theta}_n)$ for each continuous g.

<u>THEOREM 3.5</u>. *Suppose the parameter space* Θ *of an exponential family is a convex relatively closed subset of* Θ^* *and let* $\hat{g}_n = g(\hat{\theta}_n)$ *whenever the* MLE $\hat{\theta}_n$ *of* θ *exists.*

If g *is continuous then* ($\hat{g}_n$ *is consistent and*)

(3.5) $$e(\varepsilon,\theta,\{\hat{g}_n\}) = b(\varepsilon,\theta)$$

for all ε,θ *satisfying*

(3.6) $$b(\varepsilon,\theta) < K(\theta).$$

<u>REMARK 3.4</u>. The condition that Θ is relatively closed is used to show

existence of $\hat{\theta}_n$, the convexity yields unicity. The rest of the proof is essentially the demonstration that when Θ is convex and $\theta \in \Theta \cap \text{int } \Theta^*$,

$$K(\hat{\theta}(\cdot),\theta) \leq K(\cdot,\theta)$$

on a sufficiently large subset of Θ^*, implying (2.2) for $\hat{\theta}_n$ by Theorem 3.1. The full proof is given in Section 3c.

The remainder of this subsection is devoted to curved exponential families, which form an important subclass of the non-convex exponential families.

A *curved exponential family* is a k-parameter exponential family $\{P_\theta : \theta \in \Theta\}$, $k \geq 2$, such that Θ is the image under a bicontinuous map θ of an interval Ω of the real line; $\Theta = \theta(\Omega)$ is a *curve* in Θ^* (cf. Efron (1975)). Since the family $\{P_\theta : \theta \in \Theta\}$ is parametrized by the one-dimensional parameter $\omega \in \Omega$, it is usually denoted as $\{P_\omega : \omega \in \Omega\}$, where

$$dP_\omega(x) = \exp\{\theta(\omega)'x - \psi(\theta(\omega))\}d\mu(x).$$

We shall assume the following regularity conditions to hold:

(3.7) $\theta(\text{int } \Omega) \subset \text{int } \Theta^*$ and $\theta(\Omega)$ is closed in Θ^*,

(3.8) θ^{-1} is 1-1 and continuous on $\Theta \cap \text{int } \Theta^*$,

(3.9) $\dot{\theta} = \frac{d}{d\omega}\theta(\omega)$ exists and $||\dot{\theta}||$ does not vanish on compact subsets of int Ω,

and

(3.10) $\ddot{\theta}$ exists and is continuous on int Ω.

<u>EXAMPLE 3.4</u>. The family of normal densities with mean $\mu > 0$ and variance $(V\mu)^2$ (constant coefficient of variation) is a curved exponential family: Θ is a subset of the set Θ^* of Example I.2.2, defined as the image of $(0,\infty)$ under

$$\theta_1(\mu) = \frac{1}{V^2\mu}, \quad \theta_2(\mu) = \tfrac{1}{2} - \frac{1}{2V^2\mu^2}, \quad \mu > 0.$$

We shall narrow the scope of our estimation problem as outlined in Section 1 somewhat and only consider estimation of ω. To keep in line with the definitions of Section 1, take

$$g(\theta) = \omega = \theta^{-1}(\theta),$$

where, with an abuse of notation, θ stands for a *point* of Θ and θ^{-1} for the inverse of the *map* $\theta : \Omega \to \Theta$.

An important quantity in smooth one-parameter families is the (*statistical*) *curvature* γ introduced by Efron (1975), which for curved exponential families is given by

$$\gamma_\omega = \left\{\frac{\ddot\theta' \Sigma \ddot\theta}{(\dot\theta' \Sigma \dot\theta)^2} - \frac{(\dot\theta' \Sigma \ddot\theta)^2}{(\dot\theta' \Sigma \dot\theta)^3}\right\}^{\frac{1}{2}},$$

$\dot\theta$, $\ddot\theta$ and Σ evaluated at ω. (In the general situation a similar formula involving cumulants holds, see Efron (1975)).

In curved exponential families, the MLE $\hat\omega_n$ of ω is not a sufficient statistic, being of lower dimension than the minimal sufficient $\overline{X}_n$ (unless γ_ω is identically zero, in which case $\theta(\omega)$ is a straight line in Θ^* and a one-dimensional parametrization of $\{P_\theta : \theta \in \Theta\}$ should have been employed, cf. Section I.2a). Efron (1975) and Efron and Hinkley (1978) argue that γ_ω is a measure for the (second order) difference of the asymptotic variance of the MLE $\hat\omega_n$ and its Cramér-Rao bound $\{ni_\omega\}^{-1}$, where i_ω denotes the Fisher information of X_1 in P_ω. These ideas go back to Fisher (1925) and Rao (1961,1962,1963). Efron (1975) makes this precise in curved exponential families.

Quantitative results connecting the curvature with the difference of the inaccuracy rate $e(\varepsilon,\omega,\{\hat\omega_n\})$ and Bahadur's bound $b(\varepsilon,\omega)$ do not naturally arise (for fixed $\varepsilon > 0$) since the inaccuracy rate of an estimate $\tau_n(\hat\theta_n^*)$ can be determined by the values of τ_n in non-local points, whereas γ_ω is a locally-determined quantity. For the same reason, Fisher information plays no role in the large-deviation properties of estimates either.

A way to make a connection between the inaccuracy rate and Fisher information is to let ε tend to zero as in (1.10) and (1.11). By calculating expansions of $b(\varepsilon,\omega)$ and $e(\varepsilon,\omega,\{\hat\omega_n\})$, Fu (1982) obtained a "second order" optimality result. He proved

$$(3.11) \qquad \lim_{\varepsilon\to 0} \varepsilon^{-4}\{b(\varepsilon,\omega) - e(\varepsilon,\omega,\{\hat\omega_n\})\} = \tfrac{1}{8} i_\omega^2 \gamma_\omega^2 .$$

and showed that this limit is minimal, in symmetric translation families with log-concave density, over the class of translation equivariant estimates.

Note, however, that (3.11) involves taking two limits, first $n \to \infty$ and then $\varepsilon \to 0$.

Kallenberg (1983) proposed a more direct approach, letting $\varepsilon = \varepsilon_n \to 0$ and $n \to \infty$ simultaneously. When $n\varepsilon_n \to \infty$ and $\varepsilon_n^3 n^{\frac{1}{2}} = O(1)$, his results differ from Fu's.

We give a simple example to illustrate (3.11) and to show that the MLE is not necessarily admissible with respect to the inaccuracy rate when ε is large.

EXAMPLE 3.5. Let $\{P_\omega : \omega \in \Omega = [-\pi/2, \pi/2]\}$ be the family of bivariate normal distributions with mean $(\sin \omega,\ 1 - \cos \omega)$ and $\Sigma = I_2$. Note that the curvature exists for $\omega \in \text{int } \Omega$ and equals 1. This model, see fig. 3.2, was essentially introduced by Fisher (1956).

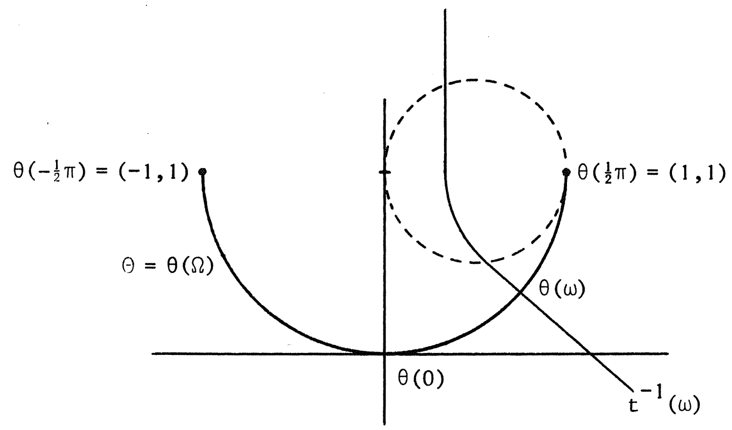

fig. 3.2. Fisher's circle model. The Fisher information and the curvature γ_ω are constant on int Ω and equal to 1.

The MLE $\hat{\omega}_n = \hat{\omega}(\overline{X}_n)$ of ω exists uniquely unless $\overline{X}_n \in \{(0,\theta_2) : \theta_2 \geq 1\}$, which has probability zero. For $\omega \in \text{int } \Omega$ the inverse images $\hat{\omega}^{-1}(\omega)$ of $\hat{\omega}$ are the open half-lines from (0,1) through $\theta(\omega)$. Using plane geometry it follows from $K(\eta,\theta) = \frac{1}{2}\|\eta-\theta\|^2$, Lemma 1.3 and Lemma 3.2 that

$$(3.12) \qquad b(\varepsilon,\omega) = \begin{cases} 2 \sin^2 \frac{1}{2}\varepsilon & \text{when } \varepsilon < \frac{1}{2}\pi + |\omega|, \\ \infty & \text{otherwise} \end{cases}$$

and that

$$(3.13)\qquad e(\varepsilon,\omega,\{\hat{\omega}_n\}) = \begin{cases} \frac{1}{2}\sin^2\varepsilon & \text{when } \varepsilon < \frac{1}{2}\pi, \\ \frac{1}{2} & \text{when } \frac{1}{2}\pi \le \varepsilon < \frac{1}{2}\pi + |\omega|, \\ \infty & \text{otherwise.} \end{cases}$$

Relation (3.11) is now easily verified by Taylor expansion, for each $\omega \in \text{int}\ \Omega$.

A better estimate $T_n = t(\overline{X}_n)$ is defined by describing the "fibres" $t^{-1}(\omega)$:

For $\omega \in [0,\pi/2)$, $t^{-1}(\omega)$ consists of halflines

$$\{(x_1,x_2) : x_2 \le 1 - \sqrt{\tfrac{1}{4} - (x_1 - \tfrac{1}{2})^2},\quad x_1 = (1-x_2)\ \text{tg}\ \omega\}$$

$$\{(x_1,x_2) : x_2 \ge 1,\quad x_1 = 1 - \cos\omega\}$$

and the circle segment with centre (1,1) joining the two halflines. A typical fibre has been drawn in fig. 3.2. For negative values of ω, the fibres form a mirror image of the positive half plane. Using plane geometry it may be verified that $e(\varepsilon,\omega,\{T_n\}) > e(\varepsilon,\omega,\{\hat{\omega}_n\})$ when $\frac{1}{2}(\pi/2 + |\omega|) < \varepsilon < \pi/2 + |\omega|$.

Note that the estimate $\{T_n\}$ is optimal with respect to asymptotic variances at each $\omega \in \text{int}\ \Omega$.

Examples of non-admissibility of the MLE in finite-parameter families are given in Rukhin (1983) and Kester (1981).

In the same "circle model" we give an example of an estimate which attains - at a fixed ω_0 and each $\varepsilon > 0$ - Bahadur's bound. At other values of ω, however, the inaccuracy rate is lower.

<u>EXAMPLE 3.6</u>. Let $\{P_\omega : \omega \in \Omega\}$ be defined as in Example 3.5. To attain Bahadur's bound we should see to it that the (Euclidean) distance of $\{(x_1,x_2) : |t(x_1,x_2) - \omega| > \varepsilon\}$ to $\theta(\omega)$ equals that of $\{\theta(\omega-\varepsilon),\theta(\omega+\varepsilon)\}$ to $\theta(\omega)$, cf. Lemma 1.3, Lemma 3.2. This can be done for a fixed $\omega_0 = 0$, say, and each $\varepsilon > 0$ by taking as the fibres $t^{-1}(\omega)$ straight lines through $\theta(\omega)$, perpendicular to $\theta(\omega) - \theta(\omega_0)$. Note that the resulting estimate is indeed consistent and that the optimality holds in ω_0 only.

Another example shows that it is possible in a curved exponential family to attain Bahadur's bound for a fixed $\varepsilon > 0$, but at each $\omega \in \text{int}\ \Omega$.

EXAMPLE 3.7. Let $\{P_\theta : \theta \in \mathbb{R}^2\}$ be the bivariate normal family with mean $\theta \in \mathbb{R}^2$ and unit covariance matrix. Take the curved subfamily $\{P_\omega : \omega \in \mathbb{R}\}$ as

$$\theta(\omega) = (f(\omega), e^{f(\omega)}),$$

where

$$f(\omega) = \begin{cases} \omega, & \omega \geq 0 \\ -2 \log (1-\omega) + 1 - \frac{1}{1-\omega}, & \omega < 0. \end{cases}$$

Note that f - and hence θ - has continuous second derivatives, thereby satisfying (3.10). Let $\varepsilon = 1$. We construct $T_n = t(\overline{X}_n)$ by its fibres: $t^{-1}(\omega)$ is the straight line through $\theta(\omega)$ which is perpendicular to $\theta(\omega) - \theta(\omega+1)$ (on a strip around $\theta(\Omega)$).

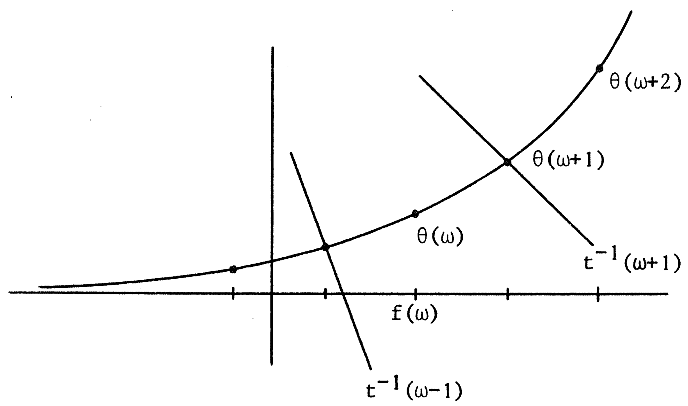

fig. 3.3. $t^{-1}(\omega+1)$ is perpendicular to $\theta(\omega+1) - \theta(\omega+2)$. The distance $\|\theta(\omega+1) - \theta(\omega)\|$ is increasing as a function of ω.

We have $b(1,\omega) = \frac{1}{2} \min (\|\theta(\omega-1) - \theta(\omega)\|^2, \|\theta(\omega+1)-\theta(\omega)\|^2) = \frac{1}{2}\|\theta(\omega-1)-\theta(\omega)\|^2$ and

$$(3.14) \qquad e(1,\omega,\{T_n\}) = \tfrac{1}{2} \min (\|E_1-\theta(\omega)\|^2, \|E_2 - \theta(\omega)\|^2)$$

where $E_1 = \{x : t(x) < \omega-1\}$, $E_2 = \{x : t(x) > \omega+1\}$. The example has been constructed so that the minimum in (3.14) occurs for E_1 and then equals b since $t^{-1}(\omega-1)$ is perpendicular to $\theta(\omega-1)-\theta(\omega)$.

These examples indicate that in the class of consistent estimates, a uniformly best estimate with respect to the inaccuracy rate does not exist.

Estimates which are optimal with respect to asymptotic variances play a prominent role in estimation, both in theory and in practice. It would therefore be interesting to investigate their performance with respect to the inaccuracy rate. Since the asymptotic variance is a local property, however, optimality in that sense has little connection with the inaccuracy rate, which is often non-locally determined, see Example 3.5. We shall impose a linearity condition to link the "non-local behaviour" of the estimates to their "local behaviour". Estimates $\{T_n\} = \{t_n(\overline{X}_n)\}$ where the inverse images under t_n are essentially hyperplanes in Λ will be called linear M-estimates (LME's), cf. Definition 3.1. We shall prove that LME's which are optimal with respect to asymptotic variances have the same inaccuracy rate as the MLE for each ε in an interval $(0,\varepsilon_0)$, due to "convergence" of these estimates to the MLE.

Moreover, we prove that LME's which do not converge to the MLE are not locally inaccuracy rate optimal (cf. (1.10)). It follows that the inaccuracy rate of these LME's is lower than that of the MLE for each ε in an interval $(0,\varepsilon_0)$, since the MLE is locally inaccuracy rate optimal, cf. Bahadur (1960b, 1967).

Note that these results imply a stronger superiority of the MLE than the property proved in Fu (1982), Theorem 3.3, for MLE's in translation families.

We now define LME's more formally. To simplify the notation we write

$$m(\omega) = \lambda(\theta(\omega))$$

when $\omega \in \Omega \cap \theta^{-1}(\Theta^1)$.

<u>DEFINITION 3.1</u>. An estimate $\{T_n\} = \{t_n(\overline{X}_n)\}$ is called a *linear M-estimate* (LME) when functions $\rho_n : \Omega \to \mathbb{R}^k$, continuous strictly monotone functions $b_n : \Omega \to \Omega$ and for each compact interval $B \subset \text{int}\,\Omega$, a positive $d = d(B)$ exist such that when $\omega \in B$ and

$$(3.15) \qquad \|x - m(\omega)\| < d,$$

the relation of $t_n(x)$ and ω is given by

$$(3.16) \qquad t_n(x) \gtreqless \omega \Leftrightarrow \rho_n(\omega)'(x - m[b_n(\omega)]) \gtreqless 0.$$

Moreover, we require that (3.15) and (3.16) hold for $x = m[b_n(\omega)]$ when $\omega \in B$. The constant d will be called the radius of T_n on B. For definiteness we assume $||\rho_n|| \equiv 1$.

In figure 3.4, the relation (3.16) is illustrated. Note that $\theta(\Omega)$ has been mapped onto $m(\Omega)$.

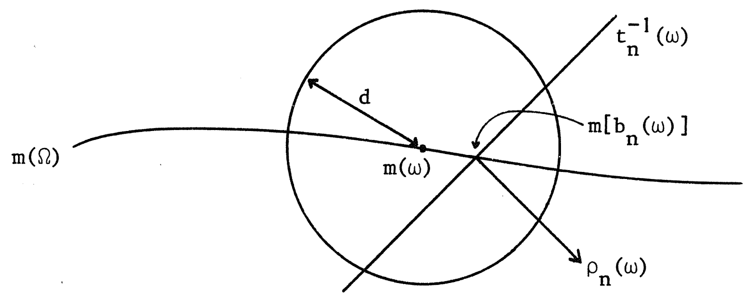

fig. 3.4. Inside the ball $\{x : ||x - m(\omega)|| < d\}$, $t_n^{-1}(\omega)$ is the hyperplane perpendicular to $\rho_n(\omega)$.

REMARK 3.5. M-estimates $T_n = t_n(X_1,\dots,X_n)$ are usually defined on the sample space X^n as the (or an appropriate) zero of $\Sigma_{i=1}^n \psi(X_i,\cdot)$, for some function ψ. The right hand side of (3.16) resembles this, but involves the sufficient reduction of X^n to the space of means. The estimates defined in definition 3.1 are called *linear* since the inverse images $t_n^{-1}(\omega)$ are (k-1)-dimensional hyperplanes (in a neighbourhood of $m(\omega)$), see fig. 3.4. An intuitively appealing consequence of the inequalities in (3.16) is that $t_n(x) < t_n(y)$ implies $t_n(x) < t_n(\frac{1}{2}(x+y)) < t_n(y)$ (when x,y and $\frac{1}{2}(x+y)$ are close enough to $m(\Omega)$).

By (3.16) we have that $t_n(m(\omega)) = b_n^{-1}(\omega)$ (therefore, the quantity $b_m^{-1}(\omega) - \omega$ could be called the Fisher bias of T_n). Lemma 3.6 provides convergence of $b_n(\omega)$ to ω for each consistent LME; the proof is deferred to Section 3c.

LEMMA 3.6. *If* $\{T_n\}$ *is both an* LME *and consistent then the functions* b_n *converge to the identity on* int Ω.

In Examples 3.5, 3.6 and 3.7, the estimates considered are LME's. In these examples it is also easily verified, that the MLE is an LME. This is more generally true:

LEMMA 3.7. *For each compact interval* B *of* int Ω *there exists a* $d > 0$ *such that* $\hat{\omega}(x)$ *exists uniquely when* $\|x - m(B)\| < d$ *and such that the* MLE $\{\hat{\omega}_n\}$ *is an* LME, *with radius* d *on* B.

The proof, given in 3c, first shows the unique existence of $\hat{\omega}(x)$. Next, (3.16) is found as the expression of the fact that the derivative $\dot{\theta}(\omega)'(x - m(\omega))$ of the likelihood equals zero for $\omega = \hat{\omega}(x)$ and is positive (negative) when $\omega < \hat{\omega}(x)$ $(\omega > \hat{\omega}(x))$. It follows that $\rho_n(\omega) = \dot{\theta}(\omega) / \|\dot{\theta}(\omega)\|$ in case of the MLE.

REMARK 3.6. LME's with $\rho_n \neq \dot{\theta} / \|\dot{\theta}\|$ are found in Examples 3.6 and 3.7: in Example 3.6 we have $\rho_n(\omega) = (\theta(\omega) - \theta(0)) / \|\theta(\omega) - \theta(0)\|$, in 3.7 it is $\rho_n(\omega) = (\theta(\omega) - \theta(\omega - 1)) / \|\theta(\omega) - \theta(\omega - 1)\|$ $((b_n(\omega) \equiv \omega$ in both examples).

The following theorem states in effect that LME's which are optimal with respect to the asymptotic variance have the same inaccuracy rate as the MLE, when ε is small enough.

THEOREM 3.8. *If an* LME $\{T_n\}$ *is optimal with respect to the asymptotic variance, i.e.*

$$(3.17) \qquad n^{\frac{1}{2}} i_\omega^{\frac{1}{2}} (T_n - \omega) \xrightarrow{\mathcal{D}_{P_\omega}} N(0,1) \text{ as } n \to \infty,$$

then for each compact interval $C \subset$ int Ω *there is an* $\varepsilon_0 > 0$ *such that*

$$(3.18) \qquad e(\varepsilon,\omega,\{T_n\}) = e(\varepsilon,\omega,\{\hat{\omega}_n\})$$

for each $\varepsilon < \varepsilon_0$ *and all* $\omega \in C$.

The proof is given in subsection c. It amounts essentially to the demonstration that (3.17) implies $\rho_n(\omega) \to \dot{\theta}(\omega) / \|\dot{\theta}(\omega)\|$ and $b_n(\omega) \to \omega$ for each $\omega \in$ int Ω, hence the estimate "converges" to the MLE. The proof is then completed by an application of Lemma 3.9 which may be of independent interest and is given here. In contrast to the condition in

Lemma 3.3, the convergence (3.19) need not be uniform.

LEMMA 3.9. *Let* $\{T_n\} = \{t_n(\overline{X}_n)\}$ *be a consistent* LME *such that*

$$(3.19) \qquad \lim_{n\to\infty} \rho_n(\omega) = \rho(\omega), \qquad \omega \in \text{int}\ \Omega.$$

(i) *For each compact interval* C *of* int Ω *there is an* $\varepsilon_0 > 0$ *such that*

$$e(\varepsilon,\omega,\{T_n\}) = K(E_-(\varepsilon) \cup E_+(\varepsilon),\theta(\omega))$$

for each $\varepsilon < \varepsilon_0$ *and all* $\omega \in C$, *where*

$$(3.20) \qquad \begin{aligned} E_-(\varepsilon) &= \{\eta \in \Theta^* : \rho(\omega-\varepsilon)'(\lambda(\eta) - m(\omega-\varepsilon)) < 0\}, \\ E_+(\varepsilon) &= \{\eta \in \Theta^* : \rho(\omega+\varepsilon)'(\lambda(\eta) - m(\omega+\varepsilon)) > 0\}. \end{aligned}$$

(ii) *Moreover,*

$$(3.21) \qquad K(E_+(\varepsilon),\theta(\omega)) = K(\eta_\varepsilon,\theta(\omega)),$$

with η_ε *satisfying* $\rho(\omega+\varepsilon)'(\lambda(\eta_\varepsilon) - m(\omega+\varepsilon)) = 0$ *and, for a* $t > 0$,

$$(3.22) \qquad \eta_\varepsilon = \theta(\omega) + t\rho(\omega+\varepsilon).$$

A similar relation holds for $E_-(\varepsilon)$.

This lemma is proved in Section 3c.

REMARK 3.7. For "convergent" LME's (satisfying $b_n(\omega) \to \omega$ and (3.19)), the inaccuracy rate is apparently determined by the function ρ, when ε is small enough.

By Lemma 3.9, LME's which converge to the MLE have the same inaccuracy rate as the MLE. The next theorem says that other LME's are not locally inaccuracy rate optimal (cf. (1.11)). Since the MLE is locally inaccuracy rate optimal, this implies that the inaccuracy rate of LME's which do not converge to the MLE is lower than the inaccuracy rate of the MLE for each ε in an interval $(0,\varepsilon_0)$. The theorem will be proved in Section 3c.

THEOREM 3.10. *Let* $\{T_n\}$ *be a consistent* LME *and let* $\omega_0 \in \text{int}\ \Omega$. *If* $\{\rho_n(\omega_0)\}$ *does not converge to* $\dot{\theta}(\omega_0)/\|\dot{\theta}(\omega_0)\|$ *then*

$$\lim_{\varepsilon\to 0} e(\varepsilon,\omega_0,\{T_n\}) \,/\, e(\varepsilon,\omega_0,\{\hat{\omega}_n\}) < 1.$$

3c. Proofs

PROOF of Lemma 3.2. Let $\tilde{b}$ satisfy $K(E,\theta_0) < \tilde{b} < K(\theta_0)$ such that (3.2) holds. Define

$$F = E \cap \Gamma(\tilde{b},\theta_0), \quad F_n = E_n \cap \Gamma(\tilde{b},\theta_0),$$

then $K(\mathrm{cl}\,F,\theta_0) = K(\mathrm{int}\,F,\theta_0) = K(E,\theta_0)$ and, by Theorem 3.1,

$$|P_{\theta_0}(\hat{\theta}_n^* \in E_n) - P_{\theta_0}(\hat{\theta}_n^* \in F_n)| \le P_{\theta_0}(\hat{\theta}_n^* \notin \Gamma(\tilde{b},\theta_0)) = e^{-n\tilde{b}+o(n)}.$$

Since $K(E,\theta_0) < \tilde{b}$ it is thus sufficient to prove

$$\lim_{n\to\infty} n^{-1} \log P_{\theta_0}(\hat{\theta}_n^* \in F_n) = -K(E,\theta_0).$$

Let $c_1 > K(E,\theta_0)$ and choose $\theta \in \mathrm{int}\,F$ with $K(\theta,\theta_0) < c_1$. Let $\delta > 0$ satisfy $U(\delta,\theta) = \{\eta : \|\eta-\theta\| < \delta\} \subset F$. Since $d_H(F_n,F) \to 0$ we have $F_n^c \subset (F^c)_{\delta/2}$ when n is large enough, which implies $U(\delta/2,\theta) \subset F_n$ by $U(\delta/2,\theta) \cap (F^c)_{\delta/2} = \emptyset$. Furthermore,

$$P_\theta(\hat{\theta}_n^* \in U(\delta/2,\theta)) \to 1 \quad \text{as } n \to \infty,$$

yielding by Lemma 1.1

$$\liminf_{n\to\infty} n^{-1} \log P_{\theta_0}(\hat{\theta}_n^* \in F_n) \ge$$
$$\liminf_{n\to\infty} n^{-1} \log P_{\theta_0}(\hat{\theta}_n^* \in U(\delta/2,\theta)) \ge -K(\theta,\theta_0) > -c_1.$$

Now choose $c_2 < K(E,\theta_0)$. Since $c_2 < K(\mathrm{cl}\,F,\theta_0)$, $d = \inf\{\|\eta-\theta\| : \eta \in \Gamma(c_2,\theta_0),\ \theta \in \mathrm{cl}\,F\}$ is positive. Thus $\Gamma(c_2,\theta_0) \cap (F)_{d/2} = \emptyset$, implying $\Gamma(c_2,\theta_0) \cap F_n = \emptyset$ when n is large enough. By Theorem 3.1 we obtain

$$P_{\theta_0}(\hat{\theta}_n^* \in F_n) \le P_{\theta_0}(\hat{\theta}_n^* \notin \Gamma(c_2,\theta_0)) = e^{-nc_2+o(n)} \quad \text{as } n \to \infty.$$

Since c_1 and c_2 can be chosen arbitrarily close to $K(E,\theta_0)$, (3.3) follows. □

PROOF of Lemma 3.3. Let $e = K(E(\varepsilon,\theta,\tau),\theta)$. We prove first that $e \le b(\varepsilon,\theta)$: Since τ is continuous and $\hat{\theta}_n^* \xrightarrow{P_\eta} \eta$ on $\mathrm{int}\,\Theta^1$, we have for $\eta \in \mathrm{int}\,\Theta^1$

$$\tau(\hat{\theta}_n^*) \xrightarrow{P_\eta} \tau(\eta),$$

and hence by (3.4), $\tau_n(\hat{\theta}_n^*) \xrightarrow{P_\eta} \tau(\eta)$ for $\eta \in \text{int}\,\Gamma(\tilde{b},\theta)$. Consistency of $\tau_n(\hat{\theta}_n^*)$ now implies $\tau(\eta) = g(\eta)$ for each $\eta \in \Theta \cap \text{int}\,\Gamma(\tilde{b},\theta)$, hence $e \le b(\varepsilon,\theta)$. Lemma 3.2 now implies assertion (iii) since (3.1) holds for $E = E(\varepsilon,\theta,\tau)$.

Now let $c > 0$ with $e + c < \tilde{b}$.

By continuity of τ and $K(\cdot,\theta)$ there exists an $\eta \in \Gamma(\tilde{b},\theta)$ such that $e < K(\eta,\theta) < e + c$ and $\|\tau(\eta) - g(\theta)\| > \varepsilon$. Moreover, there is an open U with $\eta \in U$ such that $e < K(\cdot,\theta) < \tilde{b}$ and $\|\tau(\cdot) - g(\theta)\| > \varepsilon + c_1$ on U for some positive c_1. Condition (3.4) now ensures that $\|\tau_n(\cdot) - g(\theta)\| > \varepsilon$ on U when n is large enough and hence it follows from $\hat{\theta}_n^* \xrightarrow{P_\eta} \eta$ that as $n \to \infty$

$$P_\eta(\|\tau_n(\hat{\theta}_n^*) - g(\theta)\| > \varepsilon) \ge P_\eta(\hat{\theta}_n^* \in U) \to 1.$$

Now apply Lemma 1.1 to obtain

$$\liminf_{n\to\infty} n^{-1} \log P_\theta(\|T_n - g(\theta)\| > \varepsilon) \ge -K(\eta,\theta) > -e-c.$$

Since c was arbitrarily small, (i) is proved.

It remains to prove (ii). Let $\xi < \varepsilon$. Since $\tau_n \to \tau$ uniformly on $\Gamma(\tilde{b},\theta)$, we have when n is large enough, for each $\eta \in \Gamma(\tilde{b},\theta)$,

$$\|\tau(\eta) - g(\theta)\| \le \xi \Rightarrow \|\tau_n(\theta) - g(\theta)\| \le \varepsilon,$$

implying

$$\begin{aligned} P_\theta(\|T_n - g(\theta)\| > \varepsilon) &\le P_\theta(\|\tau_n(\hat{\theta}_n^*) - g(\theta)\| > \varepsilon,\ \hat{\theta}_n^* \in \Gamma(\tilde{b},\theta)) \\ &\quad + P_\theta(\hat{\theta}_n^* \notin \Gamma(\tilde{b},\theta)) \\ &\le P_\theta(\|\tau(\hat{\theta}_n^*) - g(\theta)\| > \xi) + P_\theta(\hat{\theta}_n^* \notin \Gamma(\tilde{b},\theta)) \\ &\le e^{-nK(E(\xi,\theta,\tau),\theta)+o(n)} + e^{-n\tilde{b}+o(n)}, \end{aligned}$$

where the last inequality holds by (iii) and Theorem 3.1. The observation that $\tilde{b} > b(\varepsilon,\theta) \ge e \ge K(E(\xi,\theta,\tau),\theta)$ completes the proof. □

REMARK 3.8. The proofs of (i) and (ii) above are very similar to the proof of Lemma 3.2. The latter lemma cannot be invoked directly however, since condition (3.4) does not imply convergence in Hausdorff distance of the sets $E(\varepsilon,\theta,\tau_n)$ to $E(\varepsilon,\theta,\tau)$, see Example 3.3.

The proof of Theorem 3.5 is preceded by some technical lemmas.

LEMMA 3.11. *Let* Θ *be a convex subset of* Θ^* *and let* $\eta \in \text{int}\,\Theta^*$.
If $\hat{\theta}(\eta)$ *exists and* $\hat{\theta}(\eta) \in \text{int}\,\Theta^*$ *then*

$$K(\hat{\theta}(\eta),\theta) \leq K(\eta,\theta) \tag{3.23}$$

for each $\theta \in \Theta \cap \text{int}\,\Theta^*$.

PROOF. Let $\theta \in \Theta \cap \text{int}\,\Theta^*$. Suppose $\hat{\theta}(\eta)$ exists and $\hat{\theta}(\eta) \in \text{int}\,\Theta^* \cap \Theta$. By convexity of Θ, the function $t \mapsto K(\eta,t\hat{\theta}(\eta) + (1-t)\theta)$ is minimal for $t = 1$. Consequently, its left derivative is nonpositive at $t = 1$:

$$(\hat{\theta}(\eta) - \theta)'(\lambda(\hat{\theta}(\eta)) - \lambda(\eta)) \leq 0.$$

Application of (I.2.12) yields (3.23). □

The next lemma establishes existence of the MLE.

LEMMA 3.12. *Let* Θ *be a relatively closed convex subset of* Θ^* *and let* $\eta \in \text{int}\,\Theta^*$.

If $K(\eta,\theta) < K(\theta)$ *for some* $\theta \in \Theta$, *then the Kullback-Leibler projection* $\hat{\theta}(\eta)$ *exists and* $\hat{\theta}(\eta) \in \text{int}\,\Theta^*$, *thus* $\hat{\theta}_n$ *exists when*

$$\overline{X}_n \in \lambda\left(\bigcup_{\theta\in\Theta} \{\eta : K(\eta,\theta) < K(\theta)\}\right).$$

PROOF. Let $\eta \in \text{int}\,\Theta^*$ and $\theta \in \Theta$ satisfy $K(\eta,\theta) < K(\theta)$, then the closed Kullback-Leibler ball $\Gamma(K(\eta,\theta),\theta)$ is a subset of $\text{int}\,\Theta^*$ and compact, hence $K(\eta,\cdot)$ attains its infimum on the compact set $\Gamma(K(\eta,\theta),\theta) \cap \Theta$, say in π. Now we prove that $K(\eta,\xi) \geq K(\eta,\pi)$ for all $\xi \in \Theta$. Fix $\xi \in \Theta$, suppose $K(\eta,\xi) < \infty$ and let $\delta > 0$. Define $\xi_\alpha = \alpha\xi + (1-\alpha)\theta$ and let $\tilde{\alpha} < 1$ satisfy $K(\eta,\xi_{\tilde{\alpha}}) < K(\eta,\xi) + \delta$. Now $K(\eta,\cdot)$ attains its infimum (unique by strict convexity) on the compact convex set $\{\xi_\alpha : 0 \leq \alpha \leq \tilde{\alpha}\}$, say in ξ_{α^*}. Since $\xi_{\alpha^*} \in \text{int}\,\Theta^*$, Lemma 3.11 applies and it follows that $\xi_{\alpha^*} \in \Gamma(K(\eta,\theta),\theta) \cap \Theta$, thus

$$K(\eta,\pi) \leq K(\eta,\xi_{\alpha^*}) \leq K(\eta,\xi_{\tilde{\alpha}}) < K(\eta,\xi) + \delta.$$

Unicity of $\hat{\theta}(\eta) = \pi$ follows from the convexity of Θ and the strict convexity of $K(\eta,\cdot)$. □

PROOF of Theorem 3.5. Let $\varepsilon > 0$, $\theta \in \Theta$ satisfy (3.6) and choose $\eta \in \Theta \cap \text{int } \Theta^*$ with $\|g(\eta) - g(\theta)\| > \varepsilon$. Then, by the result of Berk (1972) mentioned in the introduction to this section (p. 33), $g(\hat{\theta}_n)$ is consistent at η, hence by Lemma 1.1 (taking $A_n = \{(x_1,\ldots,x_n) : \hat{\theta}(\bar{x}_n)$ exists and $\|g(\hat{\theta}(\bar{x}_n)) - g(\theta)\| > \varepsilon\}$, $P = P_\theta, Q = P_\eta$) we have

$$\liminf_{n\to\infty} n^{-1} \log P_\theta(\|g(\hat{\theta}_n) - g(\theta)\| > \varepsilon) \geq -K(\eta,\theta). \tag{3.24}$$

Since (3.24) holds for η's with $K(\eta,\theta)$ arbitrarily close to $b(\varepsilon,\theta)$, we obtain

$$e(\varepsilon,\theta,\{\hat{g}_n\}) \leq b(\varepsilon,\theta).$$

It remains to prove

$$\limsup_{n\to\infty} n^{-1} \log P_\theta(\|\hat{g}_n - g(\theta)\| > \varepsilon) \leq -b(\varepsilon,\theta).$$

Let $A = \{\eta \in \Theta^* : K(\eta,\theta) < b(\varepsilon,\theta)\}$ then $\hat{\theta}(\cdot)$ exists on A and $\hat{\theta}(A) \subset \text{int } \Theta^*$ by Lemma 3.12. We obtain

$$P_\theta(\|\hat{g}_n - g(\theta)\| > \varepsilon) \leq P_\theta(\|g(\hat{\theta}_n) - g(\theta)\| > \varepsilon,\ \hat{\theta}_n^* \in A) + P_\theta(\hat{\theta}_n^* \notin A). \tag{3.25}$$

However, $\hat{\theta}_n^* \in A$ implies $\hat{\theta}_n \in A$ by Lemma 3.11 which in view of the definitions of $b(\varepsilon,\theta)$ and A implies that $\|g(\hat{\theta}_n) - g(\theta)\| \leq \varepsilon$. It follows that the first term in the right hand side of (3.25) equals zero. The second term equals $\exp\{-nb(\varepsilon,\theta) + o(n)\}$ as $n \to \infty$, by Theorem 3.1. □

REMARK 3.9. The above proof contains a proof of Proposition 2.1 in this special case. That proposition could not be applied directly since it assumes consistency of $\{\tilde{P}_n\}$ for *each* $P \in \mathcal{P}$.

PROOF of Lemma 3.6. Let $\omega \in \text{int } \Omega$. Since $n^{\frac{1}{2}}(\bar{X}_n - m(\omega))$ is asymptotically normal with nonsingular covariance matrix, we have

$$P_\omega(\rho_n(b_n^{-1}(\omega))'(\bar{X}_n - m(\omega)) > 0) \to \tfrac{1}{2} \tag{3.26}$$

implying by (3.16) that

$$P_\omega(t_n(\bar{X}_n) > b_n^{-1}(\omega)) \to \tfrac{1}{2} \quad \text{as } n \to \infty.$$

Since consistency of $\{T_n\}$ implies for each $\varepsilon > 0$

$$P_\omega(t_n(\overline{X}_n) > \omega + \varepsilon) \to 0 \quad \text{as } n \to \infty,$$

we obtain $\limsup_{n\to\infty} b_n^{-1}(\omega) \le \omega$. Together with the analogous result from (3.26) with the inequality reversed we have

$$\lim_{n\to\infty} b_n^{-1}(\omega) = \omega$$

and the lemma is proved. □

PROOF of Lemma 3.7. Let B be a compact interval of int Ω and choose a compact subset A of int Θ^* such that $\theta(B) \subset$ int A and such that $\{\omega \in \Omega : \theta(\omega) \in A\}$ is a compact subset of int Ω. There exists a $d_1 > 0$ such that $\|x - m(B)\| < d_1$ implies

$$K(\lambda^{-1}(x), \Theta \cap A^c) > \tfrac{1}{2}K(\theta(B), A^c) > K(\lambda^{-1}(x), \theta(B)).$$

It follows that for $\|x - m(B)\| < d_1$, the infimum of $K(\lambda^{-1}(x), \cdot)$ over Θ is attained in a point $\theta(\hat{\omega})$, say, in the compact set $\Theta \cap A$, where $\hat{\omega} = \hat{\omega}(x)$.

We prove that $d > 0$ and $\delta > 0$ exist such that for $\|x - m(B)\| < d$,

$$(3.27) \qquad |\omega - \hat{\omega}| \le \delta \Rightarrow \frac{d^2}{d\omega^2} K(\lambda^{-1}(x), \theta(\omega)) > 0$$

and

$$(3.28) \qquad |\omega - \hat{\omega}| > \delta \Rightarrow K(\lambda^{-1}(x), \theta(\omega)) > K(\lambda^{-1}(x), \theta(\hat{\omega})),$$

implying unicity of the minimizing $\hat{\omega}(x)$. Let

$$c_1 = \inf \{\dot{\theta}(\omega)' \ddagger_\omega \dot{\theta}(\omega) : \omega \in \theta^{-1}(\Theta \cap A)\}$$

which is positive by (3.9) and nonsingularity of $\ddagger$. Now

$$\frac{d^2}{d\omega^2} K(\lambda^{-1}(x), \theta(\omega)) = -\ddot{\theta}(\omega)'(x - m(\omega)) + \dot{\theta}(\omega)' \ddagger_\omega \dot{\theta}(\omega),$$

hence it is sufficient to show that $\ddot{\theta}(\omega)'(x - m(\omega)) < c_1$ for $|\omega - \hat{\omega}| \le \delta$.

By Lemma I.2.4 it holds that for some $c_2 > 0$

$$\|x - m(\hat{\omega})\| \le c_2\|x - m(B)\| \text{ for all } x \in \lambda(A)$$

and by Lemma I.2.4 and (3.9) there is a $c_3 > 0$ such that

$$\|m(\hat{\omega}) - m(\omega)\| \le c_3|\hat{\omega} - \omega|.$$

Now choose $\delta = \frac{1}{2}c_1/(c_3c_4)$ and $d_2 = \frac{1}{2}c_1/(c_2c_4)$ where c_4 is an upper bound on $\|\ddot{\theta}(\omega)\|$ provided by (3.10), then (3.27) holds when $\|x - m(B)\| < d_2$. Next, let $|\omega - \hat{\omega}| > \delta$. By (3.8), $c_5 > 0$ exists such that

$$|\omega - \hat{\omega}| > \delta \Rightarrow \|\theta(\omega) - \theta(\hat{\omega})\| > c_5.$$

Hence,

$$(3.29) \qquad \|\lambda^{-1}(x) - \theta(\omega)\| \geq \|\theta(\hat{\omega}) - \theta(\omega)\| - \|\lambda^{-1}(x) - \theta(\hat{\omega})\| > \tfrac{1}{2}c_5$$

when $\|x - m(B)\|$ is so small that $\|\lambda^{-1}(x) - \theta(\hat{\omega})\| < \frac{1}{2}c_5$. Again by Lemma I.2.4, (3.29) implies

$$K(\lambda^{-1}(x), \theta(\omega)) > c_6 > 0.$$

Now choose $0 < d < d_2$ such that $\|x - m(B)\| < d$ implies (3.29) and $K(\lambda^{-1}(x), \theta(\hat{\omega})) \leq c_6$, establishing (3.28).

We proceed to prove $\hat{\omega}_n = \hat{\omega}(\overline{X}_n)$ to be an LME. Let $t_n = \hat{\omega}$, $\rho_n = \dot{\theta} / \|\dot{\theta}\|$ and $b_n(\omega) = \omega$; we show that (3.16) holds when d is small enough. Let $\omega_0 \in B$. By Lemma I.2.4, both $K(\lambda^{-1}(x), \theta(\hat{\omega}(x)))$ and $K(\lambda^{-1}(x), \theta(\omega_0))$ tend to zero as $\|x - m(\omega_0)\| \to 0$ hence there is a $0 < \tilde{d} < d$ such that

$$\|x - m(\omega_0)\| < \tilde{d} \Rightarrow |\hat{\omega}(x) - \omega_0| < \delta$$

implying by (3.27) that $\frac{d}{d\omega} K(\lambda^{-1}(x), \theta(\omega_0)) = -\dot{\theta}(\omega_0)'(x - m(\omega_0))$ is negative, zero or positive as required by (3.16). □

The proof of Theorem 3.8 will be preceded by a technical lemma and the proof of Lemma 3.9.

<u>LEMMA 3.13</u>. *Let* B *be a compact interval of* int Ω. *Uniformly for* LME's $\{T_n\} = \{t(\overline{X}_n)\}$ *with radius* $\geq$ d *on* B *and*

$$(3.30) \qquad \|m[b(\omega)] - b(\omega)\| < d/3$$

it holds for $\omega_k \in B$, $k = 1,2,\ldots$ *that*

$$\|\rho(\omega_k) - \rho(\omega_0)\| = O(\|m[b(\omega_k)] - m[b(\omega_0)]\|)$$

as $\omega_k \to \omega_0 \in B$.

<u>PROOF</u>. Suppose $\omega_0 < \omega_k$ for each k and define

(3.31) $$y_k = m[b(\omega_0)] + \tfrac{1}{3}d\nu_k/\|\nu_k\|,$$

where

$$\nu_k = \rho(\omega_k) - [\rho(\omega_0)'\rho(\omega_k)]\rho(\omega_0)$$

is the part of $\rho(\omega_k)$ that is perpendicular to $\rho(\omega_0)$. Since $\|y_k - m(\omega_0)\| < d$ by (3.30) and (3.31) we have $t(y_k) = \omega_0$. When k is large enough it also holds that $\|y_k - m(\omega_k)\| < d$, implying by $t(y_k) < \omega_k$ and Definition 3.1 that

$$\rho(\omega_k)'(\tfrac{1}{3}d\nu_k/\|\nu_k\| + m[b(\omega_0)] - m[b(\omega_k)]) < 0,$$

which by evaluation of $\rho(\omega_k)'\nu_k$ leads to

(3.32) $$\tfrac{1}{3}d\|\nu_k\| < \rho(\omega_k)'(m[b(\omega_k)] - m[b(\omega_0)]).$$

Now $\|\nu_k\| = \{1 - [\rho(\omega_k)'\rho(\omega_0)]^2\}^{\frac{1}{2}}$ hence (3.32) implies that $\rho(\omega_k)'\rho(\omega_0)$ tends to 1 or to −1 as $k \to \infty$. To exclude the latter possibility take $x = m[b(\omega_0)] - \frac{1}{3}d\rho(\omega_0)$, then $t(x) < \omega_0 < \omega_k$, thus, when k is large enough,

$$\rho(\omega_k)'(m[b(\omega_0)] - \tfrac{1}{3}d\rho(\omega_0) - m[b(\omega_k)]) < 0$$

which can be rewritten as

(3.33) $$\tfrac{1}{3}d\rho(\omega_k)'\rho(\omega_0) > \rho(\omega_k)'(m[b(\omega_0)] - m[b(\omega_k)]).$$

Since the right hand side of (3.33) tends to zero, only the possibility $\rho(\omega_k)'\rho(\omega_0) \to 1$ remains.

To complete the proof, observe that when $\rho(\omega_k)'\rho(\omega_0) > 0$,

$$\|\rho(\omega_k) - \rho(\omega_0)\|^2 = 2 - 2\rho(\omega_k)'\rho(\omega_0)$$
$$\leq 2(1 - [\rho(\omega_k)'\rho(\omega_0)]^2) = 2\|\nu_k\|^2$$

and combine this with (3.32). When $\omega_k < \omega_0$ the proof is analogous. □

<u>PROOF of Lemma 3.9</u>. Let C be a compact interval of int Ω, choose a compact interval B of int Ω such that $C \subset$ int B and choose a compact convex subset A of int Θ^* such that $\theta(B) \subset$ int A. Let d be the radius of $\{T_n\}$ on B and let $b_0 > 0$ satisfy $K(\eta,\theta(C)) < b_0 \Rightarrow \eta \in A$ and also $\omega \in B$, $K(\eta,\theta(\omega)) < b_0 \Rightarrow \|\lambda(\eta) - m(\omega)\| < d/2$. Choose $\varepsilon_0 > 0$ such that $K(\theta(\omega+\varepsilon_0),\theta(\omega)) < b_0$ and $K(\theta(\omega-\varepsilon_0),\theta(\omega)) < b_0$ for each $\omega \in C$ and such that $|\omega - \omega_1| \leq \varepsilon_0$ implies $K(\theta(\omega_1),\theta(\omega)) \leq b_0/2$ for each $\omega \in C$.

Now fix $\varepsilon < \varepsilon_0$ and $\omega \in C$. Write

$$e = K(E_+(\varepsilon), \theta(\omega))$$

then $e < b_0$ since $\theta(\omega+\varepsilon) \in \text{cl } E_+(\varepsilon)$.

We shall prove that

$$\lim_{n\to\infty} n^{-1} \log P_\omega(T_n > \omega+\varepsilon) = -e.$$

By Definition 3.1 and the conditions on b_0 we have, writing ω_ε for $\omega+\varepsilon$ and Γ for $\Gamma(b_0,\theta(\omega))$,

$$|P_\omega(T_n > \omega_\varepsilon) - P_\omega(\rho_n(\omega_\varepsilon)'(\overline{X}_n - m[b_n(\omega_\varepsilon)]) > 0, \overline{X}_n \in \lambda(\Gamma))|$$

$$< P_\omega(\overline{X}_n \notin \lambda(\Gamma)) = \exp\{-nb_0 + o(n)\}$$

as $n \to \infty$, where the equality holds by Theorem 3.1 since $\Gamma(b_0,\theta(\omega)) \subset A$ implies $b_0 < K(\theta(\omega))$. Thus, since $e < b_0$ it suffices to prove

$$\lim_{n\to\infty} n^{-1} \log P_\omega(\overline{X}_n \in \lambda(E_n)) = -e \tag{3.34}$$

where

$$E_n = \{\eta \in \Theta^* : \rho_n(\omega_\varepsilon)'(\lambda(\eta) - m[b_n(\omega_\varepsilon)]) > 0\} \cap \Gamma.$$

Since Γ is bounded and $\rho_n(\omega_\varepsilon) \to \rho(\omega_\varepsilon)$ by (3.19) and $b_n(\omega_\varepsilon) \to \omega_\varepsilon$ since $\{T_n\}$ is consistent, cf. Lemma 3.6, we have

$$d_H(E_n, E) \to 0 \quad \text{as } n \to \infty$$

with

$$E = \{\eta \in \Theta^* : \rho(\omega_\varepsilon)'(\lambda(\eta) - m(\omega_\varepsilon)) > 0\} \cap \Gamma.$$

Note that indeed $K(E,\theta(\omega)) = K(E_+(\varepsilon),\theta(\omega))$ since E and $E_+(\varepsilon)$ differ only outside Γ and $K(E_+(\varepsilon),\theta(\omega) = e < b_0$. Lemma 3.2 now implies (3.34) and together with the analogous result for $E_-(\varepsilon)$, (i) follows.

Now we prove (ii): Let η_ε be a point in cl E where (3.21) holds and let $\eta \in E$, then, writing $x_\alpha = (1-\alpha)\lambda(\eta_\varepsilon) + \alpha\lambda(\eta)$,

$$\rho(\omega_\varepsilon)'(x_\alpha - \lambda(\eta_\varepsilon)) > 0 \tag{3.35}$$

for each positive α. By (I.2.12) we have for $x_\alpha \in \Lambda$

$$K(\lambda^{-1}(x_\alpha),\theta(\omega)) = K(\eta_\varepsilon,\theta(\omega)) + K(\lambda^{-1}(x_\alpha),\eta_\varepsilon)$$
$$+ (\eta_\varepsilon - \theta(\omega))'(x_\alpha - \lambda(\eta_\varepsilon)).$$

Since $\lambda^{-1}(x_\alpha) \in E$ for each sufficiently small positive α, we obtain

$$(\eta_\varepsilon - \theta(\omega))'(x_\alpha - \lambda(\eta_\varepsilon)) \geq -K(\lambda^{-1}(x_\alpha),\eta_\varepsilon).$$

Now let $\alpha \downarrow 0$ then Lemma I.2.4 implies

$$(\eta_\varepsilon - \theta(\omega))'(\lambda(\eta) - \lambda(\eta_\varepsilon)) \geq 0$$

which holds for every $\eta \in E$ satisfying (3.35) (for $\alpha = 1$) whence $\eta_\varepsilon - \theta(\omega)$ is a multiple of $\rho(\omega_\varepsilon)$, proving (ii). □

PROOF of Theorem 3.8. Let C_0 be a compact interval in int Ω. Choose more compact intervals C_1, C_2, C_3, C_4 such that $C_i \subset \text{int } C_{i+1}$, $i = 0,\ldots,3$ and $C_4 \subset \text{int } \Omega$. We prove that the conditions of Lemma 3.9 hold with $\rho(\omega) = \dot{\theta}(\omega) / \|\dot{\theta}(\omega)\|$ for $\omega \in C_1$.

Since $\{T_n\}$ is consistent, $b_n(\omega) \to \omega$ for each $\omega \in C_4$, as $n \to \infty$, by Lemma 3.6. Moreover, since $\rho_n(\omega) \neq 0$ for all n,ω, we obtain using the uniform boundedness of the moments of X on the compact set $\theta(C_4) \subset \text{int } \Theta^*$ and the Berry-Esseen theorem that uniformly on C_4 as $n \to \infty$,

$$P_\omega(\rho_n(b_n^{-1}(\omega))'(\bar{X}_n - m(\omega)) \leq 0) \to \tfrac{1}{2}. \tag{3.36}$$

By the convergence of $b_n(\omega) \to \omega$ on C_4 we have $b_n^{-1}(\omega) \in C_4$ for all $\omega \in C_3$ when n is large enough. Thus, (3.16) and (3.36) imply uniformly on C_3 as $n \to \infty$,

$$P_\omega(T_n \leq b_n^{-1}(\omega)) \to \tfrac{1}{2}.$$

The asymptotic normality of $n^{\frac{1}{2}}(T_n - \omega)$ (with zero mean) yields $b_n^{-1}(\omega) = \omega + o(n^{-\frac{1}{2}})$ uniformly on C_3 whence

$$b_n(\omega) = \omega + o(n^{-\frac{1}{2}}) \tag{3.37}$$

as $n \to \infty$ holds uniformly on C_2.

Now let $\omega \in C_1$ and $u \in \mathbb{R}$ and consider

$$(3.38)\qquad P_\omega(n^{\frac{1}{2}} i_\omega^{\frac{1}{2}}(T_n - \omega) \le u) = P_\omega(\rho_n(\omega_n)'(\overline{X}_n - m[b_n(\omega_n)]) \le 0) + o(1),$$

where $\omega_n = \omega + n^{-\frac{1}{2}} i_\omega^{-\frac{1}{2}} u$ and the $o(1)$-term is bounded by the probability that $\|\overline{X}_n - \lambda(\omega_n)\| > d$, d being the radius of $\{T_n\}$ on C_4. We have

$$(3.39)\qquad \rho_n(\omega_n)'(\overline{X}_n - m[b_n(\omega_n)]) = \rho_n(\omega)'(\overline{X}_n - m(\omega))$$
$$- \rho_n(\omega)'(m[b_n(\omega_n)] - m(\omega)) + (\rho_n(\omega_n) - \rho_n(\omega))'(\overline{X}_n - m[b_n(\omega_n)]).$$

Since $\omega_n \in C_2$ when n is large enough it follows from (3.37) and the asymptotic normality of $n^{\frac{1}{2}}(\overline{X}_n - m(\omega))$ that the last two terms in (3.39) together equal

$$(3.40)\qquad -(b_n(\omega_n) - \omega)\rho_n(\omega)'\dot{m}(\omega) + o(|b_n(\omega_n) - \omega|)$$
$$+ \|\rho_n(\omega_n) - \rho_n(\omega)\| \cdot O_{P_\omega}(n^{-\frac{1}{2}})$$
$$= -n^{-\frac{1}{2}} i_\omega^{-\frac{1}{2}} \rho_n(\omega)'\dot{m}(\omega)u + o(n^{-\frac{1}{2}})$$
$$+ O(\|m[b_n(\omega_n)] - m[b_n(\omega)]\|) \cdot O_{P_\omega}(n^{-\frac{1}{2}}),$$

where Lemma 3.13 was also used. As a result of (3.39) and (3.40), (3.38) equals

$$(3.41)\qquad P_\omega(n^{\frac{1}{2}}\rho_n(\omega)'(\overline{X}_n - m(\omega)) \le i_\omega^{-\frac{1}{2}}\rho_n(\omega)'\dot{m}(\omega)u) + o(1).$$

Since the covariance Σ_ω is nondegenerate on C_1 and $\rho_n \ne 0$ we have as $n \to \infty$

$$(3.42)\qquad \frac{n^{\frac{1}{2}}\rho_n(\omega)'(\overline{X}_n - m(\omega))}{\{\rho_n(\omega)'\Sigma_\omega\,\rho_n(\omega)\}^{\frac{1}{2}}} \xrightarrow{\mathcal{D}_{P_\omega}} N(0,1).$$

Denoting as $\langle\cdot,\cdot\rangle_\omega$ and $\|\cdot\|_\omega$ the inner product and norm induced by Σ_ω, condition (3.17), (3.38), (3.41) and (3.42) imply, using $\dot{m}(\omega) = (d/d(\omega))\lambda(\theta(\omega)) = \Sigma_\omega\dot{\theta}(\omega)$ and $i_\omega = \dot{\theta}(\omega)'\Sigma_\omega\dot{\theta}(\omega) = \|\dot{\theta}(\omega)\|_\omega^2$ that, as $n \to \infty$,

$$\frac{\langle\rho_n(\omega),\dot{\theta}(\omega)\rangle_\omega}{\|\rho_n(\omega)\|_\omega\|\dot{\theta}(\omega)\|_\omega} \longrightarrow 1.$$

Since $\|\rho_n(\omega)\| = 1$, it follows that, as $n \to \infty$,

$$(3.43)\qquad \rho_n(\omega) \to \dot{\theta}(\omega)/\|\dot{\theta}(\omega)\|.$$

By (3.37) and (3.43) the condition (3.19) of Lemma 3.9 is satisfied. □

PROOF of Theorem 3.10. Since $\{T_n\}$ is consistent, $b_n(\omega) \to \omega$ on int Ω, which by monotonicity implies uniform convergence on compact subsets of int Ω. Moreover, since $\|\rho_n(\omega_0)\| = 1$, there is a subsequence $\{n_k\}$ such that $\rho_{n_k}(\omega_0)$ converges to ρ_0, say, with $\rho_0 \neq \dot\theta(\omega_0) / \|\dot\theta(\omega_0)\|$. By the uniform convergence of $\{b_n\}$ and Lemma 3.13 there exist $c_1 > 0$ and $c_2 > 0$ such that $|\omega - \omega_0| < c_2$ implies

$$(3.44) \qquad \|\rho_{n_k}(\omega) - \dot\theta(\omega_0) / \|\dot\theta(\omega_0)\| \| > c_1 .$$

Now abbreviate $\theta_0 = \theta(\omega_0)$, $\omega_\varepsilon = \omega_0 + \varepsilon$, and write $b_{max}(\varepsilon,\theta_0) = \max\{K(\theta(\omega_\varepsilon),\theta_0), K(\theta(\omega_{-\varepsilon}),\theta_0)\}$. Choose a compact interval $B \subset$ int Ω and choose $\varepsilon_0 < c_2$ such that $[\omega_0 - \varepsilon_0, \omega_0 + \varepsilon_0] \subset$ int B and such that both

$$b_{max}(\varepsilon_0) < K(\theta_0)$$

and

$$\lambda(\Gamma) \subset \{x : \|x - m(\omega)\| < \tfrac{1}{2} d(B)\},$$

where Γ abbreviates $\Gamma(b_{max}(\varepsilon_0,\theta_0),\theta_0)$. For each $\varepsilon < \varepsilon_0$ we have, as in the proof of Lemma 3.9,

$$\limsup_{k\to\infty} n_k^{-1} \log P_{\omega_0}(T_{n_k} > \omega_0 + \varepsilon)$$
$$= \limsup_{k\to\infty} n_k^{-1} \log P_{\omega_0}(\overline{X}_{n_k} \in \lambda(E_{n_k}(\varepsilon)),$$

where

$$E_{n_k}(\varepsilon) = \{\eta : \rho_{n_k}(\omega_\varepsilon)'(\lambda(\eta) - m[b_{n_k}(\omega_\varepsilon)]) > 0\} \cap \Gamma .$$

The subsequence $\{\rho_{n_k}(\omega_\varepsilon)\}$ has a further subsequence $\{\rho_{n_k'}(\omega_\varepsilon)\}$ which converges to ρ_ε, say, where ρ_ε satisfies by (3.44)

$$(3.45) \qquad \|\rho_\varepsilon - \dot\theta(\omega_0) / \|\dot\theta(\omega_0)\| \| > c_1 ,$$

implying that $E_{n_k'}(\varepsilon) \to E(\varepsilon)$ in Hausdorff distance, with

$$(3.46) \qquad E(\varepsilon) = \{\eta : \rho_\varepsilon'(\lambda(\eta) - m(\omega_\varepsilon)) > 0\} \cap \Gamma .$$

By Lemma 3.2 we obtain

$$(3.47) \qquad \limsup_{n\to\infty} n^{-1} \log P_{\omega_0}(T_n > \omega_0 + \varepsilon)$$

$$\geq \lim_{k\to\infty} n_k'^{-1} \log P_{\omega_0}(\overline{X}_{n_k'} \in \lambda(E(\varepsilon)) = -K(E(\varepsilon),\theta_0).$$

Note, that (3.47) holds for every $\varepsilon < \varepsilon_0$ for an $E(\varepsilon)$ of the form (3.46) with ρ_ε satisfying (3.45). We proceed to evaluate $K(E(\varepsilon),\theta_0)$. By Lemma 3.9 (ii) we have

$$(3.48) \qquad K(E(\varepsilon),\theta_0) = K(\eta_\varepsilon,\theta_0)$$

with

$$(3.49) \qquad \eta_\varepsilon = \theta_0 + t_\varepsilon \rho_\varepsilon.$$

Since $\rho_\varepsilon'(\lambda(\eta_\varepsilon) - m(\omega_\varepsilon)) = 0$, (I.2.12) implies

$$(3.50) \qquad K(\theta(\omega_\varepsilon),\theta_0) = K(\eta_\varepsilon,\theta_0) + K(\theta(\omega_\varepsilon),\eta_\varepsilon).$$

By Taylor expansion we have, as $\varepsilon \downarrow 0$,

$$\theta(\omega_\varepsilon) = \theta(\omega_0 + \varepsilon) = \theta(\omega_0) + \varepsilon\dot{\theta}(\omega_0) + o(\varepsilon),$$

which combined with (3.45) and (3.49) yields by plane geometry, for some $c_3 > 0$ and sufficiently small ε, $\|\theta(\omega_\varepsilon) - \eta_\varepsilon\| > c_3\varepsilon$. Lemma I.2.4 now implies for a positive c_4 and small ε's

$$(3.51) \qquad K(\theta(\omega_\varepsilon),\eta_\varepsilon) > c_4\varepsilon^2.$$

Since, also by Taylor expansion as $\varepsilon \downarrow 0$,

$$K(\theta(\omega_\varepsilon),\theta_0) = \tfrac{1}{2} i_{\omega_0}\varepsilon^2 + o(\varepsilon^2),$$

(3.50) and (3.51) imply

$$(3.52) \qquad K(\eta_\varepsilon,\theta_0) \leq (\tfrac{1}{2} i_{\omega_0} - c_4)\varepsilon^2 + o(\varepsilon^2) \quad \text{as } \varepsilon \downarrow 0.$$

Noting that the MLE does in first order attain the local inaccuracy bound $\frac{1}{2}\varepsilon^2 i_\omega$, cf. Bahadur (1960b,1967), the proof is complete by (3.47), (3.48) and (3.52). □

4. SHIFT FAMILIES

4a. Introduction

In this section, let $\mathcal{P} = \{P_\theta : \theta \in \mathbb{R}\}$ be a shift family of probability measures on $\mathbb{R}$, with densities p_θ with respect to Lebesgue measure given by

$$(4.1) \qquad p_\theta(x) = p(x-\theta), \quad x \in \mathbb{R}, \quad \theta \in \mathbb{R}.$$

The estimation problem as outlined in Section 1 will be narrowed to estimation of θ: we shall take $g(\theta) = \theta$. As a consequence, Bahadur's bound $b(\varepsilon,\theta)$ is now independent of θ and will be written as $b(\varepsilon)$.

EXAMPLE 4.1. Let $p(x)$ be the standard normal density, then $\{\overline{X}_n\}$ is inaccuracy rate optimal, cf. Example 1.1.

EXAMPLE 4.2. Let $p(x) = e^{-x} \cdot 1_{[0,\infty)}(x)$, then $\{P_\theta\}$ defined by (4.1) is the exponential shift family. We have $b(\varepsilon) = \varepsilon$, and this bound is attained by $\{X_{1:n}\}$, the smallest order statistic in the sample $X_1,\dots,X_n$. This can be shown by direct calculation of the inaccuracy rate or deduced by Proposition 2.1 from the optimality of the largest order statistic in the uniform family of Example 1.2, by taking $g(\theta) = -\log\theta$.

There is another exponentially convex shift family, as is demonstrated in the next example.

EXAMPLE 4.3. Let, for a fixed $\alpha > 0$,

$$p_\theta(x) = (\Gamma(\alpha))^{-1} \exp\{\alpha(x-\theta) - e^{x-\theta}\}, \qquad x \in \mathbb{R}.$$

This loggamma shift family is a one-parameter exponential family. In a more canonical form with canonical parameter $\nu = -e^{-\theta}$, the density p^* of the sufficient statistic $Y = e^X$ is

$$p^*(y) = \exp\{\nu y + \log(-\nu)\}(\Gamma(\alpha))^{-1} y^{\alpha-1}.$$

(The more usual parametrization of this gamma density employs $\beta = -1/\nu$ as the scale parameter.) Let

$$\hat{\nu}_n = -\alpha\left(\frac{1}{n}\sum_{i=1}^{n} e^{X_i}\right)^{-1},$$

the MLE of ν and take $g(\cdot) = -\log(-\cdot)$ in Theorem 3.5 or Remark 3.2, then $g(\hat{\nu}_n)$ is seen to be optimal with respect to the inaccuracy rate as an estimate of $\theta = g(\nu)$ (Note that $K(\nu) = \infty$ for each ν by (I.2.12) and (I.2.14)). Thus

$$T_n = \log\left(\frac{1}{n\alpha}\sum_{i=1}^{n} e^{X_i}\right)$$

attains Bahadur's bound for each ε and θ. The bound is min $\{K(P_{-\varepsilon},P_0), K(P_\varepsilon,P_0)\}$ and equals $\alpha(e^{-\varepsilon} - 1 + \varepsilon)$.

These examples seem to be the only shift families (essentially) which are exponentially convex, cf. Barndorff-Nielsen (1978), Section 1.3.

In other shift families we may therefore expect, that inaccuracy rate optimal estimates usually do not exist, cf. Section 2. Sievers (1978) came to the same conclusion, be it apparently on a more empirical basis. We give an example where Bahadur's bound is attained with a consistent estimate, in a shift family which is not exponentially convex, for a fixed θ and all $\varepsilon > 0$:

EXAMPLE 4.4. Let P_θ be the uniform $(\theta-1,\theta+1)$ distribution, then $b(\varepsilon) = \infty$ since $K(\eta,\theta) = \infty$ for all $\eta \neq \theta$. Define the estimate $\{T_n\}$ by

$$T_n = \begin{cases} 0 & \text{when } X_{1:n} \geq -1 \text{ and } X_{n:n} \leq 1 \\ X_{1:n}+1 & \text{when } X_{1:n} < -1 \\ X_{n:n}-1 & \text{when } X_{n:n} > 1, \end{cases}$$

then $\{T_n\}$ is consistent and $P_0(|T_n| > \varepsilon) = 0$ for each $\varepsilon > 0$. Note however, that the inaccuracy rate is finite for all other θ's.

A consistent estimate with $e(\varepsilon,0,\{T_n\}) = b(\varepsilon)$ for each $\varepsilon > 0$ in the double exponential shift family is given in Section 4d, Example 4.11.

It is seen in Examples 4.4 and 4.11 that Bahadur's bound can be attained with a consistent estimate. The estimates attain the bound for one θ only, however, and they are therefore not translation equivariant. (Translation) equivariance is a restriction which is usually imposed upon location estimates in shift families. Sievers (1978) found an upper bound on the inaccuracy rate for equivariant, not necessarily consistent estimates. We shall derive it here in somewhat greater generality. Note that for equivariant estimates the inaccuracy rate is independent of θ; it will be denoted as $e(\varepsilon,\{T_n\})$.

LEMMA 4.1. (Sievers). *If* p *is a density on* $\mathbb{R}$ *and* $\{T_n\}$ *is equivariant, then*

$$e(\varepsilon,\{T_n\}) \leq s(\varepsilon)$$

where

$$s(\varepsilon) = M(P_{-\varepsilon}, P_{\varepsilon})$$

which is defined in (I.2.8). *The bound* $s(\varepsilon)$ *will be called Sievers' bound.*

PROOF. In view of (I.2.8) we have to show

$$e(\varepsilon, \{T_n\}) \le \max\,(K(Q,P_{-\varepsilon}), K(Q,P_{\varepsilon})) \tag{4.2}$$

for each equivariant $\{T_n\}$ and all $Q \in P^*$, the class of pms on $\mathbb{R}$. Assume $M(P_{-\varepsilon},P_{\varepsilon}) < \infty$ and let Q satisfy $K(Q,P_{-\varepsilon}) < \infty$ and $K(Q,P_{\varepsilon}) < \infty$ then $Q \ll P_{\varepsilon}$, hence Q has a density. The equivariance of $\{T_n\}$ now implies $Q(T_n = 0) = 0$ since the Lebesque measure of the same event is zero. It follows that

$$\max\,\{\limsup_{n\to\infty} Q(T_n > 0),\ \limsup_{n\to\infty} Q(T_n < 0)\} > 0,$$

implying by corollary 1.2 that either

$$\limsup_{n\to\infty} n^{-1} \log P_{-\varepsilon}(T_n > 0) \ge -K(Q,P_{-\varepsilon})$$

or

$$\limsup_{n\to\infty} n^{-1} \log P_{\varepsilon}(T_n < 0) \ge -K(Q,P_{\varepsilon}).$$

By translation equivariance and

$$e(\varepsilon,\{T_n\}) =$$
$$\min\{-\limsup_{n\to\infty} n^{-1} \log P_0(T_n > \varepsilon),\ -\limsup_{n\to\infty} n^{-1} \log P_0(T_n < -\varepsilon)\},$$

(4.2) is established. □

REMARK 4.1. In view of Lemma I.2.2 we have

$$M(P_{-\varepsilon},P_{\varepsilon}) = -\log \inf_{0<\alpha<1} \int p^{\alpha}(x-\varepsilon)p^{1-\alpha}(x+\varepsilon)\,dx, \tag{4.3}$$

which is the expression for the bound in Sievers (1978).

Sievers (1978) only proved the bound (4.3) to hold for densities with nondecreasing $p(x-\varepsilon)/p(x+\varepsilon)$. In that case, a best equivariant estimate (minimizing $P_0(|T_n| > \varepsilon)$ for each n over the class of translation equivariant estimates) exists, cf. Ferguson (1967), Section 4.7. This estimate was also derived in Huber (1968). Sievers derives his bound as the inaccuracy rate of this best equivariant estimate. We shall find an estimate which attains Sievers' bound in a different way in Section 4b and prove it to be essentially unique in a large class of equivariant estimates, in Section 4c.

EXAMPLE 4.1. (continued). Sievers' bound also equals $\frac{1}{2}\varepsilon^2$ (and is attained by the sample mean).

EXAMPLE 4.2. (continued). Here, $s(\varepsilon) = K(P_\varepsilon, P_{-\varepsilon}) = 2\varepsilon$. This bound is attained by the estimate $\{X_{1:n} - \varepsilon\}$.

EXAMPLE 4.3. (continued). By Lemma I.2.2, $s(\varepsilon)$ is found by minimizing $\max\{K(R,P_{-\varepsilon}),K(R,P_\varepsilon)\}$ over $R \in \mathcal{P}^{P_{-\varepsilon},P_\varepsilon}$. Since $\mathcal{P} = \{P_\theta : \theta \in \mathbb{R}\}$ is an exponential family, $\mathcal{P}^{P_{-\varepsilon},P_\varepsilon}$ equals $\{P_\theta : -\varepsilon \le \theta \le \varepsilon\}$. The minimization yields $s(\varepsilon) = K(\tilde{\theta},\varepsilon)$ $(= K(\tilde{\theta},-\varepsilon))$ with $\tilde{\theta} = \log(2\varepsilon/(e^\varepsilon - e^{-\varepsilon}))$, hence

$$s(\varepsilon) = \alpha\left(\frac{2\varepsilon}{e^\varepsilon - e^{-\varepsilon}}\, e^\varepsilon - 1 - \varepsilon - \log \frac{2\varepsilon}{e^\varepsilon - e^\varepsilon}\right).$$

It is not immediately clear which estimate attains this bound. We shall return to this question in the next subsection.

In Examples 4.2 and 4.3 we see that, contrary to a remark of Sievers (1978), the bound $s(\varepsilon)$ can be *larger* than $b(\varepsilon)$. The reason is that Sievers' bound holds for equivariant estimates, which are not necessarily consistent. Indeed the estimate in Example 4.2 above is not consistent.

EXAMPLE 4.4. (continued). Sievers' bound equals $-\log(1-\varepsilon)$ when $\varepsilon < 1$ and ∞ otherwise. It is attained by $\{(X_{1:n} + X_{n:n})/2\}$. This example represents a rare occasion where an estimate attains Sievers' bound for each $\varepsilon > 0$.

4b. M-estimates

In this subsection the inaccuracy rate is derived for a class of M-estimates, essentially by means of Chernoff's theorem. As a side effect, an estimate which attains Sievers' bound will emerge in a natural way.

M-estimates are defined here as a suitable zero (or change of sign) of

$$\lambda_n(t) = \sum_{i=1}^{n} \psi(X_i - t),$$

where ψ is a function into the extended real line which attains positive as well as negative values, but not both $-\infty$ and $+\infty$. We consider two classes of functions ψ, requiring either

(4.4) ψ is nondecreasing

or

(4.5) ψ is bounded, continuous and such that λ_n has at least one zero for each n $[P_0]$.

The condition on λ_n holds when $x \cdot \psi(x)$ is nonnegative for $|x|$ large enough.

We now define M-estimates more precisely: When ψ satisfies (4.4), the M-estimate $\{T_n\} = \{T_n^{(\psi)}\}$ is defined by

$$T_n = \sup \{t : \lambda_n(t) \geq 0\}. \tag{4.6}$$

When ψ satisfies (4.5) and not (4.4), $\lambda_n(t)$ may have zeros which are meaningless as an estimate of θ; we define T_n as the zero closest to an auxiliary estimate. We take the sample median $M_n = X_{[(n+1)/2]:n}$ for this purpose: When ψ satisfies (4.5), $\{T_n\}$ is defined by

$$T_n = \begin{cases} t^+ & \text{when } t^+ - M_n \leq M_n - t^- \\ t^- & \text{when } t^+ - M_n > M_n - t^- \end{cases} \tag{4.7}$$

where

$$t^+ = \inf \{t : t \geq M_n, \lambda_n(t) = 0\},$$
$$t^- = \sup \{t : t \leq M_n, \lambda_n(t) = 0\}.$$

Note, that definitions (4.6) and (4.7) render $\{T_n\}$ translation equivariant. The inaccuracy rates of these estimates involve the log-moment generating functions of $\psi(X)$ under P_ε and $P_{-\varepsilon}$; we define

$$\gamma_\theta(\tau) = \log \int e^{\tau\psi(x)} dP_\theta(x).$$

Furthermore, we define the quantity $e_\psi(\varepsilon)$ by

$$e_\psi(\varepsilon) = \min \{-\inf_{\tau \geq 0} \gamma_{-\varepsilon}(\tau), -\inf_{\tau \leq 0} \gamma_\varepsilon(\tau)\}. \tag{4.8}$$

<u>THEOREM 4.2</u>. *Let* ψ *satisfy* (4.4) *and let* $\{T_n\}$ *be defined by* (4.6). *If* $P_\varepsilon(\psi(X_1) < 0) > 0$ *or* $P_\varepsilon(\psi(X_1) = 0) = 0$ *then*

(4.9) $\qquad e(\varepsilon,\{T_n\}) = e_\psi(\varepsilon).$

REMARK 4.2. The asymmetry of the condition above is due to the asymmetry in definition (4.6). The M-estimate could have been defined as $\inf\{t : \lambda_n(t) \le 0\}$, in which case there would have been conditions on the sign of ψ under $P_{-\varepsilon}$, in Theorem 4.2.

PROOF of Theorem 4.2. We shall prove that

$$\lim_{n\to\infty} n^{-1} \log P_0(T_n > \varepsilon) = \inf_{\tau\ge 0} \gamma_{-\varepsilon}(\tau) \tag{4.10}$$

and

$$\lim_{n\to\infty} n^{-1} \log P_0(T_n < -\varepsilon) = \inf_{\tau\le 0} \gamma_\varepsilon(\tau). \tag{4.11}$$

Since $\{T_n\}$ is translation equivariant,

$$P_0(T_n < -\varepsilon) = P_\varepsilon(T_n < 0)$$

and, as mentioned before, $P_\varepsilon(T_n = 0) = 0$. It follows now from (4.6) that

$$P_\varepsilon(\lambda_n(0) < 0) \le P_\varepsilon(T_n < 0) \le P_\varepsilon(\lambda_n(0) \le 0). \tag{4.12}$$

In cases where $P_\varepsilon(\psi(X_1) = 0) = 0$ and $P_\varepsilon(\psi(X_1) < 0) = 0$, we have $P_\varepsilon(\lambda_n(0) = 0) = 0$, hence, by (4.12), $P_\varepsilon(T_n < 0) = P_\varepsilon(\lambda_n(0) < 0) = 0$. It is easily seen that the right hand side of (4.11) also equals $-\infty$ in this case.

Now assume $P_\varepsilon(\psi(X_1) < 0) > 0$ and $\psi > -\infty$, then by Chernoff's theorem as generalized in Bahadur (1971) p. 8-9, we have

$$\lim_{n\to\infty} n^{-1} \log P_\varepsilon(\lambda_n(0) \le nc_n) = \inf_{\tau\le 0} \gamma_\varepsilon(\tau)$$

for each sequence $\{c_n\}$ with $\lim_{n\to\infty} c_n = 0$, implying (4.11) by (4.12).

When $P_\varepsilon(\psi(X_1) = -\infty) > 0$ (and hence $P_\varepsilon(\psi(X_1) = \infty) = 0$) $\gamma_\varepsilon(\tau)$ is seen to be $+\infty$ for each negative τ, hence the infimum equals 0 (at $\tau = 0$). Moreover, $P_\varepsilon(\lambda_n(0) < 0) \to 1$ as $n \to \infty$, establishing (4.11) in this case, too.

To prove (4.10), observe that (4.6) implies

$$P_{-\varepsilon}(T_n > 0) = P_{-\varepsilon}(\lambda_n(0) \ge 0),$$

and it is therefore sufficient to apply Chernoff's theorem directly. □

An example which shows that the condition of the theorem may not be omitted is obtained as follows.

EXAMPLE 4.5. Let $p(x) = e^{-x} 1_{[0,\infty)}(x)$ and let

$$\psi(x) = \begin{cases} -\infty, & x < 0 \\ 0, & 0 \le x < 2 \\ 1, & x \ge 2. \end{cases}$$

We have $P_\varepsilon(T_n < 0) = P_\varepsilon(\lambda_n(0) < 0) = 0$, and $P_{-\varepsilon}(T_n > 0) = P_{-\varepsilon}(\lambda_n(0) \ge 0) = e^{-n\varepsilon}$, implying $e(\varepsilon,\{T_n\}) = \varepsilon$.

On the other hand, it is seen that

$$\inf_{\tau\le 0} \log \int e^{\tau\psi} dP_\varepsilon = \inf_{\tau \le 0} \log \{1 - e^{-(2-\varepsilon)} + e^{\tau} e^{-(2-\varepsilon)}\}$$
$$= \log (1 - e^{-(2-\varepsilon)})$$

and $\inf_{\tau\ge 0} \log \int e^{\tau\psi} dP_{-\varepsilon} = -\varepsilon$, whence

$$e_\psi(\varepsilon) = \min \{-\log (1 - e^{-(2-\varepsilon)}), \varepsilon\},$$

which is smaller than ε when $\varepsilon = \frac{1}{2}$.

NOTE. For monotone ψ, Rubin and Rukhin (1983) recently proved (4.9) to hold for any estimate $\{T_n\}$ satisfying, for a sequence $q_n \to 0$,

$$\left|\frac{1}{n} \Sigma\, \psi(X_i - T_n)\right| \le q_n.$$

When ψ is not monotone but satisfies (4.5) and $\{T_n\}$ is defined by (4.7), the situation is more complicated. Usually however, the result (4.9) still holds provided ε is small enough. We shall first discuss the ideas involved. Since ψ is non-monotone, relation (4.12) no longer holds, but we may define a sequence of measurable subsets of the sample space, where (4.12) holds. When the probability of these subsets is large enough, we are still able to prove (4.9). The actual formulation of these notions is a bit technical:

Define for $\delta > 0$,

$$(4.13)\qquad \begin{cases} C_n(\delta) = \{(x_1,\ldots,x_n) : \lambda_n(t) \text{ is decreasing on } (-\delta,\delta)\}, \\ D_n(\delta) = \{(x_1,\ldots,x_n) : X_{[(n+1)/2]:n} \in [-\delta,\delta]\}, \\ c(\delta) = -\limsup_{n\to\infty} n^{-1} \log P_0\{\mathbb{R}^n \setminus C_n(\delta)\}, \\ d(\delta) = -\limsup_{n\to\infty} n^{-1} \log P_0\{\mathbb{R}^n \setminus D_n(\delta)\}. \end{cases}$$

Observe that $c(\cdot)$ is nonincreasing and that $d(\cdot) = e(\cdot,\{M_n\})$, the inaccuracy rate of the sample median.

THEOREM 4.3. *Let* ψ *and* $\{T_n\}$ *satisfy* (4.5) *and* (4.7), *respectively. It holds that*

$$(4.14)\qquad e(\varepsilon,\{T_n\}) \geq \min\left(e_\psi(\varepsilon), c(\delta), d\left(\frac{\delta-\varepsilon}{2}\right)\right)$$

for each $\delta \geq \varepsilon$. *Furthermore, when* $e_\psi(\varepsilon) < c(\varepsilon)$ *we have*

$$(4.15)\qquad e(\varepsilon,\{T_n\}) \leq e_\psi(\varepsilon).$$

The proof essentially shows that (4.12) holds with a large enough probability, and is given in 4d.

REMARK 4.3. Most of the non-monotone functions ψ that have been proposed for M-estimates, cf. Andrews *et al* (1972), have a central part (where the density has most of its mass) with positive, often constant slope, while the parts with negative slope correspond with the tails of the density. Suppose that the slope is $\geq \frac{1}{2}$ on $[-a,a]$ and nowhere less than -1. Now $c(\delta)$ can be estimated: Since $\lambda_n(\cdot)$ is nonincreasing on $(-\delta,\delta)$ whenever at least two thirds of the sample points are in the interval $(-a+\delta,a-\delta)$, we have

$$P_0(C_n(\delta)) \geq \sum_{j>\frac{2}{3}n} \binom{n}{j} p^j (1-p)^{n-j},$$

where $p = P_0((-a+\delta,a-\delta))$. If $p > \frac{2}{3}$ it follows that

$$c(\delta) \geq \tfrac{2}{3} \log \frac{2}{3p} + \tfrac{1}{3} \log \frac{1}{3(1-p)}.$$

Under certain general conditions, the inaccuracy rate of the sample median is (generalize in Example 6.1 in Bahadur (1971))

$$d(\delta) = \min\{-\tfrac{1}{2}\log 4p_+(1-p_+), -\tfrac{1}{2}\log 4p_-(1-p_-)\},$$

where $p_+ = P_0((\delta,\infty))$, $p_- = P_0((-\infty,-\delta))$. Thus, the conditions of Theorem 4.3 may be verified.

Rubin and Rukhin (1983) remark that for the MLE of the Cauchy shift family, (4.9) does not hold. Now this MLE is obtained as the M-estimate with

$$\psi(x) = -\frac{p'(x)}{p(x)} = \frac{2x}{1+x^2}.$$

The slope of ψ is positive on $(-1,1)$ but since $P((-1,1)) = \frac{1}{2}$, Remark 4.3 is not helpful here. The next theorem implies however that (4.9) holds in this case, when ε is sufficiently small.

THEOREM 4.4. *Assume that* p *is positive in a neighbourhood of* 0 *and that* $P((-\infty,0)) = \frac{1}{2}$. *If* ψ *satisfies* (4.5) *and is, moreover, continuously differentiable with bounded derivative such that* $|\psi'(x) - \psi'(y)| < c_1|x-y|$ *for a* $c_1 < \infty$ *and all* $x,y \in \mathbb{R}$, *and such that*

$$\int \psi'(x)p(x)dx > 0$$

then, for each sufficiently small ε, (4.9) *holds for* $\{T_n\}$ *satisfying* (4.7).

REMARK 4.4. The assumption on p ensures that $d(\delta) > 0$ for each $\delta > 0$. When the sample median is replaced by another equivariant estimate in the definitions (4.7) and (4.13), Theorem 4.3 remains valid; for Theorem 4.4, $d(\delta) > 0$ for some positive δ is again required.

Theorem 4.4 will be proved in Section 4d. Its main part is the demonstration that $c(\delta) > 0$ for some $\delta > 0$.

The uncomely expression (4.8) for the inaccuracy rate of an M-estimate may be made more transparent by defining

$$(4.16) \qquad \mathcal{P}_\psi = \{Q \in \mathcal{P}^* : \int \psi dQ = 0\}.$$

It was proved by Hoeffding (1965), that

$$-\inf_{\tau \geq 0} \gamma_{-\varepsilon}(\tau) = K(\mathcal{P}_\psi, P_{-\varepsilon}).$$

Thus, the inaccuracy rate equals the Kullback-Leibler "distance" of the "plane" $\mathcal{P}_\psi$ to the pair $\{P_{-\varepsilon}, P_\varepsilon\}$.

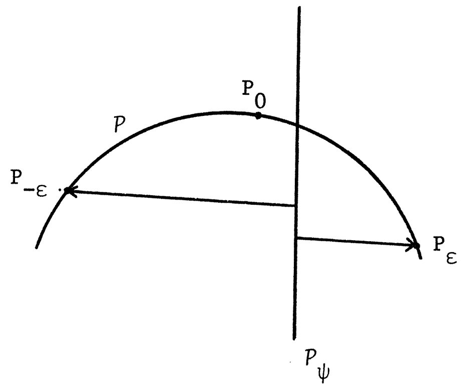

fig. 4.1. The shift family P and the "plane" P_ψ. The arrows indicate Kullback-Leibler "distance" determining the inaccuracy rate.

Now consider the expression (4.3) for Sievers' bound $s(\varepsilon)$. It can be rewritten as

$$(4.17) \qquad -\log \inf_{0<\alpha<1} \int e^{\alpha \log \frac{p(x-\varepsilon)}{p(x+\varepsilon)}} p(x+\varepsilon)dx.$$

By comparing this to the definition (4.8) of $e_\psi(\varepsilon)$, we find that $s(\varepsilon) = e_{\psi_\varepsilon}(\varepsilon)$, when ψ_ε is defined as

$$(4.18) \qquad \psi_\varepsilon(x) = \log \frac{p(x-\varepsilon)}{p(x+\varepsilon)}.$$

and either $\psi_\varepsilon < \infty$ a.e. or $\psi_\varepsilon > -\infty$ a.e. To see this, note that

$$\inf_{\tau\le 0} \log \int e^{\tau\psi_\varepsilon(x)} dP_\varepsilon(x) = \inf_{\tau\le 1} \log \int e^{\tau\psi_\varepsilon(x)} dP_{-\varepsilon}(x)$$

and use convexity of $\tau \mapsto \int e^{\tau\psi_\varepsilon(x)} dP_{-\varepsilon}(x)$ on $\mathbb{R}$. When ψ_ε is nondecreasing and either a.e. $> -\infty$ or a.e. $< \infty$, indeed $\{T_n^{(\psi_\varepsilon)}\}$ attains Sievers' bound (Sievers (1978), Thm 2.1).

Note that the condition $P_\varepsilon(\psi < 0) > 0$ or $P_\varepsilon(\psi = 0) = 0$ of Theorem 4.2 is not needed when $\psi = \psi_\varepsilon$. Example 4.6 below shows that densities p exist such that the above-mentioned condition fails for $\psi = \psi_\varepsilon$.

EXAMPLE 4.6. Let p be given by

$$p(x) = \tfrac{1}{2}(1_{[0,1)}(x) + e^{-x+1} 1_{[1,\infty)}(x)),$$

then we have, when $0 < \varepsilon < \tfrac{1}{2}$,

$$\psi_\varepsilon(x) = \begin{cases} -\infty, & x < \varepsilon \\ 0, & \varepsilon \le x < 1-\varepsilon \\ x-1+\varepsilon, & 1-\varepsilon < x \le 1+\varepsilon \\ 2\varepsilon, & x > 1+\varepsilon. \end{cases}$$

Note that (4.18) leaves ψ_ε undefined on $(-\infty,-\varepsilon)$. It seems reasonable to extend ψ_ε monotonically in that case.

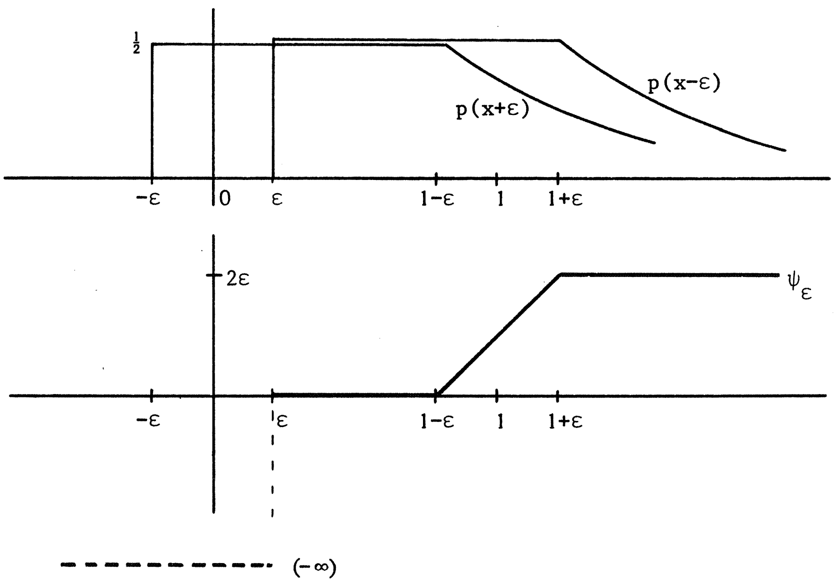

fig. 4.2. The densities p(x-ε) and p(x+ε) (above) and the function ψ_ε (below).

Since ψ_ε is nondecreasing, $\{T_n^{(\psi_\varepsilon)}\}$ attains Sievers' bound $s(\varepsilon) = e_{\psi_\varepsilon}(\varepsilon)$, though the condition of Theorem 4.2 is violated.

EXAMPLE 4.2. (continued). We have

$$\psi_\varepsilon(x) = \begin{cases} -\infty, & x < \varepsilon \\ 2\varepsilon, & x \ge \varepsilon. \end{cases}$$

The corresponding M-estimate is $\{X_{1:n} - \varepsilon\}$.

EXAMPLE 4.3. (continued). In this family,

$$\psi_\varepsilon(x) = -2\alpha\varepsilon + e^x(e^\varepsilon - e^{-\varepsilon}),$$

which is increasing. It follows that the corresponding M-estimate,

$$T_n^{(\psi_\varepsilon)} = \log\left(\frac{e^\varepsilon - e^{-\varepsilon}}{2\varepsilon\alpha n}\sum_{i=1}^n e^{X_i}\right)$$

attains Sievers' bound, which was found in the previous treatment of this example. Observe, that $T_n^{(\psi_\varepsilon)}$ and the MLE of θ differ only by the constant $\log((e^\varepsilon - e^{-\varepsilon})/2\varepsilon)$ (which is of order ε^2 as $\varepsilon \to 0$).

For densities and ε's such that ψ_ε is not monotone, but satisfies (4.5), Theorem 4.3 provides conditions under which $\{T_n^{(\psi_\varepsilon)}\}$ attains Sievers' bound. Moreover, when p is sufficiently smooth we shall prove that the conditions of Theorem 4.3 hold for $\psi = \psi_\varepsilon$ when ε is small enough, with the important implication that Sievers' bound is attainable in these situations.

THEOREM 4.5. *Let* $p > 0$ *on* $\mathbb{R}$ *with* $\int_{-\infty}^0 p(x)dx = \frac{1}{2}$. *If* p *is three times differentiable such that* $p'(x) > 0$ (< 0) *for each small (large) enough* x, *such that the first three derivatives of* $\log p$ *are bounded and such that* $\int (\log p)''pdx < 0$, *then* (4.9) *holds for* $\{T_n^{(\psi_\varepsilon)}\}$ *(defined by* (4.7)) *and* ψ_ε, *when* ε *is small enough.*

The proof is deferred to Section 4d.

EXAMPLE 4.7. The *Cauchy density*, $p(x) = (\pi(1+x^2))^{-1}$. We have

$$\psi_\varepsilon(x) = \log\frac{1+(x+\varepsilon)^2}{1+(x-\varepsilon)^2}.$$

The conditions of Theorem 4.5 hold, hence $\{T_n^{(\psi_\varepsilon)}\}$ attains Sievers' bound when ε is sufficiently small.

It has already been mentioned that the estimates which attain Sievers' bound are not necessarily consistent, see Example 4.2.

To investigate this problem it is useful to consider, if it exists,

$$\lambda(t) = \int \psi(x-t)p(x)dx.$$

Now suppose that ψ is nondecreasing and that λ exists. Define t_0 as "the value of $T_n^{(\psi)}$ at P_0", analogous to definition (4.6):

$$t_0 = \sup \{t : \lambda(t) \geq 0\}.$$

If $\lambda(t) > 0$ for each $t < t_0$, the M-estimate $T_n^{(\psi)}$ converges to t_0 $[P_0]$. (Serfling (1980), Lemma 7.2.1.A). Thus, $\{T_n^{(\psi)}\}$ is consistent iff $t_0 = 0$.

If ψ satisfies (4.5), $\lambda(t)$ exists for each t and $\lambda(t) = 0$ may have more than one root. Let t_0 be the root corresponding to definition (4.7) with $M_n = 0$ and $\lambda_n = \lambda$, and suppose λ is strictly decreasing in a neighbourhood $(t_0 - \delta, t_0 + \delta)$. If $|t_0| < \frac{1}{2}\delta$ and the sample median M_n converges to zero $[P_0]$ then $\{T_n\}$ defined by (4.7) converges to t_0 $[P_0]$. This can be proved by the method of proof of Theorem A.2 in Portnoy (1977). Again, consistency requires $t_0 = 0$.

For Sievers' estimate $\{T_n^{(\psi_\varepsilon)}\}$ we have, when both K's are finite,

$$\lambda(0) = \int \log \frac{p(x-\varepsilon)}{p(x+\varepsilon)}\, p(x)dx = K(P_0, P_{-\varepsilon}) - K(P_0, P_\varepsilon).$$

If, moreover, $p > 0$ on $\mathbb{R}$ and $\lambda_\varepsilon(t) = \int \psi_\varepsilon(x-t)p(x)dx$ exists in a neighbourhood of 0, a nondecreasing ψ_ε implies that λ_ε is continuous and decreasing in a neighbourhood of 0 (When $|\psi_\varepsilon|$ is bounded, this follows from the continuity of a convolution when one of the arguments is continuous, when $|\psi_\varepsilon|$ is not bounded, a truncation argument can be used). For p's such that ψ_ε satisfies (4.5), λ_ε is decreasing in a neighbourhood of 0 when the (central) part of ψ_ε which has a positive slope is large enough, see Remark 4.3.

For most interesting cases it therefore holds that consistency of $\{T_n^{(\psi_\varepsilon)}\}$ requires equality of the Kullback-Leibler information of P_0 with respect to $P_{-\varepsilon}$ and P_ε. For symmetric densities this obviously holds, but otherwise generally not.

The question arises whether we can find - in the general case - a bound on the inaccuracy rate for translation equivariant *consistent* estimates, and whether this bound is attained. Note that such a bound cannot be larger than the least of the bounds of Bahadur and Sievers. Equivariant consistent estimates that attain either of these bounds are therefore best with respect to the inaccuracy rate in the class of equivariant consistent estimates. Examples of such estimates are found in Examples 4.2 and 4.3, where the MLE's attain Bahadur's bound, and usually in cases where p is symmetric and $\{T_n^{(\psi_\varepsilon)}\}$ attains Sievers' bound.

When $K(P_0,P_{-\varepsilon})$ and $K(P_0,P_\varepsilon)$ are both finite, a class of (equivariant) M-estimates is obtained by taking

$$\psi^{(\alpha)}(x) = \alpha\{\log \frac{p(x)}{p(x+\varepsilon)} - K(P_0,P_{-\varepsilon})\} + (1-\alpha)\{\log \frac{p(x-\varepsilon)}{p(x)} + K(P_0,P_\varepsilon)\}.$$

Note, that indeed $\lambda(0) = 0$ for each α, hence, when $p(x)/p(x+\varepsilon)$ is non-decreasing, the corresponding M-estimate $\{T_n^{(\alpha)}\}$ is usually consistent for each $\alpha \in [0,1]$. Now α may be chosen to maximize the inaccuracy rate over the class $\{\{T_n^{(\alpha)}\} : \alpha \in [0,1]\}$. It is unknown, however, whether this maximizes the inaccuracy rate over all consistent equivariant estimates, or at least over all consistent M-estimates. In case of symmetry, $\alpha = \frac{1}{2}$ yields Sievers' estimate.

4c. Unicity properties of Sievers' estimate

In this section we study unicity properties of Sievers' estimate $\{T_n^{(\psi_\varepsilon)}\}$, cf. (4.18). We start with some preliminaries. Many estimates (and other statistics) can be written as

$$T_n(X_1,\ldots,X_n) = T(\hat{P}_n)$$

where T is a functional on $\mathcal{P}^*$ and $\hat{P}_n$ is the empirical probability measure associated with $X_1,\ldots,X_n$, assigning mass $1/n$ to each of the points $X_1,\ldots,X_n$.

The sample mean, for example, is obtained by taking the functional

$$T(Q) = \int x dQ(x). \tag{4.19}$$

Since functionals like (4.19) may not be properly defined on $\mathcal{P}^*$ (the right hand side may not exist), we also define for each $m > 0$

$$\mathcal{P}_m = \{Q \in \mathcal{P}^* : Q((-m,m)) = 1\}.$$

With this definition the functional (4.19) is defined on $\mathcal{P}_m$ for each m.

M-estimates are found as a suitable (according to definitions (4.6) or (4.7)) sign-change of

$$\lambda_Q(t) = \int \psi(x-t)dQ(x). \tag{4.20}$$

Furthermore, the class of L-estimates (linear combinations of order statistics) is defined here as generated by the functionals

(4.21) $$L(Q) = \int_0^1 J(t)Q^{-1}(t)dt,$$

where J is a weight function on (0,1) and Q^{-1} is the inverse of $Q((-\infty,\cdot])$ defined by

$$Q^{-1}(t) = \inf \{x : Q((-\infty,x]) \geq t\}.$$

Note that $L_n = L(\hat{P}_n)$ is properly defined as soon as definition (4.21) is proper for each $Q \in \bigcup_m P_m$, and that L_n is translation equivariant iff J has total mass 1. A well known example of an L-estimate is the α-trimmed mean, obtained by taking

(4.22) $$J(t) = (1-2\alpha)^{-1}\, 1_{(\alpha,1-\alpha)}(t).$$

We shall also need the concept of continuity of a functional. A functional T is weakly continuous on P^* when $Q \mapsto T(Q)$ is continuous with respect to the topology of weak convergence, cf. Billingsley (1968) or Huber (1980). For our purposes, a convenient definition of weak convergence, denoted $Q_n \xrightarrow{w} Q$, is that for each bounded and continuous real function f, we have

$$\int f(x)dQ_n(x) \to \int f(x)dQ(x) \quad \text{as } n \to \infty.$$

Let $p > 0$ on $\mathbb{R}$ and let $\{R_\alpha : \alpha \in [0,1]\} = P^{P_{-\varepsilon},P_\varepsilon}$ be the exponential family between $P_{-\varepsilon}$ and P_ε, cf. Definition I.2.1. Furthermore, let $\tilde{R} = R_{\tilde{\alpha}}$ be the unique pm in $\{R_\alpha\}$ satisfying (cf. Lemma I.2.2),

$$\max \{K(\tilde{R},P_{-\varepsilon}),K(\tilde{R},P_\varepsilon)\} = M(P_{-\varepsilon},P_\varepsilon).$$

Note that $p > 0$ on $\mathbb{R}$ implies $R_0 = P_{-\varepsilon}$, $R_1 = P_\varepsilon$ hence $\tilde{\alpha} \in (0,1)$. From (I.2.23) it then follows that

(4.23) $$\int \psi_\varepsilon(x)d\tilde{R}(x) = 0.$$

We shall now first assume that ψ_ε is nondecreasing and finite-valued. This implies that $p > 0$ on $\mathbb{R}$ and moreover, that the M-estimate $\{T_n^{(\psi_\varepsilon)}\}$ satisfies the "Cauchy mean value" property. In terms of the corresponding functional this means that for each $\gamma \in [0,1]$ and $P,Q \in P^*$, (4.24) $\quad T(P) \le T(Q) \Rightarrow T(P) \le T(\gamma P + (1-\gamma)Q) \le T(Q).$

Under some continuity assumptions we shall prove that translation equivariant estimates $T_n = T(\hat{P}_n)$ such that T satisfies (a weakened version of) (4.24) can only attain Sievers' bound on the inaccuracy rate when $\{T_n\}$ essentially equals $\{T_n^{(\psi_\varepsilon)}\}$.

The assumptions are

(4.25) T is weakly continuous on $\mathcal{P}_m$ for each m and translation equivariant,

(4.26) $T_n \xrightarrow{R_\alpha} T(R_\alpha)$ and $\alpha \mapsto T(R_\alpha)$ is continuous, for α in an open neighbourhood of $\tilde{\alpha}$

and furthermore, for each $Q \in \bigcup_m \mathcal{P}_m$ and $Q_\gamma = \gamma Q + (1-\gamma)\tilde{R}$, $\gamma \in [0,1]$ it holds that

(4.27) $\quad T_n \xrightarrow{Q_\gamma} T(Q_\gamma)$,

(4.28) $\quad \int \log \frac{p(x+\varepsilon)}{p(x+\varepsilon+\delta)}\, dQ_\gamma$ and $\int \log \frac{p(x-\varepsilon)}{p(x-\varepsilon-\delta)}\, dQ_\gamma$ tend to zero as $\delta \downarrow 0$

and lastly, a weakened form of (4.24):

(4.29) $\quad T(Q) \begin{smallmatrix} < \\ \\ > \end{smallmatrix} T(\tilde{R}) \Rightarrow T(Q_\gamma) \begin{smallmatrix} \le \\ \\ \ge \end{smallmatrix} T(\tilde{R}).$

<u>REMARK 4.5</u>. Assumptions (4.25) - (4.27) hold when T is a weakly continuous translation equivariant functional on $\mathcal{P}^*$. Note however, that the sample mean (4.19) is not weakly continuous on $\mathcal{P}^*$ while it is on $\mathcal{P}_m$ for each m. Assumption (4.28) holds when $d(\log p(x))/dx$ is a.e. bounded.

<u>THEOREM 4.6</u>. *Assume* $p > 0$ *on* $\mathbb{R}$ *and* ψ_ε *is nondecreasing. If*

(4.30) $\quad e(\varepsilon, \{T(\hat{P}_n)\}) = s(\varepsilon)$

and assumptions (4.25) - (4.29) *hold, then for each* k-*tuple* $(x_1,\ldots,x_k)$ *with corresponding empirical* $\hat{P}_k$, *we have*

(4.31) $\quad T(\hat{P}_k) \begin{smallmatrix} < \\ \\ > \end{smallmatrix} t \Rightarrow \Sigma\, \psi_\varepsilon(x_i - t) \begin{smallmatrix} \le \\ \\ \ge \end{smallmatrix} 0.$

<u>REMARK 4.6</u>. Relation (4.31) implies that

(4.32) $\inf\{t : \lambda_k(t) \le 0\} \le T(\hat{P}_k) \le \sup\{t : \lambda_k(t) \ge 0\}$

where $\lambda_k(t) = \Sigma_1^k \psi_\varepsilon(x_i - t)$. Many authors (Huber (1980), Sievers (1978), Serfling (1980)) define M-estimates this way, requiring $T(\hat{P}_k)$ to be any measurable function satisfying (4.32). We have chosen for the simpler definition (4.6), see also Remark 4.2.

In the proof, which is given in Section 4d, we show that when (4.31) fails for some k-tuple $x_1,\ldots,x_k$, a pm Q exists with $T(Q) > 0$ and $K(Q,P_{-\varepsilon}) < s(\varepsilon)$. Relation (4.30) is then contradicted by application of Corollary 1.2.

The applicability of Theorem 4.6 is, apart from the continuity assumptions, restricted by assumption (4.29), which, though apparently innocent, does not hold for most L-estimates (Leurgans (1981)), and may fail for M-estimates based on a non-monotone ψ, too.

EXAMPLE 4.8. Let the functional L be the α-trimmed mean, defined by (4.21) and (4.22). Let R be absolutely continuous and symmetric, then $L(R) = 0$. Define Q to have atoms of size a at $R^{-1}(\frac{1}{2}\alpha)$ and of size $1-a$ at a point $y > R^{-1}(1-\frac{1}{2}\alpha)$, with $\frac{1}{2} > 1-a > \alpha$ and y chosen large enough to ensure $L(Q) > 0$. Now let $\gamma = \frac{1}{2}\alpha$, then $(1-\gamma)R(R^{-1}(\frac{1}{2}\alpha)) + \gamma a < \alpha$, hence $L(Q_\gamma)$ is determined by the size of the atoms of Q, not by y. Now $a > \frac{1}{2}$ implies $L(Q_\gamma) < 0$, contradicting (4.24) (and (4.29) when R is chosen equal to $\tilde{R}$).

EXAMPLE 4.9. Let the functional T be defined as the M-functional $T^{(\psi)}$, with ψ given by fig. 4.3.

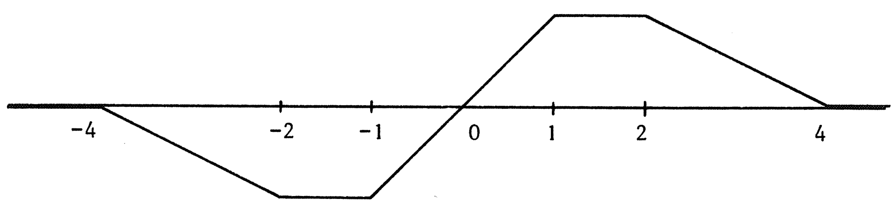

fig. 4.3. The function ψ, (a "Hampel").

The corresponding M-estimate was proposed by Hampel, cf. Andrews *et al* (1972). Let R have a positive symmetric density and choose Q to have atoms $\frac{1}{2}(1-\delta)$ at -3 and 1 and an atom of size δ at -1. Definition (4.7) yields $T(R) = 0$, $T(Q) = -1$, when $\delta > 0$. However, the median of $Q_\gamma = (1-\gamma)R + \gamma Q$ tends to 0 as $\gamma \downarrow 0$ and for $\delta < \frac{1}{5}$ it holds that

$$\int \psi(x)\,dQ_\gamma(x) > 0$$

for each $\gamma > 0$, implying $T(Q_\gamma) > 0$ when γ is small enough.

We shall now use the concept of differentials of statistical functionals to derive a necessary condition for a differentiable statistic to have inaccuracy rate $s(\varepsilon)$.

A statistic T_n of the form $T(\hat{P}_n)$ shall be called G-*differentiable at* R, when a measurable function $\chi = \chi_R$ exists such that for all $Q \in \cup P_m$

$$(4.33) \qquad \lim_{\gamma\downarrow 0} \gamma^{-1}\{T(R+\gamma(Q-R)) - T(R)\} = \int \chi(x)\,dQ(x)$$

is finite and

$$(4.34) \qquad \int \chi(x)\,dR(x) = 0.$$

G-differentiability is a modification of Gâteaux-differentiability, cf. Huber (1980), Section 2.5. Note that χ is Hampel's *Influence Curve* (Hampel (1968,1974)).

THEOREM 4.7. *Let* $p > 0$ *on* $\mathbb{R}$. *If* $T_n = T(\hat{P}_n)$ *is translation equivariant,* G-*differentiable at* $\tilde{R}$, *and* (4.27) *and* (4.28) *hold for each* $Q \in \cup P_m$, *then* (4.30) *implies that* χ *is a.e. proportional to* ψ_ε.

The proof is given in Section 4d.

Two lemmas are now presented which establish G-differentiability for classes of M- and L-estimates:

LEMMA 4.8. *Let* $T_n^{(\psi)} = T(\hat{P}_n)$ *be an* M-*estimate, and let* $T(R) = 0$. *If*

(4.35) $\lambda_R'(0)$ *exists and does not vanish*

and if for each $Q \in \cup P_m$

(4.36) $T(Q_\gamma) \to 0$ *as* $\gamma \downarrow 0$, (4.37) $\lambda_{Q_\gamma}(T(Q_\gamma)) = 0$ *for* $\gamma \in [0,\gamma_0]$

where $Q_\gamma = R + \gamma(Q - R)$, *and*

(4.38) $\lambda_Q(\cdot)$ *is continuous at* 0,

then T *is* G*-differentiable at* R *with*

(4.39) $\chi(x) = -\psi(x)/\lambda_R'(0)$.

REMARK 4.7. Serfling (1980), p 245, seems to require (4.35) only in his statement on Gâteaux-differentiability of M-estimates. To see the necessity of (4.38), take

$$\psi(x) = \begin{cases} 0, & x = 0 \\ x/|x|, & x \neq 0. \end{cases}$$

The corresponding M-estimate is the sample median. Let R be standard normal, then (4.35) holds. Let Q have atoms of size $\frac{1}{2}$ at 0 and 1, then $T(Q_\gamma) = 0$ for each $\gamma < 1$ and $\int \psi dQ = \frac{1}{2}$, contradicting (4.39) to be the influence curve.

Lemma 4.8 is proved in Section 4d.

From Lemma 4.8 and Theorem 4.7 it follows that if an M-estimate satisfies the conditions of Lemma 4.8 and Theorem 4.7, and attains Sievers' bound, this M-estimate is $T_n^{(\psi_\varepsilon)}$.

We turn our attention to L-estimates:

LEMMA 4.9. *Let* R *have a positive density on* $\mathbb{R}$ *and let* $E_R|X| < \infty$. *If the functional* L *is given by* (4.22) *with* J *bounded and continuous, then* L *is* G*-differentiable at* R *with*

$$\chi(x) = \int_0^x J(R(u))du - \int_{-\infty}^{\infty}\int_0^y J(R(u))du\, dR(y),$$

where R(u) *abbreviates* $R((-\infty,u])$. *Note that* χ *is properly defined since* J *is bounded and* $E_R|X| < \infty$.

The proof is a application of a theorem of Boos (1979) followed by integration by parts. It is given in full in the next subsection.

We would like to determine the inaccuracy rate of the unique (equivariant) L-estimate (with bounded continuous J) which may attain Sievers' bound.

By Theorem 4.7 and Lemma 4.9 this L-estimate is determined by

$$(4.40)\qquad \begin{aligned} &\int_0^x J(\tilde{R}(u))du = c(\psi_\varepsilon(x) - \psi_\varepsilon(0)),\\ &\int_0^1 J(t)dt = 1. \end{aligned}$$

The last relation ensures that L is translation equivariant. Note that proposition 4.7 only concerns *bounded* J's. By (4.40) this implies that $\psi_\varepsilon(x) = O(|x|)$ as $|x| \to \infty$.

The inaccuracy rate of an L-estimate is presently not known to have a simple form like (4.8) for M-estimates. When J equals zero on the "tails" [0,a] and [1-a,1] for some $a > 0$, Groeneboom *et al* (1979) give an expression for the essential large deviation probability, which involves minimizing over a complicated (non-convex!) subset of P^* (The tails condition was relaxed somewhat by Groeneboom and Shorack (1981)). For trimmed means, however, Groeneboom *et al* (1979) find a more explicit expression. We shall use this to show that a trimmed mean attains Sievers' bound in the double exponential family.

<u>EXAMPLE 4.10</u>. Let $p(x) = \frac{1}{2} \exp\{-|x|\}$, then by symmetry, for a given $\varepsilon > 0$,

$$d\tilde{R}(x)/dx = \begin{cases} (2(1+\varepsilon))^{-1} \exp\{-|x|+\varepsilon\}, & |x| > \varepsilon \\ (2(1+\varepsilon))^{-1}, & |x| \le \varepsilon. \end{cases}$$

By Lemma 4.1, Sievers' bound equals $K(\tilde{R},P_{-\varepsilon}) = \varepsilon - \log(1+\varepsilon)$. This is attained by the M-estimate $\{T_n^{(\psi_\varepsilon)}\}$ (Sievers (1978)), with ψ_ε given as

$$\psi_\varepsilon(x) = \begin{cases} -2\varepsilon, & x < -\varepsilon \\ 2x, & |x| \le \varepsilon \\ 2\varepsilon, & x > \varepsilon. \end{cases}$$

The "candidate" L-estimate is the $(2(1+\varepsilon))^{-1}$-trimmed mean L_n, say. By symmetry, its inaccuracy rate equals

$$-\lim_{n\to\infty} n^{-1} \log P_0(L_n > \varepsilon).$$

By Theorem 6.3 and formula (6.13) of Groeneboom *et al* (1979), this can be expressed as (writing $\alpha = (2(1+\varepsilon))^{-1}$)

$$(4.41)\qquad 2\alpha \log \alpha + (1-2\alpha)\log(1-2\alpha)$$
$$+ \inf \{\sup_{t\geq 0} f(a,b,t) : -\infty < a < b < \infty,\ b > \varepsilon\}$$

with

$$(4.42)\qquad f(a,b,t) = (1-2\alpha)[t\varepsilon - \log \int_a^b e^{tx} p(x)dx]$$
$$-\alpha[\log P_0(a) + \log(1-P_0(b))].$$

We shall show that $f(0,2\varepsilon,1)$ attains the sup and inf in (4.41). Consider

$$f(a,b,1) - f(0,2\varepsilon,1) =$$
$$(1-2\alpha)[-\log \int_a^b \tfrac{1}{2}e^{x-|x|}dx + \log \int_0^{2\varepsilon} \tfrac{1}{2}dx]$$
$$-\alpha[\log P_0(a) - \log \tfrac{1}{2} + \log(1-P_0(b)) - \log \tfrac{1}{2}\, e^{-2\varepsilon}].$$

Multiplying by $(1+\varepsilon)/\varepsilon = (1-2\alpha)^{-1}$ we obtain, when $a < 0$ and $b > \varepsilon$,

$$\log\varepsilon - \log\tfrac{1}{2} - \log\left[\frac{1-e^{2a}}{2} + b\right] - \frac{1}{2\varepsilon}\left[a - b + 2\varepsilon\right]$$
$$= -\log\left[\frac{1}{2\varepsilon}\left(\frac{1-e^{2a}}{2} + b\right)\right] - \frac{1}{2\varepsilon}\left[a-b\right] - 1$$
$$\geq \frac{1}{4\varepsilon}\left[e^{2a} - 2a - 1\right] \geq 0,$$

where the inequality $\log x \leq x-1$ was used. When $a \geq 0$, we find in a similar way that

$$f(a,b,1) \geq f(0,2\varepsilon,1).$$

It follows that

$$\inf_{a,b}\sup_{t\geq 0} f(a,b,t) \geq f(0,2\varepsilon,1)$$

and it suffices to remark that $f(0,2\varepsilon,1) \geq f(0,2\varepsilon,t)$ by symmetry and convexity of

$$t \mapsto \int_0^{2\varepsilon} e^{(t-1)(x-\varepsilon)}dx.$$

It remains to evaluate $f(0,2\varepsilon,1)$. Together with the other part of (4.41), this indeed equals $\varepsilon - \log(1+\varepsilon)$.

4d. Proofs and miscellany

This section contains, apart from the pending proofs, some examples of unexpected phenomena. We start with the proofs.

PROOF of Theorem 4.3. Let $\delta \geq \varepsilon > 0$. By (4.7) we have

$$\begin{aligned} P_0(|T_n| > \varepsilon) &\leq P_0(\{\lambda_n(-\varepsilon) < 0 \text{ or } \lambda_n(\varepsilon) > 0\} \cap C_n(\delta)) \\ &\quad + P_0(\{|M_n| > \tfrac{\delta-\varepsilon}{2}\} \cap C_n(\delta)) + P_0(\mathbb{R}^n \setminus C_n(\delta)) \\ &\leq P_0(\lambda_n(-\varepsilon) \leq 0 \text{ or } \lambda_n(\varepsilon) \geq 0) \\ &\quad + P_0(\mathbb{R}^n \setminus D_n(\tfrac{\delta-\varepsilon}{2})) + P_0(\mathbb{R}^n \setminus C_n(\delta)). \end{aligned}$$

Since, by Chernoff's theorem (Chernoff (1952)) and translation equivariance,

$$-\lim_{n\to\infty} n^{-1} \log P_0(\lambda_n(-\varepsilon) \leq 0 \text{ or } \lambda_n(\varepsilon) \geq 0) = e_\psi(\varepsilon),$$

(4.14) is proved. To prove (4.15) it is sufficient to write the inequality

$$\begin{aligned} P_0(|T_n| > \varepsilon) &\geq P_0(\{|T_n| \geq \varepsilon\} \cap C_n(\varepsilon)) \\ &\geq P_0(\{\lambda_n(-\varepsilon) \leq 0 \text{ or } \lambda_n(\varepsilon) \geq 0\} \cap C_n(\varepsilon)) \\ &\geq P_0(\lambda_n(-\varepsilon) \leq 0 \text{ or } \lambda_n(\varepsilon) \geq 0) - P_0(\mathbb{R}^n \setminus C_n(\varepsilon)). \end{aligned}$$

□

PROOF of Theorem 4.4. By the conditions on p, $d(\delta) > 0$ for each $\delta > 0$. We shall prove that

(4.43) $\quad e_\psi(\varepsilon) \to 0 \quad \text{as } \varepsilon \to 0$

and that

(4.44) $\quad c(\delta) > 0 \quad \text{for some } \delta > 0.$

The theorem then follows from Theorem 4.3. Since ψ is bounded and continuous, $\gamma_\theta(\tau)$ is continuous in θ and τ by dominated convergence. Moreover, by strict convexity of $\gamma_0(\cdot)$ we have $\gamma_0(\tau) > 0$ for each $\tau > 0$ or for each $\tau < 0$. Assume the latter, then by pointwise convergence of γ_ε to γ_0 it follows that $\inf_{\tau\leq 0} \gamma_\varepsilon(\tau) \to 0$ as $\varepsilon \to 0$, implying (4.43).

To prove (4.44), let $\int \psi' p dx = c_2 > 0$, then the derivative of

$$\tau \mapsto \log \int e^{\tau(\psi' - \frac{1}{2}c_2)} p dx$$

is positive at $\tau = 0$, implying

$$\inf_{\tau \le 0} \log \int e^{\tau(\psi' - \frac{1}{2}c_2)} p dx = c_3 < 0.$$

By Chernoff's theorem it follows that

$$-\lim_{n\to\infty} n^{-1} \log P(\frac{1}{n}\lambda_n'(0) \ge -\tfrac{1}{2}c_2) = c_3.$$

Since the Lipschitz condition on ψ' is "inherited" by $n^{-1}\lambda_n'$ we have

$$P(\frac{1}{n}\lambda_n'(0) < -\tfrac{1}{2}c_2) \le P(\frac{1}{n}\lambda_n'(t) < 0 \text{ on } (-\tfrac{1}{2}c_1^{-1}c_2, \tfrac{1}{2}c_1^{-1}c_2))$$

and (4.44) is established. □

PROOF of Theorem 4.5. First note that, since $e_{\psi_\varepsilon}(\varepsilon) = M(P_{-\varepsilon}, P_\varepsilon) < K(P_{-\varepsilon}, P_\varepsilon)$,

$$e_{\psi_\varepsilon}(\varepsilon) \to 0 \quad \text{as } \varepsilon \to 0,$$

and that $d(\delta) > 0$ for each $\delta > 0$ as in the previous proof. By continuity of p and the tails condition on p', (4.5) holds for ψ_ε (for each $\varepsilon > 0$), hence it is again sufficient to prove that $c_\varepsilon(\delta) > 0$ for some $\delta > 0$, where $c_\varepsilon(\cdot)$ is defined as $c(\cdot)$ in (4.13) by taking

$$\lambda_n(t) = \lambda_{\varepsilon,n}(t) = \sum_{i=1}^{n} \frac{1}{2\varepsilon} \psi_\varepsilon(X_i - t).$$

Since this is essentially an elaboration of the previous proof, using the mean value theorem to bound ψ, ψ' and ψ'', we omit further details. □

The proofs of Theorems 4.6 and 4.7 are preceded by a useful lemma.

LEMMA 4.10. *If* $p > 0$ *and* $\{T(\hat{P}_n)\}$ *is a translation equivariant estimate such that* (4.26) *and* (4.30) *hold, then* $T(\tilde{R}) = 0$.

PROOF. Assume $T(\tilde{R}) > 0$, then by (4.26)

$$R_\alpha(T(\hat{P}_n) > 0) \to 1 \quad \text{as } n \to \infty,$$

for an $\alpha < \tilde{\alpha}$ with $K(R_\alpha, P_{-\varepsilon}) < M(P_{-\varepsilon}, P_\varepsilon) = s(\varepsilon)$. Corollary 1.2 now implies

$$\limsup_{n\to\infty} n^{-1} \log P_{-\varepsilon}(T(\hat{P}_n) > 0) \ge -K(R_\alpha, P_{-\varepsilon}),$$

contradicting (4.30) by equivariance of $T(\hat{P}_n)$. □

<u>PROOF of Theorem 4.6</u>. Suppose (4.31) fails then without loss of generality there is a k-tuple $x_1,\dots,x_k$ with

$$T(\hat{P}_k) > 0, \quad \Sigma\, \psi_\varepsilon(x_i) < 0.$$

Let $m > \max\{|x_i|\}$, then by the monotonicity and boundedness of ψ_ε on $(-m,m)$ and by (4.25) a pm $Q \in \mathcal{P}_m$ exists such that

$$T(Q) > 0, \quad \int \psi_\varepsilon(x)dQ(x) < 0 \tag{4.45}$$

and

$$K(Q,R) < \infty.$$

By Definition I.2.1 we have

$$\int \log \frac{d\tilde{R}}{dP_{-\varepsilon}}\, dQ = \tilde{\alpha} \int \log \frac{dP_\varepsilon}{dP_{-\varepsilon}}\, dQ - \psi^{P_{-\varepsilon},P_\varepsilon}(\tilde{\alpha}). \tag{4.46}$$

Since $\tilde{\alpha} \in (0,1)$, the last term equals $K(\tilde{R},P_{-\varepsilon})$, cf. (I.2.25). It follows that (4.46) and the second part of (4.45) yield

$$\int \log \frac{d\tilde{R}}{dP_{-\varepsilon}}\, dQ < K(\tilde{R},P_{-\varepsilon}). \tag{4.47}$$

By Csiszár's (1975) lemma (cf. Lemma I.2.1), (4.47) and the convexity of $K(\cdot,P_{-\varepsilon})$ imply that

$$K(Q_\gamma,P_{-\varepsilon}) < K(\tilde{R},P_{-\varepsilon}) \tag{4.48}$$

for each $\gamma \in (0,\gamma_0)$ for some $\gamma_0 > 0$. By Lemma 4.10

$$T(\tilde{R}) = 0$$

hence (4.29) and the first part of (4.45) imply for each $\gamma \in [0,1]$

$$T(Q_\gamma) \geq 0. \tag{4.49}$$

Fix $\gamma > 0$ such that (4.48) (and (4.49)) hold. Since T is translation equivariant, (4.27) implies for each $\delta > 0$

$$Q_\gamma(\cdot-\delta)(T_n > 0) \to 1 \quad \text{as } n \to \infty, \tag{4.50}$$

where $Q_\gamma(\cdot-\delta)$ denotes the pm Q_γ shifted by the amount δ. By (4.28) and (4.48) δ can be chosen small enough to ensure

(4.51) $\quad K(Q_\gamma(.-\delta),P_{-\varepsilon}) < K(\tilde{R},P_{-\varepsilon}).$

Since $K(\tilde{R},P_{-\varepsilon}) = s(\varepsilon)$, (4.50) and (4.51) imply by Corollary 1.2 that (4.30) does not hold. □

PROOF of Theorem 4.7. This proof is similar to the previous one. Again assume that the conclusion of the statement is false, i.e. assume that χ and ψ_ε are not a.e. proportional. First we shall construct a $Q \in \cup P_m$ such that, say,

(4.52) $\quad \int \chi(x)dQ(x) > 0, \quad \int \psi_\varepsilon(x)dQ(x) < 0.$

Let $A = \{x : \psi_\varepsilon(x) < 0, \chi(x) > 0\}$. If $\tilde{R}(A) > 0$ define Q by

(4.53) $\quad \frac{dQ}{d\tilde{R}}(x) = \{\tilde{R}(A \cap (-m,m))\}^{-1} 1_{A\cap(-m,m)}(x)$

for some m with $\tilde{R}(A \cap (-m,m)) > 0$. A similar choice can be made when $R(\psi_\varepsilon > 0, \chi < 0) > 0$, resulting then in (4.52) with reversed inequalities. Now assume that ψ_ε and χ do not attain opposite signs (a.e.).

Let ψ_ε^+, ψ_ε^-, χ^+ and χ^- denote the positive and negative parts of ψ_ε and χ and assume without loss of generality that ψ_ε^+ is not a.e. proportional to χ^+ and that

(4.54) $\quad \int \psi_\varepsilon^- d\tilde{R} = \int \chi^- d\tilde{R}.$

Since $\int \psi_\varepsilon d\tilde{R} = \int \chi d\tilde{R} = 0$ (cf. (4.23),(4.34)), (4.54) holds for the positive parts, too.

Let $U_m = \{x : -m < x < m, \psi_\varepsilon(x) > \chi(x) > 0\}$ and $V_m = \{x : -m < x < m, \psi_\varepsilon(x) < 0 \text{ or } \chi(x) < 0\}$, then by (4.54) an $m < \infty$ exists such that (4.52) holds with Q defined such that $dQ/d\tilde{R}$ is proportional to

(4.55) $\quad \{-\int \psi_\varepsilon 1_{V_m} d\tilde{R}\}^{-1} 1_{V_m} + 2\{\int (\psi_\varepsilon + \chi) 1_{U_m} d\tilde{R}\}^{-1} 1_{U_m}.$

Moreover, (4.55) and, in the previous case, (4.53) imply $K(Q,\tilde{R}) < \infty$. The differentiability condition now implies that

(4.56) $\quad T(Q_\gamma) > T(\tilde{R})$

for each sufficiently small $\gamma > 0$. As in the previous proof, (4.56) and the second part of (4.52) contradict (4.30). □

PROOF of Lemma 4.8. By (4.37), (4.36) and (4.38) we have

$$\lambda_R(T(Q_\gamma)) = -\gamma\lambda_Q(0) + o(\gamma),$$

which by differentiability of λ_R (cf. (4.35)) yields

$$T(Q_\gamma)\lambda_R'(0) + o(T(Q_\gamma)) = -\gamma\lambda_Q(0) + o(\gamma).$$

Divide by $\lambda_R'(0)$ and γ and let $\gamma \downarrow 0$ to obtain G-differentiability and (4.39). □

PROOF of Lemma 4.9. In this proof, denote $Q(\cdot) = Q((-\infty,\cdot])$ for each $Q \in P^*$. Let $Q \in \cup P_m$. We shall prove that (4.33) holds with χ defined by (4.35). Define $q(t) = t(1-t)$, then $\int q(R(x))dx < \infty$ since $E_R|X| < \infty$. Moreover, when $Q_\gamma = R + \gamma(Q-R)$ we have

$$\sup_x \{|(Q_\gamma(x) - R(x))/q(R(x))|\} \to 0 \quad \text{as } \gamma \downarrow 0.$$

By Theorem 2 of Boos (1979) it follows that

$$\lim_{\gamma\downarrow 0} \gamma^{-1}(L(Q_\gamma) - L(R)) = -\int (Q(x) - R(x))J(R(x))dx. \tag{4.57}$$

Define

$$\chi_1(x) = \int_0^x J(R(u))du$$

then for each $m > m_0$ with $Q \in P_{m_0}$ we obtain, integrating by parts,

$$\begin{aligned} -\int_{-m}^{m} (Q-R)J(R)dx &= -(Q-R)\chi_1\Big|_{-m}^{m} + \int_{-m}^{m} \chi_1 d(Q-R) \\ &= -(1-R(m))\chi_1(m) - R(-m)\chi_1(-m) + \int_{-m}^{m} \chi_1 dQ - \int_{-m}^{m} \chi_1 dR. \end{aligned} \tag{4.58}$$

Now let $m \to \infty$, then the first two terms of the last member of (4.58) tend to zero; the last two tend to $\int \chi(x)dQ(x)$, completing the proof by (4.57). □

In the remainder of this section we give some examples of curious phenomena. The first is a consistent estimate in the double exponential family which attains Bahadur's bound at $\theta = 0$ for each $\varepsilon > 0$.

EXAMPLE 4.11. Let $p(x) = \frac{1}{2}\exp\{-|x|\}$. We shall construct an estimate which is based on the discretized Kullback-Leibler information of $\hat{P}_n$ with respect to P_0, and show it to be consistent and to attain Bahadur's bound at $\theta = 0$ for each $\varepsilon > 0$.

For each $k = 1,2,\ldots,$ let

$$A_{k1} = (-\infty, a_1), \quad A_{kj} = [a_{j-1}, a_j), \quad j = 2,\ldots,k-1,$$
$$A_{kk} = [a_{k-1}, \infty),$$

with $a_1,\ldots,a_{k-1}$ chosen such that $P_0(A_{kj}) = 1/k$.

Define for pms $Q \in P^*$ the discretized Kullback-Leibler information with respect to P_0 by

$$K_k(Q) = \sum_{j=1}^{k} Q(A_{kj}) \log \frac{Q(A_{kj})}{P_0(A_{kj})}. \tag{4.59}$$

Now $\theta \to K(P_\theta, P_0)$ is continuous and increasing on $[0,\infty)$; let K^{-1} be its inverse.

Define the functionals $T^{(k)}$ by

$$T^{(k)}(Q) = K^{-1}(K_k(Q)) \cdot \text{sign } (Q^{-1}(\tfrac{1}{2}))$$

and the estimate T_n as

$$T_n = T^{(k(n))}(\hat{P}_n),$$

where $k(n)$ tends to ∞ at a suitable rate to be determined later.

We prove that $\{T_n\}$ is consistent for certain sequences $\{k(n)\}$:

Fix $\theta \in (-\infty,\infty)$ and let $Z_{nkj} = \hat{P}_n(A_{kj})$ and $p_{kj} = P_\theta(A_{kj})$, then

$$\begin{aligned} |K_k(\hat{P}_n) - K_k(P_\theta)| &= \left| \sum_{j=1}^{k} [Z_{nkj} \log Z_{nkj} - p_{kj} \log kp_{kj}] \right| \\ &\le \Sigma \frac{1}{k} |kZ_{nkj} \log kZ_{nkj} - kp_{kj} \log kp_{kj}|. \end{aligned} \tag{4.60}$$

Furthermore, since P_θ is double exponential, there are constants c_1 and c_2 ($e^{-\theta}$ and e^{θ}) such that for all k,j

$$\frac{c_1}{k} \le P_\theta(A_{kj}) \le \frac{c_2}{k}, \tag{4.61}$$

implying for each $c > 0$ that

$$\begin{aligned} P_\theta(\bigcup_j \{|Z_{nkj} - p_{kj}| > \tfrac{c}{k}\}) &\le \sum_{j=1}^{k} P_\theta(|Z_{nkj} - p_{kj}| > \tfrac{c}{k}) \\ &\le \sum_{j=1}^{k} \frac{p_{kj}(1-p_{kj})}{n(c/k)^2} \le k^2 \frac{c_2}{nc^2} \end{aligned} \tag{4.62}$$

where Chebyshev's inequality was used. The upper bound of (4.62) tends to zero for each $c > 0$ provided

(4.63) $$k = k(n) = o(\sqrt{n}) \quad \text{as } n \to \infty.$$

Fix $\delta > 0$. Since $x \log x$ is uniformly continuous on $[c_1, c_2]$, a constant c_3 exists such that the last member of (4.60) is smaller than δ when $|Z_{nkj} - p_{kj}| \le c_3/k$ for each j. When $k = k(n)$ satisfies (4.63) it follows from (4.60) and (4.62) that

(4.64) $$P_\theta(|K_k(\hat{P}_n) - K_k(P_\theta)| > \delta) \to 0 \quad \text{as } n \to \infty.$$

Since $K_k(P_\theta) \to K(P_\theta, P_0)$ as $k \to \infty$, (4.64) implies, as k and $n \to \infty$ with $k = o(\sqrt{n})$,

$$P_\theta(|K_k(\hat{P}_n) - K(P_\theta, P_0)| > \delta) \to 0.$$

It follows that $K^{-1}(K_k(\hat{P}_n)) \xrightarrow{P_\theta} |\theta|$. The correct sign is implied by consistency of the sample median.

Now we prove that for each $\varepsilon > 0$

(4.65) $$e(\varepsilon, 0, \{T_n\}) = b(\varepsilon).$$

Fix $\varepsilon > 0$. By the definition of T_n we have

$$\begin{aligned} P_0(|T_n| > \varepsilon) &= P_0(K_k(\hat{P}_n) > K(P_\varepsilon, P_0)) \\ &= P_0(K_k(\hat{P}_n) > b(\varepsilon)) \le \exp\{-nb(\varepsilon) + O(k \log n)\} \end{aligned}$$

as $k = k(n) \to \infty$ and $n \to \infty$, where the last inequality was proved, using the technique of the proof of Lemma 3.1 in Groeneboom *et al* (1979), in Kester (1978), proof of (2.22).

It follows that (4.65) holds when $k = k(n) = o(n/\log n)$ as $n \to \infty$, which is implied by (4.63).

REMARK 4.8. The proof of consistency leans rather heavily on the bounds (4.61). Such bounds are available when p has exponential or heavier tails. The estimate can therefore also be constructed for the logistic and Cauchy densities.

In Sections 1 and 2 we have seen that in shift families a bound on the inaccuracy rate is attained, under some conditions, by an M-estimate which is related to that bound in a natural way. We shall now show that the analogous estimate does not attain the generalized bound in location-scale families.

Let a location-scale family $\{P_{\mu,\sigma} : \mu \in \mathbb{R},\ \sigma \in \mathbb{R}^+\}$ be given by its Lebesgue densities

$$p_{\mu,\sigma}(x) = \sigma^{-1} p(\sigma^{-1}(x-\mu)),$$

and consider location-scale equivariant estimates $\{(T_n,S_n)\}$ of (μ,σ), i.e. when $Y_i = \sigma(X_i + \mu)$ for each i, then

$$T_n(Y_1,\ldots,Y_n) = \sigma(T_n(X_1,\ldots,X_n) + \mu),$$
$$S_n(Y_1,\ldots,Y_n) = \sigma S_n(X_1,\ldots,X_n).$$

Inaccuracy rates are defined as

$$-\limsup_{n\to\infty} n^{-1} \log P_{\mu,\sigma}(|T_n - \mu| > \sigma\varepsilon)$$

and

$$-\limsup_{n\to\infty} n^{-1} \log P_{\mu,\sigma}(\sigma^{-1}S_n \notin [\rho^{-1},\rho]),$$

making them location-scale invariant, and just like in the pure location case, bounds on these rates for equivariant estimates are found as, cf. Lemma 4.1,

$$M(P_{-\varepsilon,1}, P_{\varepsilon,1}) \tag{4.66}$$

for the location estimate and

$$M(P_{0,1/\rho}, P_{0,\rho}) \tag{4.67}$$

for the scale estimate. The "candidate" M-estimates to attain these bounds are found as a suitable zero or change of sign of

$$\Sigma\,\psi\left(\frac{X_i - t}{s}\right), \quad \Sigma\,\chi\left(\frac{X_i - t}{s}\right)$$

where $\psi = \log(dP_{\varepsilon,1}/dP_{-\varepsilon,1})$ and $\chi = \log(dP_{0,\rho}/dP_{0,1/\rho})$. When p is the standard normal density, the resulting estimates

$$\overline{X}_n, \left\{\frac{\rho^2 - \rho^{-2}}{4 \log \rho}\ \frac{1}{n}\,\Sigma\,(X_i - \overline{X}_n)^2\right\}^{\frac{1}{2}}$$

indeed attain the bounds (4.66) and (4.67). For the scale estimate this follows since under $P_{0,1}$, $\Sigma_1^n (X_i - \overline{X})^2$ has the same distribution as $\Sigma_1^{n-1} X_i^2$.

In the double exponential family, however, the bound is not attained by the proposed estimates, and the results of Section 3 suggest that other M-estimates will not attain it either.

Let

$$p_{\mu,\sigma}(x) = \frac{1}{2\sigma} \exp\left\{-\left|\frac{x-\mu}{\sigma}\right|\right\},$$

then

$$\psi(x) = 2 \max(-\varepsilon, \min(x,\varepsilon)),$$

$$\chi(x) = (\rho - \tfrac{1}{\rho})|x| - 2 \log \rho$$

"should" produce M-estimates which attain the bounds (4.66) and (4.67). Let (T,S) be the functional-pair corresponding to these M-estimates, then

$$\int \chi\left(\frac{x - T(Q)}{S(Q)}\right) dQ = 0$$

implies

$$S(Q) = \frac{\rho - \rho^{-1}}{2 \log \rho} E_Q|X - T(Q)|.$$

Note that $(\rho - \rho^{-1})/(2 \log \rho)$ is larger than one for each $\rho > 1$.

Let $\tilde{R}$ be the (unique, cf. Lemma I.2.2) pm with

$$K(\tilde{R}, P_{-\varepsilon,1}) = K(\tilde{R}, P_{\varepsilon,1}) = M(P_{-\varepsilon,1}, P_{\varepsilon,1}).$$

We shall show that a pm Q exists such that $Q_\gamma = \gamma Q + (1-\gamma)\tilde{R}$ satisfies for each $\gamma \in (0,1]$

(4.68) $$\int \psi\left(\frac{x}{S(Q_\gamma)}\right) dQ_\gamma > 0,$$

implying $T(Q_\gamma) > 0$, and such that

(4.69) $$\int \psi(x) dQ < 0$$

which implies then as in the proof of Theorem 4.6 that $T_n = T(\hat{P}_n)$ does not attain the bound $M(P_{-\varepsilon,1}, P_{\varepsilon,1})$.

Let $\varepsilon = 1$ and take Q such that its density q is given by

$$q(x) = 4\alpha 1_{(-5/4,-1)}(x) + (1-\alpha)1_{(7,8)}(x)$$

with a fixed $\alpha \in (1/2, 10/19)$.

Now $E_Q|X-t| > 5/4$ for each $t \in \mathbb{R}$.

Moreover, $\tilde{R}$ is given by

$$d\tilde{R}/dx = \begin{cases} \frac{1}{4}, & |x| \le 1 \\ \frac{1}{4}e^{-|x-1|}, & |x| > 1 \end{cases}$$

and hence $E_{\tilde{R}}|X-t| \ge 5/4$ for each $t \in \mathbb{R}$, too. It follows that for each $\gamma \in (0,1]$ and $t \in \mathbb{R}$

$$E_{Q_\gamma}|X-t| > 5/4,$$

implying $S(Q_\gamma) > 5/4$. It is now easily checked that (4.68) and (4.69) hold.

REMARK 4.9. The proposed scale estimates in the normal and double exponential families above are both inconsistent. This is related to the fact that

$$\int \chi(x)\, dP_{0,1}(x) \ne 0$$

as discussed for location M-estimates at the end of Section 4b. Taking $\chi = |x| - 1$ yields a consistent scale estimate for the double exponential family but does not remedy the suboptimality of the proposed location estimate.

REMARK 4.10. It was proved in Example 4.10 that a trimmed mean (which is scale equivariant by its nature) attains the bound (4.66). This shows the bound to be sharp, in the double exponential family.

CHAPTER III

THE BAHADUR DEFICIENCY OF A TWO-SAMPLE CONDITIONAL TEST IN ONE-PARAMETER EXPONENTIAL FAMILIES

1. INTRODUCTION

We consider a full one-parameter exponential family $\{P_\theta : \theta \in \Theta^*\}$ of probability measures (pms) on $\mathbb{R}$ given by its densities with respect to a non-degenerate pm P_0 as

$$(1.1) \qquad dP_\theta(x) = \exp\{\theta x - \psi(\theta)\}dP_0(x), \qquad \theta \in \Theta^*,$$

where θ is the log-moment generating function of P_0 and Θ^* is the full parameter space, cf. Section I.2a.

Suppose we have a sample $X_1,\ldots,X_m$ from a P_{θ_1}-distribution and another sample $Y_1,\ldots,Y_n$ from P_{θ_2}, and suppose the hypothesis $\theta_1 = \theta_2$ is to be tested against the alternative $\theta_1 > \theta_2$ at level α. Since $\Sigma_1^m X_i$ and $\Sigma_1^n Y_i$ are the sufficient statistics for θ_1 and θ_2, it is convenient to abbreviate these statistics:

Define, suppressing the dependence on m and n,

$$X = \sum_{i=1}^m X_i, \quad Y = \sum_{i=1}^n Y_i.$$

Furthermore, define

$$R = X + Y.$$

$P_{\theta_1\theta_2}$, $E_{\theta_1\theta_2}$ and $\mathrm{var}_{\theta_1\theta_2}$ will denote the joint pm of X,Y and the expectation and variance with respect to $P_{\theta_1\theta_2}$; the pm induced by a function f of X and Y will be written as $P_{\theta_1\theta_2}^{f(X,Y)}$.

The uniformly most powerful unbiased (UMPU) test for the testing problem is the conditional test δ_c defined on the possible outcomes of (X,Y) by (Lehmann (1959), Sections 4.4 and 4.5)

$$(1.2) \qquad \delta_c(x,y) = \begin{cases} 1 & > \\ \gamma(x+y) & \text{when } x = k(x+y), \\ 0 & < \end{cases}$$

where $\gamma(r)$ and $k(r)$ satisfy $0 \le \gamma(r) < 1$ and

$$(1.3) \qquad \gamma(r)P_{\theta\theta}(X = k(r) \mid R = r) + P_{\theta\theta}(X > k(r) \mid R = r) = \alpha.$$

Note that since R is sufficient for θ when $\theta_1 = \theta_2 = \theta$, $\gamma(r)$ and $k(r)$ can be chosen independent of θ.

EXAMPLE 1.1. Let $X_1,\dots,X_m$ and $Y_1,\dots,Y_n$ be two random samples from Bernoulli distributions with parameters p_1 and p_2, respectively. Note that the Bernoulli distribution is written in the form (1.1) by taking $\theta = \log(p/(1-p))$, $\psi(\theta) = \log 2(e^\theta + 1)$ and $P_0(\{0\}) = P_0(\{1\}) = \frac{1}{2}$.

Testing $p_1 = p_2$ against $p_1 > p_2$ is equivalent to testing $\theta_1 = \theta_2$ against $\theta_1 > \theta_2$. The critical values and randomization probabilities $k(r)$ and $\gamma(r)$ of Fisher's exact test, cf. Example I.1.1, are determined by (1.3); the conditional distribution of X given $X + Y = r$ is hypergeometric with parameters m+n, r and m.

The aim of this chapter is to determine how much the power of the conditional test falls short of the envelope power, in terms of Bahadur efficiency and deficiency.

Let $H = \{(\theta,\theta) : \theta \in \Theta^*\}$ be the null hypothesis and let $K = \{(\theta_1,\theta_2)\}$ be a fixed point of the alternative hypothesis $H_1 = \{(\theta_1,\theta_2) : \theta_1 > \theta_2\}$. Furthermore, let $\{m_N\}$ and $\{n_N\}$ be nondecreasing sequences of integers such that $m_N + n_N = N$ for each $N = 1,2,\dots$. Define $N_+(\alpha,\beta,\theta_1,\theta_2)$ to be the smallest number N such that the MP test of H against K of size α, based on observations $X_1,\dots,X_{m_N},Y_1,\dots,Y_{n_N}$, has power at least β in K. Let $N_c(\alpha,\beta,\theta_1,\theta_2)$ be similarly defined for the conditional test. Note that N_+ and N_c also depend on the sequence of ratio's $\{m_N/N\}$.

The Bahadur efficiency e^B (of the conditional test versus the MP test) is defined as

$$(1.4) \qquad e^B = \lim_{\alpha\to 0} \frac{N_+(\alpha,\beta,\theta_1,\theta_2)}{N_c(\alpha,\beta,\theta_1,\theta_2)},$$

keeping the other parameters fixed. Since the Bahadur efficiency turns out to equal 1 in our testing problem, we consider the limiting behaviour of the deficiency, i.e. the difference of the sample sizes $N_c - N_+$.

The Bahadur deficiency is said to be of order $\mathcal{O}(D(N))$ when

$$\frac{N_c - N_+}{D(N_+)} = \mathcal{O}(1) \quad \text{as } \alpha \to 0,$$

keeping again the parameters β, θ_1 and θ_2 fixed. Deficiency of order $o(D(N))$ is similarly defined. Deficiency of order $\mathcal{O}(1)$ is also called bounded deficiency.

We shall consider families (1.1) such that P_0 is a lattice distribution or such that P_0 has a Lebesgue density satisfying the regularity condition (2.6); we shall prove that the Bahadur deficiency of the conditional test with respect to the MP test is bounded, when the ratio m_N/N remains bounded away from 0 and 1. An explicit upper bound for the asymptotic deficiency will be obtained when P_0 has a density.

Boundedness of the Bahadur deficiency implies that the Bahadur efficiency equals 1.

When $\{P_\theta : \theta \in \Theta^*\}$ is the normal shift family with known variance, the conditional test (1.2) is equal to the well known Gauss test for the difference of two normal means, which is uniformly most powerful (UMP). It follows that in this special case the deficiency is identically zero. There is another exponential family where the conditional test is UMP, the gamma scale family with known shape parameter. This fact seems to have been overlooked in textbooks. We give a proof:

Let $\{P_\theta : \theta \in \Theta^*\}$ with $\Theta^* = (-\infty, 0)$ be given by

$$dP_\theta(x) = \exp\{\theta x + \log(-\theta)\}(\Gamma(\xi))^{-1} x^{\xi-1} dx, \quad x > 0,$$

for some fixed $\xi > 0$. Let $(\theta_1, \theta_2) \in H_1$ and let, for each $\theta \in \Theta^*$, δ_θ be the MP test of (θ,θ) against (θ_1,θ_2). This test is based on the likelihood ratio $dP_{\theta_1\theta_2}/dP_{\theta\theta}$ and given by

$$\delta_\theta(X,Y) = \begin{cases} 1 \\ 0 \end{cases} \text{when } (\theta_1 - \theta)X + (\theta_2 - \theta)Y \begin{matrix} \geq \\ < \end{matrix} c_\theta$$

where c_θ is determined by $E_{\theta\theta}\delta_\theta(X,Y) = \alpha$. When δ_+ is the MP test of H against (θ_1,θ_2) with size α then, since $(\theta,\theta) \subset H$, we have

$$(1.5) \qquad \beta_+ \leq \beta_\theta$$

for each $\theta \in (-\infty,0)$. Since the conditional distribution of X/r given $X+Y = r$ is the beta $(m\xi,n\xi)$ distribution (independent of r), the conditional test (1.2) is given by

$$\delta_c(X,Y) = \begin{cases} 1 & \\ 0 & \end{cases} \text{when } X/(X+Y) \begin{matrix} \geq \\ < \end{matrix} k,$$

where k is the $(1-\alpha)$-quantile of the beta $(m\xi,n\xi)$ distribution. The power of the conditional test cannot exceed that of the MP test:

$$(1.6) \qquad \beta_c \leq \beta_+ .$$

Now choose δ such that $\delta_\theta = \delta_c$, i.e. take $\theta = k\theta_1 + (1-k)\theta_2$, then (1.5) and (1.6) imply $\beta_+ = \beta_c$: the test δ_c is MP for H against (θ_1,θ_2). Since (θ_1,θ_2) is arbitrary, δ_c is UMP.

The same proof applies to the normal shift family, with the difference that there the distribution of $X - mr/N$ given $R = r$ is independent of r. In both cases the rejection region of the conditional test has a linear boundary whence one of the tests δ_θ is the same as δ_c.

2. MAIN RESULTS

2a. Preliminaries and results

In this section we present the main results, but we start with some, mostly notational, preliminaries.

Let $\Theta^* = \{\theta : \int e^{\theta x} dP_0(x) < \infty\}$ be the full parameter space of the exponential family (1.1) and assume $0 \in \text{int}\ \Theta^*$. The first three derivatives of ψ are denoted as $\lambda = \psi'$, $\sigma^2 = \psi''$ and $\rho = \psi'''$. Note that $\rho(\theta) = E_\theta(X_1 - \lambda(\theta))^3$. Where unambiguous, we write λ_i, σ_i^2, σ_ν^2, ρ_i instead of $\lambda(\theta_i)$, $\sigma^2(\theta_i)$, $\sigma^2(\theta_\nu)$ and $\rho(\theta_i)$, see also (4.3) and (4.4).

Kullback-Leibler information will occur nearly always in a linear combination I_ν, defined by

$$(2.1) \qquad I_\nu(\theta_1,\theta_2,\theta) = \nu K(\theta_1,\theta) + (1-\nu)K(\theta_2,\theta).$$

Note that $\theta \mapsto I_\nu(\theta_1,\theta_2,\theta)$ is strictly convex since both K's in (2.1) are. In view of (I.2.11), the derivative $(\partial/\partial\theta)I_\nu(\theta_1,\theta_2,\theta)$ equals

(2.2) $\quad -\nu\lambda(\theta_1) - (1-\nu)\lambda(\theta_2) + \lambda(\theta);$

we find that θ_ν, defined as

(2.3) $\quad \theta_\nu = \lambda^{-1}(\nu\lambda(\theta_1) + (1-\nu)\lambda(\theta_2))$

is the value of θ which minimizes $I_\nu(\theta_1,\theta_2,\theta)$ for fixed θ_1,θ_2. Furthermore, we note that

(2.4) $\quad \frac{\partial}{\partial\nu} I_\nu(\theta_1,\theta_2,\theta_\nu) = K(\theta_1,\theta_\nu) - K(\theta_2,\theta_\nu).$

Our theorems will be stated for families (1.1) such that either

(2.5) $\quad P_0$ has a lattice distribution with minimal lattice $\mathbb{Z}$

or

(2.6) $\quad P_0$ has a Lebesgue density and for each compact subset Θ of int Θ^* there is a $k \geq 1$ such that the densities $p_\theta^{*k}(x)$ of the k-th convolutions of P_θ are bounded, uniformly for $\theta \in \Theta$ and $x \in \mathbb{R}$.

REMARK 2.1. The requirement that the minimal lattice be $\mathbb{Z}$ in (2.5) is imposed to avoid non-essential transformations later on. Condition (2.6) implies a regularity property of the characteristic function of P_θ which is used to obtain uniformity in the limit theorems of Section 3, cf. Lemma 3.1. Note that (2.6) is stronger than boundedness of the density p_0^{*k} for some $k \geq 1$. To see this, modify the standard normal density such that for each $x \in \mathbb{Z}$, the probability mass of $[x-\frac{1}{2},x+\frac{1}{2})$ is concentrated on a smaller interval $(x-\varepsilon_x,x+\varepsilon_x)$ and such that the resulting density $p_0(x)$ equals 1 for each $x \in \mathbb{Z}$. For any $\theta \neq 0$, the density p_θ is now unbounded.

In the sequel we assume that the ratio m_N/N, denoted by ν_N, remains bounded away from zero and one, i.e. for some $\varepsilon > 0$ it holds that

(2.7) $\quad \nu_N = m_N/N \in [\varepsilon,1-\varepsilon]$ for each $N \geq 2$.

Now let Ω be a convex compact subset of int H_1 and let $0 < \varepsilon < \frac{1}{2}$.

THEOREM 2.1. *If* (2.5) *or* (2.6) *holds then*

(2.8) $\quad N_c(\alpha,\beta,\theta_1,\theta_2) - N_+(\alpha,\beta,\theta_1,\theta_2) = \mathcal{O}(1) \quad$ as $\alpha \to 0$, *uniformly for* $\beta \in [\varepsilon, 1-\varepsilon]$ *and* $(\theta_1, \theta_2) \in \Omega$.

THEOREM 2.2. *If* (2.6) *holds then, uniformly for* $\beta \in [\varepsilon, 1-\varepsilon]$ *and* $(\theta_1, \theta_2) \in \Omega$ *as* $\alpha \to 0$,

$$N_c(\alpha,\beta,\theta_1,\theta_2) - N_+(\alpha,\beta,\theta_1,\theta_2) \le D(\nu_{N_+},\theta_1,\theta_2) + o(1), \tag{2.9}$$

where N_+ *abbreviates* $N_+(\alpha,\beta,\theta_1,\theta_2)$,

$$D(\nu,\theta_1,\theta_2) = 1 + (\min\{K(\theta_1,\theta_\nu),K(\theta_2,\theta_\nu)\})^{-1} \cdot \tfrac{1}{2}(R_\nu - 1 - \log R_\nu), \tag{2.10}$$

$$R_\nu = \frac{\nu(1-\nu)(\theta_1-\theta_2)^2\sigma_1^2\sigma_2^2}{\tau_\nu^2\sigma_\nu^2} \tag{2.11}$$

(*recall that* $\sigma_\nu^2 = \sigma^2(\theta_\nu)$) *and*

$$\tau_\nu^2 = \nu(\theta_1-\theta_\nu)^2\sigma_1^2 + (1-\nu)(\theta_2-\theta_\nu)^2\sigma_2^2 . \tag{2.12}$$

An often feasible way to prove theorems on Bahadur efficiency and deficiency is as follows. Suppose tests $\{\delta_1^{N,\alpha}\}$ and $\{\delta_2^{N,\alpha}\}$ are to be compared. One defines for each test $\alpha_i(N,\beta)$ to be the smallest size α such that $\delta_i^{N,\alpha}$ has power at least β at a fixed alternative. Asymptotic expansions for $\alpha_1(N,\beta)$ and $\alpha_2(N,\beta)$ as $N \to \infty$ are then inverted to obtain expansions for the minimal sample sizes $N_1(\alpha,\beta)$ and $N_2(\alpha,\beta)$ as $\alpha \to 0$. For the present problem this approach meets two difficulties.

In the first place, the MP test of H against the simple alternative $\{(\theta_1,\theta_2)\}$ is not explicitly known (it exists but involves a generally unknown least favourable distribution on H, cf. Lehmann (1959)). Secondly, the smallest size $\alpha_c(N,\beta)$ for the conditional test to attain power β is not easily found as a function of β: one would have to invert, for fixed $(\theta_1,\theta_2) \in H_1$,

$$\beta_c(N,\alpha) = \int E_{\theta_1\theta_2}(\delta_c^{N,\alpha} | R = r)\, dP^R_{\theta_1\theta_2}(r),$$

where $\delta_c^{N,\alpha}$ is defined by (1.2) and (1.3).

The proofs of theorems 2.1 and 2.2 circumvent these problems: In Section 4, a lower bound on $\alpha_+(N,\beta)$ is found as the size $\alpha_N(\beta)$ of the MP test of (θ,θ) against (θ_1,θ_2) for a suitable choice of θ, in Section 5 we derive an implicit expression for the critical values $k(r)$ of the conditional test, and in Section 6 these are combined to prove that the power $\beta_c(N,\alpha_N(\beta))$ exceeds $\beta - N^{-\frac{1}{2}}S$ for some finite number S. The main result is then a consequence of the following lemma.

<u>LEMMA 2.3</u>. *Consider for each* N *and* α *tests* $\delta_1^{N,\alpha}$ *and* $\delta_2^{N,\alpha}$. *Define* $\beta_i(N,\alpha)$ *to be the power of* $\delta_i^{N,\alpha}$ *at a fixed point* (θ_1,θ_2) *of* H_1 *and let* $N_i(\alpha,\beta)$ *be the smallest* N *such that* $\beta_i(N',\alpha) \geq \beta$ *for each* $N' \geq N$, $i = 1,2$. *Furthermore, let* $\alpha_i(N,\beta)$ *be the smallest size with* $\beta_i(N,\alpha) \geq \beta$, $i = 1,2$.

Now assume that $N_2(\alpha,\beta) \geq N_1(\alpha,\beta)$ *for each* α *and* β, *and that for each* N

$$(2.12) \qquad \alpha > \alpha' \Rightarrow \beta_i(N,\alpha) \geq \beta_i(N,\alpha'), \quad i = 1,2.$$

Let $\alpha_N(\beta)$ *satisfy for each* N *and* β

$$(2.13) \qquad \alpha_N(\beta) \leq \alpha_1(N,\beta)$$

and, uniformly for $\beta \in [\varepsilon,1-\varepsilon]$ *as* $N \to \infty$,

$$(2.14) \qquad \alpha_N(\beta) = N^{-\frac{1}{2}} \exp\{-NI(\nu) + N^{\frac{1}{2}}g(\nu,\beta) - E(\nu,\beta) + o(1)\},$$

where E *is continuous,* g *is continuously differentiable and* $I(\nu)$ *abbreviates* $I_\nu(\theta_1,\theta_2,\theta_\nu)$. *Moreover, let* $S(\cdot,\cdot)$ *be a continuous function such that uniformly for* $\beta \in [\varepsilon,1-\varepsilon]$ *as* $N \to \infty$

$$(2.15) \qquad \beta_2(N,\alpha_N(\beta)) \geq \beta - N^{-\frac{1}{2}}S(\nu,\beta) + o(N^{-\frac{1}{2}}).$$

For each $\varepsilon' > \varepsilon$ *we have, uniformly for* $\beta \in [\varepsilon',1-\varepsilon']$ *as* $\alpha \to 0$,

$$(2.16) \qquad N_2(\alpha,\beta) - N_1(\alpha,\beta) \leq \{\min(K(\theta_1,\theta_{\nu_1}),K(\theta_2,\theta_{\nu_2}))\}^{-1} \cdot S(\nu_1,\beta)\frac{\partial}{\partial\beta}g(\nu_1,\beta) + 1 + o(1),$$

where $\nu_1 = \nu_{N_1(\alpha,\beta)}$.

This lemma will be proved in Section 2b. Note that condition (2.12) holds for the MP test as well as for the conditional (UMPU!) test (1.2). For these tests the power is monotone in N, too.

Section 3 contains some central limit theorems which are used in Sections 4 to 6 to prove the main result.

In Section 7 we study the conditional test for different values of the sample ratio $\nu_N = m_N/N$. We shall find that the Bahadur-optimal ratio is equal to the constant ν_0 which maximizes $I_\nu(\theta_1,\theta_2,\theta_\nu)$ as a function of ν, cf. (2.4). Moreover, we prove that the conditional test with a sample ratio which does not converge to ν_0 has Bahadur efficiency lower than 1 with respect to the optimal-ratio conditional test. These results are compared with those of Albers (1974), who studied the Pitman efficiency and deficiency of these tests.

For Fisher's exact test, cf. Example 1.1, numerically determined bounds on the deficiency are presented in Section 8.

2b. Proof of Lemma 2.3.

The proof is somewhat tedious since the most significant term in (2.14), $NI(\nu)$ depends on N and on the sample ratio in a complicated way. We first present a helpful lemma:

LEMMA 2.4. *If* $\nu_N \in [\varepsilon, 1-\varepsilon]$ *for each* $N \in \mathbb{N}$ *then* $N \geq M$ *implies*

$$NI_{\nu_N}(\theta_1,\theta_2,\theta_{\nu_N}) - MI_{\nu_M}(\theta_1,\theta_2,\theta_{\nu_M}) \geq \gamma(N-M), \tag{2.17}$$

with $\gamma = \min\,(K(\theta_1,\theta_{1-\varepsilon}), K(\theta_2,\theta_\varepsilon), I_\varepsilon(\theta_1,\theta_2,\theta_\varepsilon), I_{1-\varepsilon}(\theta_1,\theta_2,\theta_{1-\varepsilon}))$.

PROOF. Abbreviate $I(\nu) = I_\nu(\theta_1,\theta_2,\theta_\nu)$, $\nu = \nu_N$, $\mu = \nu_M$. First assume $\nu \leq \mu$. By the mean value theorem we have for some ξ with $\nu \leq \xi \leq \mu$

$$\nu^{-1}I(\nu) - \mu^{-1}I(\mu) = (\nu^{-1} - \mu^{-1})\left(\frac{d}{d\alpha}\,\alpha I(\alpha^{-1})\right)_{\alpha=\xi^{-1}}. \tag{2.18}$$

Since by (2.1) and (2.4)

$$\left(\frac{d}{d\alpha}\,\alpha I(\alpha^{-1})\right)_{\alpha=\xi^{-1}} = I(\xi) - \xi I'(\xi) = K(\theta_2,\theta_\xi),$$

multiplication of (2.18) by ν and the definition of γ yield

$$I(\nu) \geq \frac{\nu}{\mu}\,I(\mu) + \left(1 - \frac{\nu}{\mu}\right)\gamma,$$

implying

$$NI(\nu) - MI(\mu) \geq \left(N\frac{m_N/N}{m_M/M} - M\right)I(\mu) + N\left(1 - \frac{m_N/N}{m_M/M}\right)\gamma$$

$$= M\left(\frac{m_N}{m_M} - 1\right)(I(\mu) - \gamma) + (N - M)\gamma \geq (N-M)\gamma.$$

The last inequality holds since $m_N \geq m_M$ and since by the concavity of I, $I(\mu) \geq \gamma$. Now assume $\nu > \mu$. By the mean value theorem,

$$(1-\nu)^{-1}I(\nu) - (1-\mu)^{-1}I(\mu) = ((1-\nu)^{-1} - (1-\mu)^{-1})\left(\frac{d}{d\alpha}\alpha I(1-\alpha^{-1})\right)_{\alpha=(1-\xi)^{-1}}.$$

As above we obtain

$$I(\nu) \geq \frac{1-\nu}{1-\mu} I(\mu) + \left(1 - \frac{1-\nu}{1-\nu}\right)\gamma$$

and (2.18) follows after substitution of $1-\nu = n_N/N$, etc. □

PROOF of Lemma 2.3. Let $\alpha > 0$ and $\beta \in [\varepsilon', 1-\varepsilon']$.

By (2.15) there is a sequence $\{\beta_N\}$ with

$$(2.19) \qquad \beta_N = \beta + N^{-\frac{1}{2}} S(\nu,\beta) + o(N^{-\frac{1}{2}}),$$

such that for each N

$$(2.20) \qquad \beta_2(N, \alpha_N(\beta_N)) \geq \beta.$$

Choose N large enough to ensure $\alpha_N(\beta_N) \leq \alpha$ then (2.20) and (2.12) imply $N_2(\alpha,\beta) < \infty$ and hence $N_1(\alpha,\beta) < \infty$. It follows that for $i = 1,2$,

$$\beta_i(N_i(\alpha,\beta),\alpha) \geq \beta > \beta_i(N_i(\alpha,\beta)-1,\alpha)$$

implying

$$(2.21) \qquad \alpha_i(N_i(\alpha,\beta),\beta) \leq \alpha \leq \alpha_i(N_i(\alpha,\beta)-1,\beta), \quad i = 1,2.$$

Abbreviating $N_i = N_i(\alpha,\beta)$ we obtain from (2.13), (2.20) and (2.21) that

$$\alpha_{N_1}(\beta) \leq \alpha_1(N_1,\beta) \leq \alpha \leq \alpha_2(N_2-1,\beta) \leq \alpha_{N_2-1}(\beta_{N_2-1})$$

which by (2.14) implies, writing ν_1 for ν_{N_1} and ν_2' for ν_{N_2-1}, that

$$(2.22) \qquad N_1 I(\nu_1) - N_1^{\frac{1}{2}} g(\nu_1,\beta) + \tfrac{1}{2}\log N_1 + E(\nu_1,\beta)$$
$$\geq (N_2-1) I(\nu_2') - (N_2-1)^{\frac{1}{2}} g(\nu_2',\beta) + \tfrac{1}{2}\log (N_2-1)$$
$$+ E(\nu_2',\beta) - S(\nu_2',\beta)\frac{\partial}{\partial\beta} g(\nu_2',\beta) + o(1) \qquad \text{as } \alpha \to 0.$$

Here, the right hand side was obtained by Taylor expansion of the terms containing β_{N_2-1}, cf. (2.19). Since g and E are bounded, (2.22) yields

$$(N_2-1) I(\nu_2') - N_1 I(\nu_1) \leq O(N_2^{\frac{1}{2}}) \qquad \text{as } \alpha \to 0,$$

and by the assumption $N_2 \geq N_1$ this implies in view of Lemma 2.4,

$$N_2 - 1 - N_1 = O(N_2^{\frac{1}{2}}) = O(N_1^{\frac{1}{2}}) \quad \text{as } \alpha \to 0.$$

It follows that $\nu_2' - \nu_1 = O(N_1^{-\frac{1}{2}})$. By Taylor expansion of the terms containing N_2 and ν_2', one now finds from (2.22) and Lemma 2.4 that

$$(2.23) \qquad N_2 - 1 - N_1 = O(1), \quad \nu_2' - \nu_1 = O(N_1^{-1}) \quad \text{as } \alpha \to 0.$$

Insertion of (2.23) into (2.22) finally yields

$$(N_2 - 1)I(\nu_2') - N_1 I(\nu_1) \le S(\nu_1,\beta)\frac{\partial}{\partial\beta}g(\nu_1,\beta) + o(1)$$

as $\alpha \to 0$. Now (2.16) is implied by the definition of $I(\nu)$. □

3. CENTRAL LIMIT THEOREMS

We need some central limit theorems with the special feature that the convergence is uniform on compact subsets of int Θ^*.

Throughout this section, let Θ be an arbitrary compact subset of int Θ^* and let $f(t,\theta) = E_\theta \exp(itX_1)$ be the characteristic function of P_θ. We shall first prove that condition (2.6) implies regularity properties of $f(t,\theta)$:

<u>LEMMA 3.1</u>. *If* (2.6) *holds then a* $k \ge 1$ *exists such that uniformly for* $\theta \in \Theta$

$$(3.1) \qquad \int |f(t,\theta)|^{2k} dt < \infty.$$

If P_0 *has a Lebesgue density then for each* $\delta > 0$ *there is an* $\eta < 1$ *such that*

$$(3.2) \qquad |f(t,\theta)| < \eta \quad \text{for all } |t| > \delta \text{ and } \theta \in \Theta.$$

<u>PROOF</u>. Let k and γ be the constants implied by condition (2.6) such that $p_\theta^{*k}(x) < \gamma$ for all $\theta \in \Theta$ and $x \in \mathbb{R}$. We have

$$\int (p_\theta^{*k}(x))^2 dx < \gamma$$

and (3.1) follows by the Plancherel indentity (cf. Feller (1971), chapter XV). To prove (3.2) assume the contrary, then there is a $\delta > 0$ and there are sequences $\{t_n\},\{\theta_n\}$ with $|t_n| > \delta$ and $\theta_n \in \Theta$ such that $|f(t_n,\theta_n)| \to 1$ as $n \to \infty$. Since Θ is compact assume $\theta_n \to \theta \in \Theta$. A slight elaboration of the proof of Theorem 2.9 in Lehmann (1959) shows that for any sequence $\{t_n\}$, $\theta_n \to \theta \in \text{int } \Theta^*$ implies

$$|f(t_n,\theta_n) - f(t_n,\theta)| \to 0 \quad \text{as } n \to \infty,$$

whence $|f(t_n,\theta)| \to 1$, too. This contradicts the fact that P_θ has a density and that $|t_n| > \delta > 0$. □

Theorems 3.2 and 3.3 concern the distribution function and the density of a weighted sum of X and Y. Let $\{a_N\}$ and $\{b_N\}$ be sequences in $\mathbb{R}$ such that $a_N^2 + b_N^2 \neq 0$. Let $T_N = a_N X + b_N Y$, $\mu_N = \nu a_N \lambda_1 + (1-\nu) b_N \lambda_2$, $\tau_N = \{\nu a_N^2 \sigma_1^2 + (1-\nu) b_N^2 \sigma_2^2\}^{\frac{1}{2}}$, where $\nu = \nu_N$, and let $T_N^* = (T_N - N\mu_N)/(\tau_N\sqrt{N})$.

THEOREM 3.2. *If* P_0 *has a Lebesgue density then uniformly for* $\theta_1,\theta_2 \in \Theta$ *and* $x \in \mathbb{R}$ *as* $N \to \infty$,

$$P_{\theta_1\theta_2}(T_N^* \le x) - \Phi(x) - N^{-\frac{1}{2}} \frac{\rho_N}{6\tau_N^3}(1-x^2)\phi(x) = o(N^{-\frac{1}{2}}), \tag{3.3}$$

where $\rho_N = \nu a_N^3 \rho_1 + (1-\nu) b_N^3 \rho_2$.

PROOF. The proof is completely analogous to that of Theorem XVI.4.1 in Feller (1971). Suppose without loss of generality $\tau_N \equiv 1$, then $|a_N|$ and $|b_N|$ are bounded. Define

$$F_N(x) = P_{\theta_1\theta_2}(T_n^* \le x),$$

$$G_N(x) = \Phi(x) + N^{-\frac{1}{2}} \frac{\rho_N}{6}(1-x^2)\phi(x),$$

and let $f_N(t)$ and $g_N(t)$ be their Fourier transforms. By Esseens smoothing lemma (cf. Feller (1971), Lemma XVI.3.2) we have for arbitrary $\varepsilon > 0$

$$|F_N(x) - G_N(x)| \le \frac{1}{\pi} \int_{-a\sqrt{N}}^{a\sqrt{N}} \left|\frac{f_N(t)-g_N(t)}{t}\right| dt + \frac{\varepsilon}{\sqrt{N}}$$

whenever the constant a is chosen so large that $24|G_N'(x)| < \varepsilon a \pi$ for all x. Note that

$$G_N'(x) = \phi(x)\{1 + N^{-\frac{1}{2}} \frac{\rho_N}{6}(x^3-3x)\}$$

is uniformly bounded for $\theta_1,\theta_2 \in \Theta$ and each N. Next consider

$$|f_N(t) - g_N(t)| = e^{-\frac{1}{2}t^2} \left| e^{Nh_N(\frac{t}{\sqrt{N}})} - 1 - \frac{\rho_N}{6\sqrt{N}}(it)^3 \right|, \tag{3.4}$$

where

$$h_N(t) = \nu \log f_1(a_N t) + (1-\nu) \log f_2(b_N t) + \tfrac{1}{2}t^2 \tag{3.5}$$

and $f_1(t) = E_{\theta_1} \exp\{it(X_1 - \lambda_1)\}$, $f_2(t) = E_{\theta_2} \exp\{it(Y_1 - \lambda_2)\}$.
The right hand side of (3.4) is now estimated using the inequality
$|e^z - 1 - w| \le (|z-w| + \frac{1}{2}|w|^2) \exp(\max(|z|,|w|))$.

Observe that, when δ is small enough, the functions $\{t \mapsto h_N'''(t): N \ge 1, \theta_1, \theta_2 \in \Theta\}$ are equicontinuous on $(-\delta,\delta)$. This follows from the uniform $(N \ge 1,\ \theta_1,\theta_2 \in \Theta, |t| < \delta)$ boundedness of the fourth derivative $h_N^{(4)}(t)$, given by

$$h_N^{(4)}(t) = \nu a_N^4\left[\left(\frac{d}{du}\right)^4 \log f_1(u)\right]_{u=a_Nt} + (1-\nu) b_N^4\left[\left(\frac{d}{du}\right)^4 \log f_2(u)\right]_{u=b_Nt}.$$

The boundedness follows from the boundedness of a_N, b_N, $f_i^{(k)}(t)$, $k \le 4$, $i = 1,2$, and from the fact that $f_i(t)$ is uniformly bounded away from zero on $(-\delta,\delta)$ when δ is sufficiently small. By the equicontinuity shown above there exists a $\delta > 0$ such that for each N and for all $\theta_1,\theta_2 \in \Theta$

$$(3.6) \qquad |h_N'''(t) - h_N'''(0)| < \varepsilon \quad \text{if } |t| < \delta.$$

Since $h_N(0) = h_N'(0) = h_N''(0) = 0$ and $h_N'''(0) = i^3\rho_N$, (3.6) implies

$$(3.7) \qquad |h_N(t) - \frac{1}{6}(it)^3\rho_N| \le \frac{1}{6}\varepsilon|t|^3 \quad \text{if } |t| < \delta.$$

Choose δ small enough to ensure not only (3.7), but also

$$|h_N(t)| < \tfrac{1}{4}t^2 \quad \text{and} \quad |\tfrac{1}{6}\rho_N t^3| < \tfrac{1}{4}t^2 \quad \text{if } |t| < \delta,$$

then it follows for $|t| < \delta\sqrt{N}$ that

$$|f_N(t) - g_N(t)| \le \left(\varepsilon\frac{|t|^3}{\sqrt{N}} + \frac{1}{72}\frac{\rho_N^2}{N}t^6\right)e^{-\frac{1}{4}t^2},$$

whence

$$\int_{-\delta\sqrt{N}}^{\delta\sqrt{N}} \left|\frac{f_N(t)-g_N(t)}{t}\right| dt = O\left(\frac{\varepsilon}{\sqrt{N}}\right).$$

It remains to estimate

$$(3.8) \qquad \int_{|t|\in(\delta\sqrt{N},a\sqrt{N})} \left|\frac{f_N(t)-g_N(t)}{t}\right| dt \le \int_{|t|\in(\delta\sqrt{N},a\sqrt{N})} \left|\frac{f_N(t)}{t}\right| dt + \int_{|t|\in(\delta\sqrt{N},a\sqrt{N})} \left|\frac{g_N(t)}{t}\right| dt.$$

The last integral is exponentially small. Since $\tau_N = 1$ which implies $\max(|a_N|,|b_N|) > d$ for some positive d, the first integral in the right hand side of (3.8) equals

$$\int_{|t|\in(\delta,a)} \left|\frac{f_1^m(a_Nt)f_2^n(b_Nt)}{t}\right| dt \le \eta^{\min(m,n)} \int_{|t|\in(\delta,a)} \left|\frac{1}{t}\right| dt = o(N^{-\frac{1}{2}}),$$

when η is the bound in (3.2) for $|t| > \delta/d$. □

THEOREM 3.3. *If* $\{P_\theta : \theta \in \Theta^*\}$ *satisfies condition* (2.6), *then uniformly for* $\theta_1, \theta_2 \in \Theta$ *and* $x \in \mathbb{R}$ *as* $N \to \infty$,

$$(3.9) \qquad p_N(x) - \phi(x) = o(1),$$

where p_N *is the density of* T_N^*.

PROOF. Again assume $\tau_N \equiv 1$. By Fourier inversion,

$$(3.10) \qquad |p_N(x) - \phi(x)| \le \frac{1}{2\pi} \int \left| f_1^m\left(\frac{a_Nt}{\sqrt{N}}\right) f_2^n\left(\frac{b_Nt}{\sqrt{N}}\right) - e^{-\frac{1}{2}t^2}\right| dt .$$

Define $h_N(t)$ as in (3.5), then uniformly for $\theta_1, \theta_2 \in \Theta$,

$$|h_N(t)| < \tfrac{1}{4}t^2 \quad \text{if } |t| < \delta.$$

By the inequality $|e^z - 1| < e^{|z|}$, this implies

$$(3.11) \qquad \left|e^{Nh_N(\frac{t}{\sqrt{N}})} - 1\right| e^{-\frac{1}{2}t^2} < e^{-\frac{1}{4}t^2} .$$

Since $Nh_N(\frac{t}{\sqrt{N}}) = O(\frac{1}{\sqrt{N}})$ by (3.7), it follows using (3.11) that

$$\int_{|t|<\delta\sqrt{N}} \left|e^{Nh_N(\frac{t}{\sqrt{N}})} - 1\right| e^{-\frac{1}{2}t^2} dt = o(1)$$

by dominated convergence.

To estimate

$$\int_{|t|>\delta\sqrt{N}} \left|f_1^m\left(\frac{a_Nt}{\sqrt{N}}\right) f_2^n\left(\frac{b_Nt}{\sqrt{N}}\right) - e^{-\frac{1}{2}t^2}\right| dt \le$$

$$\int_{|t|>\delta} |f_1^m(a_Nt)f_2^n(b_Nt)|\sqrt{N}\, dt + \int_{|t|>\delta\sqrt{N}} e^{-\frac{1}{2}t^2} dt ,$$

observe that the last integral tends to zero, and that the penultimate one is smaller than

$$\sqrt{N}\,\eta^{\min(m,n)-2k} \int_{|t|>\delta} |f_1^{2k}(a_N t)\cdot f_2^{2k}(b_N t)|\,dt\ ,$$

which tends to zero when k is the number found in Lemma 3.1, since in that case the integrand is the product of an integrable and a bounded function.

□

THEOREM 3.4. *If* P_0 *has a lattice distribution with minimal lattice* $\mathbb{Z}$, *then uniformly for* $\theta_1 \in \Theta$ *and* $x \in \mathbb{Z}$ *as* $m \to \infty$,

$$(3.12)\qquad \sigma_1\sqrt{m}\,P_{\theta_1}(X = x) - \phi\left(\frac{x-m\lambda_1}{\sigma_1\sqrt{m}}\right) = o(1).$$

PROOF. The proof is similar to that of Theorem 3.3, specialized to $a_N \equiv \sigma_1^{-1}$, $b_N \equiv 0$. Instead of (3.10) we get

$$\left|\sigma_1\sqrt{m}\,P_{\theta_1}(X = x) - \phi\left(\frac{x-m\lambda_1}{\sigma_1\sqrt{m}}\right)\right| \le \frac{1}{2\pi}\int_{-\pi\sigma_1\sqrt{m}}^{\pi\sigma_1\sqrt{m}} \left|f_1^m\left(\frac{t}{\sigma_1\sqrt{m}}\right) - e^{-\frac{1}{2}t^2}\right|dt$$
$$+ \int_{|t|>\pi\sigma_1\sqrt{m}} e^{-\frac{1}{2}t^2}\,dt\ .$$

The first integral is estimated as in the previous proof; the second one trivially tends to zero. □

The last theorem of this section is a local limit theorem for the conditional distribution of X given R = r, and partially generalizes a theorem of Hannan and Harkness (1963).

THEOREM 3.5. *Let* $(\tilde\theta_1,\tilde\theta_2)$ *be the solution of the equations*

$$(3.13)\qquad \tilde\theta_1-\tilde\theta_2 = \theta_1-\theta_2,\quad m\lambda(\tilde\theta_1)+n\lambda(\tilde\theta_2) = r,$$

and define $\tilde\lambda_i = \lambda(\tilde\theta_i)$, $\tilde\sigma_i^2 = \sigma^2(\tilde\theta_i)$, $i = 1,2$, *and* $\tilde\sigma^2 = \{(\nu\tilde\sigma_1^2)^{-1} + ((1-\nu)\tilde\sigma_2^2)^{-1}\}^{-1}$.

(i) *If* $\{P_\theta : \theta \in \Theta^*\}$ *satisfies condition* (2.6), *and* $\tilde p_N(x)$ *is the conditional density under* (θ_1,θ_2) *of* $(X-m\tilde\lambda_1)/(\tilde\sigma\sqrt{N})$ *given* R = r, *then uniformly for* $x \in \mathbb{R}$ *and all* θ_1,θ_2,r *such that* $\tilde\theta_1,\tilde\theta_2 \in \Theta$,

$$(3.14)\qquad \tilde p_N(x) - \phi(x) = o(1)\quad \text{as } N \to \infty.$$

(ii) *If* P_0 *is a lattice distribution with minimal lattice* $\mathbb{Z}$, *then uniformly for* $x \in \mathbb{Z}$ *and all* θ_1,θ_2,r *such that* $\tilde\theta_1,\tilde\theta_2 \in \Theta$,

$$\tilde{\sigma}\sqrt{N}\, P_{\theta_1\theta_2}(X = x \mid R = r) - \phi\left(\frac{x-m\tilde{\lambda}_1}{\tilde{\sigma}\sqrt{N}}\right) = o(1) \quad as\ N \to \infty.$$

PROOF. Since the distribution of $X|R = r$ depends on θ_1,θ_2 only through its difference $\theta_1-\theta_2$ we may indeed replace the parameters by $\tilde{\theta}_1$ and $\tilde{\theta}_2$. The proofs of (i) and (ii) above are completely analogous; we shall give here the proof of (i) only.

Let $\tilde{p}^X$, $\tilde{p}^Y$ and $\tilde{p}^R$ be the densities of X, Y and R under $(\tilde{\theta}_1,\tilde{\theta}_2)$. Theorem 3.3 specialized to $a_N \equiv 0$ or $b_N \equiv 0$ implies

$$(3.15)\qquad \left|\tilde{\sigma}_1\tilde{\sigma}_2\sqrt{mn}\,\tilde{p}^X(x)\tilde{p}^Y(r-x) - \phi\left(\frac{x-m\tilde{\lambda}_1}{\tilde{\sigma}_1\sqrt{m}}\right)\phi\left(\frac{r-x-n\tilde{\lambda}_2}{\tilde{\sigma}_2\sqrt{n}}\right)\right| \leq$$

$$\tilde{\sigma}_1\sqrt{m}\,\tilde{p}^X(x)\left|\tilde{\sigma}_2\sqrt{n}\,\tilde{p}^Y(r-x) - \phi\left(\frac{r-x-n\tilde{\lambda}_2}{\tilde{\sigma}_2\sqrt{n}}\right)\right|$$

$$+\phi\left(\frac{r-x-n\tilde{\lambda}_2}{\tilde{\sigma}_2\sqrt{n}}\right)\left|\tilde{\sigma}_1\sqrt{m}\,\tilde{p}^X(x) - \phi\left(\frac{x-m\tilde{\lambda}_1}{\tilde{\sigma}_1\sqrt{m}}\right)\right| = o(1),$$

and by integration we find using the uniformity in x of (3.9) that

$$(3.16)\qquad \int\left\{\tilde{\sigma}_1\tilde{\sigma}_2\sqrt{mn}\,\tilde{p}^X(x)\tilde{p}^Y(r-x) - \phi\left(\frac{x-m\tilde{\lambda}_1}{\tilde{\sigma}_1\sqrt{m}}\right)\phi\left(\frac{r-x-n\tilde{\lambda}_2}{\tilde{\sigma}_2\sqrt{n}}\right)\right\}dx = o(N^{\frac{1}{2}}).$$

In view of (3.13) and the definition of $\tilde{\sigma}$ we have

$$(3.17)\qquad \phi\left(\frac{x-m\tilde{\lambda}_1}{\tilde{\sigma}_1\sqrt{m}}\right)\phi\left(\frac{r-x-n\tilde{\lambda}_2}{\tilde{\sigma}_2\sqrt{n}}\right) = \frac{1}{\sqrt{2\pi}}\phi\left(\frac{x-m\tilde{\lambda}_1}{\tilde{\sigma}\sqrt{N}}\right)$$

thus (3.15) implies

$$(3.18)\qquad \tilde{\sigma}_1\tilde{\sigma}_2\sqrt{mn}\,\tilde{p}^X(x)\tilde{p}^Y(r-x) = \frac{1}{\sqrt{2\pi}}\phi\left(\frac{x-m\tilde{\lambda}_1}{\tilde{\sigma}\sqrt{N}}\right) + o(1)$$

and from (3.16) it follows that

$$(3.19)\qquad \tilde{\sigma}_1\tilde{\sigma}_2\sqrt{mn}\,\tilde{p}^R(r) = \frac{\tilde{\sigma}\sqrt{N}}{\sqrt{2\pi}} + o(N^{\frac{1}{2}}).$$

Let $p^{X|R=r}_{\theta_1\theta_2}(x)$ be the conditional density of X given R = r under (θ_1,θ_2), then by (3.18) and (3.19),

$$\tilde{p}_N(x) = \tilde{\sigma}\sqrt{N}\,p^{X|R=r}_{\theta_1\theta_2}(\tilde{\sigma}\sqrt{N}\cdot x + m\tilde{\lambda}_1)$$

$$= \tilde{\sigma}\sqrt{N}\,\tilde{p}^X(\tilde{\sigma}\sqrt{N}\cdot x + m\tilde{\lambda}_1)\cdot\tilde{p}^Y(r-\tilde{\sigma}\sqrt{N}\cdot x - m\tilde{\lambda}_1)/\tilde{p}^R(r) = \phi(x) + o(1). \quad \square$$

4. A LOWER BOUND FOR THE SIZE OF THE MOST POWERFUL TEST

Let δ_+ be the MP test of $H = \{(\theta,\theta) : \theta \in \Theta^*\}$ against $K = \{(\theta_1,\theta_2)\}$ with given power β in K and let $\alpha_+(N,\beta)$ be its size. Though δ_+ exists, it is in general not explicitly known - it involves a least favourable distribution, cf. Lehmann (1959), Section 3.8 -; we shall be satisfied with a (lower) bound for $\alpha_+(N,\beta)$. We find this bound by considering for each $\theta \in \Theta^*$ the MP test δ_θ of $\{(\theta,\theta)\}$ against K which has power β in K. Since $\{(\theta,\theta)\} \subset H$, the size $\alpha_\theta(N,\beta)$ of δ_θ is at most equal to $\alpha_+(N,\beta)$; a lower bound is found by maximizing $\alpha_\theta(N,\beta)$ over $\theta \in \Theta^*$.

By the fundamental lemma of Neyman and Pearson, the test δ_θ is given by

$$(4.1) \qquad \delta_\theta(x,y) = \begin{cases} 1 & > \\ \gamma_\theta \text{ when } t_\theta(x,y) & = c_\theta \\ 0 & < \end{cases}$$

where

$$t_\theta(x,y) = (\theta_1-\theta)x + (\theta_2-\theta)y$$

is an increasing function of the likelihood ratio

$$dP^{X,Y}_{\theta_1\theta_2}(x,y) \,/\, dP^{X,Y}_{\theta\theta}(x,y),$$

and c_θ and γ_θ satisfy $0 < \gamma_\theta \leq 1$ and

$$(4.2) \qquad \gamma_\theta P_{\theta_1\theta_2}(t_\theta(X,Y) = c_\theta) + P_{\theta_1\theta_2}(t_\theta(X,Y) > c_\theta) = \beta.$$

Before stating the theorem of this section we introduce more notation:

$$(4.3) \qquad \begin{cases} T_\theta = t_\theta(X,Y), \\ \mu_{\nu\theta} = N^{-1}E_{\theta_1\theta_2}T_\theta = \nu(\theta_1-\theta)\lambda_1 + (1-\nu)(\theta_2-\theta)\lambda_2, \\ \tau^2_{\nu\theta} = N^{-1}\mathrm{var}_{\theta_1\theta_2}T_\theta = \nu(\theta_1-\theta)^2\sigma_1^2 + (1-\nu)(\theta_2-\theta)^2\sigma_2^2, \\ \rho_{\nu\theta} = N^{-1}E_{\theta_1\theta_2}(T_\theta - N\mu_{\nu\theta})^3 = \nu(\theta_1-\theta)^3\rho_1 + (1-\nu)(\theta_2-\theta)^3\rho_2. \end{cases}$$

Moreover, let θ_ν be defined in (2.3) and abbreviate

$$(4.4) \qquad \mu_\nu = \mu_{\nu\theta_\nu}, \quad \tau_\nu = \tau_{\nu\theta_\nu}, \quad \rho_\nu = \rho_{\nu\theta_\nu}.$$

Finally define

$u_\beta = \Phi^{-1}(\beta)$.

<u>THEOREM 4.1</u>. *Let* $\varepsilon > 0$, *let* Θ *be a convex compact subset of* int Θ^*, *and let* $\alpha_N(\beta)$ *be the size of the MP test of* $\{(\theta_N,\theta_N)\}$ *against* $\{(\theta_1,\theta_2)\}$ *with power* β *in* (θ_1,θ_2), *where*

$$(4.5)\qquad \theta_N = \theta_\nu + N^{-\frac{1}{2}} \frac{u_\beta}{\sigma_\nu^2} \frac{\partial \tau_{\nu\theta}}{\partial\theta}\Big]_{\theta_\nu}$$

then, for each $\beta \in (0,1)$ *and each* N *with* $\theta_N \in \Theta^*$,

$$(4.6)\qquad \alpha_+(N,\beta) \geq \alpha_N(\beta).$$

If moreover, $\{P_\theta : \theta \in \Theta^*\}$ *satisfies condition* (2.6), *we have uniformly for* $\beta \in [\varepsilon, 1-\varepsilon]$ *and* $\theta_1,\theta_2 \in \Theta$ *as* $N \to \infty$,

$$(4.7)\qquad \alpha_N(\beta) = N^{-\frac{1}{2}} \exp\{-NI_\nu(\theta_1,\theta_2,\theta_\nu) - N^{\frac{1}{2}}\tau_\nu c_N^* - E_{1\nu} + o(1)\},$$

where c_N^* *is defined by*

$$(4.8)\qquad P_{\theta_1\theta_2}((T_{\theta_\nu} - N\mu_\nu)/(\tau_\nu\sqrt{N}) \geq c_N^*) = \beta,$$

and

$$(4.9)\qquad E_{1\nu} = \tfrac{1}{2}u_\beta^2 + \tfrac{1}{2}\log 2\pi\tau_\nu^2 - \tfrac{1}{2}\frac{u_\beta}{\sigma_\nu^2}\left(\frac{\partial\tau_{\nu\theta}}{\partial\theta}\Big]_{\theta_\nu}\right)^2 .$$

More generally, when $\{P_\theta : \theta \in \Theta^*\}$ *does not necessarily satisfy condition* (2.6), *then, uniformly for* $\beta \in [\varepsilon, 1-\varepsilon]$ *and* $\theta_1,\theta_2 \in \Theta$ *as* $N \to \infty$,

$$(4.10)\qquad \alpha_N(\beta) = N^{-\frac{1}{2}} \exp\{-NI_\nu(\theta_1,\theta_2,\theta_\nu) + N^{\frac{1}{2}}\tau_\nu u_\beta + O(1)\}.$$

<u>REMARK 4.1</u>. For the validity of (4.10), $\theta_N = \theta_\nu + O(N^{-\frac{1}{2}})$ is sufficient.

<u>PROOF of Theorem 4.1</u>. Statement (4.6) was proved in the introduction to this section. We shall prove (4.7) and explain the special choice (4.5) of θ_N, under the assumption that $\{P_\theta : \theta \in \Theta^*\}$ satisfies condition (2.6). Let θ be arbitrary in Θ, possibly dependent on N. Since T_θ has a density, the critical value c_θ of the test δ_θ is found from

$$(4.11)\qquad P_{\theta_1\theta_2}(T_\theta < c_\theta) = 1-\beta.$$

When c_θ^* is defined by

(4.12) $$c_\theta^* = (c_\theta - N\mu_{\nu\theta})/(\tau_{\nu\theta}\sqrt{N}),$$

(4.11) implies by Theorem 3.2 that as $N \to \infty$

(4.13) $$c_\theta^* = -u_\beta - N^{-\frac{1}{2}} \frac{\rho_{\nu\theta}}{6\tau_{\nu\theta}^3}(1-u_\beta^2) + o(N^{-\frac{1}{2}}).$$

Now $\alpha_\theta(N,\beta)$ is determined by exponential centering (cf. also Lemma I.2.6):

(4.14) $$\alpha_\theta(N,\beta) = P_{\theta\,\theta}(T_\theta \geq c_\theta) = \iint_{\{(x,y):t_\theta(x,y)\geq c_\theta\}} dP_{\theta\,\theta}^{X,Y}(x,y)$$

$$= \iint_{\{t_\theta \geq c_\theta\}} \exp\{-(\theta_1-\theta)x - (\theta_2-\theta)y + m\psi(\theta_1) + n\psi(\theta_2) - N\psi(\theta)\}\, dP_{\theta_1\theta_2}^{X,Y}(x,y)$$

$$= \iint_{\{t_\theta \geq c_\theta\}} \exp\{-NI_\nu(\theta_1,\theta_2,\theta) - t_\theta(x,y) + N\mu_{\nu\theta}\} dP_{\theta_1\theta_2}^{X,Y}(x,y)$$

$$= \exp\{-NI_\nu(\theta_1,\theta_2,\theta) + N\mu_{\nu\theta} - c_\theta\} \int_{[c_\theta,\infty)} e^{-(t-c_\theta)} dP_{\theta_1\theta_2}^{T_\theta}(t).$$

We use Theorem 3.3 with $a_N = \theta_1-\theta$ and $b_N = \theta_2-\theta$, and the dominated convergence theorem to see that

(4.15) $$\int_{[c_\theta,\infty)} e^{-(t-c_\theta)} dP_{\theta_1\theta_2}^{T_\theta}(t) = \int_{c_\theta}^{\infty} e^{-(t-c_\theta)} \frac{1}{\tau_{\nu\theta}\sqrt{N}} p_N\left(\frac{t-N\mu_{\nu\theta}}{\tau_{\nu\theta}\sqrt{N}}\right) dt$$

$$= \frac{1}{\tau_{\nu\theta}\sqrt{N}} \int_0^\infty e^{-z} p_N\left(c_\theta^* + \frac{z}{\tau_{\nu\theta}\sqrt{N}}\right) dz$$

$$= \frac{1}{\tau_{\nu\theta}\sqrt{N}} \int_0^\infty e^{-z}\phi(c_\theta^*)dz + o(N^{-\frac{1}{2}}),$$

where p_N is the density of $(T_\theta - N\mu_{\nu\theta})/(\tau_{\nu\theta}\sqrt{N})$.

Combine (4.14), (4.15) and (4.12) to get

(4.16) $$\alpha_\theta(N,\beta) = N^{-\frac{1}{2}} \exp\{-NI_\nu(\theta_1,\theta_2,\theta) - N^{\frac{1}{2}} c_\theta^* \tau_{\nu\theta} - \tfrac{1}{2}\log 2\pi\tau_{\nu\theta}^2 - \tfrac{1}{2}c_\theta^{*2} + o(1)\}.$$

Insertion of (4.5) in (4.16) yields (4.7) by means of a second order Taylor expansion. The choice (4.5) of $\{\theta_N\}$ is motivated by the fact that it minimizes the leading terms of $-\log \alpha_\theta(N,\beta)$ up to $o(1)$: we have, using (4.16) and (4.13)

$$\frac{d}{d\theta}\{NI_\nu(\theta_1,\theta_2,\theta) - N^{\frac{1}{2}}\tau_{\nu\theta}u_\beta + \tfrac{1}{2}\log 2\pi\tau_{\nu\theta}^2 - \frac{\rho_{\nu\theta}}{6\tau_{\nu\theta}^2}(1-u_\beta^2) + u_\beta^2\}$$

$$= -m\lambda(\theta_1) - n\lambda(\theta_2) + N\lambda(\theta) - N^{\frac{1}{2}}u_\beta \frac{\partial\tau_{\nu\theta}}{\partial\theta} + \mathcal{O}(1).$$

Equating this to zero yields $\theta = \theta_N + \mathcal{O}(N^{-1})$.

Now we prove (4.10) in the absence of the regularity properties of Lemma 3.1. In view of (4.2), c_θ satisfies

$$P_{\theta_1\theta_2}(T_\theta > c_\theta) \le \beta \le P_{\theta_1\theta_2}(T_\theta \ge c_\theta).$$

The Berry-Esseen theorem (cf. Feller (1971), XVI.5) shows that uniformly for $\theta,\theta_1,\theta_2 \in \Theta$ and $\beta \in [\varepsilon, 1-\varepsilon]$,

$$c_\theta^* = -u_\beta + \mathcal{O}(N^{-\frac{1}{2}}).$$

(c_θ^* is defined in (4.12)). Proceed as in the previous case and arrive at

$$\text{(4.17)} \qquad \alpha_\theta(N,\beta) = \exp\{-NI_\nu(\theta_1,\theta_2,\theta) - N^{\frac{1}{2}}c_\theta^*\tau_{\nu\theta}\} \cdot \int \delta_\theta(t)e^{-(t-c_\theta)}\,dP_{\theta_1\theta_2}^{T_\theta}(t).$$

It remains to evaluate the integral. By (4.1) we have

$$\text{(4.18)} \qquad \int_{(c_\theta,\infty)} e^{-(t-c_\theta)}\,dP(t) \le \int \delta_\theta(t)e^{-(t-c_\theta)}\,dP(t) \le \int_{[c_\theta,\infty)} e^{-(t-c_\theta)}\,dP(t),$$

denoting $P_{\theta_1\theta_2}^{T_\theta}(t)$ briefly as $P(t)$. Let b_1 be a uniform Berry-Esseen constant, i.e.

$$\text{(4.19)} \qquad \left|P(T_\theta \le t) - \Phi\left(\frac{t-N\mu_{\nu\theta}}{\tau_{\nu\theta}\sqrt{N}}\right)\right| < b_1N^{-\frac{1}{2}}$$

for each $t \in \mathbb{R}$, each N, and all $\theta,\theta_1,\theta_2 \in \Theta$, and let b_2 be a constant such that for all $\theta,\theta_1,\theta_2 \in \Theta$,

$$b_2 > \frac{(2b_1+1)\tau_{\nu\theta}}{\phi(-u_\varepsilon+1)}$$

(ε is the bound for β in the conditions of the theorem).

The left hand member of (4.18) is larger than

$$(4.20)\qquad \int_{(c_\theta, c_\theta+b_2]} e^{-(t-c_\theta)}\,dP(t) \ge e^{-b_2} P(c_\theta < T_\theta \le c_\theta + b_2)$$

$$\ge e^{-b_2}\left(\Phi\left(\frac{c_\theta+b_2-N\mu_{\nu\theta}}{\tau_{\nu\theta}\sqrt{N}}\right) - \Phi\left(\frac{c_\theta-N\mu_{\nu\theta}}{\tau_{\nu\theta}\sqrt{N}}\right) - 2b_1 N^{-\frac{1}{2}}\right)$$

$$= e^{-b_2}(\Phi(c_\theta^* + b_2/(\tau_{\nu\theta}\sqrt{N})) - \Phi(c_\theta^*) - 2b_1 N^{-\frac{1}{2}}).$$

For sufficiently large N, i.e. when $b_2/(\tau_{\nu\theta}\sqrt{N}) < \frac{1}{2}$ and $|c_\theta^*| < |u_\beta| + \frac{1}{2}$, the last member of (4.20) is larger than

$$(4.21)\qquad e^{-b_2}\left(\frac{b_2}{\tau_{\nu\theta}\sqrt{N}}\,\phi(-u_\varepsilon+1) - 2\frac{b_1}{\sqrt{N}}\right) \ge e^{-b_2} N^{-\frac{1}{2}}.$$

On the other hand (cf. (4.18)) we have

$$(4.22)\qquad \int_{[c_\theta,\infty)} e^{-(t-c_\theta)}\,dP(t) = \sum_{j=0}^{\infty} \int_{[c_\theta+j, c_\theta+j+1)} e^{-(t-c_\theta)}\,dP(t)$$

$$\le \sum_{j=0}^{\infty} e^{-j} P(c_\theta+j \le T_\theta < c_\theta+j+1)$$

$$\le \sum_{j=0}^{\infty} e^{-j}\left\{\Phi\left(c_\theta^* + \frac{j+1}{\tau_{\nu\theta}\sqrt{N}}\right) - \Phi\left(c_\theta^* + \frac{j}{\tau_{\nu\theta}\sqrt{N}}\right) + 2b_1 N^{-\frac{1}{2}}\right\}$$

$$\le \frac{1}{1-e^{-1}}\left(\frac{1}{\tau_{\nu\theta}\sqrt{2\pi}} + 2b_1\right) N^{-\frac{1}{2}}.$$

The conclusion from (4.18) - (4.22) is that the integral in (4.17) is $N^{-\frac{1}{2}} \exp\{O(1)\}$ as $N \to \infty$. The choice $\theta = \theta_N$ now maximizes the leading terms in (4.17) up to $\exp\{O(1)\}$. □

REMARK 4.2. When P_θ is the normal distribution with mean θ and unit variance, it can be shown that $\delta_+ = \delta_{\theta_\nu}$. Since in this case also $(\partial\tau_{\nu\theta}/\partial\theta)_{\theta_\nu} = 0$, equality holds in (4.6). The same equality also holds for the gamma scale family with known shape parameter, cf. Section 1.

5. THE CONDITIONAL PROBABILITY OF A LARGE DEVIATION

Since the conditional test is defined by means of the conditional probabilities in (1.3), we evaluate in this section, for general γ, r and k, $\gamma P_{\theta\,\theta}(X = k \mid R = r) + P_{\theta\,\theta}(X > k \mid R = r)$. Let Ω be an arbitrary compact subset of int H_1.

<u>THEOREM 5.1</u>. *Suppose* $\{k_N\}$, $\{r_N\}$ *are sequences such that* $\lambda^{-1}(k_N/m)$ *and* $\lambda^{-1}((r_N-k_N)/n)$ *exist and* $(\lambda^{-1}(k_N/m),\ \lambda^{-1}((r_N-k_N)/n)) \in \Omega$ *for each* N. *Denote* $\eta_1 = \lambda^{-1}(k_N/m)$, $\eta_2 = \lambda^{-1}((r_N-k_N)/n)$ *and* $\eta = \lambda^{-1}(r_N/N)$.

(i) *If* $\{P_\theta : \theta \in \Theta^*\}$ *satisfies condition* (2.6), *then as* $N \to \infty$

$$(5.1)\qquad P_{\theta\,\theta}(X \geq k_N \mid R = r_N) = $$
$$= N^{-\frac{1}{2}} \exp\{-NI_\nu(\eta_1,\eta_2,\eta) - E_{2\nu}(\eta_1,\eta_2,\eta) + o(1)\},$$

where

$$(5.2)\qquad E_{2\nu}(\eta_1,\eta_2,\eta) = \tfrac{1}{2}\log 2\pi\nu(1-\nu)(\eta_1-\eta_2)^2 \frac{\sigma^2(\eta_1)\sigma^2(\eta_2)}{\sigma^2(\eta)} .$$

(ii) *If* P_0 *is a lattice distribution with minimal lattice* $\mathbb{Z}$ *and* $\gamma_N \in [0,1)$, *then as* $N \to \infty$

$$(5.3)\qquad \gamma_N P_{\theta\,\theta}(X = k_N \mid R = r_N) + P_{\theta\,\theta}(X > k_N \mid R = r_N)$$
$$= N^{-\frac{1}{2}} \exp\{-NI(\eta_1,\eta_2,\eta) - \tfrac{1}{2}\log 2\pi\nu(1-\nu)\frac{\sigma^2(\eta_1)\sigma^2(\eta_2)}{\sigma^2(\eta)}$$
$$+ \log\left(\gamma_N + \frac{e^{-(\eta_1-\eta_2)}}{1-e^{-(\eta_1-\eta_2)}}\right) + o(1)\}.$$

<u>REMARK 5.1</u>. Note that the $o(1)$ terms in (5.1) and (5.3) are uniform in $\{\gamma_N\}$, $\{r_N\}$ and $\{k_N\}$, provided $(\eta_1,\eta_2) \in \Omega$.

<u>REMARK 5.2</u>. Statement (ii) of Theorem 5.1 is stronger than necessary; we shall use it with the two log terms replaced by $O(1)$.

<u>PROOF of Theorem 5.1</u>. First observe that there exists a convex compact subset Θ of int Θ^*, such that $(\eta_1,\eta_2) \in \Omega$ implies η_1, η_2 and $\eta \in \Theta$. We prove (5.3); the proof of (5.1) is completely analogous. Since the left hand side in (5.3) is independent of θ, replace θ by $\eta = \lambda^{-1}(r_N/N)$. For $x \in \mathbb{Z}$ we have

$$(5.4)\qquad P_{\eta\,\eta}(X = x \mid R = r_N) = P_\eta(X = x)\cdot P_\eta(Y = r_N - x)/P_{\eta\,\eta}(R = r_N) .$$

By the exponential family properties, this equals

$$\exp\{-(\eta_1-\eta)x + m\psi(\eta_1) - (\eta_2-\eta)(r_N-x) + n\psi(\eta_2) - N\psi(\eta)\}$$
$$\cdot P_{\eta_1}(X = x)P_{\eta_2}(Y = r_N - x)/P_{\eta\,\eta}(R = r_N)$$

$$= \exp\{-NI_\nu(\eta_1,\eta_2,\eta)\}\exp\{-(\eta_1-\eta_2)(x-k_N)\}P_{\eta_1\eta_2}(X=x \mid R=r_N)$$

$$\cdot P_{\eta_1\eta_2}(R=r_N)/P_{\eta\eta}(R=r_N),$$

hence the left hand side in (5.3) equals

$$\exp\{-NI_\nu(\eta_1,\eta_2,\eta)\}\{P_{\eta_1\eta_2}(R=r_N)/P_{\eta\eta}(R=r_N)\} \tag{5.5}$$

$$\cdot\Big[\gamma_N P_{\eta_1\eta_2}(X=k_N \mid R=r_N) + \sum_{x=k_N+1}^{\infty}\exp\{-(\eta_1-\eta_2)(x-k_N)\}$$

$$\cdot P_{\eta_1\eta_2}(X=x \mid R=r_N)\Big] .$$

Now Theorem 3.5 is applied: uniformly in x and r_N we have as $N \to \infty$

$$P_{\eta_1\eta_2}(X=x \mid R=r_N) = (\sigma_{\eta_1\eta_2}\sqrt{N})^{-1}\,\phi\left(\frac{x-k_N}{\sigma_{\eta_1\eta_2}\sqrt{N}}\right) + o(N^{-\frac{1}{2}})$$

where $\sigma_{\eta_1\eta_2} = \{(\nu\sigma^2(\eta_1))^{-1} + ((1-\nu)\sigma^2(\eta_2))^{-1}\}^{-\frac{1}{2}}$, whence the last factor in (5.5) equals by dominated convergence as $N \to \infty$

$$(\sigma_{\eta_1\eta_2}\sqrt{N})^{-1}\Big[\gamma_N\phi(0) + \sum_{j=1}^{\infty} e^{-(\eta_1-\eta_2)\cdot j}\phi(0) + o(1)\Big] \tag{5.6}$$

$$= (\sigma_{\eta_1\eta_2}\sqrt{2\pi N})^{-1}\Big[\gamma_N + \frac{e^{-(\eta_1-\eta_2)}}{1-e^{-(\eta_1-\eta_2)}} + o(1)\Big] .$$

Finally, the lattice-analogon of (3.19) implies as $N \to \infty$

$$\frac{P_{\eta_1\eta_2}(R=r_N)}{P_{\eta\eta}(R=r_N)} = \frac{\sigma_{\eta_1\eta_2}/\sigma(\eta_1)\sigma(\eta_2)}{\sigma_{\eta\eta}/\sigma^2(\eta)} + o(1), \tag{5.7}$$

which completes the proof, since the denominator in (5.7) equals $\sigma(\eta)/\sqrt{\nu(1-\nu)}$. □

6. THE POWER OF THE CONDITIONAL TEST

In this section, the main result (Theorems 2.1 and 2.2) will be proved. In view of Lemma 2.3 and Theorem 4.1 it is sufficient to prove that the power of the conditional test is large enough, see Lemmas 6.1 and 6.2 below.

Let Ω be an arbitrary compact subset of int $\{(\theta_1,\theta_2) \in \Theta^* \times \Theta^*: \theta_1 > \theta_2\}$ and let $\varepsilon > 0$.

<u>LEMMA 6.1</u>. *If condition* (2.5) *or* (2.6) *holds then, uniformly for* $\beta \in [\varepsilon, 1-\varepsilon]$ *and* $(\theta_1, \theta_2) \in \Omega$, *we have*

(6.1) $$\beta_c(N, \alpha_N(\beta)) = \beta + O(N^{-\frac{1}{2}}) \quad \textit{as } N \to \infty,$$

where $\alpha_N(\beta)$ *is defined in* Theorem 4.1.

<u>LEMMA 6.2</u>. *If* (2.6) *holds then, uniformly for* $\beta \in [\varepsilon, 1-\varepsilon]$ *and* $(\theta_1, \theta_2) \in \Omega$ *as* $N \to \infty$,

(6.2) $$\beta_c(N, \alpha_N(\beta)) = \beta - \tfrac{1}{2} N^{-\frac{1}{2}} \varphi(u_\beta) \tau_\nu^{-1} (R_\nu - 1 - \log R_\nu) + o(N^{-\frac{1}{2}}),$$

where τ_ν *is defined in* (4.4) *and* R_ν *in* (2.11).

Note that Lemma 6.1 is partly implied by Lemma 6.2. We shall first prove Lemma 6.2 and subsequently, by adaptation of the proof, the remaining part of Lemma 6.1. In these proofs, we omit the index ν of I_ν, θ_ν, σ_ν, $E_{1\nu}$, $E_{1\nu}$ and τ_ν, and write c^* for c_N^*.

<u>PROOF of Lemma 6.2</u>. We shall determine $\beta_c(N, \alpha_N(\beta))$ by integration of the joint density of (X,Y) over the critical region A_N of the conditional size-$\alpha_N(\beta)$ test, defined by, cf. (1.2),(1,3),

(6.3) $$A_N = \{(x,y) : P_{\theta\theta}(X \geq x \mid R = x+y) \leq \alpha_N(\beta)\}.$$

Note that the conditional probability above was found in Section 5 and $\alpha_N(\beta)$ in Section 4. Fix $\delta \in (0, \frac{1}{6})$ and define

(6.4) $$B_N = \{(x,y) : |x - m\lambda_1| \leq \sigma_1 \sqrt{m} \cdot N^{\delta}, \ |y - n\lambda_2| \leq \sigma_2 \sqrt{n} \cdot N^{\delta}\}$$

then $P_{\theta_1\theta_2}((X,Y) \notin B_N) = o(N^{-\frac{1}{2}})$, cf. Theorem XVI.7.1 of Feller (1971), implying as $N \to \infty$

(6.5) $$\beta_c(N, \alpha_N(\beta)) - P_{\theta_1\theta_2}((X,Y) \in A_N \cap B_N) = o(N^{-\frac{1}{2}}).$$

Let Ω' be another compact subset of int H_1 such that $\Omega \subset$ int Ω' then for all sufficiently large N, $(\theta_1, \theta_2) \in \Omega$ and $(x,y) \in B_N$ imply $(\lambda^{-1}(x/m), \lambda^{-1}(y/n)) \in \Omega'$.

By (6.3), (4.7) and (5.1) there exists a sequence $\{d_{1N}\}$ with $d_{1N} \to 0$ as $N \to \infty$ such that for all $(x,y) \in B_N$

$$(6.6)\qquad NI(\lambda^{-1}(x/m),\lambda^{-1}(y/n),\lambda^{-1}((x+y)/N) + \tfrac{1}{2}\log N + E_2(\lambda^{-1}(x/m),\lambda^{-1}(y/n),\lambda^{-1}((x+y)/N)) - d_{1N} \ge NI(\theta_1,\theta_2,\theta) + N^{\frac{1}{2}}\tau c^* + \tfrac{1}{2}\log N + E_1$$

implies $(x,y) \in A_N$. By Taylor expansion there is a sequence $d_{2N} \to 0$ such that (6.6) is implied by

$$(6.7)\qquad (x-m\lambda_1)(\theta_1-\theta) + (y-n\lambda_2)(\theta_2-\theta) + N^{-1}\Big\{\tfrac{1}{2}(x-m\lambda_1)^2\Big(\frac{1}{\nu\sigma_1^2} - \frac{1}{\sigma^2}\Big) + \tfrac{1}{2}(y-n\lambda_2)^2\Big(\frac{1}{(1-\nu)\sigma_2^2} - \frac{1}{\sigma^2}\Big) - (x-m\lambda_1)(y-n\lambda_2)\frac{1}{\sigma^2}\Big\} - d_{2N} \ge N^{\frac{1}{2}}\tau c^* + E_1 - E_2(\theta_1,\theta_2,\theta).$$

Now write

$$(6.8)\qquad \begin{cases} E = E_1 - E_2(\theta_1,\theta_2,\theta), \\ u = (x-m\lambda_1)/(\sigma_1\sqrt{m}), \quad v = (y-n\lambda_2)/(\sigma_2\sqrt{n}), \\ \tau_1 = (\theta_1-\theta)\sigma_1\sqrt{\nu}, \quad \tau_2 = (\theta_2-\theta)\sigma_2\sqrt{1-\nu} \end{cases}$$

(by (4.3) and (4.4) $\tau_1^2 + \tau_2^2 = \tau^2$).

In view of (6.6) - (6.8) there is a sequence $d_{3N} \to 0$ such that $(x,y) \in A_N \cap B_N$ is implied by

$$(6.9)\qquad \tau_1 u + \tau_2 v - \tau c^* \ge N^{-\frac{1}{2}}\Big\{E - \tfrac{1}{2}\Big(1-\nu\frac{\sigma_1^2}{\sigma^2}\Big)u^2 - \tfrac{1}{2}\Big(1-(1-\nu)\frac{\sigma_2^2}{\sigma^2}\Big)v^2 + \sqrt{\nu(1-\nu)}\,\frac{\sigma_1\sigma_2}{\sigma^2}uv + d_{3N}\Big\},$$

and which also satisfies $N^{-\frac{1}{2}+3\delta}/d_{3N} \to 0$.

Write (6.9) more briefly as

$$(6.10)\qquad \tau_1 u + \tau_2 v - \tau c^* \ge N^{-\frac{1}{2}}\{E - Au^2 - Bv^2 + Cuv + d_{3N}\}$$

and consider also the inequality

$$(6.11)\qquad \tau_1 u + \tau_2 v - \tau c^* \ge N^{-\frac{1}{2}}\Big\{E - A\Big(\frac{\tau c^*-\tau_2 v}{\tau_1}\Big)^2 - Bv^2 + C\Big(\frac{\tau c^*-\tau_2 v}{\tau_1}\Big)v + 2d_{3N}\Big\}.$$

For all $(x,y) \in \mathcal{B}_N$ (or $|u|,|v| < N^{\delta}$), (6.11) implies (6.10) when N is large enough. To see this, let γ_N be a uniform ($|u|,|v| < N^{\delta}$) upper bound for the absolute values of the expressions in curly brackets in (6.10) and (6.11), with $\gamma_N = \mathcal{O}(N^{2\delta})$ as $N \to \infty$. Now $|u|,|v| < N^{\delta}$, (6.11) and $|\tau_1 u + \tau_2 v - \tau c^*| > \gamma_N N^{-\frac{1}{2}}$ imply $\tau_1 u + \tau_2 v - \tau c^* > \gamma_N N^{-\frac{1}{2}}$, hence (6.10) holds. On the other hand, when $|u|,|v| < N^{\delta}$ and $|\tau_1 u + \tau_2 v - \tau c^*| \le \gamma_N N^{-\frac{1}{2}}$, it follows by $u = (\tau c^* - \tau_2 v)/\tau_1 - \mathcal{O}(N^{-\frac{1}{2}+2\delta})$ and $N^{-\frac{1}{2}+3\delta} = o(d_{3N})$, that the negation of (6.10) implies the negation of (6.11) when N is large enough. Similarly, one may also prove that $(x,y) \in A_N \cap \mathcal{B}_N$ implies (6.11), provided d_{3N} is chosen appropriately. We conclude that (rearrange (6.11) in powers of v):

(6.12)
$$(x,y) \in A_N \cap \mathcal{B}_N$$
$$\Leftrightarrow$$
$$\tau_1 u + \tau_2 v - \tau c^* \ge N^{-\frac{1}{2}}\Big\{E - A\left(\frac{\tau c^*}{\tau_1}\right)^2 + \frac{\tau c^*}{\tau_1^2}(2\tau_2 A + \tau_1 C)v - \left(\left(\frac{\tau_2}{\tau_1}\right)^2 A + B + \frac{\tau_2}{\tau_1}C\right)v^2 + o(1)\Big\}, \qquad |u|,|v| < N^{\delta}.$$

Let the variables U and V be defined by

(6.13)
$$U = (X - m\lambda_1)/(\sigma_1\sqrt{m}), \qquad V = (Y - n\lambda_2)/(\sigma_2\sqrt{n}),$$

let $F_1(u) = P_{\theta_1}(U \le u)$, $F_2(v) = P_{\theta_2}(V \le v)$, let p_1 and p_2 be the densities of F_1 and F_2, and write $c(v) = \tau_1^{-1}(\tau c^* - \tau_2 v)$. By (6.12) we have

(6.14)
$$P_{\theta_1\theta_2}((X,Y) \in A_N \cap \mathcal{B}_N) =$$
$$P_{\theta_1\theta_2}(U \ge c(V) + N^{-\frac{1}{2}}\{\tau_1^{-1}(E^1 + A^1 V - B^1 V^2)\} + o(N^{-\frac{1}{2}}),\ |U|,|V| < N^{\delta})$$
$$= \int 1 - F_1(c(v) + N^{-\frac{1}{2}}\{\tau_1^{-1}(E^1 + A^1 v - B^1 v^2)\} + o(N^{-\frac{1}{2}}))\cdot p_2(v)dv + o(N^{-\frac{1}{2}}),$$

where E^1, A^1 and B^1 abbreviate the constant and the coefficients of v and v^2 in the expression in curly brackets in (6.12). The last $o(N^{-\frac{1}{2}})$ term in (6.14) is caused by integration outside $\mathcal{B}_N$ (see (6.4)).

The integral in (6.14) is evaluated using the mean value theorem; abbreviating the coefficient of $N^{-\frac{1}{2}}$ in (6.14) as Q(v) we get

$$(6.15)\qquad \int 1 - F_1(c(v) + N^{-\frac{1}{2}}Q(v) + o(N^{-\frac{1}{2}}))p_2(v)dv$$

$$= \int 1 - F_1(c(v))p_2(v)dv$$

$$- N^{-\frac{1}{2}} \int Q(v)p_1(c(v) + \xi_N(v))p_2(v)dv + o(N^{-\frac{1}{2}})$$

for $|\xi_N(v)| < |N^{-\frac{1}{2}}Q(v) + o(N^{-\frac{1}{2}})|$. The first integral in the right hand side of (6.15) equals β by the definitions of $c(\cdot)$, c^*, U, V, τ_1 and τ_2 (cf. (4.8), (6.13) and (6.8)), the second one equals

$$(6.16)\qquad \int Q(v)\phi(c(v))\phi(v)dv + \int Q(v)\phi(c(v))(p_2(v) - \phi(v))dv$$

$$+ \int Q(v)\{p_1(c(v) + \xi_N(v)) - \phi(c(v))\}p_2(v)dv.$$

The second and third integrals in (6.16) tend to zero since they are dominated by

$$(6.17)\qquad \sup |p_2(v) - \phi(v)| \int |Q(v)|\phi(c(v))dv$$

$$+ \sup |p_1(c(v) - \xi_N(v)) - \phi(c(v))| \int |Q(v)|p_2(v)dv,$$

where the suprema are taken over $v \in \mathbb{R}$, $\theta_1, \theta_2 \in \Theta$ and $\beta \in [\varepsilon, 1-\varepsilon]$. These suprema tend to zero by Theorem 3.3 since $\xi_N(v) \to 0$. The integrals in (6.17) are uniformly bounded since all moments of P_θ are bounded for $\theta \in \Theta$.

It remains to evaluate the first integral in (6.16).
Since

$$\phi(c(v))\phi(v) = \phi(\tau_1^{-1}(\tau c^* - \tau_2 v))\phi(v) = \phi(c^*)\phi\left(\frac{v-\tau_2 c^*/\tau}{\tau_1/\tau}\right),$$

we have, substituting Q(v) and subsequently E^1, A^1, B^1, E, A, B and C,

$$\int Q(v)\phi(c(v))\phi(v)dv$$

$$= \phi(c^*)\tau^{-1}\left\{E^1 + \frac{\tau_2 c^*}{\tau}A^1 - \frac{\tau_1^2+\tau_2^2 c^{*2}}{\tau^2}B^1\right\}$$

$$= \phi(c^*)\tau^{-1}\left\{E_1 - E_2 - \tfrac{1}{2}\left(1 - \nu\frac{\sigma_1^2}{\sigma^2}\right)\left(\left(\frac{c^*\tau_1}{\tau}\right)^2 + \left(\frac{\tau_2}{\tau}\right)^2\right)\right.$$

$$- \tfrac{1}{2}\left(1 - (1-\nu)\frac{\sigma_2^2}{\sigma^2}\right)\left(\left(\frac{c^*\tau_2}{\tau}\right)^2 + \left(\frac{\tau_1}{\tau}\right)^2\right)$$

$$\left. + \frac{\sigma_1\sigma_2}{\sigma^2}\sqrt{\nu(1-\nu)}\,\frac{(c^{*2}-1)\tau_1\tau_2}{\tau^2}\right\} =$$

$$= \phi(c^*)\tau^{-1}\Big\{E_1 - E_2 - \tfrac{1}{2}(c^{*2}+1)$$
$$+ \tfrac{1}{2}\Big(\frac{c^*}{\tau\sigma}\Big)^2 (\nu(\theta_1-\theta)\sigma_1^2 + (1-\nu)(\theta_2-\theta)\sigma_2^2)^2$$
$$+ \tfrac{1}{2}\Big(\frac{\sigma_1\sigma_2}{\tau\sigma}\Big)^2 ((\theta_2-\theta) - (\theta_1-\theta))^2\nu(1-\nu)\Big\}$$

where the definitions of τ_1 and τ_2 were used. Insertion of E_1 and E_2 ((4.9) and (5.2)) now yields, since $c^* = c_N^* = -u_\beta + O(N^{-\frac{1}{2}})$ (cf. (4.13)),

$$\int Q(v)\phi(c(v))\phi(v)dv$$
$$= \phi(u_\beta)\cdot\tfrac{1}{2}\tau^{-1}\Big\{\nu(1-\nu)\frac{(\theta_1-\theta_2)^2\sigma_1^2\sigma_2^2}{\tau^2\sigma^2} - 1 - \log\nu(1-\nu)\frac{(\theta_1-\theta_2)^2\sigma_1^2\sigma_2^2}{\tau^2\sigma^2}$$
$$+ O(N^{-\frac{1}{2}})\Big\}.$$

□

PROOF of Lemma 6.1. We follow the proof of Lemma 6.2. Using (5.3) and (4.10) we obtain instead of (6.12) that a constant d exists with

$$(6.16)\qquad P_{\theta_1\theta_2}((X,Y) \in A_N \cap B_N)$$
$$\geq P_{\theta_1\theta_2}(U > c(V) + N^{-\frac{1}{2}}Q(V) + N^{-\frac{1}{2}}d,\quad |U|,|V| < N^\delta),$$

where $c(v) = \tau_1^{-1}(-\tau u_\beta - \tau_2 v)$ and $Q(v) = \tau_1^{-1}(A^1 v + B^1 v^2)$. Writing the right hand side of (6.16) as an integral with respect to F_2, we obtain

$$(6.17)\qquad \int 1 - F_1(c(v))dF_2(v)$$
$$- \int F_1(c(v) + N^{-\frac{1}{2}}Q(v) + N^{-\frac{1}{2}}d) - F_1(c(v))dF_2(v)$$

The integrand of the last integral is dominated by the number of lattice points of F_1 in an interval of length $N^{-\frac{1}{2}}(Q(v)+d)$ times the maximal mass of such a lattice point, hence by Theorem 3.4 it is dominated by

$$(6.18)\qquad \sigma_1\sqrt{m}N^{-\frac{1}{2}}\,|Q(v)+d|(\sigma\sqrt{m})^{-1}(\varphi(0)+o(1)).$$

Since $E_{\theta_2}|V|$ and $E_{\theta_2}V^2$ are bounded, the last integral in (6.17) is of order $O(N^{-\frac{1}{2}})$. The first integral of (6.17) equals

$$P_{\theta_1\theta_2}(\tau_1 U + \tau_2 V \geq -\tau u_\beta) = \beta + O(N^{-\frac{1}{2}})$$

by the Berry-Esseen theorem. □

REMARK 6.1. Several of the $\mathcal{O}(N^{-\frac{1}{2}})$ terms in the proof above can be made more explicit, cf. Remark 5.2. The estimates in Section 4 are necessarily rather crude in the lattice case, however, since T_θ then has a lattice distribution when $(\theta_1-\theta)/(\theta_2-\theta)$ is rational.

7. THE OPTIMAL RATIO OF THE SAMPLE SIZES

In this section we study the asymptotic performance of the conditional test (1.2) for different ratios of the sample sizes and compare our fixed alternatives-results with those for local alternatives proved by Albers (1974). Throughout this section, assume that P_0 is a lattice distribution with minimal lattice $\mathbb{Z}$, or that $\{P_\theta : \theta \in \Theta^*\}$ satisfies condition (2.6). We have to adapt the notation a little bit.

For all N, α, θ_1 and θ_2 let $\delta_*^{N,\alpha,\theta_1,\theta_2}$ be the conditional test (1.2) with m and n = N-m chosen such that its power $\beta_*(N,\alpha,\theta_1,\theta_2)$ in (θ_1,θ_2) is maximal. Let $\nu_*(N,\alpha,\theta_1,\theta_2)$ be the ratio m/N of the test $\delta_*^{N,\alpha,\theta_1,\theta_2}$. Note, that for fixed (θ_1,θ_2) the tests $\delta_*^{N,\alpha,\theta_1,\theta_2}$ satisfy (2.12). Furthermore, for each $\nu \in [0,1]$ let $\delta_\nu^{N,\alpha}$ be the test (1.2) with $m = [\nu N]$ and $n = N - [\nu N]$. Define $N_*(\alpha,\beta,\theta_1,\theta_2)$, $\alpha_*(N,\beta,\theta_1,\theta_2)$, $\beta_\nu(N,\alpha,\theta_1,\theta_2)$, $N_\nu(\alpha,\beta,\theta_1,\theta_2)$ and $\alpha_\nu(N,\beta,\theta_1,\theta_2)$ similarly to the definitions in Lemma 2.3.

THEOREM 7.1. (Albers (1974)). *Uniformly for* θ_2 *in a compact subset of* int Θ^*, *uniformly for* α *bounded away from zero and* β *bounded away from* α *and* 1, *as* $\theta_1 \downarrow \theta_2$,

$$(7.1) \qquad N_{\frac{1}{2}}(\alpha,\beta,\theta_1,\theta_2) - N_*(\alpha,\beta,\theta_1,\theta_2) = \mathcal{O}(1)$$

and for each constant $a > 0$, *as* $N \to \infty$,

$$(7.2) \qquad \nu_*(N,\alpha,\theta_2+aN^{-\frac{1}{2}},\theta_2) = \tfrac{1}{2} + \mathcal{O}(N^{-\frac{1}{2}}).$$

PROOF. Albers' conditions in Albers (1974), Section 2, are implied by ours. In fact, Albers gives more explicit expressions for the order terms in (7.1) and (7.2). □

The conclusion from Theorem 7.1 is that for local alternatives the choice m/N = ½ is almost optimal. For fixed alternatives the situation is different. Define ν_0 by

(7.3) $$K(\theta_1,\theta_{\nu_0}) = K(\theta_2,\theta_{\nu_0}).$$

Note that ν_0 maximizes $I_\nu(\theta_1,\theta_2,\theta_\nu)$ as a function of ν, cf. (2.4). Let Ω be a compact subset of int H_1 and let $\varepsilon > 0$.

THEOREM 7.2. *Uniformly for* (θ_1,θ_2) *in* Ω *and uniformly for* $\nu,\beta \in [\varepsilon,1-\varepsilon]$ *as* $\alpha \to 0$

(7.4) $$N_*(\alpha,\beta,\theta_1,\theta_2)/N_\nu(\alpha,\beta,\theta_1,\theta_2)$$
$$= I(\nu)/I(\nu_0) + N_*^{-\frac{1}{2}} \frac{u_\beta}{I(\nu_0)} \left(\left(\frac{I(\nu)}{I(\nu_0)} \right)^{\frac{1}{2}} \tau_{\nu_0} - \tau_\nu \right) + \mathcal{O}(N_*^{-1})$$

where $I(\nu)$ *abbreviates* $I_\nu(\theta_1,\theta_2,\theta_\nu)$ *and* N_* *stands for* $N_*(\alpha,\beta,\theta_1,\theta_2)$.

COROLLARY 7.3. *The Bahadur deficiency of* $\delta_{\nu_0}^{N,\alpha}$ *versus* $\delta_*^{N,\alpha,\theta_1,\theta_2}$ *is bounded at the fixed alternative* (θ_1,θ_2), *hence the ratio* ν_0 *is almost optimal. The Bahadur efficiency of* $\delta_{\frac{1}{2}}^{N,\alpha}$ *versus* $\delta_*^{N,\alpha,\theta_1,\theta_2}$ *is smaller than one when* $\nu_0 \neq \frac{1}{2}$.

In the sequel fix $(\theta_1,\theta_2) \in \Omega$ and delete (θ_1,θ_2) from the notation. The key lemma for the proof of Theorem 7.2 is

LEMMA 7.4. *Uniformly for* $\beta,\nu \in [\varepsilon,1-\varepsilon]$ *as* $N \to \infty$

(7.5) $$\alpha_\nu(N,\beta) = \exp\{-NI(\nu) + N^{\frac{1}{2}}\tau_\nu u_\beta - \tfrac{1}{2}\log N + \mathcal{O}(1)\}.$$

PROOF. Let $\alpha_{N\nu}(\beta)$ be defined by

$$\alpha_{N\nu}(\beta) = \exp\{-NI(\nu) + N^{\frac{1}{2}}\tau_\nu u_\beta - \tfrac{1}{2}\log N\}.$$

By Lemma 6.1 we have uniformly for $\beta,\nu \in [\varepsilon,1-\varepsilon]$ as $N \to \infty$

$$\beta_\nu(N,\alpha_{N\nu}(\beta)) = \beta + \mathcal{O}(N^{-\frac{1}{2}}).$$

Let $\{\beta_{1N}\},\{\beta_{2N}\}$ be sequences such that $\beta_{iN} = \beta + \mathcal{O}(N^{-\frac{1}{2}})$, $i = 1,2$, and for all $\nu \in [\varepsilon,1-\varepsilon]$,

$$\beta_\nu(N,\alpha_{N\nu}(\beta_{1N})) < \beta \le \beta_\nu(N,\alpha_{N\nu}(\beta_{2N})),$$

implying

$$\alpha_{N\nu}(\beta_{1N}) \le \alpha_\nu(N,\beta) \le \alpha_{N\nu}(\beta_{2N})$$

and (7.5) follows by first order Taylor expansions of $u_{\beta_{iN}}$, $i = 1,2$. □

We separate another lemma from the proof:

LEMMA 7.5. *When* $\eta > 0$ *is sufficiently small and* N *is large enough,*

$$\min_{0\le\nu\le 1} \alpha_\nu(N,\beta) = \min_{\eta<\nu<1-\eta} \alpha_\nu(N,\beta). \tag{7.6}$$

PROOF. Let $0 < \eta < \frac{1}{2}$ such that $2I(\eta) < I(\frac{1}{2})$, suppose $\nu \le \eta$. Define for each N, $M_N = \left[\frac{1-\nu}{1-\eta}N + 3\right]$, then it follows that $[\eta M_N] \ge [\nu N]$ and $M_N - [\eta M_N] \ge (1-\eta)M_N > (1-\nu)N + 1 \ge N - [\nu N]$. Since the test $\delta_\eta^{M_N,\alpha}$ is UMPU we have $\beta_\eta(M_N,\alpha) \ge \beta_\nu(N,\alpha)$, implying

$$\alpha_\eta(M_N,\beta_\nu(N,\alpha)) \le \alpha. \tag{7.7}$$

By the definition of $\alpha_\nu(N,\beta)$ it holds that

$$\beta_\nu(N,\alpha_\nu(N,\beta)) \ge \beta,$$

whence (7.7) yields

$$\alpha_\eta(M_N,\beta) \le \alpha_\nu(N,\beta).$$

Using (7.5) and the definition of M_N it follows by $2I(\eta) < I(\frac{1}{2})$ that for $\nu \le \eta$, and similarly for $\nu \ge 1-\eta$, $\alpha_\nu(N,\beta) > \alpha_{\frac{1}{2}}(N,\beta)$ when N is large enough. □

PROOF of Theorem 7.2. For all N,β we have

$$\alpha_{\nu_0}(N,\beta) \ge \alpha_*(N,\beta) \ge \min_{0\le\nu\le 1} \alpha_\nu(N,\beta). \tag{7.8}$$

The first inequality follows from

$$\beta_*(N,\alpha_{\nu_0}(N,\beta)) \ge \beta_{\nu_0}(N,\alpha_{\nu_0}(N,\beta)) \ge \beta,$$

and the second is implied by the fact that for each N the set $\{\delta_*^{N,\alpha} : \alpha > 0\}$ is a subset of $\{\delta_\nu^{N,\alpha} : \nu \in [0,1],\ \alpha > 0\}$. By Lemmas 7.5 and 7.4 the minimum in (7.8) equals for some $\eta > 0$, as $N \to \infty$,

$$\min_{\eta<\nu<1-\eta} N^{-\frac{1}{2}} \exp\{-NI(\nu) + N^{\frac{1}{2}}\tau_\nu u_\beta + O(1)\}. \tag{7.9}$$

Let the minimum be attained for $\nu = \nu_{min}(N,\beta)$ then, by equating the derivative of the leading terms in (7.9) to zero, we have

$$\nu_{min}(N,\beta) = \nu_0 + \mathcal{O}(N^{-\frac{1}{2}}) \quad \text{as } N \to \infty.$$

By Taylor expansion - note that $\frac{\partial I}{\partial \nu}\rfloor_{\nu_0} = 0$ -, (7.8) and Lemma 7.4 it follows that as $N \to \infty$,

$$\text{(7.10)} \qquad \alpha_*(N,\beta) = N^{-\frac{1}{2}} \exp\{-NI(\nu_0) + N^{\frac{1}{2}}\tau_{\nu_0}u_\beta + \mathcal{O}(1)\}.$$

Proceeding from (7.5) and (7.10) as in the proof of Lemma 2.3 we obtain, writing $N_* = N_*(\alpha,\beta)$ and $N_\nu = N_\nu(\alpha,\beta)$, that as $\alpha \to 0$

$$N_*I(\nu_0) - N_*^{\frac{1}{2}}\tau_{\nu_0}u_\beta + \tfrac{1}{2}\log N_* = N_\nu I(\nu) - N_\nu^{\frac{1}{2}}\tau_\nu u_\beta + \tfrac{1}{2}\log N_\nu + \mathcal{O}(1).$$

The theorem follows. □

8. NUMERICAL COMPLEMENTS

In the situation of Example 1.1 bounds on the deficiency have been determined by numerical computation. For some values of the alternative (p_1,p_2), the power β and the sample size N, we determined a MP test δ_{++} of (p_N,p_N) against (p_1,p_2) having power β at (p_1,p_2), where p_N is determined by relation (4.5). The sample ratio m/N was taken equal to $\frac{1}{2}$, see below.

Denoting the size of δ_{++} as α_N, we then computed the smallest $D = D(N,\beta,p_1,p_2)$ with

$$\beta_c(N+D,\alpha_N) \geq \beta.$$

Note that since δ_{++} is MP for a *smaller* null hypothesis than the MP test δ_+ of H against (p_1,p_2), D is an *upper bound* on the deficiency $N_c - N_+$. Table 8.1 actually gives the "randomized" bound D^*, defined as

$$D^* = D - \frac{\beta_c(N+D,\alpha_N) - \beta}{\beta_c(N+D,\alpha_N) - \beta_c(N+D-1,\alpha_N)}.$$

Thus, putting $D^* = D-1+R$, one has to include the (N+D)-th observation with probability R to obtain exactly power β with the conditional size $-\alpha_N$ test. Note that for moderate values of (p_1,p_2) the deficiency-bound is quite low, it seems to increase only in extreme cases but even then its constancy as a function of N is remarkable.

Furthermore, to complement Section 7 we computed the Bahadur-optimal sample ratio ν_0 and the Bahadur efficiency of the conditional test with ratio $\frac{1}{2}$ versus the test with ratio ν_0. Both are given in table 8.2 for a range of alternatives (p_1,p_2). Note that ν_0 is close to $\frac{1}{2}$ when p_1 and p_2 are moderate; for all but very extreme p's the Bahadur efficiency does not drop below 0.95.

p_1	p_2	N	β = 0.6 α_N	$\beta_c(N)$	D^*	β = 0.9 α_N	$\beta_c(N)$	D^*
0.10	0.01	20	0.247699	0.5031	9.169	0.685115	0.8730	4.896
		40	0.150003	0.5263	8.963	0.485878	0.8877	3.741
		60	0.086540	0.5297	10.658	0.333542	0.8668	10.487
		100	0.031275	0.5477	9.959	0.178110	0.8785	9.616
		150	0.009469	0.5555	10.564	0.079469	0.8826	9.648
		200	0.002964	0.5637	10.348	0.033828	0.8854	9.235
		300	0.000301	0.5696	10.908	0.006123	0.8883	9.134
		400	0.000032	0.5742	10.711	0.001014	0.8896	9.586
		500	0.000003	0.5781	10.296	0.000169	0.8911	9.567
		600	-	-	-	0.000027	0.8913	10.004
		700	-	-	-	0.000004	0.8925	9.306
0.20	0.05	20	0.217871	0.5629	3.513	0.599428	0.8844	2.615
		40	0.109422	0.5714	3.705	0.407056	0.8825	4.601
		60	0.059098	0.5771	3.545	0.285522	0.8888	3.934
		100	0.018282	0.5816	3.784	0.135545	0.8923	3.520
		150	0.004428	0.5863	3.563	0.051528	0.8938	3.495
		200	0.001103	0.5877	3.703	0.018970	0.8946	3.545
		300	0.000071	0.5900	3.650	0.002394	0.8956	3.375
		400	0.000005	0.5913	3.669	0.000280	0.8961	3.504
		500	-	-	-	0.000031	0.8966	3.617
		600	-	-	-	0.000003	0.8968	3.597
0.50	0.30	20	0.249541	0.5816	2.006	0.641540	0.8953	1.219
		40	0.148099	0.5924	1.348	0.493607	0.8993	1.029
		60	0.090976	0.5959	1.237	0.377140	0.8978	1.184
		100	0.035990	0.5952	1.263	0.218444	0.8982	1.033
		150	0.011842	0.5943	1.581	0.106783	0.8972	1.772
		200	0.004034	0.5957	1.465	0.051535	0.8982	1.420
		300	0.000483	0.5970	1.402	0.011230	0.8984	1.476
		400	0.000059	0.5971	1.439	0.002314	0.8987	1.401
		500	0.000007	0.5974	1.435	0.000457	0.8988	1.423
		600	-	-	-	0.000087	0.8990	1.402
		700	-	-	-	0.000016	0.8990	1.422
0.60	0.40	20	0.256136	0.5831	1.702	0.650192	0.8971	1.005
		40	0.155457	0.5951	1.000	0.504658	0.8997	1.000
		60	0.097319	0.5993	1.001	0.391140	0.8993	1.000
		100	0.039800	0.5960	1.012	0.231383	0.8977	1.008
		150	0.013527	0.5937	1.820	0.117941	0.8985	1.000
		200	0.004864	0.5975	1.000	0.058900	0.8998	1.000
		300	0.000635	0.5984	1.003	0.013801	0.8998	1.000
		400	0.000084	0.5964	1.380	0.003030	0.8985	1.000
		500	0.000011	0.5987	1.001	0.000651	0.8996	1.000
		600	0.000002	0.5994	1.004	0.000133	0.8988	1.095
		700	-	-	-	0.000027	0.8991	1.001

Table 8.1. The power of the conditional test and its deficiency with respect to the test δ_{++}.

p_1 \ p_2	0.001	0.002	0.005	0.01	0.02	0.05	0.1	0.2
0.002	0.472 0.997							
0.005	0.438 0.985	0.463 0.995						
0.01	0.417 0.975	0.438 0.985	0.472 0.997					
0.02	0.402 0.965	0.418 0.975	0.446 0.989	0.472 0.997				
0.05	0.389 0.957	0.399 0.964	0.419 0.976	0.439 0.986	0.464 0.995			
0.1	0.384 0.954	0.391 0.959	0.405 0.968	0.421 0.977	0.441 0.987	0.474 0.997		
0.2	0.385 0.955	0.389 0.958	0.399 0.964	0.410 0.971	0.426 0.980	0.453 0.991	0.476 0.998	
0.5	0.403 0.969	0.405 0.970	0.410 0.973	0.417 0.976	0.426 0.981	0.444 0.989	0.462 0.995	0.482 0.999

Table 8.2. The asymptotically optimal sample ratio ν_0 (upper entry) and the Bahadur efficiency (lower entry) of the conditional test with ratio $\frac{1}{2}$ with respect to the conditional test with ratio ν_0.

REFERENCES

Albers, W. (1974), *Asymptotic expansions and the deficiency concept in statistics*, Mathematisch Centrum, Amsterdam.

Andrews, D.F. *et al* (1972), *Robust estimates of location: Survey and Advances*, Princeton University Press, Princeton N.J.

Bahadur, R.R. (1960 a), *Stochastic comparison of tests*, Ann. Math. Statist. 31 276-295.

Bahadur, R.R. (1960 b), *Asymptotic efficiency of tests and estimates*, Sankhyā 22 229-252.

Bahadur, R.R. (1965), *An optimal property of the likelihood ratio statistic*, Proc. Fifth Berkeley Symp. Math. Stat. Prob. 1 13-26.

Bahadur, R.R. (1967), *Rates of convergence of estimates and test statistics*, Ann. Math. Statist. 38 303-324.

Bahadur, R.R. (1971), *Some limit theorems in statistics*, SIAM, Philadelphia.

Bahadur, R.R. (1980), *On large deviations of maximum likelihood and related estimates*, Technical report no 121, Univ. of Chicago.

Bahadur, R.R. (1983), *Large deviations of the maximum likelihood estimate in the Markov chain case*, in Recent Advances in Statistics, (M.H. Rizvi, J.S. Rustagi, D. Sigmund, Eds.), Academic Press, New York.

Bahadur, R.R., Gupta, J.C. and Zabell, S.L. (1980), *Large deviations, tests and estimates*, in Asymptotic theory of statistical tests and estimation (I.M. Chakravarti, Ed.), Academic Press, New York.

Barndorff-Nielsen, O. (1978), *Information and Exponential Families in Statistical Theory*, Wiley, New York.

Barndorff-Nielsen, O. and Cox, D.R. (1979), *Edgeworth and saddlepoint approximations with statistical applications*, J. Roy. Statist. Soc., Ser. B 41 279-312.

Basu, D. (1956), *The concept of asymptotic efficiency*, Sankhyā 17 193-196.

Berk, R.H. (1972), *Consistency and asymptotic normality of* MLE's *for exponential models*, Ann. Math. Statist. 43 193-204.

Billingsley, P. (1968), *Convergence of probability measures*, Wiley, New York.

Boos, D.D. (1979), *A differential for* L-*statistics*, Ann. Statist. 5 955-959.

Brown, L.D. (1971), *Non-local asymptotic normality of appropriate likelihood ratio tests*, Ann. Math. Statist. 42 1206-1240.

Chernoff, H. (1952), *A measure of asymptotic efficiency for tests of a hypothesis based on sums of observations*, Ann. Math. Statist. 23 493-507.

Chernoff, H. (1956), *Large sample theory - parametric case*, Ann. Math. Statist. 27 1-22.

Csiszár, I. (1975), *I-divergence geometry of probability distributions and minimization problems*, Ann. Probability 3 146-158.

Daniels, H.E. (1954), *Saddlepoint approximation in Statistics*, Ann. Math. Statist. 25 631-650.

Efron, B. (1975), *Defining the curvature of a statistical problem (with applications to second order efficiency)*, Ann. Statist. 3 1189-1242.

Efron, B. (1978), *The geometry of exponential families*, Ann. Statist. 6 362-376.

Efron, B. and Hinkley, D.V. (1978), *Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information*, Biometrika 65 457-487.

Efron, B. and Truax, D. (1968), *Large deviations theory in exponential families*, Ann. Math. Statist. 39 1402-1424.

Feller, W. (1971), *An introduction to probability theory and its applications* II, 2nd ed., Wiley, New York.

Ferguson, T.S. (1967), *Mathematical Statistics. A decision theoretic approach*, Academic Press, New York.

Fisher, R.A. (1925), *Theory of statistical estimation*, Proc. Camb. Phil. Soc. 22 700-725.

Fisher, R.A. (1956), *Statistical Methods and Scientific Inference*, Oliver and Boyd, Edinburgh.

Fu, J.C. (1973), *On a theorem of Bahadur on the rate of convergence of point estimators*, Ann. Statist. 1 745-749.

Fu, J.C. (1975), *The rate of convergence of consistent point estimators*, Ann. Statist. 3 234-240.

Fu, J.C. (1982), *Large sample point estimation: a large deviation theory approach*, Ann. Statist. 10 762-771.

Groeneboom, P. (1978), *Large deviations and asymptotic efficiency*, Mathematisch Centrum, Amsterdam.

Groeneboom, P. and Oosterhoff, J. (1981), *Bahadur efficiency and small sample efficiency*, International Statist. Review 49 127-141.

Groeneboom, P. and Shorack, G.R. (1981), *Large deviations of goodness of fit statistics and linear combinations of order statistics*, Ann. Probability 9 971-987.

Groeneboom, P., Oosterhoff, J. and Ruymgaart, F.H. (1979), *Large deviation theorems for empirical probability measures*, Ann. Probability 7 553-586.

Hampel, F.R. (1968), *Contributions to the theory of robust estimation*, Ph. D. Thesis, Univ. of California, Berkeley.

Hampel, F.R. (1974), *The influence curve and its role in robust estimation*, J. Amer. Statist. Ass. 62 1179-1186.

Hannan, J. and Harkness, W. (1963), *Normal approximation on the distribution of two independent binomials, conditional on fixed sum*, Ann. Math. Statist. 34 1593-1595.

Hodges, J.L. and Lehmann, E.L. (1970), *Deficiency*, Ann. Math. Statist. 41 783-801.

Hoefdding, W. (1965), *On probabilities of large deviations*, Proc. Fifth Berkeley Symp. Math. Statist. Prob. 1 203-219.

Huber, P.J. (1968), *Robust confidence limits*, Z. Wahrscheinlichkeits-theorie Verw. Gebiete 10 269-278.

Kallenberg, W.C.M. (1978), *Asymptotic optimality of likelihood ratio tests in exponential families*, Mathematisch Centrum, Amsterdam.

Kallenberg, W.C.M. (1981), *Bahadur deficiency of likelihood ratio tests in exponential families,* J. Multivariate Anal. 11 506-531.

Kallenberg, W.C.M. (1982), *Chernoff efficiency and deficiency,* Ann. Statist. 10 583-594.

Kallenberg, W.C.M. (1983), *A note on moderate deviation theory in estimation,* Ann. Statist. 11 498-504.

Kester, A. (1981), *Large deviation optimality of* MLE's *in exponential families,* Rapport 160, Wiskundig Seminarium der Vrije Universiteit, Amsterdam.

Kester, A. (1978), *The Bahadur slope of a test based on Kullback-Leibler information,* Rapport 76, Wiskundig Seminarium der Vrije Universiteit, Amsterdam.

Lehmann, E.L. (1959), *Testing statistical hypotheses,* Wiley, New York.

Leurgans, S. (1981), *The Cauchy mean value property and linear functions of order statistics,* Ann. Statist. 9 905-908.

Michel, R. (1979), *Third order efficiency of conditional tests for exponential families,* J. Multivariate Anal. 9 401-409.

Perng, S.-S. (1978), *Rate of convergence of estimators based on sample mean,* Ann. Statist. 6 1048-1056.

Pfanzagl, J.(1976), *Investigating the quantile of an unknown distribution,* in Contributions to Statistics (W.J. Ziegler, Ed.) 111-126, Birkhäuser Verlag, Basel.

Pinsker, M.S. (1964), *Information and information stability of random variables and processes,* Holden-Day, San Francisco.

Portnoy, S.L. (1977), *Robust estimation in dependent situations,* Ann. Statist. 5 22-43.

Rao, C.R. (1961), *Asymptotic efficiency and limiting information,* Proc. Fourth Berkeley Symp. Math. Statist. Prob. 1 531- 545.

Rao, C.R. (1962), *Efficient estimates and optimum inference procedures in large samples,* J. Roy. Statist. Soc., Ser. B 24 46-72.

Rao, C.R. (1963), *Criteria of estimation in large samples,* Sankhyā 25 189-206.

Rubin, H. and Rukhin, A.L. (1983), *Convergence rates of large deviations probabilities for point estimators*, Statistics and Probability letters 4 197-202.

Rukhin, A.L. (1983), *Convergence rates of estimators of a finite parameter: How small can error probabilities be?*, Ann. Statist. 11 202-206.

Serfling, R.J. (1980), *Approximation theorems of mathematical statistics*, Wiley, New York.

Sievers, G.L. (1978), *Estimates of location: a large deviation comparison*, Ann. Statist. 6 610-618.

Witting, H. (1966), *Mathematische Statistik*, Teubner, Stuttgart.

AUTHOR INDEX

SUBJECT INDEX

[1]) Underlined pagenumbers refer to the defining or main occurence.

MATHEMATICAL CENTRE TRACTS

1 T. van der Walt. *Fixed and almost fixed points.* 1963.

2 A.R. Bloemena. *Sampling from a graph.* 1964.

3 G. de Leve. *Generalized Markovian decision processes, part I: model and method.* 1964.

4 G. de Leve. *Generalized Markovian decision processes, part II: probabilistic background.* 1964.

5 G. de Leve, H.C. Tijms, P.J. Weeda. *Generalized Markovian decision processes, applications.* 1970.

6 M.A. Maurice. *Compact ordered spaces.* 1964.

7 W.R. van Zwet. *Convex transformations of random variables.* 1964.

8 J.A. Zonneveld. *Automatic numerical integration.* 1964.

9 P.C. Baayen. *Universal morphisms.* 1964.

10 E.M. de Jager. *Applications of distributions in mathematical physics.* 1964.

11 A.B. Paalman-de Miranda. *Topological semigroups.* 1964.

12 J.A.Th.M. van Berckel, H. Brandt Corstius, R.J. Mokken, A. van Wijngaarden. *Formal properties of newspaper Dutch.* 1965.

13 H.A. Lauwerier. *Asymptotic expansions.* 1966, out of print; replaced by MCT 54.

14 H.A. Lauwerier. *Calculus of variations in mathematical physics.* 1966.

15 R. Doornbos. *Slippage tests.* 1966.

16 J.W. de Bakker. *Formal definition of programming languages with an application to the definition of ALGOL 60.* 1967.

17 R.P. van de Riet. *Formula manipulation in ALGOL 60, part 1.* 1968.

18 R.P. van de Riet. *Formula manipulation in ALGOL 60, part 2.* 1968.

19 J. van der Slot. *Some properties related to compactness.* 1968.

20 P.J. van der Houwen. *Finite difference methods for solving partial differential equations.* 1968.

21 E. Wattel. *The compactness operator in set theory and topology.* 1968.

22 T.J. Dekker. *ALGOL 60 procedures in numerical algebra, part 1.* 1968.

23 T.J. Dekker, W. Hoffmann. *ALGOL 60 procedures in numerical algebra, part 2.* 1968.

24 J.W. de Bakker. *Recursive procedures.* 1971.

25 E.R. Paërl. *Representations of the Lorentz group and projective geometry.* 1969.

26 European Meeting 1968. *Selected statistical papers, part I.* 1968.

27 European Meeting 1968. *Selected statistical papers, part II.* 1968.

28 J. Oosterhoff. *Combination of one-sided statistical tests.* 1969.

29 J. Verhoeff. *Error detecting decimal codes.* 1969.

30 H. Brandt Corstius. *Exercises in computational linguistics.* 1970.

31 W. Molenaar. *Approximations to the Poisson, binomial and hypergeometric distribution functions.* 1970.

32 L. de Haan. *On regular variation and its application to the weak convergence of sample extremes.* 1970.

33 F.W. Steutel. *Preservation of infinite divisibility under mixing and related topics.* 1970.

34 I. Juhász, A. Verbeek, N.S. Kroonenberg. *Cardinal functions in topology.* 1971.

35 M.H. van Emden. *An analysis of complexity.* 1971.

36 J. Grasman. *On the birth of boundary layers.* 1971.

37 J.W. de Bakker, G.A. Blaauw, A.J.W. Duijvestijn, E.W. Dijkstra, P.J. van der Houwen, G.A.M. Kamsteeg-Kemper, F.E.J. Kruseman Aretz, W.L. van der Poel, J.P. Schaap-Kruseman, M.V. Wilkes, G. Zoutendijk. *MC-25 Informatica Symposium.* 1971.

38 W.A. Verloren van Themaat. *Automatic analysis of Dutch compound words.* 1972.

39 H. Bavinck. *Jacobi series and approximation.* 1972.

40 H.C. Tijms. *Analysis of (s,S) inventory models.* 1972.

41 A. Verbeek. *Superextensions of topological spaces.* 1972.

42 W. Vervaat. *Success epochs in Bernoulli trials (with applications in number theory).* 1972.

43 F.H. Ruymgaart. *Asymptotic theory of rank tests for independence.* 1973.

44 H. Bart. *Meromorphic operator valued functions.* 1973.

45 A.A. Balkema. *Monotone transformations and limit laws.* 1973.

46 R.P. van de Riet. *ABC ALGOL, a portable language for formula manipulation systems, part 1: the language.* 1973.

47 R.P. van de Riet. *ABC ALGOL, a portable language for formula manipulation systems, part 2: the compiler.* 1973.

48 F.E.J. Kruseman Aretz, P.J.W. ten Hagen, H.L. Oudshoorn. *An ALGOL 60 compiler in ALGOL 60, text of the MC-compiler for the EL-X8.* 1973.

49 H. Kok. *Connected orderable spaces.* 1974.

50 A. van Wijngaarden, B.J. Mailloux, J.E.L. Peck, C.H.A. Koster, M. Sintzoff, C.H. Lindsey, L.G.L.T. Meertens, R.G. Fisker (eds.). *Revised report on the algorithmic language ALGOL 68.* 1976.

51 A. Hordijk. *Dynamic programming and Markov potential theory.* 1974.

52 P.C. Baayen (ed.). *Topological structures.* 1974.

53 M.J. Faber. *Metrizability in generalized ordered spaces.* 1974.

54 H.A. Lauwerier. *Asymptotic analysis, part 1.* 1974.

55 M. Hall, Jr., J.H. van Lint (eds.). *Combinatorics, part 1: theory of designs, finite geometry and coding theory.* 1974.

56 M. Hall, Jr., J.H. van Lint (eds.). *Combinatorics, part 2: graph theory, foundations, partitions and combinatorial geometry.* 1974.

57 M. Hall, Jr., J.H. van Lint (eds.). *Combinatorics, part 3: combinatorial group theory.* 1974.

58 W. Albers. *Asymptotic expansions and the deficiency concept in statistics.* 1975.

59 J.L. Mijnheer. *Sample path properties of stable processes.* 1975.

60 F. Göbel. *Queueing models involving buffers.* 1975.

63 J.W. de Bakker (ed.). *Foundations of computer science.* 1975.

64 W.J. de Schipper. *Symmetric closed categories.* 1975.

65 J. de Vries. *Topological transformation groups, 1: a categorical approach.* 1975.

66 H.G.J. Pijls. *Logically convex algebras in spectral theory and eigenfunction expansions.* 1976.

68 P.P.N. de Groen. *Singularly perturbed differential operators of second order.* 1976.

69 J.K. Lenstra. *Sequencing by enumerative methods.* 1977.

70 W.P. de Roever, Jr. *Recursive program schemes: semantics and proof theory.* 1976.

71 J.A.E.E. van Nunen. *Contracting Markov decision processes.* 1976.

72 J.K.M. Jansen. *Simple periodic and non-periodic Lamé functions and their applications in the theory of conical waveguides.* 1977.

73 D.M.R. Leivant. *Absoluteness of intuitionistic logic.* 1979.

74 H.J.J. te Riele. *A theoretical and computational study of generalized aliquot sequences.* 1976.

75 A.E. Brouwer. *Treelike spaces and related connected topological spaces.* 1977.

76 M. Rem. *Associons and the closure statement.* 1976.

77 W.C.M. Kallenberg. *Asymptotic optimality of likelihood ratio tests in exponential families.* 1978.

78 E. de Jonge, A.C.M. van Rooij. *Introduction to Riesz spaces.* 1977.

79 M.C.A. van Zuijlen. *Emperical distributions and rank statistics.* 1977.

80 P.W. Hemker. *A numerical study of stiff two-point boundary problems.* 1977.

81 K.R. Apt, J.W. de Bakker (eds.). *Foundations of computer science II, part 1.* 1976.

82 K.R. Apt, J.W. de Bakker (eds.). *Foundations of computer science II, part 2.* 1976.

83 L.S. van Benthem Jutting. *Checking Landau's "Grundlagen" in the AUTOMATH system.* 1979.

84 H.L.L. Busard. *The translation of the elements of Euclid from the Arabic into Latin by Hermann of Carinthia (?), books vii-xii.* 1977.

85 J. van Mill. *Supercompactness and Wallman spaces.* 1977.

86 S.G. van der Meulen, M. Veldhorst. *Torrix I, a programming system for operations on vectors and matrices over arbitrary fields and of variable size.* 1978.

88 A. Schrijver. *Matroids and linking systems.* 1977.

89 J.W. de Roever. *Complex Fourier transformation and analytic functionals with unbounded carriers.* 1978.

90 L.P.J. Groenewegen. *Characterization of optimal strategies in dynamic games.* 1981.

91 J.M. Geysel. *Transcendence in fields of positive characteristic.* 1979.

92 P.J. Weeda. *Finite generalized Markov programming.* 1979.

93 H.C. Tijms, J. Wessels (eds.). *Markov decision theory.* 1977.

94 A. Bijlsma. *Simultaneous approximations in transcendental number theory.* 1978.

95 K.M. van Hee. *Bayesian control of Markov chains.* 1978.

96 P.M.B. Vitányi. *Lindenmayer systems: structure, languages, and growth functions.* 1980.

97 A. Federgruen. *Markovian control problems; functional equations and algorithms.* 1984.

98 R. Geel. *Singular perturbations of hyperbolic type.* 1978.

99 J.K. Lenstra, A.H.G. Rinnooy Kan, P. van Emde Boas (eds.). *Interfaces between computer science and operations research.* 1978.

100 P.C. Baayen, D. van Dulst, J. Oosterhoff (eds.). *Proceedings bicentennial congress of the Wiskundig Genootschap, part 1.* 1979.

101 P.C. Baayen, D. van Dulst, J. Oosterhoff (eds.). *Proceedings bicentennial congress of the Wiskundig Genootschap, part 2.* 1979.

102 D. van Dulst. *Reflexive and superreflexive Banach spaces.* 1978.

103 K. van Harn. *Classifying infinitely divisible distributions by functional equations.* 1978.

104 J.M. van Wouwe. *Go-spaces and generalizations of metrizability.* 1979.

105 R. Helmers. *Edgeworth expansions for linear combinations of order statistics.* 1982.

106 A. Schrijver (ed.). *Packing and covering in combinatorics.* 1979.

107 C. den Heijer. *The numerical solution of nonlinear operator equations by imbedding methods.* 1979.

108 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science III, part 1.* 1979.

109 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science III, part 2.* 1979.

110 J.C. van Vliet. *ALGOL 68 transput, part I: historical review and discussion of the implementation model.* 1979.

111 J.C. van Vliet. *ALGOL 68 transput, part II: an implementation model.* 1979.

112 H.C.P. Berbee. *Random walks with stationary increments and renewal theory.* 1979.

113 T.A.B. Snijders. *Asymptotic optimality theory for testing problems with restricted alternatives.* 1979.

114 A.J.E.M. Janssen. *Application of the Wigner distribution to harmonic analysis of generalized stochastic processes.* 1979.

115 P.C. Baayen, J. van Mill (eds.). *Topological structures II, part 1.* 1979.

116 P.C. Baayen, J. van Mill (eds.). *Topological structures II, part 2.* 1979.

117 P.J.M. Kallenberg. *Branching processes with continuous state space.* 1979.

118 P. Groeneboom. *Large deviations and asymptotic efficiencies.* 1980.

119 F.J. Peters. *Sparse matrices and substructures, with a novel implementation of finite element algorithms.* 1980.

120 W.P.M. de Ruyter. *On the asymptotic analysis of large-scale ocean circulation.* 1980.

121 W.H. Haemers. *Eigenvalue techniques in design and graph theory.* 1980.

122 J.C.P. Bus. *Numerical solution of systems of nonlinear equations.* 1980.

123 I. Yuhász. *Cardinal functions in topology - ten years later.* 1980.

124 R.D. Gill. *Censoring and stochastic integrals.* 1980.

125 R. Eising. *2-D systems, an algebraic approach.* 1980.

126 G. van der Hoek. *Reduction methods in nonlinear programming.* 1980.

127 J.W. Klop. *Combinatory reduction systems.* 1980.

128 A.J.J. Talman. *Variable dimension fixed point algorithms and triangulations.* 1980.

129 G. van der Laan. *Simplicial fixed point algorithms.* 1980.

130 P.J.W. ten Hagen, T. Hagen, P. Klint, H. Noot, H.J. Sint, A.H. Veen. *ILP: intermediate language for pictures.* 1980.

131 R.J.R. Back. *Correctness preserving program refinements: proof theory and applications.* 1980.

132 H.M. Mulder. *The interval function of a graph.* 1980.

133 C.A.J. Klaassen. *Statistical performance of location estimators.* 1981.

134 J.C. van Vliet, H. Wupper (eds.). *Proceedings international conference on ALGOL 68.* 1981.

135 J.A.G. Groenendijk, T.M.V. Janssen, M.J.B. Stokhof (eds.). *Formal methods in the study of language, part I.* 1981.

136 J.A.G. Groenendijk, T.M.V. Janssen, M.J.B. Stokhof (eds.). *Formal methods in the study of language, part II.* 1981.

137 J. Telgen. *Redundancy and linear programs.* 1981.

138 H.A. Lauwerier. *Mathematical models of epidemics.* 1981.

139 J. van der Wal. *Stochastic dynamic programming, successive approximations and nearly optimal strategies for Markov decision processes and Markov games.* 1981.

140 J.H. van Geldrop. *A mathematical theory of pure exchange economies without the no-critical-point hypothesis.* 1981.

141 G.E. Welters. *Abel-Jacobi isogenies for certain types of Fano threefolds.* 1981.

142 H.R. Bennett, D.J. Lutzer (eds.). *Topology and order structures, part 1.* 1981.

143 J.M. Schumacher. *Dynamic feedback in finite- and infinite-dimensional linear systems.* 1981.

144 P. Eijgenraam. *The solution of initial value problems using interval arithmetic; formulation and analysis of an algorithm.* 1981.

145 A.J. Brentjes. *Multi-dimensional continued fraction algorithms.* 1981.

146 C.V.M. van der Mee. *Semigroup and factorization methods in transport theory.* 1981.

147 H.H. Tigelaar. *Identification and informative sample size.* 1982.

148 L.C.M. Kallenberg. *Linear programming and finite Markovian control problems.* 1983.

149 C.B. Huijsmans, M.A. Kaashoek, W.A.J. Luxemburg, W.K. Vietsch (eds.). *From A to Z, proceedings of a symposium in honour of A.C. Zaanen.* 1982.

150 M. Veldhorst. *An analysis of sparse matrix storage schemes.* 1982.

151 R.J.M.M. Does. *Higher order asymptotics for simple linear rank statistics.* 1982.

152 G.F. van der Hoeven. *Projections of lawless sequences.* 1982.

153 J.P.C. Blanc. *Application of the theory of boundary value problems in the analysis of a queueing model with paired services.* 1982.

154 H.W. Lenstra, Jr., R. Tijdeman (eds.). *Computational methods in number theory, part I.* 1982.

155 H.W. Lenstra, Jr., R. Tijdeman (eds.). *Computational methods in number theory, part II.* 1982.

156 P.M.G. Apers. *Query processing and data allocation in distributed database systems.* 1983.

157 H.A.W.M. Kneppers. *The covariant classification of two-dimensional smooth commutative formal groups over an algebraically closed field of positive characteristic.* 1983.

158 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science IV, distributed systems, part 1.* 1983.

159 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science IV, distributed systems, part 2.* 1983.

160 A. Rezus. *Abstract AUTOMATH.* 1983.

161 G.F. Helminck. *Eisenstein series on the metaplectic group, an algebraic approach.* 1983.

162 J.J. Dik. *Tests for preference.* 1983.

163 H. Schippers. *Multiple grid methods for equations of the second kind with applications in fluid mechanics.* 1983.

164 F.A. van der Duyn Schouten. *Markov decision processes with continuous time parameter.* 1983.

165 P.C.T. van der Hoeven. *On point processes.* 1983.

166 H.B.M. Jonkers. *Abstraction, specification and implementation techniques, with an application to garbage collection.* 1983.

167 W.H.M. Zijm. *Nonnegative matrices in dynamic programming.* 1983.

168 J.H. Evertse. *Upper bounds for the numbers of solutions of diophantine equations.* 1983.

169 H.R. Bennett, D.J. Lutzer (eds.). *Topology and order structures, part 2.* 1983.

CWI TRACTS

1 D.H.J. Epema. *Surfaces with canonical hyperplane sections.* 1984.

2 J.J. Dijkstra. *Fake topological Hilbert spaces and characterizations of dimension in terms of negligibility.* 1984.

3 A.J. van der Schaft. *System theoretic descriptions of physical systems.* 1984.

4 J. Koene. *Minimal cost flow in processing networks, a primal approach.* 1984.

5 B. Hoogenboom. *Intertwining functions on compact Lie groups.* 1984.

6 A.P.W. Böhm. *Dataflow computation.* 1984.

7 A. Blokhuis. *Few-distance sets.* 1984.

8 M.H. van Hoorn. *Algorithms and approximations for queueing systems.* 1984.

9 C.P.J. Koymans. *Models of the lambda calculus.* 1984.

10 C.G. van der Laan, N.M. Temme. *Calculation of special functions: the gamma function, the exponential integrals and error-like functions.* 1984.

11 N.M. van Dijk. *Controlled Markov processes; time-discretization.* 1984.

12 W.H. Hundsdorfer. *The numerical solution of nonlinear stiff initial value problems: an analysis of one step methods.* 1985.

13 D. Grune. *On the design of ALEPH.* 1985.

14 J.G.F. Thiemann. *Analytic spaces and dynamic programming: a measure theoretic approach.* 1985.

15 F.J. van der Linden. *Euclidean rings with two infinite primes.* 1985.

16 R.J.P. Groothuizen. *Mixed elliptic-hyperbolic partial differential operators: a case-study in Fourier integral operators.* 1985.

17 H.M.M. ten Eikelder. *Symmetries for dynamical and Hamiltonian systems.* 1985.

18 A.D.M. Kester. *Some large deviation results in statistics.* 1985.