The Annals of Probability
1978, Vol. 6, No. 1, 169–173

A CURIOUS CONVERSE OF SIEVER'S THEOREM

BY JAMES LYNCH

The Pennsylvania State University

A sufficient condition for a sequence of random variables, $T_1, T_2, \cdots$, with cumulant generating functions, $\psi_1, \psi_2, \cdots$, to have a large deviation rate is that $n^{-1}\psi_n(\lambda) \to \psi(\lambda)$, where $\psi(\lambda)$ satisfies certain regularity conditions. Here it is shown that, when the large deviation rate exists and $T_1, T_2, \cdots$ are properly truncated, it is a necessary condition.

1. Introduction. Let $T_1, T_2, \cdots$ be a sequence of random variables such that T_n/n converges to zero in probability (written $T_n/n \to_P 0$). For $a > 0$, let $P_n(a) = P(T_n \geqq na)$. Then $P_n(a) \to 0$ as $n \to \infty$, and, in many cases, $P_n(a) \to 0$ exponentially fast, i.e., there exists a $\psi^*(a)$ with $0 < \psi^*(a) < \infty$ such that

$$n^{-1} \log P_n(a) \to -\psi^*(a) . \tag{1.1}$$

The function ψ^* is referred to as the *large deviation rate.*

In many instances the large deviation rate can be determined. For example, if $T_n = X_1 + \cdots + X_n$, where $X_1, X_2, \cdots$ are independent and identically distributed (i.i.d.) with mean 0, then (see Chernoff (1952), Bahadur and Rao (1960), and Bahadur (1971))

$$\psi^*(a) = \sup\{\lambda a - \psi(\lambda) : \lambda \geqq 0\} , \tag{1.2}$$

where $\psi(\lambda) = \log E(\exp(\lambda X_1))$ is the cumulant generating function (c.g.f.) of X_1. Recently, Sievers (1969) used Bahadur and Rao's type of proof of Chernoff's theorem to calculate the large deviation rate for sequences of random variables $T_1, T_2, \cdots$, which are not necessarily sums of i.i.d. random variables, when the c.g.f.'s of $\{T_n\}$ and some of their high order derivatives satisfy certain regularity conditions. More recently, Plachky (1971) and Plachky and Steinebach (1975) have improved Siever's theorem by removing these higher order derivative conditions. The Plachky and Steinebach result may be stated as follows.

THEOREM 1.1. *Let $T_1, T_2, \cdots$ be a sequence of random variables with* c.g.f.'s *$\psi_1, \psi_2, \cdots$, respectively. Let*

$$n^{-1}\psi_n(\lambda) \to \psi(\lambda) \quad \text{for} \quad \lambda \in [c, d] , \qquad 0 \leqq c < d . \tag{1.3}$$

Let $D = \{\lambda \in (c, d) : \psi'$ exists and is continuous at $\lambda\}$ and let $A = \{a : \psi'(\lambda) = a, \lambda \in D\}$. Then, if $\psi(\lambda)$ is strictly convex on $[c, d]$, (1.1) holds for $a \in A$, where

$$\psi^*(a) = \sup\{\lambda a - \psi(\lambda) : \lambda \in [c, d]\} .$$

Notice that the large deviation rate (1.2) for sums of i.i.d. random variables

Received November 15, 1976.
AMS 1970 *subject classifications.* Primary 60F10; Secondary 62E10, 62E20.
Key words and phrases. Large deviation rate.

is an immediate consequence of Theorem 1.1 with $\psi_n(\lambda) = n \log E(\exp(\lambda X_1))$, when $E(\exp(\lambda X_1)) < \infty$ for $\lambda \in [c, d]$ and X_1 is nondegenerate.

In this paper, we consider the converse problem, i.e., when does (1.1) imply (1.3)? Example 2.6, below, shows that the converse need not hold. The reason that the converse may not hold, as is illustrated in the example, is that the exponential factor in the c.g.f. of T_n artificially gives large weight to regions in the tail of the distribution of T_n which may blow up the c.g.f. as $n \to \infty$. However, this problem is removed by truncating T_n; and in Theorem 2.5, we show that the converse indeed holds for the truncated random variables.

2. The main result. Let $T_1, T_2, \cdots$ be a sequence of random variables such that $T_n/n \to_P 0$. It is assumed that

$$(2.1) \qquad n^{-1} \log P(T_n \geqq na) \to -\psi^*(a) \qquad \text{for} \quad a \in [0, d] \quad \text{and} \quad (0 < d < \infty) ,$$

where $\psi^*(d) < \infty$.

Let

$$\psi(\lambda) = \sup \{\lambda a - \psi^*(a) : a \in [0, d]\} ,$$

and let

$$\begin{aligned} S_n &= T_n \qquad \text{if} \quad T_n < nd \\ &= nd \qquad \text{if} \quad T_n \geqq nd . \end{aligned}$$

For $n = 1, 2, \cdots$ and for $\lambda \geqq 0$, let $\psi_n(\lambda)$ be the c.g.f. of S_n. Note that, for each n, ψ_n is a convex function and that the derivative of ψ_n, denoted ψ_n', exists and is finite for $\lambda > 0$. Also, since ψ_n is convex, the right derivative of ψ_n exists at 0, though it may be equal to $-\infty$, and will be denoted by $\psi_n'(0)$.

We begin with some lemmas which are required for the proof of Theorem 2.5.

LEMMA 2.1. *For $a \in [0, d]$ and for ψ^* as defined in* (2.1),

$$n^{-1} \log P(S_n \geqq na) \to -\psi^*(a) \qquad as \quad n \to \infty .$$

PROOF. For $a \in [0, d]$, $P(S_n \geqq na) = P(T_n \geqq na)$, from which the result follows immediately.

LEMMA 2.2. *For $\lambda \geqq 0$,*

$$\liminf \frac{1}{n} \psi_n(\lambda) \geqq \psi(\lambda) .$$

PROOF. Let $\lambda \geqq 0$. Then, by Markov's inequality,

$$P(S_n \geqq na) \leqq \exp(-\lambda na) E(\exp(\lambda S_n)) .$$

Hence, for $a \in [0, d]$,

$$\lambda a - \psi^*(a) \leqq \liminf n^{-1} \psi_n(\lambda) .$$

The lemma now follows from the definition of $\psi(\lambda)$ by taking the supremum of the left side of the above inequality over $a \in [0, d]$.

LEMMA 2.3. *The sequence* $\{n^{-1}\psi_n(\lambda)\}$ *is pointwise bounded, and hence, for each subsequence* $\{m\}$ *of* $\{n\}$, *there is a further subsequence* $\{k\}$ *of* $\{m\}$ *and a convex function* $\tilde{\psi}(\lambda)$ *(which may depend on the subsequence* $\{k\}$*) such that*

$$k^{-1}\psi_k(\lambda) \to \tilde{\psi}(\lambda) \qquad \text{as} \quad k \to \infty \quad \text{for} \quad \lambda \geqq 0 .$$

PROOF. Let $\lambda \geqq 0$. Then

$$n^{-1} \log P(S_n \geqq 0) \leqq n^{-1}\psi_n(\lambda) \leqq \lambda d .$$

Since $n^{-1} \log P(S_n \geqq 0) \to -\psi^*(0) \geqq -\psi^*(d) > -\infty$, it follows that the sequence $\{n^{-1}\psi_n(\lambda)\}$ is bounded for $\lambda \geqq 0$. The remainder of Lemma 2.3 is a well-known property of bounded convex functions (see Roberts and Varberg (1973), Chapter I).

LEMMA 2.4. *For* $\lambda > 0$,

$$0 \leqq \liminf n^{-1}\psi_n'(\lambda) \leqq \limsup n^{-1}\psi_n'(\lambda) \leqq d . \tag{2.2}$$

PROOF. Let $\lambda > 0$. Let ϕ_n and F_n be moment generating function and distribution function, respectively, of S_n. Then

$$\begin{aligned} n^{-1}\psi_n'(\lambda) &\leqq n^{-1} \textstyle\int_{(0,nd]} x \exp(\lambda x)\, dF_n(x)/\phi_n(\lambda) \\ &\leqq d \textstyle\int_{(0,nd]} \exp(\lambda x)\, dF_n(x)/\phi_n(\lambda) \leqq d . \end{aligned}$$

This proves the inequality on the right of (2.2).

To prove the inequality on the left, note that $T_n/n \to_P 0$ implies that $S_n/n \to_P 0$. So, for $\lambda \geqq 0$, $E(\exp(\lambda S_n/n)) = \phi_n(\lambda/n) \to 1$ as $n \to \infty$ since $\exp(\lambda x)$ is a bounded continuous function on $(-\infty, d]$. Since the function $g_n(\lambda) = \phi_n(\lambda/n)$ is convex for each n, it follows that $g_n'(\lambda) \to 0$ on $(0, \infty)$ (see Roberts and Varberg (1973), Chapter I). So,

$$\begin{aligned} n^{-1}\psi_n'(\lambda/n) &= n^{-1}\phi_n'(\lambda/n)/\phi_n(\lambda/n) \\ &= g_n'(\lambda)/g_n(\lambda) \to 0 \end{aligned}$$

on $(0, \infty)$. Thus the inequality on the left of (2.2) follows by noting that $n^{-1}\psi_n'(\lambda)$ is an increasing function of λ for each n.

THEOREM 2.5. *For* $\lambda > 0$,

$$\lim n^{-1}\psi_n(\lambda) = \psi(\lambda) .$$

PROOF. From Lemma 2.2, we need only show that

$$\limsup n^{-1}\psi_n(\lambda) \leqq \psi(\lambda) \qquad \text{for} \quad \lambda > 0 . \tag{2.3}$$

To do this we define the following quantities. Let $X_1, X_2, \cdots$ be i.i.d. normal random variables which are independent of S_n and have mean 0 and variance σ^2. Let $S_n' = \sum_1^n X_i$. Then, for $\lambda \geqq 0$,

$$\psi_{n,\sigma}(\lambda) = \log E(\exp(\lambda(S_n + S_n'))) = \psi_n(\lambda) + n(\lambda\sigma)^2/2 .$$

It follows from Lemma 2.3 that, for every subsequence $\{m\}$ of $\{n\}$, there is a further subsequence $\{k\}$ of $\{m\}$ and a convex function $\tilde{\psi}(\lambda)$ such that

$$k^{-1}\psi_k(\lambda) \to \tilde{\psi}(\lambda) \quad \text{as} \quad k \to \infty \quad \text{for} \quad \lambda \geqq 0 .$$

So,

$$k^{-1}\psi_{k,\sigma}(\lambda) \to \tilde{\psi}_\sigma(\lambda) \quad \text{as} \quad k \to \infty \quad \text{for} \quad \lambda \geqq 0 ,$$

where $\tilde{\psi}_\sigma(\lambda) = \tilde{\psi}(\lambda) + (\lambda\sigma)^2/2$ is strictly convex.

Let $D = \{\lambda \in (0, \infty)$: $\tilde{\psi}_\sigma'$ exists and is continuous at $\lambda\}$, i.e., $D = \{\lambda \in (0, \infty)$: $\tilde{\psi}'$ exists and is continuous at $\lambda\}$. It follows that D is dense in $(0, \infty)$, and furthermore, it is clear that it is independent of σ. Fix $\sigma > 0$ and let $\lambda \in D$. Let $a_{\lambda,\sigma} = \tilde{\psi}_\sigma'(\lambda)$. Then by Theorem 1.1,

$$k^{-1} \log P(S_k + S_k' \geqq ka_{\lambda,\sigma}) \to -\tilde{\psi}_\sigma^*(a_{\lambda,\sigma}) \quad \text{as} \quad k \to \infty,$$

where $\tilde{\psi}_\sigma^*(a_{\lambda,\sigma}) = \lambda a_{\lambda,\sigma} - \tilde{\psi}_\sigma(\lambda)$.

Note that $a_{\lambda,\sigma} \downarrow \tilde{\psi}'(\lambda) = a_\lambda$ as $\sigma \downarrow 0$. It follows from Lemma 2.4 that $a_\lambda \in [0, d]$ for $\lambda \in D$. We consider two cases.

CASE 1. $a_\lambda \in (0, d]$. Let $0 < \varepsilon < a_\lambda$. Then for all sufficiently small σ, say $\sigma < \sigma_0$, $0 < a_{\lambda,\sigma} - \varepsilon < a_\lambda$. Thus, since

$$\begin{aligned} P(S_k + S_k' \geqq ka_{\lambda,\sigma}) &\leqq P(S_k' \leqq k\varepsilon)P(S_k \geqq k(a_{\lambda,\sigma} - \varepsilon)) + P(S_k' \geqq k\varepsilon) \\ &\leqq P(S_k \geqq k(a_{\lambda,\sigma} - \varepsilon)) + P(S_k' \geqq k\varepsilon) , \end{aligned}$$

$$\tilde{\psi}_\sigma^*(a_{\lambda,\sigma}) \geqq \min (\psi^*(a_{\lambda,\sigma} - \varepsilon), \varepsilon^2/(2\sigma^2)) \quad \text{for} \quad 0 < \sigma < \sigma_0 .$$

That is,

$$\lambda a_{\lambda,\sigma} - \tilde{\psi}_\sigma(\lambda) \geqq \min (\psi^*(a_{\lambda,\sigma} - \varepsilon), \varepsilon^2/(2\sigma^2)) ,$$

or

$$\begin{aligned} (\lambda\sigma)^2/2 + \tilde{\psi}(\lambda) &\leqq \max (\lambda a_{\lambda,\sigma} - \psi^*(a_{\lambda,\sigma} - \varepsilon), \lambda a_{\lambda,\sigma} - \varepsilon^2/(2\sigma^2)) \\ &= \max (\lambda\varepsilon + \lambda(a_{\lambda,\sigma} - \varepsilon) - \psi^*(a_{\lambda,\sigma} - \varepsilon), \lambda a_{\lambda,\sigma} - \varepsilon^2/(2\sigma^2)) \\ &\leqq \max (\lambda\varepsilon + \psi(\lambda), \lambda a_{\lambda,\sigma} - \varepsilon^2/(2\sigma^2)) . \end{aligned}$$

Letting $\sigma^2 \to 0$ and then $\varepsilon \to 0$, we have

$$\tilde{\psi}(\lambda) \leqq \psi(\lambda) . \tag{2.4}$$

CASE 2. $a_\lambda = 0$. Then, from the argument in Case 1 with $\varepsilon = a_{\lambda,\sigma}/2$, we have

$$\tilde{\psi}(\lambda) + (\lambda\sigma)^2/2 \leqq \max (\psi(\lambda) + \lambda a_{\lambda,\sigma}/2, \lambda a_{\lambda,\sigma}) .$$

Letting $\sigma \to 0$, we get (2.4). Hence, (2.4) holds for all $\lambda \in D$, and since $\tilde{\psi}(\lambda)$ and $\psi(\lambda)$ are convex and D is dense in $(0, \infty)$, (2.4) holds for all $\lambda > 0$. Thus, since every subsequence $\{m\}$ of $\{n\}$ has a further subsequence $\{k\}$ such that (2.4) holds for $\lambda > 0$, it follows that (2.3) must hold. This completes the proof of Theorem 2.5.

We conclude the paper with an example where (2.1) holds for a sequence of random variables, $T_1, T_2, \cdots$, but $\lim n^{-1} \log E(\exp(\lambda T_n)) = \infty$ for $\lambda > 0$.

EXAMPLE 2.6. Let $X_1, X_2, \cdots$ be i.i.d. normal random variables with mean

0 and variance 1. For $n = 1, 2, \cdots$, let Y_n be a random variable independent of $X_1, \cdots, X_n$, where $Y_n = n^3$, 0, or $-n^3$ with probabilities, $\exp(-n^2)$, $1 - 2\exp(-n^2)$, and $\exp(-n^2)$, respectively. For $n = 1, 2, \cdots$, let $T_n = \sum_1^n X_i + Y_n$. For $\lambda > 0$ let $\psi_n(\lambda) = \log E(\exp(\lambda T_n))$. Then,

$$n^{-1}\psi_n(\lambda) = \lambda^2/2 + n^{-1}\log(1 + \exp(n^2(n\lambda - 1)) + o(1))$$
$$\to \infty \quad \text{as} \quad n \to \infty .$$

However, for $a > 0$,

$$\begin{aligned} P(T_n \geqq na) &= P(\textstyle\sum_1^n X_i \geqq na)(1 - 2\exp(-n^2)) \\ &\quad + \exp(-n^2)(P(\textstyle\sum_1^n X_i \geqq na - n^3) + P(\textstyle\sum_1^n X_i \geqq na + n^3)) \\ &= (P(\textstyle\sum_1^n X_i \geqq na) + \exp(-n^2))(1 + o(1)) . \end{aligned} \tag{2.5}$$

Thus, since $n^{-1}\log P(\sum_1^n X_i \geqq na) \to -a^2/2$ from (1.2), it follows from (2.5) that

$$n^{-1}\log P(T_n \geqq na) \to -a^2/2 .$$

Another interesting example which is related to the problem considered here may be found in Baum, Katz and Read (1962), page 196. There a sequence of random variables, $T_1, T_2, \cdots$, is constructed which is a martingale sequence such that

$$E(\exp(\lambda T_n)) = \infty \qquad \text{for each} \quad n \quad \text{and} \quad \lambda \neq 0 ,$$

but

$$\lim n^{-1}\log P(T_n \geqq na) = -\tfrac{1}{2} \qquad \text{for} \quad a > 0 .$$

REFERENCES

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

[2] Bahadur, R. R. and Rao, R. Ranga (1960). On deviations of the sample mean. *Ann. Math. Statist.* **31** 1015–1027.

[3] Baum, L. E., Katz, M. and Read, R. R. (1962). Exponential convergence rates for the law of large numbers. *Trans. Amer. Math. Soc.* **102** 187–199.

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

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

[6] Plachky, D. and Steinebach, J. (1975). A theorem about probabilities of large deviations with applications to queuing theory. *Period. Math. Hungar.* **5** 343–345.

[7] Roberts, A. and Varberg, D. (1973). *Convex Functions*. Academic Press, New York.

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

Department of Statistics
Pennsylvania State University
219 Pond Laboratory
University Park, Pennsylvania 16802