Z. Wahrscheinlichkeitstheorie verw. Gebiete 31, 107–113 (1975)

Large Deviation Probabilities in the Strong Law of Large Numbers *

D. Siegmund

1. Introduction

Let $x_1, x_2, \dots$ be independent random variables having a common distribution function F with mean 0. Let $s_n = \sum_1^n x_k$. The weak and strong laws of large numbers say respectively that for any $\varepsilon > 0$ the sequence of real numbers

$$P_m = P\{m^{-1} s_m > \varepsilon\} \tag{1}$$

and

$$P_m^* = P\{\sup_{n \geqq m} n^{-1} s_n > \varepsilon\} \tag{2}$$

converge to 0 as $m \to \infty$. Under conditions on the moment generating function $\int_{-\infty}^{\infty} \exp(\theta x) F(dx)$ Blackwell and Hodges (1959) in the case of lattice variables x_k and Bahadur and Ranga Rao (1960) in the general case have found an asymptotic approximation for P_m (and under further assumptions a complete asymptotic expansion). The principal result of this note gives under certain conditions the dominant term of such an asymptotic expansion for P_m^*. A precise statement appears in Section 2 and the proof in Section 3. The relation of this result to those of Strassen (1965), Müller (1968), and Robbins, Siegmund, and Wendel (1968) is discussed in Section 4.

2. Embedding *F* and Statement of Principal Theorem

Assume that $g(\xi) = \log \int_{-\infty}^{\infty} \exp(\xi y) F(\varepsilon + dy)$ is finite for ξ in some open interval I containing 0, where ε is the arbitrary (but fixed) positive number appearing in (2). From the well-known properties of the cumulant generating function g it follows that $g(0) = 0$, $g'(0) = -\varepsilon$ and g is strictly convex (unless F is degenerate at 0). Hence there exists at most one value ξ_1, necessarily positive, for which $g(\xi_1) = 0$. Assume that such a value ξ_1 exists and let ξ_0 denote the point in $(0, \xi_1)$ at which g' vanishes. Let

$$\psi(\theta) = g(\xi_0 + \theta) - g(\xi_0) \tag{3}$$

and

$$\theta_0 = -\xi_0, \quad \theta_1 = \xi_1 - \xi_0. \tag{4}$$

Then

$$\theta_0 < 0 < \theta_1, \tag{5}$$

$$\psi(0) = \psi'(0) = 0, \tag{6}$$

* Research supported in part by National Science Foundation grant NSF-GP-33570X.

and

$$\psi(\theta_0)=\psi(\theta_1). \tag{7}$$

Now for each real θ for which $\psi(\theta)<\infty$ let P_θ be the probability on the space of infinite sequences $(x_1, x_2, \ldots)$ under which the x's are independent with the common distribution function

$$P_\theta\{x_k \leqq x\} = \int_{-\infty}^{x} \exp[\theta y - \psi(\theta)] \exp[\xi_0 y - g(\xi_0)]\, F(\varepsilon + dy). \tag{8}$$

Then $P_{\theta_0}\{x_k \leqq x\} = F(x+\varepsilon)$ and hence by (2)

$$P_m^* = P_{\theta_0}\{s_n > 0 \text{ for some } n \geqq m\}. \tag{9}$$

Theorem 1. *In addition to the preceding assumptions, if either*

$$F \text{ is non-lattice} \tag{10}$$

or

$$F(\varepsilon+\cdot) \text{ is a lattice distribution supported by } \{0, \pm h, \pm 2h, \ldots\} \text{ for some } h>0 \tag{11}$$

then as $m \to \infty$

$$P_m^* \sim P_{\theta_0}\{s_m > 0\} + c\, P_{\theta_1}\{s_m \leqq 0\}, \tag{12}$$

where if (10) *holds*

$$c = |\theta_0|^{-1}\,\theta_1 \left[\exp\left\{\sum_1^\infty n^{-1}[e^{-n\psi(\theta_0)} P_0(s_n>0) - P_{\theta_0}(s_n>0)]\right\} - 1\right] \tag{13}$$

and if (11) *holds*

$$c = (\exp(|\theta_0|\, h) - 1)^{-1} (\exp(\theta_1 h) - 1) \cdot \left[\exp\left\{\sum_1^\infty n^{-1}[e^{-n\psi(\theta_0)} P_0(s_n>0) - P_{\theta_0}(s_n>0)]\right\} - 1\right]. \tag{14}$$

Asymptotic expressions for the probabilities appearing on the right hand side of (12) are given by Bahadur and Ranga Rao (1960). They will be reproduced (cf. Lemma 1) as part of the calculations which give the value of c. In the case that $F=\Phi$, the standard normal distribution function, $-\theta_0=\theta_1=\varepsilon$ and

$$P_m^* \sim \Phi(-\varepsilon m^{1/2}) \exp\left(-\sum_1^\infty n^{-1}\, \Phi(-\varepsilon n^{1/2})\right) \Big/ (1-\exp(-\varepsilon^2/2))^{1/2}. \tag{15}$$

If $w(t)$ denotes a standard Brownian motion process, then for all $m>0$ (cf. Robbins, Siegmund, and Wendel (1968))

$$P\{w(t) > \varepsilon t \text{ for some } t \geqq m\} = 2\,\Phi(-\varepsilon m^{1/2}). \tag{16}$$

For $\varepsilon=0.5$, 0.3, 0.1, or 0.04 the ratio of constants on the right hand side of (15) is respectively 1.54, 1.70, 1.89, or 1.96, and as $\varepsilon \to 0$ it converges to 2. Further remarks on the relation between (12) as $\varepsilon \to 0$ and (16) are contained in Section 4.

Embedding F in a family of measures for studying first passage and large deviation probabilities is a common technique which has been given thorough exposition by Feller (1966). The details of my presentation, which are slightly different than the usual ones, are particularly well adapted to proving (12), and I find that they make the analysis of (1) more comprehensible also.

3. Proof of Theorem 1

Let $P_\theta^{(n)}$ denote the restriction of P_θ to the σ-algebra of $x_1, x_2, \ldots, x_n$ $(n=1,2,\ldots)$. Then for any θ and θ', $P_\theta^{(n)}$ and $P_{\theta'}^{(n)}$ are mutually absolutely continuous, and by (8)

$$dP_{\theta'}^{(n)}/dP_\theta^{(n)}=\exp\left((\theta'-\theta)s_n-n[\psi(\theta')-\psi(\theta)]\right). \tag{17}$$

In particular, by (7) and (17)

$$dP_{\theta_0}^{(n)}/dP_{\theta_1}^{(n)}=\exp[-(\theta_1-\theta_0)s_n] \qquad (n=1,2,\ldots). \tag{18}$$

Also it follows easily from (8) that $\psi'(\theta)=E_\theta x_1$ and hence by (6) and the strict convexity of ψ

$$E_\theta x_1 <, =, \text{ or } >0 \quad \text{according as} \quad \theta <, =, \text{ or } >0. \tag{19}$$

Let $T_m=\inf\{n\colon n\geqq m,\ s_n>0\}$. (The inf of the empty set is $+\infty$.) Then by (9)

$$P_m^*=P_{\theta_0}\{s_m>0\}+P_{\theta_0}\{m<T_m<\infty\}. \tag{20}$$

and by (18)

$$\begin{aligned} P_{\theta_0}\{m<T_m<\infty\} &= \sum_{n=m+1}^{\infty} \int_{\{T_m=n\}} \exp[-(\theta_1-\theta_0)s_n]\,dP_{\theta_1} \\ &= \int_{\{m<T_m<\infty\}} \exp[-(\theta_1-\theta_0)S_{T_m}]\,dP_{\theta_1}. \end{aligned} \tag{21}$$

By (5) and (19) $E_{\theta_1}x_1>0$; hence $\{m<T_m<\infty\}=\{T_m>m\}=\{s_m\leqq 0\}$ a.s. P_{θ_1}, and the final integral in (21) equals

$$\int_{\{s_m\leqq 0\}} \exp[-(\theta_1-\theta_0)s_{T_m}]\,dP_{\theta_1}=P_{\theta_1}\{s_m\leqq 0\}\,E_{\theta_1}\left(\exp[-(\theta_1-\theta_0)s_{T_m}]\,|\,s_m\leqq 0\right),$$

so that by (20) and (21)

$$P_m^*=P_{\theta_0}\{s_m>0\}+P_{\theta_1}\{s_m\leqq 0\}\int_{(-\infty,0]} E_{\theta_1}\left(\exp[-(\theta_1-\theta_0)s_{T_m}]\,|\,s_m=y\right)P_{\theta_1}\{s_m\in dy\,|\,s_m\leqq 0\}. \tag{22}$$

A comparison of (22) and (12) shows that the proof will be completed by showing that the integral appearing in (22) converges to a constant c and evaluating that constant.

The conditional expectation in the integrand in (22) is actually independent of m and equals

$$E_{\theta_1}\left(\exp[-(\theta_1-\theta_0)(s_{\tau(|y|)}-|y|)]\right), \tag{23}$$

where $\tau(x)=\inf\{n\colon s_n>x\}$. Moreover, by Lemma 2 below, under P_{θ_1} the conditional distribution of s_m given that $s_m\leqq 0$ converges in the non-lattice case to a negative exponential distribution with probability density function $\theta_1 e^{\theta_1 y}$ $(y\leqq 0)$ and in the lattice case to a geometric distribution with probability mass function

$$(1-\exp(-\theta_1 h))\exp(\theta_1 h k) \qquad (k=0,-1,-2,\ldots).$$

Considered as a function of $x>0$, except at values c belonging to the denumerable set $A=\{c'\colon \sum_1^\infty P_{\theta_1}\{s_k=c'\}>0\}$, $s_{\tau(x)}-x$ is continuous with probability one and hence by dominated convergence, so is (23). (Obviously $\tau(\cdot)$ is non-decreasing and right continuous; and if $x \notin A$ but for some k $\lim_{x' \uparrow x} \tau(x') = k < \tau(x)$ then for all $x' < x$ $x' < s_k \leqq x$, so $s_k = x$ and hence

$$P_{\theta_1}\{\lim_{x' \uparrow x} \tau(x') < \tau(x)\} \leqq \sum_1^\infty P_{\theta_1}\{s_k = x\} = 0.)$$

It follows in the non-lattice case that

$$\begin{aligned}\lim_{m \to \infty} & \int_{(-\infty, 0]} E_{\theta_1}(\exp[-(\theta_1 - \theta_0)(s_{\tau(|y|)} - |y|)])\, P_{\theta_1}\{s_m \in dy \mid s_m \leqq 0\} \\ &= \theta_1 \int_{(0, \infty)} E_{\theta_1}(\exp[-(\theta_1 - \theta_0)(s_{\tau(x)} - x)]) \exp(-\theta_1 x)\, dx,\end{aligned} \tag{24}$$

and in the lattice case, since every function is continuous in the discrete topology, the limit on the left hand side of (24) exists and equals

$$(1 - \exp(-\theta_1 h)) \sum_{k=0}^\infty E_{\theta_1}(\exp[-(\theta_1 - \theta_0)(s_{\tau(hk)} - h k)]) \exp(-\theta_1 h k). \tag{25}$$

It remains to show that the right hand side of (24) and (25) equal respectively the values given in (13) and (14). This is accomplished by standard renewal theoretic arguments in conjunction with well-known results of Spitzer. Details are given only for the non-lattice case. Let $G(y) = P_{\theta_1}\{s_{\tau(0)} \leqq y\}$ $(y > 0)$ and

$$Z(x) = E_{\theta_1}(\exp[-(\theta_1 - \theta_0)(s_{\tau(x)} - x)]).$$

The standard renewal argument shows that Z satisfies the renewal equation

$$Z(x) = \int_{(x, \infty)} \exp(-(\theta_1 - \theta_0)(y - x)]\, G(dy) + \int_{(0, x]} Z(x - y)\, G(dy). \tag{26}$$

The right hand side of (24) is θ_1 times

$$Z^*(\theta_1) = \int_{(0, \infty)} \exp(-\theta_1 x)\, Z(x)\, dx,$$

which by (26) equals

$$\int_{(0, \infty)} \exp(-\theta_1 x) \int_{(x, \infty)} \exp[-(\theta_1 - \theta_0)(y - x)]\, G(dy)\, dx + Z^*(\theta_1)\, G^*(\theta_1), \tag{27}$$

where

$$G^*(\lambda) = \int_{(0, \infty)} \exp(-\lambda x)\, G(dx) \qquad (\lambda > 0).$$

Now Fubini's theorem applied to the iterated integral in (27) together with elementary computations yields

$$Z^*(\theta_1) = [(1 - G^*(\theta_1 - \theta_0))/(1 - G^*(\theta_1)) - 1]/|\theta_0|. \tag{28}$$

According to a result of Spitzer (cf. Chung, 1968, p. 262)

$$1 - G^*(\lambda) = 1 - E_{\theta_1} \exp(-\lambda s_{\tau(0)}) = \exp\left(-\sum_1^\infty n^{-1} \int_{(s_n > 0)} \exp(-\lambda s_n)\, dP_{\theta_1}\right). \tag{29}$$

From (17) and (18) it follows that

$$\int_{(s_n>0)} \exp[(\theta_0-\theta_1)\, s_n]\, dP_{\theta_1} = P_{\theta_0}(s_n>0) \tag{30}$$

and

$$\int_{(s_n>0)} \exp(-\theta_1\, s_n)\, dP_{\theta_1} = \exp(-n\,\psi(\theta_0))\, P_0(s_n>0). \tag{31}$$

Combining (28)-(31) shows that the right hand side of (24) equals

$$\theta_1\, Z^*(\theta_1) = |\theta_0|^{-1}\,\theta_1 \left[\exp\left\{\sum_1^\infty n^{-1}\,[e^{-n\psi(\theta_0)}\, P_0(s_n>0) - P_{\theta_0}(s_n>0)]\right\} - 1\right],$$

which completes the proof.

Lemma 2 below justifies the earlier assertion concerning the limiting behavior under P_{θ_1} of the conditional distribution of s_m given that $s_m \leqq 0$. Lemma 1 gives a slight generalization of the dominant term of the Bahadur-Range Rao results by essentially their method.

Let $\sigma^2 = E_0\, x_1^2$.

Lemma 1. (a) *Under the assumption* (10) *for each* $x \leqq 0$, *as* $m \to \infty$

$$P_{\theta_1}\{s_m \leqq x\} \sim (2\pi\,\sigma^2\,\theta_1^2\, m)^{-1/2} \exp(-m\,\psi(\theta_1) + \theta_1\, x). \tag{32}$$

(b) *Under the assumption* (11), *for each* $k = 0, -1, -2, \ldots$ *as* $m \to \infty$

$$P_{\theta_1}\{s_m \leqq h\,k\} \sim (2\pi\,\sigma^2\, m)^{-1/2}\, h\,(1-\exp(-\theta_1\, h))^{-1} \exp(-m\,\psi(\theta_1) + \theta_1\, h\,k). \tag{33}$$

Proof. (b) Assume that (11) is satisfied. Putting $\theta' = \theta_1$ and $\theta = 0$ in (17) and recalling (6) yields

$$P_{\theta_1}\{s_m \leqq h\,k\} = \exp(-m\,\psi(\theta_1)) \int_{(s_m \leqq hk)} \exp(\theta_1\, s_m)\, dP_0. \tag{34}$$

Under (11) the integral appearing in (34) is actually $\sum_{j \leqq k} \exp(\theta_1\, h\,j)\, P_0\{s_m = h\,j\}$ which can be split into $\sum_{j \leqq -A_m}$ and $\sum_{-A_m < j \leqq k}$, where

$$m^{-1/4}\, A_m \to \infty, \qquad m^{-1/2}\, A_m \to 0. \tag{35}$$

Then (19) and the local limit theorem for lattice distributions (cf. Feller, 1966, p. 490) yield for every $\delta > 0$ and m sufficiently large the following double inequality:

$$\begin{aligned} (2\pi\,\sigma^2\, m)^{-1/2}\, h\,(1-\delta) \sum_{-A_m < j \leqq k} \exp(\theta_1\, h\,k) &\leqq \int_{(s_m \leqq hk)} \exp(\theta_1\, s_m)\, dP_0 \\ &\leqq \exp(-\theta_1\, h\,A_m) + (2\pi\,\sigma^2\, m)^{-1/2}\, h\,(1+\delta) \sum_{-A_m < j \leqq k} \exp(\theta_1\, h\,k). \end{aligned} \tag{36}$$

By (35) the term $\exp(-\theta_1\, h\,A_m)$ in (36) is $o(m^{-1/2})$ while the two series are asymptotically $\sum_{j \leqq k} \exp(\theta_1\, h\,k) = (1-\exp(-\theta_1\, h))^{-1} \exp(\theta_1\, h\,k)$. Hence (36) together with (34) completes the proof, since $\delta > 0$ is arbitrary.

(a) Eq. (32) may be obtained formally from (33) by letting $h \to 0$ or equivalently by replacing $h \cdot$ counting measure by Lebesgue measure in the preceding proof. However, the usual local limit theorems for non-lattice distributions are not given in a form which makes this argument precise and hence it is easier to integrate (34) (with hk replaced by x) by parts and use Esseen's theorem (cf. Feller, 1966, p. 512) to prove (32).

Lemma 2. *Under* (10) *or* (11) *respectively*

$$\lim_{m\to\infty} P_{\theta_1}\{s_m \leqq x \mid s_m \leqq 0\} = \exp(\theta_1 x) \qquad (x \leqq 0) \tag{37}$$

or

$$\lim_{m\to\infty} P_{\theta_1}\{s_m \leqq hk \mid s_m \leqq 0\} = \exp(\theta_1 hk) \qquad (k=0, -1, -2, \ldots). \tag{38}$$

Proof. Since $P_{\theta_1}\{s_m \leqq x \mid s_m \leqq 0\} = P_{\theta_1}\{s_m \leqq x\}/P_{\theta_1}\{s_m \leqq 0\}$ for all $x \leqq 0$, the stated results follow at once from Lemma 1.

4. Remarks

(a) Let $x_1, x_2, \ldots$ be independent random variables having a common distribution with mean 0, variance 1, and finite moment generating function in some neighborhood of 0. Strassen (1965) obtained an asymptotic expression for $P\{s_n \geqq b(n)$ for some $n \geqq m\}$ as $m \to \infty$ for a class of curves b which among other conditions do not increase faster than $n^{3/5}$. One interesting feature of Strassen's result is that asymptotically this probability does not depend on the underlying distribution of the x's and is the same as that for standard Brownian motion. In contrast the asymptotic value given by Theorem 1 for P_m^* is highly distribution dependent, and as (15) and (16) show even for normally distributed x's the result is not the same as for Brownian motion.

(b) Let $x_1, x_2, \ldots$ be i.i.d. with mean 0 and variance 1. Independently Müller (1968) and Robbins, Siegmund, and Wendell (1968) showed that if $\varepsilon = \varepsilon(m) \to 0$ in such a way that $\varepsilon m^{1/2} \to a \geqq 0$ as $m \to \infty$, then

$$\lim_{m\to\infty} P\{\sup_{n\geqq m} n^{-1} s_n > \varepsilon\} = P\{w(t) > a\,t \text{ for some } t \geqq 1\} = 2\Phi(-a). \tag{39}$$

This leads one to conjecture that the convergence indicated by (12) holds uniformly in ε for ε near 0 and that c considered as a function of ε converge to one as $\varepsilon \to 0$. This conjecture is correct, but my proof is rather tedious and has been omitted. It utilizes Lorden's (1970) bound on $\sup_{0\leqq x<\infty} E_{\theta_1}(s_{\tau(x)} - x)$ and estimates of the error in the central limit theorem.

References

1. Bahadur, R. R., Rao, R. Ranga: On deviations of the sample mean. Ann. Math. Statist. **31**, 1015–1027 (1960)
2. Blackwell, D., Hodges, J. L., Jr.: The probability in the extreme tail of a convolution. Ann. Math. Statist. **30**, 1113–1120 (1959)
3. Chung, K. L.: A course in Probability Theory. New York: Harcourt. Brace, and World 1968
4. Feller, W.: In Introduction to Probability Theory and Its Applications. Vol. II. New York: Wiley 1966
5. Lorden, G.: On excess over the boundary. Annals Math. Statist. **41**, 520–527 (1970)
6. Müller, D.W.: Verteilungs-Invarianzprinzipien für das Gesetz der großen Zahlen. Z. Wahrscheinlichkeitstheorie verw. Gebiete **10**, 173–192 (1968)

7. Robbins, H., Siegmund, D., Wendel, J.: The limiting distribution of the last time $s_n \geqq n\varepsilon$. Proc. Nat. Acad. Sci. USA **61**, 1228–1230 (1968)
8. Straßen, V.: Almost sure behavior of sums of independent random variables and martingales, Proc. 5[th] Berkeley Sympos. Math. Statist. Probab., Vol. II, 315–345. Univ. Calif. Press 1965

D. Siegmund
Dept. of Mathematical Statistics
Columbia University
New York, N.Y. 10027
USA

(Received January 16, 1974)