stichting
mathematisch
centrum

MC

AFDELING MATHEMATISCHE STATISTIEK SW 46/76 JULI
(DEPARTMENT OF MATHEMATICAL STATISTICS)

P. GROENEBOOM

LARGE DEVIATION THEOREMS FOR EMPIRICAL DISTRIBUTION FUNCTIONS

Prepublication

2e boerhaavestraat 49 amsterdam

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

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

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

Large deviation theorems for empirical distribution functions*)

by

P. Groeneboom

ABSTRACT

Some theorems on first-order asymptotic behavior of probabilities of large deviations of multivariate empirical distribution functions are proved. One of these generalizes certain large deviation theorems of Borovkov, Hoadley, Sanov and Stone in various ways.

An information theoretical proof of a theorem of Chernoff is given.

KEY WORDS & PHRASES: *Large deviations, empirical distribution functions, Kullback-Leibler information, topologies on sets of distribution functions.*

*) This paper is not for review; it is meant for publication elsewhere.

1. INTRODUCTION

Let D_d be the space of d-dimensional distribution functions (dfs) endowed with the topology induced by the supremum metric

$$d(F,G) = \sup_{x\in \mathbb{R}^d} |F(x) - G(x)|.$$

For $F \in D_d$ we denote by μ_F the Borel measure induced by F on $\mathbb{R}^d$. Let the Kullback-Leibler information number K(G,F) of G with respect to $F(F,G\in D_d)$ be defined by

$$K(G,F) = \int_{\mathbb{R}^d} (d\mu_G/d\nu)\log\{(d\mu_G/d\nu)/(d\mu_F/d\nu)\}d\nu$$

where ν is any σ-finite measure on $\mathbb{R}^d$ dominating both μ_F and μ_G. Here and in the sequel we use the conventions

$$0 \log (0/a) = 0 \quad \text{for } a \geq 0$$

and

$$a \log (a/0) = \infty \quad \text{for } a > 0.$$

If A is a subset of D_d and $F \in D_d$, let

$$K(A,F) = \inf\{K(G,F) : G \in A\}.$$

With these notations we have the following theorem of Hoadley (1967) (specialized to the "one-sample situation").

Let $X_1, X_2, \ldots.$ *be mutually independent random variables with a common continuous* df $F \in D_1$ *and let* $\hat{F}_N$ *be the empirical* df *of the first* N *random variables* $X_1,\ldots,X_N$. *Suppose* T *is a real valued uniformly continuous functional on* D_1 *and let* $\Omega_r = \{G \in D_1 : T(G) \geq r\}$ *for each* $r \in \mathbb{R}$.

Then, if the function $t \to K(\Omega_t,F)$ *is continuous at* $t = r$ *and* $\{u_N\}$ *is a sequence of real numbers such that* $\lim_{N\to\infty} u_N = 0$

$$\lim_{N\to\infty} N^{-1}\log P\{T(\hat{F}_N) \geq r + u_N\} = -K(\Omega_r,F). \tag{1.1}$$

One of the purposes of this paper is to prove the same result under weaker conditions and thus to obtain a more general theorem.
In the next sections it will be shown that generalizations are possible in three different directions simultaneously

(i) the uniform continuity of the functional T can be weakened to continuity (or an even weaker condition).
(ii) the space of dfs D_1 may be replaced by D_d.
(iii) F may be an arbitrary df, not necessarily continuous.

Stone (1974) has given a simpler proof of Hoadley's theorem, but under the original strong conditions. Although his proof can easily be adapted to cover the d-dimensional case, it is not obvious that his approach could also be used to generalize Hoadley's theorem in the other directions.

On the other hand Borovkov (1967) has proved the following theorem.

Let F *be a continuous one-dimensional* df *and* Ω *an open set of* dfs *in* D_1. *Then, if* $K(\bar{\Omega},F) = K(\Omega,F)$, *where* $\bar{\Omega}$ *denotes the closure of* Ω:

$$(1.2) \qquad \lim_{N\to\infty} N^{-1}\log P\{\hat{F}_N \in \Omega\} = -K(\Omega,F).$$

By this theorem the *uniform* continuity of the functional T in Hoadley's theorem can be weakened to continuity, but Borovkov relies in his proofs on the rather deep methods of Fourier analysis of random walks from Borovkov (1962) for which generalisation to discontinuous or multi-dimensional dfs might prove to be difficult.

Finally Sanov (1957) has stated a large deviation theorem which is in a certain sense more general and in another sense more special than the theorems of Borovkov, Hoadley and Stone.

We shall prove a theorem (theorem 6.1 of this paper) which implies the above mentioned theorems as special cases. In our proofs we shall rely on rather simple methods which are akin to methods used in information theory (for example Csiszár (1975), Pinsker (1960)). This will clarify the relationships between the results obtained by Borovkov, Hoadley, Sanov and Stone and give a unified approach to these results which were obtained by very different methods.

The theory will be applied to give an information theoretical proof of a theorem of Chernoff (Chernoff (1952)). In a subsequent paper we shall apply our results to prove a multivariate analogue of Chernoff's theorem and to give an expression for the "exact Bahadur slope" of the trimmed mean (which can be considered as a continuous but not *uniformly* continuous functional of empirical dfs).

2. A LARGE DEVIATION THEOREM FOR EMPIRICAL DISTRIBUTION FUNCTIONS

In the sequel μ_G will denote the measure on the Borel field $\mathcal{B}$ on $\mathbb{R}^d$ induced by a d-dimensional distribution function (df) G. The empirical distribution function of a sample of N i.i.d. d-dimensional random vectors with df F will be denoted by $\hat{F}_N$.

D_d will be the space of d-dimensional dfs (which will be endowed with various different topologies in the sequel, but has at present no special topology).

By a *partition* of $\mathbb{R}^d$ is meant a *finite* partition of $\mathbb{R}^d$ consisting of $\mathcal{B}$-measurable sets. The number of sets in the partition is called the *size* of the partition.

DEFINITION 2.1. Let $F \in D_d$ and $\mathcal{P} = \{B_1,\ldots,B_p\}$ be a partition of $\mathbb{R}^d$. Then for a df $G \in D_d$

$$K_{\mathcal{P}}(G,F) = \sum_{j=1}^{p} \mu_G(B_j)\log\{\mu_G(B_j)/\mu_F(B_j)\}$$

and for a set $A \subset D_d$

$$K_{\mathcal{P}}(A,F) = \inf_{G\in A} K_{\mathcal{P}}(G,F).$$

In the sequel we shall repeatedly use without explicit reference the inequality

$$(2.1) \qquad K_{\mathcal{P}}(G,F) \leq K(G,F), \quad \text{for } F,G \in D_d \text{ and each partition } \mathcal{P}$$

which is corollary 3.2 in Kullback (1959).

LEMMA 2.1. *Let* $F \in D_d$. *Suppose* $\{V_N\}$ *is a sequence of sets contained in* D_d *and* $\{\mathcal{P}_N\}$ *a sequence of partitions of* $\mathbb{R}^d$ *of size* $m_N = o(N/\log N)$.
Then

$$(2.2) \qquad P\{\hat{F}_N \in V_N\} \le \exp\{-N(K_{\mathcal{P}_N}(V_N,F) + o(1))\},$$

where $o(1) \to 0$, *as* $N \to \infty$ *at a rate depending on* m_N *but not on the choice of the sets* V_N.

PROOF. Let $\mathcal{P}_N = \{B_{N,1},\ldots,B_{N,m_N}\}$ and

$$p_{N,j} = \mu_F(B_{N,j}), \quad \text{for } 1 \le j \le m_N \quad \text{and } N \in \mathbb{N}.$$

Then

$$P\{\hat{F}_N \in V_N\} \le P\{K_{\mathcal{P}_N}(\hat{F}_N,F) \ge K_{\mathcal{P}_N}(V_N,F)\}$$

$$= \sum \left\{ \frac{N!}{(Nz_1)!\ldots(Nz_{m_N})!} p_{N,1}^{Nz_1} \cdots p_{N,m_N}^{Nz_{m_N}} : \right.$$

$$\sum_{j=1}^{m_N} z_j \log(z_j/p_{N,j}) \ge K_{\mathcal{P}_N}(V_N,F) ,$$

$$\left. \sum_{j=1}^{m_N} z_j = 1,\ z_j \ge 0 \ \text{ and } \ Nz_j \in \mathbb{Z} \ \text{ for each } j \right\}.$$

The number of points $(z_1,\ldots,z_{m_N})$ such that

$$\sum_{j=1}^{m_N} z_j = 1,\ z_j \ge 0 \ \text{ and } \ Nz_j \in \mathbb{Z} \ \text{ for each } j \text{ is}$$

$$(2.3) \qquad \binom{N+m_N-1}{m_N-1} = \exp\{o(N)\},\ N \to \infty .$$

Moreover, by Stirling's formula

$$(2.4) \qquad \frac{N!}{(Nz_1)!\ldots(Nz_{m_N})!} \le \exp\left\{-N\left(\sum_{j=1}^{m_N} z_j \log z_j + o(1)\right)\right\}$$

where $o(1) \to 0$, as $N \to \infty$ at a rate depending on m_N, but not on the z_j's.

Hence

$$P\{\hat{F}_N \in V_N\} \le \exp\{-N(K_{\mathcal{P}_N}(V_N,F) + o(1))\}$$

where $o(1)$ tends to zero uniformly in the sets V_N by (2.3) and (2.4). □

DEFINITION 2.2. For each $N \in \mathbb{N}$ and $A \subset D_d$ the set $A^{(N)}$ is defined by

$$A^{(N)} = \{G \in A: NG(x) \in \mathbb{Z} \text{ for all } x \in \mathbb{R}^d\}$$

THEOREM 2.1. *Let* F *be a* df *in* D_d *and* Ω *a set of* dfs *in* D_d. *Suppose the following conditions hold*

(A) *For each* $c < K(\Omega,F)$ *there exist for all* N *sufficiently large a finite number* k_N *of sets* $V_{n,i} \subset D_d$ *and partitions* $\mathcal{P}_{N,i}$ *of size* $m_{N,i}$ *such that* $\max_{1 \le i \le k_N} m_{N,i} = o(N/\log N)$, $k_N = \exp(o(N))$ *and*

(A1) $K_{\mathcal{P}_{N,i}}(V_{N,i},F) > c,\ 1 \le i \le k_N$

(A2) $\Omega^{(N)} \subset \bigcup_{i=1}^{k_N} V_{N,i}$

(B) *For each* $\varepsilon > 0$ *there exist for all* N *sufficiently large a* df $G_N \in \Omega^{(N)}$ *and a partition* $\mathcal{P}_N$ *of size* $m_N = o(N/\log N)$ *satisfying*

(B1) $K_{\mathcal{P}_N}(G_N,F) < K(\Omega,F) + \varepsilon$

(B2) $\{H \in D_d: \mu_H(B) = \mu_{G_N}(B)$ *for* $B \in \mathcal{P}_N\} \subset \Omega$

then

$$\lim_{N\to\infty} N^{-1} \log P\{\hat{F}_N \in \Omega\} = -K(\Omega,F) \tag{2.5}$$

REMARK. Condition (A) is for example satisfied if $K(\Omega,F) = \sup\{K_{\mathcal{P}}(\Omega,F): \mathcal{P}$ is a partition of $\mathbb{R}^d\}$. For there then exists a partition $\mathcal{P}$ such that $K_{\mathcal{P}}(\Omega,F) > c$, for each $c < K(\Omega,F)$, hence we can take $k_N = 1$, $V_{N,1} = \Omega$. In this case the size of the partition $\mathcal{P}$ is a fixed finite number.

Other sufficient conditions for (A) and (B) will be given in the next sections. This will make clear that certain theorems of Sanov (1957), Borovkov (1967) and Stone (1974) are special cases of Theorem 2.1.

PROOF OF THEOREM 2.1. Let $c < K(\Omega,F)$ be arbitrary and let $\{k_N\}$, $\{V_{N,i}\}$ and $\{\mathcal{P}_{N,i}\}$ be sequences satisfying condition (A). Then, by Lemma 2.1:

$$\limsup_{N\to\infty} N^{-1} \log P\{\hat{F}_N \in \Omega\} \le$$
$$\le \limsup_{N\to\infty} N^{-1} \log k_N - c = -c.$$

Since $c < K(\Omega,F)$ is arbitrary, we get

$$\limsup_{N\to\infty} N^{-1} \log P\{\hat{F}_N \in \Omega\} \le - K(\Omega,F). \tag{2.6}$$

Conversely by condition (B) there exists an $N_0 \in \mathbb{N}$ such that for all $N \ge N_0$ a partition $\{B_{N,1},\ldots,B_{N,m_N}\}$ of size $m_N = o(N/\log N)$ and numbers $z_{N,j}$, $1 \le j \le m_N$ can be found satisfying

$$\sum_{j=1}^{m_N} z_{N,j} = 1,\ z_{N,j} \ge 0,\ Nz_{N,j} \in \mathbb{Z},\ 1 \le j \le m_N$$

and

$$\text{(i)} \qquad \sum_{j=1}^{m_N} z_{N,j} \log \{z_{N,j} / \mu_F(B_{N,j})\} < K(\Omega,F) + \varepsilon$$

$$\text{(ii)} \qquad \{H \in D_d\colon \mu_H(B_{N,j}) = z_{N,j},\ 1 \le j \le m_N\} \subset \Omega$$

Then for $N \ge N_1$, by Stirling's formula:

$$P\{\hat{F}_N \in \Omega\} \ge \frac{N!}{(Nz_{N,1})!\ldots(Nz_{N,m_N})!} \prod_{j=1}^{m_N} (\mu_F(B_{N,j}))^{Nz_{N,j}} \ge$$
$$\ge \exp\{-N(K(\Omega,F) + \varepsilon + o(1))\}.$$

Hence

$$\liminf_{N\to\infty} N^{-1} \log P\{\hat{F}_N \in \Omega\} \ge - K(\Omega,F) - \varepsilon$$

Thus

$$\liminf_{N\to\infty} N^{-1} \log P\{\hat{F}_N \in \Omega\} \ge - K(\Omega,F) \tag{2.7}$$

The theorem now follows from (2.6) and (2.7). □

3. A CONDITION ON THE INTERIOR OF Ω. COMPARISON WITH STONE'S CONDITIONS

In this section we show that a theorem of Stone is implied by theorem 2.1. For this purpose it will be convenient to consider on D_d the topology of convergence on all Borel sets.

DEFINITION 3.1. For each partition $P = \{B_1,\ldots,B_m\}$ of $\mathbb{R}^d$ the pseudometric d_P on D_d is defined by

$$d_P(G,H) = \max_{1\le j\le m} |\mu_G(B_j) - \mu_H(B_j)|, \quad G,H \in D_d.$$

The topology T_1 on D_d is generated by the family of pseudometrics $\{d_P\colon P$ is a partition of $\mathbb{R}^d\}$, i.e. a basis of T_1 is provided by the family of sets $\{H\colon d_P(G,H) < \delta\}$ where $G \in D_d$, $\delta > 0$ and P runs through all (finite) partitions of $\mathbb{R}^d$.

Note that the collection of sets $\{H\colon d_P(G,H) < \delta\}$ is a *basis* and not only a *subbasis* of T_1, because for each $G \in D_d$, $\varepsilon > 0$ and each finite set of partitions $\{P_1,\ldots P_k\}$ we can find a partition P and a $\delta > 0$ such that

$$d_P(H,G) < \delta \Rightarrow d_{P_i}(H,G) < \varepsilon \qquad \text{for } 1 \le i \le k.$$

It is clear that T_1 is the topology of convergence on all Borel sets (i.e. the coarsest topology on D_d for which the map $f_B\colon D_d \to \mathbb{R}$, defined by $f_B(G) = \mu_G(B)$, $G \in D_d$ is continuous for each $B \in \mathcal{B}$).

In the sequel the closure and interior of a set $A \subset D_d$ with respect to a topology T will be denoted by $\mathrm{clos}_T(A)$ and $\mathrm{int}_T(A)$, respectively.

THEOREM 3.1. *Let* F *be a* df *in* D_d *and* Ω *be a set of* dfs *in* D_d, *satisfying*

(A') $K(\Omega,F) = \sup\{K_P(\Omega,F)\colon P$ is a partition of $\mathbb{R}^d\}$

(B') $K(\Omega,F) = K(\mathrm{int}_{T_1}(\Omega),F)$.

Then

$$(3.1) \qquad \lim_{N\to\infty} N^{-1} \log P\{\hat{F}_N \in \Omega\} = -K(\Omega,F)$$

PROOF. If $K(\Omega,F) = \infty$ then also $K(\mathrm{int}_{T_1}(\Omega),F) = \infty$, using the convention $K(\emptyset,F) = \infty$. But then only condition (A) is needed in theorem 2.1, so we may suppose $K(\Omega,F) < \infty$. We shall verify that condition (B) of Theorem 2.1 is satisfied. Fix $\varepsilon > 0$. Since $K(\mathrm{int}_{T_1}(\Omega),F) = K(\Omega,F) < \infty$, $\mathrm{int}_{T_1}(\Omega) \neq \emptyset$. Hence we can find a $G \in \mathrm{int}_{T_1}(\Omega)$ satisfying $K(G,F) < K(\Omega,F) + \frac{1}{2}\varepsilon$. Since $G \in \mathrm{int}_{T_1}(\Omega)$ there exists a partition $P = \{B_1,\ldots,B_m\}$ of $\mathbb{R}^d$ and a $\delta > 0$ such that $\{H \in D_d\colon d_P(H,G) < \delta\} \subset \Omega$. It follows that there exists an $N_0 \in \mathbb{N}$ such that for each $N \geq N_0$ a df $G_N \in D_d^{(N)}$ can be found, satisfying

(i) $d_P(G_N,G) < \delta$, hence $G_N \in \Omega$ and $\{H \in D_d\colon d_P(H,G_N) = 0\} \subset \Omega$

(ii) $K_P(G_N,F) < K_P(G,F) + \frac{1}{2}\varepsilon \leq K(G,F) + \frac{1}{2}\varepsilon < K(\Omega,F) + \varepsilon$.

This shows that condition (B) of Theorem 2.1 is satisfied. □

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

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

For arbitrary $\varepsilon > 0$, there is a df $G \in \Omega$, a partition $P = \{B_1,\ldots,B_m\}$ and a $\delta > 0$ such that

(C2) $K_P(\Omega,F) \leq K_P(G,F) < K_P(\Omega,F) + \varepsilon$

(C3) $\{H \in D_d\colon d_P(H,G) < \delta\} \subset \Omega$

It turns out that if $K(\Omega,F) < \infty$ these conditions are equivalent to conditions (A') and (B') of our theorem 3.1, implying that Stone's theorem 1 is in fact equivalent to our Theorem 3.1 if $K(\Omega,F) < \infty$.

To prove the equivalence suppose that conditions (A') and (B') are satisfied and $K(\Omega,F) < \infty$. Fix $\varepsilon > 0$. By (B') there exists a $G \in \mathrm{int}_{T_1}(\Omega)$ satisfying $K(G,F) < K(\Omega,F) + \varepsilon$. Since $G \in \mathrm{int}_{T_1}(\Omega)$ there exists a partition P and a $\delta > 0$ such that $\{H \in D_d\colon d_P(H,G) < \delta\} \subset \Omega$. By (A') there exists a refinement R of P satisfying $K(\Omega,F) < K_R(\Omega,F) + \varepsilon$ (note that $K_P(H,F) \leq K_R(H,F)$ for each $H \in D_d$ if R is a refinement of P). It is clear that there exists a $\delta' > 0$ such that $H \in D_d$, $d_R(H,G) < \delta' \Rightarrow d_P(H,G) < \delta$.

Now $K_R(\Omega,F) \leq K_R(G,F) \leq K(G,F) < K(\Omega,F) + \varepsilon < K_R(\Omega,F) + 2\varepsilon$
It follows that conditions (C1) to (C3) of Stone (1974) are satisfied.

To prove the implication in the reverse direction we introduce the following definition.

DEFINITION 3.2. Let $F,G \in D_d$, $\mathcal{P} = \{B_1,\ldots,B_m\}$ be a partition of $\mathbb{R}^d$ and $K_{\mathcal{P}}(G,F) < \infty$. Then the *$\mathcal{P}_F$-linear* df G' *corresponding to* G is defined by $\mu_G'(B \cap B_j) = \mu_F(B \cap B_j)\mu_G(B_j)/\mu_F(B_j)$ if $\mu_F(B_j) > 0$ and $\mu_G'(B \cap B_j) = 0$, otherwise.

The device of $\mathcal{P}_F$-linear dfs was probably first used in large deviation problems by Sanov (1957) (for one-dimensional dfs).
It was also used by Hoadley (1967) and in the more general form of definition 3.2 by Stone.

We now can prove that the conditions (C) of Stone imply conditions (A') and (B') of our Theorem 3.1.

First of all the conditions (C) imply $K(\Omega,F) = \sup\{K_{\mathcal{P}}(\Omega,F): \mathcal{P}$ is a partition of $\mathbb{R}^d\}$ by Lemma 2.3 of Stone (1974).
Let $\varepsilon > 0$. If Stone's conditions are satisfied there exists a df $G \in \Omega$, a partition $\mathcal{P}$ of $\mathbb{R}^d$ and a $\delta > 0$ satisfying (C2) and (C3) for this ε.
Let G' be the $\mathcal{P}_F$-linear df corresponding to G.
Then $K(G',F) = K_{\mathcal{P}}(G',F) = K_{\mathcal{P}}(G,F)$. By (C3) $G' \in \mathrm{int}_{\mathcal{T}_1}(\Omega)$ and by (C2) $K(G',F) = K_{\mathcal{P}}(G,F) < K_{\mathcal{P}}(\Omega,F) + \varepsilon \leq K(\Omega,F) + \varepsilon$.
Thus $K(\mathrm{int}_{\mathcal{T}_1}(\Omega),F) < K(\Omega,F) + \varepsilon$ implying that condition (B') of Theorem 3.1 is satisfied, since $\mathrm{int}_{\mathcal{T}_1}(\Omega) \subset \Omega \Rightarrow K(\Omega,F) \leq K(\mathrm{int}_{\mathcal{T}_1}(\Omega),F)$.

4. A THEOREM OF SANOV

In this section we shall show that Theorem 11 in Sanov (1957) is a special case of our Theorem 2.1. This has some interest since there exists doubt as to the validity of the results of Sanov (see Hoeffding (1965) p.373, Hoadley (1967) p.365 and Bahadur (1971) p.12) and since for example Hoadley (1967) for this reason doesn't use Sanov's Theorem 11 but redefines Sanov's concepts of ε-neighborhood and F-distinguishability to avoid Sanov's Theorem 11.

We shall show that Sanov's theorem holds with the original definitions. This will at the same time throw some light on the relationship between the concept of F-distinguishability and Stone's conditions (C)(a question posed by Stone in Stone (1974)).

DEFINITION 4.1. Let $\mathcal{P}$ be a partition of $\mathbb{R}^d$ consisting of the sets $B_1 = (-\infty,x_1)$, $B_2 = [x_1,x_2),\ldots,B_{m-1} = [x_{m-2},x_{m-1})$, $B_m = [x_{m-1},\infty)$, where

$x_1 < x_2 < \dots_m < x_{m-1}$ and let $W_m = \bigcup_{i=1}^{m} B_i \times [a_i, b_i]$, where $0 = a_1 \le a_2 \le \dots \le a_m \le 1$, $0 \le b_1 \le \dots \le b_m = 1$ and $b_1 - a_2 \ge 0$, $b_2 - a_3 > 0$, $b_3 - a_4 > 0$, ..., $b_{m-2} - a_{m-1} > 0$, $b_{m-1} - a_m \ge 0$. Suppose $G_1 \in D_1$ is defined by $G_1(x) = a_i$, if $x \in B_i$. Then $V_m = \{G \in D_1 : (x,G(x)) \in W_m \text{ for all } x \in \mathbb{R}\} \setminus \{G_1\}$ is called an ε-*neighborhood*.

The partition P is called the *partition corresponding to* V_m.

An unsatisfactory aspect of the concept of ε-neighborhood is that ε doesn't appear in it.

DEFINITION 4.2. Let F be a df in D_1 which does not take on only a finite set of values. Then a set Ω is called F-*distinguishable* if the following conditions are satisfied:

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

(b) For every $\eta > 0$ and every $N \in \mathbb{N}$ there exists a finite number $k = k(\eta,N)$ of ε-neighborhoods $V_{m_1}, \dots, V_{m_k}$ such that $\Omega^{(N)} \subset \bigcup_{i=1}^{k} V_{m_i}$ and $K(V_{m_i},F) > K(\Omega,F) - \eta$, $1 \le i \le k$.

Moreover $\log k(\eta,N) = o(N)$ and $\max_{1\le i\le k} m_i = o(N/\log N)$.

(c) For every $\eta > 0$ there exists an ε-neighborhood $V_m \subset \Omega$ satisfying $K(V_m,F) < K(\Omega,F) + \eta$.

We show that conditions (A) and (B) of Theorem 2.1 are satisfied if Ω is F-distinguishable.

Let $\varepsilon > 0$. By condition (c) of definition 4.2 there exists an ε-neighborhood $V_m \subset \Omega$ satisfying $K(V_m,F) < K(\Omega,F) + \frac{1}{3}\varepsilon$. Choose a $G \in V_m$ satisfying

$$(4.1) \qquad K(G,F) < K(V_m,F) + \tfrac{1}{3}\varepsilon.$$

According to definition 4.1:

$$V_m = \{H \in D_1 : (x,H(x)) \in W_m\} \setminus \{G_1\}$$

where $W_m = \bigcup_{i=1}^{m} B_i \times [a_i,b_i]$, $P = \{B_1,\dots,B_m\}$ is the partition corresponding to V_m and G_1 is the "lower bound" df of V_m defined by $G_1(x) = a_i$, for $x \in B_i$, $1 \le i \le m$.

We have

$$(4.2) \qquad K_P(G,F) \le K(G,F) < K(V_m,F) + \tfrac{1}{3}\varepsilon.$$

ε-neighborhood based on the set $W_m = \bigcup_{i=1}^{m} B_i \times [a_i,b_i]$, with $a_1 = b_1 = 0$, then V_m is not open in T_1.

5. A CONDITION ON THE CLOSURE OF Ω A THEOREM OF BOROVKOV

In this section we shall show that condition (A) of Theorem 2.1 is satisfied if $K(\mathrm{clos}_{T_2}(\Omega),F) = K(\Omega,F)$, where T_2 is the topology on D_d induced by the supremum metric $d(F,G) = \sup_{x\in\mathbb{R}^d} |F(x) - G(x)|$ for F, $G \in D_d$. This will lead to easy proofs of theorems of Borovkov and Hoadley.

The idea is to prove "tightness" for a family of probability measures which have uniformly bounded Kullback-Leibler numbers with respect to a fixed df F. Actually it will be shown in Lemma 5.2 that the uniform boundedness of the numbers $K(G_m,F)$ for a sequence $\{G_m\}$ of dfs implies the convergence of a subsequence $\{G_{m_k}\}$ in the topology T_1 of convergence on all Borel sets.

This approach is akin to the information theoretical proofs of convergence of a sequence of dfs $\{G_m\}$ to a df F under the condition $\lim_{m\to\infty} K(G_m,F) = 0$ (see J.W. Linnik (1959), A. Renyi (1960), I. Csiszār (1962)). In fact, if $\lim_{m\to\infty} K(G_m,F) = 0$, then $\{G_m\}$ converges in total variation to F (Pinsker (1960)) which is a stronger kind of convergence than convergence in T_1 (see the remark following Lemma 5.1).

<u>LEMMA 5.1</u>. *T_2 is strictly coarser than T_1.*

<u>PROOF</u>. Let $\varepsilon > 0$ and G be a one-dimensional df. Then there exists a finite (possibly empty) set of points $x_i \in \mathbb{R}$ such that $\mu_G(\{x_i\}) \geq \frac{1}{2}\varepsilon$, for each i. We can therefore find a partition $P = \{B_1,\ldots,B_m\}$ of $\mathbb{R}^d$ consisting of one-point sets $\{x_i\}$ such that $\mu_G(\{x_i\}) \geq \frac{1}{2}\varepsilon$ and open or half open intervals B_j such that $\mu_G(B_j) \leq \frac{1}{2}\varepsilon$. Then $H \in D_1$, $d_P(G,H) < \varepsilon/4m \Rightarrow \sup_{x\in\mathbb{R}} |H(x) - G(x)| < \varepsilon$, which proves the statement for D_1.

Next suppose that $G \in D_d$ and that G_i, $1 \leq i \leq d$ are the one-dimensional marginals of G. For each marginal df G_i there exists by the last paragraph a partition $\{B_{i,1}\ldots,B_{i,m_i}\}$ consisting of open or half-open intervals $B_{i,j}$ such that $\mu_{G_i}(B_{i,j}) < \varepsilon/2d$ and one-point sets $B_{i,j}$ with $\mu_{G_i}(B_{i,j}) \geq \varepsilon/2d$. Let $P = \{B_1,\ldots B_m\}$ be the partition of $\mathbb{R}^d$ consisting of the product sets $B_{1,j_1} \times .. \times B_{d,j_d}$, $1 \leq j_k \leq m_i$, $1 \leq i \leq d$.

Then $H \in D_d$, $d_P(H,G) < \varepsilon/4dm \Rightarrow \sup_{x\in\mathbb{R}^d} |H(x) - G(x)| < \varepsilon$ which proves the statement for D_d.

Moreover $T_1 \neq T_2$, for convergence in the supremum metric does not imply convergence on all Borel sets (for each sequence of purely discrete dfs which converge in the supremum metric to a continuous df in D_d there exists a countable set of points in $\mathbb{R}^d$ such that the probability measure of this set is equal to one for each df in the sequence). □

We note that the topology T_1 is strictly coarser (has less open sets) than the topology generated by the total variation distance
$d(G,H) = \sup_{B\in\mathcal{B}} |\mu_G(B) - \mu_H(B)|$.

For example if we define for each $n \in \mathbb{N}$ the dfs $G_n \in D_1$ by the densities

$$g_n(x) = \begin{cases} 2, & \text{if } (2k)2^{-n} < x < (2k+1)2^{-n},\ 0 \le k < 2^{n-1} \\ 0, & \text{otherwise} \end{cases}$$

then $\lim_{n\to\infty} G_n = G$ in the topology T_1 but $\sup_{B\in B} |\mu_{G_n}(B) - \mu_G(B)| =$
$= \frac{1}{2}\int_0^1 |g_n(x) - 1|\, dx = \frac{1}{2}$ for each $n \in \mathbb{N}$, if $G(x) = x$ on $[0,1]$.

In fact we have convergence on all Borel sets, but the convergence is not uniform.

DEFINITION 5.1. A collection $\mathcal{G}$ of dfs in D_d is called *uniformly absolutely continuous* (u.a.c.) with respect to df $F \in D_d$, if for each $\varepsilon > 0$ there exists a $\delta > 0$ such that for each $G \in \mathcal{G}$ and each $B \in \mathcal{B}$: $\mu_G(B) < \delta \Rightarrow \mu_F(B) < \varepsilon$.

The next lemma gives some relationships between the topology of weak convergence and the topologies T_1 and T_2 for a class of dfs with uniformly bounded Kullback-Leibler numbers.

LEMMA 5.2. *Let* F *be an arbitrary* df *in* D_d *and let* $\mathcal{G} = \{G \in D_d : K(G,F) \le M\}$ for some fixed $M > 0$.
Then
(a) $\mathcal{G}$ is u.a.c. with respect to F.
(b) For each sequence $\{G_m\}$ in $\mathcal{G}$ which converges weakly (in law) to a df $G \in D_d$
  (i) $\lim_{m\to\infty} G_m = G$ in the topology T_1 (hence in particular in the topology T_2).
  (ii) $K(G,F) \le \liminf_{m\to\infty} K(G_m,F)$.

(c) $\mathcal{G}$ is compact in the topology $\mathcal{T}_1$.

(d) The restriction of the identity map I: $(D_d, \mathcal{T}_1) \to (D_d, \mathcal{T}_2)$ to $\mathcal{G}$ is uniformly continuous.

PROOF.

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

$$\mu_G(B) = \int_B g\, d\mu_F = \int_{B\cap\{g\le\varepsilon/2\delta\}} g d\mu_F + \int_{B\cap\{g>\varepsilon/2\delta\}} g\, d\mu_F \le (\varepsilon/2\delta)\,\mu_F(B) + (\log(\varepsilon/2\delta))^{-1} \int_{B\cap\{g>\varepsilon/2\delta\}} g \log g\, d\mu_F \le \tfrac{1}{2}\varepsilon + (M+e^{-1})(\log(\varepsilon/2\delta))^{-1} < \varepsilon,$$

where $g = dG/dF$.

(Note that the inequality $x \log x \ge -e^{-1}$ gives a lower bound for the integral $\int_B g \log g\, d\mu_F$).

It follows that $\mathcal{G}$ is u.a.c. with respect to F.

(b)(i) Suppose $\{G_m\}$ is a sequence in $\mathcal{G}$ which converges weakly to a df $G \in D_d$. By (a) there exists for each $\varepsilon > 0$ a $\delta > 0$ such that for each $m \in \mathbb{N}$ and $B \in \mathcal{B}$: $\mu_F(B) < \delta \Rightarrow \mu_{G_m}(B) < \varepsilon$.

Let $B \in \mathcal{B}$. Then there exist an open set U and a closed set K satisfying $K \subset B \subset U$ and $\mu_F(U\backslash K) < \delta$. This implies that $\sup_{m\in\mathbb{N}} \mu_{G_m}(U\backslash K) \le \varepsilon$, so by the weak convergence of $\{G_m\}$ to G:

$$\limsup_{m\to\infty} \mu_{G_m}(B) \le \limsup_{m\to\infty} \mu_{G_m}(K) + \limsup_{m\to\infty} \mu_{G_m}(B\backslash K) \le \mu_G(K) + \varepsilon \le \mu_G(B) + \varepsilon. \tag{5.1}$$

and

$$\liminf_{m\to\infty} \mu_{G_m}(B) \ge \liminf_{m\to\infty} \mu_{G_m}(U) - \limsup_{m\to\infty} \mu_{G_m}(U\backslash B) \ge \mu_G(U) - \varepsilon \ge \mu_G(B) - \varepsilon. \tag{5.2}$$

Inequalities (5.1) and (5.2) imply $\lim_{m\to\infty} \mu_{G_m}(B) = \mu_G(B)$.

(b)(ii) This follows from (b)(i) and the inequality $K(G,F) \le \liminf_{m\to\infty} K(G_m,F)$, valid if $G = \lim_{m\to\infty} G_m$ in the topology $\mathcal{T}_1$.

This relation follows from a theorem of Gelfand, Yaglom and Perez stating that $K(G,F) = \sup\{K_{\mathcal{P}}(G,F): \mathcal{P}$ is a partition of $\mathbb{R}^d\}$ (see for example Pinsker (1964), p.20)

(c) It is well known that for the topology $\mathcal{T}_1$ on D_d the notions "compact" and "sequentially compact" coincide (see for example Gänssler (1971), theorem 3.7).

Let $\{G_m\}$ be a sequence in $\mathcal{G}$. This sequence is tight because $\mathcal{G}$ is u.a.c. with respect to F and F is tight. There exists therefore a subsequence $\{G_{m_k}\}$ of $\{G_m\}$ which converges weakly to a df G. Hence by (b) $\{G_{m_k}\}$ converges to G in the topology $\mathcal{T}_1$ and
$K(G,F) \leq \liminf_{k\to\infty} K(G_{m_k},F) \leq M$.

(d) The identity map is continuous because $\mathcal{T}_2 \subset \mathcal{T}_1$ and *uniformly* continuous on $\mathcal{G}$ because $\mathcal{G}$ is compact (note that the family of pseudometrics $\{d_{\mathcal{P}}\}$ generates a uniformity on $D_d \times D_d$). □

The next lemma gives a sufficient condition for the fulfillment of condition (A') of Theorem 3.1.

LEMMA 5.3. *Let* F *be a* df *in* D_d *and* Ω *a set of* dfs *in* D_d *satisfying*

(A") $$K(\mathrm{clos}_{\mathcal{T}_2}(\Omega),F) = K(\Omega,F).$$

Then condition (A') *of Theorem* 3.1 *holds.*

PROOF. Let $\alpha = \sup\{K_{\mathcal{P}}(\Omega,F): \mathcal{P}$ is a partition of $\mathbb{R}^d\}$ and let $\eta > 0$ be such that $\alpha + \eta < K(\Omega,F)$. If $A = \{G \in D_d: K(G,F) \leq \alpha + \eta\}$, then by Lemma 4.2 the restriction to A of the identity map I: $(D_d, \mathcal{T}_1) \to (D_d, \mathcal{T}_2)$ is *uniformly* continuous. Hence, for each $m \in \mathbb{N}$ there exist a partition $\mathcal{P}_m$ and a $\delta_m > 0$, such that $\sup_{x\in\mathbb{R}^d} |G(x) - H(x)| < \frac{1}{m}$ if G, H $\in A$ and $d_{\mathcal{P}_m}(G,H) < \delta_m$. We can choose for each $m \in \mathbb{N}$ a $G_m \in \Omega$ satisfying $K_{\mathcal{P}_m}(G_m,F) \leq \alpha + \eta$.
Let the $(\mathcal{P}_m)_F$-linear function G'_m corresponding to G_m be defined as in definition 3.2. Then $K(G'_m,F) = K_{\mathcal{P}_m}(G'_m,F) = K_{\mathcal{P}_m}(G_m,F) \leq \alpha + \eta$, for each $m \in \mathbb{N}$, hence $G'_m \in A$ and because by Lemma 5.2 A is compact in the topology $\mathcal{T}_1$ there exists a $G \in A$ and a subsequence $\{G'_{m_k}\}$ of $\{G'_m\}$ satisfying $\lim_{k\to\infty} G'_{m_k} = G$ in $\mathcal{T}_1$. Then also $\lim_{k\to\infty} G'_{m_k} = G$ in $\mathcal{T}_2$ and since $d_{\mathcal{P}_m}(G'_m, G_m) = 0 < \delta_m \Rightarrow$
$\Rightarrow \sup_{x\in\mathbb{R}^d} |G_m(x) - G'_m(x)| < \frac{1}{m}$, $\lim_{k\to\infty} \sup_{x\in\mathbb{R}^d} | G_{m_k}(x) - G(x)| = 0$. It follows that $G \in \mathrm{clos}_{\mathcal{T}_2}(\Omega)$. However $G \in A \Rightarrow K(G,F) \leq \alpha + \eta < K(\Omega,F)$, a contradiction. □

Combining results of Section 3 and this section we get the following theorem.

THEOREM 5.1. *Let* F *be a* df *in* D_d *and* Ω *a set of* dfs *in* D_d *satisfying*

$$(5.3) \qquad K(\mathrm{int}_{T_1}(\Omega),F) = K(\mathrm{clos}_{T_2}(\Omega),F)$$

Then $\lim_{N\to\infty} N^{-1}\log P\{\hat{F}_N \in \Omega\} = -K(\Omega,F)$

PROOF. This follows at once from Theorem 3.1 and Lemma 5.3. □

By a theorem of Borovkov $\lim_{N\to\infty} N^{-1} \log P\{\hat{F}_N \in \Omega\} = -K(\Omega,F)$ if the underlying df F is a one-dimensional continuous df and Ω is a set of dfs in D_1 such that Ω is open in the topology T_2 and $K(\Omega,F) = K(\mathrm{clos}_{T_2}(\Omega),F)$ (see (31) in Borovkov (1967)). Obviously this is a special case of Theorem 5.1 since $T_2 \subset T_1$ implies $\mathrm{int}_{T_2}(\Omega) \subset \mathrm{int}_{T_1}(\Omega)$.

Borovkov suggests in Borovkov (1972) p.29, that Sanov's Theorem 11 is implied by this special case, although it is not entirely clear what he means by "the" theorem of Sanov. Anyhow, this implication does not seem to hold, because the ε-neighborhoods of Sanov are not necessarily open sets in T_2 (see the remark at the end of Section 4 where it is shown that an ε-neighborhood need not be open in T_1).

6. THE K-SAMPLE SITUATION. GENERALIZATION OF A THEOREM OF HOADLEY

In this section we shall consider the k-sample situation. Let $X_{i,1},\dots,X_{i,n_i}$ be i.i.d. d-dimensional random vectors with df F_i for $1 \le i \le k$ and let the sample sizes n_i tend to infinity such that $|n_i/N - \rho_i| = o(1)$, where $N = \sum_{i=1}^{k} n_i$ and $\rho_i > 0$, $1 \le i \le k$. The empirical df of the sample $\{X_{i,1},\dots,X_{i,n_i}\}$ will be denoted by $\hat{F}_{i,n_i}$.

If T is a topology on D_d, then the product topology on D_d which has a basis consisting of the product sets $A_1 \times .. \times A_k$, where $A_i \in T$ for $1 \le i \le k$, will also be denoted by T.

DEFINITION 6.1. Let $F = (F_1,\dots,F_k) \in D_d^k$ and $\rho = (\rho_1,\dots,\rho_k) \in (0,1]^k$, where $\sum_{i=1}^{k} \rho_i = 1$. Let $P = P_1 \times .. \times P_k$ be a partition of $\mathbb{R}^{dk}$ consisting of the product sets $B_{1,j_1} \times .. \times B_{k,j_k}$ where B_{i,j_i} belongs to the partition P_i of

PROOF. Only small changes in the proof of Theorem 2.1 are needed.

COROLLARY 6.1. *Let* $F = (F_1,\dots,F_k) \in D_d^k$ *and* Ω *be a subset of* D_d^k *satisfying* $I_\rho(\mathrm{int}_{T_1}(\Omega),F) = I_\rho(\mathrm{clos}_{T_2}(\Omega),F)$ *then*

$$\lim_{N\to\infty} N^{-1}\log P\{(\hat{F}_{1,n_1},\dots,\hat{F}_{k,n_k}) \in \Omega\} = -K(\Omega,F). \tag{6.4}$$

PROOF. Similar to the proof of Theorem 5.1.

DEFINITION 6.3. Suppose $T\colon D_d^k \to \mathbb{R} \cup \{-\infty,\infty\}$ is an extended real-valued function. Then, for each $r \in \mathbb{R}$, Ω_r is defined by $\Omega_r = \{G \in D_d^k\colon T(G) \ge r\}$. For $F \in D_d^k$, $r \in \mathbb{R}$ and $\rho = (\rho_1,\dots,\rho_k) \in (0,1]^k$ such that $\sum_{i=1}^k \rho_i = 1$ we define $I_\rho(r) = I_\rho(\Omega_r,F)$.

LEMMA 6.1. *Let* $T\colon D_d^k \to \mathbb{R} \cup \{-\infty,\infty\}$ *be an upper semicontinuous function. Then* $I_\rho\colon \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ *is continuous from the left.*

PROOF. If $I_\rho(r) = \infty$ for each $r \in \mathbb{R}$ the statement of the lemma is trivial, so suppose $I_\rho(r) < \infty$ for at least one $r \in \mathbb{R}$. Let $\{r_m\}$ be a sequence in $\mathbb{R}$ such that $r_m \uparrow r$ for an $r \in \mathbb{R}$ satisfying $I_\rho(r) < \infty$. I_ρ is monotonically non-decreasing on $\mathbb{R}$, hence $I_\rho(r_m) \le I_\rho(r) < \infty$ for each $m \in \mathbb{N}$ and $\lim_{m\to\infty} I_\rho(r_m)$ exists. Choose $\varepsilon > 0$. For each $m \in \mathbb{N}$ there exists a $G_m = (G_{m,1},\dots,G_{m,k}) \in D_d^k$ satisfying $I_\rho(G_m,F) < I_\rho(r_m) + \varepsilon$ and $T(G_m) \ge r_m$. Since $I_\rho(G_m,F) = \sum_{i=1}^k \rho_i\, K(G_{m,i},F_i) < I_\rho(r) + \varepsilon < \infty$ for each $m \in \mathbb{N}$ and $\rho_i > 0$, for $1 \le i \le k$, the Kullback-Leibler numbers $K(G_{m,i},F_i)$ are uniformly bounded in m for $1 \le i \le k$. By Lemma 5.2 there exists therefore a subsequence $\{G_{m_j}\}$ of $\{G_m\}$ and a $G = (G_1,\dots,G_k) \in D_n^k$ such that for each i $\lim_{j\to\infty} G_{m_j,i} = G_i$ in the topology T_1 on D_d and $K(G_i,F_i) \le \liminf_{j\to\infty} K(G_{m_j,i},F_i)$. Then also $\lim_{j\to\infty} G_{m_j} = G$ in the topology T_1 on D_d^k and $I_\rho(G,F) \le \liminf_{j\to\infty} I_\rho(G_{m_j},F) \le \liminf_{j\to\infty} I_\rho(r_{m_j}) + \varepsilon$. Since T is upper semicontinuous and $T(G_{m_j}) \ge r_{m_j}$ for each $j \in \mathbb{N}$, $T(G) \ge r$. Hence $G \in \Omega_r$ and $I_\rho(G,F) \le \liminf_{j\to\infty} I_\rho(r_{m_j}) + \varepsilon = \lim_{m\to\infty} I_\rho(r_m) + \varepsilon$.

Since $\varepsilon > 0$ is arbitrary $I_\rho(r) = \lim_{m\to\infty} I_\rho(r_m)$ is immediate from the monotonicity of I_ρ. The left continuity also holds for a point r such that $I_\rho(r) = \infty$ and $I_\rho(r') < \infty$ for $r' < r$. For if $r_m \uparrow r$ and $I_\rho(r_m)$ is uniformly bounded in m, then by the line of argument used above there exists a $G \in \Omega_r$ satisfying $I_\rho(G,F) < \infty$, a contradiction. □

THEOREM 6.2. *Let* $T: D_d^k \to \mathbb{R} \cup \{-\infty,\infty\}$ *be a continuous function on* D_d^k. *Then, if* I_ρ *is continuous from the right in* r *and if* $\{u_N\}$ *is a sequence of real numbers such that* $\lim_{N\to\infty} u_N = 0$

$$(6.5) \qquad \lim_{N\to\infty} N^{-1} \log P\{T(\hat{F}_{1,n_1},\ldots,\hat{F}_{k,n_k}) \geq r + u_N\} = -I_\rho(r)$$

PROOF. Since the function $t \to I_\rho(t)$ is monotonically non-decreasing, it has at most countably many points of discontinuity. I_ρ is continuous from the left by Lemma 6.1 and continuous from the right in r by assumption. Hence there exists for each $\varepsilon > 0$ a $\delta > 0$ such that $I_\rho(r) - \varepsilon < I_\rho(r-\delta) \leq I_\rho(r) \leq I_\rho(r+\delta) < I_\rho(r) + \varepsilon$ and I_ρ is continuous in $r-\delta$ and $r+\delta$. Obviously the continuity of I_ρ in a point $t \in \mathbb{R}$ and the continuity of T imply $I_\rho(\mathrm{clos}_{T_2}(\Omega_t),F) = I_\rho(\Omega_t,F) = I_\rho(\mathrm{int}_{T_2}(\Omega_t),F)$, since $\Omega_{t+\gamma} \subset \mathrm{int}_{T_2}(\Omega)$ for each $\gamma > 0$.

Corollary 6.1 now implies $-I_\rho(r) - \varepsilon < -I_\rho(r+\delta) =$

$$= \lim_{N\to\infty} N^{-1} \log P\{T(\hat{F}_{1,n_1},\ldots,\hat{F}_{k,n_k}) \geq r + \delta\} \leq$$
$$\leq \liminf_{N\to\infty} N^{-1} \log P\{T(\hat{F}_{1,n_1},\ldots,\hat{F}_{k,n_k}) \geq r + u_N\} \leq$$
$$\leq \limsup_{N\to\infty} N^{-1} \log P\{T(\hat{F}_{1,n_1},\ldots,\hat{F}_{k,n_k}) \geq r + u_N\} \leq$$
$$\leq \lim_{N\to\infty} N^{-1} \log P\{T(\hat{F}_{1,n_1},\ldots,\hat{F}_{k,n_k}) \geq r - \delta\} = -I_\rho(r-\delta) < -I_\rho(r) + \varepsilon.$$ Thus $\lim_{N\to\infty} N^{-1} \log P\{T(\hat{F}_{1,n_1},\ldots,\hat{F}_{k,n_k}) \geq r + u_N\} = -I_\rho(r)$. □

Hoadley's Theorem 1 in Hoadley (1967) is a special case of our Theorem 6.2. In Hoadley's theorem $D_d = D_1$, $F = (F_1,\ldots,F_k)$ consists of *continuous one-dimensional* dfs, T is a real-valued *uniformly* continuous function on D_1^k and $|n_i/N - \rho_i| = \mathcal{O}(N^{-1}\log N)$. In Hoadley's proof the set Ω_r is approached by so-called "F-strips" which are similar to the ε-neighborhoods of Sanov. This leads to very involved constructions for which generalizations to a proof of Theorem 6.2 might prove to be rather difficult.

7. CHERNOFF'S THEOREM

The foregoing theory will be applied to give an information theoretical proof of Chernoff's theorem (Chernoff (1952)). In a subsequent paper we shall give a multivariate generalization of this theorem.

THEOREM 7.1. *Let* $X_1, X_2,\ldots,$ *be a sequence of* i.i.d. *random variables with* df $F \in D_1$ *and let* Ω_r *be defined by* $\Omega_r = \{G \in D_1: \int_{\mathbb{R}} x dG(x)$ *exists and*

$\int_{\mathbb{R}} x dG(x) \geq r\}$. Then

$$(7.1) \qquad \lim_{N\to\infty} N^{-1} \log P\{N^{-1} \sum_{i=1}^{N} X_i \geq r\} = -K(\Omega_r, F).$$

This theorem is equivalent to Chernoff's theorem, since by Lemma 1 of Hoeffding (1965) $K(\Omega_r,F) = -\log(\inf_{t\geq 0}\{e^{-tr} \int_{\mathbb{R}} e^{tx} dF(x)\})$.

<u>LEMMA 7.1</u>. *Let* $F \in D_1$. *Then the mapping* $r \to K(\Omega_r,F)$, $r \in \mathbb{R}$ *is convex.*

<u>PROOF</u>. This follows from the convexity of the function $x \to x \log x$, $x \geq 0$ and the linearity of the function $G \to \int_{\mathbb{R}} x\, dG(x)$, $G \in \{H \in D_1 : \int_{\mathbb{R}} x\, dH(x) > -\infty\}$. □

<u>DEFINITION 7.1</u>. For $F \in D_1$ and $M > 0$ the (conditional) df F_M is defined by

$$(7.2) \qquad \mu_{F_M}(B) = \mu_F(B \cap [-M,M])/\mu_F([-M,M]), \quad B \in \mathcal{B}.$$

<u>LEMMA 7.2</u>. *Let* $F \in D_1$ *and* $r \in \mathbb{R}$. *Then* $K(\Omega_r,F) = \lim_{M\to\infty} K(\Omega_r,F_M)$.

<u>PROOF</u>. Let $B_M = [-M,M]$ for each $M > 0$. Choose an arbitrary $\varepsilon > 0$. There exists an $M_0 > 0$ such that $|\log \mu_F(B_M)| < \varepsilon$ for $M \geq M_0$. Hence, if $M \geq M_0$: $K(G,F) \leq K(G,F_M) + \varepsilon$ for each $G \in D$ (the inequality is trivially satisfied if $K(G,F_M) = \infty$), implying $K(\Omega_r,F) \leq \liminf_{M\to\infty} K(\Omega_r,F_M)$.

We shall prove that also

$$(7.3) \qquad \limsup_{M\to\infty} K(\Omega_r,F_M) \leq K(\Omega_r,F)$$

If $K(\Omega_r,F) = \infty$, then (7.3) is trivially satisfied. If $K(\Omega_r,F) < \infty$, but $K(\Omega_{r+\delta},F) = \infty$ for all $\delta > 0$, then it is easily seen that $K(\Omega_r,F) = -\log \mu_F(\{r\})$. So we may suppose

$$(7.4) \qquad K(\Omega_{r+\delta},F) < \infty, \qquad \text{for some } \delta > 0.$$

If (7.4) is satisfied, $K(\Omega_t,F)$ is finite in a neighborhood of $t = r$ by monotonicity. The convexity of the mapping $t \to K(\Omega_t,F)$ then implies that this mapping is continuous at $t = r$.

Let $\varepsilon > 0$. Then there exists a $\delta > 0$ such that $K(\Omega_{r+\delta},F) < K(\Omega_r,F) + \frac{1}{2}\varepsilon$. Choose a $G \in \Omega_{r+\delta}$ satisfying $K(G,F) < K(\Omega_{r+\delta},F) + \frac{1}{2}\varepsilon$. There exists an $M_0 > 0$ such that $\int_{\mathbb{R}} x\, dG_M(x) \geq r$, if $M \geq M_0$, where G_M is defined by

$\mu_{G_M}(B) = \mu_G(B \cap B_M)/\mu_G(B_M)$ for $B \in \mathcal{B}$, $B_M = [-M,M]$. Hence

$$\limsup_{M\to\infty} K(\Omega_r,F_M) \le \lim_{M\to\infty} K(G_M,F_M) = K(G,F) < K(\Omega_r,F) + \varepsilon,$$

implying (7.3). $\square$

LEMMA 7.3. *Let* $X_1,X_2,\ldots,$ *be a sequence of* i.i.d. *random variables with* df $F \in D_1$. *Then, for each* $r \in \mathbb{R}$ *and* $m \in \mathbb{N}$

$$(7.5)\qquad m^{-1} \log P\{m^{-1} \sum_{i=1}^{m} X_i \ge r\} \le \limsup_{N\to\infty} N^{-1} \log P\{N^{-1} \sum_{i=1}^{N} X_i \ge r\}.$$

PROOF. $\limsup_{N\to\infty} N^{-1} \log P\{N^{-1} \sum_{i=1}^{N} X_i \ge r\} \ge$

$$\ge \lim_{N\to\infty} (Nm)^{-1} \log\{(P\{m^{-1} \sum_{i=1}^{m} X_i \ge r\})^N\} = m^{-1} \log P\{m^{-1} \sum_{i=1}^{m} X_i \ge r\}. \quad \square$$

PROOF OF THEOREM 7.1. We first prove Theorem 7.1 under the assumption that F has compact support, i.e. $\mu_F(B_M) = 1$ for a bounded closed interval $B_M = [-M,M]$, $M > 0$.

The function $T: D_1 \to \mathbb{R}$ defined by $T(G) = \int_{[-M,M]} x\, dG(x)$ is continuous on the space D_1 endowed with the topology $\mathcal{T}_2$ induced by the supremum metric $d(G,H) = \sup_{x\in\mathbb{R}} |G(x) - H(x)|$.

We now note that since $\mu_F(B_M) = 1$:

$$P\{T(\hat{F}_N) \ge r\} = P\{\int_{[-M,M]} x\, d\hat{F}_N(x) \ge r\} =$$
$$= P\{\int_{\mathbb{R}} x\, d\hat{F}_N(x) \ge r\} = P\{N^{-1} \sum_{i=1}^{N} X_i \ge r\}$$

and $K(\Omega_r,F) = K(A_r,F)$, where $A_r = \{G: T(G) \ge r\}$.

If the function ψ: $t \to K(A_t,F)$ is continuous from the right in $t = r$, then (7.1) follows from Theorem 6.2. If the function ψ is not continuous from the right in $t = r$, then $\mu_F(\{r\}) > 0$, $F(r) = 1$ and $\lim_{N\to\infty} N^{-1} \log P\{N^{-1} \sum_{i=1}^{N} X_i \ge r\} = \log \mu_F(\{r\}) = -K(A_r,F)$ (see the proof of Lemma 7.2).

To prove the theorem without the condition that F has compact support we introduce the notation

$$P_{F_M}\{N^{-1} \sum_{i=1}^{N} X_i \geq r\} = P\{N^{-1} \sum_{i=1}^{N} X_i \geq r \mid X_i \in [-M,M],\ 1 \leq i \leq N\}.$$

Let $\varepsilon > o$ and $c = \limsup_{N\to\infty} N^{-1} \log P\{N^{-1} \sum_{i=1}^{N} X_i \geq r\}$.

There exists an $m \in \mathbb{N}$ such that $m^{-1} \log P\{m^{-1} \sum_{i=1}^{m} X_i \geq r\} \geq c - \varepsilon$.

Since $\lim_{M\to\infty} P_{F_M}\{m^{-1} \sum_{i=1}^{m} X_i \geq r\} = P\{m^{-1} \sum_{i=1}^{m} X_i \geq r\}$ there exists an $M_0 > 0$ satisfying $m^{-1} \log P_{F_M}\{m^{-1} \sum_{i=1}^{m} X_i \geq r\} \geq c - 2\varepsilon$, if $M \geq M_0$. Hence, by Lemma 7.3

$$(7.6) \qquad \limsup_{N\to\infty} N^{-1} \log P_{F_M}\{N^{-1} \sum_{i=1}^{N} X_i \geq r\} \geq c - 2\varepsilon, \text{ if } M \geq M_0.$$

By Lemma 7.2

$$(7.7) \qquad K(\Omega_r,F) = \lim_{M\to\infty} K(\Omega_r,F_M)$$

and by the first part of this proof

$$(7.8) \qquad \lim_{N\to\infty} N^{-1} \log P_{F_M}(N^{-1} \sum_{i=1}^{N} X_i \geq r\} = -K(\Omega_r,F_M).$$

Combining (7.6), (7.7) and (7.8) we get $c - 2\varepsilon \leq -K(\Omega_r,F)$, implying $c \leq -K(\Omega_r,F)$ since $\varepsilon > 0$ was arbitrarily chosen.

On the other hand we have for fixed $N \in \mathbb{N}$ and $M > 0$

$$N^{-1} \log P\{N^{-1} \sum_{i=1}^{N} X_i \geq r\} \geq N^{-1} \log P_{F_M}\{N^{-1} \sum_{i=1}^{N} X_i \geq r\} + \log \mu_F([-M,M]).$$

Hence, for M sufficiently large $\liminf_{N\to\infty} N^{-1} \log P\{N^{-1} \sum_{i=1}^{N} X_i \geq r\} \geq$

$\geq \liminf_{N\to\infty} N^{-1} \log P_{F_M}\{N^{-1} \sum_{i=1}^{N} X_i \geq r\} + \log \mu_F([-M,M])$

$\geq -K(\Omega_r,F_M) + \log \mu_F([-M,M])$, where the last inequality follows from the first part of the proof.

Thus, by Lemma 7.2

$$\liminf_{N\to\infty} N^{-1} \log P\{N^{-1} \sum_{i=1}^{N} X_i \geq r\} \geq$$

$$\geq - \lim_{M\to\infty} K(\Omega_r,F_M) + \lim_{M\to\infty} \log \mu_F([-M,M]) = - K(\Omega_r,F). \quad \square$$

8. ACKNOWLEDGEMENTS

I wish to thank F.H. Ruymgaart for suggesting the problem and possible applications of a generalization of Hoadley's theorem and J. Oosterhoff for a careful reading of the manuscript and many useful and stimulating suggestions.

REFERENCES

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

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

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

[4] BOROVKOV, A.A. (1972) *Limit theorems for random walks with boundaries*, Proc. 6th Berkeley Symp. Math. Statist. Prob. 111 19-30.

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

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

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

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

[9] HOEFFDING, W. (1965) *On probabilities of large deviations*, Proc. 5th Berkeley Symp. Math. Statist. Prob. 1 203-219.

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

[11] KULLBACK, S. (1959) *Information theory and statistics*, Wiley, New York.

[12] PINSKER, M.S. (1964) *Information and information stability of random variables and processes*, Holden-Day San Francisco (Translation of Informatsiya i informatsionnaya ustiochivost' sluchainykh velichin i protsessor Ac. of Science, USSR Moscow (1960)).

[13] SANOV, I.N. (1957)*On the probability of large deviations of random variables (in Russian)* , Mat. Sbornik N.S. 42 (84) 11-44 (English translation in Sel. Transl. Math. Statist. Prob. 1 (1961) 213-244).

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