stichting
mathematisch
centrum

AFDELING MATHEMATISCHE STATISTIEK SW 49/76 DECEMBER
(DEPARTMENT OF MATHEMATICAL STATISTICS)

P. GROENEBOOM, J. OOSTERHOFF & F.H. RUYMGAART

LARGE DEVIATION THEOREMS FOR EMPIRICAL PROBABILITY MEASURES

Preprint

2e boerhaavestraat 49 amsterdam

Printed at the Mathematical Centre, 49, 2e Boerhaavestraat, Amsterdam.

The Mathematical Centre, founded the 11-th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O).

AMS(MOS) subject classification scheme (1970): 60F10

Large deviation theorems for empirical probability measures *)

by

P. Groeneboom, J. Oosterhoff & F.H. Ruymgaart

ABSTRACT

Some theorems on first-order asymptotic behavior of probabilities of large deviations of empirical probability measures are proved. These theorems extend previous results due to Borovkov, Hoadley and Stone. A multivariate analogue of Chernoff's theorem and a large deviation result for trimmed means are obtained as particular applications of the general theory.

KEY WORDS & PHRASES: *large deviations, empirical probability measures, Kullback-Leibler information, trimmed means.*

---

*) This report will be submitted for publication elsewhere.

1. INTRODUCTION

Let S be a polish (separable complete metric) 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 $\mathcal{B}$. For $P, Q \in \Lambda$ the Kullback-Leibler information number $K(Q,P)$ is defined by

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

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

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

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

Throughout this paper $X_1, X_2, \ldots$ is a sequence of i.i.d. random variables taking values in S according to a pm $P \in \Lambda$. For each positive integer 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

$$(1.1) \qquad d(Q,R) = \sup_{x\in\mathbb{R}} |Q([-\infty,x])-R([-\infty,x])|, \qquad Q,R \in \Lambda_1.$$

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,*

$$\text{(1.2)} \qquad \lim_{n\to\infty} n^{-1} \log \Pr\{T(\hat{P}_n) \geq r + u_n\} = -K(\Omega_r,P).$$

In section 3 it will be shown that Hoadley's theorem can be generalized in three different directions simultaneously:

(i) the set of pms Λ_1 may be replaced by the set Λ of pms on a polish 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 $S = \mathbb{R}$) and the range space of T may be different from $\mathbb{R}$

(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 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* $K(\mathrm{cl}_\rho(\Omega),P) = K(\Omega,P)$ *(where* cl_ρ *denotes closure in the topology* ρ*),*

$$\text{(1.3)} \qquad \lim_{n\to\infty} n^{-1} \log \Pr\{\hat{P}_n \in \Omega\} = -K(\Omega,P).$$

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 rather deep methods of Fourier analysis of random walks in Borovkov (1962) for which generalization to more general pms seems to be difficult.

In this paper 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 $Q_n \underset{\tau}{\to} Q$, 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 $\mathrm{cl}_\tau(\Omega)$ and $\mathrm{int}_\tau(\Omega)$, respectively.

With this notation we shall prove (Theorem 3.1)

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

$$(1.4) \qquad K(\mathrm{int}_\tau(\Omega),P) = K(\mathrm{cl}_\tau(\Omega),P).$$

Then (1.3) *holds true.*

Since in the particular case S = $\mathbb{R}$ the topology τ is finer than ρ (Lemma 2.1), any ρ-continuous functional T : $\Lambda_1 \to \mathbb{R}$ is a fortiori τ-continuous. Hence our results on τ-continuous functionals T imply the corresponding (weaker) results for ρ-continuous functionals. In fact, by this line of argument the generalized form of Hoadley's theorem mentioned above easily follows from Theorem 3.1.

After some crucial lemma's in section 2 the basic theorems are obtained in section 3. The theory includes theorems of Borovkov, Hoadley, Stone and Sethuraman as particular cases and thus provides a unified approach to these results which were obtained by rather different methods. In section 4 a large deviation result for linear functions of empirical pms is proved. This result yields a multivariate analogue of Chernoff's (1952) celebrated large deviation theorem as a particular case. Finally we prove in section 5 a large deviation theorem for a class of linear combinations of order statistics (L-estimators). This leads to a large deviation theorem for trimmed means under minimal conditions.

Although for some results (e.g. Lemma 3.1) the assumption that S is a polish space and that $\mathcal{B}$ is the Borel σ-field of subsets of S is unnecessarily restrictive, we have not tried to relax this assumption since it seems to be satisfied in most applications.

Recently Sievers (1976) proved (1.3) under conditions essentially different from ours. Since Sievers' methods are based on a likelihood ratio approximation, his results cannot be fitted into our framework.

2. PRELIMINARIES

In this section some notation is introduced and a few preliminary results are proved which will play an essential role in the subsequent sections. By a *partition* $\mathcal{P}$ of S is meant a *finite* partition of S consisting of Borel sets. Such partitions are the starting point of the multinomial approximation on which the proof of Lemma 3.1 in section 3 is based. For $P,Q \in \Lambda$ and a partition $\mathcal{P} = \{B_1,\ldots,B_m\}$ of S define

$$(2.1) \qquad K_{\mathcal{P}}(Q,P) = \sum_{j=1}^{m} Q(B_j)\log\{Q(B_j)/P(B_j)\}$$

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

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

Without explicit reference the relation

$$(2.2) \qquad K(Q,P) = \sup\{K_{\mathcal{P}}(Q,P)\colon \mathcal{P} \text{ is a partition of } S\}$$

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

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

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

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

LEMMA 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 $\mathcal{P} = \{B_1,\ldots,B_m\}$ 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 $d_{\mathcal{P}}(Q,R) < \frac{1}{2}\varepsilon/m$, 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 $Q_i(B_{i,j}) < \frac{1}{2}\varepsilon$. Let $\mathcal{P}$ 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

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

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

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 $\bar{\mathbb{R}}$ is the usual topology generated by the sets $[-\infty,x)$, $(x,\infty]$, $x \in \mathbb{R}$.

<u>LEMMA 2.2</u>. *Let* $P \in \Lambda$. *Then the function* $Q \to K(Q,P)$, $Q \in \Lambda$, *is* τ-*lower semicontinuous*.

<u>PROOF</u>. Let $P,Q \in \Lambda$ and let c be an arbitrary real number such that $c < K(Q,P)$. By (2.2) there exists a partition $\mathcal{P}$ of S such that $K_{\mathcal{P}}(Q,P) > c$. Clearly there exists a $\delta > 0$ such that

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

proving the lemma. $\square$

A collection Γ of pms in Λ is called *uniformly absolutely continuous* (u.a.c.) 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 2.3. *Let* $P \in \Lambda$ *be an arbitrary pm and let* $\Gamma = \{Q \in \Lambda\colon K(Q,P) \leq M\}$ *for some finite* $M > 0$. *Then*

(a) Γ *is* u.a.c. *with respect to* P

(b) *if a sequence* $\{Q_n\}$ *in* Γ *converges weakly to a* $Q \in \Lambda$, *then* $Q_n \underset{\tau}{\to} Q$

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

PROOF.

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

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

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

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

where $q = dQ/dP$ (note that the inequality $x \log x \geq -e^{-1}$ provides an upper bound $M + e^{-1}$ for the integral $\int_C q \log q dP$ for any set $C \in \mathcal{B}$). It follows that Γ is u.a.c. with respect to P.

(b) Suppose $\{Q_n\}$ is a sequence in Γ converging weakly to a $Q \in \Lambda$. Let $\varepsilon > 0$. By (a) there exists $\delta > 0$ such that for each $n \in \mathbb{N}$ and $B \in \mathcal{B}$: $P(B) < \delta \Rightarrow Q_n(B) < \varepsilon$. Fix $B \in \mathcal{B}$. Since $\mathcal{B}$ is the σ-field of Borel sets of a metric space, each pm on $(S,\mathcal{B})$ is regular (cf. Billingsley (1968) Theorem 1.1). Hence there exists an open set U and a closed set K satisfying $K \subset B \subset U$ and $P(U\backslash K) < \delta$. This implies $\sup_{n\in\mathbb{N}} Q_n(U\backslash K) \leq \varepsilon$ and hence, by the weak convergence of $\{Q_n\}$ to Q,

$$\limsup_{n\to\infty} Q_n(B) \leq \limsup_{n\to\infty} Q_n(K) + \limsup_{n\to\infty} Q_n(B\backslash K)$$

$$\leq Q(K) + \varepsilon \leq Q(B) + \varepsilon$$

and

$$\liminf_{n\to\infty} Q_n(B) \geq \liminf_{n\to\infty} Q_n(U) - \limsup_{n\to\infty} Q_n(U/B) \geq$$

$$\geq Q(U) - \varepsilon \geq Q(\bar{B}) - \varepsilon \geq Q(B) - \varepsilon.$$

These inequalities imply $\lim_{n\to\infty} Q_n(B) = Q(B)$.

(c) By Theorem 2.6 of Gänssler (1971) the notions "compact" and "sequentially compact" coincide for the topology τ. Let $\{Q_n\}$ be a sequence in Γ. This sequence is tight because Γ is u.a.c. with respect to P and because P is tight since S is a polish space (cf. Billingsley (1968) Theorem 1.4). Let $\{Q_{n_k}\}$ be a subsequence of $\{Q_n\}$ converging weakly to $Q \in \Lambda$. By (b) $Q_{n_k} \to_\tau Q$ and by Lemma 2.2 $K(Q,P) \leq \liminf_{k\to\infty} K(Q_{n_k},P) \leq M$ proving sequential compactness. □

Lemma 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 $P, Q \in \Lambda$, let $\mathcal{P} = \{B_1,\ldots,B_m\}$ be a partition of S and let $K_{\mathcal{P}}(Q,P) < \infty$. Then the $\mathcal{P}_P$-linear pm Q' corresponding to Q is defined by

$$(2.3) \qquad Q'(B\cap B_i) = \begin{cases} P(B\cap B_i)Q(B_i)/P(B_i) & \text{if } P(B_i) > 0 \\ 0 & \text{otherwise.} \end{cases}$$

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 $\mathcal{P}$-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 (2.2) and plays a crucial role in the next sections.

Lemma 2.4. *Let* $P \in \Lambda$ *and* $\Omega \subset \Lambda$ *satisfy*

$$(2.4) \qquad K(\mathrm{cl}_\tau(\Omega),P) = K(\Omega,P).$$

Then

$$(2.5) \qquad K(\Omega,P) = \sup\{K_{\mathcal{P}}(\Omega,P)\colon \mathcal{P} \text{ is a partition of } S\}.$$

PROOF. Let $\alpha = \sup\{K_{\mathcal{P}}(\Omega,P)\colon \mathcal{P}$ is a partition of $S\}$ and suppose (2.5) does not hold, i.e. there exists an $\eta > 0$ such that $\alpha + \eta < K(\Omega,P)$ (see (2.2)). Put $\Gamma = \{Q \in \Lambda\colon K(Q,P) \le \alpha + \eta\}$. The set of all (finite) partitions $\mathcal{P}$, ordered by $\mathcal{P} > \mathcal{R}$ iff $\mathcal{P}$ is finer than $\mathcal{R}$, is a directed set. Choose for each partition $\mathcal{P}$ a pm $Q_{\mathcal{P}} \in \Omega$ satisfying $K_{\mathcal{P}}(Q_{\mathcal{P}},P) \le \alpha + \eta$. Let $Q'_{\mathcal{P}}$ be the $P_{\mathcal{P}}$-linear pm corresponding to $Q_{\mathcal{P}}$. Then

$$K(Q'_{\mathcal{P}},P) = K_{\mathcal{P}}(Q_{\mathcal{P}},P) \le \alpha + \eta$$

and hence $Q'_{\mathcal{P}} \in \Gamma$ for each partition $\mathcal{P}$. Since Γ is compact in the topology τ by Lemma 2.3, there exists a $\bar{Q} \in \Gamma$ such that $\bar{Q}$ is a cluster point of the net $\mathcal{N} = \{Q'_{\mathcal{P}}\ \colon \mathcal{P}$ is a partition of $S\}$.

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

$$Q_{\mathcal{T}}(B) = \sum_{A\in\mathcal{T},A\subset B} Q_{\mathcal{T}}(A) = \sum_{A\in\mathcal{T},A\subset B} Q'_{\mathcal{T}}(A) = Q'_{\mathcal{T}}(B).$$

Hence $d_{\mathcal{P}}(Q_{\mathcal{T}},\bar{Q}) = d_{\mathcal{P}}(Q'_{\mathcal{T}},\bar{Q}) < \varepsilon$, implying that $\bar{Q}$ is also a cluster point of the net $\{Q_{\mathcal{P}}\colon \mathcal{P}$ is a partition of $S\}$. Since $Q_{\mathcal{P}} \in \Omega$ for each $\mathcal{P}$, $\bar{Q} \in \mathrm{cl}_\tau(\Omega)$. However, $\bar{Q} \in \Gamma \Rightarrow K(\bar{Q},P) \le \alpha + \eta < K(\Omega,P)$ in contradiction to (2.4) and (2.5) follows. □

REMARK 2.1 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 (2.4) in Lemma 2.4 connot be replaced by $K(scl_\tau(\Omega),P) = K(\Omega,P)$. Let Ω be the set of all pms on $\mathbb{R}$ with countable support and let P be a non-atomic pm on $\mathbb{R}$. Then $\sup\{K_{\mathcal{P}}(\Omega,P): \mathcal{P}$ 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.

REMARK 2.2 It can easily be shown by counter examples that condition (2.4) is not necessary for (2.5) even if $K(\Omega,P) < \infty$.

3. BASIC RESULTS

Our large deviation results concerning probabilities $\Pr\{\hat{P}_n \in \Omega\}$ have as a starting point Lemma 3.1 which exploits multinomial approximations to the distributions of the empirical pms $\{\hat{P}_n\}$ induced by the sequence $X_1, X_2, \ldots$ Without explicit reference it will always be assumed that the set $\Omega \subset \Lambda$ is such that $\Pr\{\hat{P}_n \in \Omega\}$ is well defined for each $n \in \mathbb{N}$. This is certainly true if the intersection of Ω and the set of pms with finite support is measurable with respect to the Borel σ-field generated by the (relative) topology of weak convergence. It is easily seen that the lemma remains valid for arbitrary sets S and arbitrary σ-fields $\mathcal{B}$ containing all singletons.

LEMMA 3.1 *Let* $P \in \Lambda$ *and let* Ω *be a subset of* Λ *satifying*

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

(B) $K(\Omega,P) = K(int_\tau(\Omega),P)$.

Then

$$(3.1) \qquad \lim_{n\to\infty} n^{-1} \log \Pr\{\hat{P}_n \in \Omega\} = -K(\Omega,P).$$

PROOF. To prove the lemma it is first shown that condition (A) implies

$$(3.2) \qquad \limsup_{n\to\infty} n^{-1} \log \Pr\{\hat{P}_n \in \Omega\} \le -K(\Omega,P).$$

Let $c < K(\Omega,P)$. By condition (A) there exists a partition $\mathcal{P}$ of S such that $K_{\mathcal{P}}(\Omega,P) > c$. Let $\mathcal{P} = \{B_1,\ldots,B_m\}$ and let $p_j = P(B_j)$, $1 \le j \le m$. Then

$$\begin{aligned}
\Pr\{\hat{P}_n \in \Omega\} &\le \Pr\{K_{\mathcal{P}}(\hat{P}_n,P) \ge K_{\mathcal{P}}(\Omega,P)\} \\
&= \textstyle\sum^* n!/\{(nz_{n,1})!\ldots(nz_{n,m})!\}\cdot \prod_{i=1}^{m} p_i^{nz_{n,i}} \\
&= \textstyle\sum^* n!\{\prod_{i=1}^{m} (nz_{n,i})!\}^{-1} \prod_{i=1}^{m} z_{n,i}^{nz_{n,i}} \cdot \\
&\qquad \cdot \exp\{-n\textstyle\sum_{i=1}^{m} z_{n,i}\log(z_{n,i}/p_i)\},
\end{aligned}$$

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

$$\sum_{i=1}^{m} z_{n,i} = 1, \quad z_{n,i} \ge 0,\ nz_{n,i} \in \mathbb{Z} \quad (1\le i\le m)$$

and

$$\sum_{i=1}^{m} z_{n,i}\log(z_{n,i}/p_i) \ge K_{\mathcal{P}}(\Omega,P).$$

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

$$\binom{n+m-1}{m-1} = \exp(\mathcal{O}(\log n)), \qquad \text{as } n \to \infty.$$

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

$$n!/\{(nz_{n,1})!\ldots(nz_{n,m})!\} \le \exp\{-n\sum_{i=1}^{m} z_{n,i} \log z_{n,i} + \mathcal{O}(\log n)\}.$$

Hence $\Pr\{\hat{P}_n \in \Omega\} \le \exp\{-nK_{\mathcal{P}}(\Omega,P) + \mathcal{O}(\log n)\}$, implying

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

as $n \to \infty$. Since $c < K(\Omega,P)$ is arbitrary, (3.2) follows.

Conversely we prove that condition (B) implies

$$\liminf_{n\to\infty} n^{-1}\log \Pr\{\hat{P}_n \in \Omega\} \geq -K(\Omega,P). \tag{3.3}$$

Assume $K(\Omega,P) < \infty$ since otherwise (3.3) is trivial. Fix $\varepsilon > 0$. In view of condition (B) $\mathrm{int}_\tau(\Omega)$ is not empty and a pm $Q \in \mathrm{int}_\tau(\Omega)$ exists satisfying $K(Q,P) < K(\Omega,P) + \frac{1}{2}\varepsilon$. Since $Q \in \mathrm{int}_\tau(\Omega)$, a partition $\mathcal{P} = \{B_1,\ldots,B_m\}$ of S and $\delta > 0$ can be found such that $\{R \in \Lambda : d_{\mathcal{P}}(R,Q) < \delta\} \subset \Omega$. It follows that for all sufficiently large n there exist pms $Q_n \in \Lambda$ satisfying

(i) $nQ_n(B_i) \in \mathbb{Z}$, $1 \leq i \leq m$
(ii) $d_{\mathcal{P}}(Q_n,Q) < \delta$, hence $Q_n \in \Omega$ and $\{R \in \Lambda : d_{\mathcal{P}}(R,Q_n) = 0\} \subset \Omega$
(iii) $K_{\mathcal{P}}(Q_n,P) < K_{\mathcal{P}}(Q,P) + \frac{1}{2}\varepsilon \leq K(Q,P) + \frac{1}{2}\varepsilon < K(\Omega,P) + \varepsilon$.

Put $z_{n,i} = Q_n(B_i)$, $1 \leq i \leq m$. Then for all sufficiently large n

$$\begin{aligned}\Pr\{\hat{P}_n \in \Omega\} &\geq \Pr\{d_{\mathcal{P}}(\hat{P}_n,Q_n) = 0\} \\ &= n!/\{(nz_{n,1})!\ldots(nz_{n,m})!\}.\prod_{i=1}^{m} (P(B_i))^{nz_{n,i}},\end{aligned}$$

where $\sum_{i=1}^m z_{n,i} = 1$, $z_{n,i} \geq 0$, $nz_{n,i} \in \mathbb{Z}$ $(1\leq i\leq m)$ and

$$\sum_{i=1}^{m} z_{n,i} \log\{z_{n,i}/P(B_i)\} < K(\Omega,P) + \varepsilon.$$

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

$$\Pr\{\hat{P}_n \in \Omega\} \geq \exp\{-n(K(\Omega,P) + \varepsilon + o(1))\}$$

and (3.3) easily follows, which completes the proof. □

Stone (1974) proves (3.1) under the conditions (in our notation)

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

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

(C2) $K_{\mathcal{P}}(\Omega,P) \leq K_{\mathcal{P}}(Q,P) < K_{\mathcal{P}}(\Omega,P) + \varepsilon$
(C3) $\{R \in \Lambda : d_{\mathcal{P}}(R,Q) < \delta\} \subset \Omega$.

It turns out that if $K(\Omega,P) < \infty$ these conditions are equivalent to conditions (A) and (B) of our Lemma 3.1, implying that Stone's theorem is in fact equivalent to Lemma 3.1 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 $Q \in \mathrm{int}_\tau(\Omega)$ exists satisfying $K(Q,P) < K(\Omega,P) + \frac{1}{2}\varepsilon$. Since $Q \in \mathrm{int}_\tau(\Omega)$, there exist a partition $\mathcal{T}$ and $\delta > 0$ such that $\{R \in \Lambda : d_{\mathcal{T}}(R,Q) < \delta\} \subset \Omega$. By (A) there exists a partition $\mathcal{P}$ which is finer than $\mathcal{T}$ and satisfies $K(\Omega,P) < K_{\mathcal{P}}(\Omega,P) + \frac{1}{2}\varepsilon$ (note that $K_{\mathcal{T}}(R,P) \le K_{\mathcal{P}}(R,P)$ for each pm R if $\mathcal{P}$ is finer than $\mathcal{T}$). Hence

$$K_{\mathcal{P}}(\Omega,P) \le K_{\mathcal{P}}(Q,P) \le K(Q,P) < K(\Omega,P) + \tfrac{1}{2}\varepsilon < K_{\mathcal{P}}(\Omega,P) + \varepsilon.$$

Moreover, for small enough $\delta' > 0$ the implication $R \in \Lambda$, $d_{\mathcal{P}}(R,Q) < \delta' \Rightarrow$ $\Rightarrow d_{\mathcal{T}}(R,Q) < \delta$ 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 $\mathcal{P}$ of S and $\delta > 0$ satisfy (C2) and (C3) for this ε. Let Q' be the $\mathcal{P}_P$-linear pm corresponding to Q (see (2.3)). Then (C3) implies $Q' \in \mathrm{int}_\tau(\Omega)$ and (C2) yields

$$K(Q',P) = K_{\mathcal{P}}(Q',P) = K_{\mathcal{P}}(Q,P) < K_{\mathcal{P}}(\Omega,P) + \varepsilon \le K(\Omega,P) + \varepsilon.$$

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

The present method of proof of Lemma 3.1 is well suited to prove (3.1) under weaker conditions. It can for example be shown by an elaboration of the proof that Sanov's (1957) condition that Ω be F-distinguishable is indeed sufficient for (3.1). (Some obscure points in Sanov's (1957) paper have raised doubt as to the validity of his Theorem 11, cf. Hoadley (1967), Bahadur (1971).)

Combining Lemma 2.4 and Lemma 3.1 we have

THEOREM 3.1 *Let* $P \in \Lambda$ *and let* Ω *be a subset of* Λ *satisfying*

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

Then (3.1) *holds.*

Borovkov has shown (see (31) in Borovkov (1967)) that (3.1) holds if P is a non-atomic pm on $\mathbb{R}$, Ω is a ρ-open set and $K(\Omega,P) = K(\mathrm{cl}_\rho(\Omega),P)$. This is a particular case of Theorem 3.1 in view of Lemma 2.1.

REMARK 3.1 Suppose $B \subset S$ is a closed 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 3.1 remains valid if (3.4) is replaced by the weaker condition

$$K(\mathrm{int}_{\tau_B}(\Omega \cap \Lambda_B),P) = K(\mathrm{cl}_{\tau_B}(\Omega \cap \Lambda_B),P).$$

This result is an immediate consequence of Theorem 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 $\Pr\{\hat{P}_n \in \Omega\} = \Pr\{\hat{P}_n \in \Omega \cap \Lambda_B\}$). This closed set B may even be replaced by an arbitrary set $B \in \mathcal{B}$ as an inspection of the proofs of Lemmas 2.4 and 3.1 shows; however, this result will not be used in the sequel.

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 sufficient condition for the existence of such a pm Q is given in the next lemma.

LEMMA 3.2 *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 equality. Let $\eta > 0$. Because Ω is τ-closed the set $\Omega \cap \{Q \in \Lambda : K(Q,P) \le K(\Omega,P) + \eta\}$ is compact by Lemma 2.3. By Lemma 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. □

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 3.1 by considering sets Ω induced by an extended real-valued function $T: \Lambda \to \overline{\mathbb{R}}$. For a fixed function $T: \Lambda \to \overline{\mathbb{R}}$, let

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

We first prove a technical lemma.

LEMMA 3.3 *Let* $P \in \Lambda$ *and let* $T: \Lambda \to \overline{\mathbb{R}}$ *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 $\kappa: \mathbb{R} \to \overline{\mathbb{R}}$ 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 3.2 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 $K(Q,P) \leq M\}$ 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 2.2 and 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 $K(Q,P) \leq \liminf_{j\to\infty} K(Q_{m_j},P) < \infty$. 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 $\kappa(r) \leq K(Q,P) \leq \lim_{j\to\infty} K(Q_{m_j},P) = \lim_{m\to\infty} \kappa(r_m) \leq \kappa(r)$. 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 $\kappa(r) = \infty$. □

THEOREM 3.2 *Let* P *be a pm in* Λ *and let* $T: \Lambda \to \overline{\mathbb{R}}$ *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* $\lim_{n\to\infty} u_n = 0$,

$$\lim_{n\to\infty} n^{-1} \log \Pr\{T(\hat{P}_n) \geq r + u_n\} = -K(\Omega_r,P). \tag{3.6}$$

(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 3.3 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 $\kappa(r) - \varepsilon < \kappa(r-\delta) \leq \kappa(r) \leq \kappa(r+\delta) < \kappa(r) + \varepsilon$, where κ is continuous at $r - \delta$ and $r + \delta$.

The continuity of T at each $Q \in \Gamma$ implies $cl_\tau(\Omega_t) \cap \Gamma = \Omega_t \cap \Gamma$. Hence

$$K(cl_\tau(\Omega_t),P) = K(cl_\tau(\Omega_t) \cap \Gamma,P) = K(\Omega_t \cap \Gamma,P) = K(\Omega_t,P).$$

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

$$K(\Omega_t,P) = K(\Omega_t \cap \Gamma,P) = K(int_\tau(\Omega_t) \cap \Gamma,P) = K(int_\tau(\Omega_t),P),$$

since $\Gamma \cap \Omega_{t+\gamma} \subset \{Q \in \Gamma\colon T(Q) > t\} \subset \Gamma \cap int_\tau(\Omega_t)$ for each $\gamma > 0$. Hence by Theorem 3.1

$$\begin{aligned}
-\kappa(r) - \varepsilon < -\kappa(r+\delta) &= \lim_{n\to\infty} n^{-1} \log \Pr\{T(\hat{P}_n) \geq r + \delta\} \\
&\leq \liminf_{n\to\infty} n^{-1} \log \Pr\{T(\hat{P}_n) \geq r + u_n\} \\
&\leq \limsup_{n\to\infty} n^{-1} \log \Pr\{T(\hat{P}_n) \geq r + u_n\} \\
&\leq \limsup_{n\to\infty} n^{-1} \log \Pr\{T(\hat{P}_n) \geq r - \delta\} \\
&= -\kappa(r-\delta) < -\kappa(r) + \varepsilon.
\end{aligned}$$

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. □

<u>REMARK 3.2</u> Theorem 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.

<u>EXAMPLE 3.1</u>. Let $\mathcal{F}$ be a class of continuous $\mathbb{R}^d$ - valued functions defined on the polish space S and compact in the compact-open topology. Let $P \in \Lambda$ be a pm such that the one-dimensional marginals of Pf^{-1} are non-atomic for each $f \in \mathcal{F}$ and let $d(Qf^{-1},Rf^{-1})$ be the distance between Qf^{-1} and Rf^{-1} defined in Lemma 2.1.

Sethuraman (1964) proves that for each ε, $0 < \varepsilon < 1$,

$$\lim_{n\to\infty} n^{-1} \log \Pr\{\sup_{f\in\mathcal{F}} d(\hat{P}_n f^{-1},Pf^{-1}) \geq \varepsilon\} = -\kappa(\varepsilon), \tag{3.7}$$

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 $T(Q) = \sup_{f\in\mathcal{F}} d(Qf^{-1},Pf^{-1})$ is τ-continuous at each $Q \in \Lambda$ satisfying $K(Q,P) < \infty$ and hence that (3.7) follows from Theorem 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 $f_U \in \mathcal{F}$ can be found satisfying

$$d(Q_U f_U^{-1},Qf_U^{-1}) \geq \varepsilon \tag{3.8}$$

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

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 Q is tight on S 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 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 $\{f_U\colon U \in \mathcal{D}\}$ 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

$$
\begin{aligned}
d(Q_U f_U^{-1},Qf_U^{-1}) &\le \max\{Q(S\backslash K),Q_U(S\backslash K)\} + \\
&\quad + \sup_{x\in\mathbb{R}^d} |Q_U\{s \in K\colon f_U(s) \le x\}-Q\{s \in K\colon f_U(s) \le x\}| \\
&< \sup_{x\in\mathbb{R}^d} |Q_U\{s \in K\colon f(s) \le x\}-Q\{s \in K\colon f(s) \le x\}| + \tfrac{1}{2}\varepsilon \\
&< d(Q_U f^{-1},Qf^{-1}) + \tfrac{3}{4}\varepsilon < \varepsilon.
\end{aligned}
$$

This contradicts (3.8) and hence T is τ-continuous at Q. Let $\Omega_\varepsilon = \{Q \in \Lambda\colon T(Q) \ge \varepsilon\}$ 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 (3.7) follows from Theorem 3.2.

For one sample Theorem 1 in Hoadley (1967) is a particular case of our Theorem 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 proves a more general theorem in [11] 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 $N = \Sigma_{i=1}^k n_i$ and $\nu_i > 0$, $1 \le i \le k$. (We remark in passing that the condition $n_i/N - \nu_i = O(N^{-1}\log N)$ 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 $\Sigma_{i=1}^k \nu_i = 1$. 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 $\Omega \subset \Lambda^k$

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

and

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

By making small changes in the proofs of Theorems 3.1 and 3.2 one obtains the following corollaries.

COROLLARY 3.1. *Let* $P = (P_1, \ldots, P_k) \in \Lambda^k$ *and* $\Omega \subset \Lambda^k$ *satisfy*

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

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 3.2. *Let* $P = (P_1, \ldots, P_k) \in \Lambda^k$, *let* $T\colon \Lambda^k \to \overline{\mathbb{R}}$ *be continuous at each* $Q \in \Gamma = \{R \in \Lambda^k\colon I_\nu(R,P) < \infty\}$ *and let* $\Omega_t = \{Q \in \Lambda^k\colon T(Q) \ge t\}$, $t \in \mathbb{R}$. *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* $u_N \to 0$,

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

4. LINEAR FUNCTIONS OF EMPIRICAL PMS AND A MULTIVARIATE ANOLOGUE OF CHERNOFF'S THEOREM

Several important statistics are in fact linear functions of empirical pms. For example, if $S = \mathbb{R}$, the sample mean $n^{-1} \Sigma_{i=1}^n X_i$ may be written as $T(\hat{P}_n) = \int_{\mathbb{R}} x \, d\hat{P}_n(x)$, 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. $T(\alpha Q+(1-\alpha)R) = \alpha T(Q) + (1-\alpha)T(R)$, $0 \le \alpha \le 1$. 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 $Q_B(C) = Q(C \mid B)$, $C \in \mathcal{B}$. 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 4.1. *Let* $P \in \Lambda$ *and let* $B_1 \subset B_2 \subset \cdots$ *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 equality 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 $K(\Omega,P) \le \liminf_{m\to\infty} K(\Omega,P_m)$. To prove the lemma it still must be shown that conversely

$$\text{(4.1)} \qquad K(\Omega,P) \ge \limsup_{m\to\infty} K(\Omega,P_m).$$

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 (4.1). $\square$

THEOREM 4.1. *Let* $P \in \Lambda$, *let* E *be a real topological vector space and let* $B_1 \subset B_2 \subset \ldots$ *be an increasing sequence of closed subsets of* S *such that* $\lim_{m\to\infty} P(B_m) = 1$. *Let* $\Psi_m = \{Q \in \Lambda\colon Q(B_m) = 1\}$ *for* $m \in \mathbb{N}$ *and let* $\Lambda^* = \bigcup_{m=1}^{\infty} \Psi_m$. *Let* $T : \Lambda^* \to E$ *be a function whose restriction to* Ψ_m *is linear and 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*

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

PROOF. Assume without loss of generality that $P(B_1) > 0$. Let $P_m = P_{B_m}$, $m \in \mathbb{N}$. By Lemma 4.1 $K(T^{-1}(A^0),P) = \lim_{m\to\infty} K(T^{-1}(A^0),P_m)$. 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

$$(4.2) \qquad K(T^{-1}(A^0),P_m) = K(T^{-1}(\bar{A}),P_m) \qquad \text{for each } m \in \mathbb{N}.$$

Fix $\varepsilon > 0$ and $m \in \mathbb{N}$. There exists a pm $Q \in T^{-1}(\bar{A})$ which satisfies $K(Q,P_m) < K(T^{-1}(\bar{A}),P_m) + \varepsilon$. 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 $K(Q_\alpha,P_m) \le \alpha K(Q,P_m) + (1-\alpha)K(R,P_m)$, $\alpha \in (0,1)$, by the convexity of the mapping $Q' \to K(Q',P_m)$, $Q' \in \Lambda$. It follows that $K(T^{-1}(A^0),P_m) \le \lim_{\alpha\uparrow 1} K(Q_\alpha,P_m) < K(T^{-1}(\bar{A}),P_m) + \varepsilon$, proving (4.2).

Let $\Omega = T^{-1}(A)$, let $\Psi_m^* = \{Q \in \Psi_m\colon K(Q,P) < \infty\}$ 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 $\Psi_m^* \cap T^{-1}(\bar{A}) \supset \Psi_m^* \cap \mathrm{cl}_{\tau_m}(\Omega \cap \Psi_m) \supset$

$\Psi_m^* \cap \mathrm{int}_{\tau_m}(\Omega \cap \Psi_m) \supset \Psi_m^* \cap T^{-1}(A^0)$. Hence, by (4.2)

$$K(\mathrm{cl}_{\tau_m}(\Omega \cap \Psi_m),P_m) = K(\mathrm{int}_{\tau_m}(\Omega \cap \Psi_m),P_m), \quad \text{for each } m \in \mathbb{N}. \tag{4.3}$$

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 $k^{-1} \log \Pr\{T(\hat{P}_k) \in A\} \ge \gamma - \varepsilon$. 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

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

Hence for $m \ge m_0$

$$\begin{aligned} &\limsup_{n\to\infty} n^{-1} \log \Pr\{T(\hat{P}_n) \in A \mid B_m\} \\ &\ge \lim_{j\to\infty} (kj)^{-1} \log(\Pr\{T(\hat{P}_k) \in A \mid B_m\})^j \\ &= k^{-1} \log \Pr\{T(\hat{P}_k) \in A \mid B_m\} \ge \gamma - 2\varepsilon. \end{aligned} \tag{4.4}$$

The first inequality in (4.4) follows from the convexity of A, the linearity of T on Ψ_m and the property $\hat{P}_n = j^{-1} \Sigma_{i=1}^{j} \hat{P}_{k,i}$, where $n = jk$ and $\hat{P}_{k,i}$ is the empirical pm of the random variables $X_{(i-1)k+1},\dots,X_{ik}$, $1 \le i \le j$.

By (4.3), Theorem 3.1 and Remark 3.1

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

Lemma 4.1 and (4.4) now imply

$$\begin{aligned} \gamma - 2\varepsilon &\le \lim_{m\to\infty} \lim_{n\to\infty} n^{-1} \log \Pr\{T(\hat{P}_n) \in A \mid B_m\} \\ &= -\lim_{m\to\infty} K(T^{-1}(A),P_m) = -K(T^{-1}(A),P). \end{aligned}$$

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

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

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

Hence, by the first part of the proof and Lemma 4.1

$$\begin{aligned} &\liminf_{n\to\infty} n^{-1} \log \Pr\{T(\hat{P}_n) \in A\} \geq \\ &\geq \lim_{m\to\infty} [\liminf_{n\to\infty} n^{-1} \log \Pr\{T(\hat{P}_n) \in A \mid B_m\} + \log P(B_m)] \\ &= \lim_{m\to\infty} -K(T^{-1}(A),P_m) = -K(T^{-1}(A),P). \quad \square \end{aligned}$$

Next, we specialize S and E to $S = E = \mathbb{R}^d$. Let $\Lambda^* = \{Q \in \Lambda: Q$ has compact support$\}$. We define $T : \Lambda^* \to \mathbb{R}^d$ by $T(Q) = \int_{\mathbb{R}^d} x dQ(x)$, $Q \in \Lambda^*$. In Chernoff (1952) the following large deviation theorem was proved for the case $d = 1$:

$$\lim_{n\to\infty} n^{-1} \log \Pr\{T(\hat{P}_n) \geq r\} = -\sup_{t\geq 0} \{tr - \log \int_{\mathbb{R}} e^{tx} dP(x)\}$$

for any $r \in \mathbb{R}$ and $P \in \Lambda_1$ (as was noted in the beginning of this section, $T(\hat{P}_n)$ is equal to the sample mean $n^{-1} \Sigma_{i=1}^n X_i$). With the help of Theorem 4.1 we shall generalize this theorem to the case $d > 1$.

Our results are in a certain sense complementary to those of Sievers (1975), who gives sufficient conditions to reduce limits of the form $\lim_{n\to\infty} n^{-1} \log \Pr\{T_n \in B\}$, $B \subset \mathbb{R}^d$, $B \in \mathcal{B}$ to limits of the form $\lim_{n\to\infty} n^{-1} \log \Pr\{T_n^{(1)} * x_1,\ldots,T_n^{(d)} * x_d\}$, where the $*$'s are either $\geq$ or $\leq$ and $T_n = (T_n^{(1)},\ldots,T_n^{(d)})$ is a random variable taking values in $\mathbb{R}^d$. Here we shall give explicit expressions for the latter limits in the case that T_n is the sample mean.

We introduce the following notation. The i-th component of a vector $x \in \mathbb{R}^d$ is denoted by $x^{(i)}$ and the inner product of two vectors $x,y \in \mathbb{R}^d$ by $x'y$. The following ordering relations on $\mathbb{R}^d$ will be used: $x \geq y$ iff $x^{(i)} \geq y^{(i)}$ $(i \leq i \leq d)$ and $x > y$ iff $x^{(i)} > y^{(i)}$ $(1 \leq i \leq d)$. Furthermore $\mathbb{R}^d_+ = \{x \in \mathbb{R}^d : x \geq 0\}$. We denote the complement of a set $A \subset \mathbb{R}^d$ by A^c, its interior by A^0, its closure by $\bar{A}$ and its boundary by ∂A (always in the Euclidean topology). For $Q \in \Lambda^*$ the integral $\int_{\mathbb{R}^d} x dQ(x)$ denotes the vector of marginal means of Q. To avoid confusion, the letters α, β, γ, δ and ε will always denote real numbers.

For $r \in \mathbb{R}^d$ and $P \in \Lambda$ we define

$$\Omega_r = \{Q \in \Lambda^*: \int_{\mathbb{R}^d} x dQ(x) \geq r\}$$

and

$$A_P = \{s \in \mathbb{R}^d\colon\ K(\Omega_s,P) < \infty\}.$$

With these notations the following theorem will be proved.

THEOREM 4.2. *Let* $P \in \Lambda$ *and* $r \in (\partial A_P)^c$. *Then, for each sequence* $\{u_n\}$ *in* $\mathbb{R}^d$ *such that* $\lim_{n\to\infty} u_n = 0$,

$$\text{(4.5)} \qquad \lim_{n\to\infty} n^{-1} \log \Pr\{n^{-1} \sum_{i=1}^{n} X_i \geq r + u_n\} = -K(\Omega_r,P)$$

and

$$\text{(4.6)} \qquad K(\Omega_r,P) = \sup_{t\in\mathbb{R}^d_+} \{t'r - \log \int_{\mathbb{R}^d} e^{t'x} dP(x)\}.$$

Moreover, the supremum on the right-hand side of (4.6) is achieved if $r \in \overset{0}{A}_P$.

Theorem 4.2 generalizes Chernoff's theorem to d-dimensional vectors, but does not cover the case $r \in \partial A_P$. Relation (4.6) extends results by Hoeffding (1965) and Csiszár (1975, Theorem 3.3) who both considered sets Ω_r of the type $\{Q \in \Lambda\colon \int_{\mathbb{R}^d} x dQ(x) = r\}$ assuming finiteness of the moment generating function of P in a neighborhood of the origin.

The following example demonstrates that (4.5) may fail if r is a boundary point of A_P.

EXAMPLE 4.1 Let d = 2 and define the pm P by $P(\{a\}) = P(\{b\}) = \frac{1}{2}$, where $a = (1,0)$ and $b = (0,1)$. Let $r = (\frac{1}{2},\frac{1}{2})$, hence $r \in \partial A_P$. Since $\Pr\{n^{-1} \Sigma_{i=1}^n X_i \geq r\} = \binom{n}{\frac{1}{2}n} 2^{-n}$ for n even and = 0 for n odd, the limit in the left-hand member of (4.5), with $u_n = 0$, does not exist in this case (the limes inferior is $-\infty$, the limes superior is 0). It is easily verified that $K(\Omega_r,P) = 0$.

The next theorem provides some more information about the exceptional case $r \in \partial A_P$. It asserts the existence of a supporting hyperplane through r of the support of P with some special properties.

THEOREM 4.3 *Let* $P \in \Lambda$ *and* $r \in \partial A_P$. *Then there exists a hyperplane* $\mathbb{H}_s(r) = \{x \in \mathbb{R}^d\colon s'x = s'r\}$ *through* r *and a corresponding half-space*

$\mathbb{H}_s^*(r) = \{x \in \mathbb{R}^d\colon s'x > s'r\}$ *where* $s \in \mathbb{R}_+^d$ *and* $s \neq 0$, *with the following properties*

(i) $P(\mathbb{H}_s^*(r)) = 0$ *and* $P(\mathbb{H}_s^*(p)) > 0$ *for each* $p < r$

(ii) *If* $r \in A_P \cap \partial A_P$, *then* $P(\mathbb{H}_s(r)) > 0$

(iii) *If* $r \in A_P^c \cap \partial A_P$ *and* $P(\mathbb{H}_s(r)) = 0$, *then (4.5) and (4.6) hold*

(iv) *If* $P(\mathbb{H}_s(r)) = P(\{r\}) > 0$, *then (4.5) and (4.6) hold provided* $u_n = 0$ *for all large* $n \in \mathbb{N}$.

Consider the case $d = 1$. If $r \in \partial A_P$, then the hyperplane $\mathbb{H}_s(r)$ of Theorem 4.3 reduces to the point $\{r\}$ and either (iii) or (iv) are satisfied. Hence Theorem 4.2 and 4.3 together contain the original one-dimensional theorem of Chernoff.

If P is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^d$, case (ii) of Theorem 4.3 cannot occur and (iii) is in force. Hence Theorems 4.2 and 4.3 yield

COROLLARY 4.1. *Let* $P \in \Lambda$ *and suppose* P *is absolutely continuous with respect to Lebesgue measure on* $\mathbb{R}^d$. *Then (4.5) and (4.6) hold for each* $r \in \mathbb{R}^d$ *and each sequence* $\{u_n\}$ *in* $\mathbb{R}^d$ *tending to the zero vector.*

Henceforth the sets $B_m \subset \mathbb{R}^d$ and $\Psi_m \subset \Lambda$ for $m \in \mathbb{N}$ are defined by

$$B_m = \{x \in \mathbb{R}^d\colon \ |x_i| \leq m,\ 1 \leq i \leq d\}$$

and

$$\Psi_m = \{Q \in \Lambda\colon Q(B_m) = 1\}.$$

For any $m \in \mathbb{N}$ and $Q \in \Lambda$ such that $Q(B_m) > 0$ the conditional pm Q_m is defined by $Q_m(B) = Q(B \mid B_m)$, $B \in \mathcal{B}$.

Before proving the theorems we first establish two lemmas.

LEMMA 4.2 *Let* $P \in \Lambda$. *Then* A_P *is convex and the function* $s \to K(\Omega_s, P)$, $s \in \mathbb{R}^d$, *is convex and hence continuous on* A_P^0.

PROOF. This is an easy consequence of the convexity of the function $\alpha \to \alpha \log \alpha$, $\alpha > 0$ and the linearity of the function $Q \to \int_{\mathbb{R}^d} x dQ(x)$ on Λ^*. □

LEMMA 4.3. *Let Γ be a non-empty convex subset of Λ^* and let $p \in \mathbb{R}^d$. Consider the system of* d *inequalities*

$$(4.7) \qquad \int_{\mathbb{R}^d} x dQ(x) > p .$$

Then either there is a solution $Q \in \Gamma$ of (4.7) or, alternatively, there exists $t \in \mathbb{R}_+^d$, $t \neq 0$, such that

$$(4.8) \qquad t' \int_{\mathbb{R}^d} x dQ(x) \le t'p \qquad \textit{for all } Q \in \Gamma .$$

PROOF. This is Theorem 1 in Fan, Glicksberg and Hoffman (1957), specialized to the present situation. □

PROOF OF THEOREM 4.2 It is clear that $T : \Lambda^* \to \mathbb{R}^d$ defined by $T(Q) = \int_{\mathbb{R}^d} x dQ(x)$, $Q \in \Lambda^*$, is linear and τ-continuous on Ψ_m for each $m \in \mathbb{N}$. Let $r \in A_P^0$ and let $C_r = \{x \in \mathbb{R}^d : x \ge r\}$. Then $C_r^0 = \{x \in \mathbb{R}^d : x > r\}$ and $K(T^{-1}(C_r^0),P) < \infty$ by Lemma 4.2. Hence the conditions of Theorem 4.1 are satisfied with $A = C_r$ and $B_m = [-m,m]^d$, implying that (4.5) is satisfied if $u_n = 0$, $n \in \mathbb{N}$. If $r \in \bar{A}_P^c$, then obviously $K(\Omega_r,P) = \infty$ and the left-hand member of (4.5) is equal to $-\infty$ by Markov's inequality and (4.6), cf. the proof of Theorem 4.3 (iii).

We proceed to prove (4.6). First consider the case that $r \in A_P^0$. Let $Q \in \Omega_r$, $K(Q,P) < \infty$, $q = dQ/dP$ and $t \in \mathbb{R}_+^d$. Following Hoeffding (1967), we note that by Jensen's inequality

$$\begin{aligned} K(Q,P) &\ge K(Q,P) + t'(r - \int_{\mathbb{R}^d} x dQ(x)) \\ &= t'r - \int_{q>0} \log\{e^{t'x}/q(x)\} dQ(x) \\ &\ge t'r - \log \int_{\mathbb{R}^d} e^{t'x} dP(x) \end{aligned}$$

and hence

$$K(\Omega_r,P) \ge \sup_{t \in \mathbb{R}_+^d} \{t'r - \log \int_{\mathbb{R}^d} e^{t'x} dP(x)\}.$$

It still must be shown that conversely

$$(4.9)\qquad K(\Omega_r,P) \le \sup_{t\in\mathbb{R}^d_+}\{t'r - \log \int_{\mathbb{R}^d} e^{t'x}dP(x)\}.$$

First suppose that P has compact support, i.e. $P(B_m) = 1$ for sufficiently large $m \in \mathbb{N}$. Since Ψ_m is τ-closed and the restriction of T to Ψ_m is τ-continuous, $\Omega_r \cap \Psi_m$ is τ-closed and hence, by Lemma 3.2, there exists a pm $\bar{Q} \in \Omega_r$ such that

$$(4.10)\qquad K(\bar{Q},P) = K(\Omega_r,P).$$

The supporting hyperplane theorem, the convexity of the function $t \to K(\Omega_t,P)$ and its monotonicity in each argument $t^{(i)}$ imply the existence of $s \in \mathbb{R}^d_+$ such that

$$(4.11)\qquad K(\Omega_t,P) \ge K(\Omega_r,P) + s'(t-r) \qquad \text{for all } t \in A_P.$$

Let $\beta(s) = \int_{\mathbb{R}^d} e^{s'x}dP(x)$ and let the pm Q be defined by its density $q = dQ/dP$ given by $q(x) = e^{s'x}/\beta(s)$, $x \in \mathbb{R}^d$. Then

$$(4.12)\qquad K(Q,P) = s'\int_{\mathbb{R}^d} xdQ(x) - \log \beta(s).$$

Application of (4.10) and (4.11), with $t = \int_{\mathbb{R}^d} xdQ(x)$, yields

$$(4.13)\qquad K(Q,P) \ge K(\bar{Q},P) + s'(\int_{\mathbb{R}^d} xdQ(x)-r).$$

Since

$$K(\bar{Q},P) - K(\bar{Q},Q) = \int_{\mathbb{R}^d} \log q(x)d\bar{Q}(x) = s'\int_{\mathbb{R}^d} xd\bar{Q}(x) - \log \beta(s)$$

we have by (4.12) and (4.13)

$$\begin{aligned}(4.14)\qquad K(\bar{Q},Q) &= K(\bar{Q},P) - s'\int_{\mathbb{R}^d} xd\bar{Q}(x) + \log \beta(s)\\ &= K(\bar{Q},P) - K(Q,P) + s'(\int_{\mathbb{R}^d} xdQ(x) - \int_{\mathbb{R}^d} xd\bar{Q}(x))\\ &\le s'(r-\int_{\mathbb{R}^d} xd\bar{Q}(x)) \le 0.\end{aligned}$$

It follows that $K(\bar{Q},Q) = 0$, hence $\bar{Q} = Q$ and therefore

$$(4.15)\qquad K(\Omega_r,P) = K(Q,P) = s'r - \log \int_{\mathbb{R}^d} e^{s'x} dP(x).$$

This proves (4.9) for P with compact support. (We note in passing that by (4.14) $s^{(i)} > 0$ implies $\int_{\mathbb{R}^d} x^{(i)} dQ(x) = r^{(i)}$.)

There is also another line of argument to reach this conclusion. One first proves that the function $t \to t'r - \int_{\mathbb{R}^d} e^{t'x} dP(x)$ attains its supremum on the set $\mathbb{R}^d_+$ for some $s \in \mathbb{R}^d_+$, defines Q with this s as before and shows by considering partial derivatives that $Q \in \Omega_r$ and finally by Jensen's inequality that (4.15) is indeed satisfied. However, the present proof seems to be more direct and continues to hold if $r \in A_P \cap \partial A_P$.

Now let $P \in \Lambda$ be arbitrary. For each $m \in \mathbb{N}$ such that $P(B_m) > 0$ and $r \in A^0_{P_m}$ there exists by (4.11) $s_m \in \mathbb{R}^d_+$ satisfying

$$s'_m(t-r) \le K(\Omega_t,P_m) - K(\Omega_r,P_m)$$

for each $t \in A_{P_m}$. Hence in view of Lemma 4.1

$$\limsup_{m\to\infty} s'_m(t-r) \le K(\Omega_t,P) - K(\Omega_r,P)$$

for each $t > r$, $t \in A_P$, implying that $\{s_m\}$ has a convergent subsequence $\{s_{m_n}\}$. Let $\lim_{n\to\infty} s_{m_n} = s$. By Lemma 4.1, (4.15) and Fatou's lemma

$$\begin{aligned} K(\Omega_r,P) &= \lim_{n\to\infty} K(\Omega_r,P_{m_n}) \\ &= \lim_{n\to\infty} \{s'_{m_n} r - \log \int_{\mathbb{R}^d} \exp(s'_{m_n} x) dP_{m_n}(x)\} \\ &\le s'r - \log \int_{\mathbb{R}^d} \exp(s'x) dP(x). \end{aligned}$$

Thus (4.9) is proved in this case too and (4.6) follows for $r \in A^0_P$.

It remains to prove (4.6) in the case $r \in (A^c_P)^0$. Let $p \in (A^c_P)^0$, $p < r$. Apply Lemma 4.3 with $\Gamma = \{Q \in \Lambda^*: K(Q,P) < \infty\}$. Since (4.7) does not hold, there exists $s \in \mathbb{R}^d_+$, $s \ne 0$, such that

(4.16) $$s' \int_{\mathbb{R}^d} x dQ(x) \leq s'p \quad \text{for all } Q \in \Gamma.$$

It follows that (with the notation of Theorem 4.3)

(4.17) $$P(\mathbb{H}_s^*(p)) = 0.$$

For suppose that (4.17) does not hold. Let A be a compact subset of $\mathbb{H}_s^*(p)$ such that $P(A) > 0$, and let Q be the conditional pm defined by $Q(B) = P(B|A)$, $B \in \mathcal{B}$. Then $K(Q,P) = -\log P(A) < \infty$ and $s' \int_{\mathbb{R}^d} x dQ(x) > s'p$, in contradiction to (4.16). Hence

$$\sup_{t \in \mathbb{R}_+^d} \{t'r - \log \int_{\mathbb{R}^d} e^{t'x} dP(x)\} \geq$$

$$\geq \lim_{\alpha \to \infty} - \log \int_{\mathbb{R}^d} \exp\{\alpha s'(x-p) + \alpha s'(p-r)\} dP(x) =$$

$$= \infty = K(\Omega_r, P)$$

and the proof of Theorem 4.2 is complete. □

PROOF OF THEOREM 4.3. Let $r \in \partial A_P$ and put $\Gamma = \{Q \in \Lambda^*: K(Q,P) < \infty\}$. Applying Lemma 4.3 with $p = r$, (4.7) is obviously not satisfied and hence there exists $s \in \mathbb{R}_+^d$, $s \neq 0$, such that

$$s' \int_{\mathbb{R}^d} x dQ(x) \leq s'r \quad \text{for all } Q \in \Gamma.$$

It will be demonstrated that for this vector s, $\mathbb{H}_s(r)$ and $\mathbb{H}_s^*(r)$ have the required properties.

(i) The proof of $P(\mathbb{H}_s^*(r)) = 0$ is similar to the derivation of (4.17) from (4.16). Let $p < r$, hence $p \in A_P^0$. Then $P(\mathbb{H}_s^*(p)) > 0$. For otherwise every pm $Q \in \Gamma$ would satisfy $s' \int_{\mathbb{R}^d} x dQ(x) \leq s'p$, in contradiction to the existence of a pm $Q \in \Gamma$ with the property $\int_{\mathbb{R}^d} x dQ(x) > p$.

(ii) Suppose $r \in A_P \cap \partial A_P$. In that case a pm $Q \in \Gamma$ exists such that $\int_{\mathbb{R}^d} x dQ(x) \geq r$. Hence $Q(\mathbb{H}_s(r) \cup \mathbb{H}_s^*(r)) > 0$ and therefore, as a consequence of $Q << P$ and (i), $P(\mathbb{H}_s(r)) > 0$.

(iii) Let $r \in A_P^c \cap \partial A_P$ and $P(\mathbb{H}_s(r)) = 0$. In this case $K(\Omega_r, P) = \infty$ since $r \in A_P^c$. Moreover, since $P(\mathbb{H}_s^*(r) \cup \mathbb{H}_s(r)) = 0$,

$$\sup_{t \in \mathbb{R}_+^d} \{t'r - \log \int_{\mathbb{R}^d} e^{t'x} dP(x)\} \geq$$

$$\geq - \lim_{\alpha \to \infty} \log \int_{\mathbb{R}^d} \exp\{\alpha s'(x-r)\} dP(x) = \infty$$

by dominated convergence and (4.6) is proved. Finally, by Markov's inequality, for any $t \in \mathbb{R}_+^d$ and $u_n \in \mathbb{R}^d$,

$$\Pr\{n^{-1} \sum_{i=1}^n X_i \geq r + u_n\} \leq \Pr\{\sum_{i=1}^n t'X_i \geq nt'(r+u_n)\}$$

$$\leq E \exp\{\sum_{i=1}^n t'X_i\}/\exp\{nt'(r+u_n)\}$$

$$= (\int_{\mathbb{R}^d} \exp\{t'(x-r-u_n)\} dP(x))^n.$$

Hence, if $\lim_{n \to \infty} u_n = 0$,

$$\lim_{n \to \infty} n^{-1} \log \Pr\{n^{-1} \sum_{i=1}^n X_i \geq r + u_n\} \leq$$

$$\leq - \sup_{t \in \mathbb{R}_+^d} (-\log \int_{\mathbb{R}^d} \exp\{t'(x-r)\} dP(x)) = -\infty$$

and (4.5) is established.

(iv) Let $\gamma = P(\mathbb{H}_s(r)) = P(\{r\}) > 0$. Since in this case $Q \in \Gamma \cap \Omega_r$ iff $Q(\{r\}) = 1$, $K(\Omega_r, P) = -\log \gamma$. It is also readily seen that $\Pr\{n^{-1} \Sigma_{i=1}^n X_i \geq r\} = \Pr\{X_i = r, 1 \leq i \leq n\} = \gamma^n$ and hence $\lim_{n \to \infty} n^{-1} \log \Pr\{n^{-1} \Sigma_{i=1}^n X_i \geq r\} = \log \gamma$, proving (4.5) for $u_n = 0$. By dominated convergence

$$\sup_{t \in \mathbb{R}_+^d} \{t'r - \log \int_{\mathbb{R}^d} e^{t'x} dP(x)\} \geq$$

$$\geq -\lim_{\alpha \to \infty} \log \int_{\mathbb{R}^d} \exp\{\alpha s'(x-r)\} dP(x) = -\log \gamma.$$

The reverse inequality is obtained by Markov's inequality, as in the last lines of the proof of (iii). Thus (4.6) is also established and the proof of the theorem is complete. □

5. 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. If $G \in D$, the corresponding pm in Λ_1 will be denoted by P_G. 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.

For convenience of notation we write $K(G,F)$ instead of $K(P_G,P_F)$ and similarly we write $K(\Omega,F)$ to denote $\inf_{G\in\Omega} K(P_G,P_F)$ if Ω is a subset of D. For $G \in D$ the inverse G^{-1} is defined in the usual way by $G^{-1}(u) = \inf\{x \in \mathbb{R}: G(x) \geq u\}$.

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

$$(5.1) \qquad T(\hat{F}_n) = \int_0^1 J(u)\hat{F}_n^{-1}(u)du$$

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

$$(5.2) \qquad T(\hat{F}_n) = \sum_{i=1}^{n} c_{n,i}\, X_{i:n}$$

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 calles L-estimators, cf. Huber (1972). For a more recent discussion we refer to Bickel and Lehmann (1975).

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

$$(5.3) \qquad \Omega_t = \{G \in D: \int_0^1 J(u)G^{-1}(u)du \geq t,\ \int_0^1 |J(u)G^{-1}(u)|\, du < \infty\},$$

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

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

THEOREM 5.1. *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* $\lim_{n\to\infty} u_n = 0$,

$$\lim_{n\to\infty} n^{-1} \log \Pr\{T(\hat{F}_n) \geq r + u_n\} = -K(\Omega_r, F) \tag{5.4}$$

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

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

(a) $J \geq 0$ *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,* $J \geq 0$ 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 (5.4) holds provided conditions (ii) and (c) are satisfied and* $u_n \leq 0$ *for all large* $n \in \mathbb{N}$.

REMARK 5.1. Condition (ii) of Theorem 5.1 is satisfied if P_F has compact support or if $0 < \alpha < \beta < 1$.

REMARK 5.2. The second part of Theorem 5.1 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 5.1. 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

$$(5.5)\qquad T(G) = \int_0^1 J(u)1_B(G^{-1}(u))G^{-1}(u)du,\ G \in D.$$

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 $1_B(G_n^{-1})G_n^{-1}.1_A$, $n \in \mathbb{N}$, 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 (5.4) is now completed by an application of Theorem 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 3.2 and the fact that a ρ-closed set is also τ-closed.

First suppose that conditions (a) and (b) are satisfied. Since $P_G << P_F$, 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

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

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

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

Then $G_\varepsilon^{-1} \ge G^{-1}$ 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 $\int_0^1 J(u)G^{-1}(u)du = b \ge r$ since $G \in \Omega_r$. Suppose r is a continuity point of F. Then $P_G << P_F$ implies $P_G(\{r\}) = 0$ and hence $b > r$, since b is a discontinuity 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

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

Then $G_\varepsilon(r-0) = G(r-0) - \varepsilon \le \alpha - \varepsilon$, hence $G_\varepsilon \in \Omega_r$. Considering the partition $\mathcal{P} = \{(-\infty,r), [r,\infty)\}$ of $\mathbb{R}$ it follows immediately that there is a τ-open neighborhood of G_ε contained in Ω_r. Hence $K(\mathrm{int}_\tau(\Omega_r),F) \le K(G_\varepsilon,F)$, for each $\varepsilon > 0$. Since $\lim_{\varepsilon\downarrow 0} K(G_\varepsilon,F) = K(G,F)$, we have $K(\mathrm{int}_\tau(\Omega_r),F) \le \lim_{\varepsilon\downarrow 0} K(G_\varepsilon,F) = K(\Omega_r,F)$, i.e. $K(\mathrm{int}_\tau(\Omega_r),F) = K(\Omega_r,F)$. The τ-continuity of T implies that Ω_r is τ-closed and hence Theorem 3.1 yields that (5.4) 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 3.3) implies that (5.4) 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 $K(G',F) \le K(G,F)$, hence $K(G,F) = K(G',F) = -\log P_F([r,\infty))$. Since Ω_r is

τ-closed, Lemma 2.4 implies that condition (A) of Lemma 3.1 is satisfied. Hence $\limsup_{n\to\infty} n^{-1} \log \Pr\{\hat{F}_n \in \Omega_r\} \le \log P_F([r,\infty))$ (see (3.2)). It is clear that conversely $\liminf_{n\to\infty} n^{-1} \log \Pr(\hat{F}_n \in \Omega_r) \ge$ $\ge \liminf_{n\to\infty} n^{-1} \log \Pr\{X_{1:n} \ge r\} = \log P_F([r,\infty))$. Thus (5.4) holds provided $u_n = 0$ for all large $n \in \mathbb{N}$. By the same argument as before (5.4) also holds if $u_n \le 0$ for all large $n \in \mathbb{N}$. □

REMARK 5.3. The continuity of a function which is essentially equivalent to the function T in (5.5) has been pointed out by Bickel and 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 of this paper. Note that Hoadley's (1967) Theorem 1 would not suffice to prove (5.4) 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 5.1 is not immediately applicable, but the next theorem may be of use.

THEOREM 5.2. *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 (5.3) for* $t \in \mathbb{R}$. *Then, for each sequence of real numbers* $\{u_n\}$ *such that* $\lim_{n\to\infty} u_n = 0$,

$$(5.6) \qquad \lim_{n\to\infty} n^{-1} \log \Pr\{\int_0^1 J_n(u)\, \hat{F}_n^{-1}(u)du \ge r + u_n\} = -K(\Omega_r, F)$$

if J, F, α *and* β *satisfy conditions (i) and (ii) of Theorem 5.1 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. In accordance with section 4 we write $B_m = [-m,m]$ and denote by G_m the conditional df defined by

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

Let $D^* = \{G \in D: P_G(B_m) = 1$ 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 $K(G,F) \le K(\Omega_r,F) + \eta$. 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 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 $n \ge N_0$

$$(5.7) \qquad \left|\int_0^1 J_n(u)\, \hat{F}_n^{-1}(u)du - \int_0^1 J(u)\hat{F}_n^{-1}(u)du\right| \le m \int_0^1 |J_n(u) - J(u)|\, du < \frac{1}{4}\varepsilon$$

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

$$\Pr\{\hat{F}_n \in A \mid \hat{F}_n^{-1}(u) \in B_m,\ u \in (0,1)\} = \Pr\{\hat{F}_n \in A \mid B_m\}$$
$$\text{if } P_F(B_m) > 0.$$

With this notation we have for each large $m \in \mathbb{N}$:

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

$$\ge \liminf_{n\to\infty} n^{-1} \log \Pr\{\int_0^1 J_n(u)\hat{F}_n^{-1}(u)du \ge r + u_n \mid B_m\} + \log P_F(B_m)$$

$$\ge \liminf_{n\to\infty} n^{-1} \log \Pr\{\int_0^1 J(u)\hat{F}_n^{-1}(u)du \ge r + \frac{1}{2}\varepsilon \mid B_m\} + \log P_F(B_m)$$

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

The last inequality holds by Theorem 5.1, since we may choose a continuity point $r_m \in (r+\frac{1}{2}\varepsilon, r+\varepsilon)$ 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

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

Hence by condition (i)

$$(5.8) \qquad \liminf_{n\to\infty} n^{-1} \log \Pr\{ \int_0^1 J_n(u)\hat{F}_n^{-1}(u)du \geq r + u_n\} \geq -K(\Omega_r,F).$$

Next we show that conversely

$$(5.9) \qquad \limsup_{n\to\infty} n^{-1} \log \Pr\{ \int_0^1 J_n(u)\hat{F}_n^{-1}(u)du \geq r + u_n\} \leq -K(\Omega_r,F).$$

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:

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

since (5.7) holds again for large n if $\hat{F}_n^{-1}(u) \in B_m$ for $u \in (\alpha,\beta)$. This result implies (5.9) by Theorem 5.1 and Lemma 3.3 (also if $K(\Omega_r,F) = \infty$) and the present theorem follows from (5.8) and (5.9). □

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

$$(5.10) \qquad T_n = (n-2[\alpha n])^{-1} \sum_{i=[\alpha n]+1}^{n-[\alpha n]} X_{i:n}, \quad n \in \mathbb{N},$$

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 5.3 *Let* $r \in \mathbb{R}$ *let* $F \in D$ *be continuous at* r *and let* T_n *be the* α*-trimmed mean given by (5.10). Then, for each sequence* $\{u_n\}$ *such that* $\lim_{n\to\infty} u_n = 0$,

$$(5.11) \qquad \lim_{n\to\infty} n^{-1} \log \Pr\{T_n \geq r + u_n\} = -K(\Omega_r^\alpha, F),$$

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 (5.11) continues to hold provided* $u_n \leq 0$ *for all large* $n \in \mathbb{N}$.

PROOF. 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 $J = (1-2\alpha)^{-1} 1_{(\alpha,1-\alpha)}$. If F is continuous at r, then (5.11) follows since in this case (c) and (d) of Theorem 5.1 and hence the conditions of Theorem 5.2 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 5.1 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 $\varepsilon > 0$ 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 $K(G_\delta,F) < K(G,F) + \varepsilon = K(\Omega_r^\alpha,F)+\varepsilon$ if $\delta < \delta_\varepsilon$, 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 $\mathcal{P}$ denote the partition $\{(-\infty,r), \{r\}, (r,\infty)\}$ of $\mathbb{R}$. Choosing appropriate $\delta_n \in (\frac{1}{2}\delta_\varepsilon, \delta_\varepsilon)$ it follows that there exists a sequence $\{G_n\} = \{G_{\delta_n}\}$ such that for all $n > (\frac{1}{2}\alpha\delta_\varepsilon)^{-1}$

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

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

(3) $K_{\mathcal{P}}(G_n,F) < K(\Omega_r^\alpha,F) + \varepsilon$.

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 3.1 yield

$$\Pr\{T_n \ge r+u_n\} \ge \Pr\{\hat{F}_n \in \Omega_{r,n}\} \ge \Pr\{d_{\mathcal{P}}(P_{\hat{F}_n}, P_{G_n}) = 0\} \ge$$

$$\ge \exp\{-n(K(\Omega_r^\alpha, F) + \varepsilon + o(1))\}$$

as $n \to \infty$, implying

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

On the other hand (5.9) continues to hold in the present case, with Ω_r^α in in lieu of Ω_r, since the second part of the proof of Theorem 5.2 does not use condition (i). This completes the proof of the last statement of the theorem. □

The actual computation of the infimum $K(\Omega_r^\alpha, F)$ in (5.11) 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

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

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 and is closely related to the proof of (4.6) in Theorem 4.2. Let

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

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

(5.12) $$K(\Omega_r^\alpha, F) = \inf\{K(\Omega_r^\alpha(a,b), F)\colon 0 < F(a) < F(b) < 1,\ F(b) > F(r)\}.$$

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

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

where $\phi(t) = \int_a^b xe^{tx}dF(x) \Big/ \int_a^b e^{tx}dF(x)$, $t \geq 0$. 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

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

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

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

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

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

Hence

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

Similarly, by Jensen's inequality,

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

and

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

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

Thus

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

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 (5.12)

$$(5.13) \qquad K(\Omega_r^\alpha,F) = g(\alpha) + \inf \{f_\alpha(a,b): 0 < F(a) < F(b) < 1,\ F(b) > F(r)\}.$$

REMARK 5.4. We briefly indicate another route to the result (5.11). Let T_n be defined by (5.10) and let $n_\alpha = n - 2[\alpha n]$, for each $n \in \mathbb{N}$. Then we may write

$$E \exp(n_\alpha t T_n) = E(E\{\exp(t \sum_{i=[\alpha n]+1}^{n-[\alpha n]} X_{i:n}) \mid X_{[\alpha n]:n}, X_{n-[\alpha n]+1:n}\}).$$

Suppose that F has density f with respect to Lebesgue measure. If f satisfies certain smoothness conditions, it follows from this representation that

$$(5.14) \qquad \lim_{n\to\infty} n^{-1} \log E \exp(n_\alpha t(T_n-r)) =$$

$$= -\inf_{-\infty<a<b<\infty} \{(1-2\alpha)tr - \alpha \log F(a) - \alpha \log (1-F(b))$$

$$- (1-2\alpha) \log \int_a^b \exp(tx) f(x)dx\}.$$

By Theorem 1 of Sievers (1969) (see also Plachky (1971) and Plachky and Steinebach (1975)):

(5.15) $$\lim_{n\to\infty} n^{-1} \log \Pr\{T_n \geq r\} = -\inf_{t\geq 0} E \exp(n_\alpha t(T_n - r))$$

provided the sequence of moment generating functions $E \exp(n_\alpha t(T_n-r))$ enjoys certain convergence properties.

By (5.13) the expression on the right-hand side of (5.14) is equal to $-K(\Omega_r^\alpha,F)$ (note that the infima over t and a, b are interchanged). Although this alternative approach requires stronger regularity conditions it may lead to evaluation of higher order terms in an expansion of large deviation probabilities of the trimmed mean.

REFERENCES

[1] BAHADUR, R.R. (1971), *Some limit theorems in statistics*, SIAM, Philadelphia.

[2] BICKEL, P.J. & E.L. LEHMANN (1975), *Descriptive statistics for non-parametric models. II. Location*, Ann. Statist. 3, 1045-1069.

[3] BILLINGSLEY, P. (1968), *Convergence of probability measures*, Wiley, New-York.

[4] 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 translations in Sel. Transl. Math. Statist. Prob. 5 (1965), 315-372).

[5] BOROVKOV, A.A. (1967), *Boundary-value problems for random walks and large deviations in function spaces*, Theory Prob. Applications XII, 4, 575-595.

[6] CHERNOFF, H. (1952), *A measure of asymptotic efficiency for tests of a hypothesis based on sums of observations*, Ann. Math. Statist. 23, 493-507.

[7] CSISZÁR, I. (1962), *Informationstheoretische Konvergenzbegriffe im Raum der Wahrscheinlichkeitsverteilungen*, A. Magyar Tud. Ak. Mat. Kut. Int. Közl. 7, 137-158.

[8] CSISZÁR, I. (1975), *I-divergence geometry of probability distributions and minimization problems*, Ann. Prob. 3, 146-158.

[9] FAN, K.I., I. GLICKSBERG & A.J. HOFFMAN (1957), *Systems of inequalities involving convex functions*, Amer. Math. Soc. Proc. 8, 617-622.

[10] GÄNSSLER, P. (1971), *Compactness and sequential compactness in spaces of measures*, Z. Wahrscheinlichkeitstheorie verw. Geb. 17, 124-146.

[11] HOADLEY, A.B. (1967), *On the probability of large deviations of functions of several empirical cdf's*, Ann. Math. Statist. 38, 360-381.

[12] HOEFFDING, W. (1967), *On probabilities of large deviations*, Proc. 5th Berkeley Symp. Math. Statist. Prob. 1, 203-219.

[13] HUBER, P.J. (1972), *Robust statistics: a review*, Ann. Math. Statist. 43, 1041-1067.

[14] PINSKER, M.S. (1964), *Information and information stability of random variables and processes*, Holden-Day, San Francisco.

[15] PLACHKY, D. (1971), *On a theorem of G.L. Sievers*, Ann. Math. Statist. 42, 1442-1443.

[16] PLACHKY, D. & J. STEINEBACH (1975), *A theorem about probabilities of large deviations with an application to queuing theory*, Period. Math. Hungar. 6, 343-345.

[17] RÉNYI, A. (1961), *On measures of entropy and information*, Proc. 4th Berkeley Symp. Math. Statist. Prob. 1, 547- 561.

[18] 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).

[19] SETHURAMAN, J. (1964), *On the probability of large deviations of families families of sample means*, Ann. Math. Statist. 35, 1304-1316.

[20] SIEVERS, G.L. (1969), *On the probability of large deviations and exact slopes*, Ann. Math. Statist. 40, 1908-1921.

[21] SIEVERS, G.L. (1975), *Multivariate probabilities of large deviations*, Ann. Statist. 3, 897-905.

[22] SIEVERS, G.L. (1976), *Probabilities of large deviations for empirical measures*, Ann. Statist. 4, 766-770.

[23] STONE, M. (1974), *Large deviations of empirical probability measures*, Ann. Statist. 2, 362-366.