LARGE DEVIATIONS AND ASYMPTOTIC EFFICIENCIES

P. GROENEBOOM

ACKNOWLEDGEMENTS

CONTENTS

GENERAL INTRODUCTION

Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed (i.i.d.) random variables taking values in $\mathbb{R}$ according to a distribution function F and let $\hat{F}_n$ be the empirical distribution function of $X_1, \ldots, X_n$. SANOV (1957) has given conditions under which

$$(0.1) \qquad \log \Pr\{\hat{F}_n \in \Omega_n\} = -nK(\Omega_n, F) + O(n), \qquad n \to \infty,$$

where Ω_n is a set of distribution functions and $K(\Omega_n, F)$ is the Kullback-Leibler (information) number of Ω_n with respect to F (for definitions see chapter I). If $\lim_{n \to \infty} K(\Omega_n, F) = c > 0$, then (0.1) implies that $\Pr\{\hat{F}_n \in \Omega_n\}$ comverges to zero exponentially fast. In this case the probabilities $\Pr\{\hat{F}_n \in \Omega_n\}$ are called *probabilities of large deviations*.

More generally, let S be a Hausdorff topological space, let Λ be the set of probability measures on the Borel field $\mathcal{B}$ of S and let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables taking values in S according to a probability measure $P \in \Lambda$. Let $\hat{P}_n$ be the empirical probability measure of the random variables $X_1, \ldots, X_n$. The so-called *Sanov problem* is the problem of finding conditions under which

$$(0.2) \qquad \log \Pr\{\hat{P}_n \in \Omega_n\} = -nK(\Omega_n, P) + O(n), \qquad n \to \infty,$$

where Ω_n is a set of probability measures and $K(\Omega_n, P)$ is the Kullback-Leibler number of Ω_n with respect to P. This problem has been studied by many authors, see e.g. BOROVKOV (1967), HOADLEY (1967), HOEFFDING (1967), STONE (1974), DONSKER & VARADHAN (1975,1976), SIEVERS (1976) and BAHADUR & ZABELL (1979).

A new approach to the Sanov problem is the subject of the first chapter of this monograph. Some results of this chapter have first appeared in GROENEBOOM (1976); a large part of chapter I is also contained in GROENEBOOM, OOSTERHOFF & RUYMGAART (1979), denoted by GOR (1979) in the sequel. The key idea in our attack on the Sanov problem is the introduction of a suitable topology on the set of probability measures. This topology, to be called the τ-topology, is the topology of convergence on all Borel sets. It provides a natural link between multinomial approximations to the probabilities

$Pr\{\hat{P}_n \in \Omega_n\}$ and approximations of the form

$$\inf_{\underline{Q} \in \Omega_n} \sum_{i=1}^{m} Q(B_i) \, \log(Q(B_i)/P(B_i))$$

to the Kullback-Leibler numbers $K(\Omega_n, P)$, where $\{B_1, \ldots, B_m\}$ is a partition of S into Borel measurable sets. The τ-topology is finer than the weak (or vague) topology, but coarser than the topology induced by the total variation metric. Since this topology is in general not first countable, we shall occasionally have to deal with nets and filters.

Incidentally, in chapter I we also give a proof of Theorem 11 in SANOV (1957), generally called "Sanov's theorem". Sanov's own proof of this theorem is somewhat obscure (see section 1.4).

We use our results on the Sanov problem to derive asymptotic expressions for probabilities of large deviations of linear or continuous functions of empirical probability measures. For example, if $S = \mathbb{R}$ and Λ^* is the set of probability measures on $\mathbb{R}$ with compact support, sample means are given by $T(\hat{P}_n)$, where the linear function $T : \Lambda^* \to \mathbb{R}$ is defined by

$$T(Q) = \int_{\mathbb{R}} x \, dQ(x), \quad Q \in \Lambda^*.$$

Defining $\Omega_r = T^{-1}([r, \infty))$, we obtain from our results concerning the Sanov problem

$$\log Pr\{n^{-1} \sum_{i=1}^{n} X_i \geq r\} = \log Pr\{T(\hat{P}_n) \geq r\}$$

$$= \log Pr\{\hat{P}_n \in \Omega_r\} = -nK(\Omega_r, P) + \mathcal{O}(n), \quad n \to \infty.$$

This asymptotic relation is a version of a theorem on probabilities of large deviations of sample means in CHERNOFF (1952). In fact, there is an extensive literature on probabilities of large deviations of sample means (for references see BAHADUR & ZABELL (1979)) and it is interesting that many results in this field can be derived from Sanov-type theorems. Conversely, it is also possible to derive Sanov-type theorems from theorems on probabilities of large deviations of sample means; this last approach is followed in BAHADUR & ZABELL (1979).

Whereas sample means are *linear* functions of empirical probability measures, trimmed means and signed rank statistics provide examples of

continuous functions of empirical probability measures (if Λ is endowed with the τ-topology). Theorems on probabilities of large deviations of these statistics will be given in section 1.6 and chapter II, respectively.

Our interest in probabilities of large deviations of signed rank statistics is motivated by the role played by large deviation probabilities in the definition of Bahadur efficiencies of signed rank tests. More precisely, Bahadur efficiencies of such tests can be derived from the exponential rate of convergence to zero of probabilities of large deviations of the signed rank statistics under the null hypothesis. This derivation is given in chapter II. In the case of samples from purely discrete distributions a comparison by means of Bahadur efficiency of different methods for dealing with ties in signed rank tests is of particular interest (see section 2.3).

In chapter III the approach is quite different. Here we derive asymptotic efficiencies and deficiencies of some well-known multivariate tests. Whereas the (relative) asymptotic *efficiency* of two tests describes the first order asymptotic behavior of *ratios* of sample sizes required to attain certain criteria, the asymptotic *deficiency* describes the asymptotic behavior of *differences* of sample sizes required to attain such criteria. Although it might be possible to apply the theory of chapter I to the computation of Bahadur efficiencies of multivariate tests, this approach does not appear to be promising. Moreover, the results of chapter I only yield first order terms in an expansion of logarithms of large deviation probabilities, whereas higher order terms are needed in the computation of deficiencies. Hence we rely on Laplace's method for the expansion of multi-dimensional integrals to describe the asymptotic behavior of probabilities of large and moderate deviations under the null hypothesis and we use the method of the stationary phase to derive asymptotic expansions of the distribution functions under fixed and local alternatives.

Some doubt exists as to the value of the concepts of Bahadur efficiency and deficiency. The main objection to these concepts is that they do not give equal weight to error probabilities of the first and second kind. For example, Bahadur efficiency is based on the exponential rate of convergence to zero of error probabilities of the first kind, when the alternative and power are kept fixed and the sample size tends to infinity. However, there exist intimate connections between Bahadur efficiency and Pitman efficiency (see e.g. chapter II), although in the usual definition of Pitman efficiency the error probabilities of the first and second kind both tend to limits bounded away from 0 and 1 as the sample size tends to infinity. We show

in chapter III that there also exists a (somewhat more complicated) rela-
tionship between Bahadur deficiency and Hodges-Lehmann deficiency for the
tests studied here.

Of course, for the practical statistician it is important to know how
well the asymptotic efficiencies and deficiencies estimate the actual ratios
or differences of sample sizes required to attain certain criteria of test
performance. Some results in this direction will be reported in GROENEBOOM
& OOSTERHOFF (1980).

The present monograph is a virtually unaltered version of the author's
thesis written under the supervision of Prof. J. Oosterhoff. After the com-
pletion of this thesis, a paper by HSIEH (1979) appeared, in which the exact
slopes of the tests discussed in chapter III are derived. However, the me-
thods of proof used in HSIEH (1979) are quite different from the methods
used in chapter III and this chapter also contains higher order terms of the
asymptotic expansions whereas HSIEH (1979) does not consider higher order
terms. Perhaps the most interesting result of chapter III is that (non-local)
asymptotic optimality of the likelihood ratio test does not hold when higher
order terms are taken into consideration.

CHAPTER I

LARGE DEVIATIONS OF
EMPIRICAL PROBABILITY MEASURES

1.1. INTRODUCTION

Let S be a Hausdorff space and let $\mathcal{B}$ be the σ-field of Borel sets in
S. Let Λ be the set of all probability measures (pms) on $\mathcal{B}$; the abbrevia-
tion pm(s) is used in analogy with the notation df(s) for distribution
function(s). For $P,Q \in \Lambda$ the Kullback-Leibler information number $K(Q,P)$ is
defined by

$$(1.1.1) \qquad K(Q,P) = \begin{cases} \int_S q \, \log q \, dP & \text{if } Q \ll P \\ \\ \infty & \text{otherwise,} \end{cases}$$

where $q = dQ/dP$. Here and in the sequel we use the conventions $\log 0 = -\infty$,
$0 \cdot (\pm\infty) = 0$ and $\log(a/0) = \infty$ if $a \geq 0$. If Ω is a subset of Λ and $P \in \Lambda$ we
define

$$(1.1.2) \qquad K(\Omega,P) = \inf_{Q \in \Omega} K(Q,P).$$

By convention $K(\Omega,P) = \infty$ if Ω is empty.

Throughout this chapter $X_1, X_2, \ldots$ is a sequence of i.i.d. random vari-
ables taking values in S according to a pm $P \in \Lambda$. For each positive inte-
ger n the empirical pm based on $X_1, \ldots, X_n$ is denoted by $\hat{P}_n$, i.e. $\hat{P}_n(B)$ is
the fraction of X_j's, $1 \leq j \leq n$, with values in the set $B \in \mathcal{B}$.

Let $S = \mathbb{R}$ and let Λ_1 be the set of pms on $(\mathbb{R}, \mathcal{B})$, endowed with the
topology ρ induced by the supremum metric

$$(1.1.3) \qquad d(Q,R) = \sup_{x \in \mathbb{R}} |Q((-\infty,x]) - R((-\infty,x])|, \qquad Q,R \in \Lambda_1.$$

Then we have the following theorem of HOADLEY (1967) specialized to the "one-sample case".

Let $P \in \Lambda_1$ be a non-atomic pm. Let T be a real-valued function on Λ_1, uniformly continuous in the topology ρ. Define

$$\Omega_r = \{Q \in \Lambda_1 : T(Q) \geq r\}$$

for each $r \in \mathbb{R}$. Then, if the function $t \to K(\Omega_t, P)$, $t \in \mathbb{R}$, is continuous at $t = r$ and $\{u_n\}$ is a sequence of real numbers tending to zero,

$$(1.1.4) \quad \lim_{n \to \infty} n^{-1} \log \Pr\{T(\hat{P}_n) \geq r + u_n\} = -K(\Omega_r, P).$$

In section 1.3 it will be shown that Hoadley's theorem can be generalized in three different directions simultaneously:

(i)   the set Λ_1 may be replaced by the set Λ of pms on a Hausdorff space S

(ii)   the uniform continuity of the function T can be weakened to continuity
       (in a convenient topology which is finer than ρ if $S = \mathbb{R}$) at each pm
       Q such that $K(Q, P) < \infty$

(iii) $P \in \Lambda$ may be an arbitrary pm, not necessarily non-atomic.

STONE (1974) has given a simpler proof of Hoadley's theorem, but under the original strong conditions. His proof can easily be adapted to cover the case of d-dimensional random variables, but other generalizations are less obvious.

A related theorem in the spirit of SANOV (1957) has been obtained by BOROVKOV (1967):

Let $P \in \Lambda_1$ be a non-atomic pm. Then, if Ω is a ρ-open subset of Λ_1 and $K(cl_\rho(\Omega), P) = K(\Omega, P)$ (where cl_ρ denotes closure in the topology ρ),

$$(1.1.5) \quad \lim_{n \to \infty} n^{-1} \log \Pr\{\hat{P}_n \in \Omega\} = -K(\Omega, P).$$

By this theorem the *uniform* continuity (in ρ) of the functional T in Hoadley's theorem can be weakened to continuity, but Borovkov relies in his proof on methods of Fourier analysis of random walks in BOROVKOV (1962) for which generalization to more general pms seems to be difficult.

In this chapter the approach to large deviations based on multinomial approximations is systematically developed. It turns out that a natural

topology on the set Λ of pms on $(S,\mathcal{B})$ is the topology τ of convergence on all Borel sets, i.e. the coarsest topology for which the map $Q \to Q(B)$, $Q \in \Lambda$, is continuous for all $B \in \mathcal{B}$. In this topology a sequence of pms $\{Q_n\}$ in Λ converges to a pm $Q \in \Lambda$, notation $Q_n \to_\tau Q$, iff $\lim_{n\to\infty} \int_S f dQ_n = \int_S f dQ$ for each bounded $\mathcal{B}$-measurable function $f: S \to \mathbb{R}$. The closure and the interior of a set $\Omega \subset \Lambda$ in the topology τ will be denoted by $cl_\tau(\Omega)$ and $int_\tau(\Omega)$, respectively.

With this notation we shall prove (Theorem 1.3.1)

Let $P \in \Lambda$ and let Ω be a subset of Λ satisfying

$$(1.1.6) \qquad K(int_\tau(\Omega),P) = K(cl_\tau(\Omega),P).$$

Then (1.1.5) *holds.*

This is a generalization of Theorem 4.5 of DONSKER & VARADHAN (1976) who obtained some related inequalities under stronger conditions. In particular they assumed that S is a polish space and that the set Ω is either open or closed in the weak topology. By the weak topology we mean the topology with subbasis elements
$\{Q \in \Lambda: |\int f dQ - \int f dQ_0| < \varepsilon\}$, $Q_0 \in \Lambda$, $f \in C_B(S)$, where $C_B(S)$ is the set of bounded continuous functions $f: S \to \mathbb{R}$. We avoid the name "topology of weak convergence" since S is merely a Hausdorff space and hence weak convergence in Λ may not be properly defined because limits are not necessarily unique. The functions f appearing in this definition are bounded and continuous; therefore the weak topology is coarser than the previously defined topology τ.

In the particular case $S = \mathbb{R}$ the topology τ is finer than ρ (Lemma 1.2.1) which in turn is finer than the weak topology. Hence any ρ-continuous (weakly continuous) functional $T: \Lambda_1 \to \mathbb{R}$ is a fortiori τ-continuous and our results on τ-continuous functionals T imply the corresponding (weaker) results for ρ-continuous (weakly continuous) functionals. In fact, by this line of argument the generalized form of Hoadley's theorem mentioned above easily follows from Theorem 1.3.1.

After some crucial lemmas in section 1.2 the basic theorems are obtained in section 1.3. The theory includes theorems of Borovkov, Donsker & Varadhan, Hoadley, Sethuraman and Stone as particular cases and thus provides a unified approach to these results which were obtained by rather different methods. Section 1.4 is devoted to a proof of the celebrated

4

theorem of Sanov mentioned in the general introduction. In section 1.5 a
large deviation result for linear functions of empirical pms is proved.
This result generalizes CHERNOFF's (1952) theorem. Counterparts of some
of the results in sections 1.3 and 1.5 have been obtained independently and
by different methods by BAHADUR & ZABELL (1979). Finally, in section 1.6
a large deviation theorem for a class of linear combinations of order sta-
tistics (L-estimators) is proved. This leads to a large deviation theorem
for trimmed means under minimal conditions.

1.2. PRELIMINARIES

In this section some notation is introduced and a few preliminary re-
sults are proved which will play an essential role in the subsequent sec-
tions. By a *partition* P of the Hausdorff space S is meant a *finite* parti-
tion of S consisting of Borel sets. Such partitions are the starting point
of the multinomial approximation on which the proof of Lemma 1.3.1 in sec-
tion 1.3 is based. The number of sets in a partition is called the *size* of
the partition. For $P,Q \in \Lambda$ and a partition $P = \{B_1,\ldots,B_m\}$ of S define

$$(1.2.1) \qquad K_P(Q,P) = \sum_{j=1}^{m} Q(B_j) \log\{Q(B_j)/P(B_j)\},$$

and for a set $\Omega \subset \Lambda$

$$K_P(\Omega,P) = \inf_{Q \in \Omega} K_P(Q,P).$$

Without explicit reference the relation

$$(1.2.2) \qquad K(Q,P) = \sup\{K_P(Q,P): P \text{ is a partition of } S\}$$

(see e.g. PINSKER (1964), section 2.4) will repeatedly be used. We shall
say that a partition P is *finer than* a partition R iff for each $B \in P$
there exists a $C \in R$ such that $B \subset C$.

For each partition $P = \{B_1,\ldots,B_m\}$ of S the pseudo-metric d_P on Λ is
defined by

$$d_P(Q,R) = \max_{1 \le j \le m} |Q(B_j) - R(B_j)|, \qquad Q,R \in \Lambda.$$

The topology τ of convergence on all Borel sets of S is generated by the family $\{d_p\colon P$ is a partition of $S\}$. A basis of this topology is provided by the collection of sets $\{R \in \Lambda\colon d_p(R,Q) < \delta\}$ where $Q \in \Lambda$, $\delta > 0$ and P runs through all partitions of S. Note that this collection is a basis and not merely a subbasis of τ.

LEMMA 1.2.1. *Let* $S = \mathbb{R}^d$. *Then the topology* ρ *induced by the supremum metric* $d(Q,R) = \sup_{x \in \mathbb{R}^d} |Q((-\infty,x]) - R((-\infty,x])|$, $Q,R \in \Lambda$, *is strictly coarser than the topology* τ.

PROOF. Since convergence in ρ of a sequence of pms does not imply convergence on all Borel sets (a sequence of purely atomic pms may converge in ρ to a non-atomic pm), it must be shown that $\rho \subseteq \tau$.

Let $\varepsilon > 0$ and let Q be a pm on $\mathbb{R}$. Then there exists a finite (possibly empty) set of points with Q-probability $\geq \frac{1}{2}\varepsilon$. Hence there exists a partition $P = \{B_1,\ldots,B_m\}$ of $\mathbb{R}$ consisting of singletons B_i such that $Q(B_i) \geq \frac{1}{2}\varepsilon$ and open or half open intervals B_j such that $Q(B_j) < \frac{1}{2}\varepsilon$. If R is a pm on $\mathbb{R}$ such that $d_p(Q,R) < \frac{1}{2}\varepsilon/m$, then $d(Q,R) < \varepsilon$, which proves the lemma for pms on $\mathbb{R}$.

Next suppose that Q is a pm on $\mathbb{R}^d$ (d>1). Let Q_i, $1 \leq i \leq d$, be the one-dimensional marginals of Q. For each Q_i there exists by the previous paragraph a partition $\{B_{i,1},\ldots,B_{i,m_i}\}$ of $\mathbb{R}$ consisting of singletons $B_{i,j}$ with $Q_i(B_{i,j}) \geq \frac{1}{2}\varepsilon$ and open or half open intervals $B_{i,j}$ with $Q_j(B_{i,j}) < \frac{1}{2}\varepsilon/d$. Let P be the partition consisting of the product sets $B_{1,j_1} \times \ldots \times B_{d,j_d}$, $1 \leq j_i \leq m_i$, $1 \leq i \leq d$, and let $m = \max_{1 \leq i \leq d} m_i$. The implication

$$d_p(Q,R) < \tfrac{1}{2}\varepsilon/dm \Rightarrow d(Q,R) < \varepsilon$$

proves the lemma for $S = \mathbb{R}^d$. □

A function T defined on Λ will be called τ-continuous if it is continuous with respect to the topology τ on Λ and the given topology on the range space. The definition of τ-(lower,upper) semicontinuity is similar. The topology of the extended real line $\overline{\mathbb{R}}$ is the usual topology generated by the sets $[-\infty,x)$, $(x,\infty]$, $x \in \mathbb{R}$.

LEMMA 1.2.2. *Let* $P \in \Lambda$. *Then the function* $Q \to K(Q,P)$, $Q \in \Lambda$, *is* τ-*lower semicontinuous.*

6

PROOF. Let $P, Q \in \Lambda$ and let c be an arbitrary real number such that
$c < K(Q,P)$. By (1.2.2) there exists a partition $\mathcal{P}$ of S such that $K_{\mathcal{P}}(Q,P) >$
$> c$. Clearly there exists $\delta > 0$ such that

$$d_{\mathcal{P}}(R,Q) < \delta \Rightarrow K(R,P) \geq K_{\mathcal{P}}(R,P) > c,$$

proving the lemma. □

A collection Γ of pms in Λ is called *uniformly absolutely continuous*
with respect to a pm $P \in \Lambda$ if for each $\varepsilon > 0$ there exists $\delta > 0$ such that
for each $Q \in \Gamma$ and each $B \in \mathcal{B}$, $P(B) < \delta \Rightarrow Q(B) < \varepsilon$.

In the next lemma some topological properties are established of a
class $\Gamma \subset \Lambda$ with uniformly bounded Kullback-Leibler numbers.

LEMMA 1.2.3. *Let* $P \in \Lambda$ *and let* $\Gamma = \{Q \in \Lambda: K(Q,P) \leq c\}$ *for some finite*
$c \geq 0$. *Then*

(a) Γ *is uniformly absolutely continuous with respect to* P

(b) Γ *is both compact and sequentially compact in the topology* τ.

PROOF.

(a) Let $\varepsilon > 0$. Let $\delta > 0$ be such that $\frac{1}{2}\varepsilon \log(\frac{1}{2}\varepsilon/\delta) > c+e^{-1}$. Then, for each
$Q \in \Gamma$ and each $B \in \mathcal{B}$ satisfying $P(B) < \delta$,

$$Q(B) = \int_B q\, dP = \int_{B \cap \{q \leq \frac{1}{2}\varepsilon/\delta\}} q\, dP + \int_{B \cap \{q > \frac{1}{2}\varepsilon/\delta\}} q\, dP$$

$$\leq \frac{1}{2}\varepsilon\delta^{-1}P(B) + (\log(\frac{1}{2}\varepsilon/\delta))^{-1} \int_{B \cap \{q > \frac{1}{2}\varepsilon/\delta\}} q \log q\, dP$$

$$< \frac{1}{2}\varepsilon + (c+e^{-1})(\log(\frac{1}{2}\varepsilon/\delta))^{-1} < \varepsilon,$$

where $q = dQ/dP$ (note that the inequality $x \log x \geq -e^{-1}$ provides an
upper bound $c + e^{-1}$ for the integral $\int_C q \log q\, dP$ for any set $C \in \mathcal{B}$).

(b) Let M be the collection of all set functions $\mu: \mathcal{B} \to [0,1]$ endowed with
the topology τ_1 of setwise convergence (note that τ is the correspond-
ing relative topology on Λ). Using the property that a Hausdorff space
is compact iff each ultrafilter converges, we first prove that M is τ_1-
compact. Consider an ultrafilter $U = \{U_\alpha: \alpha \in I\}$ on M. For each $B \in \mathcal{B}$
the image of U under the map $\mu \to \mu(B)$ is an ultrafilter on $[0,1]$ and

hence converges to a (unique) point, say $c_B \in [0,1]$. Let $\mu_0 \in M$ be defined by $\mu_0(B) = c_B$, $B \in \mathcal{B}$. Since $\mu_0 \in \cap_{\alpha \in I} \mathrm{cl}_{\tau_1}(U_\alpha)$, the ultrafilter $\mathcal{U}$ converges to μ_0, proving τ_1-compactness of M.

In order to show that Γ is τ-compact it suffices to prove that Γ is a τ_1-closed subset of M. Let $\mu \in \mathrm{cl}_{\tau_1}(\Gamma)$. Clearly μ is an additive set function. To prove σ-additivity consider a sequence $\{B_n\}$ of disjoint Borel sets. Fix $\epsilon > 0$. By part (a) there exists $\delta > 0$ such that $B \in \mathcal{B}$, $P(B) < \delta \Rightarrow Q(B) < \epsilon$ for each $Q \in \Gamma$. Choose k so large that $P(\cup_{n=k}^{\infty} B_n) = \sum_{n=k}^{\infty} P(B_n) < \delta$. Since $\mu \in \mathrm{cl}_{\tau_1}(\Gamma)$ it follows that

$$\left| \mu\left(\overset{\infty}{\underset{n=1}{\cup}} B_n \right) - \sum_{n=1}^{k} \mu(B_n) \right| = \mu\left(\overset{\infty}{\underset{n=k+1}{\cup}} B_n \right) \le \epsilon,$$

implying that $\mu(\cup_{n=1}^{\infty} B_n) = \sum_{n=1}^{\infty} \mu(B_n)$. Hence $\mu \in \Lambda$. Now Lemma 1.2.2 implies $\mu \in \Gamma$ and thus Γ is τ-compact.

Finally Γ is also sequentially compact in τ since by Theorem 2.6 of GÄNSSLER (1971) the notions "compact" and "sequentially compact" coincide for the topology τ. $\square$

Lemma 1.2.3 is closely related to the information – theoretical proofs of convergence of a sequence of pms $\{Q_n\}$ to P under the condition $K(Q_n,P) \to 0$, as $n \to \infty$ (see RÉNYI (1961) and CSISZÁR (1962)). In fact, if $K(Q_n,P) \to 0$ then $\{Q_n\}$ converges to P in the total variation metric (cf. PINSKER (1964)), which is a stronger type of convergence than convergence in τ (the convergence has to be *uniform* on all Borel sets).

Let $P,Q \in \Lambda$ and let $\mathcal{P} = \{B_1,\ldots,B_m\}$ be a partition of S. Then the $\mathcal{P}_\mathcal{P}$-*linear* pm Q' corresponding to Q is defined by

$$(1.2.3) \qquad Q'(B \cap B_i) = \begin{cases} P(B \cap B_i)Q(B_i)/P(B_i) & \text{if } P(B_i) > 0 \\ \\ Q(B \cap B_i) & \text{if } P(B_i) = 0, \end{cases}$$

$i = 1,\ldots,m$; $B \in \mathcal{B}$. The usefulness of this concept lies in its property

$$K(Q',P) = K_\mathcal{P}(Q',P) = K_\mathcal{P}(Q,P).$$

The device of $\mathcal{P}_\mathcal{P}$-linear pms was, as far as we know, first used in large deviation problems by SANOV (1957) for pms on $\mathbb{R}$. It was also used by HOADLEY (1967) and in the more general form of the preceding definition by STONE

(1974).

The next lemma generalizes relation (1.2.2) and plays a crucial role in the next sections.

LEMMA 1.2.4. *Let* P $\in$ Λ *and* Ω $\subset$ Λ *satisfy*

(1.2.4) $K(cl_\tau(\Omega),P) = K(\Omega,P).$

Then

(1.2.5) $K(\Omega,P) = \sup\{K_P(\Omega,P): P \text{ is a partition of } S\}.$

PROOF. Let $\alpha = \sup\{K_P(\Omega,P): P \text{ is a partition of } S\}$ and suppose (1.2.5) does not hold, i.e. there exists an $\eta > 0$ such that $\alpha+\eta < K(\Omega,P)$ (see (1.2.2)). Put $\Gamma = \{Q \in \Lambda: K(Q,P) \leq \alpha+\eta\}$. The set of all (finite) partitions P, ordered by $P > R$ iff P is finer than R, is a directed set. Choose for each partition P a pm $Q_P \in \Omega$ satisfying $K_P(Q_P,P) \leq \alpha+\eta$. Let Q'_P be the P_P-linear pm corresponding to Q_P. Then

$$K(Q'_P,P) = K_P(Q_P,P) \leq \alpha+\eta$$

and hence $Q'_P \in \Gamma$ for each partition P. Since Γ is compact in the topology τ by Lemma 1.2.3 there exists a $\bar{Q} \in \Gamma$ such that $\bar{Q}$ is a cluster point of the net $N = \{Q'_P: P \text{ is a partition of } S\}$.

Consider the open neighborhood $\{R \in \Lambda: d_P(R,\bar{Q}) < \varepsilon\}$ of $\bar{Q}$. Since $\bar{Q}$ is a cluster point of the net N there exists a partition $T > P$ such that $d_P(Q'_T,\bar{Q}) < \varepsilon$. If $B \in P$, then

$$Q_T(B) = \sum_{A\in T, A\subset B} Q_T(A) = \sum_{A\in T, A\subset B} Q'_T(A) = Q'_T(B).$$

Hence $d_P(Q_T,\bar{Q}) = d_P(Q'_T,\bar{Q}) < \varepsilon$, implying that $\bar{Q}$ is also a cluster point of the net $\{Q_P: P \text{ is a partition of } S\}$. Since $Q_P \in \Omega$ for each P, $\bar{Q} \in cl_\tau(\Omega)$. However, $\bar{Q} \in \Gamma \Rightarrow K(\bar{Q},P) \leq \alpha+\eta < K(\Omega,P)$ in contradiction to (1.2.4) and so (1.2.5) follows. $\square$

REMARK 1.2.1. Lemma 1.2.4 is in fact a minimax theorem since in view of (1.2.2) the result (1.2.5) can also be written as

$$\sup_{P} \inf_{Q \in \Omega} K_p(Q,P) = \inf_{Q \in \Omega} \sup_{P} K_p(Q,P).$$

REMARK 1.2.2. The following example shows that (1.2.4) is not necessary for (1.2.5), even if $K(\Omega,P) < \infty$. Let $S = [-1,\infty) \subset \mathbb{R}$, let $\Omega_1 = \{Q \in \Lambda: \int_S x dQ(x) > 0\}$ and let $P,Q_1 \in \Lambda$ be defined by $P(\{-1\}) = P(\{0\}) = \frac{1}{2}$ and $Q_1(\{-1\}) = 1$, respectively. Define $\Omega = \Omega_1 \cup Q_1$. It is easily seen that $K(\Omega_1,P) = \sup_p K_p(\Omega_1,P) = \infty$ and hence by (1.2.2) $K(\Omega,P) = K(Q_1,P) = \log 2 = \sup_p K_p(Q_1,P) = \sup_p K_p(\Omega,P)$. Obviously $P \in cl_\tau(\Omega)$ and therefore $K(cl_\tau(\Omega),P) = 0$. Thus (1.2.4) is violated but (1.2.5) holds true.

REMARK 1.2.3. Let $scl_\tau(\Omega)$ denote the sequential closure of Ω, i.e. $Q \in scl_\tau(\Omega)$ if there exists a sequence $\{Q_n\}$ in Ω such that $Q_n \to_\tau Q$. We show that (1.2.4) in Lemma 1.2.4 cannot be replaced by $K(scl_\tau(\Omega),P) = K(\Omega,P)$. Let Ω be the set of all purely atomic pms on $\mathbb{R}$ and let P be a non-atomic pm on $\mathbb{R}$. Then $\sup\{K_p(\Omega,P): P \text{ is a partition of } \mathbb{R}\} = 0$, but $K(\Omega,P) = K(scl_\tau(\Omega),P) = \infty$ since $\Omega = scl_\tau(\Omega)$. In this case $cl_\tau(\Omega) = \Lambda_1 =$ the set of all pms on $\mathbb{R}$. This shows that there are pms in Λ_1 which can be "reached" by nets in Ω but not by sequences in Ω.

By convention the *support* supp(Q) of a pm $Q \in \Lambda$ is the set of points $x \in S$ such that each neighborhood of x has positive Q-probability. Note that $Q(\text{supp}(Q))$ may be smaller than one. However, we shall say that $Q \in \Lambda$ has finite support $\{x_1,...,x_k\}$ if $Q(\{x_i\}) > 0$, $i = 1,...,k$, and $\sum_{i=1}^k Q(\{x_i\}) = 1$. In general, let us call a pm Q *Lindelöf inner regular* if $Q(B) = \sup\{Q(V): V \subset B, V \text{ Lindelöf}\}$ for all open sets $B \subset S$ (a set is called *Lindelöf* if each open cover has a countable subcover). A pm with this property assigns probability one to its support by a line of argument similar to the proof of Lemma 2.3 in BAHADUR & ZABELL (1979). This regularity condition is certainly satisfied if S is second countable.

LEMMA 1.2.5. *Let* $P \in \Lambda$. *Each pm which has finite support contained in the support of* P *belongs to the weak closure of* $\{Q \in \Lambda: K(Q,P) < \infty\}$.

PROOF. Let $Q_0 \in \Lambda$ and $\text{supp}(Q_0) = \{x_1,...,x_k\} \subset \text{supp}(P)$. We prove that each weakly open neighborhood V of Q_0 contains a pm Q_V such that $K(Q_V,P) < \infty$. Let

$$V = \{Q \in \Lambda: |\int f_j dQ - \int f_j dQ_0| < \epsilon, \quad j = 1,...,J\},$$

where $f_1, \ldots, f_J \in C_B(S)$. Choose neighborhoods U_{ji} of x_i in S such that $x \in U_{ji} \Rightarrow |f_j(x) - f_j(x_i)| < \epsilon$, $i = 1, \ldots, k$; $j = 1, \ldots, J$, where for each j the sets $U_{j1}, \ldots, U_{jk}$ are disjoint. Now put $\tilde{U}_i = \cap_{j=1}^{J} U_{ji}$, $i = 1, \ldots, k$, and define $Q_V \in \Lambda$ by

$$Q_V(B) = \sum_{i=1}^{k} Q_0(\{x_i\}) P(B \cap \tilde{U}_i) / P(\tilde{U}_i), \quad B \in \mathcal{B}.$$

Note that $P(\tilde{U}_i) > 0$ because $x_i \in \text{supp}(P)$. Obviously $K(Q_V, P) < \infty$. Moreover, $Q_V \in V$ since for $j = 1, \ldots, J$

$$\left| \int f_j dQ_V - \int f_j dQ_0 \right| \leq \sum_{i=1}^{k} \left| \int_{\tilde{U}_i} (f_j - f_j(x_i)) dQ_V \right| < \epsilon. \quad \square$$

This lemma does not continue to hold if the weak closure is replaced by the τ-closure since the τ-closure of $\{Q \in \Lambda : K(Q,P) < \infty\}$ does not contain any pm which is not absolutely continuous with respect to P. This illustrates the difference between the weak topology and the topology τ.

1.3. BASIC RESULTS

In the sequel we discuss probabilities of events of the form $\{\hat{P}_n \in \Omega\}$, $\Omega \subset \Lambda$, where the empirical pms $\{\hat{P}_n\}$ are induced by the sequence $X_1, X_2, \ldots$. The problem which events $\{\hat{P}_n \in \Omega\}$ are $\mathcal{B}^n$-measurable for all n is (at least partially) solved by

PROPOSITION 1.3.1. *Let S be a completely regular space. Let $\tilde{\Lambda}$ denote the set of pms in Λ with finite support and rational point masses. Then $\{\hat{P}_n \in \Omega\} \in \mathcal{B}^n$ for all $n \in \mathbb{N}$ iff $\Omega \cap \tilde{\Lambda} \in \tilde{W}$, where $\tilde{W}$ is the σ-field induced by W on $\tilde{\Lambda}$ and W is the Borel σ-field on Λ generated by the weak topology.*

PROOF. For $n \in \mathbb{N}$ let $\Lambda(n)$ denote the set of pms in Λ with finite support and point masses which are multiples of n^{-1} and let $W(n)$ denote the σ-field induced by W on $\Lambda(n)$.

We first prove that $\{\hat{P}_n \in \Omega\} \in \mathcal{B}^n \Longleftrightarrow \Omega \cap \Lambda(n) \in W(n)$. Consider the map $\hat{P}_n : S^n \to \Lambda(n)$ where $\hat{P}_n(x_1, \ldots, x_n)$ is the pm assigning mass n^{-1} to each x_i, $i = 1, \ldots, n$ (since the x_i's need not be distinct, there may be less than n different point masses). Let $\mathcal{B}(n)$ denote the σ-field on $\Lambda(n)$ induced by the surjection $\hat{P}_n$. Obviously $\{\hat{P}_n \in \Omega\} \in \mathcal{B}^n \Longleftrightarrow \Omega \cap \Lambda(n) \in \mathcal{B}(n)$. We show that $\mathcal{B}(n) = W(n)$.

$B(n)$ is a Borel σ-field generated by the topology with basis elements
$V(x_1^0,\ldots,x_n^0) = \{\hat{P}_n(x_1,\ldots,x_n): x_i \in U(x_i^0),\ i = 1,\ldots,n\}$ where $(x_1^0,\ldots,x_n^0)$
$\in S^n$ and $U(x_1^0),\ldots,U(x_n^0)$ are neighborhoods in S of $x_1^0,\ldots,x_n^0$, respectively,
which are disjoint for distinct x_i^0's. On the other hand, sets
$V(Q_0;f_1,\ldots,f_J) = \{Q \in \Lambda(n): |\int f_j dQ - \int f_j dQ_0| < \epsilon,\ j = 1,\ldots,J\}$, where $Q_0 \in$
$\Lambda(n)$ and $f_1,\ldots,f_j \in C_B(S)$, are basis elements of the (relative) weak
topology on $\Lambda(n)$. If the neighborhoods $U(x_i^0)$ are small enough, $V(x_1^0,\ldots,x_n^0)$
$\subset V(Q_0;f_1,\ldots,f_J)$. Conversely, for given $U(x_1^0),\ldots,U(x_n^0)$ choose $Q_0 \in \Lambda(n)$
such that $Q_0(\{x_1^0,\ldots,x_n^0\}) = 1$ and let for $i = 1,\ldots,n$ the continuous func-
tions f_i satisfy $0 \le f_i \le 1$, $f_i(x_i^0) = 1$ and $f_i(x) = 0$ if $x \notin U(x_i^0)$; such
functions exist since S is completely regular. Let $0 < \epsilon < n^{-1}$. Then
$Q \in \Lambda(n)$, $|\int f_i dQ - \int f_i dQ_0| < \epsilon \Rightarrow Q(U(x_i^0)) \ge Q_0(\{x_i^0\})$. Since this implica-
tion holds for all i, it follows that $V(Q_0;f_1,\ldots,f_n) \subset V(x_1^0,\ldots,x_n^0)$. Hence
the topologies generating $B(n)$ and $W(n)$ coincide.

It remains to prove $\Omega \cap \Lambda(n) \in W(n)$ for all $n \Leftrightarrow \Omega \cap \tilde{\Lambda} \in \tilde{W}$. The impli-
cation $\Leftarrow$ is trivial. To prove $\Rightarrow$, let $\Omega_n \in W$ be such that $\Omega \cap \Lambda(n) =$
$\Omega_n \cap \Lambda(n)$, $n \in \mathbb{N}$. Fix $m \in \mathbb{N}$. If the pm $Q \in \tilde{\Lambda} \in \Omega^c$, then $Q \in \Lambda(s) \cap \Omega^c$ for
some $s \in \mathbb{N}$, implying $Q \notin \Omega_{sm}$. Hence $\cap_{s=1}^{\infty} \Omega_{sm} \cap \tilde{\Lambda} \subset \Omega$ and thus
$\cup_{m=1}^{\infty} \cap_{s=1}^{\infty} \Omega_{sm} \cap \tilde{\Lambda} \subset \Omega \cap \tilde{\Lambda}$. Conversely, $\cup_{m=1}^{\infty} \cap_{s=1}^{\infty} \Omega_{sm} \cap \tilde{\Lambda} \supset \cup_{m=1}^{\infty} \Omega \cap \Lambda(m) =$
$\Omega \cap \tilde{\Lambda}$. It follows that $\Omega \cap \tilde{\Lambda} \in \tilde{W}$ and the proof is complete. $\square$

The collection of sets $\Omega \subset \Lambda$ satisfying $\Omega \cap \tilde{\Lambda} \in \tilde{W}$ is quite rich, much
richer than W. Henceforth it will be assumed without explicit reference
that $Pr\{\hat{P}_n \in \Omega\}$ is well defined for all $n \in \mathbb{N}$. However, in Remark 1.3.1 we
briefly return to this matter.

For each $n \in \mathbb{N}$ and $\Omega \subset \Lambda$ the set $\Omega^{(n)}$ is defined by

$$\Omega^{(n)} = \{Q \in \Omega: nQ(B) \in \mathbb{Z} \quad \text{for all } B \in B\}.$$

Our large deviation results concerning probabilities $Pr\{\hat{P}_n \in \Omega\}$ have
as starting point Lemma 1.3.1 which exploits multinomial approximations to
the distributions of the empirical pms $\hat{P}_n$. It is easily seen that the lemma
remains valid for arbitrary sets S and arbitrary σ-fields B containing all
singletons.

LEMMA 1.3.1. *Let* P $\in \Lambda$ *and let* Ω *be a subset of* Λ. *Consider the following*
conditions

12

(A) *for each c < K(Ω,P) and for all sufficiently large n there exist sets*
$\Gamma_{n,1},\ldots,\Gamma_{n,k_n} \subset \Lambda$ *and partitions* $P_{n,1},\ldots,P_{n,k_n}$ *of S with sizes*
$m_{n,1},\ldots,m_{n,k_n}$ *respectively such that*

(A1) $$\Omega^{(n)} \subset \bigcup_{i=1}^{k_n} \Gamma_{n,i}$$

(A2) $$K_{P_{n,i}}(\Gamma_{n,i},P) > c, \qquad 1 \le i \le k_n$$

(A3) $$\max_{1 \le i \le k_n} m_{n,i} = O(n/\log n)$$

   and

   $$\log k_n = O(n), \qquad n \to \infty.$$

(B) *For each ε > 0 and each sufficiently large n there exists a pm* $Q_n \in \Omega^{(n)}$
   and a partition P_n *of size* m_n *such that*

(B1) $$K_{P_n}(Q_n,P) < K(\Omega,P) + \epsilon$$

(B2) $$\{Q \in \Lambda : d_{P_n}(Q,Q_n) = 0\} \subset \Omega$$

(B3) $$m_n = O(n/\log n).$$

   If (A) *is satisfied,*

(1.3.1) $$\limsup_{n \to \infty} n^{-1} \log \Pr\{\hat{P}_n \in \Omega\} \le - K(\Omega,P).$$

   If (B) *is satisfied,*

(1.3.2) $$\liminf_{n \to \infty} n^{-1} \log \Pr\{\hat{P}_n \in \Omega\} \ge - K(\Omega,P).$$

   Hence, if both (A) *and* (B) *are satisfied,*

(1.3.3) $$\lim_{n \to \infty} n^{-1} \log \Pr\{\hat{P}_n \in \Omega\} = - K(\Omega,P).$$

PROOF. To prove the lemma it is first shown that condition (A) implies
(1.3.1). Let c < K(Ω,P). By condition (A1) we have

$$\Pr\{\hat{P}_n \in \Omega\} \leq \sum_{i=1}^{k_n} \Pr\{\hat{P}_n \in \Gamma_{n,i}\}.$$

Let

$$P_{n,i} = \{B_{n,i,1}, \ldots, B_{n,i,m_{n,i}}\}$$

and let

$$p_{n,i,j} = P(B_{n,i,j}), \qquad 1 \leq i \leq k_n, \ 1 \leq j \leq m_{n,i}.$$

Then

$$\Pr\{\hat{P}_n \in \Omega\} \leq \sum_{i=1}^{k_n} \Pr\{K_{P_{n,i}}(\hat{P}_n, P) \geq K_{P_{n,i}}(\Gamma_{n,i}, P)\}$$

$$= \sum_{i=1}^{k_n} \sum^{i_*} n! \prod_{j=1}^{m_{n,i}} \{((nz_{n,i,j})!)^{-1} p_{n,i,j}^{nz_{n,i,j}}\}$$

$$= \sum_{i=1}^{k_n} \sum^{i_*} n! \prod_{j=1}^{m_{n,i}} \{((nz_{n,i,j})!)^{-1} z_{n,i,j}^{nz_{n,i,j}}\} \cdot$$

$$\cdot \exp\{-n \sum_{j=1}^{m_{n,i}} z_{n,i,j} \log(z_{n,i,j}/p_{n,i,j})\},$$

where $\sum^{i_*}$ denotes summation over all $(z_{n,i,1}, \ldots, z_{n,i,m_{n,i}})$ such that

$$\sum_{j=1}^{m_{n,i}} z_{n,i,j} = 1, \ z_{n,i,j} \geq 0, \ nz_{n,i,j} \in \mathbb{Z}, \qquad 1 \leq j \leq m_{n,i},$$

and

$$\sum_{j=1}^{m_{n,i}} z_{n,i,j} \log(z_{n,i,j}/p_{n,i,j}) \geq K_{P_{n,i}}(\Gamma_{n,i}, P).$$

The number of points $(z_{n,i,1}, \ldots, z_{n,i,m_{n,i}})$ satisfying the first condition is equal to

$$\binom{n+m_{n,i}-1}{m_{n,i}-1} = \exp\{O(n)\}, \qquad n \to \infty,$$

uniformly in i, $1 \leq i \leq k_n$. Moreover, by Stirling's formula, as $n \to \infty$,

$$n!/\{(nz_{n,i,1})! \ldots \ldots (nz_{n,i,m_{n,i}})!\} =$$

$$= \exp\{-n \sum_{j=1}^{m_{n,i}} z_{n,i,j} \log z_{n,i,j} + \mathcal{O}(n)\}$$

again uniformly in i, $1 \leq i \leq k_n$. Hence

$$\Pr\{\hat{P}_n \in \Omega\} \leq \sum_{i=1}^{k_n} \exp\{-nK_{P_{n,i}}(\Gamma_{n,i}, P) + \mathcal{O}(n)\}$$

$$\leq k_n \exp\{-nc + \mathcal{O}(n)\},$$

implying

$$n^{-1} \log \Pr\{\hat{P}_n \in \Omega\} \leq -c + \mathcal{O}(1),$$

as $n \to \infty$. Since $c < K(\Omega, P)$ is arbitrary, (1.3.1) follows.

Conversely we prove that condition (B) implies (1.3.2). Assume $K(\Omega, P) < \infty$, since otherwise (1.3.2) is trivial. Fix $\varepsilon > 0$. By condition (B) there exists an $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ a pm $Q_n \in \Omega^{(n)}$ and a partition $P_n = \{B_{n,1}, \ldots, B_{n,m_n}\}$ of size $m_n = \mathcal{O}(n/\log n)$ can be found such that (B_1) and (B_2) are satisfied.
Put

$$z_{n,j} = Q_n(B_{n,j}), \qquad 1 \leq j \leq m_n.$$

Then for $n \geq n_0$

$$\Pr\{\hat{P}_n \in \Omega\} \geq \Pr\{d_{P_n}(\hat{P}_n, Q_n) = 0\}$$

$$= n!/\{(nz_{n,1})! \ldots (nz_{n,m_n})!\} \cdot \prod_{j=1}^{m_n} P(B_{n,j})^{nz_{n,j}},$$

where

$$\sum_{j=1}^{m_n} z_{n,j} \log\{z_{n,j}/P(B_{n,j})\} < K(\Omega, P) + \varepsilon.$$

Hence, again by Stirling's formula, as $n \to \infty$,

$$\Pr\{\hat{P}_n \in \Omega\} \geq \exp\{-nK(\Omega, P) - n\varepsilon + \mathcal{O}(n)\},$$

and (1.3.2) follows, completing the proof. □

REMARK 1.3.1. If Ω is an arbitrary subset of Λ, the event $\{\hat{P}_n \in \Omega\}$ is not necessarily measurable. But the proof of Lemma 1.3.1 is based on the inclusion

$$\{d_{P_n}(\hat{P}_n, Q_n) = 0\} \subset \Omega \subset \bigcup_{i=1}^{k_n} \{K_{P_{n,i}}(\hat{P}_n, P) \geq K_{P_{n,i}}(\Gamma_{n,i}, P)\}$$

where the sets on the left and the right are measurable. Hence, if $\overline{P^n}$ ($\underline{P}^n$) denotes the outer (inner) measure corresponding to the product measure P^n on B^n, the proof of the lemma shows that under the conditions (A) and (B)

$$\lim_{n \to \infty} n^{-1} \log \overline{P^n}\{\hat{P}_n \in \Omega\} = \lim_{n \to \infty} n^{-1} \log \underline{P}^n\{\hat{P}_n \in \Omega\} = -K(\Omega, P)$$

for *any* set $\Omega \subset \Lambda$. In this sense Lemma 1.3.1 continues to hold for arbitrary sets Ω. Similar remarks apply to all other results of this section.

LEMMA 1.3.2. *Let* $P \in \Lambda$ *and let* Ω *be a subset of* Λ. *Sufficient conditions for* (1.3.1) *and* (1.3.2) *are respectively given by*

(A') $K(\Omega, P) = \sup\{K_{\mathcal{P}}(\Omega, P) : \mathcal{P} \text{ is a partition of } S\}$

(B') $K(\Omega, P) = K(\text{int}_\tau(\Omega), P)$.

PROOF. If (A') is satisfied there exists for each $c < K(\Omega, P)$ a partition $\mathcal{P}$ such that $K_{\mathcal{P}}(\Omega, P) > c$. Hence (1.3.1) holds by the first part of Lemma 1.3.1 with $k_n = 1$ and $\Gamma_{n,1} = \Omega$.

To verify (1.3.2) assume $K(\Omega, P) < \infty$. Fix $\varepsilon > 0$. By (B') we can find a $Q \in \text{int}_\tau(\Omega)$ satisfying $K(Q, P) < K(\Omega, P) + \frac{1}{2}\varepsilon$. Since $Q \in \text{int}_\tau(\Omega)$ there exists a partition $\mathcal{P} = \{B_1, \ldots, B_m\}$ of S and a $\delta > 0$ such that

$$\{R \in \Lambda : d_{\mathcal{P}}(R, Q) < \delta\} \subset \Omega.$$

It follows that for all sufficiently large n there exist pms $Q_n \in \Lambda^{(n)}$ satisfying

(i) $d_{\mathcal{P}}(Q_n, Q) < \delta$, hence $Q_n \in \Omega$ and $\{R \in \Lambda : d_{\mathcal{P}}(R, Q_n) = 0\} \subset \Omega$

(ii) $K_{\mathcal{P}}(Q_n, P) < K_{\mathcal{P}}(Q, P) + \frac{1}{2}\varepsilon \leq K(Q, P) + \frac{1}{2}\varepsilon < K(\Omega, P) + \varepsilon$.

Hence condition (B) of Lemma 1.3.1 is satisfied, implying (1.3.2). □

STONE (1974) proves (1.3.3) under the conditions (in our notation)

(C1) $K(\Omega,P) < \infty$

For each $\varepsilon > 0$ there are a pm $Q \in \Omega$, a partition P of S and $\delta > 0$ such that

(C2) $K_p(\Omega,P) \leq K_p(Q,P) < K_p(\Omega,P) + \varepsilon$

(C3) $\{R \in \Lambda: d_p(R,Q) < \delta\} \subset \Omega$.

It turns out that if $K(\Omega,P) < \infty$ these conditions are equivalent to conditions (A') and (B') of our Lemma 1.3.2, implying that Stone's theorem is in fact equivalent to Lemma 1.3.2 if $K(\Omega,P) < \infty$.

To prove the equivalence suppose that conditions (A') and (B') are fulfilled and $K(\Omega,P) < \infty$. Fix $\varepsilon > 0$. By (B') a pm $Q \in \text{int}_\tau(\Omega)$ exists satisfying $K(Q,P) < K(\Omega,P) + \frac{1}{2}\varepsilon$. Since $Q \in \text{int}_\tau(\Omega)$, there exists a partition T and $\delta > 0$ such that $\{R \in \Lambda: d_T(R,Q) < \delta\} \subset \Omega$. By (A') there exists a partition P which is finer than T and satisfies $K(\Omega,P) < K_p(\Omega,P) + \frac{1}{2}\varepsilon$ (note that $K_T(R,P) \leq K_p(R,P)$ for each pm R if P is finer than T). Hence

$$K_p(\Omega,P) \leq K_p(Q,P) \leq K(Q,P) < K(\Omega,P) + \frac{1}{2}\varepsilon < K_p(\Omega,P) + \varepsilon.$$

Moreover, for small enough $\delta' > 0$ the implication $R \in \Lambda$, $d_p(R,Q) < \delta' \Rightarrow \Rightarrow d_T(R,Q) < \delta$ holds. It follows that conditions (C2) and (C3) of Stone are satisfied.

Conversely, suppose that Stone's conditions (C1) to (C3) hold. Then by Lemma 2.3 of STONE (1974), condition (A') also holds. Let $\varepsilon > 0$. Let a pm $Q \in \Omega$, a partition P of S and $\delta > 0$ satisfy (C2) and (C3) for this ε. Let Q' be the P_p-linear pm corresponding to Q (see (1.2.3)). Then (C3) implies Q' $\in \text{int}_\tau(\Omega)$ and (C2) yields

$$K(Q',P) = K_p(Q',P) = K_p(Q,P) < K_p(\Omega,P) + \varepsilon \leq K(\Omega,P) + \varepsilon.$$

Thus $K(\text{int}_\tau(\Omega),P) < K(\Omega,P) + \varepsilon$ for each $\varepsilon > 0$ and condition (B') follows.

Combining Lemma 1.2.4 and Lemma 1.3.2 we have

THEOREM 1.3.1. *Let* P $\in \Lambda$ *and let* Ω *be a subset of* Λ *satisfying*

(1.3.4) $K(\text{int}_\tau(\Omega),P) = K(\text{cl}_\tau(\Omega),P)$.

Then (1.3.3) *holds.*

Borovkov has shown (see (31) in BOROVKOV (1967)) that (1.3.3) holds if P is a non-atomic pm on $\mathbb{R}$, Ω is a ρ-open set and $K(\Omega,P) = K(cl_\rho(\Omega),P)$. This is a particular case of Theorem 1.3.1 in view of Lemma 1.2.1.

In their work on large deviations of Markov processes, DONSKER & VARADHAN (1975,1976) have shown that in the i.i.d. case (1.3.1) (or (1.3.2)) hold under the conditions that Ω be weakly closed (or open, respectively) and S be a polish space. Since the weak topology is coarser than the topology τ, their result is contained in Lemma 1.3.2 together with Lemma 1.2.4.

REMARK 1.3.2. Suppose $B \subset S$ is an arbitrary Borel set satisfying $P(B) = 1$. Let $\Lambda_B = \{Q \in \Lambda: Q(B) = 1\}$ and let τ_B denote the relative τ-topology on Λ_B. Then Theorem 1.3.1 remains valid if (1.3.4) is replaced by the weaker condition

$$K(int_{\tau_B}(\Omega \cap \Lambda_B),P) = K(cl_{\tau_B}(\Omega \cap \Lambda_B),P).$$

This result is an immediate consequence of Theorem 1.3.1 (replace S by B, Λ by Λ_B and τ by τ_B and note that $K(\Omega \cap \Lambda_B,P) = K(\Omega,P)$ and $Pr\{\hat{P}_n \in \Omega\} = Pr\{\hat{P}_n \in \Omega \cap \Lambda_B\}$).

REMARK 1.3.3. Theorem 1.3.1 does not hold if $int_\tau(\Omega)$ and $cl_\tau(\Omega)$ are replaced by the interior and closure of Ω w.r.t. the topology π induced by the total variation distance $\delta(Q,R) = sup\{|Q(B) - R(B)|: B \in \mathcal{B}\}$. For example, let $S = \mathbb{R}$, let Ω be the set of pms which are purely atomic and let P be a non-atomic pm. Then $K(cl_\pi(\Omega),P) = K(int_\pi(\Omega),P) = \infty$, but $lim_{n\to\infty} n^{-1} \log Pr\{\hat{P}_n \in \Omega\} = 0$, since $Pr\{\hat{P}_n \in \Omega\} = 1$ for all n. Note that $\Omega = cl_\pi(\Omega)$, since $Q_n \to_\pi Q$ for some $Q \in \Lambda_1$ and $Q_n \in \Omega (n \in \mathbb{N})$ implies that Q is concentrated on the countable union of the countable sets carrying the mass of the Q_n's.

However, $cl_\tau(\Omega) = \Lambda_1$ (see Remark 1.2.3).

COROLLARY 1.3.1. *Let S be a polish space and let* $P \in \Lambda$. *Then* (1.3.3) *holds for each weakly open convex subset* Ω *of* Λ.

PROOF. By Theorem 6.3 in PARTHASARATHY (1967) the space Λ endowed with the weak topology is a polish space. In the appendix of BAHADUR & ZABELL (1979) it is proved that each pm Q defined on the σ-algebra of Borel subsets of a polish space satisfies

$$Q(A) = \sup\{Q(K): K \subset A, K \text{ is convex and compact}\},$$

for each open convex set A; this property is called *convex-tightness* of the pm Q. Hence, for all $n \in \mathbb{N}$ the pm Q_n defined on the σ-algebra C of Borel subsets of Λ (with the weak topology) by

$$Q_n(C) = \Pr\{\hat{P}_n \in C\}, \qquad C \in C,$$

is convex-tight.

Let Ω be a convex and weakly open subset of Λ. Since a weakly open subset of Λ is also τ-open, Lemma 1.3.2 implies (1.3.2). Fix $\varepsilon > 0$. By convex-tightness, there exists for each n a convex weakly closed subset $\hat{C}_n \subset \Omega$ such that

$$n^{-1} \log \Pr\{\hat{P}_n \in \Omega\} \leq n^{-1} \log \Pr\{\hat{P}_n \in C_n\} + \varepsilon.$$

Thus, by Lemma 1.3.2, (A') and the convexity of C_n,

$$n^{-1} \log \Pr\{\hat{P}_n \in \Omega\} \leq n^{-1} \log \Pr\{\hat{P}_n \in C_n\} + \varepsilon \leq$$

$$\leq \lim_{k \to \infty} \sup (nk)^{-1} \log \Pr\{\hat{P}_{nk} \in C_n\} + \varepsilon$$

$$\leq - K(C_n, P) + \varepsilon \leq - K(\Omega, P) + \varepsilon.$$

Here the second inequality follows from the superadditivity of the function $m \to \log \Pr\{\hat{P}_m \in C_n\}$, $m \in \mathbb{N}$. This property easily follows from the convexity of C_n and the relation $\hat{P}_n = j^{-1} \sum_{i=1}^{j} \hat{P}_{k,i}$, where $n = jk$ and $\hat{P}_{k,i}$ is the empirical pm of the random variables $X_{(i-1)k+1}, \ldots, X_{ik}$, $1 \leq i \leq j$. Since $\varepsilon > 0$ is arbitrary it follows that

$$\lim_{n \to \infty} n^{-1} \log \Pr\{\hat{P}_n \in \Omega\} = - K(\Omega, P). \quad \square$$

A similar result under slightly different conditions is given in GOR (1979). Corollary 1.3.1 has first been proved by BAHADUR & ZABELL (1979). Their (quite different) proof is based on a Chernoff-type theorem for sample means.

The condition that Ω is weakly open and convex cannot be replaced by

the condition that Ω is τ-open and convex, even if $K(\text{int}_\tau(\Omega),P) < \infty$. To see
this, let $S = [0,1]$, let P be Lebesgue measure on S and let $\Omega =$
$= \{P \in \Lambda: P((0,\frac{1}{2})) > \frac{3}{4}\} \cup \{P \in \Lambda: P \text{ has at least one point mass}\}$. Then Ω is
τ-open and convex and the events $\{\hat{P}_n \in \Omega\}$ are measurable. Obviously
$\Pr\{\hat{P}_n \in \Omega\} = 1$ for all n and $\lim_{n\to\infty} n^{-1} \log \Pr\{\hat{P}_n \in \Omega\} = 0$ although $K(\Omega,P) =$
$= \frac{3}{4}\log 3 - \log 2 > 0$.

To determine the infimum $K(\Omega,P)$ appearing in the preceding results one
usually tries to find a pm $Q \in \Omega$ for which this infimum is attained. A suf-
ficient condition for the existence of such a pm Q is given in the next lemma.

LEMMA 1.3.3. *Let* $P \in \Lambda$ *and let* Ω *be a non-empty* τ*-closed set of pms in* Λ.
Then there exists a pm $Q \in \Omega$ *such that* $K(Q,P) = K(\Omega,P)$.

PROOF. We assume $K(\Omega,P) < \infty$ since otherwise any $Q \in \Omega$ achieves the equali-
ty. Let $\eta > 0$. Because Ω is τ-closed the set $\Omega \cap \{Q \in \Lambda: K(Q,P) \leq K(\Omega,P) +$
$+ \eta\}$ is compact by Lemma 1.2.3. By Lemma 1.2.2 the map $Q \to K(Q,P)$, $Q \in \Lambda$,
is τ-lower semicontinuous. Since a lower semicontinuous function attains its
infimum on a compact set, the proof is complete. $\square$

A similar result is proved in CSISZÁR (1975), where Ω is required to
be convex and closed in the topology of the total variation metric.

Next we specialize Theorem 1.3.1 by considering sets Ω induced by an
extended real-valued function $T: \Lambda \to \bar{\mathbb{R}}$. For a fixed function $T: \Lambda \to \bar{\mathbb{R}}$,
let

$$\Omega_t = \{Q \in \Lambda: T(Q) \geq t\}, \quad t \in \mathbb{R}.$$

We first prove a technical lemma.

LEMMA 1.3.4. *Let* $P \in \Lambda$ *and let* $T: \Lambda \to \bar{\mathbb{R}}$ *be a function which is* τ*-upper
semicontinuous on the set* $\Gamma = \{Q \in \Lambda: K(Q,P) < \infty\}$. *Then the function*
$t \to K(\Omega_t,P)$, $t \in \mathbb{R}$, *is continuous from the left.*

PROOF. Let $\kappa: \mathbb{R} \to \bar{\mathbb{R}}$ denote the function defined by $t \to K(\Omega_t,P)$, $t \in \mathbb{R}$.
Let $\{r_m\}$ be a sequence in $\mathbb{R}$ such that $r_m \uparrow r$ for some $r \in \mathbb{R}$ satisfying
$\kappa(r) < \infty$. Since κ is nondecreasing $\kappa(r_m) \leq \kappa(r) < \infty$ for each $m \in \mathbb{N}$ and
$\lim_{m\to\infty} \kappa(r_m)$ exists. For each $m \in \mathbb{N}$ there exists by Lemma 1.3.3 a pm
$Q_m \in \Omega_{r_m}$ such that $K(Q_m,P) = \kappa(r_m)$ (note that $\{Q \in \Lambda: T(Q) \geq t$ and $K(Q,P) \leq$
$\leq M\}$ is τ-closed for each $t \in \mathbb{R}$ and $M \geq 0$). Since $K(Q_m,P) \leq \kappa(r)$ for each

m, Lemmas 1.2.2 and 1.2.3 imply the existence of a subsequence $\{Q_{m_j}\}$ of $\{Q_m\}$ and a pm $Q \in \Lambda$ such that $Q_{m_j} \to_\tau Q$ and $K(Q,P) \leq \lim\inf_{j \to \infty} K(Q_{m_j},P) < \infty$. It follows that $T(Q) \geq r$ since T is upper semicontinuous on Γ and since $T(Q_{m_j}) \geq r_{m_j}$ for each $j \in \mathbb{N}$. Hence $Q \in \Omega_r$ and $\kappa(r) \leq K(Q,P) \leq$

$\leq \lim_{j \to \infty} K(Q_{m_j},P) = \lim_{m \to \infty} \kappa(r_m) \leq \kappa(r)$. Thus $\lim_{m \to \infty} \kappa(r_m) = \kappa(r)$ follows.

The left continuity also holds for a point $r \in \mathbb{R}$ such that $\kappa(r) = \infty$ and $\kappa(r') < \infty$ for all $r' < r$. For if $\{\kappa(r_m)\}_{m=1}^{\infty}$ is uniformly bounded for a sequence $\{r_m\}$ with $r_m \uparrow r$, then by the preceding line of argument there exists a pm $Q \in \Omega_r$ satisfying $K(Q,P) < \infty$ in contradiction to $\kappa(r) = \infty$. $\quad\Box$

THEOREM 1.3.2. *Let* $P \in \Lambda$ *and let* $T: \Lambda \to \bar{\mathbb{R}}$ *be a function which is* τ-*continuous at each* $Q \in \Gamma = \{R \in \Lambda: K(R,P) < \infty\}$. *Then, if the function* $t \to K(\Omega_t,P)$, $t \in \mathbb{R}$, *is continuous from the right at* $t = r$ *and if* $\{u_n\}$ *is a sequence of real numbers such that* $\lim_{n \to \infty} u_n = 0$,

$$(1.3.6) \qquad \lim_{n \to \infty} n^{-1} \log \Pr\{T(\hat{P}_n) \geq r+u_n\} = -K(\Omega_r,P).$$

(Note that the continuity property of T *is stronger than the property* "T *is continuous on* Γ".)

PROOF. Again define the function κ by $\kappa(t) = K(\Omega_t,P)$. Since κ is nondecreasing it has at most countably many discontinuities. It is continuous from the left by Lemma 1.3.4 and continuous from the right at $t = r$ by assumption.

Let $K(\Omega_r,P) < \infty$. Then there exists for each $\varepsilon > 0$ a $\delta > 0$ such that $\kappa(r) - \varepsilon < \kappa(r-\delta) \leq \kappa(r) \leq \kappa(r+\delta) \leq \kappa(r) + \varepsilon$, where κ is continuous at $r-\delta$ and $r+\delta$.

The continuity of T at each $Q \in \Gamma$ implies $\mathrm{cl}_\tau(\Omega_t) \cap \Gamma = \Omega_t \cap \Gamma$. Hence

$$K(\mathrm{cl}_\tau(\Omega_t),P) = K(\mathrm{cl}_\tau(\Omega_t) \cap \Gamma,P) = K(\Omega_t \cap \Gamma,P) = K(\Omega_t,P).$$

Moreover, if κ is continuous from the right at t,

$$K(\Omega_t,P) = K(\Omega_t \cap \Gamma,P) = K(\mathrm{int}_\tau(\Omega_t) \cap \Gamma,P) = K(\mathrm{int}_\tau(\Omega_t),P),$$

since $\Gamma \cap \Omega_{t+\gamma} \subset \{Q \in \Gamma: T(Q) > t\} \subset \Gamma \cap \mathrm{int}_\tau(\Omega_t)$ for each $\gamma > 0$. Hence by Theorem 1.3.1

$$-\kappa(r) - \varepsilon < -\kappa(r+\delta) = \lim_{n\to\infty} n^{-1} \log \Pr\{T(\hat{P}_n) \geq r+\delta\}$$

$$\leq \liminf_{n\to\infty} n^{-1} \log \Pr\{T(\hat{P}_n) \geq r+u_n\}$$

$$\leq \limsup_{n\to\infty} n^{-1} \log \Pr\{T(\hat{P}_n) \geq r+u_n\}$$

$$\leq \limsup_{n\to\infty} n^{-1} \log \Pr\{T(\hat{P}_n) \geq r-\delta\}$$

$$= -\kappa(r-\delta) < -\kappa(r) + \varepsilon.$$

Thus

$$\lim_{n\to\infty} n^{-1} \log \Pr\{T(\hat{P}_n) \geq r+u_n\} = -\kappa(r) = -K(\Omega_r, P).$$

The case $K(\Omega_r, P) = \infty$ may be dealt with along the same lines. The details are omitted. $\square$

REMARK 1.3.4. Theorem 1.3.2 continues to hold if T is an $\mathbb{R}^d$-valued function and r and $\{u_n\}$ are vectors in $\mathbb{R}^d$. The proof is quite similar.

EXAMPLE 1.3.1. Let F be a class of continuous $\mathbb{R}^d$-valued functions defined on the Hausdorff space S and compact in the compact-open topology. Let $P \in \Lambda$ be tight and assume that the one-dimensional marginals of Pf^{-1} are non-atomic for each $f \in F$. Let $d(Qf^{-1}, Rf^{-1})$ be the distance between Qf^{-1} and Rf^{-1} defined in Lemma 1.2.1.

SETHURAMAN (1964) proves (in the case that S is a polish space) that for each ε, $0 < \varepsilon < 1$,

$$(1.3.7) \qquad \lim_{n\to\infty} n^{-1} \log \Pr\{\sup_{f\in F} d(\hat{P}_n f^{-1}, Pf^{-1}) \geq \varepsilon\} = -\kappa(\varepsilon),$$

where

$$\kappa(\varepsilon) = \min_{0<p\leq 1-\varepsilon} \{(p+\varepsilon)\log((p+\varepsilon)/p) + (1-p-\varepsilon)\log((1-p-\varepsilon)/(1-p))\}.$$

Here we prove that the function $T: \Lambda \to \mathbb{R}$ defined by $T(Q) = \sup_{f\in F} d(Qf^{-1}, Pf^{-1})$ is τ-continuous at each $Q \in \Gamma$ satisfying $K(Q,P) < \infty$ and hence that (1.3.7) follows from Theorem 1.3.2.

Let $Q \in \Lambda$ satisfy $K(Q,P) < \infty$ and suppose that T is not continuous at

Q. Then there exists an $\varepsilon > 0$ such that for each τ-open neighborhood U of Q a pm $Q_U \in U$ and a function $f_U \in F$ can be found satisfying

$$(1.3.8) \qquad d(Q_U f_U^{-1}, Q f_U^{-1}) \geq \varepsilon.$$

(Note that for all pms $R, R' \in \Lambda$ one has $|T(R)-T(R')| \leq \sup_{f \in F} d(Rf^{-1}, R'f^{-1})$.)
Let the set $\mathcal{D} = \{U: U \text{ is a } \tau\text{-open neighborhood of } Q\}$ be directed by $U > V$ iff $U \subset V$. With this (partial) ordering on the set $\mathcal{D}$, $\{f_U: U \in \mathcal{D}\}$ and $\{Q_U: U \in \mathcal{D}\}$ are nets in F and Λ respectively. Since F is compact in the compact-open topology, the net $\{f_U: U \in \mathcal{D}\}$ has a cluster point $f \in F$.

Let for $x = (x^{(1)}, \ldots, x^{(d)}) \in \mathbb{R}^d$ the norm of x be defined by $\|x\| = \max_{1 \leq i \leq d} |x^{(i)}|$ and let $x \leq y$ iff $x^{(i)} \leq y^{(i)}$, $1 \leq i \leq d$. Since P is tight and $K(Q,P) < \infty$, Q is tight and hence there exists a compact set $K \subset S$ such that $Q(S \backslash K) < \frac{1}{4}\varepsilon$. The pm Qf^{-1} has non-atomic marginals since Pf^{-1} has non-atomic marginals and $Q \ll P$. Hence there exists an $\eta > 0$ such that

$$|Q\{s \in K: f(s) \leq x\} - Q\{s \in K: f(s) \leq y\}| < \frac{1}{4}\varepsilon$$

if $\|x-y\| < \eta$. By Lemma 1.2.1 we can choose a τ-open neighborhood U_0 of Q such that $d(Rf^{-1}, Qf^{-1}) < \frac{1}{4}\varepsilon$ and $R(S \backslash K) < \frac{1}{4}\varepsilon$ if $R \in U_0$. Since f is a cluster point of the net $\{f_U: U \in \mathcal{D}\}$ there exists a τ-open neighborhood $U \subset U_0$ of Q such that $\sup_{s \in K} \|f_U(s) - f(s)\| < \eta$. Because $Q_U \in U \subset U_0$ one has

$$d(Q_U f_U^{-1}, Q f_U^{-1}) \leq \max\{Q(S\backslash K), Q_U(S\backslash K)\} +$$

$$+ \sup_{x \in \mathbb{R}^d} |Q_U\{s \in K: f_U(s) \leq x\} - Q\{s \in K: f_U(s) \leq x\}|$$

$$< \sup_{x \in \mathbb{R}^d} |Q_U\{s \in K: f(s) \leq x\} - Q\{s \in K: f(s) \leq x\}| + \frac{1}{2}\varepsilon$$

$$< d(Q_U f^{-1}, Q f^{-1}) + \frac{3}{4}\varepsilon < \varepsilon.$$

This contradicts (1.3.8) and hence T is τ-continuous at Q. Let $\Omega_\varepsilon = \{Q \in \Lambda: T(Q) \geq \varepsilon\}$ for $0 < \varepsilon < 1$. It has been shown by HOEFFDING (1967) that $K(\Omega_\varepsilon, P) = \kappa(\varepsilon)$ and that κ is continuous in ε for $0 < \varepsilon < 1$. Thus (1.3.7) follows from Theorem 1.3.2.

For one sample Theorem 1 in HOADLEY (1967) is a particular case of our Theorem 1.3.2. In Hoadley's theorem $S = \mathbb{R}$, P is a non-atomic pm on $\mathbb{R}$ and

T is a real-valued *uniformly* continuous function with respect to the topology ρ.

Actually HOADLEY (1967) proves a more general theorem where T is not merely a function of one but of several empirical pms. This setup is of interest in problems concerning k samples. The results obtained so far in this section can also be generalized to the k-sample case. We briefly indicate how this works out.

Let $X_{i,1},\ldots,X_{i,n_i}$ be i.i.d. random variables taking values in S according to a pm $P_i \in \Lambda$, $1 \le i \le k$, and assume that the sample sizes n_i tend to infinity in such a way that $\lim_{N\to\infty} n_i/N = \nu_i$, where $N = \sum_{i=1}^{k} n_i$ and $\nu_i > 0$, $1 \le i \le k$. (We remark in passing that the condition $n_i/N - \nu_i = = \mathcal{O}(N^{-1} \log N)$ in HOADLEY (1967) is unnecessarily restrictive.) The empirical pm of the i-th sample will be denoted by $\hat{P}_{i,n_i}$, $1 \le i \le k$. Λ is endowed with the topology τ and Λ^k is given the product topology.

Let $P = (P_1,\ldots,P_k) \in \Lambda^k$ and $\nu = (\nu_1,\ldots,\nu_k) \in (0,1]^k$ where $\sum_{i=1}^{k} \nu_i = 1$. Let $\mathcal{P} = \mathcal{P}_1 \times \ldots \times \mathcal{P}_k$ be a partition of S^k consisting of product sets $B_{1,j_1} \times \ldots \times B_{k,j_k}$ where B_{i,j_i} belongs to a partition $\mathcal{P}_i$ of S for $1 \le i \le k$. Then we define for $Q = (Q_1,\ldots,Q_k) \in \Lambda^k$ and a set $\Omega \subset \Lambda^k$

$$I_\nu(Q,P) = \sum_{i=1}^{k} \nu_i K(Q_i,P_i), \qquad I_\nu(\Omega,P) = \inf_{Q \in \Omega} I_\nu(Q,P)$$

and

$$I_{\nu,\mathcal{P}}(Q,P) = \sum_{i=1}^{k} \nu_i K_{\mathcal{P}_i}(Q_i,P_i), \qquad I_{\nu,\mathcal{P}}(\Omega,P) = \inf_{Q \in \Omega} I_{\nu,\mathcal{P}}(Q,P).$$

By making small changes in the proofs of Theorems 1.3.1 and 1.3.2 one obtains the following corollaries.

COROLLARY 1.3.2. *Let* $P = (P_1,\ldots,P_k) \in \Lambda^k$ *and* $\Omega \subset \Lambda^k$ *satisfy*

$$I_\nu(\text{int}(\Omega),P) = I_\nu(\text{cl}(\Omega),P).$$

Then

$$\lim_{N\to\infty} N^{-1} \log \Pr\{(\hat{P}_{1,n_1},\ldots,\hat{P}_{k,n_k}) \in \Omega\} = -I_\nu(\Omega,P).$$

COROLLARY 1.3.3. *Let* $P = (P_1,\ldots,P_k) \in \Lambda^k$, *let* $T: \Lambda^k \to \bar{\mathbb{R}}$ *be continuous at each* $Q \in \Gamma = \{R \in \Lambda^k: I_\nu(R,P) < \infty\}$ *and let* $\Omega_t = \{Q \in \Lambda^k: T(Q) \ge t\}$, $t \in \mathbb{R}$.

Then, if the function $t \to I_\nu(\Omega_t, P)$ *is continuous from the right at* $t = r$
and if $\{u_N\}$ *is a sequence of real numbers such that* $u_N \to 0$,

$$\lim_{N \to \infty} N^{-1} \log \Pr\{T(\hat{P}_{1,n_1}, \ldots, \hat{P}_{k,n_k}) \geq r + u_N\} = -I_\nu(\Omega_r, P).$$

1.4. A THEOREM OF SANOV

In this section we shall show that Theorem 11 in SANOV (1957) is a
special case of our Lemma 1.3.1. This is of interest since some obscure
points in SANOV's (1957) paper have raised doubt as to the validity of his
Theorem 11 (cf. HOADLEY (1967), BAHADUR (1971)). We shall show that Sanov's
theorem holds with the original definitions.

Let $S = \mathbb{R}$ and let D be the set of distribution functions (dfs) on $\mathbb{R}$.
If $G \in D$, the corresponding pm in Λ_1 will be denoted by P_G. For convenience
of notation we write $K(G, F)$ instead of $K(P_G, P_F)$ and $K(\Omega, F)$ to denote
$\inf\{K(G, F): G \in \Omega\}$ if $\Omega \subset D$ (with similar conventions for $K_P(G, F)$ and
$K_P(\Omega, F)$). If $\Omega \subset D$, then $\Omega^{(n)} = \{G \in \Omega: nG(x) \in \mathbb{Z} \text{ for all } x \in \mathbb{R}\}$. A set
$\Omega \subset D$ will be called τ-open if the set of pms $\{P_G \in \Lambda_1: G \in \Omega\}$ is open in
the topology τ defined on Λ_1. The topology τ on D is defined by these τ-open
sets.

Suppose that P is a partition of $\mathbb{R}$ consisting of the sets
$B_1 = (-\infty, x_1)$, $B_2 = [x_1, x_2), \ldots,$ $B_{m-1} = [x_{m-2}, x_{m-1})$, $B_m = [x_{m-1}, \infty)$, where
$-\infty < x_1 < x_2 < \ldots < x_{m-1} < \infty$. Let the set W_m be defined by
$W_m = \bigcup_{i=1}^m (B_i \times [a_i, b_i])$, where $0 = a_1 \leq a_2 \leq \ldots \leq a_m \leq 1$, $0 \leq b_1 \leq b_2 \leq \ldots$
$\leq b_m = 1$ and $b_1 - a_2 \geq 0$, $b_2 - a_3 > 0, \ldots, b_{m-2} - a_{m-1} > 0$, $b_{m-1} - a_m \geq 0$.
Finally let $G_1 \in D$ be defined by $G_1(x) = a_i$, if $x \in B_i$. Then $V_m =$
$= \{G \in D: (x, G(x)) \in W_m \text{ for all } x \in \mathbb{R}\} \setminus \{G_1\}$ is called an ε-*neighborhood*.
The partition P is called the *partition corresponding to* V_m.

DEFINITION 1.4.1. Let $F \in D$ assume infinitely many values. A set $\Omega \subset D$ is
called F-*distinguishable* if the following conditions hold

(a) $K(\Omega, F) < \infty$

(b) for each $\eta > 0$ and each $n \in \mathbb{N}$ there exists a finite number $k_n = k_n(\eta)$
of ε-neighborhoods $V_{m_1}, \ldots, V_{m_{k_n}}$ such that $\Omega^{(n)} \subset \bigcup_{i=1}^{k_n} V_{m_i}$ and $K(V_{m_i}, F) >$
$> K(\Omega, F) - \eta$, $1 \leq i \leq k_n$. Moreover, $\log k_n(\eta) = 0(n)$ and $\max_{1 \leq i \leq k_n} m_i =$
$= 0(n/\log n)$, as $n \to \infty$.

(c) for each $\eta > 0$ there exists an ε-neighborhood $V_m \subset \Omega$ satisfying

$$K(V_m,F) < K(\Omega,F) + \eta.$$

The following theorem is Theorem 11 in SANOV (1957).

THEOREM 1.4.1. _Let_ $F \in D$ _assume infinitely many values, let_ $\Omega \subset D$ _be F-distinguishable and let_ $X_1,X_2,\ldots$ _be a sequence of i.i.d. random variables with df F. Then, if_ $\hat{F}_n$ _is the empirical df of_ $X_1,\ldots,X_n$,

$$\lim_{n\to\infty} n^{-1} \log \Pr\{\hat{F}_n \in \Omega\} = -K(\Omega,F).$$

PROOF. We shall first show that condition (B) of Lemma 1.3.1 is satisfied. Fix $\varepsilon > 0$. By condition (c) of Definition 1.4.1 there exists an ε-neighborhood $V_m \subset \Omega$ satisfying

$$(1.4.1) \qquad K(V_m,F) < K(\Omega,F) + \frac{1}{3}\varepsilon.$$

Choose $G \in V_m$ such that $K(G,F) < K(V_m,F) + \frac{1}{3}\varepsilon$. If P is the partition corresponding to V_m we have

$$(1.4.2) \qquad K_P(G,F) \leq K(G,F) < K(V_m,F) + \frac{1}{3}\varepsilon.$$

By the definition of ε-neighborhoods it is clear that for sufficiently large n there exists a df $G_n \in V_m^{(n)}$ satisfying

$$(1.4.3) \qquad |K_P(G_n,F) - K_P(G,F)| < \frac{1}{3}\varepsilon.$$

Moreover

$$(1.4.4) \qquad \{H \in D: d_P(P_H,P_{G_n}) = 0\} \subset V_m \subset \Omega.$$

It follows from (1.4.1) to (1.4.4) that condition (B) of Lemma 1.3.1 is satisfied.

On the other hand condition (b) of Definition 1.4.1 implies condition (A) of Lemma 1.3.1, for the ε-neighborhoods V_{m_i} of condition (b) correspond to sets of pms $\Gamma_{n,i} = \{P_G \in \Lambda_1 : G \in V_{m_i}\}$ in condition (A) of Lemma 1.3.1 and we have $K_P(V_m,F) = K(V_m,F)$ if P is the partition corresponding to an ε-neighborhood V_m.

To prove this last property we note that if $G \in V_m$ and if P'_G is the P_{P_F}-linear pm corresponding to P_G, then the df G' induced by P'_G is an element of V_m and therefore

$$K_p(G,F) = K_p(G',F) = K(G',F) ,$$

implying

$$K_p(V_m,F) \geq K(V_m,F) .$$

Since always $K_p(V_m,F) \leq K(V_m,F)$ it follows that $K_p(V_m,F) = K(V_m,F)$. Hence condition (A) of Lemma 1.3.1 is satisfied. $\square$

REMARK 1.4.1. From the preceding proof it is clear that the condition "F assumes infinitely many values" is redundant in Sanov's theorem.

The crucial property of the ε-neighborhoods V_m seems to be that $K(V_m,F) = K_p(V_m,F)$ for the partition P corresponding to V_m. It was shown in Lemma 1.3.1 that we do not have to impose this condition on the sets V_m and that we can look at the quantities $K_p(V_m,F)$ *directly* for suitably chosen sets V_m which do not necessarily have the property $K_p(V_m,F) = K(V_m,F)$.

1.5. LINEAR FUNCTIONS OF EMPIRICAL PMS

Several important statistics are in fact linear functions of empirical pms. For example, if $S = \mathbb{R}$, the sample mean $n^{-1} \sum_{i=1}^{n} X_i$ may be written as $T(\hat{P}_n)$, where T is defined by

$$T(Q) = \int_{\mathbb{R}} x \, dQ(x)$$

for all $Q \in \Lambda$ with bounded support. Note that T is a *linear* function, i.e. $T(\alpha Q + (1-\alpha)R) = \alpha T(Q) + (1-\alpha)T(R)$, $0 \leq \alpha \leq 1$. Although T is not τ-continuous at any pm Q, T is τ-continuous on each set $\{Q \in \Lambda: Q([-M,M]) = 1\}$, where M is a fixed positive number. This property suggests that large deviation theorems might be obtained by first truncating the underlying pm and subsequently taking limits, letting the support of the truncated pm tend to S. It turns out that this kind of truncation is more convenient than truncation of functionals T. Slightly different truncation arguments are systematically used in BAHADUR (1971) and HOADLEY (1967).

For the purpose of truncation we introduce conditional pms. If $B \subset S$

is a Borel set and $Q \in \Lambda$ satisfies $Q(B) > 0$, the *conditional* pm Q_B is defined by $Q_B(C) = Q(C|B)$, $C \in \mathcal{B}$. For $\Gamma \subset \Lambda$ and $B \in \mathcal{B}$ with $P(B) > 0$, we write $\Pr\{\hat{P}_n \in \Gamma \mid B\}$ to denote $\Pr\{\hat{P}_n \in \Gamma \mid X_i \in B, 1 \le i \le n\}$.

The following lemma explains why truncation is a useful approach.

LEMMA 1.5.1. *Let* $P \in \Lambda$ *and let* $B_1 \subset B_2 \subset \dots$ *be an increasing sequence of Borel sets in S such that* $\lim_{m \to \infty} P(B_m) = 1$. *Let* $\Lambda^* = \{Q \in \Lambda: Q(B_m) = 1$ *for an* $m \in \mathbb{N}\}$. *Then, for each subset* Ω *of* Λ^*

$$\lim_{m \to \infty} K(\Omega, P_{B_m}) = K(\Omega, P).$$

PROOF. Fix $\varepsilon > 0$. Let $m_0 \in \mathbb{N}$ be so large that $|\log P(B_{m_0})| < \varepsilon$. Write $P_m = P_{B_m}$, $m \in \mathbb{N}$. Then

$$K(Q, P) \le K(Q, P_m) + \varepsilon \quad \text{for all } Q \in \Lambda \text{ and } m \ge m_0.$$

The inequality is trivially true if $K(Q, P_m) = \infty$ and is a consequence of $K(Q, P) - K(Q, P_m) = -\log P(B_m)$ if $K(Q, P_m) < \infty$. It follows that $K(\Omega, P) \le$
$\le \liminf_{m \to \infty} K(\Omega, P_m)$. To prove the lemma it still must be shown that conversely

$$(1.5.1) \qquad K(\Omega, P) \ge \limsup_{m \to \infty} K(\Omega, P_m).$$

The inequality is obvious if $K(\Omega, P) = \infty$. Hence assume $K(\Omega, P) < \infty$ and let $Q \in \Omega$ satisfy $K(Q, P) < K(\Omega, P) + \varepsilon$. Since $Q \in \Lambda^*$, there exists an $m_0 \in \mathbb{N}$ such that $Q(B_{m_0}) = 1$. Hence

$$\limsup_{m \to \infty} K(\Omega, P_m) \le \lim_{m \to \infty} K(Q, P_m) = K(Q, P) < K(\Omega, P) + \varepsilon$$

implying (1.5.1). $\square$

THEOREM 1.5.1. *Let* $P \in \Lambda$, *let* E *be a real Hausdorff topological vector space and let* $B_1 \subset B_2 \subset \dots$ *be an increasing sequence of Borel sets of S such that* $\lim_{m \to \infty} P(B_m) = 1$. *Let* $\Psi_m = \{Q \in \Lambda: Q(B_m) = 1\}$ *for* $m \in \mathbb{N}$ *and let* $\Lambda^* = \bigcup_{m=1}^{\infty} \Psi_m$. *Let* $T: \Lambda^* \to E$ *be a function whose restriction* $T|\Psi_m$ *is linear and* τ*-continuous at each* $Q \in \Psi_m$ *such that* $K(Q, P) < \infty$, *for each* $m \in \mathbb{N}$.

If A is a convex subset of E with closure $\bar{A}$ *and interior* A^0 *satisfying* $K(T^{-1}(A^0), P) < \infty$, *then*

(1.5.2) $\lim_{n \to \infty} n^{-1} \log \Pr\{T(\hat{P}_n) \in A\} = -K(T^{-1}(A),P).$

PROOF. Assume without loss of generality that $P(B_1) > 0$. Let $P_m = P_{B_m}$, $m \in \mathbb{N}$. By Lemma 1.5.1 $K(T^{-1}(A^0),P) = \lim_{m \to \infty} K(T^{-1}(A^0),P_m)$. Hence we may also assume without loss of generality that $K(T^{-1}(A^0),P_m) < \infty$ for each $m \in \mathbb{N}$. We shall first prove

(1.5.3) $K(T^{-1}(A^0),P_m) = K(T^{-1}(\bar{A}),P_m)$ for each $m \in \mathbb{N}$.

Fix $\epsilon > 0$ and $m \in \mathbb{N}$. There exists a pm $Q \in T^{-1}(\bar{A})$ which satisfies $K(Q,P_m) < K(T^{-1}(\bar{A}),P_m) + \epsilon$. There also exists a pm $R \in T^{-1}(A^0)$ such that $K(R,P_m) < \infty$. Let $Q_\alpha = \alpha Q + (1-\alpha)R$, $0 < \alpha < 1$. Since $Q, R \in \Psi_m$ and T is linear on Ψ_m, $Q_\alpha \in T^{-1}(A^0)$ for each $\alpha \in (0,1)$. Moreover $K(Q_\alpha,P_m) \leq$ $\leq \alpha K(Q,P_m) + (1-\alpha)K(R,P_m)$, $\alpha \in (0,1)$, by the convexity of the mapping $Q' \to K(Q',P_m)$, $Q' \in \Lambda$. It follows that $K(T^{-1}(A^0),P_m) \leq \lim_{\alpha \uparrow 1} K(Q_\alpha,P_m) <$ $< K(T^{-1}(\bar{A}),P_m) + \epsilon$, proving (1.5.3).

Let $\Omega = T^{-1}(A)$, let $\Psi_m^* = \{Q \in \Psi_m : K(Q,P) < \infty\}$ and let τ_m denote the relative τ-topology on Ψ_m, $m \in \mathbb{N}$. Since the restriction of T to Ψ_m is τ_m-continuous at each $Q \in \Psi_m^*$, one has $\Psi_m^* \cap T^{-1}(\bar{A}) \supset \Psi_m^* \cap cl_{\tau_m}(\Omega \cap \Psi_m) \supset \Psi_m^* \cap$ $\cap int_{\tau_m}(\Omega \cap \Psi_m) \supset \Psi_m^* \cap T^{-1}(A^0)$. Hence, by (1.5.3)

(1.5.4) $K(cl_{\tau_m}(\Omega \cap \Psi_m),P_m) = K(int_{\tau_m}(\Omega \cap \Psi_m),P_m)$,   for each $m \in \mathbb{N}$.

Let $\gamma = \limsup_{n \to \infty} n^{-1} \log \Pr\{T(\hat{P}_n) \in A\}$ and let $k \in \mathbb{N}$ be such that $k^{-1} \log \Pr\{T(\hat{P}_k) \in A\} \geq \gamma - \epsilon$. Since $\lim_{m \to \infty} \Pr\{T(\hat{P}_k) \in A \mid B_m\} =$ $= \Pr\{T(\hat{P}_k) \in A\}$ there exists $m_0 \in \mathbb{N}$ such that

$k^{-1} \log \Pr\{T(\hat{P}_k) \in A \mid B_m\} \geq \gamma - 2\epsilon$ for all $m \geq m_0$.

Hence for $m \geq m_0$

(1.5.5) $\limsup_{n \to \infty} n^{-1} \log \Pr\{T(\hat{P}_n) \in A \mid B_m\}$

$\geq \lim_{j \to \infty} (kj)^{-1} \log (\Pr\{T(\hat{P}_k) \in A \mid B_m\})^j$

$= k^{-1} \log \Pr\{T(\hat{P}_k) \in A \mid B_m\} \geq \gamma - 2\epsilon.$

The first inequality in (1.5.5) follows by the same superadditivity argument as used in the proof of Corollary 1.3.1.

By (1.5.4), Theorem 1.3.1 and Remark 1.3.2

$$\lim_{n \to \infty} n^{-1} \log \Pr\{T(\hat{P}_n) \in A \mid B_m\}$$

$$= \lim_{n \to \infty} n^{-1} \log \Pr\{\hat{P}_n \in \Omega \mid B_m\} = -K(\Omega, P_m).$$

Lemma 1.5.1 and (1.5.5) now imply

$$\gamma - 2\varepsilon \leq \lim_{m \to \infty} \lim_{n \to \infty} n^{-1} \log \Pr\{T(\hat{P}_n) \in A \mid B_m\}$$

$$= -\lim_{m \to \infty} K(T^{-1}(A), P_m) = -K(T^{-1}(A), P).$$

Thus $\gamma \leq -K(T^{-1}(A), P)$.

Conversely, for any $m, n \in \mathbb{N}$

$$n^{-1} \log \Pr\{T(\hat{P}_n) \in A\} \geq n^{-1} \log \Pr\{T(\hat{P}_n) \in A \mid B_m\} + \log P(B_m).$$

Hence, by the first part of the proof and Lemma 1.5.1

$$\liminf_{n \to \infty} n^{-1} \log \Pr\{T(\hat{P}_n) \in A\}$$

$$\geq \lim_{m \to \infty} [\liminf_{n \to \infty} n^{-1} \log \Pr\{T(\hat{P}_n) \in A \mid B_m\} + \log P(B_m)]$$

$$= \lim_{m \to \infty} -K(T^{-1}(A), P_m) = -K(T^{-1}(A), P). \quad \square$$

COROLLARY 1.5.1. *In Theorem 1.5.1 let* $T|\Psi_n$ *be linear and weakly continuous for each* $n \in \mathbb{N}$. *Then* (1.5.2) *holds for each subset* A *of* E *and* $P \in \Lambda$ *satisfying one of the following conditions:*

(i)   A *is convex and* $K(T^{-1}(A^0), P) < \infty$

(ii) A *is open,* P *is Lindelöf inner regular and* $K(T^{-1}(A), P) = \infty$

(iii) A *is open and convex and* P *is Lindelöf inner regular.*

PROOF. Under condition (i) the result follows from Theorem 1.5.1 since weak continuity implies τ-continuity of $T|\Psi_n$. Since condition (iii) implies either (i) or (ii), it remains to consider condition (ii).

Suppose A is open and $K(T^{-1}(A),P) = \infty$. Assume without loss of generality $P(B_1) > 0$. For each $m \in \mathbb{N}$ let P_m be the conditional pm P_{B_m}. We first show that $P_m(\text{supp}(P_m)) = 1$.

Let $A_m = B_m \cap (\text{supp}(P)\setminus\text{supp}(P_m))$. Since $P(\text{supp}(P)) = 1$ by the inner regularity of P, it suffices to prove $P(A_m) = 0$. For all $x \in A_m$ let U_x be a neighborhood of x in S such that $P_m(U_x) = 0$ and put $U = \bigcup_{x \in A_m} U_x$. Fix $\varepsilon > 0$. Again by the inner regularity of P there exists a Lindelöf subspace $V \subset U$ satisfying $P(V) > P(U)-\varepsilon$. Since V may be covered by countably many sets U_x, it is seen that $P_m(V) = 0$ and hence

$$P(A_m) \leq P(U \cap B_m) = P((U\setminus V) \cap B_m) < \varepsilon,$$

implying $P(A_m) = 0$.

Next we prove that $T^{-1}(A)$ does not contain pms with finite support in $B_m \cap \text{supp}(P_m)$, implying $\Pr\{T(\hat{P}_n) \in A \mid B_m\} = 0$ for all $m,n \in \mathbb{N}$ and hence $\Pr\{T(\hat{P}_n) \in A\} = 0$ for all $n \in \mathbb{N}$ in accordance with (1.5.2).

Fix $m \in \mathbb{N}$ and let $Q_0 \in \Lambda$ have finite support $\text{supp}(Q_0) \subset B_m \cap \text{supp}(P_m)$. Suppose $T(Q_0) \in A$. The weak continuity of $T|\Psi_m$ implies that there is a weak neighborhood V of Q_0 such that $T(V \cap \Psi_m) \subset A$. By Lemma 1.2.5 (with P replaced by P_m) the set V contains a pm Q_V such that $K(Q_V,P_m) < \infty$. It follows that $Q_V \in \Psi_m$ and hence $K(T^{-1}(A),P_m) < \infty$, in contradiction to $K(T^{-1}(A),P_m) = K(T^{-1}(A),P) = \infty$. Therefore $T(Q_0) \notin A$, as required. $\square$

REMARK 1.5.1. The events $\{\hat{P}_n \in T^{-1}(A)\}$ in Theorem 1.5.1 and Corollary 1.5.1 are not necessarily $\mathcal{B}^n$-measurable. If these events are not $\mathcal{B}^n$-measurable we still have

$$\lim_{n\to\infty} n^{-1} \log \overline{P^n}\{T(\hat{P}_n) \in A\} = \lim_{n\to\infty} n^{-1} \log \underline{P^n}\{T(\hat{P}_n) \in A\} =$$

$$= -K(T^{-1}(A),P),$$

where $\overline{P^n}(\underline{P^n})$ denotes the outer (inner) measure corresponding to the product measure P^n on $\mathcal{B}^n$ (cf. Remark 1.3.1). Proposition 1.3.1 implies that $\{\hat{P}_n \in T^{-1}(A)\}$ is $\mathcal{B}^n$-measurable if $A \in \mathcal{B}$, S is completely regular and T is weakly continuous.

Consider the particular case that S is a locally convex (Hausdorff) topological vector space. BAHADUR & ZABELL (1979) have shown that for each

convex open set A ⊂ S

(1.5.6) $\lim_{n\to\infty} n^{-1} \log \Pr\{n^{-1} \sum_{i=1}^{n} X_i \in A\}$

exists and is equal to -K(M(A),P), where M(A) ⊂ Λ is the set of pms with
expectation in A (their Theorems 2.1, 2.3 and 3.3). This theorem, together
with other methods to evaluate the limit, is derived under the condition
that the pm P and its convolutions satisfy certain inner regularity condi-
tions.

Another version of their result can also be deduced from Corollary 1.5.1.
For this purpose integrals of functions taking values in vector spaces are
needed. Let E be a copy of the locally convex topological vector space S,
let E' be the dual of E (i.e. the space of continuous real-valued linear
functionals on E) and let E'* be the algebraic dual of E' (i.e. the space
of real-valued linear functionals on E'). For y ∈ E, y' ∈ E', y'* ∈ E'*
write <y,y'> = y'(y) and <y'*,y'> = y'*(y'). Finally, let B be a *compact* sub-
set of S, let Ψ_B be the set of pms on the Borel σ-field of B and let C(B)
denote the space of continuous functions f: B → E. Then the *integral* of
f ∈ C(B) with respect to a pm Q ∈ Ψ_B, denoted by $\int_B fdQ$ or $\int_B f(x)dQ(x)$,
is an element of E'* defined by the relation

$$< \int_B fdQ,y'> = \int_B <f(x),y'> dQ(x)$$

for each y' ∈ E' (cf. BOURBAKI (1965), p.74-82).

Let Ê be the completion of E induced by the uniformity compatible with
the topology on E. Each element y ∈ Ê can be identified with an element of
E'* by identifying y with the linear form y' → <y,y'> on E' (where E' is
identified with Ê').

With this identification we have the following two fundamental proper-
ties of the integral for each f ∈ C(B): (a) the closure of the convex hull
of f(B) in Ê is equal to the set {$\int_B fdQ$: Q ∈ Ψ_B}, and (b) the map
φ: Q → $\int_B fdQ$ is the unique weakly continuous linear mapping from Ψ_B into Ê
such that φ(Q) = $\sum_{i=1}^{k} f(x_i)Q(\{x_i\})$ for each pm Q with finite support
{$x_1,\ldots,x_k$} (cf. BOURBAKI (1965), loc.cit.).

Now suppose that S = Ê and that P ∈ Λ is tight. Then there exists an
increasing sequence of compact subsets B_1 ⊂ B_2 ⊂... of S such that
$\lim_{n\to\infty} P(B_n)$ = 1. In the notation of Theorem 1.5.1 define T: Λ* → Ê by

$$(1.5.7) \qquad T(Q) = \int_{B_m} x \, dQ(x), \qquad Q \in \Psi_m \quad (m \in \mathbb{N}).$$

Since $\int_{B_m} <x,y'> dQ(x) = \int_{B_n} <x,y'> dQ(x)$ if $Q(B_m) = Q(B_n) = 1$, the value of $T(Q)$ does not depend on the choice of B_m. Moreover, by property (b) mentioned above, $T(\hat{P}_n) = n^{-1} \sum_{i=1}^{n} X_i$ and $T|\Psi_n$ is linear and weakly continuous for each $n \in \mathbb{N}$. Hence Corollary 1.5.1 implies that (1.5.6) exists and is equal to $-K(T^{-1}(A),P)$ under the conditions (i), (ii) or (iii). Apart from pms with non-compact support, the set $T^{-1}(A)$ coincides with the set $M(A)$ appearing in the result of BAHADUR & ZABELL (1979).

Note that we required that $S = \hat{E}$ and that P is tight. If P is convex-tight (for the definition see p.18), then by property (a) above T can be defined as a function $T: \Lambda^* \to E$ (instead of $\hat{E}$) since the sets B_n can be chosen compact and convex in this case.

Thus we have proved

COROLLARY 1.5.2. *Let S be a locally convex (Hausdorff) topological vector space and let $P \in \Lambda$ be convex-tight. Then the limit (1.5.6) exists and is equal to $-K(T^{-1}(A),P)$, with T defined by (1.5.7), if $A \subset S$ and P satisfy one of the conditions*

(i) A *is convex and* $K(T^{-1}(A^0),P) < \infty$

(ii) A *is open,* P *is Lindelöf inner regular and* $K(T^{-1}(A),P) = \infty$

(iii) A *is open and convex and* P *is Lindelöf inner regular.*

REMARK 1.5.2. Results related to Corollary 1.5.2 are given in BOROVKOV & MOGUL'SKII (1978).

REMARK 1.5.3. We briefly return to measurability questions. If $A \in \mathcal{B}$, the events $\{\hat{P}_n \in \Psi_m, T(\hat{P}_n) \in A\}$ are $\mathcal{B}^n$-measurable, with Ψ_m and T defined as in (1.5.7). This follows from Proposition 1.3.1, the weak continuity of $T|\Psi_m$ and the compactness (and hence complete regularity) of B_m. Thus, if $A \in \mathcal{B}$ and $\{n^{-1} \sum_{i=1}^{n} X_i \in A\} \in \mathcal{B}^n$, we may write

$$Pr\{n^{-1} \sum_{i=1}^{n} X_i \in A\}$$

$$(1.5.8) \qquad = Pr\{n^{-1} \sum_{i=1}^{n} X_i \in A, \; X_i \in \bigcup_{m=1}^{\infty} B_m, \; i = 1, \ldots, n\}$$

$$= Pr\{T(\hat{P}_n) \in A, \; \hat{P}_n \in \bigcup_{m=1}^{\infty} \Psi_m\}.$$

The events $\{n^{-1} \sum_{i=1}^{n} X_i \in A\}$ are not necessarily B^n-measurable, even if $A \in B$. However, if P is tight and the σ-algebra B^n is enlarged to the σ-algebra $\overline{B^n}$ containing all subsets of sets of P-measure zero in B^n, then $\{n^{-1} \sum_{i=1}^{n} X_i \in A\} \in \overline{B^n}$, for each $A \in B$. Hence if the pm P^n is (uniquely) extended to a pm on $\overline{B^n}$, then $Pr\{n^{-1} \sum_{i=1}^{n} X_i \in A\}$ is well-defined for all $A \in B$.

In BAHADUR & ZABELL (1979) $X_1, X_2, \ldots$ are measurable transformations from a probability space (Ω, A, P) to a locally convex Hausdorff topological vector space S, endowed with the σ-algebra of Borel sets B. The pm P on A induces a pm μ on B defined by

$$\mu(B) = P(X_1 \in B), \qquad B \in B.$$

It is assumed that, for each $n \in \mathbb{N}$ and $B_1, \ldots, B_n$ in B,

$$P(\omega : X_i(\omega) \in B_i, \quad i = 1, \ldots, n) = \prod_{i=1}^{n} \mu(B_i)$$

and that $\{\omega : n^{-1} \sum_{i=m}^{m+n-1} X_i(\omega) \in B\} \in A$ for each $m, n = 1, 2, \ldots$. Let for $m, n = 1, 2, \ldots$ the function $T_{m,n} : S^{m+n-1} \longrightarrow S$ be defined by

$$T_{m,n}(x) = n^{-1} \sum_{i=m}^{m+n-1} x_i, \qquad x = (x_1, \ldots, x_{m+n-1}) \in S^{m+n-1}.$$

Then $C_{m,n} = \{T_{m,n}^{-1}(B) : B \in B\}$ is a σ-algebra on S^{m+n-1} which is not necessarily contained in B^{m+n-1}. Hence the (convolution) pm $\mu_{m,n}$ on B defined by

$$\mu_{m,n}(B) = P(T_{m,n}(X_1, \ldots, X_{m+n-1}) \in B), \qquad B \in B$$

need not be uniquely determined by the product measure μ^{m+n-1} on B^{m+n-1}. Therefore BAHADUR & ZABELL (1979) impose additional conditions on the pms $\mu_{m,n}$ to ensure that $\mu_{m,n}(B)$ is well-defined for each $B \in B$ and that the collection $\{\mu_{m,n}\}_{m,n \in \mathbb{N}}$ has certain intuitively reasonable regularity properties.

However, if the pm μ on B is tight, it follows that the pms $\mu_{m,n}$ on B are determined by the completion of the measure space (S^{m+n-1}, μ^{m+n-1}). In the light of this remark it seems that the tightness conditions on the convolutions $\mu_{m,n}$ are redundant.

34

1.6. LINEAR COMBINATIONS OF ORDER STATISTICS

In this section $X_1, X_2, \ldots$ are real-valued i.i.d. random variables with distribution function (df) F. Instead of Λ_1, the set of pms on $(\mathbb{R}, \mathcal{B})$, we shall consider the set D of one dimensional dfs. We shall employ the same notation as in section 1.4. Moreover, for $G \in D$ the inverse G^{-1} is defined in the usual way by $G^{-1}(u) = \inf\{x \in \mathbb{R} : G(x) \geq u\}$. A set of dfs A in D will be called τ-open (or ρ-open) if the set of pms $\{P_G \in \Lambda_1 : G \in A\}$ is open in the topology τ (or ρ) defined on Λ_1. The topologies τ and ρ on D are defined by these τ-open and ρ-open sets respectively. Obviously all results on large deviations for pms on $\mathbb{R}$ lead to corresponding results for dfs on $\mathbb{R}$, so we freely use the theory of the preceding sections.

Suppose $J: [0,1] \to \mathbb{R}$ is an L-integrable function, i.e. $\int_0^1 |J(u)| du < \infty$. We consider linear combinations of order statistics of the form

$$(1.6.1) \qquad T(\hat{F}_n) = \int_0^1 J(u) \hat{F}_n^{-1}(u) du,$$

where $\hat{F}_n$ denotes the empirical df of $X_1, \ldots, X_n$, or in a perhaps more familiar notation

$$(1.6.2) \qquad T(\hat{F}_n) = \sum_{i=1}^n c_{n,i} X_{i:n},$$

where $c_{n,i} = \int_{(i-1)/n}^{i/n} J(u) du$ and $X_{i:n}$ is the i-th order statistic of $X_1, \ldots, X_n$. These statistics are sometimes called L-estimators, cf. HUBER (1972). For a more recent discussion we refer to BICKEL & LEHMANN (1975).

Related to the statistics $T(\hat{F}_n)$ are the sets

$$(1.6.3) \qquad \Omega_t = \{G \in D: \int_0^1 J(u) G^{-1}(u) du \geq t, \int_0^1 |J(u) G^{-1}(u)| du < \infty\},$$

where $t \in \mathbb{R}$.

The following large deviation theorem is a consequence of the preceding theory.

THEOREM 1.6.1. *Let* $F \in D$, *let* $J: [0,1] \to \mathbb{R}$ *be an L-integrable function and let* $[\alpha, \beta]$ *be the smallest closed interval containing the support of J. Then, for each sequence* $\{u_n\}$ *of real numbers such that* $\lim_{n \to \infty} u_n = 0$,

$$(1.6.4) \qquad \lim_{n \to \infty} n^{-1} \log \Pr\{T(\hat{F}_n) \geq r + u_n\} = -K(\Omega_r, F)$$

if J,F and $r \in \mathbb{R}$ *satisfy the conditions*

(i) $t \to K(\Omega_t, F), t \in \mathbb{R}$, *is continuous from the right at* $t = r$

(ii) $-\infty < \sup\{x \in \mathbb{R} : F(x) \le \alpha\} \le \inf\{x \in \mathbb{R} : F(x) \ge \beta\} < \infty$.

Moreover, (i) *is certainly satisfied if one of the following pairs of conditions holds:*

(a) *J* ≥ 0 *on an interval* (γ, δ) *and* $\int_\gamma^\delta J(u)\,du > 0$

(b) *F is continuous,*

or

(c) *the support of J is an interval,* $J \ge 0$ *and* $\int_0^1 J(u)\,du > 0$

(d) *F is continuous at* $r_1 = r / \int_0^1 J(u)\,du$.

Finally, if r_1 *is a discontinuity point of F then* (1.6.4) *holds provided comditions* (ii) *and* (c) *are satisfied and* $u_n \le 0$ *for all large* $n \in \mathbb{N}$.

REMARK 1.6.1. Condition (ii) of Theorem 1.6.1 is satisfied if P_F has compact support or if $0 < \alpha < \beta < 1$. A partial extension of Theorem 1.6.1 for the case that (ii) is not satisfied is given in OOSTERHOFF (1978).

REMARK 1.6.2. The second part of Theorem 1.6.1 illustrates a phenomenon known from proofs of asymptotic normality of linear combinations of order statistics: with strong conditions on the underlying df F only weak conditions on the score functions are needed and vice versa.

PROOF OF THEOREM 1.6.1. Let $A = [\alpha, \beta]$ let B be the smallest interval containing the support of P_F and let 1_A and 1_B denote the indicator functions of A and B respectively. Then

$$T(\hat{F}_n) = \int_0^1 J(u)\, 1_B(\hat{F}_n^{-1}(u))\, \hat{F}_n^{-1}(u)\,du$$

with probability one. Define the function $T: D \to \mathbb{R}$ by

(1.6.5) $T(G) = \int_0^1 J(u)\, 1_B(G^{-1}(u))\, G^{-1}(u)\,du,$ $G \in D$.

The function T is ρ-continuous. For a proof consider a sequence of dfs $\{G_n\}$, such that $G_n \to_\rho G$ for a df $G \in D$. Then $G_n^{-1} \to G^{-1}$ except perhaps on a countable number of discontinuity points of G^{-1}. Together with condition (ii) this implies that the functions $1_B(G_n^{-1}) G_n^{-1} \cdot 1_A$, $n \in \mathbb{N}$, are uniformly bounded on the interval $[0,1]$. Hence $\lim_{n\to\infty} T(G_n) = T(G)$ by dominated convergence

implying that T is ρ-continuous. The proof of (1.6.4) is now completed by an application of Theorem 1.3.2, since ρ-continuity implies τ-continuity.

In the proof of the other statements of the theorem we may assume that $K(\Omega_r,F) < \infty$, since otherwise condition (i) is trivially satisfied. Let $G \in \Omega_r$ satisfy $K(G,F) = K(\Omega_r,F)$. The existence of G is assured by Lemma 1.3.3 and the fact that a ρ-closed set is also τ-closed.

First suppose that conditions (a) and (b) are satisfied. Since $P_G \ll P_F$, G is continuous. Let (γ,δ) be an interval satisfying condition (a) and let γ_1 and $\varepsilon > 0$ be numbers such that $\gamma_1 \in (\gamma,\delta)$ and $\varepsilon < \min\{\gamma_1-\gamma,\delta-\gamma_1\}$. Let $c = G^{-1}(\gamma)$, $d = G^{-1}(\delta)$, $c_1 = G^{-1}(\gamma_1)$ and let the df G_ε be defined by its P_G-density $g_\varepsilon = dP_{G_\varepsilon}/dP_G$ given by

$$g_\varepsilon(x) = \begin{cases} (\gamma_1-\gamma-\varepsilon)/(\gamma_1-\gamma), & x \in (c,c_1) \\ (\delta-\gamma_1+\varepsilon)/(\delta-\gamma_1), & x \in [c_1,d) \\ 1 & , \quad \text{elsewhere.} \end{cases}$$

Then $G_\varepsilon^{-1}(u) > G^{-1}(u)$, $u \in (\gamma,\delta)$ and $G_\varepsilon^{-1} = G^{-1}$ elsewhere. Note that G_ε is derived from G by moving some probability mass of P_G to the right on the interval (c,d). Since $J(u) \geq 0$ for $u \in (\gamma,\delta)$ and $\int_\gamma^\delta J(u)du > 0$, $\int_\gamma^\delta J(u)G_\varepsilon^{-1}(u)du > \int_\gamma^\delta J(u)G^{-1}(u)du$. Hence $T(G_\varepsilon) > T(G)$. Since $\lim_{\varepsilon\downarrow 0} K(G_\varepsilon,F) = K(G,F)$, (i) follows.

Next suppose that conditions (c) and (d) are satisfied. Without loss of generality assume $\int_0^1 J(u)du = 1$ and hence $r_1 = r$. Let again $G \in \Omega_r$ satisfy $K(G,F) = K(\Omega_r,F) < \infty$. First suppose that $G^{-1}(\alpha+0) < G^{-1}(\beta)$. Then there exists $\gamma \in (\alpha,\beta)$ such that $G^{-1}(\gamma+h) > G^{-1}(\gamma)$ for each $h > 0$. Let $c = G^{-1}(\gamma)$ (hence $0 < G(c) = \gamma < 1$) and let for $0 < \varepsilon < \min\{\gamma,1-\gamma\}$ the df G_ε be defined by its P_G-density $g_\varepsilon = dP_{G_\varepsilon}/dP_G$ given by

$$g_\varepsilon(x) = \begin{cases} (\gamma-\varepsilon)/\gamma & , \quad x \leq c \\ (1-\gamma+\varepsilon)/(1-\gamma), & x > c. \end{cases}$$

Then $G_\varepsilon^{-1} \geq G^{-1}$ and $G_\varepsilon^{-1}(u) > G^{-1}(u)$ for each u in a left-hand neighborhood of γ. Hence $\int_0^1 J(u)G_\varepsilon^{-1}(u)du > \int_0^1 J(u)G^{-1}(u)du$ for each $\varepsilon > 0$. Since $\lim_{\varepsilon\downarrow 0} K(G_\varepsilon,F) = K(G,F)$, condition (i) follows.

It remains to consider the case that $G^{-1}(\alpha+0) = G^{-1}(\beta) = b$, say. Then $\int_0^1 J(u)G^{-1}(u)du = b \geq r$ since $G \in \Omega_r$. Suppose r is a continuity point

of F. Then $P_G \ll P_F$ implies $P_G(\{r\}) = 0$ and hence $b > r$, since b is a dis-
continuity point of G. It follows that $K(\Omega_t, F) = K(\Omega_r, F)$ for all $t \in (r,b)$,
implying (i).

Now suppose that r is a discontinuity point of F and that $b = r$. Note
that $G(r-0) \le \alpha$ in this case. If $G(r-0) > 0$ we proceed as follows. For
$0 < \varepsilon < G(r-0)$ define the df G_ε by its density $g_\varepsilon = dP_{G_\varepsilon}/dP_G$ given by

$$g_\varepsilon(x) = \begin{cases} (G(r-0)-\varepsilon)/G(r-0) & , \quad x < r \\[2ex] (1-G(r-0)+\varepsilon)/(1-G(r-0)), & x \ge r. \end{cases}$$

Then $G_\varepsilon(r-0) = G(r-0) - \varepsilon \le \alpha-\varepsilon$, hence $G_\varepsilon \in \Omega_r$. Considering the partition
$P = \{(-\infty,r),[r,\infty)\}$ of $\mathbb{R}$ it follows immediately that there is a τ-open
neighborhood of G_ε contained in Ω_r. Hence $K(\text{int}_\tau(\Omega_r),F) \le K(G_\varepsilon,F)$, for each
$\varepsilon > 0$. Since $\lim_{\varepsilon \downarrow 0} K(G_\varepsilon,F) = K(G,F)$, we have $K(\text{int}_\tau(\Omega_r),F) \le$
$\le \lim_{\varepsilon \downarrow 0} K(G_\varepsilon,F) = K(\Omega_r,F)$, i.e. $K(\text{int}_\tau(\Omega_r),F) = K(\Omega_r,F)$. The τ-continuity
of T implies that Ω_r is τ-closed and hence Theorem 1.3.1 yields that (1.6.4)
holds provided $u_n = 0$ for all large $n \in \mathbb{N}$. The left continuity of the
function $t \to K(\Omega_t,F)$ (Lemma 1.3.4) implies that (1.6.4) also holds if $u_n \le 0$
for all large $n \in \mathbb{N}$ (consider a sequence $\{t_m\}$ in $\mathbb{R}$ such that $t_m \uparrow r$ and
$t \to K(\Omega_t,F)$ is continuous at t_m for each $m \in \mathbb{N}$).

Finally suppose $G(r-0) = 0$. Let the df G' be defined by $P_{G'}(B) =$
$= P_F(B \cap [r,\infty))/P_F([r,\infty))$, for each Borel set B. Then $G' \in \Omega_r$ and $K(G',F) \le$
$\le K(G,F)$, hence $K(G,F) = K(G',F) = -\log P_F([r,\infty))$. Since Ω_r is τ-closed,
Lemma 1.2.4 implies that condition (A') of Lemma 1.3.2 is satisfied. Hence
$\lim \sup_{n \to \infty} n^{-1} \log \Pr\{\hat{F}_n \in \Omega_r\} \le \log P_F([r,\infty))$. It is clear that conversely
$\lim \inf_{n \to \infty} n^{-1} \log \Pr\{\hat{F}_n \in \Omega_r\} \ge \lim \inf_{n \to \infty} n^{-1} \log \Pr\{X_{1:n} \ge r\} =$
$= \log P_F([r,\infty))$. Thus (1.6.4) holds provided $u_n = 0$ for all large $n \in \mathbb{N}$.
By the same argument as before (1.6.4) also holds if $u_n \le 0$ for all large
$n \in \mathbb{N}$. $\square$

REMARK 1.6.3. The continuity of a function which is essentially equivalent
to the function T in (1.6.5) has been pointed out by BICKEL & LEHMANN (1975).
In fact there exists an interesting link between robust statistics and the
theory of large deviations, since robustness of statistics $T(\hat{F}_n)$ may be
defined by continuity of the corresponding functionals T on D with respect
to some suitably chosen topology and since large deviations of these types
of "continuous" functionals of empirical dfs can be tackled by the methods

38

of this chapter. Note that HOADLEY's (1967) Theorem 1 would not suffice to prove (1.6.4) since T is in general not *uniformly* ρ-continuous (and F is not assumed to be continuous).

In applications the weight function J appearing in the definition of the statistic $T(\hat{F}_n)$ may also depend on n. In this case Theorem 1.6.1 is not immediately applicable, but the next theorem may be of use.

THEOREM 1.6.2. *Let* $F \in D$, *let* J_n *($n \in \mathbb{N}$) and J be L-integrable functions defined on* $[0,1]$ *and let* $[\alpha,\beta]$ *be the smallest closed interval containing the support of J and the support of each* J_n. *Let* Ω_t *be defined by (1.6.3) for* $t \in \mathbb{R}$. *Then, for each sequence of real numbers* $\{u_n\}$ *such that* $\lim_{n\to\infty} u_n = 0$,

$$(1.6.6) \quad \lim_{n\to\infty} n^{-1} \log \Pr\left\{ \int_0^1 J_n(u)\hat{F}_n^{-1}(u)\,du \geq r+u_n \right\} = -K(\Omega_r, F)$$

if J, F, α and β satisfy conditions (i) *and* (ii) *of Theorem 1.6.1 and if the sequence* $\{J_n\}$ *satisfies*

(iii) $\lim_{n\to\infty} \int_0^1 |J_n(u) - J(u)|\,du = 0$.

PROOF. The proof proceeds by a truncation argument. We write $B_m = [-m,m]$ and denote by G_m the conditional df defined by

$$P_{G_m}(B) = P_G(B|B_m), \quad B \in \mathcal{B}, \text{ if } G \in D \text{ and } P_G(B_m) > 0.$$

Let $D^* = \{G \in D: P_G(B_m) = 1 \text{ for some } m \in \mathbb{N}\}$. By condition (i) there exists for each $\eta > 0$ a $\delta > 0$ and a df $G \in \Omega_{r+\delta}$ satisfying $K(G,F) \leq K(\Omega_r, F) + \eta$. Since $G_m \in \Omega_r$ for large m and $\lim_{m\to\infty} K(G_m,F) = K(G,F)$, it follows that $K(\Omega_r, F) = K(\Omega_r \cap D^*, F)$. Hence by Lemma 1.4.1 $\lim_{m\to\infty} K(\Omega_r, F_m) = K(\Omega_r, F)$. Fix $\varepsilon > 0$. Then there exists $N_0 = N_0(m,\varepsilon)$ such that for all $n \geq N_0$

$$(1.6.7) \quad \left| \int_0^1 J_n(u)\hat{F}_n^{-1}(u)\,du - \int_0^1 J(u)\hat{F}_n^{-1}(u)\,du \right| \leq m \int_0^1 |J_n(u)-J(u)|\,du < \tfrac{1}{4}\varepsilon$$

if $\hat{F}_n^{-1}(u) \in B_m$, $u \in (0,1)$. For convenience of notation we shall write

$$\Pr\{\hat{F}_n \in A | \hat{F}_n^{-1}(u) \in B_m, u \in (0,1)\} = \Pr\{\hat{F}_n \in A | B_m\}$$

if $P_F(B_m) > 0$. With this notation we have for each large $m \in \mathbb{N}$:

$$\liminf_{n \to \infty} n^{-1} \log \Pr\left\{\int_0^1 J_n(u)\hat{F}_n^{-1}(u)du \geq r+u_n\right\}$$

$$\geq \liminf_{n \to \infty} n^{-1} \log \Pr\left\{\int_0^1 J_n(u)\hat{F}_n^{-1}(u)du \geq r+u_n\big|B_m\right\} + \log P_F(B_m) \geq$$

$$\geq \liminf_{n \to \infty} n^{-1} \log \Pr\left\{\int_0^1 J(u)\hat{F}_n^{-1}(u)du \geq r+\tfrac{1}{2}\varepsilon\big|B_m\right\} + \log P_F(B_m)$$

$$\geq -K(\Omega_{r+\varepsilon},F_m) + \log P_F(B_m).$$

The last inequality holds by Theorem 1.6.1, since we may choose a continuity point $r_m \in (r+\tfrac{1}{2}\varepsilon, r+\varepsilon)$ of the function $t \to K(\Omega_t, F_m)$.

Since $\lim_{m \to \infty} K(\Omega_{r+\varepsilon}, F_m) = K(\Omega_{r+\varepsilon}, F)$, we have

$$\liminf_{n \to \infty} n^{-1} \log \Pr\left\{\int_0^1 J_n(u)\hat{F}_n^{-1}(u)du \geq r+u_n\right\} \geq -K(\Omega_{r+\varepsilon}, F).$$

Hence by condition (i)

$$(1.6.8) \qquad \liminf_{n \to \infty} n^{-1} \log \Pr\left\{\int_0^1 J_n(u)\hat{F}_n^{-1}(u)du \geq r+u_n\right\} \geq -K(\Omega_r, F).$$

Next we show that conversely

$$(1.6.9) \qquad \limsup_{n \to \infty} n^{-1} \log \Pr\left\{\int_0^1 J_n(u)\hat{F}_n^{-1}(u)du \geq r+u_n\right\} \leq -K(\Omega_r, F).$$

Fix $\varepsilon > 0$. There exists an $m \in \mathbb{N}$ such that for all $n \in \mathbb{N}$, $\Pr\{\hat{F}_n^{-1}(\alpha+0) \notin B_m\} < \varepsilon^n$ and $\Pr\{\hat{F}_n^{-1}(\beta) \notin B_m\} < \varepsilon^n$ (this may be seen for example by an application of Chernoff's theorem to the binomial representation of the probabilities $\Pr\{\hat{F}_n^{-1}(\alpha+0) \notin B_m\}$ and $\Pr\{\hat{F}_n^{-1}(\beta) \notin B_m\}$).
Hence for large n:

$$\Pr\left\{\int_0^1 J_n(u)\hat{F}_n^{-1}(u)du \geq r+u_n\right\}$$

$$\leq \Pr\left\{\int_0^1 J_n(u)\hat{F}_n^{-1}(u)du \geq r+u_n \text{ and } \hat{F}_n^{-1}(u) \in B_m, \; u \in (\alpha,\beta)\right\} + 2\varepsilon^n$$

$$\leq \Pr\left\{\int_0^1 J(u)\hat{F}_n^{-1}(u)du \geq r-\varepsilon\right\} + 2\varepsilon^n,$$

since (1.6.7) holds again for large n if $\hat{F}_n^{-1}(u) \in B_m$ for $u \in (\alpha,\beta)$.

This result implies (1.6.9) by Theorem 1.6.1 and Lemma 1.3.4 (also if $K(\Omega_r, F) = \infty$) and the present theorem follows from (1.6.8) and (1.6.9). $\square$

REMARK 1.6.4. Related results for particular cases have been obtained by BOOK (1974) and STEINEBACH (1977). BOOK (1974) gives large deviation results for Winsorized means of samples from an exponential distribution and STEINEBACH (1977) considers the more general case of linear combinations of order statistics from uniform and exponential distributions. The proofs in these papers are based on the relation between probabilities of large deviations and moment generating functions.

For $0 < \alpha < \frac{1}{2}$, the α-*trimmed mean* of $X_1, \ldots, X_n$ is defined by

$$(1.6.10) \qquad T_n = (n-2[\alpha n])^{-1} \sum_{i=[\alpha n]+1}^{n-[\alpha n]} X_{i:n}, \qquad n \in \mathbb{N},$$

where $[x]$ denotes the largest integer $\leq x$. As an application of the previous theorems we prove the following large deviation result for α-trimmed means.

THEOREM 1.6.3. *Let* $r \in \mathbb{R}$, *let* $F \in D$ *be continuous at* r *and let* T_n *be the* α-*trimmed mean given by* (1.6.10). *Then, for each sequence* $\{u_n\}$ *such that* $\lim_{n \to \infty} u_n = 0$,

$$(1.6.11) \qquad \lim_{n \to \infty} n^{-1} \log \Pr\{T_n \geq r+u_n\} = -K(\Omega_r^\alpha, F),$$

where

$$\Omega_r^\alpha = \{G \in D: \int_\alpha^{1-\alpha} G^{-1}(u)\,du \geq (1-2\alpha)r\}.$$

If F *is discontinuous at* r, *then* (1.6.11) *continues to hold provided* $u_n \leq 0$ *for all large* $n \in \mathbb{N}$.

PROOF. We write the statistic T_n in the form $\int_0^1 J_n(u)\hat{F}_n^{-1}(u)\,du$ with $J_n = n(n-2[\alpha n])^{-1} 1_{A_n}$, where $A_n = ([\alpha n]/n, 1-[\alpha n]/n)$. Let $J = (1-2\alpha)^{-1} \cdot 1_{(\alpha, 1-\alpha)}$. If F is continuous at r, then (1.6.11) follows since in this case (c) and (d) of Theorem 1.6.1 and hence the conditions of Theorem 1.6.2 are fulfilled.

Now suppose that F is discontinuous at r. Let $G \in \Omega_r^\alpha$ satisfy $K(G,F) = K(\Omega_r^\alpha, F)$ (such G exists!). It was shown in the course of the proof of Theorem 1.6.1 that the function $t \to K(\Omega_t^\alpha, F)$ is continuous at r (and hence

the above proof remains valid) unless $G^{-1}(\alpha+0) = G^{-1}(1-\alpha) = r$.

It remains to consider this exceptional case. Fix $\epsilon > 0$ and let $\Omega_{r,n} = \{H \in D: \int_0^1 J_n(u)H^{-1}(u)du \geq r\}$, $n \in \mathbb{N}$. For $0 < \delta < 1$ let $G_\delta \in D$ be defined by $G_\delta(x) = (1-\delta)G(x)$ if $x < r$ and $G_\delta(x) = (1-\delta)G(x) + \delta$ if $x > r$, implying $G_\delta(r-0) \leq \alpha-\delta\alpha$ and $G_\delta(r) \geq 1-\alpha+\delta\alpha$. Note that $K(G_\delta,F) < K(G,F) + \epsilon = K(\Omega_r^\alpha,F) + \epsilon$ if $\delta < \delta_\epsilon$, say. Moreover, $A_n \subset (\alpha-\delta\alpha,1-\alpha+\delta\alpha)$ and hence $G_\delta \in \Omega_{r,n}$ if $n > (\alpha\delta)^{-1}$. Let P denote the partition $\{(-\infty,r),\{r\}, (r,\infty)\}$ of $\mathbb{R}$. Choosing appropriate $\delta_n \in (\frac{1}{2}\delta_\epsilon,\delta_\epsilon)$ it follows that there exists a sequence $\{G_n\} = \{G_{\delta_n}\}$ such that for all $n > (\frac{1}{2}\alpha\delta_\epsilon)^{-1}$

(1) $nG_n(r-0) \in \mathbb{Z}$ and $nG_n(r) \in \mathbb{Z}$

(2) $G_n \in \Omega_{r,n}$ and $\{H \in D: d_P(P_H,P_{G_n}) = 0\} \subset \Omega_{r,n}$

(3) $K_P(G_n,F) < K(\Omega_r^\alpha,F) + \epsilon$.

Hence, if $u_n \leq 0$ for all large n, the same arguments that were used in the last part of the proof of Lemma 1.3.2 yield

$$\Pr\{T_n \geq r+u_n\} \geq \Pr\{\hat{F}_n \in \Omega_{r,n}\} \geq \Pr\{d_P(P_{\hat{F}_n},P_{G_n}) = 0\}$$

as $n \to \infty$, implying

$$\liminf_{n\to\infty} n^{-1} \log \Pr\{T_n \geq r+u_n\} \geq -K(\Omega_r^\alpha,F).$$

On the other hand (1.6.9) continues to hold in the present case, with Ω_r^α in lieu of Ω_r, since the second part of the proof of Theorem 1.6.2 does not use condition (i). This completes the proof of the last statement of the theorem. $\square$

The actual computation of the infimum $K(\Omega_r^\alpha,F)$ in (1.6.11) is not easy. We shall derive a more explicit expression for $K(\Omega_r^\alpha,F)$ under the assumption that F is continuous. In this case any df H such that $K(H,F) < \infty$ is also continuous and

$$\int_\alpha^{1-\alpha} H^{-1}(u)du = \int_a^b xdH(x)$$

where $a = H^{-1}(\alpha)$, $b = H^{-1}(1-\alpha)$ and $-\infty < a < b < \infty$. We also assume $F(r) < 1$ since otherwise $K(\Omega_r^\alpha,F) = \infty$.

The minimization procedure is performed in two steps. Let

$$\Omega_r^\alpha(a,b) = \{H \in D: (1-2\alpha)^{-1} \int_a^b x\,dH(x) \geq r, \quad H(a) = \alpha, \quad H(b) = 1-\alpha\}$$

for $-\infty < a < b < \infty$. In view of the continuity of F

$$(1.6.12) \qquad K(\Omega_r^\alpha, F) = \inf\{K(\Omega_r^\alpha(a,b), F): 0 < F(a) < F(b) < 1, \; F(b) > F(r)\}.$$

Consider the function $t \to tr - \log \int_a^b e^{tx}\,dF(x)$, $t \geq 0$. This function achieves its maximum on $[0,\infty)$ at a point $s = s(a,b)$ defined by

$$s = \begin{cases} 0 & \text{if } \int_a^b x\,dF(x)/(F(b)-F(a)) \geq r \\[2ex] \phi^{-1}(r) & \text{otherwise,} \end{cases}$$

where $\phi(t) = \int_a^b xe^{tx}\,dF(x)/\int_a^b e^{tx}\,dF(x)$, $t \geq 0$. Note that in the second case the equation $\phi(t) = r$ has a unique positive root s since $\phi(0) < r$, $\lim_{t\to\infty} \phi(t) > r$ and $\phi'(t) > 0$ for all $t \geq 0$.

Let $G \in D$ be defined by its density $g = dP_G/dP_F$ given by

$$g(x) = \begin{cases} \alpha/F(a) & , \quad x < a \\ (1-2\alpha)e^{sx}/\int_a^b e^{sx}\,dF(x), & a \leq x \leq b \\ \alpha/(1-F(b)) & , \quad x > b. \end{cases}$$

Then $G \in \Omega_r^\alpha(a,b)$ and

$$K(G,F) = 2\alpha\log\alpha + (1-2\alpha)\log(1-2\alpha) - \alpha\log F(a) - \alpha\log(1-F(b)) +$$

$$+ (1-2\alpha)sr - (1-2\alpha)\log \int_a^b e^{sx}\,dF(x).$$

Let $H \in \Omega_r^\alpha(a,b)$, $K(H,F) < \infty$ and $h = dP_H/dP_F$. By Jensen's inequality

$$sr - \log\{(1-2\alpha)^{-1} \int_a^b e^{sx}\,dF(x)\}$$

$$\leq sr - \log\{(1-2\alpha)^{-1} \int_a^b \exp(sx - \log h(x))\,dH(x)\} \leq$$

$$\leq s\{r - (1-2\alpha)^{-1} \int_a^b x\,dH(x)\} + (1-2\alpha)^{-1} \int_a^b \log h(x)\,dH(x).$$

Hence

$$\int_a^b \log h(x)\,dH(x) \ge (1-2\alpha)\{sr + \log(1-2\alpha) - \log \int_a^b e^{sx}\,dF(x)\}.$$

Similarly, by Jensen's inequality,

$$\int_{-\infty}^a \log h(x)\,dH(x) \ge H(a)\ \log\{H(a)/F(a)\} = \alpha\ \log(\alpha/F(a))$$

and

$$\int_b^\infty \log h(x)\,dH(x) \ge (1-H(b))\ \log\{(1-H(b))/(1-F(b))\} =$$

$$= \alpha\ \log\{\alpha/(1-F(b))\}.$$

Thus

$$K(H,F) = \int_{\mathbb{R}} \log h(x)\,dH(x) \ge K(G,F),$$

implying $K(\Omega_r^\alpha(a,b),F) = K(G,F)$.

Now define the functions

$$f_\alpha(a,b) = (1-2\alpha)s(a,b)r - \alpha\log F(a) - \alpha \log(1-F(b)) +$$

$$- (1-2\alpha)\ \log \int_a^b \exp(s(a,b)x)\,dF(x)$$

and

$$g(\alpha) = 2\alpha\ \log\ \alpha + (1-2\alpha)\log(1-2\alpha).$$

Then, by (1.6.12)

$$(1.6.13) \qquad K(\Omega_r^\alpha,F) = g(\alpha) + \inf\{f_\alpha(a,b): 0 < F(a) < F(b) < 1,\ F(b) > F(r)\}.$$

CHAPTER II

BAHADUR EFFICIENCIES OF SIGNED RANK TESTS

2.1. INTRODUCTION

Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables taking values in $\mathbb{R}$ according to some pm $P \in \Lambda_1$. The *hypothesis of symmetry* is given by

$$H_0: \quad P((a,b)) = P((-b,-a)), \quad \text{for each interval } (a,b) \subset (0,\infty).$$

We shall consider signed rank tests for H_0 based on signed rank test statistics without regression constants. If P is non-atomic these statistics are of the following form

$$(2.1.1) \qquad T_n = \int_{(0,\infty)} J_n(\hat{H}_n(x)) d\hat{F}_n(x),$$

where $\hat{F}_n$ is the empirical df of $X_1, \ldots, X_n$, $\hat{H}_n$ is the empirical df of $|X_1|, \ldots, |X_n|$ and $J_n: (0,1] \to \mathbb{R}$ is a score function.

For example, $J_n(u) \equiv 1$ yields the sign test statistic and $J_n(u) = $ $= n^{-1} \sum_{i=1}^{n} i \cdot 1_{((i-1)/n, i/n]}(u)$ yields the Wilcoxon signed rank test statistic. Large deviation theorems for these statistics will be given in section 2.2. If P is purely atomic and hence ties are present, there exist several methods for dealing with ties and zeros leading to different types of signed rank statistics. We shall be concerned with probabilities of large deviations of these statistics in section 2.3.

The large deviation theorems to be presented in the next two sections will be relatively straightforward consequences of the theory which was developed in chapter I. They will enable us to determine the exact slopes of signed rank tests. The main results of section 2.2 have first been derived by HO(1974) under slightly stronger conditions and by rather different methods (Theorems 2.2.1, 2.2.2 and 2.2.4). In Theorem 2.2.3 a sufficient

condition is given under which the Bahadur efficiency coincides locally with Pitman efficiency.

As far as we know, the results of Section 2.3 are new. They yield the exact slopes of conditional signed rank tests when the underlying pms have finite support under essentially the same conditions as in the non-atomic case. Related results have appeared in an unpublished Ph.D. thesis by CHAKRABARTI (1978).

2.2. EXACT SLOPES IN THE NON-ATOMIC CASE

Let for each $n \in \mathbb{N}$ the *score function* $J_n : (0,1] \to \mathbb{R}$ be defined by

$$(2.2.1) \qquad J_n(u) = \sum_{i=1}^{n} a_{in} 1_{((i-1)/n, i/n]}(u), \quad u \in (0,1],$$

where the a_{in}'s, $1 \leq i \leq n$, are real numbers. For each $Q \in \Lambda_1$ the df H_Q is defined by

$$(2.2.2) \qquad H_Q(x) = Q([-x,x]) 1_{[0,\infty)}(x), \quad x \in \mathbb{R},$$

and the pm $\bar{Q}$, with support contained in $[-1,1]$, by

$$(2.2.3) \qquad \bar{Q}(B) = \begin{cases} Q(\{x > 0: H_Q(x) \in B\}), & B \in \mathcal{B},\ B \subset (0,1], \\ Q(\{x < 0: -H_Q(-x) \in B\}), & B \in \mathcal{B},\ B \subset [-1,0), \\ Q(\{0\}), & \text{if } B = \{0\}. \end{cases}$$

Note that $\bar{P}$ is the distribution of the random variable $H_P(|X_1|) \text{sign}(X_1)$, where $\text{sign}(x) = 1_{(0,\infty)}(x) - 1_{(-\infty,0)}(x)$. The empirical pm $\hat{P}_n$ induces an empirical pm $\hat{\bar{P}}_n$ of the (dependent) random variables $H_{\hat{P}_n}(|X_i|) \text{sign}(X_i)$, $i = 1,\ldots,n$ (note that $H_{\hat{P}_n} = \hat{H}_n$). In the sequel we use the convention $H_Q^{-1}(0) = 0$.

With this notation we can write the statistics (2.1.1) with score functions J_n as defined by (2.2.1) in the following way

$$(2.2.4) \qquad T_n = \int_{(0,1]} J_n(u) d\hat{\bar{P}}_n(u).$$

This representation of T_n suggests the study of the functionals $T_{J_m} : \Lambda_1 \to \mathbb{R}$ defined by

(2.2.5) $T_{J_m}(Q) = \int_{(0,1]} J_m(u)d\bar{Q}(u), \qquad Q \in \Lambda_1,$

where J_m is a score function satisfying (2.2.1).

LEMMA 2.2.1. *Let* $J_m: (0,1] \to \mathbb{R}$ *be a score function satisfying* (2.2.1). *Then* $T_{J_m}: \Lambda_1 \to \mathbb{R}$ *is* ρ-*continuous at each non-atomic pm* $Q \in \Lambda_1$.

PROOF. Suppose that $Q \in \Lambda_1$ is non-atomic and that $Q_n \to_\rho Q, n \to \infty$, for a sequence $\{Q_n\}$ in Λ_1. Let F_n and F be the dfs corresponding to Q_n and Q, respectively. Then $F_n \to F$ in the supremum metric and hence $H_{Q_n} \to H_Q$, also in the supremum metric. It follows that $H_{Q_n}^{-1} \to H_Q^{-1}$ except perhaps on a countable set D of discontinuity points. Fix $\varepsilon > 0$ and first suppose that $F(0) < 1$. Let $a_0 \in (0,1)$ and $a_1, a_2 \in (0,1) \setminus D$ be points such that $b_i = F(H_Q^{-1}(a_i))$, $i = 0,1,2$ satisfy $b_0 - \varepsilon < b_1 < b_0 < b_2 < b_0 + \varepsilon$. Since $F_n(H_{Q_n}^{-1}(a_i)) \to b_i$, $i = 1,2$, (by the uniform convergence of $\{F_n\}$ to F), we have

$$b_1 \leq \liminf_{n \to \infty} F_n(H_{Q_n}^{-1}(a_0)) \leq \limsup_{n \to \infty} F_n(H_{Q_n}^{-1}(a_0)) \leq b_2.$$

It follows that $F_n(H_{Q_n}^{-1}(a_0)) \to F(H_Q^{-1}(a_0))$ for all $a_0 \in (0,1)$. If $F(0) = 1$, then $F_n(H_{Q_n}^{-1}(a)) \to 1$, for all $a \in (0,1)$.

In a similar way one can prove that $F_n(H_{Q_n}^{-1}(a)) \to F(H_Q^{-1}(a))$, if $a = 0$ or $a = 1$ and that $F_n(-H_{Q_n}^{-1}(-a)) \to F(-H_Q^{-1}(-a))$, if $a \in [-1,0)$. Hence $\bar{Q}_n \to_\rho \bar{Q}$, $n \to \infty$, and thus, in view of (2.2.1),

$$T_{J_m}(Q_n) \to T_{J_m}(Q), \qquad n \to \infty,$$

implying the ρ-continuity of T_{J_m} at Q. $\square$

The following somewhat technical lemma will enable us to reduce probabilities of large deviations of T_n under H_0 to probabilities of large deviations of the statistics

(2.2.6) $\int_{(0,1]} J_m(u)d\hat{\bar{P}}_n(u),$

where J_m is a *fixed* score function. By Lemma 2.2.1 we can apply Theorem 1.3.2 to the statistics of type (2.2.6) and thus obtain expressions for the asymptotic behavior of probabilities of large deviations.

We introduce the following notation. The set $\Gamma \subset \Lambda_1$ is defined by

(2.2.7) $\Gamma = \{Q \in \Lambda_1: Q$ has a density q w.r.t. Lebesgue measure such that
$$q(u) + q(-u) = 1, \quad u \in (0,1]\}.$$

Note that $\bar{Q} \in \Gamma$ for each non-atomic pm $Q \in \Lambda_1$. If $J: (0,1) \to \mathbb{R}$ is an L-integrable function, then $\int_0^1 J(u)dQ(u)$ is well-defined for each $Q \in \Gamma$, since $\int_0^1 |J(u)|dQ(u) \leq \int_0^1 |J(u)|du < \infty$.
Hence we may define

(2.2.8) $\Gamma(t,J) = \{Q \in \Gamma: \int_0^1 J(u)dQ(u) \geq t\}$,

where $t \in \mathbb{R}$. Furthermore, the function $f_J: \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ is defined by

(2.2.9) $f_J(t) = \inf\{K(Q,R): Q \in \Gamma(t,J)\}$,

where R is the pm in Λ_1 corresponding to the uniform density on $[-1,1]$ and where $f_J(t) = \infty$ if $\Gamma(t,J) = \emptyset$. Finally we define for L-integrable functions $J: (0,1) \to \mathbb{R}$,

$$m(J) = \frac{1}{2} \int_0^1 J(u)du$$

(2.2.10)

$$M(J) = \int_0^1 J(u)1_{(0,\infty)}(J(u))du.$$

LEMMA 2.2.2. *Let* $J: (0,1) \to \mathbb{R}$ *be an L-integrable function. Then*

(i)   *the function* f_J *is finite, convex and non-decreasing on the interval* $(m(J),M(J))$.

(ii)   *if* $m(J) < r < M(J)$, *then*

(2.2.11) $f_J(r) = \lambda r - \frac{1}{2} \lambda \int_0^1 J(u)du - \int_0^1 \log \cosh(\frac{1}{2} \lambda J(u))du$,

    where $\lambda = \lambda(r)$ *is the unique solution of the equation*

$$(2.2.12) \qquad \int_0^1 J(u)\{1 + \exp(-\lambda J(u))\}^{-1} du = r$$

(iii) *Let r and ε > 0 satisfy*

$$m(J) < r-\varepsilon < r+\varepsilon < M(J).$$

If $\widetilde{J}$ is an L-integrable function such that $\int_0^1 |J(u)-\widetilde{J}(u)|du \le \varepsilon$, then

$$(2.2.13) \qquad m(\widetilde{J}) \le m(J) + \varepsilon < r < M(J) - \varepsilon \le M(\widetilde{J})$$

and

$$(2.2.14) \qquad f_J(r-\varepsilon) \le f_{\widetilde{J}}(r) \le f_J(r+\varepsilon).$$

PROOF.

(ii) Assume $m(J) < M(J)$, since otherwise the implication is trivially satisfied. By the method of Langrangian multipliers it is seen that the infimum at the right-hand side of (2.2.9) is attained by the pm $Q \in \Gamma$ with the following density q

$$q(u) = \begin{cases} \{1+\exp(-\lambda J(u))\}^{-1}, & u \in (0,1), \\ \{1+\exp(\lambda J(-u))\}^{-1}, & u \in (-1,0), \\ 0, & \text{otherwise}, \end{cases}$$

where λ is a solution of equation (2.2.12). Since by assumption $M(J) > m(J)$, the function J is not equal to zero on a set of positive Lebesgue measure, implying that the function

$$\phi: \lambda \to \int_0^1 J(u)\{1 + \exp(-\lambda J(u))\}^{-1} du, \quad \lambda \in [0,\infty),$$

is strictly increasing. By Lebesgue's dominated convergence theorem ϕ is also continuous. Since $\phi(0) = m(J)$ and $\lim_{\lambda \to \infty} \phi(\lambda) = M(J)$, there exists exactly one solution λ of (2.2.12).

(i) By (ii) $f_J(r)$ is finite if $m(J) < r < M(J)$. The convexity of the function f_J follows from the convexity of the mapping $Q \to K(Q,P)$ and the linearity of the mapping $Q \to \int_0^1 J(u)dQ(u)$, $Q \in \Gamma$. Finally, the monotonicity of f_J follows from the inclusion relation

$\Gamma(r',J) \subset \Gamma(r,J)$, if $r \leq r'$.

(iii) Let $Q \in \Gamma$ and let q be the density of Q. Then

$$\int_0^1 |J(u) - \tilde{J}(u)| q(u) du \leq \int_0^1 |J(u) - \tilde{J}(u)| du \leq \varepsilon.$$

Hence

$$\int_0^1 J(u) q(u) du - \varepsilon \leq \int_0^1 \tilde{J}(u) q(u) du \leq \int_0^1 J(u) q(u) du + \varepsilon,$$

implying (2.2.13) and (2.2.14). $\quad\square$

In the following theorem an asymptotic expression for probabilities of large deviations under H_0 of the signed rank statistics (2.1.1) is given (in the non-atomic case).

THEOREM 2.2.1. *Let* P *be a non-atomic pm in* Λ_1 *satisfying* H_0 *and let* $X_1, X_2, \ldots$ *be a sequence of i.i.d. random variables taking values in* $\mathbb{R}$ *according to* P. *Let* T_n *be defined by* (2.1.1), $n = 1, 2, \ldots$, *where the score functions* J_n *satisfy* (2.2.1) *and the condition*

$$(2.2.15) \qquad \lim_{n \to \infty} \int_0^1 |J_n(u) - J(u)| du = 0,$$

for some L-*integrable function* $J: (0,1) \to \mathbb{R}$. *If* $m(J) < r < M(J)$ *then, for each sequence* $\{u_n\}$ *of real numbers such that* $\lim_{n \to \infty} u_n = 0$,

$$\lim_{n \to \infty} n^{-1} \log \Pr\{T_n \geq r + u_n\} = -f_J(r),$$

where $f_J(r)$ *is defined by* (2.2.11) *and* (2.2.12).

PROOF. Without loss of generality we may assume that P corresponds to the uniform density on $[-1,1]$, since the distribution of T_n is the same for each non-atomic pm, symmetric about 0.

Fix $\varepsilon > 0$ such that

$$m(J) < r-2\varepsilon < r+2\varepsilon < M(J).$$

By (2.2.15) there exists an $m = m(\varepsilon)$ such that

$$\int\limits_{0}^{1} |J_n(u) - J(u)|\, du < \frac{1}{4}\, \varepsilon, \qquad n \geq m.$$

Since with probability one the empirical df $\hat{H}_n$ runs through all values $0, 1/n, \ldots, 1$, we have with probability one

$$\int\limits_{(0,1]} |J_n(u) - J_m(u)|\, d\hat{\tilde{P}}_n(u) \leq n^{-1} \sum_{i=1}^{n} |J_n(i/n) - J_m(i/n)| \leq$$

$$\leq \int\limits_{0}^{1} |J_n(u) - J_m(u)|\, du + (m/n) \max_{1 \leq i \leq m} |J_m(i/m) - J_m((i-1)/m)|.$$

Hence there exists an $n_0 \geq m$ such that with probability one

$$(2.2.16) \qquad |T_n - T_{J_m}(\hat{P}_n)| < \varepsilon,\ n \geq n_0,$$

where T_{J_m} is defined as in (2.2.5).

Let

$$\Omega_t = \{Q \in \Lambda_1 : T_{J_m}(Q) \geq t\}, \qquad t \in \mathbb{R}.$$

By Theorem 1.3.2 and Lemma 2.2.1

$$(2.2.17) \qquad \lim_{n \to \infty} n^{-1} \log \Pr\{T_{J_m}(\hat{P}_n) \geq r + u_n\} = -K(\Omega_r, P),$$

if $t \to K(\Omega_t, P)$ is continuous from the right at $t = r$ and $\lim_{n \to \infty} u_n = 0$.

Thus, by (2.2.16) and (2.2.17),

$$(2.2.18)$$
$$-K(\Omega_{r+\varepsilon}, P) \leq \liminf_{n \to \infty} n^{-1} \log \Pr\{T_n \geq r + u_n\}$$

$$\leq \limsup_{n \to \infty} n^{-1} \log \Pr\{T_n \geq r + u_n\} \leq -K(\Omega_{r-\varepsilon}, P),$$

if the function $t \to K(\Omega_t, P)$, $t \in \mathbb{R}$, is continuous from the right at $t = r - \varepsilon$ and $t = r + \varepsilon$, and $\lim_{n \to \infty} u_n = 0$. We shall show that $K(\Omega_t, P) = f_{J_m}(t)$ for $m(J_m) < t < M(J_m)$.

Let Q be a pm in Λ_1 such that $K(Q, P) < \infty$, then Q is absolutely continuous w.r.t. Lebesgue measure. Suppose q and $\bar{q}$ are the densities of Q and $\bar{Q}$, respectively. Then

$$T_{J_m}(Q) = \int\limits_0^1 J_m(u)\bar{q}(u)\,du$$

and

$$K(Q,P) = \int\limits_{-1}^1 q(u)\log(2q(u))\,du$$

$$= \int\limits_0^1 (q(H_Q^{-1}(u))/\{q(H_Q^{-1}(u))+q(-H_Q^{-1}(u))\})\log(2q(H_Q^{-1}(u)))\,du$$

$$+ \int\limits_{-1}^0 (q(-H_Q^{-1}(-u))/\{q(H_Q^{-1}(-u))+q(-H_Q^{-1}(-u))\})\log(2q(-H_Q^{-1}(-u)))\,du$$

$$= \int\limits_{-1}^1 \bar{q}(u)\log(2\bar{q}(u))\,du + \int\limits_0^1 (q(x)+q(-x))\log\{q(x)+q(-x)\}\,dx$$

$$\geq \int\limits_{-1}^1 \bar{q}(u)\log(2\bar{q}(u))\,du = K(\bar{Q},P).$$

Since $\bar{\bar{Q}} = \bar{Q}$ it follows that

$$(2.2.19) \quad K(\Omega_t,P) = \inf\{K(\bar{Q},P): Q \in \Lambda_1, \int\limits_0^1 J_m(u)\,d\bar{Q}(u) \geq t\}$$

$$= \inf\{K(Q,P): Q \in \Gamma(t,J_m)\} = f_{J_m}(t).$$

By Lemma 2.2.2 and the choice of ε and $m = m(\varepsilon)$ we have $m(J_m) \leq$ $\leq m(J) + \varepsilon < r-\varepsilon < r+\varepsilon < M(J)-\varepsilon \leq M(J_m)$. Hence the function $t \to K(\Omega_t,P)$, $t \in \mathbb{R}$, is continuous at $t = r + \varepsilon$ and $t = r - \varepsilon$ by Lemma 2.2.2 and (2.2.19). The theorem now follows from (2.2.18). $\square$

Theorem 2.2.1 has been proved by HO(1974) by methods which are similar to those used in WOODWORTH (1970). Ho also considers the more general case of signed rank statistics with regression constants. For this case the results are less explicit.

We now discuss the concept of *Bahadur efficiency* of signed rank tests in the non-atomic case. Let T_n be defined by (2.1.1) and let G_n be the df of T_n under H_0, when the df of X_i, $i = 1,2,\ldots$, is continuous. The *level*

52

attained by T_n is defined by

$$(2.2.20) \quad L_n(T_n) = 1-G_n(T_n-).$$

Suppose that the limit

$$(2.2.21) \quad c(P) = - \lim_{n \to \infty} 2n^{-1} \log L_n(T_n)$$

exists with probability one, if $X_1, X_2, \ldots$ are distributed according to the non-atomic pm P. Then $c(P)$ is called the *exact slope* of the sequence $\{T_n\}$ at P. If the sequence $\{T_n^{(i)}\}$ has the exact slope $c_i(P)$ at P, $i = 1,2$, and if $c_1(P)$ and $c_2(P)$ are not both equal to zero, then the *Bahadur efficiency* of $\{T_n^{(1)}\}$ w.r.t. $\{T_n^{(2)}\}$ at P is defined by $c_1(P)/c_2(P)$ (with the conventions $0/\infty = 0$ and $a/0 = \infty$ if $a > 0$). In chapter III other (equivalent) definitions of Bahadur efficiency and exact slopes will be used. For the proof of the equivalence of these definitions (under a very weak regularity condition) we refer to BAHADUR (1967).

The following "strong law of large numbers for signed rank statistics" will be of use in the computation of exact slopes.

LEMMA 2.2.3. *Let* $X_1, X_2, \ldots$ *be distributed according to a non-atomic pm* $P \in \Lambda_1$. *If* T_n *is defined by* (2.1.1) *and if the score functions* J_n *satisfy the conditions of Theorem* 2.2.1, *then with probability one*

$$T_n \to \int_0^1 J(u)d\bar{P}(u), \quad \text{as } n \to \infty.$$

PROOF. By (2.2.15)

$$(2.2.22) \quad \limsup_{n \to \infty} \int_0^1 |J_n(u) - J(u)|d\bar{P}(u) \le \lim_{n \to \infty} \int_0^1 |J_n(u) - J(u)|du = 0.$$

By the Glivenko-Cantelli theorem we have with probability one

$$\hat{P}_n \to_\rho P, \quad \text{as } n \to \infty,$$

Hence, by Lemma 2.2.1

(2.2.23) $T_{J_m}(\hat{P}_n) \to T_{J_m}(P)$,   as $n \to \infty$,

with probability one, if m is a *fixed* natural number. The result now follows from (2.2.22) and (2.2.23) (see (2.2.16)).   □

   Combination of Theorem 2.2.1 and Lemmas 2.2.2 and 2.2.3 gives

THEOREM 2.2.2. *Let the score functions* J_n *satisfy the conditions of Theorem 2.2.1. Let* $X_1, X_2, \ldots$ *be distributed according to* P, *where* P *is a non-atomic pm in* Λ_1 *satisfying*

(2.2.24) $m(J) < \int_0^1 J(u)\, d\bar{P}(u) < M(J)$.

Then the exact slope of the sequence of signed rank statistics $\{T_n\}$ *at* P *is given by* $2f_J(\int_0^1 J(u)\, d\bar{P}(u))$, *where the function* f_J *is defined by* (2.2.11).

PROOF. By (2.2.24) and Lemma 2.2.2 the function f_J is continuous at $\int_0^1 J(u)\, d\bar{P}(u)$.

   Since by Lemma 2.2.3

$$T_n \to \int_0^1 J(u)\, d\bar{P}(u), \qquad \text{as } n \to \infty,$$

with probability one, it follows from Theorem 2.2.1 that

$$-2n^{-1}\log L_n(T_n) \to 2f_J\left(\int_0^1 J(u)\, d\bar{P}(u)\right), \qquad \text{as } n \to \infty,$$

also with probability one.   □

   We shall apply Theorem 2.2.2 to one-sample tests for location. Let $X_1, X_2, \ldots$ be distributed according to a df F_θ, where $F_\theta(x) = F(x-\theta)$ for some unknown continuous df F symmetric about 0 and where θ is an unknown location parameter. The null hypothesis $\theta = 0$ is to be tested against the alternative $\theta > 0$. Since F is unknown, this hypothesis coincides with the hypothesis

   H_c: P is non-atomic and satisfies H_0,

(c for "continuous"), where P is the pm corresponding to F.

54

We shall consider the *sign test*, *Wilcoxon's signed rank test* and *van der Waerden's signed rank test*. These statistics can (apart from standardizations) be written in the form (2.1.1) with score functions of type (2.2.1). The score functions of these tests are respectively given by

$$J_n^{(1)}(u) = 1, \quad u \in (0,1],$$

$$J_n^{(2)}(u) = n^{-1} \sum_{i=1}^{n} i \cdot 1_{((i-1)/n, i/n]}(u), \quad u \in (0,1],$$

$$J_n^{(3)}(u) = \sum_{i=1}^{n} \Phi^{-1}(\tfrac{1}{2} + \tfrac{1}{2} i/(n+1)) \cdot 1_{((i-1)/n, i/n]}(u), \quad u \in (0,1],$$

where Φ is the standard normal df. Let us denote the corresponding test statistics by $T_n^{(1)}$, $T_n^{(2)}$ and $T_n^{(3)}$, respectively. For these statistics the conditions of Theorem 2.2.2 are satisfied with the limiting score functions $J^{(i)}$ defined by $J^{(1)}(u) = 1$, $J^{(2)}(u) = u$ and $J^{(3)}(u) = \Phi^{-1}(\tfrac{1}{2} + \tfrac{1}{2} u)$, $u \in (0,1)$.

We define for $i = 1,2,3$,

$$(2.2.25) \quad \mu_i(F,\theta) = \int_0^1 J^{(i)}(u) d\bar{P}_\theta,$$

where F is a symmetric and continuous df and P_θ is the pm corresponding to F_θ. If $\tfrac{1}{2} < F(\theta) < 1$, the exact slope of the *sign test* at F_θ is given by

$$c_1(F,\theta) = 2F(\theta) \log(2F(\theta)) + 2(1-F(\theta)) \log(2(1-F(\theta))).$$

This result easily follows from Theorem 2.2.2. By Theorem 2.2.2 we also obtain the following results. If $\tfrac{1}{4} < \mu_2(F,\theta) < \tfrac{1}{2}$, the exact slope of *Wilcoxon's signed rank* test is given by

$$c_2(F,\theta) = 2\lambda(2\mu_2(F,\theta) - \tfrac{1}{2}) - 2 \log \cosh(\tfrac{1}{2}\lambda),$$

where $\lambda > 0$ is the unique solution of the equation

$$\int_0^1 u\{1 + \exp(-\lambda u)\}^{-1} du = \mu_2(F,\theta).$$

Finally, if $(2\pi)^{-\frac{1}{2}} < \mu_3(F,\theta) < 2(2\pi)^{-\frac{1}{2}}$, the exact slope of *van der Waerden's signed rank test* is given by

$$c_3(F,\theta) = 2 \left\{ \lambda\mu_3(F,\theta) - \lambda(2\pi)^{-\frac{1}{2}} - \int_0^1 \log\cosh\left(\tfrac{1}{2}\lambda\Phi^{-1}\left(\tfrac{1}{2}+\tfrac{1}{2}u\right)\right) du \right\},$$

where $\lambda > 0$ is the unique solution of the equation

$$\int_0^1 \Phi^{-1}\left(\tfrac{1}{2}+\tfrac{1}{2}u\right)\{1 + \exp\left(-\lambda\Phi^{-1}\left(\tfrac{1}{2}+\tfrac{1}{2}u\right)\right)\}^{-1} du = \mu_3(F,\theta).$$

It is clear that in general the computation of these exact slopes must be accomplished by numerical methods. Tables and graphs of the Bahadur efficiencies of the three rank tests w.r.t. one another have been given by KLOTZ (1965) for normal, logistic and double exponential location alternatives. Graphs of the efficiencies of the three tests w.r.t. the most powerful test for these same types of alternatives are given in GROENEBOOM & OOSTERHOFF (1977).

We shall show that under some regularity conditions the Bahadur efficiencies tend to the Pitman efficiencies as $\theta \downarrow 0$. To this end we need the following result of PITMAN (1949).

LEMMA 2.2.4. *Let* $\mathbb{P}_\theta$ *be the probability distribution of the random sequence* $S = (X_1, X_2, \ldots)$, *where* $\theta \in [0,\infty)$. *Let* $T_n = T_n(X_1, \ldots, X_n)$, $n \in \mathbb{N}$, *be a test statistic for testing* $\theta = 0$ *against* $\theta > 0$, *rejecting the hypothesis for large values of* T_n. *Suppose*

(i) $\mathbb{P}_\theta(T_n \geq a)$ *is non-decreasing in* θ *for each* $a \in \mathbb{R}$ *and* $\theta \geq 0$.

(ii) $\lim_{n\to\infty} \mathbb{P}_{\theta_n}(n^{\frac{1}{2}}(T_n - \mu(\theta_n))/\sigma(\theta_n) \leq x) = \Phi(x)$,
      for $x \in \mathbb{R}$ *and all* $\theta_n = k.n^{-\frac{1}{2}}$, $k \geq 0$, *where the function* $\theta \to \mu(\theta)$,
      $\theta \geq 0$, *has a right-hand derivative* $\mu'(0)$ *at* $\theta = 0$ *and the function*
      $\theta \to \sigma(\theta)$, $\theta \geq 0$, *satisfies* $\lim_{\theta\downarrow 0}\sigma(\theta) = \sigma(0) > 0$.

If $\{\tilde{T}_n\}$ *is a second sequence of statistics satisfying the same conditions, but with* μ *and* σ *replaced by* $\tilde{\mu}$ *and* $\tilde{\sigma}$, *then the Pitman efficiency of* $\{\tilde{T}_n\}$ *w.r.t.* $\{T_n\}$ *for testing* $\theta = 0$ *against* $\theta > 0$ *is given by*

$$e = (\tilde{\mu}'(0)/\tilde{\sigma}(0))^2/(\mu'(0)/\sigma(0))^2.$$

THEOREM 2.2.3. *Let* $\{T_n\}$ *be a sequence of signed rank statistics such that the score functions* J_n *and* J *satisfy* (2.2.15). *Let* F *be a continuous df, symmetric about* 0, *such that for sufficiently small* $\theta > 0$,

$$m(J) < \mu(F,\theta) < M(J),$$

where $\mu(F,\theta) = \int_0^1 J(u) d\bar{P}_\theta(u)$ and P_θ is the pm corresponding to F_θ. Moreover, suppose that the following conditions hold

(i) $\int_0^1 J^2(u) du < \infty$

(ii)   the right-hand derivative $\mu'(F,0)$ of the function $\theta \to \mu(F,\theta)$ at $\theta = 0$ exists

(iii) $\{T_n\}$ and F satisfy the conditions of Lemma 2.2.4 with $\mu(\theta) = \mu(F,\theta)$ and $\sigma^2(\theta) = \int_0^1 (J(u)-\mu(\theta))^2 d\bar{P}_\theta$.

If $\{\tilde{T}_n\}$ is a second sequence of statistics satisfying the same conditions, but with μ, σ, J_n and J replaced by $\tilde{\mu}, \tilde{\sigma}, \tilde{J}_n$ and $\tilde{J}$ respectively, then the Pitman efficiency of $\{\tilde{T}_n\}$ w.r.t. $\{T_n\}$ for testing H_c against F_θ, $\theta > 0$, is equal to $\lim_{\theta \downarrow 0} e(\theta)$, where $e(\theta)$ is the Bahadur efficiency of $\{\tilde{T}_n\}$ w.r.t. $\{T_n\}$ at θ.

PROOF. Without loss of generality we may assume that $\int_0^1 J(u) du = 0$. By condition (ii) we have

$$\mu(\theta) = \theta\mu'(0) + o(\theta), \qquad \theta \downarrow 0,$$

where $\mu'(0)$ denotes the right-hand derivative of the function $\theta \to \mu(\theta)$ at $\theta = 0$. Since by condition (i)

$$\int_0^1 J(u)\{1 + \exp(-\lambda J(u))\}^{-1} du = \frac{1}{4}\lambda \int_0^1 J^2(u) du + o(\lambda), \qquad \lambda \downarrow 0,$$

the solution λ of (2.2.12) with $r = \mu(\theta)$ satisfies

$$\lambda = 4\theta\mu'(0)/\sigma^2(J) + o(\theta), \qquad \theta \downarrow 0,$$

where $\sigma^2(J) = \int_0^1 J^2(u) du$. Hence

$$f_J(\mu(\theta)) = 2(\theta\mu'(0))^2/\sigma^2(J) + o(\theta^2), \qquad \theta \downarrow 0.$$

Similarly,

$$f_{\tilde{J}}(\tilde{\mu}(\theta)) = 2(\theta\tilde{\mu}'(0))^2/\sigma^2(\tilde{J}) + o(\theta^2), \qquad \theta \downarrow 0,$$

implying

$$\lim_{\theta \downarrow 0} f_{\widetilde{J}}(\widetilde{\mu}(\theta))/f_J(\mu(\theta)) = (\widetilde{\mu}'(0)/\sigma(\widetilde{J}))^2/(\mu'(0)/\sigma(J))^2.$$

The theorem now follows from Lemma 2.2.4. □

Theorem 2.2.3 implies that the Bahadur efficiencies of the sign test, Wilcoxon's signed rank test and van der Waerden's signed rank test w.r.t. one another tend to the respective Pitman efficiencies if $\theta \downarrow 0$ and F is a normal, logistic or double exponential df. It can be shown using WIEAND (1976) that in these circumstances the Pitman efficiencies are also the limits of the *approximate* Bahadur efficiencies as $\theta \downarrow 0$ (for the definition of this concept see e.g. BAHADUR (1960)). It follows that in this case the Pitman efficiency coincides with the limiting approximate and exact Bahadur efficiency as $\theta \downarrow 0$.

Sufficient conditions under which the exact and approximate Bahadur efficiencies of linear rank tests for the two-sample problem coincide local-ly for $\theta \downarrow 0$ have been given by KREMER (1979).

Although locally van der Waerden's test, Wilcoxon's test and the sign test are asymptotically most powerful for testing H_c against normal, logistic and double exponential shift alternatives respectively, these optimality properties are not preserved for fixed alternatives. In fact, the Bahadur efficiency of Wilcoxon's signed rank test w.r.t. van der Waerden's signed rank test is greater than one for alternatives Φ_θ with θ greater than 1.1 (see KLOTZ (1965)). Nevertheless, the best possible exact slope for testing H_c is attained by signed rank tests at each fixed alternative under some weak regularity conditions. This is shown by the next theorem.

THEOREM 2.2.4. *Let* P $\in \Lambda_1$ *be a non-atomic pm not satisfying* H_0 *and let* $\bar{p}$ *be a density of* $\bar{P}$ *w.r.t. Lebesgue measure. Let the function* J: $(0,1) \to \mathbb{R}$ *be defined by*

$$J(u) = \begin{cases} \log(\bar{p}(u)/\bar{p}(-u)), & \text{if } \bar{p}(u)\bar{p}(-u) > 0 \\ \\ 0, & \text{otherwise.} \end{cases}$$

Suppose that the score functions J_n *and* J *satisfy* (2.2.15) *and*

$$m(J) < \int_0^1 J(u)\,d\bar{P}(u) < M(J).$$

Then the sequence $\{T_n\}$ *defined by* (2.2.4) *attains the best possible slope for testing* H_c *against* P. *This exact slope is given by*

$$2K(P,\Omega_0) \;=\; 2 \int_{-1}^{1} \bar{p}(u)\log(2\bar{p}(u))\,du,$$

where $\Omega_0 = \{Q \in \Lambda_1 : Q \text{ satisfies } H_c\}$.

PROOF. For each $Q \in \Lambda_1$ we define a pm $\tilde{Q} \in \Lambda_1$ by

$$\tilde{Q}(B) \;=\; Q(\{x \in \mathbb{R} : |x| \in B\}), \qquad B \in \mathcal{B}.$$

Let $Q \in \Omega_0$ and let μ be a σ-finite measure dominating P, Q, $\tilde{P}$ and $\tilde{Q}$. Suppose that p, q, $\tilde{p}$ and $\tilde{q}$ are the densities of P, Q, $\tilde{P}$ and $\tilde{Q}$ w.r.t. μ and let $K(P,Q) < \infty$. Then

$$K(P,Q) \;=\; \int_{\mathbb{R}} p(x)\log(p(x)/q(x))\,d\mu(x)$$

$$=\; \int_{\mathbb{R}} p(x)\log(2p(x)/\{p(x)+p(-x)\})\,d\mu(x)$$

$$+\; \int_{\mathbb{R}} p(x)\log(\{\tfrac{1}{2}(p(x)+p(-x))\}/q(x))\,d\mu(x)$$

$$=\; \int_{-1}^{1} \bar{p}(u)\log(2\bar{p}(u))\,du \;+\; \int_{[0,\infty)} \tilde{p}(x)\log(\tilde{p}(x)/\tilde{q}(x))\,d\mu(x)$$

$$\geq\; \int_{-1}^{1} \bar{p}(u)\log(2\bar{p}(u))\,du,$$

by the symmetry of the density q and the non-negativity of $K(\tilde{P},\tilde{Q})$. Hence

$$K(P,\Omega_0) \;=\; \int_{-1}^{1} \bar{p}(u)\log(2\bar{p}(u))\,du \;=\; \int_{\mathbb{R}} p(x)\log(2p(x)/\{p(x)+p(-x)\})\,d\mu(x)$$

and the pm $R \in \Omega_0$ with μ-density $r(x) = \tfrac{1}{2}\{p(x) + p(-x)\}$ is the "least favorable" pm for testing H_c against P. By a result of Stein the Bahadur slope of the sequence of most powerful (MP) tests for testing R against P is equal to $2K(P,R)$ (cf. BAHADUR (1971), Lemma 6.1). Since the power of any test of H_c against P cannot exceed the power of the MP test for testing R against P, $2K(P,R)$ is the best possible exact slope at P. By Theorem 2.2.2

this is also the slope of $\{T_n\}$ at P. $\square$

Theorem 2.2.4 has first been proved by HO(1974) under somewhat stronger conditions (Ho assumes that P has a density w.r.t. Lebesque measure and that $\{J_n\}$ is of uniformly bounded variation on (0,1)). The optimality property of signed rank tests expressed by Theorem 2.2.4 is called the *asymptotic sufficiency* in the Bahadur sense of the vector of signs and ranks. A similar result for linear rank tests has been proved for the two-sample problem by HÁJEK (1974) and for the independence problem by GROENEBOOM *et al* (1976).

It follows from Theorem 2.2.4 that a rank test satisfying the conditions of Theorem 2.2.1 is optimal in the sense of Bahadur efficiency for testing H_c against each non-atomic P such that $\bar{P}$ has a density $\bar{p}$ satisfying

$$\bar{p}(u) = \{1 + \exp(-J(u))\}^{-1}, \quad u \in (0,1).$$

Hence the *sign test* is optimal against each non-atomic P such that the density $\bar{p}$ of $\bar{P}$ satisfies

$$\bar{p}(u) = \{1 + e^{-1}\}^{-1}, \quad u \in (0,1).$$

For *Wilcoxon's signed rank test* this pm P is characterized by

$$\bar{p}(u) = \{1 + e^{-u}\}^{-1}, \quad u \in (0,1)$$

and for *van der Waerden's signed rank test* by

$$\bar{p}(u) = \{1 + \exp(-\Phi^{-1}(\tfrac{1}{2}+\tfrac{1}{2}u))\}^{-1}, \quad u \in (0,1).$$

2.3. EXACT SLOPES FOR PMS WITH FINITE SUPPORT

Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables taking values in the set $D = \{x \in \mathbb{Z} : |x| \leq m\}$ according to some $P_a \in \Lambda_1$ (a for "atomic"), where $m \in \mathbb{N}$. The null hypothesis to be tested is

$$H_a : P_a \quad \text{satisfies } H_0.$$

There are several types of statistics for testing H_a corresponding to different ways of dealing with ties and zeros. To facilitate the definition of these statistics we introduce an infinite sequence of i.i.d. auxiliary random variables $U_1, U_2, \ldots$ which have a uniform distribution on $(-\frac{1}{2}, \frac{1}{2})$. We suppose that the X_i's and U_i's are stochastically independent. Let

$$(2.3.1) \qquad Y_i = X_i + U_i, \qquad i = 1, 2, \ldots,$$

let $\hat{P}_n$ be the empirical pm of $Y_1, \ldots, Y_n$ and let $\hat{H}_n$ be the empirical df of $|Y_1|, \ldots, |Y_n|$. *Score functions* J_n are defined as in (2.2.1).

Signed rank statistics with *randomized ranks* and *Pratt's method for handling ties at zero* (cf. PRATT (1959)) are defined by

$$(2.3.2) \qquad T_n^{(1)} = \int_{(\frac{1}{2}, \infty)} J_n(\hat{H}_n(x)) \, d\hat{P}_n(x).$$

Signed rank statistics with *randomized ranks* and *zeros discarded* are defined by

$$(2.3.3) \qquad T_n^{(2)} = \int_{(\frac{1}{2}, \infty)} J_n(\hat{H}_n(x) - \hat{H}_n(\tfrac{1}{2})) \, d\hat{P}_n(x).$$

Signed rank statistics with *Pratt's method for handling ties at zero* and the *average score method for other ties* are defined by

$$(2.3.4) \qquad T_n^{(3)} = \sum_{k=1}^{m} \frac{\hat{P}_n((k-\frac{1}{2}, k+\frac{1}{2}])}{P_{\hat{H}_n}((k-\frac{1}{2}, k+\frac{1}{2}])} \int_{(k-\frac{1}{2}, k+\frac{1}{2}]} J_n(\hat{H}_n(x)) \, dP_{\hat{H}_n}(x).$$

Finally, signed rank statistics with *zeros discarded* and the *average score method for other ties* are defined by

$$(2.3.5) \qquad T_n^{(4)} = \sum_{k=1}^{m} \frac{\hat{P}_n((k-\frac{1}{2}, k+\frac{1}{2}])}{P_{\hat{H}_n}((k-\frac{1}{2}, k+\frac{1}{2}])} \int_{(k-\frac{1}{2}, k+\frac{1}{2}]} J_n(\hat{H}_n(x) - \hat{H}_n(\tfrac{1}{2})) \, dP_{\hat{H}_n}(x).$$

A *tie at* i, $i = 0, \ldots, m$, of the sample $Y_1, \ldots, Y_n$ is defined by

$$\tau_{n,i} = \#\{j \in \mathbb{N} : |Y_j| \in (i-\tfrac{1}{2}, i+\tfrac{1}{2}), \; j \le n\},$$

where $\#V$ denotes the number of elements of a (finite) set V. The *vector of ties* of the sample $Y_1, \ldots, Y_n$ is defined by $\tau_n = (\tau_{n,0}, \ldots, \tau_{n,m})$.

Let $A_0 = (-\frac{1}{2},\frac{1}{2})$ and $A_i = (-i-\frac{1}{2},-i+\frac{1}{2}) \cup (i-\frac{1}{2},i+\frac{1}{2})$, $i = 1,\ldots,m$. Furthermore, let P_i be the uniform pm on A_i, i.e. $P_0((a,b)) = b-a$, if $(a,b) \subset (-\frac{1}{2},\frac{1}{2})$ and $P_i((a,b)) = \frac{1}{2}(b-a)$, if $(a,b) \subset (i-\frac{1}{2},i+\frac{1}{2})$ or $(a,b) \subset (-i-\frac{1}{2},-i+\frac{1}{2})$, $i = 1,\ldots,m$. If we condition on the vector of ties τ_n and if P_a satisfies H_a, then $Y_1,\ldots,Y_n$ may be considered as coming from $m + 1$ samples of sizes $\tau_{n,0},\ldots,\tau_{n,m}$ from the probability distributions $P_0,\ldots,P_m$, respectively. This observation will be helpful in proving large deviation theorems for the statistics $T_n^{(1)},\ldots,T_n^{(4)}$. To avoid trivialities we define $\Pr\{T_n^{(i)} \in A \mid \tau_n = t_n\} = 0$, if $A \in \mathcal{B}$ and $\Pr\{\tau_n = t_n\} = 0$, $i = 1,\ldots,4$.

For the statistics $T_n^{(1)}$ we have the following result.

<u>THEOREM 2.3.1.</u> *Let H_a be satisfied and define $p_i = \Pr\{|X_1| = i\}$, $i = 0,1,\ldots,m$, with $p_0 < 1$. Suppose that the score functions J_n satisfy (2.2.15) for some L-integrable function J. Let $\{t_n\} = \{(t_{n,0},\ldots,t_{n,m})\}$ be a sequence of $(m+1)$-tuples of non-negative integers such that $\sum_{i=0}^{m} t_{n,i} = n$ and*

(2.3.6) $\displaystyle \lim_{n\to\infty} n^{-1} t_{n,i} = p_i$, $i = 0,\ldots,m$.

If

(2.3.7) $\displaystyle \frac{1}{2}\int_{p_0}^{1} J(u)\,du < r < \int_{p_0}^{1} J(u)\, 1_{(0,\infty)}(J(u))\,du$

and $\{u_n\}$ is a sequence of real numbers tending to 0, then

$$\lim_{n\to\infty} n^{-1} \log \Pr\{T_n^{(1)} \geq r+u_n \mid \tau_n = t_n\} = -f_J(r,p_0),$$

where

(2.3.8) $\displaystyle f_J(r,p_0) = \lambda r - \frac{1}{2}\lambda \int_{p_0}^{1} J(u)\,du - \int_{p_0}^{1} \log\cosh(\tfrac{1}{2}\lambda J(u))\,du$

and $\lambda > 0$ is the unique solution of the equation

(2.3.9) $\displaystyle \int_{p_0}^{1} J(u)\{1 + \exp(-\lambda J(u))\}^{-1}\,du = r.$

<u>PROOF.</u> First suppose that $p_0 > 0$. For $i = 0,1,\ldots,m$ let $\mathcal{B}_i$ denote the Borel field on A_i and let Γ_i be the set of pms on $(A_i,\mathcal{B}_i)$. Without loss of generality we may assume that $p_i > 0$, $i = 1,\ldots,m$. Let $\nu = (\nu_0,\ldots,\nu_m)$, where

$\nu_i > 0$, $i = 0,\ldots,m$ and $\sum_{i=0}^{m} \nu_i = 1$. Furthermore, let $Q = (Q_0,\ldots,Q_m)$, where Q_i is a pm in Γ_i. The df $H_{Q,\nu}$ is defined by

$$(2.3.10) \quad H_{Q,\nu}(x) = \sum_{i=0}^{m} \nu_i Q_i([-x,x] \cap A_i) 1_{[0,\infty)}(x)$$

and the pms $\bar{Q}_i \in \Lambda_1$ are defined by

$$(2.3.11) \quad \bar{Q}_i(B) = \begin{cases} Q_i(\{x > 0: H_{Q,\nu}(x) \in B, \; x \in A_i\}), & B \in \mathcal{B}, \; B \subset (0,1] \\ Q_i(\{x < 0: -H_{Q,\nu}(-x) \in B, \; x \in A_i\}), & B \in \mathcal{B}, \; B \subset [-1,0) \\ Q_i(\{0\}), & \text{if } B = \{0\}, \; i = 0. \end{cases}$$

$\bar{Q}_i$ is the distribution of $H_{Q,\nu}(|Y_1|) 1_{A_i}(Y_1) \, \text{sgn}(Y_1)$ if Q_i is the conditional distribution of Y_1 given $Y_1 \in A_i$, $i = 0,\ldots,m$. Let Γ_i be endowed with the ρ-topology. The product topology on $\Gamma = \Pi_{i=0}^{m} \Gamma_i$ will also be called ρ-topology.

With these definitions it can be proved along the lines of Lemma 2.2.1 that the mapping $T_{\nu,J_k} : \Gamma \to \mathbb{R}$ defined by

$$(2.3.12) \quad T_{\nu,J_k}(Q) = \sum_{i=1}^{m} \nu_i \int_{(0,1]} J_k(u) \, d\bar{Q}_i(u),$$

is ρ-continuous at each $Q = (Q_0,\ldots,Q_m)$ such that Q_i is non-atomic, $0 \le i \le m$, for each *fixed* score function J_k.

Let $\hat{P}_{n,i}$ be the empirical pm of a sample of size $t_{n,i}$ of random variables distributed according to P_i, where P_i is the uniform distribution on A_i, $i = 0,\ldots,m$.

Let $\hat{\bar{P}}_{n,i}$ be defined by (2.3.11) with Q_i replaced by $\hat{P}_{n,i}$, ν replaced by $n^{-1} t_n$ and $H_{Q,\nu}$ replaced by $\hat{H}_n$. Then we have

$$(2.3.13) \quad T_n^{(1)} = n^{-1} \sum_{i=1}^{m} t_{n,i} \int_{(0,1]} J_n(u) \, d\hat{\bar{P}}_{n,i}(u),$$

conditional on $\tau_n = t_n$. Hence

$$\Pr\{T_n^{(1)} \ge r + u_n \mid \tau_n = t_n\}$$

$$= \Pr\{n^{-1} \sum_{i=1}^{m} t_{n,i} \int_{(0,1]} J_n(u) \, d\hat{\bar{P}}_{n,i}(u) \ge r + u_n\}.$$

Let $p = (p_0,\ldots,p_m)$ and let

$$\Omega_t = \Omega_t(T_{p,J_k}) = \{Q \in \Gamma: T_{p,J_k}(Q) \geq t\}.$$

Moreover, for $P = (P_0, \ldots, P_m)$ and $Q = (Q_0, \ldots, Q_m)$ define

$$I_p(Q,P) = \sum_{i=0}^{m} P_i \, K(Q_i, P_i)$$

and

$$I_p(\Omega,P) = \inf\{I_p(Q,P): Q \in \Omega\}, \qquad \text{for } \Omega \subset \Gamma.$$

Then, by an easy extension of Corollary 1.3.3

$$\lim_{n\to\infty} n^{-1} \log \Pr\{T_{p,J_k}((\hat{P}_{n,0}, \ldots, \hat{P}_{n,m})) \geq r+u_n\} = -I_p(\Omega_r,P),$$

if $t \to I_p(\Omega_t,P)$ is continuous from the right at $t = r$ and if $u_n \to 0$, as $n \to \infty$.

Let

$$\alpha(t,J) = \inf\{I_p(Q,P): T_{p,J}(Q) \geq t\},$$

where $T_{p,J}$ is defined by

$$(2.3.14) \qquad T_{p,J}(Q) = \begin{cases} \sum\limits_{i=1}^{m} P_i \int\limits_{(0,1]} J(u) \, d\bar{\bar{Q}}_i(u), & \text{if } I_p(Q,P) < \infty \\[2ex] 0, & \text{otherwise.} \end{cases}$$

By the Lagrangian multiplier technique it can be shown that, if r satisfies (2.3.7), one has $\alpha(r,J) = f_J(r,p_0)$, where $f_J(r,p_0)$ is defined by (2.3.8). For small $\varepsilon > 0$ and sufficiently large $k = k(\varepsilon)$ we have

$$\alpha(r-2\varepsilon,J) < I_p(\Omega_{r-\varepsilon}(T_{p,J_k}),P)$$

and

$$\alpha(r+2\varepsilon,J) > I_p(\Omega_{r+\varepsilon}(T_{p,J_k}),P),$$

implying that the function

$$t \to I_p(\Omega_t(T_{p,J_k}),P), \qquad t \in \mathbb{R},$$

is continuous from the right at $t = r \pm \varepsilon$ (cf. Lemma 2.2.2). The proof is now completed in the same way as the proof of Theorem 2.2.1 using the continuity of the function $\nu \to I_\nu(\Omega_r(T_{\nu,J_k}),P)$ at $\nu = p$.

If $p_i = 0$ for some i, $0 \le i \le m$, the proof is essentially the same. We only have to adapt the method of proof used above for a smaller number of samples. □

REMARK 2.3.1. If $p_0 = 0$, the number $f_J(r,p_0)$ reduces to $f_J(r)$ as defined by (2.2.11).

In a similar way we obtain the following theorems for the statistics $T_n^{(2)}$, $T_n^{(3)}$ and $T_n^{(4)}$.

THEOREM 2.3.2. Let $J, \{J_n\}$, $\{t_n\}$ and $\{u_n\}$ satisfy the conditions of Theorem 2.3.1 and suppose that H_a holds. Then, if

$$\tfrac{1}{2} \int_0^{1-p_0} J(u)\,du < r < \int_0^{1-p_0} J(u) 1_{(0,\infty)}(J(u))\,du$$

we have

$$\lim_{n \to \infty} n^{-1} \log \Pr\{T_n^{(2)} \ge r+u_n \mid \tau_n = t_n\} = -g_J(r,p_0),$$

where

$$(2.3.15) \qquad g_J(r,p_0) = \lambda r - \tfrac{1}{2}\lambda \int_0^{1-p_0} J(u)\,du - \int_0^{1-p_0} \log \cosh(\tfrac{1}{2}\lambda J(u))\,du$$

and $\lambda > 0$ is the unique solution of the equation

$$(2.3.16) \qquad \int_0^{1-p_0} J(u)\{1 + \exp(-\lambda J(u))\}^{-1}\,du = r.$$

THEOREM 2.3.3. Let $J, \{J_n\}$, $\{t_n\}$ and $\{u_n\}$ satisfy the conditions of Theorem 2.3.1 and suppose that H_a holds.
Let

$$q_i = \sum_{j=0}^{i} p_j, \qquad i = 0,\ldots,m, \quad p = (p_0,\ldots,p_m).$$

(i) *If*

$$\tfrac{1}{2}\int_{p_0}^{1} J(u)\,du < r < \int_{p_0}^{1} J(u)\, 1_{(0,\infty)}(J(u))\,du,$$

then

$$\lim_{n\to\infty} n^{-1}\log \Pr\{T_n^{(3)} \ge r+u_n \mid \tau_n = t_n\} = -h_J(r,p),$$

where

(2.3.17) $$h_J(r,p) = \lambda r - \tfrac{1}{2}\lambda\int_{p_0}^{1} J(u)\,du - \sum_{i=1}^{m} p_i \log \cosh(\tfrac{1}{2}\lambda p_i^{-1}\int_{q_{i-1}}^{q_i} J(u)\,du)$$

and $\lambda > 0$ *is the unique solution of the equation*

(2.3.18) $$\sum_{i=1}^{m}\int_{q_{i-1}}^{q_i} J(u)\,du\{1 + \exp(-\lambda p_i^{-1}\int_{q_{i-1}}^{q_i} J(v)\,dv)\}^{-1} = r.$$

(ii) *If*

$$\tfrac{1}{2}\int_{0}^{1-p_0} J(u)\,du < r < \int_{0}^{1-p_0} J(u)\, 1_{(0,\infty)}(J(u))\,du,$$

then

$$\lim_{n\to\infty} n^{-1}\log \Pr\{T_n^{(4)} \ge r+u_n \mid \tau_n = t_n\} = -k_J(r,p),$$

where

(2.3.19) $$k_J(r,p) = \lambda r - \tfrac{1}{2}\lambda\int_{0}^{1-p_0} J(u)\,du - \sum_{i=1}^{m} p_i \log \cosh(\tfrac{1}{2}\lambda p_i^{-1}\int_{q_{i-1}-p_0}^{q_i-p_0} J(u)\,du),$$

and $\lambda > 0$ *is the unique solution of the equation*

(2.3.20) $$\sum_{i=1}^{m}\int_{q_{i-1}-p_0}^{q_i-p_0} J(u)\,du\{1 + \exp(-\lambda p_i^{-1}\int_{q_{i-1}-p_0}^{q_i-p_0} J(v)\,dv)\}^{-1} = r.$$

To define the Bahadur efficiency of *conditional* signed rank tests we use the approach of BAHADUR & RAGHAVACHARI (1972). Let $S = [-m-\tfrac{1}{2},m+\tfrac{1}{2}]^{\infty}$ and let $Y = (Y_1,Y_2,\dots)$, where Y_i is defined by (2.3.1), $i = 1,2,\dots$. Then Y takes values in S according to the product measure $\mathbb{P} = P \times P \times \dots$ for some pm P on $[-m-\tfrac{1}{2},m+\tfrac{1}{2}]$. Since $Y_i = X_i + U_i$, P is absolutely continuous

w.r.t. Lebesgue measure. The statistics T_n defined by (2.3.2) to (2.3.5) and the vectors of ties τ_n will be considered as functions of Y. For each $n \in \mathbb{N}$, B_n will be the σ-field on S induced by the mapping $y \to (y_1, \ldots, y_n)$, $y \in S$. Clearly T_n and τ_n are B_n-measurable for each $n \in \mathbb{N}$. The σ-field on S induced by the mapping $y \to \tau_n(y)$, $y \in S$, will be denoted by C_n (C_n is the *conditioning* field for the statistics T_n).

Let $F_n(t,y) = \mathbb{P}(T_n < t \mid C_n)(y)$ be a regular version of the conditional df of T_n given τ_n. Then the *level attained by* $T_n \mid C_n$ is defined by

(2.3.21) $L_n(y) = 1 - F_n(T_n(y)-, y)$.

Hence, for fixed y, $L_n(y)$ denotes a version of the conditional probability given C_n of T_n being as large as or larger than the observed value $T_n(y)$ if H_0 holds.

Let $X_1, X_2, \ldots$ be distributed according to some pm P_a on $\{j \in \mathbb{Z} : |j| \leq m\}$ which does not satisfy H_a. Then $Y_1, Y_2, \ldots$ are distributed according to some *absolutely continuous* pm P which does not satisfy H_0. Let $\mathbb{P} = P \times P \times \ldots$ be the induced product measure on S. Then, if

$$c(P_a) = - \lim_{n \to \infty} 2n^{-1} \log L_n(Y)$$

exists with $\mathbb{P}$-probability one, we call $c(P_a)$ the *conditional exact slope* of $\{T_n\}$ at P_a. If the sequence $\{T_n^{(i)}\}$ has the conditional exact slope $c_i(P_a)$ at P_a, $i = 1, 2$, and if $c_1(P_a)$ and $c_2(P_a)$ are not both equal to zero, then the *conditional Bahadur efficiency* of $\{T_n^{(1)}\}$ w.r.t. $\{T_n^{(2)}\}$ at P_a is defined by $c_1(P_a)/c_2(P_a)$

THEOREM 2.3.4. *Suppose* P_a *does not satisfy* H_a. *Let* $p = (p_0, \ldots, p_m)$, $p_i = P_a(\{-i, i\})$, $i = 1, \ldots, m$, $p_0 = P_a(\{0\})$, *with* $p_0 < 1$ *and let* J *and* $\{J_n\}$ *satisfy* (2.2.15). *Let* P *be the distribution of* $X_1 + U_1$, *where* U_1 *has a uniform df on* $(-\tfrac{1}{2}, \tfrac{1}{2})$ *and is independent of* X_1. *Furthermore, let* H_P *be defined by* (2.2.2) *and let the statistics* $T_n^{(1)}, \ldots, T_n^{(4)}$ *be defined by* (2.3.2) *to* (2.3.5). *Finally, let* $c_i(P_a)$ *be the conditional exact slope of* $\{T_n^{(i)}\}$ *at* P_a, $i = 1, \ldots, 4$, *and let* $b_1(P_a)$ *and* $b_2(P_a)$ *be defined by*

$$b_1(P_a) = \int_{\tfrac{1}{2}}^{\infty} J(H_P(x)) \, dP(x)$$

$$b_2(P_a) = \int_{\frac{1}{2}}^{\infty} J(H_p(x) - H_p(\tfrac{1}{2})) dP(x).$$

We have

(i) *If*

$$\frac{1}{2} \int_{p_0}^{1} J(u) du < b_1(P_a) < \int_{p_0}^{1} J(u) 1_{(0,\infty)}(J(u)) du,$$

then

$$c_1(P_a) = 2f_J(b_1(P_a), p_0)$$

and

$$c_3(P_a) = 2h_J(b_1(P_a), (p_0, \ldots, p_m)),$$

where f_J and h_J are defined by (2.3.8) and (2.3.17), respectively.

(ii) *If*

$$\frac{1}{2} \int_{0}^{1-p_0} J(u) du < b_2(P_a) < \int_{0}^{1-p_0} J(u) 1_{(0,\infty)}(J(u)) du,$$

then

$$c_2(P_a) = 2g_J(b_2(P_a), p_0)$$

and

$$c_4(P_a) = 2k_J(b_2(P_a), (p_0, \ldots, p_m)),$$

where g_J and k_J are defined by (2.3.15) and (2.3.19), respectively.

<u>PROOF</u>. Let $p_i > 0$, $i = 0, \ldots, m$. We may write

$$b_1(P_a) = \sum_{i=1}^{m} p_i^{-1} P_a(\{i\}) \int_{H_p(i-\frac{1}{2})}^{H_p(i+\frac{1}{2})} J(u) du$$

and

$$b_2(P_a) = \sum_{i=1}^{m} p_i^{-1} P_a(\{i\}) \int_{H_p(i-\frac{1}{2})-p_0}^{H_p(i+\frac{1}{2})-p_0} J(u) du.$$

68

Let P_i be the pm on $(A_i, \mathcal{B}_i)$ defined by

$$P_i(B) = p_i^{-1} P(B), \qquad B \in \mathcal{B}_i,$$

and (with a notation analogous to the notation used in the proof of Theorem 2.3.1) let

$$\hat{\bar{P}}_{n,i}(B) = \tau_{n,i}^{-1} \cdot \#\{j \le n: \, \mathrm{sign}(Y_j)\hat{H}_n(|Y_j|) \in B, Y_j \in A_i\}, \qquad B \in \mathcal{B},$$

where $i = 0, \ldots, m$, and $\tau_{n,i}$ is the tie of $Y_1, \ldots, Y_n$ at i. Then

$$T_n^{(1)} = n^{-1} \sum_{i=1}^{m} \tau_{n,i} \int_{(0,1]} J_n(u) \, d\hat{\bar{P}}_{n,i}(u)$$

and

$$b_1(P_a) = \sum_{i=1}^{m} p_i \int_{(0,1]} J(u) \, d\bar{P}_i(u).$$

Let

$$W = \{\nu \in [0,1]^{m+1}: \, \nu = (\nu_0, \ldots, \nu_m), \, \sum_{i=0}^{m} \nu_i = 1\}$$

be endowed with the (relative) Euclidean topology and let Γ be defined as in the proof of Theorem 2.3.1 (endowed with the ρ-topology). Moreover, let $U \times \Gamma$ be endowed with the product topology. Then the mapping

$$(\nu, Q) \to \sum_{i=1}^{m} \nu_i \int_{(0,1]} J_n(u) \, d\bar{Q}_i(u), \quad \nu = (\nu_0, \ldots, \nu_m), \quad Q = (Q_0, \ldots, Q_m),$$

is continuous at $((p_0, \ldots, p_m), (P_0, \ldots, P_m))$ for each fixed n. Since

$$\sum_{i=1}^{m} p_i \int_{(0,1]} |J_n(u) - J(u)| \, d\bar{P}_i(u) \to 0, \qquad n \to \infty,$$

it follows that $T_n^{(1)} \to b_1(P_a)$, as $n \to \infty$, with probability one under P. Likewise, also with probability one under P,

$$T_n^{(3)} \to b_1(P_a), \qquad \text{as } n \to \infty,$$

and for $i = 2, 4$,

$$T_n^{(i)} \to b_2(P_a), \qquad \text{as } n \to \infty.$$

Only trivial changes in the proof are necessary if $p_i = 0$ for some i. The result now follows from Theorems 2.3.1, 2.3.2 and 2.3.3. $\quad\square$

The theorem implies that the exact slope of a signed rank test with randomized ranks never exceeds the exact slope of a signed rank test based on the average score method if the score functions J_n and the method of dealing with zeros are the same for the two tests. We shall prove this for tests based on statistics of type (2.3.2) and (2.3.4) and omit the (quite similar) proof for tests based on statistics of type (2.3.3) and (2.3.5).

Let $p_i = H_p(i+\tfrac{1}{2}) - H_p(i-\tfrac{1}{2}) > 0$ and $q_i = H_p(i+\tfrac{1}{2})$, $i = 0,\ldots,m$. By Jensen's inequality we have for fixed $\lambda \geq 0$,

$$(2.3.22) \qquad \sum_{i=1}^{m} p_i \log \cosh(\tfrac{1}{2}\lambda p_i^{-1} \int_{q_{i-1}}^{q_i} J(u)\,du) \leq \int_{p_0}^{1} \log \cosh(\tfrac{1}{2}\lambda J(u))\,du.$$

Furthermore, if $\lambda_1 > 0$ satisfies (2.3.9),

$$\lambda_1 r - \tfrac{1}{2}\lambda_1 \int_{p_0}^{1} J(u)\,du - \int_{p_0}^{1} \log \cosh(\tfrac{1}{2}\lambda_1 J(u))\,du$$

(2.3.23)

$$= \max_{\lambda > 0}\{\lambda r - \tfrac{1}{2}\lambda \int_{p_0}^{1} J(u)\,du - \int_{p_0}^{1} \log \cosh(\tfrac{1}{2}\lambda J(u))\,du\}.$$

Similarly, if $\lambda_3 > 0$ satisfies (2.3.18),

$$\lambda_3 r - \tfrac{1}{2}\lambda_3 \int_{p_0}^{1} J(u)\,du - \sum_{i=1}^{m} p_i \log \cosh(\tfrac{1}{2}\lambda_3 p_i^{-1} \int_{q_{i-1}}^{q_i} J(u)\,du)$$

(2.3.24)

$$= \max_{\lambda > 0}\{\lambda r - \tfrac{1}{2}\lambda \int_{p_0}^{1} J(u)\,du - \sum_{i=1}^{m} p_i \log \cosh(\tfrac{1}{2}\lambda p_i^{-1} \int_{q_{i-1}}^{q_i} J(u)\,du)\}.$$

Hence, by (2.3.22), (2.3.23) and (2.3.24)

$$f_J(r,p_0) \leq h_J(r,p),$$

where $f_J(r,p_0)$ is defined by (2.3.8) and $h_J(r,p)$ is defined by (2.3.17). This implies by Theorem 2.3.4 that $c_1(P_a) \leq c_3(P_a)$ if P_a is a fixed

70

alternative to H_a satisfying

$$\tfrac{1}{2} \int_{p_0}^{1} J(u)\,du < b_1(P_a) < \int_{p_0}^{1} J(u)\,1_{(0,\infty)}(J(u))\,du.$$

Analogous results on Pitman efficiencies of signed rank tests with different methods of dealing with ties are given in VORLÍČKOVÁ (1972) and CONOVER (1973a,1973b). In fact, under conditions analogous to those of Theorem 2.2.3 Pitman efficiencies are limits of Bahadur efficiencies as the alternatives tend to the null hypothesis.

CHAPTER III

THE EFFICIENCIES OF SOME TESTS FOR
THE MULTIVARIATE LINEAR HYPOTHESIS

3.1. INTRODUCTION

Let the row vectors of the $N \times m$ random matrix Y be distributed independently according to an m-variate normal distribution with covariance matrix Σ. Suppose

$$\mathcal{E}Y = XB,$$

where X is an $N \times q$ (design-) matrix of rank $q < N$ and B is an unknown $q \times m$ matrix of regression coefficients. We write $n = N-q$. The multivariate linear hypothesis is given by

$$H_0 : AB = 0,$$

where A is an $n_1 \times q$ matrix of rank $n_1 \leq q$. We shall consider tests of H_0 based on the matrix of sums of squares and cross-products due to the hypothesis

$$S_{h,N} = Y'X(X'X)^{-1}A'[A(X'X)^{-1}A']^{-1}A(X'X)^{-1}X'Y$$

and the matrix of sums of squares and cross-products due to error

$$S_{e,N} = Y'(I_N - X(X'X)^{-1}X')Y,$$

where I_N denotes the $N \times N$ identity matrix. The matrix $S_{e,N}$ has a central Wishart distribution $W_m(n,\Sigma)$ and the matrix $S_{h,N}$ has a non-central Wishart distribution $W_m(n_1,\Sigma,N\Omega)$, where

$$\Omega = N^{-1}B'A'[A(X'X)^{-1}A']^{-1}AB\Sigma^{-1}.$$

As an example consider the one-way analysis of variance set-up with $X = I_q \otimes J_k$, $N = qk$, $J_k = \overline{(1,\ldots,1)}'$ and $B = (\mu_1',\ldots,\mu_q')'$, where $\mu_i = (\mu_{i1},\ldots,\mu_{im})$ denotes the mean of the observations from the i-th group, $i = 1,\ldots,q$, and $\otimes$ denotes the Kronecker product of two matrices. The hypothesis H_0 of equality of the means $\mu_1,\ldots,\mu_q$ is given by

$$H_0 : AB = 0,$$

where
$$A = \begin{pmatrix} 1 & 0\ldots0 & -1 \\ 0 & 1\ldots0 & -1 \\ & \ddots & \\ 0\ldots\ldots1 & & -1 \end{pmatrix}$$

An easy computation yields:

$$\Omega = q^{-1} \sum_{i=1}^{q} (\mu_i-\bar{\mu})'(\mu_i-\bar{\mu})\Sigma^{-1},$$

where $\bar{\mu} = q^{-1} \sum_{i=1}^{q} \mu_i$. Note that for fixed alternatives $B = (\mu_1',\ldots,\mu_q')'$ the matrix $N\Omega = \mathcal{O}(N)$, as $N \to \infty$.

Well-known tests for testing H_0 are based on the statistics

(3.1.1) $\quad t_N^{(1)} = \text{tr } S_{h,N}(S_{h,N}+S_{e,N})^{-1}$

(3.1.2) $\quad t_N^{(2)} = \text{tr } S_{h,N} S_{e,N}^{-1}$

(3.1.3) $\quad t_N^{(3)} = \text{largest latent root of } S_{h,N} S_{e,N}^{-1}$

and

(3.1.4) $\quad t_N^{(4)} = -\log|S_{e,N}(S_{e,N}+S_{h,N})^{-1}|,$

where $|A|$ denotes the determinant of the matrix A. The tests based on $t_N^{(1)}$ to $t_N^{(4)}$ are (generally) called Pillai's test, Hotelling's T_0^2, Roy's largest root test and the likelihood ratio (LR) test, respectively.
In each case, H_0 is rejected for large values of the test statistic.

Numerical comparisons of the power functions have been made by PILLAI & JAYACHANDRAN (1967) in the case of 2-dimensional normal distributions.

The general picture which emerges from these comparisons is that Pillai's test behaves rather well against alternatives for which the latent roots of the non-centrality matrix $N\Omega$ are small or close to one another, whereas Hotelling's T_0^2 and Roy's largest root test are more powerful against alternatives for which $N\Omega$ has one large root and one small root. The LR test seems to occupy an intermediate position.

Asymptotic expansions of the distribution functions of the test statistics $t_N^{(1)}$, $t_N^{(2)}$ and $t_N^{(4)}$ under H_0 and under contiguous alternatives (i.e. tr $N\Omega$ bounded away from 0 and ∞ as $N \to \infty$) have been given by FUJIKOSHI (1970). He uses these expansions to compute approximate powers in the 3-dimensional case (m=3), with the general conclusion that Pillai's test is more powerful than the LR test and the LR test is more powerful than Hotelling's T_0^2 if the roots of $N\Omega$ are close to one another, while the reverse seems to be true if the largest latent root of $N\Omega$ dominates the other roots.

ROY, GNANADESIKAN & SRIVASTAVA (1971) have done some simulation experiments for the 2-dimensional case. These experiments essentially yield the same picture of the power behavior of the tests with the additional finding that Roy's test seems to be more powerful than the other tests when there is only large latent root of $N\Omega$ different from zero.

Finally, as a finite sample result, we mention that Pillai's test is locally best among invariant level-α tests of H_0 which depend only on the latent roots of $S_{h,N}S_{e,N}^{-1}$ (see SCHWARTZ (1967a), Theorem 1) and that each of the four tests is admissible (SCHWARTZ (1967b)).

All these results suggest that the power functions of the tests depend critically on the *product* of the sample size N and the matrix Ω. Different pictures of the relationships between the tests arise in the following situations:

(i) Ω fixed, $N \to \infty$.

In this situation Bahadur efficiencies and deficiencies can be computed (see section 3.3), showing a marked superiority of the LR test. The LR test is only "deficient" with respect to some of the other tests on certain halflines of the parameter space.

(ii) $\Omega = \Omega_N$ depends on the sample size N, $\{tr \ \Omega_N\}$ is a bounded sequence and N tr $\Omega_N \to \infty$ (note that (i) is a special case of (ii)).

To give an example of what happens in this case, we mention that the deficiency of the LR test w.r.t. Roy's largest root test is of order $\theta_{1,N}^{-1} \log (N\theta_{1,N})$, if $\theta_{1,N} \to 0$, $N\theta_{1,N} \to \infty$ and $N\theta_{j,N} = \mathcal{O}(1)$, $j \geq 2$, where

$\theta_{1,N} \geq \ldots \geq \theta_{m,N}$ are the latent roots of Ω_N (this is the case of non-contiguous local alternatives). Hence, if $\theta_{1,N} = k \cdot N^{-1+\varepsilon}$ for some constant $k > 0$ and some small $\varepsilon > 0$, the deficiency is "almost" of order N (see section 3.4).

(iii) $N\Omega = N\Omega_N = M$, where M is some fixed positive semidefinite matrix. This is the familiar case of contiguous alternatives. The Pitman efficiencies of Pillai's test, Hotelling's T_0^2 and the LR test w.r.t. one another are equal to one. In section 3.4 the Hodges-Lehmann deficiencies of these tests are computed, showing that Pillai's test is superior to the other two tests if $|\theta_{1,N}/\theta_{j,N} - 1| < \varepsilon$, for all $N \in \mathbb{N}$, $j = 2,\ldots,m$ and small $\varepsilon > 0$, while Hotelling's T_0^2 is better than the other two tests when $\theta_{j,N}/\theta_{1,N} < \varepsilon$ for all $N \in \mathbb{N}$, $j = 2,\ldots,m$ and small $\varepsilon > 0$, where $\theta_{1,N},\ldots,\theta_{m,N}$ are again the latent roots of Ω_N. In contrast to this, the LR-test is only deficient in the sense of Bahadur deficiency w.r.t. Pillai's test if $\theta_1/\theta_j = 1$, $j = 2,\ldots,m$ and only deficient in this sense to Hotelling's T_0^2 if $\theta_j/\theta_1 = 0$, $j = 2,\ldots,m$, where $\theta_1,\ldots,\theta_m$ are the latent roots of the fixed non-centrality matrix Ω (cf. (i)).

The distribution of Roy's test statistic under contiguous alternatives is rather complicated; presumably the Pitman efficiency of Roy's test w.r.t. the other tests is less than one if $\theta_{1,N} = \ldots = \theta_{m,N}$ and greater than one if $\theta_{j,N}/\theta_{1,N} = 0$, $j = 2,\ldots,m$.

(iv) N fixed, $\Omega \to 0$.

In this case Pillai's test is superior by the result of SCHWARTZ (1967a).

The result of SCHWARTZ (1967a) provides an analytical explanation of the numerical finding by PILLAI & JAYACHANDRAN (1967) that Pillai's test is the best test against alternatives for which the latent roots of Ω are small. It is the purpose of this chapter to treat the situations mentioned under (i), (ii) and (iii). Thus we obtain a more complete analytical description of the asymptotic power behavior of the four tests considered. The results not only support the numerical results described before, but also explain why the product of the sample size N and the magnitude of tr Ω is so important in determining the power of the tests. The role played by the ratios of the latent roots of Ω also becomes more transparent.

The results are based on theorems on probabilities of large and moderate deviations of the test statistics under H_0. These theorems are given in section 3.2. We shall also need expansions of the distribution functions of the test statistics under fixed and local alternatives. These

expansions are given in the appendix (section 3.5). Some efforts are made to prove the validity of the expansions rigorously, since the known expansions in this field are only formal ones.

3.2. PROBABILITIES OF LARGE AND MODERATE DEVIATIONS OF THE TEST STATISTICS UNDER THE NULL HYPOTHESIS

Using the notation of section 3.1, we assume that $S_{e,N}$ has the Wishart distribution $W_m(n,I)$ and that $S_{h,N}$ has the Wishart distribution $W_m(n_1,I)$ under H_0, where I is the identity matrix. Replacing Σ by I causes no loss of generality, since each statistic is a function of $S_{h,N}S_{e,N}^{-1}$. Moreover we shall assume that n_1 is a *fixed* number and that $n_1 \geq m$, since the results for the case $n_1 < m$ can be obtained by the well-known substitutions $n_1 \rightarrow m$, $m \rightarrow n_1$, and $n \rightarrow n+n_1-m$. Unless otherwise specified, we assume $m \geq 2$.

First asymptotic expressions for large and moderate deviation probabilities of Pillai's test statistic will be given. The matrix

$$(S_{h,N} + S_{e,N})^{-\frac{1}{2}} S_{h,N} (S_{h,N} + S_{e,N})^{-\frac{1}{2}}$$

has under H_0 the multivariate beta density

(3.2.1) $f_{n,n_1}(s) = c_n |S(s)|^{\frac{1}{2}n_1-p} |I-S(s)|^{\frac{1}{2}n-p}$,

where

(3.2.2) $c_n = \Gamma_m(\frac{1}{2}(n_1+n)) / \{\Gamma_m(\frac{1}{2}n_1)\Gamma_m(\frac{1}{2}n)\}$

(3.2.3) $\Gamma_m(a) = \pi^{\frac{1}{4}m(m-1)} \prod_{i=1}^{m} \Gamma(a-\frac{1}{2}(i-1))$

(3.2.4) $p = \frac{1}{2}(m+1)$,

and $S(s)$ is a (symmetric) positive definite $m \times m$ matrix whose upper triangle is given by the vector $s = (s_{11}, \ldots, s_{mm}, s_{12}, s_{13}, \ldots, s_{m-1,m})' \in \mathbb{R}^{mp}$. The df of $t_N^{(1)}$, defined by (3.1.1), is given by

(3.2.5) $F_N^{(1)}(x) = \int_{\{tr\ S(s) \leq x, 0 < S(s) < I\}} f_{n,n_1}(S(s))ds$,

where $0 < S(s) < I$ means that the matrices $S(s)$ and $I-S(s)$ are both positive definite and where ds denotes the Lebesgue measure $ds_{11} \cdots ds_{mm} ds_{12} \cdots ds_{m-1,m}$ on $\mathbb{R}^{mp}$.

In the sequel m will always denote the order of the matrix $S(s)$ (and hence the dimension of the multivariate observations) and p will denote $\frac{1}{2}(m+1)$. The space of positive definite $m \times m$ matrices will be denoted by S_m.

THEOREM 3.2.1. *Let* $F_N^{(1)}$ *be the df of Pillai's test statistic* $t_N^{(1)}$ *under* H_0. *Let* $\{x_n\}$ *be a sequence of real numbers in the interval* $(0,1)$, *bounded away from 1 and satisfying* $nx_n^2 \to \infty$ *as* $n \to \infty$. *Then, as* $n \to \infty$,

$$(3.2.6) \quad 1 - F_N^{(1)}(mx_n) \sim c_n 2^{\frac{1}{2}m}(mn\pi)^{-\frac{1}{2}}(\frac{2\pi}{n})^{\frac{1}{2}mp}(1-x_n)^{\frac{1}{2}mn}x_n^{m(\frac{1}{2}n_1-p)},$$

where c_n *is given by* (3.2.2).

PROOF. By (3.2.1) and (3.2.5) we have

$$1 - F_N^{(1)}(mx_n) =$$

$$= c_n \int_{\{0<S(s)<I,\text{tr } S(s)\geq mx_n\}} |S(s)|^{\frac{1}{2}n_1-p}|I-S(s)|^{\frac{1}{2}n-p}ds$$

$$= c_n x_n^{\frac{1}{2}mn_1} \int_{\{0<x_n S(s)<I,\text{tr } S(s)\geq m\}} |S(s)|^{\frac{1}{2}n_1-p}|I-x_n S(s)|^{\frac{1}{2}n-p}ds.$$

By the transformation

$$\begin{cases} u_{11} = \text{tr } S(s) \\ u_{ij} = s_{ij} \quad , (i,j) \neq (1,1), \end{cases}$$

the last integral can be written

$$(3.2.7) \quad c_n x_n^{\frac{1}{2}mn_1} \int_{\{0<x_n S(u)<I,u_{11}\geq m\}} |S(u)|^{\frac{1}{2}n_1-p}|I-x_n S(u)|^{\frac{1}{2}n-p}du,$$

where $u = (u_{11},u_{22},\ldots,u_{mm},u_{12},\ldots,u_{m-1,m})'$ and (with a slight abuse of notation) $S(u) = S(s(u))$.

Let $E = \{u \in \mathbb{R}^{mp} : s(u)$ represents the upper triangle of a matrix $S(u) \in S_m\}$ and let $D_n = \{u \in E : 0 < x_n S(u) < I, u_{11} \geq m\}$. We define the functions $g_n : D_n \to \mathbb{R}$ and $h_n : D_n \to \mathbb{R}$ by

$$g_n(u) = \log|I-x_n S(u)|, \quad u \in D_n$$

and

$$h_n(u) = |S(u)|^{\frac{1}{2}n\frac{1}{1}-p}|I-x_n S(u)|^{-p}, \qquad u \in D_n.$$

The function g_n attains a maximum in D_n at $u_0 = (\overbrace{m,1,\ldots,1}^{m},0,\ldots,0)$' and the asymptotic behavior (as $n \to \infty$) of the integral (3.2.7) is completely determined by the behavior of the integrand in a neighborhood of u_0.

Let for $\varepsilon > 0$, $K_\varepsilon = [m,m+\varepsilon]$, $L_\varepsilon = [1-\varepsilon,1+\varepsilon]$, $M_\varepsilon = [-\varepsilon,\varepsilon]$ and

$$W_\varepsilon = K_\varepsilon \times L_\varepsilon^{m-1} \times M_\varepsilon^{\frac{1}{2}m(m-1)}.$$

Fix $\varepsilon > 0$ such that $W_\varepsilon \subset D_n$ (this inclusion is satisfied for all sufficiently small ε and all n) and let $\phi_n : W_\varepsilon \to \mathbb{R}$ be defined by

$$\phi_n(u) = h_n(u) / \frac{\partial}{\partial u_{11}} g_n(u), \qquad u \in W_\varepsilon.$$

Since $\partial/\partial u_{11} \, g_n(u) = -x_n a_{11}(u)$, where $a_{11}(u)$ is the element in the first row and first column of $(I-x_n S(u))^{-1}$, the function ϕ_n is well defined on W_ε (note that $a_{11}(u) > 0$, $u \in W_\varepsilon$, since $(I-x_n S(u))^{-1}$ is positive definite for these values of u). We have

$$(3.2.8) \qquad \int_{W_\varepsilon} h_n(u) \exp\{\tfrac{1}{2}n g_n(u)\}du = 2n^{-1} \int_{W_\varepsilon} \frac{\partial}{\partial u_{11}} (\phi_n(u) \exp\{\tfrac{1}{2}n g_n(u)\})du$$

$$-2n^{-1} \int_{W_\varepsilon} \exp\{\tfrac{1}{2}n g_n(u)\} \frac{\partial}{\partial u_{11}} \phi_n(u)du.$$

Let $u^{(2)} = (u_{22},\ldots,u_{m-1,m})$' and let $W_\varepsilon^{(2)} = L_\varepsilon^{m-1} \times M_\varepsilon^{\frac{1}{2}m(m-1)}$. Then, by Fubini's theorem, the first integral at the right-hand side of (3.2.8) is equal to

$$2n^{-1} \int_{W_\varepsilon^{(2)}} (\phi_n(m+\varepsilon,u^{(2)}) \exp\{\tfrac{1}{2}n g_n(m+\varepsilon,u^{(2)})\}$$

$$-\phi_n(m,u^{(2)}) \exp\{\tfrac{1}{2}n g_n(m,u^{(2)})\})du^{(2)}.$$

In a neighborhood of $u_0^{(2)} = (\overbrace{1,\ldots,1}^{m-1},0,\ldots,0)$' one has the expansion

$(3.2.9)$ $g_n(m,u^{(2)}) = m \log (1-x_n) + \log|I-x_n(1-x_n)^{-1}(S(m,u^{(2)})-I)|$

$= m \log(1-x_n) - \tfrac{1}{2}x_n^2(1-x_n)^{-2}\text{tr}(S(m,u^{(2)})-I)^2(1+\mathcal{O}(1)), \quad u^{(2)} \to u_0^{(2)}.$

Hence, by Laplace's method for multi-dimensional integrals (cf. BLEISTEIN & HANDELSMAN (1975), Ch. 8, Section 3)

$(3.2.10)$ $\displaystyle - \int_{W_\varepsilon^{(2)}} \phi_n(m,u^{(2)})\exp\{\tfrac{1}{2}ng_n(m,u^{(2)})\}du^{(2)}$

$\displaystyle = \int_{W_\varepsilon^{(2)}} (x_n a_{11}(u))^{-1}h_n(m,u^{(2)})\exp\{\tfrac{1}{2}ng_n(m,u^{(2)})\}du^{(2)}$

$\displaystyle \sim x_n^{-1}(1-x_n)^{-mp+1} \int_{W_\varepsilon^{(2)}} \exp\{\tfrac{1}{2}ng_n(m,u^{(2)})\}du^{(2)}$

$\displaystyle \sim x_n^{-1}(1-x_n)^{\frac{1}{2}mn-mp+1} \int_{W_\varepsilon^{(2)}} \exp\{-\tfrac{1}{4}nx_n^2(1-x_n)^{-2}\text{tr}(S(m,u^{(2)})-I)^2\}du^{(2)}$

$\displaystyle \sim x_n^{-mp}(1-x_n)^{\frac{1}{2}mn}(\tfrac{1}{2}n)^{-\frac{1}{2}(mp-1)} \int_{\mathbb{R}^{mp-1}} \exp(-\tfrac{1}{2}k(w))dw,$

where $k(w) = k(w_1,\dots,w_{mp-1})$ is the quadratic form

$(3.2.11)$ $\displaystyle k(w) = 2\sum_{i=1}^{m-1} w_i^2 + 2\sum_{i<j<m} w_i w_j + 2\sum_{i=m}^{mp-1} w_i^2.$

The last asymptotic equivalence in $(3.2.10)$ is obtained by making a shift to the origin (followed by an application of Laplace's method).

Since $(2\pi)^{-\frac{1}{2}(mp-1)}m^{\frac{1}{2}}2^{\frac{1}{4}m(m-1)}\exp\{-\tfrac{1}{2}k(w)\}$ is the density of an $(mp-1)$-variate normal distribution, one has

$(3.2.12)$ $\displaystyle \int_{\mathbb{R}^{mp-1}} \exp\{-\tfrac{1}{2}k(w)\}dw = (2\pi)^{\frac{1}{2}(mp-1)}m^{-\frac{1}{2}}2^{-\frac{1}{4}m(m-1)}$

and hence by $(3.2.10)$ and $(3.2.12)$

$$(3.2.13) \quad - \int\limits_{W_\epsilon^{(2)}} \phi_n(m,u^{(2)}) \exp\{\tfrac{1}{2}ng_n(m,u^{(2)})\}du^{(2)}$$

$$\sim (2\pi)^{\frac{1}{2}(mp-1)} m^{-\frac{1}{2}} 2^{-\frac{1}{2}m(m-1)} x_n^{-mp}(1-x_n)^{\frac{1}{2}mn}(\tfrac{1}{2}n)^{-\frac{1}{2}(mp-1)}.$$

It is easily seen that the integrals

$$\int\limits_{W_\epsilon^{(2)}} \phi_n(m+\epsilon,u^{(2)}) \exp\{\tfrac{1}{2}ng_n(m+\epsilon,u^{(2)})\}du^{(2)},$$

$$\int\limits_{W_\epsilon} \exp\{\tfrac{1}{2}ng_n(u)\} \frac{\partial}{\partial u_{11}} \phi_n(u)\,du,$$

and $\quad \tfrac{1}{2}n \int\limits_{D_n\backslash W_\epsilon} h_n(u)\exp\{\tfrac{1}{2}ng_n(u)\}du,$

are of lower order than the integral in (3.2.13), as $n \to \infty$.
The result now follows from (3.2.7), (3.2.8) and (3.2.13). $\square$

REMARK 3.2.1. By Stirling's formula one has

$$(3.2.14) \quad c_n \sim \Gamma_m(\tfrac{1}{2}n_1)^{-1}(\tfrac{1}{2}n)^{\frac{1}{2}mn_1}, \qquad n \to \infty.$$

REMARK 3.2.2. We briefly indicate a different proof of Theorem 3.2.1. The moment generating function of $nt_N^{(1)}$ is given by the hypergeometric function of matrix argument

$$m_n(t) = {}_1F_1(\tfrac{1}{2}n_1;\tfrac{1}{2}(n+n_1);ntI).$$

For simplicity we shall only consider the case $x_n = x \in (0,1)$ for all $n \in \mathbb{N}$.

For large n and fixed $t > 1$ the "peak" of the integrand of the integral

$$m_n(\tfrac{1}{2}t) = c_n \int\limits_{0<S(s)<I} \exp\{\tfrac{1}{2}nt\,\mathrm{tr}\,S(s)\}\,|S(s)|^{\frac{1}{2}n_1-p}|I-S(s)|^{\frac{1}{2}n-p}ds$$

is reached at the interior point $t^{-1}(1-t)(\overbrace{1,\ldots,1}^{m},0,\ldots,0)'$ of the domain of integration. Hence, by Laplace's method

$$(3.2.15) \quad m_n(\tfrac{1}{2}t) \sim k_n\, e^{\frac{1}{2}mnt} t^{-\frac{1}{2}mn}(t^{-1}(1-t))^{m(\frac{1}{2}n_1-p)}, \qquad n \to \infty,$$

where

$$k_n = c_n \, 2^{\frac{1}{2}m} (2\pi n^{-1})^{\frac{1}{2}mp} \, e^{-\frac{1}{2}mn}.$$

The moment generating function $m_n(\frac{1}{2}t)$ can be inverted to obtain an asymptotic expression for the density $f_n(mnx)$ of $nt_N^{(1)}$

$$(3.2.16) \qquad f_n(mnx) \sim \frac{k_n}{4\pi i} \int_{c-i\infty}^{c+i\infty} e^{\frac{1}{2}mn(1-x)s} \, s^{-\frac{1}{2}mn} (s^{-1}(1-s))^{m(\frac{1}{2}n_1-p)} ds,$$

where $c = (1-x)^{-1}$ is the saddlepoint of $m_n(\frac{1}{2}s)\exp(-\frac{1}{2}mnxs)$ for large n. But the right-hand side of (3.2.16) is in fact a well-known integral representation of a *one-dimensional* hypergeometric function (see ERDÉLYI et al. (1953), p. 273). Hence

$$f_n(mnx) \sim \frac{1}{2}k_n(\frac{1}{2}mn(1-x))^{\frac{1}{2}mn-1}\Gamma(\frac{1}{2}mn)^{-1} \, {}_1F_1(-m(\frac{1}{2}n_1-p); \, \frac{1}{2}mn; \, \frac{1}{2}mn(1-x))$$

$$\sim \frac{1}{2}k_n(\frac{1}{2}mn(1-x))^{\frac{1}{2}mn-1}\Gamma(\frac{1}{2}mn)^{-1} \, x^{m(\frac{1}{2}n_1-p)},$$

by ERDÉLYI et al. (1953), p.280, (18), implying

$$1 - F_N^{(1)}(mx) = mn \int_x^\infty f_n(mny)\,dy$$

$$\sim c_n \, 2^{\frac{1}{2}m}(mn\pi)^{-\frac{1}{2}}(2\pi n^{-1})^{\frac{1}{2}mp}(1-x)^{\frac{1}{2}mn} \, x^{m(\frac{1}{2}n_1-p)}.$$

We note that in using this method to find asymptotic expressions for large deviation probabilities, it is necessary to look for a saddlepoint of the function $m_n(\frac{1}{2}s)\exp(-\frac{1}{2}mnsx)$ before inverting. For example, in the domain $0 \le t < 1$, the function $m_n(\frac{1}{2}t)$ satisfies the relation

$$(3.2.17) \qquad m_n(\frac{1}{2}t) \sim (1-t)^{-\frac{1}{2}mn_1}, \qquad n \to \infty.$$

Inversion of the right-hand side of (3.2.17) leads to tail probabilities of a $\chi^2_{mn_1}$-distribution (the limiting distribution under H_0), contradicting (3.2.16).

REMARK 3.2.3. In the statistical literature only the probabilities $1-F_N^{(1)}(mx)$ for *fixed* $x \in (0,1)$ (and large N) are generally called *large deviation* probabilities. The case $x_n \to 0$, $nx_n \to \infty$, corresponds to probabilities of

moderate deviations.

We next turn to Hotelling's T_0^2 and Roy's largest root test. It was seen in the proof of Theorem 3.2.1 that the asymptotic behavior of certain probabilities of large and moderate deviations of Pillai's test statistic is determined by an integral over the neighborhood of a "critical point" which is located in a smooth piece of the boundary of the domain of integration. We shall see that the asymptotic behavior of these probabilities in the case of Hotelling's T_0^2 and Roy's largest root test is likewise determined by an integral over the neighborhood of a critical point. But this time the critical point does not belong to a smooth piece of the boundary, but it is located in a "corner" of the boundary.

We shall need the density of the latent roots of the matrix $S_{h,N} S_{e,N}^{-1}$ under H_0. This density is given by

$$(3.2.18) \quad c_n \pi^{\frac{1}{2}m^2} \Gamma_m(\tfrac{1}{2}m)^{-1} \prod_{i=1}^{m} \lambda_i^{\frac{1}{2}(n_1-m-1)} \prod_{i=1}^{m} (1+\lambda_i)^{-\frac{1}{2}(n_1+n)} \prod_{i<j} (\lambda_i - \lambda_j),$$

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m \geq 0,$$

where c_n is given by (3.2.2) (see ANDERSON (1958), Ch.13).

For Hotelling's T_0^2 we have the following result.

THEOREM 3.2.2. *Let* $F_N^{(2)}$ *be the df of Hotelling's* T_0^2 *statistic* $t_N^{(2)}$ *under* H_0. *Let* $\{x_n\}$ *be a bounded sequence of real numbers satisfying* $nx_n^2 \to \infty$, *as* $n \to \infty$. *Then, as* $n \to \infty$,

$$(3.2.19) \quad 1 - F_N^{(2)}(x_n) \sim c_n \pi^{\frac{1}{2}m} \Gamma(\tfrac{1}{2}m)^{-1} \Gamma_{m-1}(\tfrac{1}{2}(n_1-1)) (2/n)^{\frac{1}{2}(n_1-1)(m-1)+1}$$

$$\circ \ x_n^{-\frac{1}{2}(m-2)(n_1-2)} (1+x_n)^{-\frac{1}{2}(n-1)+\frac{1}{2}(m-2)(n_1-1)},$$

where c_n *is given by* (3.2.2).

PROOF. Let the function $G_n : (0,\infty) \to \mathbb{R}$ be defined by

$$(3.2.20) \quad G_n(x) = \int \cdots \cdots \int_{\{\sum_{i=1}^{m}\lambda_i \geq x, \lambda_1 \geq \ldots \geq \lambda_m \geq 0\}} \prod_{i=1}^{m} \lambda_i^{\frac{1}{2}n_1-p} \prod_{i=1}^{m} (1+\lambda_i)^{-\frac{1}{2}(n_1+n)}$$

$$\circ \prod_{i<j} (\lambda_i - \lambda_j) d\lambda_1 \ldots d\lambda_m$$

$$= x^{\frac{1}{2}mn_1} \int \cdots \cdots \int_{\{\sum_{i=1}^{m}\lambda_i \geq 1, \lambda_1 \geq \ldots \geq \lambda_m \geq 0\}} \prod_{i=1}^{m} \lambda_i^{\frac{1}{2}n_1-p}$$

$$\cdot \prod_{i=1}^{m} (1+x\lambda_i)^{-\frac{1}{2}(n_1+n)}$$

$$\cdot \prod_{i<j} (\lambda_i - \lambda_j) d\lambda_1 \ldots d\lambda_m.$$

We make the following change of variables in the last integral of (3.2.20)

$$\begin{cases} u_1 = \sum_{i=1}^{m} \lambda_i - 1 \\ u_i = \lambda_i - \lambda_{i+1}, \quad 1 < i < m \\ u_m = \lambda_m \end{cases}$$

The Jacobian of this transformation is equal to one and the inverse transformation is given by

$$\begin{cases} \lambda_1 = u_1 + 1 - \sum_{i=2}^{m} (i-1)u_i \\ \lambda_i = \sum_{j=i}^{m} u_j, \quad i \geq 2 \end{cases}$$

By this change of variables $G_n(x)$ can be written as

$$x^{\frac{1}{2}mn_1} \int \cdots \cdots \int_{\{u_1+1 \geq \sum_{i=2}^{m} iu_i, u_i \geq 0, i=1,\ldots,m\}} \{(u_1+1-\sum_{i=2}^{m} (i-1)u_i)$$

$$\cdot \prod_{i=2}^{m} (\sum_{j=i}^{m} u_j)\}^{\frac{1}{2}n_1-p}\{(1+x(u_1+1-\sum_{i=2}^{m} (i-1)u_i))$$

$$\cdot \prod_{i=2}^{m} (1+x\sum_{j=i}^{m} u_j)\}^{-\frac{1}{2}(n_1+n)} \prod_{1<i\leq j<m} (\sum_{k=i}^{j} u_k)$$

$$\cdot \prod_{j=2}^{m} (u_1+1-\sum_{i=2}^{m} (i-1)u_i-\sum_{k=j}^{m} u_k) du_1 \ldots du_m.$$

Let

$$\phi_n(u_1,\ldots,u_m) = \log(1+x_n(u_1+1-\sum_{i=2}^{m} (i-1)u_i))+\sum_{i=2}^{m} \log(1+x_n\sum_{j=i}^{m} u_j).$$

The function ϕ_n is defined for all $(u_1,\ldots,u_m)$' such that $u_1 + 1 \geq \sum_{i=2}^{m} i u_i$ and $u_i \geq 0$, $i = 1,\ldots,m$.

Moreover

$$\phi_n(u_1,\ldots,u_m) = \log(1+x_n) + (1+x_n)^{-1} x_n u_1 (1+o(1))$$

$$+ (1+x_n)^{-1} x_n^2 \sum_{i=2}^{m} (i-1) u_i (1+o(1)),$$

$$u_i \downarrow 0, \qquad i = 1,\ldots,m.$$

Hence, by Laplace's method,

$$(3.2.21) \qquad G_n(x_n) \sim x_n^{\frac{1}{2}mn_1} (1+x_n)^{-\frac{1}{2}(n+n_1)} \int_0^\infty \exp\{-\tfrac{1}{2}n(1+x_n)^{-1} x_n u_1\} du_1$$

$$\cdot \int_0^\infty \cdots \int_0^\infty \exp\{-\tfrac{1}{2}n(1+x_n)^{-1} x_n^2 \sum_{i=2}^{m} (i-1) u_i\}$$

$$\cdot \prod_{i=2}^{m} \left(\sum_{j=i}^{m} u_j\right)^{\frac{1}{2}n_1-p} \prod_{1<i\leq j<m} \left(\sum_{k=i}^{j} u_k\right) du_2 \cdots du_m.$$

Inverting the previous transformation, the integral over the variables $u_2,\ldots,u_m$ can be written as

$$(3.2.22) \qquad \int_{\lambda_2 \geq \ldots \geq \lambda_m \geq 0} \cdots \int \exp\{-\tfrac{1}{2}n x_n^2 (1+x_n)^{-1} \sum_{i=2}^{m} \lambda_i\}$$

$$\cdot \prod_{i=2}^{m} \lambda_i^{\frac{1}{2}n_1-p} \prod_{1<i<j\leq m} (\lambda_i - \lambda_j) d\lambda_2 \cdots d\lambda_m$$

$$= (2n^{-1} x_n^{-2}(1+x_n))^{\frac{1}{2}(n_1-1)(m-1)} \int_{\lambda_2 \geq \ldots \geq \lambda_m \geq 0} \cdots \int \exp(-\sum_{i=2}^{m} \lambda_i)$$

$$\cdot \prod_{i=2}^{m} \lambda_i^{\frac{1}{2}n_1-p} \prod_{1<i<j\leq m} (\lambda_i - \lambda_j) d\lambda_2 \cdots d\lambda_m$$

$$= (2n^{-1} x_n^{-2}(1+x_n))^{\frac{1}{2}(n_1-1)(m-1)} \Gamma_{m-1}(\tfrac{1}{2}(m-1)) \Gamma_{m-1}(\tfrac{1}{2}(n_1-1)) \pi^{-\frac{1}{2}(m-1)^2}.$$

By (3.2.21) and (3.2.22) one has

$$(3.2.23) \quad G_n(x_n) \sim (1+x_n)^{-\frac{1}{2}(n-1)+\frac{1}{2}(m-2)(n_1-1)}$$

$$\cdot \, x_n^{-\frac{1}{2}(m-2)(n_1-2)} (2/n)^{\frac{1}{2}(n_1-1)(m-1)+1}$$

$$\cdot \, \Gamma_{m-1}(\tfrac{1}{2}(m-1)) \Gamma_{m-1}(\tfrac{1}{2}(n_1-1)) \pi^{-\frac{1}{2}(m-1)^2} ,$$

and by (3.2.18)

$$(3.2.24) \quad 1 - F_N^{(2)}(x_n) = c_n \pi^{\frac{1}{2}m^2} \Gamma_m(\tfrac{1}{2}m)^{-1} G_n(x_n).$$

The result now follows by (3.2.23) and (3.2.24). □

REMARK 3.2.4. In Theorems 3.2.1 and 3.2.2 it is assumed that $nx_n^2 \to \infty$. Actually, the sequences $\{x_n\}$ such that $nx_n^2 \to x \in (0, \infty)$ separate two regions where different types of asymptotic relations hold. If $nx_n^2 \to 0$, but $nx_n \to \infty$, the tails $1 - F_N^{(i)}(x_n)$, $i = 1, 2$, behave as the tails of a $\chi^2_{mn_1}$ - distribution (the limiting distribution under H_0).

For example, in the case of Pillai's test statistic $t_N^{(1)}$, this is easily established by inspection of the simultaneous distribution of the latent roots of $S_{h,N}(S_{h,N}+S_{e,N})^{-1}$ (ANDERSON (1958), p.314, (47)). One has, as $nx_n \to \infty$ and $nx_n^2 \to 0$,

$$1 - F_N^{(1)}(x_n) = c_n \, \pi^{\frac{1}{2}m^2} \Gamma_m(\tfrac{1}{2}m)^{-1}$$

$$\cdot \int \cdots \int_{\{1 \geq \lambda_1 \geq \ldots \geq \lambda_m \geq 0, \, \sum_{i=1}^m \lambda_i \geq x_n\}} \prod_{i=1}^m \{\lambda_i^{\frac{1}{2}n_1-p}(1-\lambda_i)^{\frac{1}{2}n-p}\}$$

$$\cdot \prod_{i<j} (\lambda_i - \lambda_j) \, d\lambda_1 \ldots d\lambda_m$$

$$= c_n \, \pi^{\frac{1}{2}m^2} \Gamma_m(\tfrac{1}{2}m)^{-1} x_n^{\frac{1}{2}mn_1}$$

$$\cdot \int \cdots \int_{\{x_n^{-1} \geq \lambda_1 \geq \ldots \geq \lambda_m \geq 0, \, \sum_{i=1}^m \lambda_i \geq 1\}} \prod_{i=1}^m \{\lambda_i^{\frac{1}{2}n_1-p}(1-x_n\lambda_i)^{\frac{1}{2}n-p}\}$$

$$\cdot \prod_{i<j} (\lambda_i - \lambda_j) \, d\lambda_1 \ldots d\lambda_m$$

$$\sim c_n \, \pi^{\frac{1}{2}m^2} \Gamma_m(\tfrac{1}{2}m)^{-1} x_n^{\frac{1}{2}mn_1}$$

$$\cdot \int \cdots \cdots \int \prod_{i=1}^{m} \{\lambda_i^{\frac{1}{2}n_1-p} e^{-\frac{1}{2}nx_n\lambda_i}\}$$

$$\{\lambda_1 \geq \ldots \geq \lambda_m \geq 0, \sum_{i=1}^{m} \lambda_i \geq 1\}$$

$$\cdot \prod_{i<j} (\lambda_i - \lambda_j) d\lambda_1 \cdots d\lambda_m$$

$$\sim \Gamma(\tfrac{1}{2}mn_1)^{-1} e^{-\frac{1}{2}nx_n} (\tfrac{1}{2}nx_n)^{\frac{1}{2}mn_1-1}.$$

The last asymptotic equivalence is obtained by using (3.2.14) and the fact that

$$\prod_{i=1}^{m} \{\lambda_i^{\frac{1}{2}n_1-p} e^{-\frac{1}{2}\lambda_i}\} \prod_{i<j} (\lambda_i - \lambda_j)$$

is (apart from a scale factor) equal to the density of latent roots of a matrix A having a Wishart $W(n_1, I)$ distribution.

By similar methods one obtains

$$1 - F_N^{(2)}(x_n) \sim \Gamma(\tfrac{1}{2}mn_1)^{-1} e^{-\frac{1}{2}nx_n} (\tfrac{1}{2}nx_n)^{\frac{1}{2}mn_1-1},$$

as $nx_n \to \infty$ and $nx_n^2 \to 0$.

An asymptotic expression for large and moderate deviation probabilities in the case of Roy's largest root test statistic $t_N^{(3)}$ is given by the following theorem

THEOREM 3.2.3. *Let* $F_N^{(3)}$ *be the df of Roy's largest root statistic under* H_0. *Let* $\{x_n\}$ *be a bounded sequence of real numbers satisfying* $nx_n \to \infty$, *as* $n \to \infty$. *Then, as* $n \to \infty$,

$$(3.2.25) \quad 1 - F_N^{(3)}(x_n) \sim c_n \Gamma(\tfrac{1}{2}m)^{-1} \pi^{\frac{1}{2}m} \Gamma_{m-1}(\tfrac{1}{2}(n_1-1))(2/n)^{\frac{1}{2}(n_1-1)(m-1)+1}$$

$$\cdot x_n^{\frac{1}{2}(n_1+m-3)} (1+x_n)^{-\frac{1}{2}(n_1+n)+1},$$

where c_n *is given by* (3.2.2)

PROOF. Let $G_n : (0,\infty) \to \mathbb{R}$ be defined by

$$(3.2.26) \quad G_n(x) = \int \cdot \cdot \cdot \cdot \cdot \cdot \cdot \int_{\{\lambda_1 \geq x, \lambda_1 \geq \ldots \geq \lambda_m \geq 0\}} \prod_{i=1}^{m} \{\lambda_i^{\frac{1}{2}n_1 - p}(1+\lambda_i)^{-\frac{1}{2}(n_1+n)}\}$$

$$\cdot \prod_{i<j} (\lambda_i - \lambda_j) d\lambda_1 \ldots d\lambda_m =$$

$$= x^{\frac{1}{2}mn_1} \int \cdot \cdot \cdot \cdot \cdot \cdot \cdot \int_{\{\lambda_1 \geq 1, \lambda_1 \geq \ldots \geq \lambda_m \geq 0\}} \prod_{i=1}^{m} \{\lambda_i^{\frac{1}{2}n_1 - p}(1+x\lambda_i)^{-\frac{1}{2}(n_1+n)}\}$$

$$\cdot \prod_{i<j} (\lambda_i - \lambda_j) d\lambda_1 \ldots d\lambda_m.$$

By the change of variables

$$\begin{cases} u_1 = \lambda_1 - 1 \\ u_i = \lambda_i - \lambda_{i+1}, \quad 1 < i < m \\ u_m = \lambda_m \end{cases}$$

we have

$$G_n(x) = x^{\frac{1}{2}mn_1} \int \cdot \cdot \cdot \int_{u_1 + 1 \geq \sum_{i=2}^{m} u_i} \{(u_1+1) \prod_{i=2}^{m} (\sum_{j=i}^{m} u_j)\}^{\frac{1}{2}n_1 - p}$$

$$\cdot \{(1+x(u_1+1)) \prod_{i=2}^{m} (1+x \sum_{j=i}^{m} u_j)\}^{-\frac{1}{2}(n_1+n)}$$

$$\cdot \prod_{1 < i \leq j < m} (\sum_{k=i}^{j} u_k) \prod_{j=2}^{m} (u_1 + 1 - \sum_{h=j}^{m} u_h) du_1 \ldots du_m.$$

Define for non-negative $u_1, \ldots, u_m$

$$\phi_n(u_1, \ldots, u_m) = \log(1+x_n+x_n u_1) + \sum_{i=2}^{m} \log(1+x_n \sum_{j=i}^{m} u_j).$$

For $u_1, \ldots, u_m$ near zero we have

$$\phi_n(u_1, \ldots, u_m) = \log(1+x_n) + u_1 x_n (1+x_n)^{-1}(1+o(1))$$

$$+ x_n \sum_{i=2}^{m} (i-1) u_i (1+o(1)), \quad u_i \downarrow 0, \quad i = 1, \ldots, m.$$

By the same argument as in the proof of Theorem 3.2.2 we obtain

$$G_n(x_n) \sim x_n^{\frac{1}{2}mn_1}(1+x_n)^{-\frac{1}{2}(n_1+n)}\int_0^\infty \cdots \int_0^\infty \exp\{-\frac{1}{2}nx_n(1+x_n)^{-1}u_1$$

$$-\frac{1}{2}nx_n\sum_{i=2}^m (i-1)u_i\}$$

$$\cdot \prod_{i=2}^m (\sum_{j=i}^m u_j)^{\frac{1}{2}n_1-p} \prod_{1<i\le j<m}(\sum_{k=i}^j u_k)du_1\ldots du_m$$

$$\sim (2/n)^{\frac{1}{2}(n_1-1)(m-1)+1}x_n^{\frac{1}{2}(n_1+m-3)}(1+x_n)^{-\frac{1}{2}(n_1+n)+1}$$

$$\cdot \Gamma_{m-1}(\tfrac{1}{2}(m-1))\Gamma_{m-1}(\tfrac{1}{2}(n_1-1))/\pi^{\frac{1}{2}(m-1)^2}.$$

The result now follows from the relation

$$1 - F_N^{(3)}(x_n) = c_n\pi^{\frac{1}{2}m^2}\Gamma_m(\tfrac{1}{2}m)^{-1}G_n(x_n). \qquad \square$$

REMARK 3.2.5. Note that (3.2.25) holds for all bounded sequences $\{x_n\}$ such that $nx_n \to \infty$ (and not only for sequences $\{x_n\}$ satisfying $nx_n^2 \to \infty$). If $m = 1$, $nx_n \to \infty$ and $nx_n^2 \to 0$, then $1 - F_N^{(3)}(x_n)$ behaves by (3.2.25) as the tail of a $\chi_{n_1}^2$ - distribution.

We have seen that in the case of Pillai's test, Hotelling's T_0^2 and Roy's largest root test the asymptotic behavior of the tails of the null distribution is determined by certain critical points on the boundary of the domain of integration. The behavior of the LR test is somewhat different; in this case there is in fact a whole surface of critical points. Accordingly, the constant appearing in the leading term of an asymptotic expansion of large deviation probabilities is more complicated. To facilitate the statement of Theorem 3.2.4, this constant is given below. Let for $x \in (0,\infty)$

$$(3.2.27) \quad c(x) = \Gamma(\tfrac{1}{2}n_1)^{-m}\int \cdots \int_{\{\sum_{i=1}^{m-1}y_i\le x, y_i>0, i=1,\ldots,m\}} (1-e^{-x+\sum_{i=1}^{m-1}y_i})^{\frac{1}{2}n_1-1}$$

$$\cdot \prod_{i=1}^{m-1}\{e^{\frac{1}{2}iy_i}(1-e^{-y_i})^{\frac{1}{2}n_1-1}\}dy_1\ldots dy_{m-1}.$$

It will be shown in the proof of Theorem 3.2.4 that $c(x) \sim x^{\frac{1}{2}mn_1-1}/\Gamma(\tfrac{1}{2}mn_1)$,

as $x \downarrow 0$.

THEOREM 3.2.4. *Let* $F_N^{(4)}$ *be the df of the LR statistic* $t_N^{(4)}$ *under* H_0.

(i) *If* $\{x_n\}$ *is a sequence of real numbers such that* $\lim_{n \to \infty} x_n = x \in (0,\infty)$, *then*

(3.2.28) $\quad 1 - F_N^{(4)}(x_n) \sim c(x) (\tfrac{1}{2}n)^{\frac{1}{2}mn_1-1} e^{-\frac{1}{2}nx_n}, \quad$ *as* $n \to \infty$,

(ii) *If* $\{x_n\}$ *is a sequence of real numbers such that* $x_n \downarrow 0$ *and* $nx_n \to \infty$, *then*

(3.2.29) $\quad 1 - F_N^{(4)}(x_n) \sim \Gamma(\tfrac{1}{2}mn_1)^{-1} (\tfrac{1}{2}nx_n)^{\frac{1}{2}mn_1-1} e^{-\frac{1}{2}nx_n}, \quad$ *as* $n \to \infty$.

PROOF. The statistic $u_N = \exp(-t_N^{(4)})$ has under H_0 the distribution of a $U_{m,n_1,n}$ – variable (see ANDERSON (1958), Ch.8). Hence

(3.2.30) $\quad 1 - F_N^{(4)}(x_n) = \Gamma_m(\tfrac{1}{2}(n_1+n))\Gamma_m(\tfrac{1}{2}n)^{-1}\Gamma(\tfrac{1}{2}n_1)^{-m}$

$$\cdot \int \cdots \cdots \cdots \int_{\{\prod_{i=1}^m y_i \le \exp(-x_n), 0 < y_i < 1\}} \prod_{i=1}^m \{y_i^{\frac{1}{2}(n-1-i)}$$

$$\cdot (1-y_i)^{\frac{1}{2}n_1-1}\} dy_1 \cdots dy_m.$$

By the change of variables $y_i \to e^{-x_n u_i}$ one has

(3.2.31) $\quad 1 - F_N^{(4)}(x_n) = \Gamma_m(\tfrac{1}{2}(n_1+n))\Gamma_m(\tfrac{1}{2}n)^{-1}\Gamma(\tfrac{1}{2}n_1)^{-m}$

$$\cdot x_n^m \int \cdots \cdots \int_{\{\sum_{i=1}^m u_i \ge 1, u_i > 0\}} \prod_{i=1}^m \{e^{-\frac{1}{2}(n+1-i)x_n u_i}$$

$$\cdot (1-e^{-x_n u_i})^{\frac{1}{2}n_1-1}\} du_1 \cdots du_m.$$

Next, by the change of variables

$$\begin{cases} y_1 = \sum_{i=1}^m u_i \\ y_i = u_i, \ i > 1 \end{cases}$$

and integration by parts we obtain

$$(3.2.32) \quad \int \cdots \int_{\{\sum_{i=1}^{m} u_i \geq 1, u_i > 0\}} \prod_{i=1}^{m} \{e^{-\frac{1}{2}(n+1-i)x_n u_i}(1-e^{-x_n u_i})^{\frac{1}{2}n_1-1}\} du_1 \cdots du_m$$

$$= \int_1^{\infty} dy_1 \int \cdots \int_{\{\sum_{i=2}^{m} y_i \leq y_1, y_i > 0\}} e^{-\frac{1}{2}nx_n y_1}(1-e^{-x_n(y_1-\sum_{i=2}^{m} y_i)})^{\frac{1}{2}n_1-1}$$

$$\cdot \prod_{i=2}^{m} \{e^{\frac{1}{2}(i-1)x_n y_i}(1-e^{-x_n y_i})^{\frac{1}{2}n_1-1}\} dy_2 \cdots dy_m$$

$$= 2(nx_n)^{-1}e^{-\frac{1}{2}nx_n} \int \cdots \int_{\{\sum_{i=2}^{m} y_i \leq 1, y_i > 0\}} \{1-e^{-x_n(1-\sum_{i=2}^{m} y_i)}\}^{\frac{1}{2}n_1-1}$$

$$\cdot \prod_{i=2}^{m} \{e^{\frac{1}{2}(i-1)x_n y_i}(1-e^{-x_n y_i})^{\frac{1}{2}n_1-1}\} dy_2 \cdots dy_m$$

$$+ 2(nx_n)^{-1} \int_1^{\infty} e^{-\frac{1}{2}nx_n y_1} g_n'(y_1) dy_1,$$

where

$$g_n(y_1) = \int \cdots \int_{\{\sum_{i=2}^{m} y_i \leq y_1, y_i > 0\}} \{1-e^{-x_n(y_1-\sum_{i=2}^{m} y_i)}\}^{\frac{1}{2}n_1-1}$$

$$\cdot \prod_{i=2}^{m} \{e^{\frac{1}{2}(i-1)x_n y_i}(1-e^{-x_n y_i})^{\frac{1}{2}n_1-1}\} dy_2 \cdots dy_m.$$

Integration by parts of the second term at the right-hand side of (3.2.32) shows that this second term is of lower order than the first term as $n \to \infty$. Since

$$\Gamma_m(\tfrac{1}{2}(n+n_1))/\Gamma_m(\tfrac{1}{2}n) \sim (\tfrac{1}{2}n)^{\frac{1}{2}n_1 m},$$

result (i) follows from (3.2.31) and (3.2.32).

If $x_n \downarrow 0$, one has by Lebesgue's dominated convergence theorem

$$(3.2.33) \qquad \int \cdots \cdots \int_{\{\sum_{i=1}^{m-1} y_i \le 1, y_i > 0\}} \prod_{i=1}^{m-1} \{e^{\frac12 i x_n y_i} \{x_n^{-1}(1-e^{-x_n y_i})\}^{\frac12 n_1 - 1}\}$$

$$\cdot \{x_n^{-1}(1-e^{-x_n(1-\sum_{i=1}^{m-1} y_i)})\}^{\frac12 n_1 - 1} dy_1 \cdots dy_{m-1}$$

$$\to \int \cdots \cdots \int_{\{\sum_{i=1}^{m-1} y_i \le 1, y_i > 0\}} \prod_{i=1}^{m-1} y_i^{\frac12 n_1 - 1} (1-\sum_{i=1}^{m-1} y_i)^{\frac12 n_1 - 1} dy_1 \cdots dy_{m-1}.$$

It is easily shown (using induction) that this last expression equals

$$\Gamma(\tfrac12 n_1)^m / \Gamma(\tfrac12 n_1 m).$$

Thus (ii) follows from (3.2.31) to (3.2.33). $\square$

REMARK 3.2.6. If $x_n \downarrow 0$ and $nx_n \to \infty$, part (ii) of Theorem 3.2.4 yields that $1 - F_N^{(4)}(x_n)$ behaves as the tail of a $\chi^2_{mn_1}$-distribution (compare with Remark 3.2.4).

3.3. BAHADUR EFFICIENCY AND DEFICIENCY AT FIXED ALTERNATIVES

Consider the four statistics $t_N^{(1)}, \ldots, t_N^{(4)}$ defined by (3.1.1) to (3.1.4). Without loss of generality we may assume that $S_{e,N}$ has the central Wishart distribution $W_m(n, I)$ and that $S_{h,N}$ has the non-central Wishart distribution $W_m(n_1, I, N\Theta)$, where Θ is a diagonal matrix $\mathrm{diag}(\theta_1, \ldots, \theta_m)$ with $\theta_1 \ge \ldots \ge \theta_m \ge 0$. It is clear that the distributions of the statistics only depend on the alternative through the matrix Θ. Probabilities under H_0 will be denoted by P_0 and probabilities under the alternative Θ by P_Θ. We shall (continue to) assume that n_1 is a fixed number satisfying $n_1 \ge m$ and that the sample size $N = n + O(1)$, as $N \to \infty$.

Let for $\beta \in (0,1)$, $N \in \mathbb{N}$ and $i = 1, \ldots, 4$ the constant $c_N^{(i)} = c_N^{(i)}(\beta, \Theta)$ be chosen such that

$$(3.3.1) \qquad P_\Theta(t_N^{(i)} \ge c_N^{(i)}) = \beta$$

and let $\alpha_N^{(i)}(\beta, \Theta)$ be defined by

$$(3.3.2) \qquad \alpha_N^{(i)}(\beta,\Theta) = P_0(t_N^{(i)} \geq c_N^{(i)}).$$

It will be shown that for $i = 1,\ldots,4$ the limit

$$(3.3.3) \qquad c_i(\Theta) = -\lim_{N \to \infty} N^{-1} \log \alpha_N^{(i)}(\beta,\Theta)$$

exists and that this limit does not depend on β. The number $2c_i(\Theta)$ is called the *exact (Bahadur) slope* of the test sequence $\{t_N^{(i)}\}$ at the alternative Θ. The *Bahadur efficiency* of the test sequence $\{t_N^{(i)}\}$ w.r.t. the test sequence $\{t_N^{(j)}\}$ is given by the ratio $c_i(\Theta)/c_j(\Theta)$ (see also chapter II).

An interpretation of the Bahadur efficiency in terms of sample sizes required to attain a certain criterion runs as follows. Define for $0 < \alpha < \beta < 1$ and $\Theta \neq 0$ the numbers $N^{(i)}(\alpha,\beta,\Theta)$, $i = 1,\ldots,4$, by

$$(3.3.4) \qquad N^{(i)}(\alpha,\beta,\Theta) = \min \{N \in \mathbb{N} : \alpha_N^{(i)}(\beta,\Theta) \leq \alpha\},$$

i.e. $N^{(i)}(\alpha,\beta,\Theta)$ is the smallest sample size of a level α test based on $\{t_N^{(i)}\}$ which achieves power β at Θ. Then the Bahadur efficiency of $\{t_N^{(j)}\}$ w.r.t. $\{t_N^{(i)}\}$ is given by

$$\lim_{\alpha \downarrow 0} N^{(i)}(\alpha,\beta,\Theta)/N^{(j)}(\alpha,\beta,\Theta).$$

(see e.g. BAHADUR (1967)).

The exact Bahadur slope of the sequences of test statistics $\{t_N^{(i)}\}$, $i = 1,\ldots,4$ at Θ are given by the following theorem

THEOREM 3.3.1. *Let the diagonal matrix Θ represent a fixed alternative to H_0 and let $e_1(\Theta),\ldots,e_4(\Theta)$ be the exact slopes at Θ of Pillai's test, Hotelling's T_0^2, Roy's largest root test and the LR test respectively. Then*

$$(3.3.5) \qquad e_1(\Theta) = -m \log\{m^{-1} \mathrm{tr}(I+\Theta)^{-1}\},$$

$$(3.3.6) \qquad e_2(\Theta) = \log(1+\mathrm{tr}\,\Theta),$$

$$(3.3.7) \qquad e_3(\Theta) = \log(1+\theta_1),$$

and

(3.3.8) $e_4(\Theta) = \log|I+\Theta|$.

PROOF. By Proposition 3.5.1 and Remark 3.5.1 of the appendix, we have the following expansions of the critical values $c_N^{(i)}$

(3.3.9) $c_N^{(1)} = \text{tr}\{\Theta(I+\Theta)^{-1}\} + O(N^{-\frac{1}{2}})$, $N \to \infty$

(3.3.10) $c_N^{(2)} = \text{tr }\Theta + O(N^{-\frac{1}{2}})$, $N \to \infty$

(3.3.11) $c_N^{(3)} = \theta_1 + O(N^{-\frac{1}{2}})$, $N \to \infty$

(3.3.12) $c_N^{(4)} = \log|I+\Theta| + O(N^{-\frac{1}{2}})$, $N \to \infty$.

The theorem now follows easily from Theorems 3.2.1 to 3.2.4 and (3.3.9) to (3.3.12).   □

The exact slope of the LR test can also be computed by an application of theorems in BAHADUR & RAGHAVACHARI (1970). These theorems can be used to show that the LR test attains the optimal slope at each alternative (see HSIEH (1979)).

We have, by Jensen's inequality, $c_1(\Theta) \le c_4(\Theta)$ with equality if and only if $\theta_1 = \ldots = \theta_m$. Hence the Bahadur efficiency of the LR test w.r.t. Pillai's test is strictly greater than one unless the roots $\theta_1, \ldots, \theta_m$ are equal.

Furthermore we have $|I+\Theta| \ge 1+\text{tr }\Theta \ge 1+\theta_1$, where both inequalities are strict unless $\theta_i = 0$, $i \ge 2$. It follows that the Bahadur efficiency of the LR test w.r.t. Hotelling's T_0^2 and Roy's test is strictly greater than one unless only one latent root of Θ is different from zero.

This has the remarkable consequence that Pillai's test, Hotelling's T_0^2 and Roy's test are inadmissible in the sense of Bahadur efficiency. However, this is a kind of "first order inadmissibility" and it will be shown that in cases where the Bahadur efficiency of the LR test w.r.t. one of the three other tests is equal to one, the other test is strictly better when we take second order terms into consideration.

The relationships between the exact slopes partly reflect the numerical results of PILLAI & JAYACHANDRAN (1967) and FUJIKOSHI (1970) and the simulation results of ROY, GNANADESIKAN & SRIVASTAVA (1971); when tr Θ is kept constant, Pillai's test has the greatest exact slope if the latent roots

$\theta_1, \ldots, \theta_m$ are equal and Hotelling's T_0^2 and Roy's test have the greatest exact slope if only one root is not equal to zero. But the results on Bahadur efficiency seem to be in favor of the LR test, in contrast to the numerical result mentioned above, where the LR test only occupied an inter-mediate position. We shall give an explanation of this fact in section 3.4.

To make a more precise comparison in cases where the Bahadur efficiency is equal to one, we introduce the concept of *Bahadur deficiency*. Let for $i = 1, \ldots, 4$, $0 < \alpha < \beta < 1$ and $\Theta \neq 0$, the number $N^{(i)}(\alpha, \beta, \Theta)$ be defined by (3.3.4). Suppose that there exists a positive non-decreasing function $f = f_\Theta : \mathbb{N} \to \mathbb{R}$ such that $\lim_{n \to \infty} f(n) = \infty$ and

$$(3.3.13) \qquad \lim_{\alpha \downarrow 0} \{N^{(j)}(\alpha, \beta, \Theta) - N^{(i)}(\alpha, \beta, \Theta)\}/f(N^{(i)}(\alpha, \beta, \Theta)) = 1,$$

for each $\beta \in (0,1)$. Then we say (in analogy to the definition of Hodges-Lehmann deficiency in the case of contiguous alternatives) that the *Bahadur deficiency* of $\{t_N^{(j)}\}$ w.r.t. $\{t_N^{(i)}\}$ is asymptotically equivalent to $f(N)$. In other words, the additional number of observations required to attain the same power β and the same level α for test $\{t_N^{(j)}\}$ as for test $\{t_N^{(i)}\}$ at Θ is asymptotically equivalent to $f(N^{(i)}(\alpha, \beta, \Theta))$ as $\alpha \downarrow 0$ (a slightly weaker definition of Bahadur deficiency is given in KALLENBERG (1978)).

We briefly defer the definition of Bahadur deficiency in cases where

$$0 \leq \lim_{\alpha \downarrow 0} \sup \{N^{(j)}(\alpha, \beta, \Theta) - N^{(i)}(\alpha, \beta, \Theta)\} < \infty.$$

The Bahadur deficiency of the LR test w.r.t. the other tests in cases where the Bahadur efficiency equals one, is given by the following theorem.

THEOREM 3.3.2.

(i) *Let the latent roots* $\theta_1, \ldots, \theta_m$ *of* Θ *be such that* $\theta_1 = \ldots = \theta_m = \theta > 0$. *Then the Bahadur deficiency of the LR test w.r.t. Pillai's test is asymptotically equivalent to* $(mp-1) \log N/\{m \log(1+\theta)\}$.

(ii) *Let the latent roots* $\theta_1, \ldots, \theta_m$ *of* Θ *be such that* $\theta_1 > 0$ *and* $\theta_2 = \ldots = \theta_m = 0$. *Then the Bahadur deficiency of the LR test w.r.t. Hotelling's* T_0^2 *and Roy's largest root test is asymptotically equivalent to* $(n_1-1)(m-1) \log N/\log(1+\theta_1)$.

REMARK 3.3.1. For computational purposes it is convenient to note that the Bahadur deficiency of $\{t_N^{(j)}\}$ w.r.t. $\{t_N^{(i)}\}$ is asymptotically equivalent to

$$2e(\theta)^{-1}\{\log \alpha_N^{(j)}(\beta,\theta) - \log \alpha_N^{(i)}(\beta,\theta)\}$$

if $e(\theta)$ is the exact slope both of $\{t_N^{(i)}\}$ and $\{t_N^{(j)}\}$ at θ.

PROOF OF THEOREM 3.3.2.

(i) It follows from Proposition 3.5.1 of the appendix that $c_N^{(1)}$ defined by (3.3.1) has the expansion

$$(3.3.14) \qquad c_N^{(1)} = m\theta(1+\theta)^{-1}+N^{-\frac{1}{2}}\tau_1 u_\beta + \mathcal{O}(N^{-1}), \qquad N \to \infty,$$

where $\tau_1 = (1+\theta)^{-2}\{2m\theta(2+\theta)\}^{\frac{1}{2}}$ and $u_\beta = \Phi^{-1}(\beta)$ with Φ the standard normal df. The critical value $c_N^{(4)}$ (also defined by (3.3.1)) has by Proposition 3.5.1 the expansion

$$(3.3.15) \qquad c_N^{(4)} = m \log(1+\theta)+N^{-\frac{1}{2}}\tau_4 u_\beta + \mathcal{O}(N^{-1}), \qquad N \to \infty,$$

where $\tau_4 = (1+\theta)^{-1}\{2m\theta(2+\theta)\}^{\frac{1}{2}}$.

Hence, by (3.3.14) and Theorem 3.2.1

$$(3.3.16) \qquad \log \alpha_N^{(1)}(\beta,\theta) = -\tfrac{1}{2}Nm \log(1+\theta)-\tfrac{1}{2}(1+\theta)^{-1}\{2Nm\theta(2+\theta)\}^{\frac{1}{2}}u_\beta +$$

$$+\tfrac{1}{2}(mn_1-mp-1) \log N+\mathcal{O}(1), \qquad N \to \infty.$$

Similarly, by (3.3.15) and Theorem 3.2.4

$$(3.3.17) \qquad \log \alpha_N^{(4)}(\beta,\theta) = -\tfrac{1}{2}Nm \log(1+\theta)-\tfrac{1}{2}(1+\theta)^{-1}\{2Nm\theta(2+\theta)\}^{\frac{1}{2}}u_\beta +$$

$$+\tfrac{1}{2}(mn_1-2) \log N+\mathcal{O}(1), \qquad N \to \infty.$$

The result now follows from Remark 3.3.1.

(ii) By Proposition 3.5.1 we have

$$(3.3.18) \qquad c_N^{(2)} = \theta_1 + N^{-\frac{1}{2}}\tau u_\beta + \mathcal{O}(N^{-1}), \qquad N \to \infty,$$

where $\tau = \{2\theta_1(2+\theta_1)\}^{\frac{1}{2}}$. Proposition 3.5.2 yields

$$(3.3.19) \qquad c_N^{(3)} = \theta_1 + N^{-\frac{1}{2}}\tau u_\beta + \mathcal{O}(N^{-1}), \qquad N \to \infty.$$

Thus, by (3.3.18), (3.3.19) and Theorems 3.2.2 and 3.2.3

$$(3.3.20) \quad \log \alpha_N^{(2)}(\beta,\theta) = -\tfrac{1}{2}N \log (1+\theta_1) - \tfrac{1}{2}N^{\frac{1}{2}}\tau u_\beta +$$

$$+\tfrac{1}{2}\{mn_1-(n_1-1)(m-1)-2\} \log N + O(1) =$$

$$= \log \alpha_N^{(3)}(\beta,\theta) + O(1), \quad N \to \infty.$$

Since by Proposition 3.5.1

$$(3.3.21) \quad c_N^{(4)} = \log (1+\theta_1) + N^{-\frac{1}{2}}\tau u_\beta + O(N^{-1}), \quad N \to \infty,$$

Theorem 3.2.4 implies

$$(3.3.22) \quad \log \alpha_N^{(4)}(\beta,\theta) = -\tfrac{1}{2}N \log (1+\theta_1) - \tfrac{1}{2}N^{\frac{1}{2}}\tau u_\beta + \tfrac{1}{2}(mn_1-2)\log N + O(1),$$

$$N \to \infty.$$

The result now follows from Remark 3.3.1. $\square$

Thus, Pillai's test, Hotelling's T_0^2 and Roy's largest root test are strictly better than the LR test in the sense of Bahadur *deficiency*, in cases where the Bahadur efficiency w.r.t. the LR test is equal to one. In KALLENBERG (1978) it is shown that in typical testing problems in multivariate exponential families the Bahadur deficiency of the LR test w.r.t. the optimal test is of order $O(\log N)$, in accordance with the results of Theorem 3.3.2.

It is clear from the proof of Theorem 3.3.2 (cf. (3.3.20)) that the difference between $N^{(2)}(\alpha,\beta,\theta)$ and $N^{(3)}(\alpha,\beta,\theta)$ remains bounded as $\alpha \downarrow 0$ if $\theta_1 > 0$ and $\theta_2 = \ldots = \theta_m = 0$. In order to give a measure of efficiency of the two tests w.r.t. one another in this case, we introduce the concept of *finite* Bahadur deficiency. To this end we need an artifice which is similar to a technique by which (finite) Hodges-Lehmann deficiencies are defined.

Let $\{t_N\}_{N \in \mathbb{N}}$ be a test sequence for H_0 and let c_N and $\alpha_N(\beta,\theta)$ defined by (3.3.1) and (3.3.2) with $t_N^{(i)}$, $c_N^{(i)}$ and $\alpha_N^{(i)}(\beta,\theta)$ replaced by t_N, c_N and $\alpha_N(\beta,\theta)$, respectively. We shall use the abbreviation N_α to denote $N(\alpha,\beta,\theta)$. Define

(3.3.23) $\bar{N}_\alpha = p(\alpha)N_\alpha + (1-p(\alpha))(N_\alpha-1)$,

where $p(\alpha) = \log(\alpha_{N_\alpha-1}(\beta,\Theta)/\alpha)/\log(\alpha_{N_\alpha-1}(\beta,\Theta)/\alpha_{N_\alpha}(\beta,\Theta))$.

Let (t_N,c_N) be a test based on t_N which rejects H_0 for values of t_N larger than or equal to c_N and let $t_{\bar{N}_\alpha}$ be a test based on $(t_{N_\alpha},c_{N_\alpha})$ and $(t_{N_\alpha-1},c_{N_\alpha-1})$ which chooses $(t_{N_\alpha},c_{N_\alpha})$ with probability $p(\alpha)$ and $(t_{N_\alpha-1},c_{N_\alpha-1})$ with probability $1-p(\alpha)$. The test $t_{\bar{N}_\alpha}$ has power β at Θ and the logarithm of its size has expectation $\log \alpha$. The extension of $\{t_N\}_{N\in\mathbb{N}}$ to $\{t_{N_\alpha}\}_{\alpha\in(0,1)}$ serves the purpose of making the sample size, required to obtain power β for the alternative Θ at the significance level α, a continuous function of α.

Let $\bar{N}_\alpha^{(2)}$ and $\bar{N}_\alpha^{(3)}$ be defined by (3.3.23) with $\{t_N\}$ replaced by $\{t_N^{(2)}\}$ and $\{t_N^{(3)}\}$, respectively. Then, if only one latent root of Θ is different from zero, $\lim_{\alpha\downarrow0}\{\bar{N}_\alpha^{(2)}-\bar{N}_\alpha^{(3)}\}$ exists and is independent of β. We shall call this limit the *Bahadur deficiency* of Hotelling's T_0^2 w.r.t. Roy's largest root test.

REMARK 3.3.2. Although this definition of Bahadur deficiency is mathematically the most convenient one, it is not appealing from a statistical point of view. In statistical practice one would fix the size of the test and would "randomize" the sample size in such a way that the (expected) power of the test has some desired value β. In fact, this method is followed in the definition of Hodges-Lehmann deficiency. However, in the present context, where the alternative Θ and the power β of the test are fixed, there exists a simple relationship between the sample size N and the "performance criterion" $\log \alpha_N^{(i)}(\beta,\Theta)$, since $\log \alpha_N^{(i)}(\beta,\Theta) \sim - N c_i(\Theta)$, $N \to \infty$, $i = 1,\dots,4$. This has motivated our definition of (finite) Bahadur deficiency.

THEOREM 3.3.3. *Suppose that Θ satisfies $\theta_1 > 0$ and $\theta_2 = \dots = \theta_m = 0$. The Bahadur deficiency of Hotelling's T_0^2 w.r.t. Roy's largest root test at Θ is given by*

(3.3.24) $(m-1)(n_1-1)\{\log(\theta_1^{-1}(1+\theta_1))-(1+\theta_1)^{-1}\}/\log(1+\theta_1)$.

PROOF. By Proposition 3.5.1

$$(3.3.25) \quad c_N^{(2)} = \theta_1 + N^{-\frac{1}{2}}\tau u_\beta + N^{-1}\{mn_1 + (m+1+N-n)\theta_1 +$$

$$+ \tau^{-2}(u_\beta^2 - 1)(4\theta_1 + 8\theta_1^2 + \tfrac{8}{3}\theta_1^3)\} + O(N^{-3/2}), \qquad N \to \infty$$

where $\tau = \{2\theta_1(2+\theta_1)\}^{\frac{1}{2}}$, and by Proposition 3.5.2

$$(3.3.26) \quad c_N^{(3)} = \theta_1 + N^{-\frac{1}{2}}\tau u_\beta + N^{-1}\{n_1 + m - 1 + (m+1+N-n)\theta_1$$

$$+ \tau^{-2}(u_\beta^2 - 1)(4\theta_1 + 8\theta_1^2 + \tfrac{8}{3}\theta_1^3)\} + O(N^{-3/2}).$$

Hence by Theorems 3.2.2 and 3.2.3

$$(3.3.27) \quad \log \alpha_N^{(2)}(\beta,\theta) - \log \alpha_N^{(3)}(\beta,\theta) =$$

$$= \tfrac{1}{2}(m-1)(n_1-1)\{\log(\theta_1^{-1}(1+\theta_1)) - (1+\theta_1)^{-1}\} + o(1), \qquad N \to \infty,$$

implying (3.3.24). □

The theorem implies that Roy's test is superior for this type of alternative in the sense of Bahadur deficiency. This confirms the simulation results of ROY, GNANADESIKAN & SRIVASTAVA (1971). It is also seen that the deficiency tends to zero as $\theta_1 \to \infty$.

3.4. DEFICIENCIES AT MOVING ALTERNATIVES

It was seen in section 3.3 that the LR test is superior to the other tests in the sense of Bahadur *efficiency* and that the LR test is only inferior to some of the other tests in the sense of Bahadur *deficiency* on certain halflines in the parameter space. We shall show that the situation is somewhat different for moving alternatives and that the regions of the parameter space on which the LR test is deficient are enlarged if the alternatives tend to the null hypothesis. The rate at which these regions are enlarged is inversely proportional to the rate at which the alternatives move to the null hypothesis.

First we shall introduce a deficiency concept in the case of moving alternatives Θ_N such that $N \operatorname{tr} \Theta_N \to \infty$ and $\{\Theta_N\}$ is a bounded sequence. Let $\alpha_N^{(i)}(\beta,\Theta_q)$ be the level of a test based on $t_N^{(i)}$ with power $\beta \in (0,1)$ at the alternative Θ_q. Let $N^{(i)}(\alpha,\beta,\Theta_q)$ be defined by (3.3.4) with Θ replaced

by Θ_q and let $\{\Theta_N\}$ be a bounded sequence of alternatives such that $N \, \text{tr} \, \Theta_N \to \infty$. If there exists a positive non-decreasing function $f : \mathbb{N} \to \mathbb{R}$ such that $\lim_{N \to \infty} f(N) = \infty$ and

$$(3.4.1) \qquad N^{(j)}(\alpha_n^{(i)}(\beta,\Theta_n),\beta,\Theta_n) - n \sim f(n), \qquad n \to \infty,$$

with f independent of $\beta \in (0,1)$, we shall say that the (non-finite) *deficiency* of $\{t_N^{(j)}\}$ w.r.t. $\{t_N^{(i)}\}$ is asymptotically equivalent to $f(N)$. In other words, the additional number of observations required to attain for a test based on the sequence $\{t_N^{(j)}\}$ the same power β and the same level α as for a test based on the statistic $t_N^{(i)}$ is asymptotically equivalent to $f(N)$.

The following theorem is an extension of Theorem 3.3.2.

THEOREM 3.4.1. *Let* $\{\Theta_N\} = \{\text{diag}(\theta_{1,N},\ldots,\theta_{m,N})\}$ *be a bounded sequence of alternatives such that* $N \, \text{tr} \, \Theta_N^2 \to \infty$, *as* $N \to \infty$.

(i) *If* $\theta_{1,N} - \theta_{j,N} = O(N^{-1})$, $j = 2,\ldots,m$, *then the deficiency of the LR test w.r.t. Pillai's test is asymptotically equivalent to*
$(mp-1)(\log(N\theta_{1,N}^2))/\{m \log(1+\theta_{1,N})\}$.

(ii) *If* $\theta_{j,N} = O(N^{-1})$, $j = 2,\ldots,m$, *then the deficiency of the LR test w.r.t. Hotelling's* T_0^2 *is asymptotically equivalent to*
$(m-1)(n_1-1)(\log(N\theta_{1,N}^2))/\log(1+\theta_{1,N})$.

PROOF.

(i) Since $\theta_{j,N} = \theta_{1,N} + O(N^{-1})$, $j > 1$, we have

$$\text{tr}\{\Theta_N(I+\Theta_N)^{-1}\} = m\theta_{1,N}(1+\theta_{1,N})^{-1} + O(N^{-1}), \qquad N \to \infty,$$

and

$$\tau_{1,N}^2 = 2m\theta_{1,N}(2+\theta_{1,n})(1+\theta_{1,N})^{-4} + O(N^{-1}),$$

where $\tau_{1,N}$ is defined by (3.5.4) in the appendix. Hence $c_N^{(1)}$ may be expanded in the following way:

$$(3.4.2) \qquad c_N^{(1)} = m\theta_{1,N}(1+\theta_{1,N})^{-1} + u_\beta N^{-\frac{1}{2}}\{2m\theta_{1,N}(2+\theta_{1,N})\}^{\frac{1}{2}}(1+\theta_{1,N})^{-2}$$

$$+ O(N^{-1}), \qquad N \to \infty.$$

Thus, by Theorem 3.2.1,

$$(3.4.3) \quad \log \alpha_N^{(1)}(\beta, \Theta_N) = -\tfrac{1}{2}mN \log(1+\theta_{1,N}) - \tfrac{1}{2}N^{\frac{1}{2}}u_\beta(1+\theta_{1,N})^{-1}\{2m\theta_{1,N}(2+\theta_{1,N})\}^{\frac{1}{2}}$$

$$+\tfrac{1}{2}(mn_1-2)\log(N\theta_{1,N}) - \tfrac{1}{2}(mp-1)\log(N\theta_{1,N}^2) + \mathcal{O}(1),$$

$$N \to \infty.$$

Furthermore, by (3.5.8) in the appendix,

$$(3.4.4) \quad c_N^{(4)} = \log|I+\Theta_N| + N^{-\frac{1}{2}}\tau_{4,N}u_\beta + \mathcal{O}(N^{-1}), \qquad N \to \infty$$

and hence by Theorem 3.2.4

$$(3.4.5) \quad \log \alpha_N^{(4)}(\beta, \Theta_N) = -\tfrac{1}{2}Nc_N^{(4)} + \tfrac{1}{2}(mn_1-2)\log(Nc_N^{(4)}) + \mathcal{O}(1)$$

$$= -\tfrac{1}{2}mN \log(1+\theta_{1,N}) - \tfrac{1}{2}N^{\frac{1}{2}}u_\beta(1+\theta_{1,N})^{-1}\{2m\theta_{1,N}(2+\theta_{1,N})\}^{\frac{1}{2}}$$

$$+\tfrac{1}{2}(mn_1-2)\log(N\theta_{1,N}) + \mathcal{O}(1), \qquad N \to \infty.$$

Comparison of (3.4.3) and (3.4.5) yields (i).

(ii) Since the method of proof is similar to that used in (i), we shall only give the essential steps and omit details. By Theorems 3.2.2 and 3.2.4 and Proposition 3.5.1 one has

$$(3.4.6) \quad \log \alpha_N^{(2)}(\beta, \Theta_N) = -\tfrac{1}{2}N \log(1+\theta_{1,N}) - \tfrac{1}{2}u_\beta(1+\theta_{1,N})^{-1}\{2N\theta_{1,N}(2+\theta_{1,N})\}^{\frac{1}{2}}$$

$$+\tfrac{1}{2}\{mn_1 - (m-1)(n_1-1)-2\}\log N$$

$$-\tfrac{1}{2}(m-2)(n_1-2)\log\theta_{1,N} + \mathcal{O}(1), \qquad N \to \infty$$

and

$$(3.4.7) \qquad \log \alpha_N^{(4)}(\beta,\Theta_N) = -\tfrac{1}{2}N \log(1+\theta_{1,N}) - \tfrac{1}{2}u_\beta(1+\theta_{1,N})^{-1}\{2N\theta_{1,N}(2+\theta_{1,N})\}^{\tfrac{1}{2}}$$

$$+\tfrac{1}{2}(mn_1-2)\log(N\theta_{1,N}) + O(1), \qquad N \to \infty.$$

The result follows from (3.4.6) and (3.4.7). □

Theorems 3.2.3 and 3.2.4 yield asymptotic expressions for the tail probabilities $\alpha_N^{(3)}(\beta,\Theta_N)$ and $\alpha_N^{(4)}(\beta,\Theta_N)$ which are uniformly valid for all bounded sequences $\{\Theta_N\}$ such that $N \operatorname{tr} \Theta_N \to \infty$. This leads to the following result

THEOREM 3.4.2. *Let* $\{\Theta_N\} = \{\operatorname{diag}(\theta_{1,N},\ldots,\theta_{m,N})\}$ *be a bounded sequence of alternatives such that* $N\theta_{1,N} \to \infty$ *and* $\theta_{j,N} = O(N^{-1})$, $j \geq 2$, *as* $N \to \infty$. *Then the deficiency of the LR test w.r.t. Roy's largest root test is asymptotically equivalent to* $(m-1)(n_1-1)(\log(N\theta_{1,N}))/\log(1+\theta_{1,N})$.

PROOF. By Theorem 3.2.3 and Proposition 3.5.2

$$(3.4.8) \qquad \log \alpha_N^{(3)}(\beta,\Theta_N) = -\tfrac{1}{2}N \log(1+\theta_{1,N}) - \tfrac{1}{2}u_\beta(1+\theta_{1,N})^{-1}\{2N\theta_{1,N}(2+\theta_{1,N})\}^{\tfrac{1}{2}}$$

$$-\tfrac{1}{2}(mn_1-2-(m-1)(n_1-1)\} \log N$$

$$+\tfrac{1}{2}(n_1+m-3)\log \theta_{1,N} + O(1), \qquad N \to \infty.$$

The expansion of $\log \alpha_N^{(4)}(\beta,\Theta_N)$ is given by (3.4.5) in the proof of Theorem 3.4.1 (by Theorem 3.2.4 and Proposition 3.5.1 this expansion is uniformly valid for all sequences $\{\Theta_N\}$ such that $\Theta_N \to 0$ and $N \operatorname{tr} \Theta_N \to \infty$). The theorem now follows from (3.4.7) and (3.4.8). □

COROLLARY 3.4.1. *Under the conditions of Theorem* 3.4.1 (ii) *the deficiency of Hotelling's* T_0^2 *w.r.t. Roy's largest root test is asymptotically equivalent to* $-(m-1)(n_1-1)(\log \theta_{1,N})/\log(1+\theta_{1,N})$.

These results imply that the regions of the parameter space on which the LR test is deficient are in a certain sense "enlarged" if $\Theta_N \to 0$. For in Theorems 3.4.1 and 3.4.2 the ratios $\theta_{j,N}/\theta_{1,N}$ either have to satisfy the condition

$$(3.4.9) \qquad |\theta_{j,N}/\theta_{1,N}-1| = O((N\theta_{1,N})^{-1}), \qquad j = 2,\ldots,m$$

or the condition

$$(3.4.10) \qquad \theta_{j,N}/\theta_{1,N} = O((N\theta_{1,N})^{-1}), \qquad j = 2,\ldots,m$$

and it is seen that the rate at which the upper bounds at the right hand sides of (3.4.9) and (3.4.10) tend to zero is inversely proportional to the rate at which $\theta_{1,N}$ tends to zero. It is also interesting that the deficiency becomes of higher order if $\Theta_N \to 0$. For example, the deficiency of Hotelling's T_0^2 w.r.t. Roy's test is no longer finite under condition (ii) of Theorem 3.4.1 if $\Theta_N \to 0$.

If $\mathrm{tr}\ \Theta_N \to 0$ at a faster rate than specified by the conditions of Theorem 3.4.1, the differences between $\alpha_N^{(1)}(\beta,\Theta_N)$, $\alpha_N^{(2)}(\beta,\Theta_N)$ and $\alpha_N^{(4)}(\beta,\Theta_N)$ become smaller. Roy's test seems to behave somewhat differently. We do not attempt to give a complete treatment of all possible cases, but comment briefly on some cases of particular interest.

If $N\ \mathrm{tr}\ \Theta_N \to \infty$ and $N\ \mathrm{tr}\ \Theta_N^2 \to 0$, we have

$$(3.4.11) \qquad \lim_{N\to\infty} \alpha_N^{(i)}(\beta,\Theta_N)/\alpha_N^{(j)}(\beta,\Theta_N) = 1, \qquad i,j = 1,2,4.$$

This follows from Proposition 3.5.1, Theorem 3.2.4 and Remark 3.2.4. However, if $\theta_{j,N} = O(N^{-1})$, $j \geq 2$, $N\theta_{1,N} \to \infty$ and $N\theta_{1,N}^2 \to 0$ we have by Theorem 3.4.2 for $j = 1,2,4$

$$(3.4.12) \qquad \alpha_N^{(j)}(\beta,\Theta_N)/\alpha_N^{(3)}(\beta,\Theta_N) \sim (N\theta_{1,N})^{(m-1)(n_1-1)}, \qquad N \to \infty,$$

implying that the deficiency of Pillai's test, Hotelling's T_0^2 and the LR test w.r.t. Roy's test is asymptotically equivalent to $(m-1)(n_1-1)\theta_{1,N}^{-1} \log(N\theta_{1,N})$. Hence Roy's test is better than the other tests against these sequences of alternatives.

If $\Theta_N \to 0$ and $N\ \mathrm{tr}\ \Theta_N \to c \in (0,\infty)$, we are in the familiar case of contiguous alternatives. Asymptotic expansions of the dfs of Pillai's test, Hotelling's T_0^2 and the LR test under H_0 and under alternatives Θ_N such that $N\Theta_N = \Lambda$, where $\Lambda = \mathrm{diag}(\lambda_1,\ldots,\lambda_m)$ does not depend on N, have been given by FUJIKOSHI (1970). These expansions can be used to compute Hodges-Lehmann deficiencies (as introduced by HODGES & LEHMANN (1970)). It makes sense to compute these deficiencies, since the Pitman efficiencies of Pillai's test, Hotelling's T_0^2 and the LR test w.r.t. one another are equal to one.

We first compute the Hodges-Lehmann deficiency of the LR test w.r.t.

Pillai's test for some fixed size $\alpha \in (0,1)$. Let $\{\Theta_N\}$ be a sequence of alternatives such that $N\Theta_N = \Lambda = \mathrm{diag}(\lambda_1,\ldots,\lambda_m)$ and let $2\delta = \mathrm{tr}\,\Lambda$, $4\varepsilon = \mathrm{tr}\,\Lambda^2$. It follows from Theorem 5.2 in FUJIKOSHI (1970) that the power at Θ_N of a test based on $t_N^{(1)}$ with size α has the asymptotic expansion

$$(3.4.13) \quad \beta_N^{(1)}(\Theta_N) = \Pr\{\chi_f^2(\delta) > x_N\} - (4N)^{-1}\{(m+n_1+1)f\,\Pr\{\chi_f^2(\delta) > x_N\}$$

$$-2(m+n_1+1)f\,\Pr\{\chi_{f+2}^2(\delta) > x_N\}+\{(m+n_1+1)f-4(m+n_1+1)\delta$$

$$-4\varepsilon\}\Pr\{\chi_{f+4}^2(\delta) > x_N\}+4(m+n_1+1)\delta\,\Pr\{\chi_{f+6}^2(\delta) > x_N\}$$

$$+4\varepsilon\,\Pr\{\chi_{f+8}^2(\delta) > x_N\}\}+\mathcal{O}(N^{-2}), \qquad N \to \infty,$$

where $f = mn_1$, x_N is chosen such that $P_0\{Nt_N^{(1)} > x_N\} = \alpha$ and where $\chi_n^2(\delta)$ has a non-central chi-squared distribution with n degrees of freedom and non-centrality parameter δ. Here FUJIKOSHI (1970) is followed. Often 2δ is called the non-centrality parameter. The power at Θ_N of a test based on $t_N^{(4)}$ with size α has the asymptotic expansion

$$(3.4.14) \quad \beta_N^{(4)}(\Theta_N) = \Pr\{\chi_f^2(\delta) > x_\alpha\} + (2N)^{-1}\{(m+n_1+1)\delta\,\Pr\{\chi_{f+2}^2(\delta) > x_\alpha\}$$

$$-\{(m+n_1+1)\delta-2\varepsilon\}\Pr\{\chi_{f+4}^2(\delta) > x_\alpha\}$$

$$-2\varepsilon\,\Pr\{\chi_{f+6}^2(\delta) > x_\alpha\}\} + \mathcal{O}(N^{-2}), \qquad N \to \infty,$$

where x_α is given by $\Pr\{\chi_f^2 > x_\alpha\} = \alpha$ and δ, ε and f are as before. The expansion (3.4.14) follows from FUJIKOSHI (1970), p. 106 and 107.
Since $\frac{d}{d\eta}\Pr\{\chi_f^2(\eta) > x_\alpha\} = \Pr\{\chi_{f+2}^2(\eta) > x_\alpha\} - \Pr\{\chi_f^2(\eta) > x_\alpha\}$, one has by Taylor expansion

$$(3.4.15) \quad \beta_{N+k}^{(4)}(\Theta_N) = \beta_{N+k}^{(4)}(\Theta_{N+k}) + k\delta N^{-1}\{\Pr\{\chi_{f+2}^2(\delta) > x_\alpha\}$$

$$-\Pr\{\chi_f^2(\delta) > x_\alpha\}\} + \mathcal{O}(N^{-2})$$

$$= \beta_N^{(4)}(\Theta_N) + k\delta N^{-1}\{\Pr\{\chi_{f+2}^2(\delta) > x_\alpha\}$$

$$-\Pr\{\chi_f^2(\delta) > x_\alpha\}\} + \mathcal{O}(N^{-2}), \qquad N \to \infty,$$

where k is some fixed integer. The "critical value" x_N in (3.4.13) has the expansion

$$(3.4.16) \quad x_N = x_\alpha - (m+n_1+1)\{2(f+2)N\}^{-1}x_\alpha(x_\alpha-f-2) + O(N^{-2}), \quad N \to \infty.$$

(See FUJIKOSHI (1970), p.97.)

Let $g_{\delta,f}$ denote the density of a non-central chi-squared distribution with f degrees of freedom and non-centrality parameter δ, let

$$(3.4.17) \quad a_1 = (m+n_1+1)\{2(f+2)\}^{-1}x_\alpha(x_\alpha-f-2)g_{\delta,f}(x_\alpha)$$

and let a_2 be the coefficient of N^{-1} in the expansion (3.4.13) with x_N replaced by x_α. Then, by (3.4.13) and (3.4.16)

$$(3.4.18) \quad \beta_N^{(1)}(\Theta_N) = \Pr\{\chi_f^2(\delta) > x_\alpha\} + N^{-1}(a_1+a_2) + O(N^{-2}).$$

Let c be the coefficient of N^{-1} in the expansion (3.4.14) and let d be defined by

$$(3.4.19) \quad d = \delta(\Pr\{\chi_{f+2}^2(\delta) > x_\alpha\} - \Pr\{\chi_f^2(\delta) > x_\alpha\}).$$

It then follows from (3.4.14), (3.4.15) and (3.4.18) that the Hodges-Lehmann deficiency of the LR test w.r.t. Pillai's test is given by

$$(3.4.20) \quad d_1(\alpha,\delta,\epsilon) = (a_1+a_2-c)/d.$$

Note that this deficiency is *finite* in contrast to the Bahadur deficiency.

The computation of the Hodges-Lehmann deficiency of Hotelling's T_0^2 w.r.t. the LR test may also be based on the expansions given in FUJIKOSHI (1970). Since the computations are quite similar we only give the result. We define, using the same notation as before,

$$(3.4.21) \quad b_1 = -\{2(f+2)\}^{-1}x_\alpha\{(m+n_1+1)x_\alpha+(m-n_1+1)(f+2)\}g_{\delta,f}(x_\alpha),$$

$$(3.4.22) \quad b_2 = \tfrac{1}{4}\{(n_1-m-1)f\,\Pr\{\chi_f^2(\delta) > x_\alpha\} - 2n_1(mn_1-2\delta)\Pr\{\chi_{f+2}^2(\delta) > x_\alpha\}$$

$$+\{f(m+n_1+1)-4(m+2n_1+1)\delta+4\epsilon\}\Pr\{\chi_{f+4}^2(\delta) > x_\alpha\} + 4\{(m+n_1+1)\delta$$

$$-2\epsilon\}\Pr\{\chi_{f+6}^2(\delta) > x_\alpha\} + 4\epsilon\,\Pr\{\chi_{f+8}^2(\delta) > x_\alpha\}\}.$$

The Hodges-Lehmann deficiency of the LR test w.r.t. Hotelling's T_0^2 is given by

$$(3.4.23) \quad d_2(\alpha,\delta,\epsilon) = (b_1+b_2-c)/d,$$

where c and d have the same meaning as in (3.4.20). It follows that the Hodges-Lehmann deficiency of Hotelling's T_0^2 w.r.t. Pillai's test is given by

$$(3.4.24) \quad d_3(\alpha,\delta,\epsilon) = (a_1+a_2-b_1-b_2)/d.$$

Numerical computations on the basis of (3.4.20), (3.4.23) and (3.4.24) confirm the numerical results on power behavior in table 4 on p. 108 of FUJIKOSHI (1970): if $\lambda_1,\ldots,\lambda_m$ (i.e. $N\theta_{1,N},\ldots,N\theta_{m,N}$) are close to one another we have $\{t_N^{(1)}\} > \{t_N^{(4)}\} > \{t_N^{(2)}\}$, meaning that Pillai's test is more powerful than the LR test and the LR test is more powerful than Hotelling's T_0^2, whereas if λ_1 is dominant w.r.t. $\lambda_2,\ldots,\lambda_m$ the inequalities are reversed.

These results also explain why the LR test occupies an intermediate position in numerical comparisons on power behavior in contrast to the optimality in the sense of Bahadur efficiency. In numerical investigations it is convenient to consider moderate values of $N\theta_{i,N}$. But it is clear from our analytical results that the LR test compares more favorably w.r.t. the other tests for sequences of alternatives $\{\theta_N\}$ such that $N \operatorname{tr} \theta_N \to \infty$, especially for sequences of fixed alternatives $\{\theta_N\} = \{\theta,\theta,\ldots\}$.

3.5. APPENDIX

We summarize the notation which will be used in this section. The symbol u_β will denote $\Phi^{-1}(1-\beta)$, where Φ is the standard normal df, N denotes sample size, n_1 and n denote the degrees of freedom of the "hypothesis matrix" $S_{h,N}$ and the "error matrix" $S_{e,N}$, respectively. The alternative to H_0 is specified by the non-centrality matrix θ_N of the non-central Wishart distribution $W_m(n_1,I,N\theta_N)$. The matrix θ_N has the structure $\theta_N = = \operatorname{diag}(\theta_{1,N},\ldots,\theta_{m,N})$ with $\theta_{1,N} \geq \theta_{2,N} \geq \ldots \geq \theta_{m,N}$. We shall use the abbreviation etr(A) to denote exp(tr A).

<u>PROPOSITION 3.5.1.</u> *Let $\{\Theta_N\}$ be a bounded sequence of alternatives such that*

$$(3.5.1) \qquad N \text{ tr } \Theta_N \to \infty, \qquad \text{as } N \to \infty.$$

Let the critical values $c_N^{(1)}$, $c_N^{(2)}$ and $c_N^{(4)}$ be defined by (3.3.1) for Pillai's test, Hotelling's T_0^2 and the LR test, respectively. Then, for Pillai's test

$$(3.5.2) \qquad c_N^{(1)} = \text{tr}\{\Theta_N(I+\Theta_N)^{-1}\} + N^{-\frac{1}{2}}\tau_{1,N}u_\beta + N^{-1}[(n_1-m-1)a_{2,N} + a_{1,N}a_{2,N} + a_{3,N}$$

$$+ (N-n)(a_{1,N}-a_{2,N}) + 4\tau_{1,N}^{-2} \text{ tr}\{\Theta_N(I+\Theta_N)^{-4}(I-4\Theta_N(I+\Theta_N)^{-1}$$

$$+ \frac{11}{3}\Theta_N^2(I+\Theta_N)^{-2} - \Theta_N^3(I+\Theta_N)^{-3})\}(u_\beta^2-1)] + O(N^{-3/2}\tau_{1,N}^{-1})$$

where

$$(3.5.3) \qquad a_{i,N} = \text{tr}\{(I+\Theta_N)^{-i}\}, \qquad i = 1,2,3,$$

and

$$(3.5.4) \qquad \tau_{1,N} = \{2(a_{2,N}-a_{4,N})\}^{\frac{1}{2}}.$$

For Hotelling's T_0^2 we have

$$(3.5.5) \qquad c_N^{(2)} = \text{tr } \Theta_N + N^{-\frac{1}{2}}\tau_{2,N}u_\beta + N^{-1}[mn_1 + (m+1+N-n)b_{1,N} +$$

$$\tau_{2,N}^{-2}(4b_{1,N}+8b_{2,N}+ \frac{8}{3} b_{3,N})(u_\beta^2-1)] + O(N^{-3/2}\tau_{2,N}^{-1}),$$

where

$$(3.5.6) \qquad b_{i,N} = \text{tr}(\Theta_N^i), \qquad i = 1,2,3,$$

and

$$(3.5.7) \qquad \tau_{2,N} = (4b_{1,N}+2b_{2,N})^{\frac{1}{2}}.$$

For the LR test we have

$$(3.5.8) \quad c_N^{(4)} = \log|I+\Theta_N| + N^{-\frac{1}{2}}\tau_{4,N}u_\beta + N^{-1}[mn_1 + (N-n-n_1)d_{1,N} +$$

$$+ \tfrac{1}{2}d_{2,N} + \tfrac{1}{2}d_{1,N}^2 + \tau_{4,N}^{-2}(4d_{1,N} - 8d_{2,N} + \tfrac{20}{3}d_{3,N} - 2d_{4,N})(u_\beta^2-1)]$$

$$+ \mathcal{O}(N^{-3/2}\tau_{4,N}^{-1}),$$

where

$$(3.5.9) \quad d_{i,N} = \mathrm{tr}\{\Theta_N^i(I+\Theta_N)^{-i}\}, \qquad i = 1,\ldots,4,$$

$$(3.5.10) \quad \tau_{4,N} = (4d_{1,N} - 2d_{2,N})^{\frac{1}{2}}.$$

PROOF. We only prove (3.5.5) and omit the quite similar (though slightly more involved) proofs of (3.5.2) and (3.5.8). The proof is an elaboration of arguments used in FUJIKOSHI (1977) and SUGIURA (1973).

We have

$$N^{-1}S_{h,N} = \Theta_N + N^{-\frac{1}{2}}(Y\Theta_N^{\frac{1}{2}} + \Theta_N^{\frac{1}{2}}Y') + N^{-1}(YY' + ZZ'),$$

where the $m \times m$ matrix $Y = (y_{ij})$ and the $m \times (n_1-m)$ matrix $Z = (z_{ij})$ have independent standard normal variables as their elements. Moreover, let

$$(3.5.11) \quad V_N = N^{\frac{1}{2}}(N^{-1}S_{e,N} - I).$$

We define the norm $\|A\|$ of an $m \times m$ matrix A by $\|A\| = \{\mathrm{tr}\, A'A\}^{\frac{1}{2}}$. Note that $\|AB\| \le \|A\| \|B\|$ (cf. BELLMAN (1960), Exercise 1 p.162). Let B_N denote the event

$$(3.5.12) \quad B_N = \{\|Y\| + \|ZZ'\| + \|V_N\| \le (N\tau_{2,N}^2)^{\frac{1}{24}}\}.$$

Then

$$(3.5.13) \quad P_{\Theta_N}(B_N) = 1 - \mathcal{O}(\exp\{-(cN\tau_{2,N}^2)^{\frac{1}{12}}\}) = 1 - \mathcal{O}((N\tau_{2,N}^2)^{-\infty}),$$

where c is a positive constant and $\mathcal{O}((N\tau_{2,N}^2)^{-\infty})$ means "$\mathcal{O}((N\tau_{2,N}^2)^{-k})$ for each $k > 0$".

On B_N the statistic $\tau_{2,N}^{-2}S_{h,N}S_{e,N}^{-1}$ has the convergent power series expansion

$$(3.5.14) \quad \tau_{2,N}^{-2} S_{h,N} S_{e,N}^{-1} = \tau_{2,N}^{-2}\Theta_N + \sum_{k=1}^{\infty} \epsilon_N^k A_{k,N},$$

where

$$\epsilon_N = (N\tau_{2,N}^2)^{-\frac{1}{2}},$$

$$(3.5.15) \quad A_{1,N} = \tau_{2,N}^{-1}(Y\Theta_N^{\frac{1}{2}}+\Theta_N^{\frac{1}{2}}Y'-\Theta_N V_N),$$

$$A_{2,N} = YY' + ZZ' - \tau_{2,N}A_{1,N}V_N,$$

$$A_{k,N} = -\tau_{2,N}A_{k-1,N}V_N, \qquad k = 3,4,\dots .$$

The series is convergent on B_N since by (3.5.12) and (3.5.15)

$$\epsilon_N^k \|A_{k,N}\| \le c_1 \epsilon_N^{\frac{11}{12}k}$$

for some constant $c_1 > 0$ and all $N \in \mathbb{N}$. By (3.5.14) one also has the following convergent power-series expansion on B_N

$$(3.5.16) \quad \tau_{2,N}^{-2} \operatorname{tr} S_{h,N} S_{e,N}^{-1} = \tau_{2,N}^{-2} \operatorname{tr} \Theta_N + \sum_{k=1}^{3} \epsilon_N^k \operatorname{tr} A_{k,N} + R_N,$$

where $|R_N| = |\sum_{k=4}^{\infty} \epsilon_N^k \operatorname{tr} A_{k,N}| \le c_2 \epsilon_N^3$, for some constant $c_2 > 0$ and all $N \in \mathbb{N}$.

Let F_N be the df of $N^{\frac{1}{2}}\tau_{2,N}^{-1} \operatorname{tr}(S_{h,N}S_{e,N}^{-1}-\Theta_N)$ and let G_N be the df of $\sum_{k=1}^{3} \epsilon_N^{k-1} \operatorname{tr} A_{k,N}$.
Then, by (3.5.13)

$$F_N(x) \ge P_{\Theta_N}(\{\sum_{k=1}^{\infty} \epsilon_N^{k-1} \operatorname{tr} A_{k,N} \le x\} \cap B_N)$$

$$\ge G_N(x-c_2\epsilon_N^2) - \mathcal{O}(\epsilon_N^\infty),$$

and similarly

$$F_N(x) \le G_N(x+c_2\epsilon_N^2) + \mathcal{O}(\epsilon_N^\infty).$$

We shall prove that for each $c > 0$

$$G_N(x\pm c\epsilon_N^2) = G_N(x) + \mathcal{O}(\epsilon_N^2), \qquad N \to \infty,$$

108

implying

(3.5.17) $F_N(x) = G_N(x) + O(\varepsilon_N^2)$, $x \in \mathbb{R}$.

Let $\psi_N(t) = E \; \text{etr}(it \sum_{k=1}^{3} \varepsilon_N^{k-1} A_{k,N})$. Since for fixed $t \in \mathbb{R}$

$$E \; \text{etr}(it \; \tau_{2,N} A_{1,N}) = E \, \text{etr}(2it \; Y\Theta_N^{\frac{1}{2}}) E \; \text{etr}(-it \; \Theta_N V_N)$$

$$= \text{etr}(-2t^2 \Theta_N) \, |I+2it \; N^{-\frac{1}{2}} \Theta_N|^{-\frac{1}{2}n} \text{etr}(it \; N^{\frac{1}{2}} \Theta_N)$$

$$= \text{etr}\{-t^2(2\Theta_N+\Theta_N^2)\}\{1+it(N-n)N^{-\frac{1}{2}}\text{tr} \; \Theta_N - \frac{4}{3}N^{-\frac{1}{2}}(it)^3 \text{tr} \; \Theta_N^3 + O(N^{-1}\text{tr} \; \Theta_N^2)\}\;,$$

the statistic $\sum_{k=1}^{3} \varepsilon_N^{k-1} \; \text{tr} \; A_{k,N}$ has a standard normal df as its limiting distribution. For $|t| < \varepsilon_N^{-\frac{1}{4}}$ the function $\psi_N(t)$ has the expansion

(3.5.18) $\psi_N(t) = E \; \text{etr}(it \; A_{1,N})\{1+it \; \varepsilon_N \; \text{tr} \; A_{2,N}$

$$+ it \; \varepsilon_N^2 \; \text{tr} \; A_{3,N} + \tfrac{1}{2}(it \; \varepsilon_N)^2 (\text{tr} \; A_{2,N})^2\}$$

$$+ O(\varepsilon_N^{5/2} \; \max(|t|^2, |t|^3))\;.$$

This follows from the fact that on B_N

$$\sum_{j=2}^{\infty} \frac{1}{j!} \{\text{tr}(\varepsilon_N A_{2,N}+\varepsilon_N^2 A_{3,N})t\}^j - \tfrac{1}{2}(t\varepsilon_N \; \text{tr} \; A_{2,N})^2$$

$$= O(\varepsilon_N^{5/2} \; \max(|t|^2, |t|^3))\;.$$

The expectations in (3.5.18) can be evaluated in the same way as in SUGIURA (1973), implying that for $|t| < \varepsilon_N^{-\frac{1}{4}}$ the function $\psi_N(t)$ has the expansion

(3.5.19) $\psi_N(t) = e^{-\frac{1}{2}t^2}\{1+\varepsilon_N \sum_{k=1}^{2} \gamma_k(\Theta_N) (it)^{2k-1} + O(\varepsilon_N^2 \; \max(|t|^2, |t|^6))\}$

$$+ \max(|t|^2, |t|^3) O(\varepsilon_N^{5/2})\;,$$

where

$$\gamma_1(\Theta_N) = mn_1 + (m+1+N-n)b_{1,N},$$

(3.5.20)

$$\gamma_2(\Theta_N) = \tau_{2,N}^{-2}(4b_{1,N}+8b_{2,N}+\tfrac{8}{3}b_{3,N}),$$

with $b_{i,N} = tr(\Theta_N^i)$, $i = 1,2,3$.

Next, for $t \geq \varepsilon_N^{-\frac{1}{4}}$ we have

(3.5.21) $\quad \psi_N(t) = O(\varepsilon_N^\infty).$

This can be proved in the following way. Consider for each $N \in \mathbb{N}$

(3.5.22) $\quad h_N(y_{11}) = \sum_{k=1}^{3} \varepsilon_N^{k-1} \, tr \, A_{k,N}$

as a function of y_{11}, with fixed y_{ij}, $(i,j) \neq (1,1)$, fixed Z and V_N.
Then

$$h_N'(y_{11}) = 2\tau_{2,N}^{-1}\theta_{1,N}^{\frac{1}{2}} + 2\varepsilon_N(y_{11}-\theta_{1,N}^{\frac{1}{2}}v_{11,N})$$

$$-2\varepsilon_N^2(v_{11,N}y_{11}+\sum_{j=2}^{m} v_{1j,N}y_{j1}-\theta_{1,N}^{\frac{1}{2}}\sum_{k=1}^{m} v_{1k,N}^2)\tau_{2,N},$$

where $V_N = (v_{ij,N})$. It follows that there exist constants d and e such that
for all N, $h_N'(y_{11}) \geq e > 0$ if $|y_{ij}| \leq d\,\varepsilon_N^{-1}$ and $|v_{ij,N}| \leq d\,\varepsilon_N^{-1}$,
$i,j = 1,\ldots,m$.
Putting $k_N(y_{11}) = \dfrac{d}{dy_{11}}\{(2\pi)^{-\frac{1}{2}}\exp(-\frac{1}{2}y_{11}^2)/h_N'(y_{11})\}$ we have by integration by
parts

$$(2\pi)^{-\frac{1}{2}} \int_{-d\varepsilon_N^{-1}}^{d\varepsilon_N^{-1}} \exp\{-\tfrac{1}{2}y_{11}^2 + it\,h_N(y_{11})\}dy_{11}$$

$$= \left[(2\pi)^{-\frac{1}{2}}\exp\{-\tfrac{1}{2}y_{11}^2 + it\,h_N(y_{11})\}/(it\,h_N'(y_{11}))\right]_{-d\varepsilon_N^{-1}}^{d\varepsilon_N^{-1}}$$

$$- (it)^{-1} \int_{-d\varepsilon_N^{-1}}^{d\varepsilon_N^{-1}} k_N(y_{11})\exp\{it\,h_N(y_{11})\}dy_{11}$$

$$= O(|t|^{-1}\varepsilon_N^\infty) + O(|t|^{-1}) = O(|t|^{-1}).$$

The asymptotic relation holds *uniformly* for

110

$$|y_{ij}| \leq d\epsilon_N^{-1}, \quad (i,j) \neq (1,1),$$

(3.5.23)

$$|v_{ij}| \leq d\epsilon_N^{-1}, \quad |z_{ij}| \leq d\epsilon_N^{-1}, \qquad i,j = 1,\ldots,m.$$

Repeated integration by parts shows that

$$(2\pi)^{-\frac{1}{2}} \int_{-d\epsilon_N^{-1}}^{d\epsilon_N^{-1}} \exp\{-\tfrac{1}{2}y_{11}^2 + it\, h_N(y_{11})\} dy_{11} = O(|t|^{-\infty}) = O(\epsilon_N^{\infty}),$$

again uniformly under (3.5.23). Since

$$P_{\Theta_N}(\{\sum_{i,j=1}^{m} (|y_{ij}|+|z_{ij}|+|v_{ij,N}|) \leq d\epsilon_N^{-1}\}) = 1 - O(\epsilon_N^{\infty}),$$

we have

$$\psi_N(t) = E(E(\text{etr}(it \sum_{k=1}^{3} \epsilon_N^{k-1} A_{k,N} \mid y_{ij}, (i,j) \neq (1,1), Z, V_N)) =$$

$$= O(\epsilon_N^{\infty}).$$

This proves (3.5.21).

Let

(3.5.24) $$\widetilde{\phi}_N(t) = e^{-\frac{1}{2}t^2}\{1+\epsilon_N \sum_{k=1}^{2} \gamma_k(\Theta_N)(it)^{2k-1}\}, \qquad t \in \mathbb{R}.$$

By (3.5.19)

(3.5.25) $$\int_{|t|<\epsilon_N^{-\frac{1}{4}}} |(\psi_N(t)-\widetilde{\phi}_N(t))/t|\, dt = O(\epsilon_N^2),$$

and by (3.5.21) and (3.5.24)

(3.5.26) $$\int_{\epsilon_N^{-\frac{1}{4}}<|t|<\epsilon_N^{-2}} |(\psi_N(t)-\widetilde{\phi}_N(t))/t|\, dt$$

$$\leq \int_{\epsilon_N^{-\frac{1}{4}}<|t|<\epsilon_N^{-2}} |\psi_N(t)/t|\, dt + \int_{\epsilon_N^{-\frac{1}{4}}<|t|<\epsilon_N^{-2}} |\widetilde{\phi}_N(t)/t|\, dt = O(\epsilon_N^{\infty}).$$

The function $\widetilde{\phi}_N$ is the Fourier transform of the function $\widetilde{F}_N$ defined by

$$(3.5.27) \quad \widetilde{F}_N(x) = \Phi(x) - \varepsilon_N\{\gamma_1(\Theta_N)\phi(x) + \gamma_2(\Theta_N)\phi''(x)\},$$

where $\phi = \Phi'$. By Esseen's smoothing Lemma (ESSEEN (1945), Theorem 2a, p.32),

$$(3.5.28) \quad |G_N(x) - \widetilde{F}_N(x)| \le \pi^{-1}\int_{-\varepsilon_N^{-2}}^{\varepsilon_N^{-2}} |(\phi_N(t) - \widetilde{\phi}_N(t))/t|dt + O(\varepsilon_N^2) = O(\varepsilon_N^2),$$

uniformly in $x \in \mathbb{R}$. Hence $G_N(x + c\varepsilon_N^2) = G_N(x) + O(\varepsilon_N^2)$ for each constant $c \in \mathbb{R}$ and by (3.5.17), (3.5.27) and (3.5.28)

$$F_N(x) = \widetilde{F}_N(x) + O(\varepsilon_N^2) = \Phi(x) - \varepsilon_N\{\gamma_1(\Theta_N)\phi(x) + \gamma_2(\Theta_N)\phi''(x)\} + O(\varepsilon_N^2),$$

implying (3.5.5). $\square$

PROPOSITION 3.5.2. Let $\{\Theta_N\}$ be a bounded sequence of alternatives such that
(i) $N \, \mathrm{tr} \, \Theta_N \to \infty$, as $N \to \infty$
and
(ii) $\theta_{2,N}/\theta_{1,N} \le c < 1$, for N sufficiently large.
Let the critical value $c_N^{(3)}$ of Roy's largest root test be defined by (3.3.1). Then $c_N^{(3)}$ has the expansion

$$(3.5.29) \quad c_N^{(3)} = \theta_{1,N} + N^{-\frac{1}{2}}\tau_{3,N}u_\beta + N^{-1}\{n_1 + (m+1+N-n)\theta_{1,N} +$$

$$+ \sum_{j=2}^{m} (\theta_{1,N} - \theta_{j,N})^{-1}(\theta_{1,N} + \theta_{j,N} + \theta_{1,N}\theta_{j,N})$$

$$+ \tau_{3,N}^{-2}(4\theta_{1,N} + 8\theta_{1,N}^2 + \frac{8}{3}\theta_{1,N}^3)(u_\beta^2 - 1)\} + O(N^{-3/2}\tau_{3,N}^{-1}),$$

where

$$(3.5.30) \quad \tau_{3,N} = \{2\theta_{1,N}(2+\theta_{1,N})\}^{\frac{1}{2}}.$$

PROOF. Let

$$(3.5.31) \quad \varepsilon_N = \{N\tau_{3,N}^2\}^{-\frac{1}{2}},$$

$$D_{1,N} = \tau_{3,N}^{-1}(Y\Theta_N^{\frac{1}{2}} + \Theta_N^{\frac{1}{2}}Y' - \Theta_N V_N),$$

$$D_{2,N} = YY' + ZZ' - \tau_{3,N}D_{1,N}V_N'$$

$$D_{k,N} = -\tau_{3,N}D_{k-1,N}V_N' \qquad k = 3,4,\ldots,$$

where Y, Z and V_N are defined as in the proof of Proposition 3.5.1. On the set B_N, defined by (3.5.12) with $\tau_{2,N}$ replaced by $\tau_{3,N}$, the statistic $\tau_{3,N}^{-2} S_{h,N} S_{e,N}^{-1}$ has the convergent power series expansion

$$(3.5.32) \qquad \tau_{3,N}^{-2} S_{h,N} S_{e,N}^{-1} = \tau_{3,N}^{-2}\Theta_N + \sum_{k=1}^{\infty} \varepsilon_N^k D_{k,N}.$$

Let $\Lambda_N = \tau_{3,N}^{-2}\Theta_N$ and let $\lambda_{1,N} \geq \ldots \geq \lambda_{m,N}$ be the diagonal elements of Λ_N. On B_N the largest latent root $\ell_{1,N}$ of $\tau_{3,N}^{-2} S_{h,N} S_{e,N}^{-1}$ can be expanded in the following way

$$(3.5.33) \qquad \ell_{1,N} = \tau_{3,N}^{-2}\theta_{1,N} + \varepsilon_N d_{11}^{(1)} + \varepsilon_N^2\{d_{11}^{(2)} + \sum_{j>1}\mu_{1j}(d_{1j}^{(1)})^2\} +$$

$$\varepsilon_N^3\{d_{11}^{(3)} + \sum_{j>1}\mu_{1j}(d_{1j}^{(1)}d_{j1}^{(2)} + d_{1j}^{(2)}d_{j1}^{(1)} - \mu_{1j}d_{11}^{(1)}d_{12}^{(1)}d_{21}^{(1)} +$$

$$+ \sum_{i,j\neq 1}\mu_{1i}\mu_{1j}d_{1i}^{(1)}d_{ij}^{(1)}d_{j1}^{(1)}\} + R_N,$$

where $d_{ij}^{(k)}$ is the ij-th element of the matrix $D_{k,N}$, $\mu_{1j} = \theta_{1,N}(\theta_{1,N}-\theta_{j,N})^{-1}$ and $|R_N| \leq c_1\varepsilon_N^4$, for some constant c_1. The coefficients in (3.5.33) can be computed by a perturbation method (see BELLMAN (1960) or FUJIKOSHI (1978)). Let F_N be the df of $\varepsilon_N^{-1}(\ell_{1,N}-\tau_{3,N}^{-2}\theta_{1,N})$ and let G_N be the df of $\varepsilon_N^{-1}(\ell_{1,N}-\tau_{3,N}^{-2}\theta_{1,N}-R_N)$. Then, by the same arguments as used in the proof of Proposition 3.5.1

$$F_N(x) = G_N(x) + \mathcal{O}(\varepsilon_N^2)$$

and

$$G_N(x) = \Phi(x) - \varepsilon_N\{h_1(\Theta_N)\phi(x) + h_2(\Theta_N)\phi''(x)\} + \mathcal{O}(\varepsilon_N^2),$$

where

$$(3.5.34) \qquad h_1(\Theta_N) = n_1 + (m+1+N-n)\theta_{1,N} + \sum_{j>1}(\theta_{1,N}-\theta_{j,N})^{-1}(\theta_{1,N}+\theta_{j,N}+\theta_{1,N}\theta_{j,N}),$$

$$h_2(\Theta_N) = \frac{4}{3}\theta_{1,N}(3+6\theta_{1,N}+2\theta_{1,N}^2)\tau_{3,N}^{-2}.$$

The proof can be completed in the same way as the proof of Proposition 3.5.1. □

REMARK 3.5.1. If condition (ii) is not satisfied, the expansion (3.5.29) is not correct. However, if (i) holds, we still have

$$(3.5.35) \qquad c_N^{(3)} = \theta_{1,N} + O(N^{-\frac{1}{2}}),$$

since (3.5.32) remains valid under this condition.

REFERENCES

[1] ANDERSON, T.W. (1958), *An introduction to multivariate statistical analysis*, Wiley, New York.

[2] BAHADUR, R.R. (1960), *Stochastic comparison of tests*, Ann. Math. Statist. 31, 276-295.

[3] BAHADUR, R.R. (1967), *Rates of convergence of estimates and test statistics*, Ann. Math. Statist. 38, 303-324.

[4] BAHADUR, R.R. (1971), *Some limit theorems in statistics*, SIAM, Philadelphia.

[5] BAHADUR, R.R. & M. RAGHAVACHARI (1972), *Some asymptotic properties of likelihood ratios on general sample spaces*, Proc. Sixth Berkeley Symp. Math. Statist. Prob. 1, 129-152.

[6] BAHADUR, R.R. & S.L. ZABELL (1979), *Large deviations of the sample mean in general vector spaces*, Ann. Probability 7, 587-621.

[7] BÁRTFAI, P. (1977), *On the multivariate Chernoff theorem*, Preprint of the Math. Inst. Hung. Acad. Sci..

[8] BELLMAN, R. (1960), *Introduction to matrix analysis*, McGraw-Hill, New York.

[9] BICKEL, P.J. & E.L. LEHMANN (1975), *Descriptive statistics for nonparametric models*, II. *Location*, Ann. Statist. 3, 1045-1069.

[10] BLEISTEIN, N. & R.A. HANDELSMAN (1975), *Asymptotic expansions of integrals*, Holt, Rinehart and Winston, New York.

[11] BOOK, S.A. (1974), *Large deviations of the Winsorized mean of an exponential distribution*, Naval Res. Logist. Quart., 21, 549-557.

[12] BOROVKOV, A.A. (1962), *New limit theorems in boundary-value problems for sums of independent terms* (in Russian), Sibirsk. Math. Zh. 3, 645-695 (English translation in Sel. Transl. Math. Statist. Prob. 5 (1965), 315-372.)

[13] BOROVKOV, A.A. (1967), *Boundary-value problems for random walks and large deviations in function spaces*, Theory Prob. Applications XII, 4, 575-595.

[14] BOROVKOV, A.A. & A.A. MOGUL'SKII (1978), *On probabilities of large de-
        viations in topological spaces,* I (in Russian), Sibirsk. Math. Zh.
        19, 988-1004.

[15] BOURBAKI, N. (1965), *Eléments de mathématique, Livre VI, Intégration,*
        Ch. 1-4, Hermann, Paris.

[16] CHAKRABARTI, R. (1978), *Large deviation large sample efficiency for*
        nonparametric tests of symmetry in the presence of ties, Unpub-
        lished thesis, University of Wisconsin, Madison.

[17] CHERNOFF, H. (1952), *A measure of asymptotic efficiency for tests of a*
        hypothesis based on sums of observations, Ann. Math. Statist. 23,
        493-507.

[18] CONOVER, W.J. (1973a), *Rank tests for one sample, two samples, and* k
        samples without the assumption of a continuous function, Ann.
        Statist. 1, 1105-1125.

[19] CONOVER, W.J. (1973b), *On methods of handling ties in the Wilcoxon*
        signed-rank test, Journal of the American Statistical Associa-
        tion 68, 985-988.

[20] CSISZÁR, I. (1962), *Informationstheoretische Konvergenzbegriffe im Raum*
        der Wahrscheinlichkeitsverteilungen, A. Magyar Tud. Ak. Mat. Kut.
        Int. Közl. 7, 137-158.

[21] CSISZÁR, I. (1975), *I-divergence geometry of probability distributions*
        and minimization problems, Ann. Prob. 3, 146-158.

[22] DONSKER, M.D. & S.R.S. VARADHAN (1975), *Asymptotic evaluation of cer-*
        tain Markov process expectations for large time I, Comm. Pure
        Appl. Math. 28, 1-47.

[23] DONSKER, M.D. & S.R.S. VARADHAN (1976), *Asymptotic evaluation of cer-*
        tain Markov process expectations for large time III, Comm. Pure
        Appl. Math. 29, 389-461.

[24] ERDÉLYI, A. (Ed.) (1953), *Higher transcendental functions,* Vol. I,
        McGraw-Hill, New York.

[25] ESSEEN, C.G. (1945), *Fourier analysis of distribution functions, a*
        mathematical study of the Laplace-Gaussian law, Acta Mathematica
        77, 1-125.

116

[26] FAN, K.I., I. GLICKSBERG & A.J. HOFFMAN (1957), *Systems of inequalities involving convex functions*, Amer. Math. Soc. Proc. 8, 617-622.

[27] FUJIKOSHI, Y. (1970), *Asymptotic expansions of the distributions of test statistics in multivariate analysis*, J. Sci. Hiroshima Univ., Ser. A-I, 34, 153-247.

[28] FUJIKOSHI, Y. (1977), *Asymptotic expansions for the distributions of some multivariate tests*, Multivariate Analysis IV (P.R. Krishnaiah, Ed.) 55-71, North-Holland Publishing Company, Amsterdam.

[29] FUJIKOSHI, Y. (1978), *Asymptotic expansions for the distribution of some functions of the latent roots of matrices in three situations*, J. Multivariate Anal. 8, 63-72.

[30] GÄNSSLER, P. (1971), *Compactness and sequential compactness in spaces of measures*, Z. Wahrscheinlichkeitstheorie verw. Geb. 17, 124-146.

[31] GROENEBOOM, P., Y. LEPAGE & F.H. RUYMGAART (1976), *Rank tests for independence with best strong exact Bahadur slope*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 36, 119-127.

[32] GROENEBOOM, P. (1976), *Large deviation theorems for empirical distribution functions*, Report SW 46/76, Mathematical Centre, Amsterdam.

[33] GROENEBOOM, P. & J. OOSTERHOFF (1977), *Bahadur efficiency and probabilities of large deviations*, Statistica Neerlandica 31, 1-24.

[34] GROENEBOOM, P., J. OOSTERHOFF & F.H. RUYMGAART (1979), *Large deviation theorems for empirical probability measures*, Ann. Probability 7, 553-586.

[35] GROENEBOOM, P. & J. OOSTERHOFF (1980), *Bahadur efficiency and small-sample efficiency*, Manuscript in preparation.

[36] HÁJEK, J. (1974), *Asymptotic sufficiency of the vector of ranks in the Bahadur sense*, Ann. Statist. 2, 75-83.

[37] HO, NGUYEN VAN (1974), *Asymptotic efficiency in the Bahadur sense for the signed rank tests*, Proc. Prague Symp. on Asymptotic Statistics (3-6 September 1973) 2, 127-156.

[38] HOADLEY, A.B. (1967), *On the probability of large deviations of functions of several empirical cdf's*, Ann. Math. Statist. 38, 360-381.

[39] HODGES, J.L. & E.L. LEHMANN (1970), *Deficiency*, Ann. Math. Statist. <u>43</u>, 1041-1067.

[40] HOEFFDING, W. (1965), *Asymptotically optimal tests for multinomial distributions*, Ann. Math. Statist. <u>36</u>, 369-401.

[41] HOEFFDING, W. (1967), *On probabilities of large deviations*, Proc. Fifth Berkeley Symp. Math. Stat. Prob. <u>1</u>, 203-219.

[42] HSIEH, H.K. (1979), *Exact Bahadur efficiencies for tests of the multivariate linear hypothesis*, Ann. Statist. <u>7</u>, 1231-1245.

[43] HUBER, P.J. (1972), *Robust statistics: a review*, Ann. Math. Statist. <u>43</u>, 1041-1067.

[44] KALLENBERG, W.C.M. (1978), *Bahadur deficiency of likelihood ratio tests in exponential families*, Report no. 77, Department of Mathematics, Vrije Universiteit, Amsterdam.

[45] KLOTZ, J. (1963), *Small sample power and efficiency for the one sample Wilcoxon and normal scores tests*, Ann. Math. Statist. <u>34</u>, 624-632.

[46] KLOTZ, J. (1965), *Alternative efficiencies for signed rank tests*, Ann. Math. Statist. <u>36</u>, 1759-1766.

[47] KREMER, E. (1979), *Approximate and local Bahadur efficiency of linear rank tests in the two-sample problem*, Ann. Statist. <u>7</u>, 1246-1255.

[48] LANFORD, D.E. (1972), *Entropy and equilibrium states in classical statistical mechanics*, Statistical mechanics and mathematical problems, ed. A. Lenard, Lecture Notes in Physics <u>20</u>, 1-113, Springer-Verlag, Berlin.

[49] OOSTERHOFF, J. (1978), *Large deviations of multivariate L-estimators with monotone weight functions*, Report SW 63/78, Mathematical Centre, Amsterdam.

[50] PARTHASARATHY, K. (1967), *Probability measures on metric spaces*, Academic Press, New York.

[51] PETROV, V.V. (1965), *On the probabilities of large deviations for sums of independent random variables*, Theor. Probability Appl. <u>10</u>, 287-298.

[52] PILLAI, K.C. & K. JAYACHANDRAN (1967), *Power comparisons of tests of two multivariate hypotheses bases on four criteria,* Biometrika 54, 195-210.

[53] PINSKER, M.S. (1964), *Information and information stability of random variables and processes,* Holden-Day, San Francisco.

[54] PITMAN, E.J.G. (1949), *Lecture Notes on nonparametric statistical inference,* unpublished.

[55] PLACHKY, D. (1971), *On a theorem of G.L. Sievers,* Ann. Math. Statist. 42, 1442-1443.

[56] PLACHKY, D. & J. STEINEBACH (1975), *A theorem about probabilities of large deviations with an application to queuing theory,* Period. Math. Hungar 6, 343-345.

[57] PRATT, J.W. (1959), *Remarks on zeros and ties in the Wilcoxon signed rank procedures,* Journal of the American Statistical Association, 54, 655-667.

[58] RAGHAVACHARI, M. (1970), *On a theorem of Bahadur on the rate of convergence of test statistics,* Ann. Math. Statist. 41, 1695-1699.

[59] RÉNYI, A. (1961), *On measures of entropy and information,* Proc. 4th Berkeley Symp. Math. Statist. Prob. 1, 547-561.

[60] ROY, S.N., R. GNANADESIKAN & J.N. SRIVASTAVA (1971), *Analysis and design of certain quantitative multiresponse experiments,* Pergamon Press, Oxford.

[61] SANOV, I.N. (1957), *On the probability of large deviations of random variables* (in Russian), Mat. Sbornik N.S. 42 (84), 11-44 (English translation in Sel. Transl. Math. Statist. Prob. 1 (1961), 213-244).

[62] SCHWARTZ, R.E. (1967a), *Locally minimax tests,* Ann. Math. Statist. 37, 340-359.

[63] SCHWARTZ, R.E. (1967b), *Admissible tests in multivariate analysis of variance,* Ann. Math. Statist. 37, 698-710.

[64] SETHURAMAN, J. (1964), *On the probability of large deviations of families of sample means,* Ann. Math. Statist. 35, 1304-1316.

[65] SIEVERS, G.L. (1969), *On the probability of large deviations and exact slopes*, Ann. Math. Statist. 40, 1908-1921.

[66] SIEVERS, G.L. (1975), *Multivariate probabilities of large deviations*, Ann. Statist. 3, 897-905.

[67] SIEVERS, G.L. (1976), *Probabilities of large deviations for empirical probability measures*, Ann. Statist. 2, 362-366.

[68] STEINEBACH, J. (1977), *Exponential convergence properties of linear estimators under exponential and uniform distribution*, Metrika 24, 137-161.

[69] STONE, M. (1974), *Large deviations of empirical probability measures*, Ann. Statist. 2, 363-366.

[70] SUGIURA, N. (1973), *Further asymptotic formulas for the non-null distributions of three statistics for multivariate linear hypothesis*, Ann. Inst. Statist. Math. 25, 153-163.

[71] TUSNÁDY, G. (1977), *On asymptotically optimal tests*, Ann. Statist. 5, 385-393.

[72] VORLÍČKOVÁ, D. (1972), *Asymptotic properties of rank tests of symmetry under discrete distributions*, Ann. Math. Statist. 43, 2013-2018.

[73] WIEAND, H.S. (1976), *A condition under which the Pitman and Bahadur approaches to efficiency coincide*, Ann. Statist. 4, 1003-1011.

[74] WOODWORTH, G.G. (1970), *Large deviations and Bahadur efficiency of linear rank statistics*, Ann. Math. Statist. 41, 251-283.

AUTHOR INDEX

SUBJECT INDEX

OTHER TITLES IN THE SERIES MATHEMATICAL CENTRE TRACTS

A leaflet containing an order-form and abstracts of all publications mentioned below is available at the Mathematisch Centrum, Tweede Boerhaavestraat 49,Amsterdam,1091AL The Netherlands. Orders should be sent to the same address.

MCT 1 T. VAN DER WALT, *Fixed and almost fixed points*, 1963. ISBN 90 6196 002 9.

MCT 2 A.R. BLOEMENA, *Sampling from a graph*, 1964. ISBN 90 6196 003 7.

MCT 3 G. DE LEVE, *Generalized Markovian decision processes, part I: Model and method*, 1964. ISBN 90 6196 004 5.

MCT 4 G. DE LEVE, *Generalized Markovian decision processes, part II: Probabilistic background*, 1964. ISBN 90 6196 005 3.

MCT 5 G. DE LEVE, H.C. TIJMS & P.J. WEEDA, *Generalized Markovian decision processes, Applications*, 1970. ISBN 90 6196 051 7.

MCT 6 M.A. MAURICE, *Compact ordered spaces*, 1964. ISBN 90 6196 006 1.

MCT 7 W.R. VAN ZWET, *Convex transformations of random variables*, 1964. ISBN 90 6196 007 X.

MCT 8 J.A. ZONNEVELD, *Automatic numerical integration*, 1964. ISBN 90 6196 008 8.

MCT 9 P.C. BAAYEN, *Universal morphisms*, 1964. ISBN 90 6196 009 6.

MCT 10 E.M. DE JAGER, *Applications of distributions in mathematical physics*, 1964. ISBN 90 6196 010 X.

MCT 11 A.B. PAALMAN-DE MIRANDA, *Topological semigroups*, 1964. ISBN 90 6196 011 8.

MCT 12 J.A.TH.M. VAN BERCKEL, H. BRANDT CORSTIUS, R.J. MOKKEN & A. VAN WIJNGAARDEN, *Formal properties of newspaper Dutch*, 1965. ISBN 90 6196 013 4.

MCT 13 H.A. LAUWERIER, *Asymptotic expansions*, 1966, out of print; replaced by MCT 54 and 67.

MCT 14 H.A. LAUWERIER, *Calculus of variations in mathematical physics*, 1966. ISBN 90 6196 020 7.

MCT 15 R. DOORNBOS, *Slippage tests*, 1966. ISBN 90 6196 021 5.

MCT 16 J.W. DE BAKKER, *Formal definition of programming languages with an application to the definition of ALGOL 60*, 1967. ISBN 90 6196 022 3.

MCT 17 R.P. VAN DE RIET, *Formula manipulation in ALGOL 60, part 1*, 1968. ISBN 90 6196 025 8.

MCT 18 R.P. VAN DE RIET, *Formula manipulation in ALGOL 60, part 2*, 1968. ISBN 90 6196 038 X.

MCT 19 J. VAN DER SLOT, *Some properties related to compactness*, 1968. ISBN 90 6196 026 6.

MCT 20 P.J. VAN DER HOUWEN, *Finite difference methods for solving partial differential equations*, 1968. ISBN 90 6196 027 4.

MCT 21 E. WATTEL, *The compactness operator in set theory and topology*, 1968. ISBN 90 6196 028 2.

MCT 22 T.J. DEKKER, *ALGOL 60 procedures in numerical algebra, part 1*, 1968. ISBN 90 6196 029 0.

MCT 23 T.J. DEKKER & W. HOFFMANN, *ALGOL 60 procedures in numerical algebra, part 2*, 1968. ISBN 90 6196 030 4.

MCT 24 J.W. DE BAKKER, *Recursive procedures*, 1971. ISBN 90 6196 060 6.

MCT 25 E.R. PAERL, *Representations of the Lorentz group and projective geometry*, 1969. ISBN 90 6196 039 8.

MCT 26 EUROPEAN MEETING 1968, *Selected statistical papers, part I*, 1968. ISBN 90 6196 031 2.

MCT 27 EUROPEAN MEETING 1968, *Selected statistical papers, part II*, 1969. ISBN 90 6196 040 1.

MCT 28 J. OOSTERHOFF, *Combination of one-sided statistical tests*, 1969. ISBN 90 6196 041 X.

MCT 29 J. VERHOEFF, *Error detecting decimal codes*, 1969. ISBN 90 6196 042 8.

MCT 30 H. BRANDT CORSTIUS, *Excercises in computational linguistics*, 1970. ISBN 90 6196 052 5.

MCT 31 W. MOLENAAR, *Approximations to the Poisson, binomial and hypergeometric distribution functions*, 1970. ISBN 90 6196 053 3.

MCT 32 L. DE HAAN, *On regular variation and its application to the weak convergence of sample extremes*, 1970. ISBN 90 6196 054 1.

MCT 33 F.W. STEUTEL, *Preservation of infinite divisibility under mixing and related topics*, 1970. ISBN 90 6196 061 4.

MCT 34 I. JUHÁSZ, A. VERBEEK & N.S. KROONENBERG, *Cardinal functions in topology*, 1971. ISBN 90 6196 062 2.

MCT 35 M.H. VAN EMDEN, *An analysis of complexity*, 1971. ISBN 90 6196 063 0.

MCT 36 J. GRASMAN, *On the birth of boundary layers*, 1971. ISBN 90 6196 064 9.

MCT 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. ISBN 90 6196 065 7.

MCT 38 W.A. VERLOREN VAN THEMAAT, *Automatic analysis of Dutch compound words*, 1971. ISBN 90 6196 073 8.

MCT 39 H. BAVINCK, *Jacobi series and approximation*, 1972. ISBN 90 6196 074 6.

MCT 40 H.C. TIJMS, *Analysis of (s,S) inventory models*, 1972. ISBN 90 6196 075 4.

MCT 41 A. VERBEEK, *Superextensions of topological spaces*, 1972. ISBN 90 6196 076 2.

MCT 42 W. VERVAAT, *Success epochs in Bernoulli trials (with applications in number theory)*, 1972. ISBN 90 6196 077 0.

MCT 43 F.H. RUYMGAART, *Asymptotic theory of rank tests for independence*, 1973. ISBN 90 6196 081 9.

MCT 44 H. BART, *Meromorphic operator valued functions*, 1973. ISBN 90 6196 082 7.

MCT 45   A.A. BALKEMA, *Monotone transformations and limit laws*, 1973.
             ISBN 90 6196 083 5.

MCT 46   R.P. VAN DE RIET, *ABC ALGOL, A portable language for formula manipu-*
             lation systems, part 1: The language, 1973. ISBN 90 6196 084 3.

MCT 47   R.P. VAN DE RIET, *ABC ALGOL, A portable language for formula manipu-*
             lation systems, part 2: The compiler, 1973. ISBN 90 6196 085 1.

MCT 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. ISBN 90 6196 086 X.

MCT 49   H. KOK, *Connected orderable spaces*, 1974. ISBN 90 6196 088 6.

MCT 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. ISBN 90 6196 089 4.

MCT 51   A. HORDIJK, *Dynamic programming and Markov potential theory*, 1974.
             ISBN 90 6196 095 9.

MCT 52   P.C. BAAYEN (ed.), *Topological structures*, 1974. ISBN 90 6196 096 7.

MCT 53   M.J. FABER, *Metrizability in generalized ordered spaces*, 1974.
             ISBN 90 6196 097 5.

MCT 54   H.A. LAUWERIER, *Asymptotic analysis, part 1*, 1974. ISBN 90 6196 098 3.

MCT 55   M. HALL JR. & J.H. VAN LINT (Eds), *Combinatorics, part 1: Theory*
             of designs, finite geometry and coding theory, 1974.
             ISBN 90 6196 099 1.

MCT 56   M. HALL JR. & J.H. VAN LINT (Eds), *Combinatorics, part 2: graph*
             theory, foundations, partitions and combinatorial geometry,
             1974. ISBN 90 6196 100 9.

MCT 57   M. HALL JR. & J.H. VAN LINT (Eds), *Combinatorics, part 3: Combina-*
             torial group theory, 1974. ISBN 90 6196 101 7.

MCT 58   W. ALBERS, *Asymptotic expansions and the deficiency concept in sta-*
             tistics, 1975. ISBN 90 6196 102 5.

MCT 59   J.L. MIJNHEER, *Sample path properties of stable processes*, 1975.
             ISBN 90 6196 107 6.

MCT 60   F. GÖBEL, *Queueing models involving buffers*, 1975. ISBN 90 6196 108 4.

* MCT 61   P. VAN EMDE BOAS, *Abstract resource-bound classes, part 1*.
             ISBN 90 6196 109 2.

* MCT 62   P. VAN EMDE BOAS, *Abstract resource-bound classes, part 2*.
             ISBN 90 6196 110 6.

MCT 63   J.W. DE BAKKER (ed.), *Foundations of computer science*, 1975.
             ISBN 90 6196 111 4.

MCT 64   W.J. DE SCHIPPER, *Symmetric closed categories*, 1975. ISBN 90 6196
             112 2.

MCT 65   J. DE VRIES, *Topological transformation groups 1 A categorical ap-*
             proach, 1975. ISBN 90 6196 113 0.

MCT 66   H.G.J. PIJLS, *Locally convex algebras in spectral theory and eigen-*
             function expansions, 1976. ISBN 90 6196 114 9.

* MCT 67 H.A. LAUWERIER, *Asymptotic analysis, part 2.*
          ISBN 90 6196 119 X.

  MCT 68 P.P.N. DE GROEN, *Singularly perturbed differential operators of
          second order,* 1976. ISBN 90 6196 120 3.

  MCT 69 J.K. LENSTRA, *Sequencing by enumerative methods,* 1977.
          ISBN 90 6196 125 4.

  MCT 70 W.P. DE ROEVER JR., *Recursive program schemes: semantics and proof
          theory,* 1976. ISBN 90 6196 127 0.

  MCT 71 J.A.E.E. VAN NUNEN, *Contracting Markov decision processes,* 1976.
          ISBN 90 6196 129 7.

  MCT 72 J.K.M. JANSEN, *Simple periodic and nonperiodic Lamé functions and
          their applications in the theory of conical waveguides,*1977.
          ISBN 90 6196 130 0.

  MCT 73 D.M.R. LEIVANT, *Absoluteness of intuitionistic logic,* 1979.
          ISBN 90 6196 122 x.

  MCT 74 H.J.J. TE RIELE, *A theoretical and computational study of general-
          ized aliquot sequences,* 1976. ISBN 90 6196 131 9.

  MCT 75 A.E. BROUWER, *Treelike spaces and related connected topological
          spaces,* 1977. ISBN 90 6196 132 7.

  MCT 76 M. REM, *Associons and the closure statement,* 1976. ISBN 90 6196 135 1.

  MCT 77 W.C.M. KALLENBERG, *Asymptotic optimality of likelihood ratio tests in
          exponential families,* 1977 ISBN 90 6196 134 3.

  MCT 78 E. DE JONGE, A.C.M. VAN ROOIJ, *Introduction to Riesz spaces,* 1977.
          ISBN 90 6196 133 5.

  MCT 79 M.C.A. VAN ZUIJLEN, *Empirical distributions and rankstatistics,* 1977.
          ISBN 90 6196 145 9.

  MCT 80 P.W. HEMKER, *A numerical study of stiff two-point boundary problems,*
          1977. ISBN 90 6196 146 7.

  MCT 81 K.R. APT & J.W. DE BAKKER (Eds), *Foundations of computer science II,*
          part 1, 1976. ISBN 90 6196 140 8.

  MCT 82 K.R. APT & J.W. DE BAKKER (Eds), *Foundations of computer science II,*
          part 2, 1976. ISBN 90 6196 141 6.

  MCT 83 L.S. VAN BENTEM JUTTING, *Checking Landau's "Grundlagen" in the
          AUTOMATH system,* 1979 ISBN 90 6196 147 5.

  MCT 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.
          ISBN 90 6196 148 3.

  MCT 85 J. VAN MILL, *Supercompactness and Wallman spaces,* 1977.
          ISBN 90 6196 151 3.

  MCT 86 S.G. VAN DER MEULEN & M. VELDHORST, *Torrix I,* 1978.
          ISBN 90 6196 152 1.

* MCT 87 S.G. VAN DER MEULEN & M. VELDHORST, *Torrix II,*
          ISBN 90 6196 153 x.

  MCT 88 A. SCHRIJVER, *Matroids and linking systems,* 1977.
          ISBN 90 6196 154 8.

MCT 89   J.W. DE ROEVER, *Complex Fourier transformation and analytic functionals with unbounded carriers*, 1978. ISBN 90 6196 155 6.

* MCT 90   L.P.J. GROENEWEGEN, *Characterization of optimal strategies in dynamic games*, . ISBN 90 6196 156 4.

MCT 91   J.M. GEYSEL, *Transcendence in fields of positive characteristic*, 1979. ISBN 90 6196 157 2.

MCT 92   P.J. WEEDA, *Finite generalized Markov programming*, 1979. ISBN 90 6196 158 0.

MCT 93   H.C. TIJMS (ed.) & J. WESSELS (ed.), *Markov decision theory*, 1977. ISBN 90 6196 160 2.

MCT 94   A. BIJLSMA, *Simultaneous approximations in transcendental number theory*, 1978 . ISBN 90 6196 162 9.

MCT 95   K.M. VAN HEE, *Bayesian control of Markov chains*, 1978. ISBN 90 6196 163 7.

* MCT 96   P.M.B. VITÁNYI, *Lindenmayer systems: structure, languages, and growth functions*, . ISBN 90 6196 164 5.

* MCT 97   A. FEDERGRUEN, *Markovian control problems; functional equations and algorithms*, . ISBN 90 6196 165 3.

MCT 98   R. GEEL, *Singular perturbations of hyperbolic type*, 1978. ISBN 90 6196 166 1

MCT 99   J.K. LENSTRA, A.H.G. RINNOOY KAN & P. VAN EMDE BOAS, *Interfaces between computer science and operations research*, 1978. ISBN 90 6196 170 X.

MCT 100 P.C. BAAYEN, D. VAN DULST & J. OOSTERHOFF (Eds), *Proceedings bicentennial congress of the Wiskundig Genootschap, part 1*, 1979. ISBN 90 6196 168 8.

MCT 101 P.C. BAAYEN, D. VAN DULST & J. OOSTERHOFF (Eds), *Proceedings bicentennial congress of the Wiskundig Genootschap, part 2*, 1979. ISBN 90 9196 169 6.

MCT 102 D. VAN DULST, *Reflexive and superreflexive Banach spaces*, 1978. ISBN 90 6196 171 8.

MCT 103 K. VAN HARN, *Classifying infinitely divisible distributions by functional equations*, 1978 . ISBN 90 6196 172 6.

MCT 104 J.M. VAN WOUWE, *Go-spaces and generalizations of metrizability*, 1979. ISBN 90 6196 173 4.

* MCT 105 R. HELMERS, *Edgeworth expansions for linear combinations of order statistics*, . ISBN 90 6196 174 2.

MCT 106 A. SCHRIJVER (Ed.), *Packing and covering in combinatorics*, 1979. ISBN 90 6196 180 7.

MCT 107 C. DEN HEIJER, *The numerical solution of nonlinear operator equations by imbedding methods*, 1979. ISBN 90 6196 175 0.

MCT 108 J.W. DE BAKKER & J. VAN LEEUWEN (Eds), *Foundations of computer science III, part 1*, 1979. ISBN 90 6196 176 9.

MCT 109 J.W. DE BAKKER & J. VAN LEEUWEN (Eds), *Foundations of computer science III*, part 2, 1979. ISBN 90 6196 177 7.

MCT 110 J.C. VAN VLIET, *ALGOL 68 transput*, part I: *Historical Review and Discussion of the Implementation Model*, 1979. ISBN 90 6196 178 5.

MCT 111 J.C. VAN VLIET, *ALGOL 68 transput*, part II: *An implementation model*, 1979. ISBN 90 6196 179 3.

MCT 112 H.C.P. BERBEE, *Random walks with stationary increments and Renewal theory*, 1979. ISBN 90 6196 182 3.

MCT 113 T.A.B. SNIJDERS, *Asymptotic optimality theory for testing problems with restricted alternatives*, 1979. ISBN 90 6196 183 1.

MCT 114 A.J.E.M. JANSSEN, *Application of the Wigner distribution to harmonic analysis of generalized stochastic processes*, 1979. ISBN 90 6196 184 x.

MCT 115 P.C. BAAYEN & J. VAN MILL (Eds), *Topological Structures II*, part 1, 1979. ISBN 90 6196 185 5.

MCT 116 P.C. BAAYEN & J. VAN MILL (Eds), *Topological Structures II*, part 2, 1979. ISBN 90 6196 186 6.

MCT 117 P.J.M. KALLENBERG, *Branching processes with continuous state space*, 1979. ISBN 90 6196 188 2.

MCT 118 P. GROENEBOOM, *Large deviations and Asymptotic efficiencies*, 1980. ISBN 90 6196 190 4

AN ASTERISK BEFORE THE NUMBER MEANS "TO APPEAR"