The Annals of Statistics
1976, Vol. 4, No. 4, 766–770

PROBABILITIES OF LARGE DEVIATIONS FOR EMPIRICAL MEASURES[1]

BY GERALD L. SIEVERS
Western Michigan University

Sanov's theorem on the asymptotic behavior of probabilities of large deviations for an empirical probability distribution is established under different conditions than previously given by Sanov, Hoadley and Stone. The new conditions are based on likelihood ratio approximations rather than on multinomial approximations. It is shown that these conditions are strictly more general than those of Stone.

1. Introduction. Let $\mathscr{X}$ be an arbitrary set and $\mathscr{B}$ be a σ-field of subsets of $\mathscr{X}$ which contains all singletons. Let p be a probability measure on $\mathscr{B}$. Let $X_1, X_2, \cdots$ denote a sequence of i.i.d. random variables which take values in $\mathscr{X}$ according to p. For each positive integer n, let $(\mathscr{X}^{(n)}, \mathscr{B}^{(n)})$ be the sample space of $X_{(n)} = (X_1, \cdots, X_n)$ and $p^{(n)}$ the corresponding product measure.

If q is a probability measure on $\mathscr{B}$ and $q \ll p$, let $\lambda_q(x)$ be a $\mathscr{B}$-measurable function, $0 \leqq \lambda_q(x) < \infty$, such that $dq = \lambda_q(x)\, dp$ and let $I(q, p) = \int_{\mathscr{X}} \log \lambda_q(x)\, dq$. If q is not absolutely continuous with respect to p let $I(q, p) = \infty$. For $\mathbf{x} = (x_1, \cdots, x_n) \in \mathscr{X}^{(n)}$, let $\lambda_{nq}(\mathbf{x}) = \prod_{i=1}^{n} \lambda_q(x_i)$.

For each positive integer n, let $p_n = p_n(\cdot \mid X_1, \cdots, X_n)$ be the empirical measure defined by $p_n(B) = n^{-1}$ (the number of subscripts j such that $X_j \in B$, $1 \leqq j \leqq n$) for $B \in \mathscr{B}$.

Let A denote a set of probability measures on $\mathscr{B}$ such that $P(p_n \in A \mid p)$ is well-defined for each n and $I(A, p) = \inf_{q \in A} I(q, p) < \infty$. Let $A_n = \{\mathbf{x} \in \mathscr{X}^{(n)} : p_n(\cdot \mid x_1, \cdots, x_n) \in A\}$.

In Section 2 sufficient conditions are presented for the equation

$$P(p_n \in A \mid p) = \exp[-nI(A, p) + o(n)]\,. \tag{1.1}$$

These conditions were motivated by a consideration of likelihood ratio tests of $H_0: p$ vs. $H_1: q$ (or $H_1: q \in A$) which have exact slopes $I(q, p)$. For (1.1) to hold, it would be sufficient for A_n, viewed as a critical region of a test of H_0 vs. H_1, to be similar in some sense to the likelihood ratio tests for suitable $q \in A$ "nearest" to p ($I(q, p)$ near $I(A, p)$). Condition I (of Section 2) requires that the power of this test be asymptotically bounded away from 0 at such alternatives. Viewed another way, condition I requires for certain $q \in A$ with $q \ll p$, that the

Received July 1974; revised August 1975.
[1] This research was supported in part by Western Michigan University (Faculty Research Fellowship).
AMS 1970 *subject classifications*. Primary 60F10; Secondary 62F20, 62G20.
Key words and phrases. Large deviations, exact slope, empirical probability measures, likelihood ratio.

discrete probability measures in A "near" q be sufficiently dense. This bears some resemblance to Stone's (1974) condition (C3).

Condition II, with $B_n = A_n$, requires that some $q \in A$, $q \ll p$, nearly dominates A in that for all points of A_n, λ_{nq} is sufficiently large. Using $B_n \subset A_n$ covers some cases where λ_{nq} may not be sufficiently large over all of A_n but points in $A_n - B_n$ can be ignored in view of II(ii).

In Section 3, it is shown that conditions I and II are more general than Stone's (1974) conditions (C2) and (C3). In Section 4, a well-known example is considered which provides nontrivial cases where conditions I and II can hold but Stone's conditions do not hold directly. Stone indicates that his conditions can be applied to this type of example by using a truncation argument.

2. Main results. Consider the following conditions:

(I) there exists a sequence of probability measures $\{q_m\}_{m=1}^{\infty}$ such that $q_m \in A$ and $q_m \ll p$ for all m and

(i) $\lim_{m\to\infty} I(q_m, p) = I(A, p)$
(ii) $\liminf_{n\to\infty} P(p_n \in A \mid q_m) > 0$ for all m,

(II) for all $\varepsilon > 0$, there exists $q \in A$, $q \ll p$, and a sequence $\{B_n\}_{n=1}^{\infty}$ such that $B_n \subset A_n$ and B_n is $\mathscr{B}^{(n)}$-measurable for each n and with

$$c_{nq} = \inf_{\mathbf{x} \in B_n} \lambda_{nq}(\mathbf{x}) ,$$

(i) $I(A, p) - \varepsilon \leqq \liminf_{n\to\infty} n^{-1} \log c_{nq}$
(ii) $n^{-1} \log \int_{A_n - B_n} dp^{(n)} \leqq -n^{-1} \log c_{nq} + \varepsilon$

for all n sufficiently large.

Note that condition I holds if there exists $q \in A$, $q \ll p$, such that $I(q, p) = I(A, p)$ and I(ii) holds for $q_m = q$. Also, if $B_n = A_n$ then II(ii) trivially holds.

THEOREM 2.1. *If $I(A, p) < \infty$ and conditions* I *and* II *hold, then* (1.1) *holds.*

PROOF. First show that condition I implies

$$\liminf_{n\to\infty} n^{-1} \log P(p_n \in A \mid p) \geqq -I(A, p) . \tag{2.2}$$

For each m, consider $\{p_n \in A\} = \{X_{(n)} \in A_n\}$ as a critical region for testing the simple hypotheses H_0: p vs. H_1: q_m. The significance level of the test is $P(p_n \in A \mid p)$ and the power is $P(p_n \in A \mid q_m)$. Condition I(ii) implies that the power is asymptotically bounded away from 0 and applying an argument like (6.10) of Lemma 6.1 of Bahadur (1971), it follows that

$$\liminf_{n\to\infty} n^{-1} \log P(p_n \in A \mid p) \geqq -I(q_m, p) .$$

Then I(i) implies (2.2).

Next it is shown that condition II implies

$$\limsup_{n\to\infty} n^{-1} \log P(p_n \in A \mid p) \leqq -I(A, p) . \tag{2.3}$$

Let $\varepsilon > 0$. Then

$$\int_{B_n} dp^{(n)} = \int_{B_n} \lambda_{nq}^{-1}\, dq^{(n)} \leqq c_{nq}^{-1} \leqq c_{nq}^{-1} e^{n\varepsilon}$$

and

$$P(p_n \in A \mid p) = \int_{A_n - B_n} dp^{(n)} + \int_{B_n} dp^{(n)}$$
$$\leqq 2 \max \{\int_{A_n - B_n} dp^{(n)}, c_{nq}^{-1} e^{n\varepsilon}\} .$$

Then II(ii) implies

$$n^{-1} \log P(p_n \in A \mid p) \leqq n^{-1} \log 2 + \varepsilon - n^{-1} \log c_{nq}$$

and II(i) implies

$$\limsup_{n\to\infty} n^{-1} \log P(p_n \in A \mid p) \leqq -I(A, p) + 2\varepsilon .$$

Since ε is arbitrary, (2.3) follows.

3. Relation to Stone's conditions. In this section it is shown that conditions I and II are more general than Stone's conditions (C2) and (C3).

Let D_k denote a k-class partition $\mathscr{X}_1^k \cup \cdots \cup \mathscr{X}_k^k$ of $\mathscr{X}$ such that $\mathscr{X}_i^k$ is $\mathscr{B}$-measurable and $p(\mathscr{X}_i^k) > 0$ for $i = 1, \cdots, k$. Let

$$I(q, D_k) = \sum_{i=1}^k q(\mathscr{X}_i^k) \log [q(\mathscr{X}_i^k)/p(\mathscr{X}_i^k)] ,$$
$$I(A, D_k) = \inf_{q \in A} I(q, D_k) ,$$
$$I(A, k) = \sup_{D_k} \{I(A, D_k) \mid k\} ,$$
$$I(A, \sup) = \sup_k I(A, k) \qquad \text{and for} \quad \delta > 0$$
$$N_\delta(q \mid D_k) = \{q' : \max_{1\leq i\leq k} |q(\mathscr{X}_i^k) - q'(\mathscr{X}_i^k)| \leqq \delta\} .$$

Assume that Stone's conditions hold. Then for any $\varepsilon > 0$ there exists $q' \in A$, an integer k, a partition D_k and $\delta > 0$ such that

(C2) $I(A, D_k) \leqq I(q', D_k) \leqq I(A, D_k) + \varepsilon$
(C3) $N_\delta(q' \mid D_k) \subset A$.

Next proceed to construct the q, $\{B_n\}$ and $\{q_m\}$ required for conditions I and II. Define a probability measure $q \in N_0(q' \mid D_k)$ by

$$q(\mathscr{X}_i^k) = q'(\mathscr{X}_i^k) \qquad \text{and}$$
$$q(B_i) = p(B_i) q'(\mathscr{X}_i^k)/p(\mathscr{X}_i^k)$$

for $\mathscr{B}$-measurable $B_i \subset \mathscr{X}_i^k$, $i = 1, \cdots, k$. Then $q \in A$, $q \ll p$, $I(q, D_k) = I(q', D_k)$, $I(q, D_k) = I(q, p)$ and $N_\delta(q \mid D_k) = N_\delta(q' \mid D_k)$.

To show condition I, note that for each positive integer m there is a probability measure q_m on $\mathscr{B}$ (using $\varepsilon = 1/m$ and $q_m = q$ in the previous discussion) such that $q_m \in A$, $q_m \ll p$,

$$I(q_m, p) = I(q_m, D_k) \leqq I(A, D_k) + (1/m) \qquad \text{(from C2)}$$
$$\leqq I(A, p) + (1/m)$$

and

$$I(A, p) \leqq I(q_m, p) .$$

Hence I(i) follows. Also, since $N_\delta(q_m \mid D_k) \subset A$ and

$$P[p_n \in N_\delta(q_m \mid D_k) \mid q_m] \to 1$$

as $n \to \infty$ for each m, we have I(ii).

To show condition II, use q as previously defined and the following construction for $\{B_n\}$. There exists N such that for each positive integer $n \geqq N$ there is a probability measure q_n which assigns probability $1/n$ to n points of $\mathscr{X}$, $q_n \in N_\delta(q \mid D_k) \subset A$ and

$$(3.1) \qquad |\textstyle\sum_{i=1}^k q_n(\mathscr{X}_i^k) \log [q(\mathscr{X}_i^k)/p(\mathscr{X}_i^k)] - I(q, D_k)| < \varepsilon .$$

(Note if $q(\mathscr{X}_i^k) = 0$ take q_n such that $q_n(\mathscr{X}_i^k) = 0$ with $0 \log 0 = 0$.) Now let

$$B_n = \{\mathbf{x} \in \mathscr{X}^{(n)} : p_n(\cdot \mid x_1, \cdots, x_n) \in N_0(q_n \mid D_k)\} .$$

Then $B_n \subset A_n$ and

$$(3.2) \qquad \begin{aligned} n^{-1} \log c_{nq} &= n^{-1} \log (\inf_{\mathbf{x} \in B_n} \lambda_{nq}(\mathbf{x})) \\ &= \textstyle\sum_{i=1}^k q_n(\mathscr{X}_i^k) \log [q(\mathscr{X}_i^k)/p(\mathscr{X}_i^k)] \end{aligned}$$

since $\lambda_q(x) = q(\mathscr{X}_i^k)/p(\mathscr{X}_i^k)$ if $x \in \mathscr{X}_i^k$ for $i = 1, \cdots, k$. From (3.1), (3.2) and (C2) it follows that

$$\begin{aligned} n^{-1} \log c_{nq} &\leqq I(q, D_k) + \varepsilon \\ &\leqq I(A, D_k) + 2\varepsilon . \end{aligned}$$

Now II(i) follows, since from (3.1) and (3.2)

$$\begin{aligned} n^{-1} \log c_{nq} &\geqq I(q, D_k) - \varepsilon \\ &= I(q, p) - \varepsilon \\ &\geqq I(A, p) - \varepsilon \end{aligned}$$

for all $n \geqq N$.

To show II(ii), first note that from Stone's Lemma 2.1

$$\begin{aligned} \int_{A_n - B_n} dp^{(n)} &\leqq \int_{A_n} dp^{(n)} = P(p_n \in A \mid p) \\ &\leqq \exp[-nI(A, k) + O(\log n)] \end{aligned}$$

where O depends only on k. Then for n sufficiently large

$$\begin{aligned} n^{-1} \log \int_{A_n - B_n} dp^{(n)} &\leqq -I(A, k) + \varepsilon \\ &\leqq -I(A, D_k) + \varepsilon \\ &\leqq n^{-1} \log c_{nq} + 3\varepsilon \end{aligned}$$

from (3.3). Since ε is arbitrary, II(ii) follows.

4. An example. Suppose $Y = Y(x)$ is a real-valued, $\mathscr{B}$-measurable function on $\mathscr{X}$ with m.g.f. $\phi(t) = \int_{\mathscr{X}} \exp(tY(x))\, dp$ which satisfies the standard conditions of Bahadur (1971), Chapter 2. Let $A = \{q : \int_{\mathscr{X}} Y\, dq$ exists and is $\geqq 0\}$. Then $P[p_n \in A \mid p] = P[\sum_{i=1}^n Y(X_i) \geqq 0 \mid p]$. Now Bahadur's standard conditions imply that there exists $\tau > 0$ such that $\phi'(\tau) = 0$. Then the probability measure $q = q_\tau$ defined by $dq = [\phi(\tau)]^{-1} \exp(\tau Y(x))\, dp$ satisfies $E_q(Y) = \phi'(\tau) = 0$ and

$\operatorname{Var}_q(Y) > 0$. Also $I(q, p) = -\log \phi(\tau) = I(A, p)$ follows from Theorem 4.2 of Bahadur (1971).

It will be shown that conditions I and II are satisfied and hence (1.1) holds. Note

$$A_n = \{\mathbf{x} \in \mathscr{X}^{(n)} : \sum_{i=1}^n Y(x_i) \geqq 0\} \qquad \text{and}$$
$$\lambda_{nq}(\mathbf{x}) = [\phi(\tau)]^{-n} \exp[\tau \sum_{i=1}^n Y(x_i)] .$$

Now using $B_n = A_n$,

$$c_{nq} = \inf_{\mathbf{x} \in B_n} \lambda_{nq}(\mathbf{x}) = [\phi(\tau)]^{-n} .$$

Then $I(A, p) \leqq I(q, p) = -\log \phi(\tau) = +n^{-1} \log c_{nq}$ for all n and condition II holds.

If $q_m = q$ for all m, then, since $P[\sum_{i=1}^n Y(X_i) \geqq 0 \mid q] \to \frac{1}{2}$ as $n \to \infty$ by the central limit theorem and $I(A, p) = I(q, p)$, condition I follows.

REMARK. In this example, the application of Theorem 2.1 is simplified in that the infimum of λ_{nq} over the entire set A_n is large enough for condition II(i) to hold. Then using $B_n = A_n$, condition II(ii) trivially holds.

It is clear that (1.1) continues to hold if a single point is added to each A_n. These points can be chosen so that c_{nq} tends to 0 when $B_n = A_n$ and condition II would fail. However, with the option $B_n \neq A_n$, condition II would hold.

REFERENCES

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

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

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

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

DEPARTMENT OF MATHEMATICS
WESTERN MICHIGAN UNIVERSITY
KALAMAZOO, MICHIGAN 49001