The Annals of Probability
1979, Vol. 7, No. 4, 553–586

LARGE DEVIATION THEOREMS FOR EMPIRICAL PROBABILITY MEASURES

By P. Groeneboom, J. Oosterhoff and F. H. Ruymgaart

Mathematisch Centrum,
Vrije Universiteit and Katholieke Universiteit Nijmegen,

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.

1. Introduction. Let S be a Hausdorff space and let $\mathfrak{B}$ be the σ-field of Borel sets in S. Let Λ be the set of all probability measures (pms) on $\mathfrak{B}$; the abbreviation 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

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

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 $a \geqslant 0$. 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 $X_1, \cdots, X_n$ is denoted by $\hat{P}_n$, i.e., $\hat{P}_n(B)$ is the fraction of X_j's, $1 \leqslant j \leqslant n$, with values in the set $B \in \mathfrak{B}$.

Let $S = \mathbb{R}$ and let Λ_1 be the set of pms on $(\mathbb{R}, \mathfrak{B})$, 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 nonatomic pm. Let T be a real-valued function on Λ_1, uniformly continuous in the topology ρ. Define

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

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

Received March 25, 1977; revised May 24, 1978.
AMS 1970 *subject classification.* Primary 60F10.
Key words and phrases. Large deviations, empirical probability measures, Kullback-Leibler information, trimmed means.

$\{u_n\}$ is a sequence of real numbers tending to zero,

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

In Section 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 $S = \mathbb{R}$) and the range space of T may be different from $\mathbb{R}$;

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

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 nonatomic 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),

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

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 $(S, \mathfrak{B})$ 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 $B \in \mathfrak{B}$. 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 $\mathfrak{B}$-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

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

Then (1.3) *holds true.*

This is a generalization of Theorem 4.5 of Donsker and 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 (also called the vague topology) we mean the topology with subbasis elements $\{Q \in \Lambda : |\int f dQ - \int f dQ_0| < \varepsilon\}$, $Q_0 \in \Lambda$, $f \in C_B(S)$; 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 2.1) which in turn is finer than the weak topology. Hence any ρ-continuous (weakly continuous) functional $T : \Lambda_1 \rightarrow \mathbb{R}$ 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 3.1.

After some crucial lemmas in Section 2 the basic theorems are obtained in Section 3. The theory includes theorems of Borovkov, Donsker and Varadhan, 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. Section 5 is devoted to this subject. Finally we prove in Section 6 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.

In this framework Chernoff-type theorems are derived from Sanov-type theorems. In a very penetrating paper Bahadur and Zabell (1979) go the other way. They first obtain very general Chernoff-type theorems for sample means taking values in open convex sets and then, as an application, derive a result of type (1.3).

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 the Hausdorff space 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 $\mathcal{P} = \{B_1, \cdots, B_m\}$ of S define

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

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

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

(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 $\mathcal{P} = \{B_1, \cdots, B_m\}$ of S the pseudometric $d_{\mathcal{P}}$ on Λ is defined by

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

The topology τ of convergence on all Borel sets of S is generated by the family $\{d_{\mathcal{P}} : \mathcal{P}$ is a partition of $S\}$. A basis of this topology is provided by the collection of sets $\{R \in \Lambda : d_{\mathcal{P}}(R, Q) < \delta\}$ 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 nonatomic 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 $\geqslant \frac{1}{2}\varepsilon$. Hence there exists a partition $\mathcal{P} = \{B_1, \cdots, B_m\}$ of $\mathbb{R}$ consisting of singletons B_i such that $Q(B_i) \geqslant \frac{1}{2}\varepsilon$ 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, $1 \leqslant i \leqslant d$, be the one-dimensional marginals of Q. For each Q_i there exists by the previous paragraph a partition $\{B_{i,1}, \cdots, B_{i,m_i}\}$ of $\mathbb{R}$ consisting of singletons $B_{i,j}$ with $Q_i(B_{i,j}) \geqslant \frac{1}{2}\varepsilon$ and open or half open intervals $B_{i,j}$ with $Q_i(B_{i,j}) < \frac{1}{2}\varepsilon/d$. Let $\mathcal{P}$ be the partition consisting of the product sets $B_{1,j_1} \times \cdots \times B_{d,j_d}$, $1 \leqslant j_i \leqslant m_i$, $1 \leqslant i \leqslant d$, and let $m = \max_{1 \leqslant i \leqslant d} m_i$. 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 $\overline{\mathbb{R}}$ is the usual topology generated by the sets $[-\infty, x)$, $(x, \infty]$, $x \in \mathbb{R}$.

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

PROOF. 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

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

proving the lemma. $\square$

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 $B \in \mathfrak{B}$, $P(B) < \delta \Rightarrow Q(B) < \varepsilon$.

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 $P \in \Lambda$ and let $\Gamma = \{Q \in \Lambda : K(Q, P) \leqslant c\}$ for some finite $c \geqslant 0$. 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 $B \in \mathfrak{B}$ satisfying $P(B) < \delta$,

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

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

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

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

(b) Let M be the collection of all set functions $\mu : \mathfrak{B} \to [0, 1]$ endowed with the topology τ_1 of setwise convergence (note that τ is the corresponding 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 $\mathfrak{U} = \{U_\alpha : \alpha \in I\}$ on M. For each $B \in \mathfrak{B}$ the image of $\mathfrak{U}$ 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 $\mu_0(B) = c_B$, $B \in \mathfrak{B}$. Since $\mu_0 \in \bigcap_{\alpha \in I} \text{cl}_{\tau_1}(U_\alpha)$, the ultrafilter $\mathfrak{U}$ 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 $\mu \in \text{cl}_{\tau_1}(\Gamma)$. Clearly μ is an additive set function. To prove σ-additivity consider a sequence $\{B_n\}$ of disjoint Borel sets. Fix $\varepsilon > 0$. By part (a) there exists $\delta > 0$ such that $B \in \mathfrak{B}$, $P(B) < \delta \Rightarrow Q(B) < \varepsilon$ for each $Q \in \Gamma$. Choose k so large that $P(\cup_{n=k}^{\infty} B_n) = \Sigma_{n=k}^{\infty} P(B_n) < \delta$. Since $\mu \in \text{cl}_{\tau_1}(\Gamma)$ it follows that

$$\left|\mu\left(\bigcup_{n=1}^{\infty} B_n\right) - \Sigma_{n=1}^{k} \mu(B_n)\right| = \mu\left(\bigcup_{n=k+1}^{\infty} B_n\right) \leqslant \varepsilon,$$

implying that $\mu(\cup_{n=1}^{\infty} B_n) = \Sigma_{n=1}^{\infty} \mu(B_n)$. Hence $\mu \in \Lambda$. Now Lemma 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 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 $\mathcal{P} = \{B_1, \cdots, B_m\}$ be a partition of S. Then the $\mathcal{P}_P$-linear pm Q' corresponding to Q is defined by

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

$i = 1, \cdots, m$; $B \in \mathcal{B}$. 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 $\mathcal{P}$-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*

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

Then

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

PROOF. Let $\alpha = \sup\{K_{\mathcal{P}}(\Omega, P) : \mathcal{P}$ 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 $\Gamma = \{Q \in \Lambda : K(Q, P) \leq \alpha + \eta\}$. 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 $K_{\mathcal{P}}(Q_{\mathcal{P}}, P) \leq \alpha + \eta$. Let $Q'_{\mathcal{P}}$ be the $\mathcal{P}_P$-linear pm corresponding to $Q_{\mathcal{P}}$. Then

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

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 $\mathcal{N} = \{Q'_{\mathcal{P}} : \mathcal{P}$ is a partition of $S\}$.

Consider the open neighborhood $\{R \in \Lambda : d_{\mathcal{P}}(R, \bar{Q}) < \varepsilon\}$ 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

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

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 $\{Q_{\mathcal{P}} : \mathcal{P}$ is a partition of $S\}$. Since $Q_{\mathcal{P}} \in \Omega$ for each $\mathcal{P}$, $\bar{Q} \in \text{cl}_\tau(\Omega)$. However, $\bar{Q} \in \Gamma \Rightarrow K(\bar{Q}, P) \leq \alpha + \eta < K(\Omega, P)$ in contradiction to (2.4) and so (2.5) follows. □

REMARK 2.1. Lemma 2.4 is in fact a minimax theorem since in view of (2.2) the result (2.5) can also be written as

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

REMARK 2.2. The following example shows that (2.4) is not necessary for (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 $K(\Omega_1, P) = \sup_{\mathcal{P}}K_{\mathcal{P}}(\Omega_1, P) = \infty$ and hence by (2.2) $K(\Omega, P) = K(Q_1, P) = \log 2 = \sup_{\mathcal{P}}K_{\mathcal{P}}(Q_1, P) = \sup_{\mathcal{P}}K_{\mathcal{P}}(\Omega, P)$. Obviously $P \in \mathrm{cl}_\tau(\Omega)$ and therefore $K(\mathrm{cl}_\tau(\Omega), P) = 0$. Thus (2.4) is violated but (2.5) holds true.

REMARK 2.3. Let $\mathrm{scl}_\tau(\Omega)$ denote the sequential closure of Ω, i.e., $Q \in \mathrm{scl}_\tau(\Omega)$ 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 cannot be replaced by $K(\mathrm{scl}_\tau(\Omega), P) = K(\Omega, P)$. Let Ω be the set of all pms on $\mathbb{R}$ with countable carrier and let P be a nonatomic pm on $\mathbb{R}$. Then $\sup\{K_{\mathcal{P}}(\Omega, P) : \mathcal{P}$ is a partition of $\mathbb{R}\} = 0$, but $K(\Omega, P) = K(\mathrm{scl}_\tau(\Omega), P) = \infty$ since $\Omega = \mathrm{scl}_\tau(\Omega)$. In this case $\mathrm{cl}_\tau(\Omega) = \Lambda_1 =$ 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* $\mathrm{supp}(Q)$ 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 $Q(\mathrm{supp}(Q))$ may be smaller than one. However, we shall say that $Q \in \Lambda$ has finite support $\{x_1, \cdots, x_k\}$ if $Q(\{x_i\}) > 0$, $i = 1, \cdots, k$, and $\Sigma_{i=1}^k Q(\{x_i\}) = 1$. In general, let us call a pm Q *Lindelöf inner regular* if $Q(B) = \sup\{Q(V) : V \subset B, V$ 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 and Zabell (1979). This regularity condition is certainly satisfied if S is second countable.

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

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

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

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

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

$$|\int f_j dQ_V - \int f_j dQ_0| \leqslant \Sigma_{i=1}^k |\int_{\tilde{U}_i}(f_j - f_j(x_i))dQ_V| < \varepsilon. \qquad \square$$

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.

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 $X_1, X_2, \cdots$. The problem of which events $\{\hat{P}_n \in \Omega\}$ are $\mathfrak{B}^n$-measurable for all n is (at least partially) solved by

PROPOSITION 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 $\{\hat{P}_n \in \Omega\} \in \mathfrak{B}^n$ for all $n \in \mathbb{N}$ iff $\Omega \cap \tilde{\Lambda} \in \tilde{\mathcal{W}}$, where $\tilde{\mathcal{W}}$ is the σ-field induced by $\mathcal{W}$ on $\tilde{\Lambda}$ and $\mathcal{W}$ 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 $\mathcal{W}(n)$ denote the σ-field induced by $\mathcal{W}$ on $\Lambda(n)$.

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

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

It remains to prove $\Omega \cap \Lambda(n) \in \mathcal{W}(n)$ for all $n \Leftrightarrow \Omega \cap \tilde{\Lambda} \in \tilde{\mathcal{W}}$. The implication $\Leftarrow$ is trivial. To prove $\Rightarrow$, let $\Omega_n \in \mathcal{W}$ be such that $\Omega \cap \Lambda(n) = \Omega_n \cap \Lambda(n)$, $n \in \mathbb{N}$. Fix $m \in \mathbb{N}$. If the pm $Q \in \tilde{\Lambda} \cap \Omega^c$, 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 $\Omega \cap \tilde{\Lambda} \in \tilde{\mathcal{W}}$ and the proof is complete. $\square$

The collection of sets $\Omega \subset \Lambda$ satisfying $\Omega \cap \tilde{\Lambda} \in \tilde{\mathcal{W}}$ is quite rich, much richer than $\mathcal{W}$. 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 3.1 we briefly return to this matter.

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 $\{\hat{P}_n\}$. 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. Consider the conditions*
(A) $K(\Omega, P) = \sup\{K_{\mathcal{P}}(\Omega, P) : \mathcal{P}$ *is a partition of* $S\}$,
(B) $K(\Omega, P) = K(\text{int}_\tau(\Omega), P)$.
If (A) is satisfied,

$$\limsup_{n\to\infty} n^{-1}\log \Pr\{\hat{P}_n \in \Omega\} \leqslant -K(\Omega, P); \tag{3.1}$$

if (B) is satisfied,

$$\liminf_{n\to\infty} n^{-1}\log \Pr\{\hat{P}_n \in \Omega\} \geqslant -K(\Omega, P). \tag{3.2}$$

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

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

PROOF. To prove the lemma it is first shown that condition (A) implies (3.1). 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 $\mathcal{P} = \{B_1, \cdots, B_m\}$ and let $p_j = P(B_j)$, $1 \leqslant j \leqslant m$. Then

$$\begin{aligned}\Pr\{\hat{P}_n \in \Omega\} &\leqslant \Pr\{K_{\mathcal{P}}(\hat{P}_n, P) \geqslant K_{\mathcal{P}}(\Omega, P)\}\\ &= \Sigma^* n!/\{(nz_{n,1})! \cdots (nz_{n,m})!\}\cdot \Pi_{i=1}^m p_i^{nz_{n,i}}\\ &= \Sigma^* n!\,\{\Pi_{i=1}^m (nz_{n,i})!\}^{-1}\Pi_{i=1}^m z_{n,i}^{nz_{n,i}}\cdot\\ &\quad\cdot\exp\{-n\Sigma_{i=1}^m z_{n,i}\log(z_{n,i}/p_i)\},\end{aligned}$$

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

$$\Sigma_{i=1}^m z_{n,i} = 1, \qquad z_{n,i} \geqslant 0, \qquad nz_{n,i} \in \mathbb{Z} \qquad 1 \leqslant i \leqslant m$$

and

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

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

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

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

$$n!/\{(nz_{n,1})! \cdots (nz_{n,m})!\} \leqslant \exp\{-n\Sigma_{i=1}^m z_{n,i}\log z_{n,i} + O(\log n)\}.$$

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

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

as $n \to \infty$. Since $c < K(\Omega, P)$ is arbitrary, (3.1) follows.

Conversely we prove that condition (B) implies (3.2). Assume $K(\Omega, P) < \infty$ since otherwise (3.2) is trivial. Fix $\varepsilon > 0$. In view of condition (B) $\text{int}_\tau(\Omega)$ is not empty and a pm $Q \in \text{int}_\tau(\Omega)$ exists satisfying $K(Q, P) < K(\Omega, P) + \frac{1}{2}\varepsilon$. Since $Q \in \text{int}_\tau(\Omega)$, a partition $\mathcal{P} = \{B_1, \cdots, B_m\}$ 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) $nQ_n(B_i) \in \mathbb{Z}$, $1 \leqslant i \leqslant m$;

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

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

$$\Pr\{\hat{P}_n \in \Omega\} \geqslant \Pr\{d_{\mathcal{P}}(\hat{P}_n, Q_n) = 0\}$$
$$= n!/\{(nz_{n,1})! \cdots (nz_{n,m})!\} \cdot \Pi_{i=1}^m (P(B_i))^{nz_{n,i}},$$

where $\Sigma_{i=1}^m z_{n,i} = 1$, $z_{n,i} \geqslant 0$, $nz_{n,i} \in \mathbb{Z} (1 \leqslant i \leqslant m)$ and

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

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

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

and (3.2) easily follows, which completes the proof. □

REMARK 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 3.1 is based on the inclusion $\{d_{\mathcal{P}}(\hat{P}_n, Q_n) = 0\} \subset \Omega \subset \{K_{\mathcal{P}}(\hat{P}_n, P) \leqslant K_{\mathcal{P}}(\Omega, P)\}$ where the sets on the left and right are measurable. Hence, if $\overline{P^n}(\underline{P}^n)$ denotes the outer (inner) measure corresponding to the product measure P^n on S^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 3.1 continues to hold for arbitrary sets Ω. Similar remarks apply to all other results of this section.

Stone (1974) proves (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 $\mathcal{P}$ of S and $\delta > 0$ such that

(C2) $K_{\mathcal{P}}(\Omega, P) \leqslant K_{\mathcal{P}}(Q, P) < K_{\mathcal{P}}(\Omega, P) + \varepsilon$;

(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 $Q \in \text{int}_\tau(\Omega)$ exists satisfying $K(Q, P) < K(\Omega, P) + \frac{1}{2}\varepsilon$. Since $Q \in \text{int}_\tau(\Omega)$, there exists a partition $\mathcal{T}$ 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 $K_{\mathcal{T}}(R, P) \leqslant K_{\mathcal{P}}(R, P)$

for each pm R if $\mathcal{P}$ is finer than $\mathcal{T}$). Hence

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

Moreover, for small enough $\delta' > 0$ the implication $R \in \Lambda$, $d_{\mathcal{P}}(R, Q) < \delta' \Rightarrow d_{\mathcal{T}}(R, Q) < \delta$ 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 $Q' \in \text{int}_\tau(\Omega)$ and (C2) yields

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

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

The present method of proof of Lemma 3.1 is well suited to prove (3.3) 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 (3.3). (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*

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

Then (3.3) *holds.*

Borovkov has shown (see (31) in Borovkov (1967)) that (3.3) holds if P is a nonatomic pm on $\mathbb{R}$, Ω is a ρ-open set and $K(\Omega, P) = K(\text{cl}_\rho(\Omega), P)$. This is a particular case of Theorem 3.1 in view of Lemma 2.1.

In their work on large deviations of Markov processes, Donsker and Varadhan (1975, 1976) have shown that in the i.i.d. case (3.1) (or (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 3.1 together with Lemma 2.4.

REMARK 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 3.1 remains valid if (3.4) is replaced by the weaker condition

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

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 $\Pr\{\hat{P}_n \in \Omega\} = \Pr\{\hat{P}_n \in \Omega \cap \Lambda_B\}$).

A convex set $\Omega \subset \Lambda$ will be called *strongly τ-convex* if for each $Q \in \text{cl}_\tau(\Omega)$ and each $R \in \text{int}_\tau(\Omega)$ it holds that $\{\alpha Q + (1 - \alpha)R : 0 < \alpha < 1\} \subset \text{int}_\tau(\Omega)$. If $\text{int}_\tau(\Omega) = \varnothing$, convexity and strong τ-convexity are equivalent.

COROLLARY 3.1. *Let $P \in \Lambda$ and let Ω be a subset of Λ. Then* (3.3) *holds if one of the following conditions is satisfied*:

(i) Ω *is strongly τ-convex and* $K(\mathrm{int}_\tau(\Omega), P) < \infty$;

(ii) Ω *is weakly open*, $P(\mathrm{supp}(P)) = 1$ *and* $K(\Omega, P) = \infty$.

PROOF. First suppose that condition (i) is satisfied. Fix $\varepsilon > 0$. Let $Q \in \mathrm{cl}_\tau(\Omega)$ be such that $K(Q, P) < K(\mathrm{cl}_\tau(\Omega), P) + \varepsilon$ and let $R \in \mathrm{int}_\tau(\Omega)$ be such that $K(R, P) < \infty$. Define $Q_\alpha = \alpha Q + (1 - \alpha)R$, $0 < \alpha < 1$. Then $Q_\alpha \in \mathrm{int}_\tau(\Omega)$ and the convexity of the map $Q' \to K(Q', P)$ implies

$$K(\mathrm{int}_\tau(\Omega), P) \leqslant \lim_{\alpha\uparrow 1} K(Q_\alpha, P) \leqslant \lim_{\alpha\uparrow 1}\{\alpha K(Q, P) + (1 - \alpha)K(R, P)\}$$
$$= K(Q, P) < K(\mathrm{cl}_\tau(\Omega), P) + \varepsilon.$$

Hence $K(\mathrm{int}_\tau(\Omega), P) = K(\mathrm{cl}_\tau(\Omega), P)$. Application of Theorem 3.1 completes the proof of this case.

Next suppose that condition (ii) holds. By Lemma 2.5 the weak closure of $\{Q \in \Lambda : K(Q, P) < \infty\}$ contains all empirical pms with their (finite) support contained in the support of P. Since $P(\mathrm{supp}(P)) = 1$, $\Pr\{\hat{P}_n \in \Omega\} = 0$ for all $n \in \mathbb{N}$ implying (3.3). □

Condition (i) of the preceding corollary cannot be replaced by the condition that Ω is τ-open, convex and $K(\mathrm{int}_\tau(\Omega), P) < \infty$. To see this, let $S = [0, 1]$, let P be Lebesgue measure on S and let $\Omega = \{Q \in \Lambda : Q((0, \frac{1}{2})) > \frac{3}{4}\} \cup \{Q \in \Lambda : Q$ 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 $K(\Omega, P) = \frac{3}{4}\log 3 - \log 2 > 0$.

However, in the case that Ω is *weakly* open and convex Bahadur and Zabell (1979) have proved (3.3) assuming that S is a Polish space. Their proof is based on a Chernoff-type theorem for sample means.

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 nonempty 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 $\Omega \cap \{Q \in \Lambda : K(Q, P) \leqslant K(\Omega, P) + \eta\}$ 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

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

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 $\kappa(r_m) \leqslant \kappa(r) < \infty$ 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 $\{Q \in \Lambda : T(Q) \geqslant t$ and $K(Q, P) \leqslant M\}$ is τ-closed for each $t \in \mathbb{R}$ and $M \geqslant 0$). Since $K(Q_m, P) \leqslant \kappa(r)$ 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 $K(Q, P) \leqslant \liminf_{j\to\infty} K(Q_{m_j}, P) < \infty$. It follows that $T(Q) \geqslant r$ since T is upper semicontinuous on Γ and since $T(Q_{m_j}) \geqslant r_{m_j}$ for each $j \in \mathbb{N}$. Hence $Q \in \Omega_r$ and $\kappa(r) \leqslant K(Q, P) \leqslant \lim_{j\to\infty} K(Q_{m_j}, P) = \lim_{m\to\infty}\kappa(r_m) \leqslant \kappa(r)$. 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 $P \in \Lambda$ 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* $\lim_{n\to\infty} u_n = 0$,

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

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

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

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

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

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

Thus

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

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

REMARK 3.3. 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.

EXAMPLE 3.1. Let $\mathfrak{F}$ 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 nonatomic for each $f \in \mathfrak{F}$. Let $d(Qf^{-1}, Rf^{-1})$ be the distance between Qf^{-1} and Rf^{-1} defined in Lemma 2.1.

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

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

where

$$\begin{aligned} \kappa(\varepsilon) = \min_{0<p\leqslant 1-\varepsilon}\{&(p+\varepsilon)\log((p+\varepsilon)/p) \\ &+ (1-p-\varepsilon)\log((1-p-\varepsilon)/(1-p))\}. \end{aligned}$$

Here we prove that the function $T : \Lambda \to \mathbb{R}$ defined by $T(Q) = \sup_{f\in\mathfrak{F}} d(Qf^{-1}, Pf^{-1})$ is τ-continuous at each $Q \in \Gamma$ 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 $f_U \in \mathfrak{F}$ can be found satisfying

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

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

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

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

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 $R(S \setminus K) < \frac{1}{4}\varepsilon$ if $R \in U_0$. Since f is a cluster point of the net $\{f_U : U \in \mathfrak{D}\}$ 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

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

This contradicts (3.8) and hence T is τ-continuous at Q. Let $\Omega_\varepsilon = \{Q \in \Lambda : T(Q) \geqslant \varepsilon\}$ 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 nonatomic 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 $X_{i,1}, \cdots, X_{i,n_i}$ be i.i.d. random variables taking values in S according to a pm $P_i \in \Lambda$, $1 \leqslant i \leqslant k$, 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 $\nu_i > 0$, $1 \leqslant i \leqslant k$. (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 ith sample will be denoted by $\hat{P}_{i,n_i}$, $1 \leqslant i \leqslant k$. Λ is endowed with the topology τ and Λ^k is given the product topology.

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

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

and

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

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

COROLLARY 3.2. *Let* $P = (P_1, \cdots, P_k) \in \Lambda^k$ *and* $\Omega \subset \Lambda^k$ *satisfy*

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

Then

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

COROLLARY 3.3. *Let* $P = (P_1, \cdots, P_k) \in \Lambda^k$, *let* $T : \Lambda^k \to \overline{\mathbb{R}}$ *be continuous at each* $Q \in \Gamma = \{R \in \Lambda^k : I_\nu(R, P) < \infty\}$ *and let* $\Omega_t = \{Q \in \Lambda^k : T(Q) \geqslant t\}$, $t \in \mathbb{R}$. *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* $u_N \to 0$,

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

4. Linear functions of empirical probability measures. 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 $T(\hat{P}_n)$, 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., $T(\alpha Q + (1 - \alpha)R) = \alpha T(Q) + (1 - \alpha)T(R)$, $0 \leqslant \alpha \leqslant 1$. 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 carrier 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 $\Pr\{\hat{P}_n \in \Gamma | B\}$ to denote $\Pr\{\hat{P}_n \in \Gamma | X_i \in B, 1 \leqslant i \leqslant n\}$.

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

$$K(Q, P) \leqslant K(Q, P_m) + \varepsilon \qquad \text{for all } Q \in \Lambda \text{ and } m \geqslant m_0.$$

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 $K(\Omega, P) \leqslant \liminf_{m\to\infty} K(\Omega, P_m)$. To prove the lemma it still must be shown that conversely

$$K(\Omega, P) \geqslant \limsup_{m\to\infty} K(\Omega, P_m). \tag{4.1}$$

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

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

implying (4.1). □

THEOREM 4.1. *Let $P \in \Lambda$, let E be a real Hausdorff topological vector space and let $B_1 \subset B_2 \subset \cdots$ 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

$$\lim_{n\to\infty} n^{-1} \log \Pr\{T(\hat{P}_n) \in A\} = -K(T^{-1}(A), P). \tag{4.2}$$

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

$$K(T^{-1}(A^0), P_m) = K(T^{-1}(\bar{A}), P_m) \qquad \text{for each } m \in \mathbb{N}. \tag{4.3}$$

Fix $\varepsilon > 0$ and $m \in \mathbb{N}$. There exists a pm $Q \in T^{-1}(\bar{A})$ which satisfies $K(Q, P_m) < K(T^{-1}(\bar{A}), P_m) + \varepsilon$ and 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 $Q, R \in \Psi_m$, T is linear on Ψ_m and E is a topological vector space, $Q_\alpha \in T^{-1}(A^0)$ for each α. Proceeding as in the proof of Corollary 3.1 one obtains (4.3).

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 $\Psi_m^* \cap T^{-1}(\bar{A}) \supset \Psi_m^* \cap \mathrm{cl}_{\tau_m}(\Omega \cap \Psi_m) \supset \Psi_m^* \cap \mathrm{int}_{\tau_m}(\Omega \cap \Psi_m) \supset \Psi_m^* \cap T^{-1}(A^0)$. Hence, by (4.3)

$$K(\mathrm{cl}_{\tau_m}(\Omega \cap \Psi_m), P_m) = K(\mathrm{int}_{\tau_m}(\Omega \cap \Psi_m), P_m) \qquad \text{for each } m \in \mathbb{N}. \tag{4.4}$$

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 $k^{-1} \log \Pr\{T(\hat{P}_k) \in A\} \geqslant \gamma - \varepsilon$. Since $\lim_{m\to\infty} \Pr\{T(\hat{P}_k) \in A | B_m\} = \Pr\{T(\hat{P}_k) \in A\}$

there exists $m_0 \in \mathbb{N}$ such that

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

Hence for $m \geqslant m_0$

$$(4.5) \quad \limsup_{n\to\infty} n^{-1}\log \Pr\{T(\hat{P}_n) \in A | B_m\} \geqslant \lim_{j\to\infty}(kj)^{-1}\log(\Pr\{T(\hat{P}_k) \in A | B_m\})^j = k^{-1}\log \Pr\{T(\hat{P}_k) \in A | B_m\} \geqslant \gamma - 2\varepsilon.$$

The first inequality in (4.5) 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 $X_{(i-1)k+1}, \cdots, X_{ik}$, $1 \leqslant i \leqslant j$.

By (4.4), Theorem 3.1 and Remark 3.2

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

Lemma 4.1 and (4.5) now imply

$$\gamma - 2\varepsilon \leqslant \lim_{m\to\infty}\lim_{n\to\infty} n^{-1}\log \Pr\{T(\hat{P}_n) \in A | B_m\} = -\lim_{m\to\infty} K(T^{-1}(A), P_m) = -K(T^{-1}(A), P).$$

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

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

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

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

$$\liminf_{n\to\infty} n^{-1}\log \Pr\{T(\hat{P}_n) \in A\} \geqslant \lim_{m\to\infty}\left[\liminf_{n\to\infty} n^{-1}\log \Pr\{T(\hat{P}_n) \in A | B_m\} + \log P(B_m)\right] = \lim_{m\to\infty} - K(T^{-1}(A), P_m) = -K(T^{-1}(A), P). \qquad \square$$

COROLLARY 4.1. *In Theorem* 4.1 *let* $T|\Psi_n$ *be linear and weakly continuous for each* $n \in \mathbb{N}$. *Then* (4.2) *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 4.1 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 $U = \cup_{x \in A_m} U_x$. 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

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

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 $\Pr\{T(\hat{P}_n) \in A | B_m\} = 0$ 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 (4.2).

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 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. □

Consider the particular case that S is a locally convex (Hausdorff) topological vector space. Bahadur and Zabell (1979) have shown that for each convex open set $A \subset S$

$$\lim_{n\to\infty} n^{-1} \log \Pr\{n^{-1}\Sigma_{i=1}^{n} X_i \in A\} \tag{4.6}$$

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

Another version of their result can also be deduced from Corollary 4.1. 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 $y \in E$, $y' \in E'$, $y'^* \in E'^*$ write $\langle y, y' \rangle = y'(y)$ and $\langle y'^*, y' \rangle = y'^*(y')$. Finally, let B be a *compact* subset 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 $f: B \to E$. Then the *integral* of $f \in C(B)$ with respect to a pm $Q \in \Psi_B$, denoted by $\int_B f dQ$ or $\int_B f(x) dQ(x)$, is an element of E'^* defined by the relation

$$\langle \textstyle\int_B f dQ, y' \rangle = \int_B \langle f(x), y' \rangle dQ(x)$$

for each $y' \in E'$ (cf. Bourbaki (1965), pages 74–82).

Let $\hat{E}$ be the completion of E induced by the uniformity compatible with the topology on E. Each element $y \in \hat{E}$ can be identified with an element of E'^* by identifying y with the linear form $y' \to \langle y, y' \rangle$ on E' (where E' is identified with $\hat{E}'$).

With this identification we have the following two fundamental properties of the integral for each $f \in C(B)$: (a) the closure of the convex hull of $f(B)$ in $\hat{E}$ is equal to the set $\{\int_B f dQ\colon Q \in \Psi_B\}$, and (b) the map $\phi : Q \to \int_B f dQ$ is the unique weakly continuous linear mapping from Ψ_B into $\hat{E}$ such that $\phi(Q) = \Sigma_{i=1}^k f(x_i)Q(\{x_i\})$ for each pm Q with finite support $\{x_1, \cdots, x_k\}$ (cf. Bourbaki (1965), loc. cit.).

Now suppose $P \in \Lambda$ is tight. Then there exists an increasing sequence of compact subsets $B_1 \subset B_2 \subset \cdots$ of S such that $\lim_{n\to\infty} P(B_n) = 1$. In the notation of Theorem 4.1 define $T : \Lambda^* \to \hat{E}$ by

$$T(Q) = \int_{B_m} x dQ(x), \qquad Q \in \Psi_m m \in \mathbb{N}. \tag{4.7}$$

Since $\int_{B_m} \langle x, y' \rangle dQ(x) = \int_{B_n} \langle x, y' \rangle dQ(x)$ 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, $T(\hat{P}_n) = n^{-1}\Sigma_{i=1}^n X_i$ and $T|\Psi_n$ is linear and weakly continuous for each $n \in \mathbb{N}$. Hence Corollary 4.1 implies that (4.6) exists and is equal to $-K(T^{-1}(A), P)$ under the conditions (i), (ii) or (iii). Apart from pms with noncompact support, the set $T^{-1}(A)$ coincides with the set $M(A)$ appearing in the result of Bahadur and Zabell (1979).

Note that we required that $S = \hat{E}$ and P is tight. If in addition P is convex-tight, i.e., $\sup\{P(K)\colon K$ compact and convex$\} = 1$, 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 4.2. *Let S be a locally convex (Hausdorff) topological vector space and let $P \in \Lambda$ be convex-tight. Then the limit* (4.6) *exists and equals* $-K(T^{-1}(A), P)$, *with T defined by* (4.7), *if $A \in \mathfrak{B}$ and P satisfy 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.*

5. A d-dimensional Chernoff-type theorem. Consider d-dimensional i.i.d. random variables $X_1, X_2, \ldots$ taking values in $S = \mathbb{R}^d$ $(d \geq 1)$. Let $\Lambda^* = \{Q \in \Lambda\colon Q$ has compact support$\}$. In this section the map $T : \Lambda^* \to \mathbb{R}^d$ is defined 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 $d = 1$

$$\lim_{n\to\infty} n^{-1}\log \Pr\{n^{-1}\Sigma_{i=1}^n X_i \geq r\} = -\sup_{t\geq 0}\{\text{tr} - \log \int_{\mathbb{R}} e^{tx} dP(x)\}$$

for any $r \in \mathbb{R}$ and $P \in \Lambda_1$ (as was noted in the previous section, the sample mean $n^{-1}\Sigma_{i=1}^n X_i$ is equal to $T(\hat{P}_n)$). With the help of Corollary 4.2 we shall generalize this theorem to the case $d > 1$.

Related results about limits of the form (4.6) have been obtained by Lanford (1972) who considered open convex sets $A \subset \mathbb{R}^d$ and more recently by Bartfai (1977) who assumed $A \subset \mathbb{R}^d$ to be open and the moment generating function of X_1 to be finite in a neighborhood of the origin.

Our results are in a certain sense complementary to those of Sievers (1975), who gives sufficient conditions to reduce limits of the form $\lim_{n\to\infty} n^{-1}\log \Pr\{T_n \in A\}$, $A \subset \mathbb{R}^d$, $A \in \mathfrak{B}$ to limits of the form $\lim_{n\to\infty} n^{-1}\log \Pr\{T_n^{(1)} * x_1, \cdots, T_n^{(d)} * x_d\}$, where the $*$'s are either $\geqslant$ or $\leqslant$ and $T_n = (T_n^{(1)}, \cdots, T_n^{(d)})$ 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 ith 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: $x \geqslant y$ iff $x^{(i)} \geqslant y^{(i)}(1 \leqslant i \leqslant d)$ and $x > y$ iff $x^{(i)} > y^{(i)}(1 \leqslant i \leqslant d)$. Furthermore $\mathbb{R}_+^d = \{x \in \mathbb{R}^d : x \geqslant 0\}$. 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

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

and

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

With these notations the following theorem will be proved.

THEOREM 5.1. *Let $P \in \Lambda$ and $r \in (\partial A_P)^c$. Then, for each sequence $\{u_n\}$ in $\mathbb{R}^d$ such that* $\lim_{n\to\infty} u_n = 0$,

$$\lim_{n\to\infty} n^{-1}\log \Pr\{n^{-1}\textstyle\sum_{i=1}^n X_i \geqslant r + u_n\} = -K(\Omega_r, P) \tag{5.1}$$

and

$$K(\Omega_r, P) = \sup_{t\in\mathbb{R}_+^d}\{t'r - \log \textstyle\int_{\mathbb{R}^d} e^{t'x} dP(x)\}. \tag{5.2}$$

Moreover, the supremum on the right-hand side of (5.2) *is achieved if* $r \in A_P^0$.

Theorem 5.1 generalizes Chernoff's theorem to d-dimensional vectors, but does not cover the case $r \in \partial A_P$. Relation (5.2) extends results by Hoeffding (1967) and Csiszár (1975, Theorem 3.3) who both considered sets Ω_r of the type $\{Q \in \Lambda : \int_{\mathbb{R}^d} x dQ(x) = r\}$ assuming finiteness of the moment generating function of P in a neighborhood of the origin.

The following example demonstrates that (5.1) may fail if r is a boundary point of A_P.

EXAMPLE 5.1. 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 $\Pr\{n^{-1}\sum_{i=1}^n X_i \geqslant r\} = \binom{n}{\frac{1}{2}n} 2^{-2}$ for n even and $= 0$ for n odd, the limit in the left-hand member of (5.1), 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 5.2. *Let $P \in \Lambda$ and $r \in \partial A_P$. Then there exist a hyperplane $\mathbb{H}_s(r) = \{x \in \mathbb{R}^d : s'x = s'r\}$ through r and a corresponding half-space $\mathbb{H}_s^*(r) = \{x \in \mathbb{R}^d : s'x > s'r\}$, 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* (5.1) *and* (5.2) *hold;*
(iv) *if $P(\mathbb{H}_s(r)) = P(\{r\}) > 0$, then* (5.1) *and* (5.2) *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 5.2 reduces to the point $\{r\}$ and either (iii) or (iv) is satisfied. Hence Theorems 5.1 and 5.2 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 5.2 cannot occur and (iii) holds. Hence Theorems 5.1 and 5.2 yield

COROLLARY 5.1. *Let $P \in \Lambda$ and suppose P is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^d$. Then* (5.1) *and* (5.2) *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

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

and

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

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

Before proving the theorems we first establish two lemmas.

LEMMA 5.1. *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 5.2. *Let Γ be a nonempty convex subset of Λ^* and let $p \in \mathbb{R}^d$. Consider the system of d inequalities*

$$\int_{\mathbb{R}^d} x\,dQ(x) > p. \tag{5.3}$$

Then either there is a solution $Q \in \Gamma$ of (5.3) *or, alternatively, there exists $t \in \mathbb{R}_+^d$, $t \neq 0$, such that*

$$t'\int_{\mathbb{R}^d} x\,dQ(x) \leq t'p \qquad \text{for all } Q \in \Gamma. \tag{5.4}$$

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

PROOF OF THEOREM 5.1. Since any pm on the Borel sets of $\mathbb{R}^d$ is tight, Corollary 4.2 implies that (5.1) is satisfied if $r \in A_P^0$ and $u_n = 0$ for all $n \in \mathbb{N}$. By a similar argument as we used in proving Theorem 3.2, more general sequences $\{u_n\}$ may be dealt with. If $r \in (\bar{A}_P)^c$ then obviously $K(\Omega_r, P) = \infty$. Moreover, choosing $r^0 \in A_P^c$ such that $r + u_n > r^0$ for all large n, $\Pr\{n^{-1}\Sigma_{i=1}^n X_i \geqslant r + u_n\} \leqslant \Pr\{n^{-1}\Sigma_{i=1}^n X_i > r^0\}$ and another application of Corollary 4.2 (invoking condition (ii)) yields (5.1) in this case too.

We proceed to prove (5.2). 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

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

and hence

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

It still must be shown that conversely

$$K(\Omega_r, P) \leqslant \sup_{t \in \mathbb{R}_+^d}\left\{t'r - \log \int_{\mathbb{R}^d} e^{t'x}\,dP(x)\right\}. \tag{5.5}$$

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

$$K(\bar{Q}, P) = K(\Omega_r, P). \tag{5.6}$$

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

$$K(\Omega_t, P) \geqslant K(\Omega_r, P) + s'(t - r) \qquad \text{for all } t \in A_P. \tag{5.7}$$

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

$$K(Q, P) = s'\int_{\mathbb{R}^d} x\,dQ(x) - \log \beta(s). \tag{5.8}$$

Application of (5.6) and (5.7), with $t = \int_{\mathbb{R}^d} x\,dQ(x)$, yields

$$K(Q, P) \geqslant K(\bar{Q}, P) + s'\left(\int_{\mathbb{R}^d} x\,dQ(x) - r\right). \tag{5.9}$$

Since

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

we have by (5.8) and (5.9)

$$(5.10)\quad \begin{aligned} K(\overline{Q}, Q) &= K(\overline{Q}, P) - s'\textstyle\int_{\mathbb{R}^d} x\, d\overline{Q}(x) + \log \beta(s) \\ &= K(\overline{Q}, P) - K(Q, P) + s'\left(\textstyle\int_{\mathbb{R}^d} x\, dQ(x) - \int_{\mathbb{R}^d} x\, d\overline{Q}(x)\right) \\ &\leqslant s'\left(r - \textstyle\int_{\mathbb{R}^d} x\, d\overline{Q}(x)\right) \leqslant 0. \end{aligned}$$

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

$$(5.11)\qquad K(\Omega_r, P) = K(Q, P) = s'r - \log \textstyle\int_{\mathbb{R}^d} e^{s'x}\, dP(x).$$

This proves (5.5) for P with compact support. (We note in passing that, by (5.10), $s^{(i)} > 0$ 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 (5.11) is indeed satisfied. However, the present proof seems to be more direct.

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 (5.7) $s_m \in \mathbb{R}^d_+$ satisfying

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

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

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

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, (5.11), and Fatou's lemma

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

Thus (5.5) is proved in this case too and (5.2) follows for $r \in A^0_P$.

It remains to prove (5.2) in the case $r \in (A^c_P)^0$. Let $p \in (A^c_P)^0$, $p < r$. Apply Lemma 5.2 with $\Gamma = \{Q \in \Lambda^* : K(Q, P) < \infty\}$. Since (5.3) does not hold, there exists $s \in \mathbb{R}^d_+$, $s \neq 0$, such that

$$(5.12)\qquad s'\textstyle\int_{\mathbb{R}^d} x\, dQ(x) \leqslant s'p \qquad \text{for all } Q \in \Gamma.$$

It follows that (with the notation of Theorem 5.2)

$$(5.13)\qquad P(\mathbb{H}^*_s(p)) = 0.$$

For suppose that (5.13) 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 $Q(B) = P(B \mid A)$, $B \in \mathcal{B}$. Then $K(Q, P) = -\log P(A) < \infty$ and $s'\int_{\mathbb{R}^d} x\, dQ(x) > s'p$, in contradiction to (5.12). Hence

$$\sup_{t\in\mathbb{R}^d_+}\{t'r - \log \int_{\mathbb{R}^d} e^{t'x}dP(x)\}$$
$$\geqslant \lim_{\alpha\to\infty} - \log \int_{\mathbb{R}^d}\exp\{\alpha s'(x-p) + \alpha s'(p-r)\}dP(x)$$
$$= \infty = K(\Omega_r, P)$$

and the proof of Theorem 5.1 is complete. □

PROOF OF THEOREM 5.2. Let $r \in \partial A_P$ and put $\Gamma = \{Q \in \Lambda^* : K(Q, P) < \infty\}$. Applying Lemma 5.2 with $p = r$, (5.3) is obviously not satisfied and hence there exists $s \in \mathbb{R}^d_+$, $s \neq 0$, such that

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

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 (5.13) from (5.12). 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 $s'\int_{\mathbb{R}^d} x\,dQ(x) \leqslant s'p$, 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 $\int_{\mathbb{R}^d} x\,dQ(x) \geqslant r$. Hence $Q(\mathbb{H}_s(r) \cup \mathbb{H}_s^*(r)) > 0$ and therefore, as a consequence of $Q \ll P$ 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 $P(\mathbb{H}_s^*(r) \cup \mathbb{H}_s(r)) = 0$,

$$\sup_{t\in\mathbb{R}^d_+}\{t'r - \log \int_{\mathbb{R}^d} e^{t'x}dP(x)\}$$
$$\geqslant -\lim_{\alpha\to\infty}\log \int_{\mathbb{R}^d}\exp\{\alpha s'(x-r)\}dP(x) = \infty$$

by dominated convergence and (5.2) is proved. Finally, by Markov's inequality, for any $t \in \mathbb{R}^d_+$ and $u_n \in \mathbb{R}^d$,

$$\Pr\{n^{-1}\Sigma_{i=1}^n X_i \geqslant r + u_n\} \leqslant \Pr\{\Sigma_{i=1}^n t'X_i \geqslant nt'(r+u_n)\}$$
$$\leqslant E\exp\{\Sigma_{i=1}^n t'X_i\}/\exp\{nt'(r+u_n)\}$$
$$= \left(\int_{\mathbb{R}^d}\exp\{t'(x - r - u_n)\}dP(x)\right)^n.$$

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

$$\lim_{n\to\infty} n^{-1}\log \Pr\{n^{-1}\Sigma_{i=1}^n X_i \geqslant r + u_n\}$$
$$\leqslant -\sup_{t\in\mathbb{R}^d_+}\left(-\log \int_{\mathbb{R}^d}\exp\{t'(x-r)\}dP(x)\right) = -\infty$$

and (5.1) 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 $\Pr\{n^{-1}\Sigma_{i=1}^n X_i \geqslant r\} = \Pr\{X_i = r, 1 \leqslant i \leqslant n\} = \gamma^n$ and hence $\lim_{n\to\infty} n^{-1}\log \Pr\{n^{-1}\Sigma_{i=1}^n X_i \geqslant r\} = \log\gamma$, proving (5.1) for $u_n = 0$. By dominated convergence

$$\sup_{t\in\mathbb{R}^d_+}\{t'r - \log \int_{\mathbb{R}^d} e^{t'x}dP(x)\}$$
$$\geqslant -\lim_{\alpha\to\infty}\log \int_{\mathbb{R}^d}\exp\{\alpha s'(x-r)\}dP(x) = -\log\gamma.$$

The reverse inequality is obtained by Markov's inequality, as in the last lines of the proof of (iii). Thus (5.2) is also established and the proof of the theorem is complete. □

6. Linear combinations of order statistics. In this section $X_1, X_2, \cdots$ are real-valued i.i.d. random variables with distribution function (df)F. Instead of Λ_1, the set of pms on $(\mathbb{R}, \mathfrak{B})$, we shall consider the set D of one dimensional df's. If $G \in D$, the corresponding pm in Λ_1 will be denoted by P_G. A set of df's 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 df's 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 $G^{-1}(u) = \inf\{x \in \mathbb{R} : G(x) \geqslant u\}$.

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

$$T(\hat{F}_n) = \int_0^1 J(u) \hat{F}_n^{-1}(u) du, \tag{6.1}$$

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

$$T(\hat{F}_n) = \Sigma_{i=1}^n c_{n,i} X_{i:n}, \tag{6.2}$$

where $c_{n,i} = \int_{(i-1)/n}^{i/n} J(u) du$ and $X_{i:n}$ is the ith order statistic of $X_1, \cdots, X_n$. These statistics are sometimes called 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

$$\Omega_t = \left\{G \in D : \int_0^1 J(u) G^{-1}(u) du \geqslant t,\ \int_0^1 |J(u) G^{-1}(u)| du < \infty\right\}, \tag{6.3}$$

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

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

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

$$\lim_{n\to\infty} n^{-1} \log \Pr\{T(\hat{F}_n) \geqslant r + u_n\} = -K(\Omega_r, F) \tag{6.4}$$

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

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

(a) *$J \geqslant 0$ 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,* $J \geqslant 0$ *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* (6.4) *holds provided conditions* (ii) *and* (c) *are satisfied and* $u_n \leqslant 0$ *for all large* $n \in \mathbb{N}$.

REMARK 6.1. Condition (ii) of Theorem 6.1 is satisfied if P_F has compact support or if $0 < \alpha < \beta < 1$.

REMARK 6.2. The second part of Theorem 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 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

$$T(G) = \int_0^1 J(u) 1_B(G^{-1}(u))G^{-1}(u)du, \qquad G \in D. \tag{6.5}$$

The function T is ρ-continuous. For a proof consider a sequence of df's $\{G_n\}$, 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 (6.4) 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 $P_G \ll P_F$, 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

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

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

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

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 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 $G(r - 0) \leq \alpha$ 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

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

Then $G_\varepsilon(r - 0) = G(r - 0) - \varepsilon \leq \alpha - \varepsilon$, 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 $K(\text{int}_\tau(\Omega_r), F) \leq K(G_\varepsilon, F)$, for each $\varepsilon > 0$. Since $\lim_{\varepsilon\downarrow 0} K(G_\varepsilon, F) = K(G, F)$, we have $K(\text{int}_\tau(\Omega_r), F) \leq \lim_{\varepsilon\downarrow 0} K(G_\varepsilon, F) = K(\Omega_r, F)$, 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 3.1 yields that (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 3.3) implies that (6.4) also holds if $u_n \leq 0$ 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 $K(G', F) \leq K(G, F)$, 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 $\limsup_{n\to\infty} n^{-1} \cdot \log \Pr\{\hat{F}_n \in \Omega_r\} \leq \log P_F([r, \infty))$. It is clear that conversely $\liminf_{n\to\infty} n^{-1} \cdot \log \Pr\{\hat{F}_n \in \Omega_r\} \geq \liminf_{n\to\infty} n^{-1} \log \Pr\{X_{1:n} \geq r\} = \log P_F([r, \infty))$. Thus (6.4) holds provided $u_n = 0$ for all large $n \in \mathbb{N}$. By the same argument as before (6.4) also holds if $u_n \leq 0$ for all large $n \in \mathbb{N}$. □

REMARK 6.3. The continuity of a function which is essentially equivalent to the function T in (6.5) 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 df's can be tackled by the methods of this paper. Note that Hoadley's (1967) Theorem 1 would not suffice to prove (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 6.1 is not immediately applicable, but the next theorem may be of use.

THEOREM 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* (6.3) *for $t \in \mathbb{R}$. Then, for each sequence of real numbers $\{u_n\}$ such that* $\lim_{n\to\infty} u_n = 0$,

$$(6.6)\qquad \lim_{n\to\infty} n^{-1}\log \Pr\left\{\int_0^1 J_n(u)\hat{F}_n^{-1}(u)du \geq r + u_n\right\} = -K(\Omega_r, F)$$

if J, F, α and β satisfy conditions (i) *and* (ii) *of Theorem* 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. In accordance with Section 5 we write $B_m = [-m, m]$ and denote by G_m the conditional df defined by

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

Let $D^* = \{G \in D : P_G(B_m) = 1 \text{ for some } m \in \mathbb{N}\}$. By condition (i) there exist 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 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 $n \geq N_0$

$$(6.7)\qquad \left|\int_0^1 J_n(u)\hat{F}_n^{-1}(u)du - \int_0^1 J(u)\hat{F}_n^{-1}(u)du\right| \leq m\int_0^1 |J_n(u) - J(u)|du < \tfrac{1}{4}\varepsilon$$

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 $m \in \mathbb{N}$:

$$\begin{aligned}
&\liminf_{n\to\infty} n^{-1}\log \Pr\left\{\int_0^1 J_n(u)\hat{F}_n^{-1}(u)du \geq r + u_n\right\}\\
&\quad\geq \liminf_{n\to\infty} n^{-1}\log \Pr\left\{\int_0^1 J_n(u)\hat{F}_n^{-1}(u)du \geq r + u_n\,|\,B_m\right\} + \log P_F(B_m)\\
&\quad\geq \liminf_{n\to\infty} n^{-1}\log \Pr\left\{\int_0^1 J(u)\hat{F}_n^{-1}(u)du \geq r + \tfrac{1}{2}\varepsilon\,|\,B_m\right\} + \log P_F(B_m)\\
&\quad\geq -K(\Omega_{r+\varepsilon}, F_m) + \log P_F(B_m).
\end{aligned}$$

The last inequality holds by Theorem 6.1, 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

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

Hence by condition (i)

$$\liminf_{n\to\infty} n^{-1}\log \Pr\{\textstyle\int_0^1 J_n(u)\hat{F}_n^{-1}(u)du \geq r + u_n\} \geq -K(\Omega_r, F). \tag{6.8}$$

Next we show that conversely

$$\limsup_{n\to\infty} n^{-1}\log \Pr\{\textstyle\int_0^1 J_n(u)\hat{F}_n^{-1}(u)du \geq r + u_n\} \leq -K(\Omega_r, F). \tag{6.9}$$

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:

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

since (6.7) holds again for large n if $\hat{F}_n^{-1}(u) \in B_m$ for $u \in (\alpha, \beta)$. This result implies (6.9) by Theorem 6.1 and Lemma 3.3 (also if $K(\Omega_r, F) = \infty$) and the present theorem follows from (6.8) and (6.9). □

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

$$T_n = (n - 2[\alpha n])^{-1}\Sigma_{i=[\alpha n]+1}^{n-[\alpha n]} X_{i:n}, \qquad n \in \mathbb{N}, \tag{6.10}$$

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 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* (6.10). *Then, for each sequence $\{u_n\}$ such that* $\lim_{n\to\infty} u_n = 0$,

$$\lim_{n\to\infty} n^{-1}\log \Pr\{T_n \geq r + u_n\} = -K(\Omega_r^\alpha, F), \tag{6.11}$$

where

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

If F is discontinuous at r, then (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 $J_n = n \cdot (n - 2[\alpha n])^{-1} 1_{A_n}$, 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 (6.11) follows since in this case (c) and (d) of Theorem 6.1 and hence the conditions of Theorem 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 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 $\varepsilon > 0$ and let $\Omega_{r,n} = \{H \in D : \int_0^1 J_n(u)H^{-1}(u)du \geqslant r\}$, $n \in \mathbb{N}$. 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 $G_\delta(r-0) \leqslant \alpha - \delta\alpha$ and $G_\delta(r) \geqslant 1 - \alpha + \delta\alpha$. 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 $u_n \leqslant 0$ for all large n, the same arguments that were used in the last part of the proof of Lemma 3.1 yield

$$\Pr\{T_n \geqslant r + u_n\} \geqslant \Pr\{\hat{F}_n \in \Omega_{r,n}\} \geqslant \Pr\{d_{\mathcal{P}}(P_{\hat{F}_n}, P_{G_n}) = 0\}$$
$$\geqslant \exp\{-n(K(\Omega_r^\alpha, F) + \varepsilon + o(1))\}$$

as $n \to \infty$, implying

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

On the other hand (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 6.2 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 (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

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

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 (5.2) in Theorem 5.1. Let

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

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

$$K(\Omega_r^\alpha, F) = \inf\{K(\Omega_r^\alpha(a, b), F) : 0 < F(a) < F(b) < 1, \quad F(b) > F(r)\}. \tag{6.12}$$

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

$$s = 0 \qquad \text{if } \int_a^b x dF(x)/(F(b) - F(a)) \geqslant r$$
$$= \phi^{-1}(r) \quad \text{otherwise,}$$

where $\phi(t) = \int_a^b x e^{tx}dF(x)/\int_a^b e^{tx}dF(x)$, $t \geqslant 0$. 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 $\phi'(t) \ngtr 0$ for all $t \geqslant 0$.

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

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

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

$$\begin{aligned} K(G, F) = 2\alpha\log\alpha + (1-2\alpha)\log(1-2\alpha) - \alpha\log F(a) - \alpha\log(1-F(b)) \\ + (1-2\alpha)sr - (1-2\alpha)\log\textstyle\int_a^b e^{sx}dF(x). \end{aligned}$$

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

$$\begin{aligned} &sr - \log\{(1-2\alpha)^{-1}\textstyle\int_a^b e^{sx}dF(x)\} \\ &\quad \leqslant sr - \log\{(1-2\alpha)^{-1}\textstyle\int_a^b \exp(sx - \log h(x))dH(x)\} \\ &\quad \leqslant s\{r - (1-2\alpha)^{-1}\textstyle\int_a^b x\,dH(x)\} + (1-2\alpha)^{-1}\textstyle\int_a^b \log h(x)dH(x). \end{aligned}$$

Hence

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

Similarly, by Jensen's inequality,

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

and

$$\begin{aligned} \textstyle\int_b^\infty \log h(x)dH(x) &\geqslant (1-H(b))\log\{(1-H(b))/(1-F(b))\} \\ &= \alpha\log\{\alpha/(1-F(b))\}. \end{aligned}$$

Thus

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

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

Now define the functions

$$\begin{aligned} f_\alpha(a, b) = (1-2\alpha)s(a, b)r - \alpha\log F(a) - \alpha\log(1-F(b)) \\ - (1-2\alpha)\log\textstyle\int_a^b \exp(s(a, b)x)dF(x) \end{aligned}$$

and

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

Then, by (6.12)

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

REMARK 6.4. We briefly indicate another route to the result (6.11). Let T_n be defined by (6.10) and let $n_\alpha = n - 2[\alpha n]$, for each $n \in \mathbb{N}$. Then we may write

$$E\exp(n_\alpha tT_n) = E\left(E\left\{\exp\left(t\textstyle\sum_{i=[\alpha n]+1}^{n-[\alpha n]}X_{i:n}\right)\middle| X_{[\alpha n]:n}, X_{n-[\alpha n]+1:n}\right\}\right).$$

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

$$(6.14)\quad \lim_{n\to\infty} n^{-1}\log E\exp(n_\alpha t(T_n - r)) = -\inf_{-\infty<a<b<\infty}\{(1-2\alpha)tr - \alpha\log F(a) - \alpha\log(1-F(b)) - (1-2\alpha)\log \textstyle\int_a^b \exp(tx)f(x)dx\}.$$

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

$$(6.15)\quad \lim_{n\to\infty} n^{-1}\log \Pr\{T_n \geqslant r\} = -\inf_{t\geqslant 0}\lim_{n\to\infty} n^{-1}\log E\exp(n_\alpha t(T_n - r)),$$

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

By (6.13) the expression on the right-hand side of (6.14) 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.

Acknowledgment. We are very grateful to R. R. Bahadur who kindly sent us a preprint of Bahadur and Zabell (1979). The Corollaries 3.1, 4.1 and 4.2 were inspired by this paper. We are also indebted to I. Csiszár who first suggested that S need not be a Polish space (as assumed in a first draft). We wish to thank the referee for several illuminating suggestions; in particular the Remarks 2.1 and 3.1 are the result of his comments.

REFERENCES

[1] Bahadur, R. R. (1971). *Some Limit Theorems in Statistics*. SIAM, Philadelphia.

[2] Bahadur, R. R. and Zabell, S. L. (1979). Large deviations of the sample mean in general vector spaces. *Ann. Probability* **7** 587–621.

[3] Bártfai, P. (1977). On the multivariate Chernoff theorem. Preprint of the Math. Inst. Hung. Acad. Sci.

[4] Bickel, P. J. and Lehmann, E. L. (1975). Descriptive statistics for nonparametric models, II. Location. *Ann. Statist.* **3** 1045–1069.

[5] Borovkov, A. A. (1962). New limit theorems in boundary-value problems for sums of independent terms. *Sibirsk. Math. Z.* **3** 645–695 (English translations in *Sel. Transl. Math. Statist. Prob.* **5** (1965), 315–372).

[6] Borovkov, A. A. (1967). Boundary-value problems for random walks and large deviations in function spaces. *Theor. Probability Appl.* **12** 575–595.

[7] Bourbaki, N. (1965). *Eléments de mathématique, Livre* VI, Intégration. (Chapters 1–4.) Hermann, Paris.

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

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

[10] Csiszár, I. (1975). *I*-divergence geometry of probability distributions and minimization problems. *Ann. Probability* **3** 146–158.

[11] Donsker, M. D. and Varadhan, S. R. S. (1975). Asymptotic evaluation of certain Markov process expectations for large time I. *Comm. Pure Appl. Math.* **27** 1–47.

[12] DONSKER, M. D. and VARADHAN, S. R. S. (1976). Asymptotic evaluation of certain Markov process expectations for large time III. *Comm. Pure Appl. Math.* **29** 389–461.

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

[14] GÄNSSLER, P. (1971). Compactness and sequential compactness in spaces of measures. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **17** 124–146.

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

[16] HOEFFDING, W. (1967). On probabilities of large deviations. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **1** 203–219. Univ. of California Press.

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

[18] LANFORD, D. E. (1972). Entropy and equilibrium states in classical statistical mechanics, Statistical mechanics and mathematical problems (ed. A. Lenard), *Lecture notes in Physics* **20** 1–113. Springer-Verlag, Berlin.

[19] PINSKER, M. S. (1964). *Information and Information Stability of Random Variables and Processes.* Holden-Day, San Francisco.

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

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

[22] RÉNYI, A. (1961). On measures of entropy and information. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1** 547–561.

[23] SANOV, I. N. (1957). On the probability of large deviations of random variables. *Mat. Sb.* **42** 11–44 (English translation in *Sel. Transl. Math. Statist. Prob.* **1** (1961), 213–244).

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

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

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

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

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

P. GROENEBOOM
DEPARTMENT OF STATISTICS
MATHEMATISCH CENTRUM
2E BOERHAAVESTRAAT 49
AMSTERDAM
THE NETHERLANDS

J. OOSTERHOFF
DEPARTMENT OF MATHEMATICS
VRIJE UNIVERSITEIT
DE BOELELAAN 1081
AMSTERDAM
THE NETHERLANDS

F. H. RUYMGAART
DEPARTMENT OF MATHEMATICS
KATHOLIEKE UNIVERSITEIT
TOERNOOVELD
NIJMEGEN
THE NETHERLANDS