Z. Wahrscheinlichkeitstheorie verw. Geb. 26, 43–49 (1973)

A Large Deviation Theorem for Weighted Sums*

Stephen A. Book

Asymptotic representations are derived for large deviation probabilities of weighted sums of independent, identically distributed random variables. The main theorem generalizes a 1952 theorem of Chernoff which asserts that

$$n^{-1}\log P(S_n>cn)\to -\log\rho,$$

where S_n is the partial sum of a sequence of independent, identically distributed random variables $X_1, X_2, \ldots$ and ρ is a constant depending on X_1. The main result is similar in form to, but different in focus from, a particular case of Feller's (1969) theorem on large deviations for triangular arrays.

1. Preliminaries

If $X_1, X_2, \ldots$ is a sequence of independent, identically distributed nondegenerate random variables and $\{a_{nk}: 1\leqq k\leqq n,\ 1\leqq n<\infty\}$ is a double array of non-negative real numbers such that $\sum_{k=1}^{n} a_{nk}^2=1$ for all n, we want to study the asymptotic behavior of $P\left(S_n>c\sum_{k=1}^{n} a_{nk}\right)$, where $S_n=\sum_{k=1}^{n} a_{nk}X_k$ and $c>0$. That probability is estimated with the same degree of precision, i.e. logarithmic, as in the large deviation theorems of Chernoff [3], Sethuraman [9], Pinsky [8], Sievers [10], and Feller [6]. All these results, including the one in the present article, represent the asymptotic behavior less precisely than, for example, Cramér [4], Bahadur and Ranga Rao [1], and Book [2]. The theorems in [2] on weighted sums, while more precise than those here, apply at present only in the absolutely continuous case.

In this section, we state the conditions on the random variables and the weights, and we derive a representation of the large deviation probability as an integral of an "associated" distribution function. Paralleling the proof of Chernoff's theorem on pp. 1017–1018 of Bahadur and Ranga Rao [1], we need to know that the associated sums are asymptotically normal, and we show this in Section 2. Section 3 contains the proof of the main theorem and its reduction to Chernoff's theorem in the simpler case. The theorem and proof are compared in Section 4 with Feller's large deviation theorem for triangular arrays.

We require that the a_{nk}'s be such that the sum $S_n=\sum_{k=1}^{n} a_{nk}X_k$ is not dominated by a relatively few terms. This requirement produces an infinitesimal array, in

* This paper is based on work done for the author's doctoral dissertation written under Prof. Donald R. Truax of the University of Oregon, Eugene.

view of the fact that $\sum_{k=1}^{n} a_{nk}^2 = 1$ for all n. In particular, we ask

Condition I. There exist numbers α and θ, $0<\alpha\leqq 1$, $0<\theta\leqq 1$, such that, for every sufficiently large n, at least αn of the a_{nk}'s exceed or equal $\theta\sigma_n$, where $\sigma_n = \max\{a_{nk}: 1\leqq k\leqq n\}$.

Our sequence $X_1, X_2, \ldots$ of independent, identically distributed nondegenerate random variables has $E(X_1)=0$, $E(X_1^2)=1$, and moment-generating function $\phi(t)$ which we assume to exist in a nondegenerate interval $|t|<B$. We define the function $Q(t)=\phi'(t)/\phi(t)$, and we set, for simplicity of notation, $A_n=\sum_{k=1}^{n} a_{nk}$. We can determine the asymptotic representation for $P(S_n>cA_n)$ if, by analogy with a condition of Chernoff and Bahadur-Ranga Rao, we ask that ϕ satisfy

Condition II. Q assumes the value $c/\alpha\theta$ at some point, and $B_0=\theta^{-1}Q^{-1}(c/\alpha\theta)$ lies in the domain of ϕ.

Condition II holds, for example, if $\phi(t)<\infty$ for all real t and $P(X_1>c/\alpha\theta)>0$. Note that the larger the range of X_1 is, the smaller the values that α and θ are allowed to be.

As in Section 1 of [2], we define random variables $Y_{nk}=a_{nk}(X_k-c)$ and observe that $P(S_n>cA_n)=P\left(\sum_{k=1}^{n} Y_{nk}>0\right)$. The moment-generating function (m.g.f.) of Y_{nk} is $\phi_{nk}(h)=\exp(-hca_{nk})\,\phi(ha_{nk})$, where $\phi_{nk}(h)$ exists for $|h|<Ba_{nk}^{-1}$. Because each $a_{nk}\geqq 0$, all $\phi_{nk}(h)$, $1\leqq k\leqq n$, exist for $|h|<B\sigma_n^{-1}$. We restrict h to this interval from now on. We define an "associated" distribution function (d.f.) $\bar{H}_{nk}(y)$ by

$$d\bar{H}_{nk}(y)=\frac{e^{hy}}{\phi_{nk}(h)}\,dH_{nk}(y)$$

for each h, $0<h<B\sigma_n^{-1}$, where $H_{nk}(y)$ is the d.f. of Y_{nk}, and we denote by $\bar{Y}_{nk}$ a random variable distributed according to this d.f. We set $\bar{S}_n=\sum_{k=1}^{n}\bar{Y}_{nk}$ and $A_n=\sum_{k=1}^{n} a_{nk}$. We have the following formula, whose proof is identical in form with the proof of the analogous Lemma 2 on p. 1017 of Bahadur and Ranga Rao [1]:

Lemma 1. *If* $\bar{H}_n(y)=P(\bar{S}_n\leqq y)$, *then*

$$P(S_n>cA_n)=\exp(-hcA_n)\left[\prod_{k=1}^{n}\phi(ha_{nk})\right] I_n(h),$$

where $I_n(h)=h\int_0^\infty e^{-hy}[\bar{H}_n(y)-\bar{H}_n(0)]\,dy$.

In order to make use of the formula of Lemma 1, we must study the behavior of the quantity $\bar{H}_n(y)-\bar{H}_n(0)$. We want to find conditions under which the quantity can be approximated by the corresponding normal probability $\Phi(y)-\Phi(0)$, so that it will be bounded away from 0 for each $y>0$ as $n\to\infty$.

2. Asymptotic Normality of $\bar{S}_n$

The "associated" random variable $\bar{Y}_{nk}$ has m.g.f. $\bar{\phi}_{nk}(t)=\phi_{nk}(t+h)/\phi_{nk}(h)$, where $\phi_{nk}(h)=\exp(-h\,c\,a_{nk})\,\phi(h\,a_{nk})$. Therefore $\bar{S}_n$ has m.g.f.

$$\bar{\phi}_n(t)=\prod_{k=1}^{n}[\phi_{nk}(t+h)/\phi_{nk}(h)].$$

The moment-generating properties of the m.g.f. imply that

$$E(\bar{S}_n)=\sum_{k=1}^{n}\bar{\phi}'_{nk}(0) \quad\text{and}\quad \operatorname{Var}(\bar{S}_n)=\sum_{k=1}^{n}[\bar{\phi}''_{nk}(0)-(\bar{\phi}'_{nk}(0))^2].$$

Upon computing these derivatives from the expression for $\bar{\phi}_{nk}(t)$ and recalling that $Q(t)=\phi'(t)/\phi(t)$, we obtain explicit expressions as follows:

Lemma 2.

$$E(\bar{S}_n)=\sum_{k=1}^{n}a_{nk}\,Q(h\,a_{nk})-c\,A_n,$$

and

$$\operatorname{Var}(\bar{S}_n)=\sum_{k=1}^{n}a_{nk}^2\,Q'(h\,a_{nk}).$$

It turns out that, under Conditions I and II, there exists a sequence $\{h_n\colon 1\leqq n<\infty\}$ of h's to appear in the definition of the associated random variables such that $E(\bar{S}_n)=0$ and the variances $\operatorname{Var}(\bar{S}_n)$ are uniformly bounded away from 0 and ∞. In particular, we have the following two results, whose proofs can be found in Section 2 of [2]:

Lemma 3. *Under Conditions I and II, there exists, for every positive integer n, a solution $h=h_n$ of the equation $E(\bar{S}_n)=0$, and the solution satisfies the inequalities*

$$b_0=Q^{-1}(c\,\alpha\,\theta^2)\leqq h_n\,\sigma_n\leqq\theta^{-1}\,Q^{-1}(c/\alpha\,\theta)=B_0.$$

Lemma 4. *Under Conditions I and II, there exist numbers $d_0^2>0$ and $D_0^2<\infty$ such that $d_0^2\leqq\operatorname{Var}(\bar{S}_n)\leqq D_0^2$.*

From now on, we assume that the associated random variables are constructed with $h=h_n$. The following result, based on the continuity theorem for moment-generating functions of Curtiss [5], establishes the asymptotic normality of the associated sums. For simplicity of notation, we set $\bar{\sigma}_n^2=\operatorname{Var}(\bar{S}_n)$.

Lemma 5. *Under Conditions I and II, $\mathscr{L}(\bar{S}_n/\bar{\sigma}_n)\to N(0,1)$ as $n\to\infty$.*

Proof. $\bar{S}_n$ has m.g.f.

$$\bar{\phi}_n(t)=\exp(-t\,c\,A_n)\prod_{k=1}^{n}[\phi((t+h_n)\,a_{nk})/\phi(h_n\,a_{nk})],$$

which exists when $(t+h_n)\,\sigma_n<B$, so when $|t|<\sigma_n^{-1}(B-h_n\,\sigma_n)$, and so when $|t|\leqq B-B_0$ since $h_n\,\sigma_n\leqq B_0$ and $\sigma_n<1$. The m.g.f. $\gamma_n(t)$ of $\bar{S}_n/\bar{\sigma}_n$ is given by

$$\gamma_n(t)=\bar{\phi}_n(t/\bar{\sigma}_n)=e^{-(t\,c\,A_n)/\bar{\sigma}_n}\prod_{k=1}^{n}\frac{\phi((t\,\bar{\sigma}_n^{-1}+h_n)\,a_{nk})}{\phi(h_n\,a_{nk})},$$

and exists for $|t|\leqq\bar{\sigma}_n(B-B_0)$, so for $|t|\leqq d_0(B-B_0)$ by Lemma 4. Using a Taylor series expansion with remainder about $t=0$ of $\log\phi((t\,\bar{\sigma}_n^{-1}+h_n)\,a_{nk})$, we have, for

$|t| \leqq d_0(B-B_0)$,

$$\begin{aligned}\log \gamma_n(t) &= -(t\,c\,A_n)/\bar{\sigma}_n + \sum_{k=1}^{n} \log \phi\left((t\,\bar{\sigma}_n^{-1} + h_n)\, a_{nk}\right) \\ &\quad - \sum_{k=1}^{n} \log \phi(h_n\, a_{nk}) \\ &= -(t\,c\,A_n)/\bar{\sigma}_n + \sum_{k=1}^{n} t\,\bar{\sigma}_n^{-1}\, a_{nk}\, Q(h_n\, a_{nk}) \\ &\quad + \sum_{k=1}^{n} 2^{-1}\, t^2\, \bar{\sigma}_n^{-2}\, a_{nk}^2\, Q'(h_n\, a_{nk}) \\ &\quad + \sum_{k=1}^{n} 6^{-1}\, t^3\, \bar{\sigma}_n^{-3}\, a_{nk}^3\, Q''\left(\theta_{nk}\, t\,\bar{\sigma}_n^{-1}\, a_{nk} + h_n\, a_{nk}\right),\end{aligned}$$

where $|\theta_{nk}| < 1$. Now by Lemma 2 and the choice of h_n, the expansion reduces to

$$\log \gamma_n(t) = t^2/2 + (t^3/6\,\bar{\sigma}_n^3) \sum_{k=1}^{n} a_{nk}^3\, Q''(t_{nk}),$$

where $t_{nk} = (\theta_{nk}\, t\,\bar{\sigma}_n^{-1} + h_n)\, a_{nk}$. To apply the theorem of Curtiss [5], it remains only to show that the last term tends to 0 for all t in a nondegenerate closed interval about the origin. But for $|t| \leqq d_0(B-B_0)/2$, we have

$$\begin{aligned}|Q''(t_{nk})| &\leqq \max\{|Q''(x)|:\ |x| \leqq d_0(B-B_0)/2\,\bar{\sigma}_n + h_n\, a_{nk}\} \\ &\leqq \max\{|Q''(x)|:\ |x| \leqq (B+B_0)/2\} = M < \infty,\end{aligned}$$

because Q'' is continuous for $|t| \leqq d_0(B-B_0)/2$. Therefore, for t fixed,

$$\begin{aligned}\left|(t^3/6\,\bar{\sigma}_n^3) \sum_{k=1}^{n} a_{nk}^3\, Q''(t_{nk})\right| &\leqq (|t|^3/6\,\bar{\sigma}_n^3)\, M \sum_{k=1}^{n} a_{nk}^3 \\ &\leqq (|t|^3\, M/6\, d_0^3)\, \sigma_n \sum_{k=1}^{n} a_{nk}^2\end{aligned}$$

tends to 0 as $n \to \infty$ because $1 = \sum_{k=1}^{n} a_{nk}^2 \geqq \alpha\, n\, \theta\, \sigma_n^2$ implies that $\sigma_n \to 0$ as $n \to \infty$.

3. The Large Deviation Theorem

It follows from Lemma 5 that, as $n \to \infty$, $\bar{H}_n(y\,\bar{\sigma}_n) - \bar{H}_n(0)$ tends to $\Phi(y) - \Phi(0)$. This situation yields us the following fact:

Lemma 6. *Under Conditions I and II,* $\lim_{n\to\infty} (h_n\, A_n)^{-1} \log I_n(h_n) = 0$.

Proof. For any $\varepsilon > 0$, Lemma 4 implies that

$$\begin{aligned}I_n(h_n) &\geqq h_n \int_{\varepsilon}^{\infty} e^{-h_n y}\, [\bar{H}_n(y) - \bar{H}_n(0)]\, dy \geqq [\bar{H}_n(\varepsilon) - \bar{H}_n(0)]\, e^{-h_n \varepsilon} \\ &\geqq e^{-h_n \varepsilon}\, P(0 < \bar{S}_n/\bar{\sigma}_n \leqq \varepsilon\, D_0^{-1}).\end{aligned}$$

It then follows from Lemma 5 that $\liminf_{n\to\infty}(h_n A_n)^{-1}\log I_n(h_n)=0$, since $A_n \geqq \alpha n \theta \sigma_n \to \infty$ as $n\to\infty$ because $1=\sum_{k=1}^{n} a_{nk}^2 \leqq n\sigma_n^2$, and $h_n\to\infty$ as $n\to\infty$ because $h_n\sigma_n \geqq b_0>0$ by Lemma 3 and $\sigma_n\to 0$. On the other hand,

$$I_n(h_n)\leqq h_n \int_0^\infty \exp(-h_n y)\,dy=1,$$

so that $\limsup_{n\to\infty}(h_n A_n)^{-1}\log I_n(h_n)\leqq 0$.

The main theorem now follows, setting $L(t)=t\,Q(t)-\log\phi(t)$.

Theorem. *Under Conditions I and II,* $P(S_n>c\,A_n)=e^{-r_n+o(r_n)}$, *where* $r_n=c\,b_n h_n A_n$ *for a bounded sequence of positive numbers* b_n *such that*

$$0<(\alpha/c\,B_0)\,L(\theta\, b_0)\leqq b_n\leqq 1-(\alpha/c\,B_0)\log\phi(\theta\, b_0)<1.$$

Proof. From Lemma 1 and the conclusion of Lemma 6, we need be concerned only with

$$\log J_n=\log\left\{\exp(-h_n\,c\,A_n)\prod_{k=1}^{n}\phi(h_n a_{nk})\right\}$$

$$=-h_n\,c\,A_n+\sum_{k=1}^{n}\log\phi(h_n a_{nk}).$$

On the one hand,

$$\log J_n\geqq -h_n\,c\,A_n+\alpha\, n\log\phi(h_n\theta\sigma_n)\geqq -h_n\,c\,A_n+\alpha\, n\log\phi(\theta\, b_0),$$

so that

$$\liminf_{n\to\infty}(c\,h_n A_n)^{-1}\log J_n\geqq -1+(\alpha/c\,B_0)\log\phi(\theta b_0),$$

because $\alpha\, n(c\,h_n A_n)^{-1}\geqq \alpha\, n(c\,n B_0)^{-1}=\alpha/c\,B_0$, as $h_n A_n\leqq h_n\sum_{k=1}^{n} a_{nk}\leqq h_n\, n\,\sigma_n\leqq n B_0$.

On the other hand, by Lemma 2 and the definition of h_n,

$$\log J_n=-h_n\sum_{k=1}^{n} a_{nk}\,Q(h_n a_{nk})+\sum_{k=1}^{n}\log\phi(h_n a_{nk})$$

$$=-\sum_{k=1}^{n} L(h_n a_{nk})\leqq -\alpha\, n\, L(\theta\, b_0)$$

because $L(t)$ is positive for all positive t and is increasing in t, since the (non-degenerate) associated distributions have variances $Q'(t)$. It follows that

$$\limsup_{n\to\infty}(c\,h_n A_n)^{-1}\log J_n\leqq -(\alpha/c\,B_0)\,L(\theta\, b_0).$$

Therefore, $-(c\,h_n A_n)^{-1}\log P(S_n>c\,A_n)=b_n+\delta_n$, where b_n is bounded as in the conclusion of the theorem, and $\delta_n\to 0$ as $n\to\infty$. Therefore

$$\log P(S_n>c\,A_n)=-r_n+o(r_n),$$

and the theorem is proved.

As a corollary, we obtain

Chernoff's Theorem. *If* $S_n = \sum_{k=1}^{n} X_k$ *and there exists a* $\tau > 0$ *such that* $Q(\tau) = c$, *then*

$$\lim_{n\to\infty} n^{-1} \log P(S_n > cn) = -\log \rho,$$

where $\rho = e^{-c\tau} \phi(\tau)$.

Proof. In this case, we would have $\alpha = 1$ and $\theta = 1$ in Condition I and so Condition II is equivalent to the existence of a $\tau > 0$ with $Q(\tau) = c$. Then $A_n = n^{\frac{1}{2}}$, $\sigma_n = n^{-\frac{1}{2}}$, and $S_n = n^{-\frac{1}{2}} \sum_{k=1}^{n} X_k$ in the main theorem. The bounds b_0 and B_0 of Lemma 3 have the common value τ so that $h_n = \tau n^{\frac{1}{2}}$. The bounds on b_n in the main theorem have the common value $(c\tau)^{-1}(c\tau - \log \phi(\tau)) = (c\tau)^{-1}(-\log \rho)$, and therefore $r_n = -n \log \rho$. Chernoff's theorem follows.

4. Comparison with Feller's Theorem

If Theorem 4, p. 14, of Feller's paper [6] is applied to the weighted sum case, it reads as follows in our notation:

Feller's Theorem. *Under some conditions on* X_1 *and the weights, if* $\{h_n^*: 1 \leq n < \infty\}$ *is a sequence of numbers such that*

$$r_n^* = h_n^* \sum_{k=1}^{n} a_{nk} Q(h_n^* a_{nk}) - \sum_{k=1}^{n} \log \phi(h_n^* a_{nk}) \to \infty$$

as $n \to \infty$, *and*

$$z_n^* = \sum_{k=1}^{n} a_{nk} Q(h_n^* a_{nk}),$$

then

$$P(S_n > z_n^*) = \exp(-r_n^* + o(r_n^*)).$$

In formulating his theorem, Feller was primarily interested in deriving the analogues of the law of the iterated logarithm in [7], so he wanted the sequence of h_n^*'s to give r_n^*'s and z_n^*'s in accordance with this goal. The h_n^*'s are used only as a bridge between z_n^* and r_n^*; they hold no intrinsic interest, so no attempt is made to study their behavior, or even their existence for particular z_n^*'s and r_n^*'s. On the other hand, Feller's results are much more wide-ranging, in the sense that any sequence of h_n^*'s, which results in $r_n^* \to \infty$ as $n \to \infty$, is considered.

In the present paper, only the sequence $z_n = c A_n$ is considered, since the goal is an extension of Chernoff's large deviation theorem. The method of proof and the conditions are, of course, quite a bit simpler than Feller's, and really only a generalization of Bahadur and Ranga Rao's technique on pp. 1017–1018 of [1]. Based on the choice of $z_n = c A_n$, the existence and some properties of the corresponding h_n's are established, and then the r_n's are determined later. This produces a result which reduces in a straightforward manner to Chernoff's original theorem, since the h_n's can be precisely calculated in that case.

Our Condition II, an analogue of the only condition in the Chernoff and Bahadur-Ranga Rao articles, guarantees that the equation $z_n^* = c A_n$ has a solution $h = h_n^*$, and provides some estimate of the sequence of h_n^*'s. The conditions of

Feller's theorem, on the other hand, do not always guarantee that the parametric equations for z_n^* and r_n^* in terms of h_n^* are solvable. In fact, for his more general results, he includes as a specific condition that the equations be solvable, as well as that $r_n^*(h_n) \to \infty$ as $n \to \infty$. In our more restricted situation, Condition II itself assures this, while assuming nothing more than the natural analogue of Chernoff's original condition.

Acknowledgment. The author would like to thank Prof. D. R. Truax for several discussions on questions leading to the results presented here.

References

1. Bahadur, R. R., Ranga Rao, R.: On deviations of the sample mean. Ann. Math. Statistics **31**, 1015–1027 (1960).
2. Book, S. A.: Large deviation probabilities for weighted sums. Ann. Math. Statistics **43**, 1221–1234 (1972).
3. Chernoff, H.: A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. Ann. Math. Statistics **23**, 493–507 (1952).
4. Cramér, H.: Sur un nouveau théorème-limite de la théorie des probabilitiés. Actualités Scientifiques et Industrielles, No. **736**, 5–23 (1938).
5. Curtiss, J. H.: A note on the theory of moment generating functions. Ann. Math. Statistics **13**, 430–433 (1942).
6. Feller, W.: Limit theorems for probabilities of large deviations. Z. Wahrscheinlichkeitstheorie verw. Gebiete **14**, 1–20 (1969).
7. Feller, W.: General analogues to the law of the iterated logarithm. Z. Wahrscheinlichkeitstheorie verw. Gebiete **14**, 21–26 (1969).
8. Pinsky, M.: An elementary derivation of Khintchine's estimate for large deviations. Proc. Amer. Math. Soc. **22**, 288–290 (1969).
9. Sethuraman, J.: Probabilities of deviations. S. N. Roy Memorial Vol., Univ. of N. Carolina Press, 655–672 (1970).
10. Sievers, G. L.: On the probability of large deviations and exact slopes. Ann. Math. Statistics **40**, 1908–1921 (1969).

Stephen A. Book
Department of Mathematics
California State College
Dominguez Hills, Calif. 90747
USA

(Received April 25, 1972)