Strong large deviations for arbitrary sequences of random variables

Joutard, Cyrille
Université Montpellier 3, Département de Mathématiques
Route de Mende
34199 Montpellier, France
E-mail: cjoutard@math.univ-montp2.fr

1 Introduction

Let $X_1, X_2, \ldots, X_n, \ldots$ be a sequence of independent and identically distributed real random variables with zero mean and finite variance. For $a > 0$, the probability $P(\overline{X}_n \geq a)$ converges to 0. More precisely, we have the following logarithmic equivalent for this deviation probability :

$$\lim_{n\to\infty} \frac{1}{n} \log P(\overline{X}_n \geq a) = -I(a),$$

where I is the Fenchel-Legendre dual of the cumulant generating function of X_1 (I is usually called the rate function). This result is a consequence of the large deviation principle (LDP) satisfied by $\overline{X}_n$ and gives only an asymptotic equivalent for $\log P(\overline{X}_n \geq a)$ (for the general definition of a large deviation principle, we refer to [6]). In some cases, one may want to get asymptotic expressions for $P(\overline{X}_n \geq a)$. Bahadur and Rao [1] were among the first to establish such expressions for the sample mean. Such results are referred to as strong large deviation results (see Chaganty and Sethuraman [4]).

In addition to the theorems of [1] and [4] (who generalized the Bahadur-Rao Theorem on the sample mean to an arbitray sequence of random variables), several results pertaining to strong large deviations in asymptotic statistic can be found in the literature. Blackwell and Hodges [2] treated the lattice case of the Bahadur-Rao result on the sample mean. Generalizing the Bahadur-Rao result, Book [3] obtained a strong large deviation theorem for weighted sums of i.i.d. random variables. Chaganty and Sethuraman [5] proved a multidimensional version of their earlier result.

This paper provides strong large deviation results for an arbitray sequence of random variables Z_n. Some assumptions on the normalized cumulant generating function are assumed. We consider both the case where Z_n is absolutely continuous (or its distribution has an absolutely continuous component) and the case where Z_n is lattice-valued. Our results require, in particular, an asymptotic expansion of the normalized cumulant generating function. The proofs use techniques from [1] and [4] who also obtained strong large deviation theorems for an arbitrary sequence of random variables T_n. Note, however, that their large deviation expressions cannot generally be computed explicitly in a general frame. That is, one cannot (generally) derive an explicit asymptotic expression for the tail probability $P(T_n \geq c)$ that is a function of n. We illustrate some of our theorems with several statistical applications : the sample variance, the Wilcoxon signed-rank statistic and the Kendall's tau statistic. The paper is organized as follows. In Section 2, we introduce the framework and assumptions, before giving the main results and discussing the statistical applications. Section 3 deals with the lemmas needed for the proofs of the main results which are deferred to Section 4.

2 Main Results

2.1 Notation and assumption

Let Z_n be a sequence of random variables and let b_n be a sequence of real positive numbers such that $\lim_{n\to\infty} b_n = \infty$. Let ϕ_n be the moment generating function (m.g.f.) of $b_n Z_n$,

$$\phi_n(t) = E\{\exp(t b_n Z_n)\}, \quad t \in R,$$

and let φ_n be the normalized cumulant generating function (c.g.f.) of Z_n,

$$\varphi_n(t) = b_n^{-1} \log E\{\exp(tb_n Z_n)\}.$$

Assume that there exists a differentiable function φ in $]-\alpha,\alpha[$, $\alpha > 0$, such that $\lim_{n\to\infty} \varphi_n(t) = \varphi(t)$, for all $t \in]-\alpha,\alpha[$. Let a be a real such that $|a - \varphi'(0)| > 0$ and there exists $\tau_a \in \{t \in R : 0 < |t| < \alpha\}$, such that $\varphi'(\tau_a) = a$. The parameter τ_a is used to make an exponential change of measure which allows to sharpen the large deviation result (see the proofs in Section 4).
This paper deals with strong large deviation results for Z_n, by obtaining an asymptotic equivalent for the tail probability $P(Z_n \geq a)$, where $a > \varphi'(0)$ (the real a does not necessarily depend on n). We distinguish the cases where Z_n is absolutely continuous (or its distribution has an absolutely continuous component) and Z_n is lattice-valued. To establish the strong large deviation results, we need to assume several assumptions, in particular on the (normalized) c.g.f. φ_n and on the m.g.f ϕ_n :

$(A.1)$ The c.g.f φ_n is an analytic function in $D_C := \{z \in C : |z| < \alpha\}$, and there exists $M > 0$ such that $|\varphi_n(z)| < M$ for all $z \in D_C$ and n large enough.

$(A.2)$ There exist $\alpha_0 \in]0, \alpha - \tau_a[$ and a function H such that for each $t \in]\tau_a - \alpha_0, \tau_a + \alpha_0[$ and for n large enough,

$$\varphi_n(t) = \varphi(t) + b_n^{-1} H(t) + o\left(b_n^{-1}\right), \tag{1}$$

where the function φ is three times continuously differentiable in $]\tau_a - \alpha_0, \tau_a + \alpha_0[$, $\varphi''(\tau_a) > 0$, and H is continuously differentiable in $]\tau_a - \alpha_0, \tau_a + \alpha_0[$.

$(A.3)$ There exists $\delta_0 > 0$ such that

$$\sup_{\delta<|t|\leq\lambda|\tau_a|} \left| \frac{\phi_n(\tau_a + it)}{\phi_n(\tau_a)} \right| = o\left(\frac{1}{\sqrt{b_n}}\right),$$

for any given δ and λ such that $0 < \delta < \delta_0 < \lambda$.

Assumption $(A.1)$ is needed in the proof of Lemma 3.2, where we make use of Cauchy's inequality to bound the remaining term of a Taylor expansion. Assumption $(A.2)$ guarantees the existence of an asymptotic expression for the (normalized) cumulant generating function. This assumption is necessary to establish the strong large deviation results with rate functions that do not depend on n. It is also used to prove Lemma 3.1 and Lemma 3.2. Assumption $(A.3)$ is a version of Condition 3.16 of [4]. It implies a necessary condition on the characteristic function of the random variable V_n (defined in (6)) and is required to apply Theorem 2.3 in [4] (see the proof of Theorem 2.1 in Section 4). It plays a similar role to that of the Cramer's condition.

2.2 Theorems

In what follows, we give the main results. The first theorem deals with the case of absolutely continuous variables.

Theorem 2.1 *Assume that Z_n is absolutely continuous (or its distribution has an absolutely continuous component). Let a be a real such that $a > \varphi'(0)$ and let assumptions $(A.1) - (A.3)$ hold. Then, for n large enough,*

$$P(Z_n \geq a) = \frac{\exp(-b_n\varphi^*(a) + H(\tau_a))}{\sigma_a \tau_a (2\pi b_n)^{1/2}} [1 + o(1)] \tag{2}$$

where $\tau_a > 0$ is such that $\varphi'(\tau_a) = a$. Further, $\varphi^*(a) = \tau_a a - \varphi(\tau_a)$ and $\sigma_a^2 = \varphi''(\tau_a)$.

Now let us consider the case where b_nZ_n is lattice. Recall that a random variable Y is said to be lattice if it takes values in a subset of the lattice set $\{d_0 + ks, k \in Z\}$. The real d_0 is called the displacement and the positive real s is the span of Y. Denote by d_n and s_n the displacement and the span of the statistic b_nZ_n, respectively. The following assumption is required and replaces (A.3) (see [4]) :

$(A'.3)$ There exists $\delta_1 > 0$ such that for $0 < \delta < \delta_1$

$$\sup_{\delta \le |t| \le \pi/s_n} \left| \frac{\phi_n(\tau_a + it)}{\phi_n(\tau_a)} \right| = o\left(\frac{1}{\sqrt{b_n}}\right).$$

The next theorem assumes that the span s_n goes to zero as $n \to \infty$. As noted in [4, Remark 3.4], in this case, Assumption $(A'.3)$ implies Assumption $(A.3)$. Thus we obtain the same results as those of Theorem 2.1.

Theorem 2.2 *Let assumptions $(A.1)-(A.2)$ and $(A'.3)$ hold. Assume that the span s_n of the lattice-valued random variable b_nZ_n goes to zero as n tends to infinity. Then, for $a > \varphi'(0)$ and for n large enough,*

$$P(Z_n \ge a) = \frac{\exp(-b_n\varphi^*(a) + H(\tau_a))}{\sigma_a\tau_a(2\pi b_n)^{1/2}}[1 + o(1)]$$

where $\tau_a > 0$ is such that $\varphi'(\tau_a) = a$. Further, $\varphi^*(a) = \tau_a a - \varphi(\tau_a)$ and $\sigma_a^2 = \varphi''(\tau_a)$.

2.3 Examples

We present three examples to illustrate the theorems of the preceding section.

Example 1. The sample variance. We consider the sample variance. This statistic has the following well known expression :

$$Z_n = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

Assuming that the X_i's have a normal distribution $\mathcal{N}(\mu; \sigma^2)$, $\sigma^2 > 0$, we know that $\sigma^{-2}\sum_{i=1}^{n}(X_i - \overline{X})^2$ follows a chi-square distribution with $n-1$ degrees of freedom. A probability of large deviation for the sample variance was studied by Sievers [9]. Here, we give a strong large deviation result by applying Theorem 2.1 (with $b_n = n$).

Corollary 2.1 *Let Z_n be defined as above. Then for a real a such that $a > \sigma^2$ and n large enough,*

$$P(Z_n \ge a) = \frac{\exp(-(n+1)\varphi_{SV}^*(a))}{\sigma(4a\pi n)^{1/2}}[1 + o(1)], \tag{3}$$

where $\varphi_{SV}^*(a) = \frac{1}{2}\left(\frac{a}{\sigma^2} - \log\left(\frac{a}{\sigma^2}\right) - 1\right) > 0$.

Example 2. The Wilcoxon signed-rank statistic. A large deviations result for the Wilcoxon signed-rank statistic was obtained by Klotz [8]. The asymptotic expression (4) will follow from Theorem 2.2 (with $b_n = n$).

Let $\{X_1, \ldots, X_n\}$ be a sequence of i.i.d. continuous random variables having distribution function F and let R_i be the rank of $|X_i|$, $i = 1, \ldots, n$. In other words, if one arranges $|X_1|, |X_2|, \ldots, |X_n|$ in increasing order of magnitude, R_i denotes the rank of $|X_i|$. Assume that the random variables X_i are symmetric about their median m. The Wilcoxon signed-rank statistic W_n is defined as the sum of the quantities R_i corresponding to the positive $X_i's$, that is,

$$W_n = \sum_{i=1}^{n} I_{\{X_i>0\}} R_i.$$

The statistic W_n is used to test the null hypothesis $H_0 : m = 0$, and can also be written as :

$$W_n = \sum_{1\le i\le j\le n} I_{\{\frac{1}{2}(X_i+X_j)>0\}}.$$

Letting $Z_n = \frac{W_n}{n^2}$, we have the following result.

Corollary 2.2 *Let Z_n be defined as above. Then for a real $a > 1/4$ and n large enough,*

$$(4) \qquad P(Z_n \ge a) = \frac{\exp(-n\varphi_W^*(a) + H_W(\tau_a))}{\sigma_a(2\pi n)^{1/2}}[1 + o(1)]$$

where $\tau_a > 0$, is such that $\int_0^1 \frac{x}{1+\exp(-\tau_a x)}dx = a$, $H_W(t) = \frac{1}{2}\log\left(\frac{\exp(t)+1}{2}\right)$ and $\sigma_a^2 = \int_0^1 \frac{x^2\exp(\tau_a x)}{(1+\exp(\tau_a x))^2}dx$. Further, $\varphi_W^*(a) = \tau_a a - \varphi_W(\tau_a)$ where

$$\varphi_W(t) = \int_0^1 \log\left(\frac{e^{tx}+1}{2}\right)dx.$$

Example 3. The Kendall's tau statistic. Sievers [9] gave a large deviation result for this non-parametric test of independence. An application of Theorem 2.2 (with $b_n = n$) will yield the strong large deviation result (5). Let $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ be a sequence of i.i.d. continuous random couples having distribution function $F(x, y)$ and let F_X and F_Y be the marginal distributions. The Kendall's tau statistic Z_n can be defined as follows :

$$Z_n = 2\sum_{1\le i<j\le n} \frac{(I_{\{X_i\ge X_j\}} - I_{\{X_i\le X_j\}})(I_{\{Y_i\ge Y_j\}} - I_{\{Y_i\le Y_j\}})}{n(n-1)}.$$

It was first used by Kendall to test the null hypothesis $H_0 : F(x, y) = F_X(x)F_Y(y)$, for all x, y. We have the following corollary.

Corollary 2.3 *Let Z_n be defined as above. Then for a real $a \in]0, 1[$ and n large enough,*

$$(5) \qquad P(Z_n \ge a) = \frac{\exp(-n\varphi_K^*(a) + H_K(\tau_a))}{\sigma_a(2\pi n)^{1/2}}[1 + o(1)]$$

where $\tau_a > 0$ is such that $1 - \frac{1}{\tau_a} + 4\int_0^1 \frac{x}{\exp(4\tau_a x)-1}dx = a$,

$$H_K(t) = 2t - 1 + \frac{3}{2}\log(1 - e^{-4t}) - \frac{1}{2}\log(4t) - \int_0^1 \log(1 - e^{-4tx})dx$$

and

$$\sigma_a^2 = \frac{1}{\tau_a^2} - 16\int_0^1 \frac{x^2\exp(4\tau_a x)}{(\exp(4\tau_a x) - 1)^2}dx.$$

Further, $\varphi_K^*(a) = \tau_a a - \varphi_K(\tau_a)$, where

$$\varphi_K(t) = t + 1 - \log(4t) + \int_0^1 \log(1 - e^{-4tx})dx.$$

3 Lemmas

In this section, we present two lemmas needed for the proofs of Theorems 2.1 and 2.2 (the proofs of these lemmas will be omitted). To do so we first introduce some notation. Denote the distribution function of $b_n Z_n$ by K_n. Let a be a real such that $a > \varphi'(0)$ and there exists $\tau_a \in]0, \alpha[$ satisfying $\varphi'(\tau_a) = a$. Using an exponential change of measure, let

$$H_n(u) = \int_{-\infty<y<u} \exp(y\tau_a - b_n\varphi_n(\tau_a))dK_n(y)$$

be the distribution function of $b_n Z_n^*$. Define the random variable V_n as follows :

$$V_n = \frac{\sqrt{b_n}(Z_n^* - a)}{\sigma_a}, \tag{6}$$

where we recall that $\sigma_a^2 = \varphi''(\tau_a) > 0$ (Assumption $(A.2)$). The following lemma shows the asymptotic normality of V_n.

Lemma 3.1 *Let Assumption $(A.2)$ hold. Then, the statistic V_n converges in distribution to a standard normal random variable.*

The proof of the next lemma is similar to that of [4, Lemma 3.1] (in particular, we make use of [4, Theorem 2.6]).

Lemma 3.2 *Let f_n be the characteristic function of V_n and assume that assumptions $(A.1) - (A.2)$ are satisfied. Then, there exist $\delta > 0$, $\gamma > 0$ and $n_0 \in N^*$ such that,*

$$\sup_{n\geq n_0} |f_n(t)|I(|t| \leq \delta\sqrt{b_n}\sigma_a) \leq \exp(-\gamma t^2). \tag{7}$$

4 Proofs

We give the proofs of the theorem of Section 2.

Proof of Theorem 2.1. The beginning of the proof of (2) follows the same lines as in [1]. Let $a > \varphi'(0)$. We can write the Fenchel-Legendre transform of φ as follows :

$$\varphi^*(a) := \sup_{t\in R}\{ta - \varphi(t)\} = \tau_a a - \varphi(\tau_a),$$

where $\tau_a \in]0, \alpha[$ is such that $\varphi'(\tau_a) = a$. Recall that nU_n^* is a random variable with distribution function

$$H_n(u) = \int_{-\infty<y<u} \exp(y\tau_a - b_n\varphi_n(\tau_a))dK_n(y).$$

Using Assumption $(A.2)$, the right tail probability may now be written as follows :

$$\begin{aligned}
P(Z_n \geq a) &= E\{\exp(-\tau_a b_n Z_n^* + b_n\varphi_n(\tau_a))I_{\{Z_n^*\geq a\}}\} \\
&= \exp(b_n\varphi_n(\tau_a) - b_n\tau_a a)E\{\exp(-\tau_a b_n(Z_n^* - a))I_{\{Z_n^*\geq a\}}\} \\
&= \exp(b_n\varphi_n(\tau_a) - b_n\tau_a a)E\{\exp(-\tau_a\sigma_a\sqrt{b_n}V_n)I_{\{V_n\geq 0\}}\} \\
&= e^{b_n(\varphi(\tau_a)-\tau_a a)+H(\tau_a)+o(1)}E\{\exp(-\tau_a\sigma_a\sqrt{b_n}V_n)I_{\{V_n\geq 0\}}\} \\
&= e^{-b_n\varphi^*(a)+H(\tau_a)}E\{\exp(-\tau_a\sigma_a\sqrt{b_n}V_n)I_{\{V_n\geq 0\}}\}(1+o(1))
\end{aligned}$$

where

$$V_n = \frac{\sqrt{b_n}(Z_n^* - a)}{\sigma_a} \quad \text{and} \quad \sigma_a = \sqrt{\varphi''(\tau_a)} > 0.$$

It remains to prove that

$$\lim_{n\to\infty} \tau_a\sigma_a\sqrt{b_n}E\{\exp(-\tau_a\sigma_a\sqrt{b_n}V_n)I_{\{V_n\geq 0\}}\} = \frac{1}{\sqrt{2\pi}}. \tag{8}$$

To do this, we will apply [4, Theorem 2.7] to the sequence of random variables $\{V_n, n \geq 1\}$. Lemma 3.1 and Lemma 3.2 show that V_n converges in distribution to a standard normal variable and that (7) holds, respectively. Besides, it is easy to see that

$$\sup_{\delta\sqrt{b_n}\sigma_a<|t|\leq\lambda\tau_a\sigma_a\sqrt{b_n}} |f_n(t)| = \sup_{\delta<|t|\leq\lambda\tau_a} \left|\frac{\phi_n(\tau+it)}{\phi_n(\tau)}\right|,$$

where f_n is the characteristic function of V_n and ϕ_n the m.g.f. of b_nZ_n. Hence, by Assumption $(A.3)$, we have for n large enough,

$$\sup_{\delta\sqrt{b_n}\sigma_a<t\leq\lambda\tau_a\sigma_a\sqrt{b_n}} |f_n(t)| = o(b_n^{-1/2}). \tag{9}$$

The convergence in distribution of Z_n, (7) and (9) allow us to verify the conditions of [4, Theorem 2.3]. Denote the density of V_n (or pseudo density if V_n does not possess a proper density function) by q_n. By [4, Theorem 2.3], there exists a constant $M_0 > 0$ such that

$$\sup_y q_n(y) \leq M_0, \tag{10}$$

and if $z_n \to z$, then

$$q_n(z_n) \to (\sqrt{2\pi})^{-1}e^{-z^2/2}. \tag{11}$$

[4, Theorem 2.7] follows directly from (10) and (11). Finally, [4, Theorem 2.7] implies (8), which ends the proof.

Proof of Theorem 2.2. Theorem 2.2 follows from Theorem 2.1, since Assumption $(A'.3)$ implies Assumption $(A.3)$ (in view of the fact that $s_n \to 0$ as $n \to \infty$).

References

[1] Bahadur, R., Rao, R. (1960). On deviations of the sample mean. *Annals of Mathematical Statistics*, **31**, 1015–1027.

[2] Blackwell, D., Hodges, J.L. (1959). The probability in the extreme tail of a convolution. *Annals of Mathematical Statistics*, **30**, 1113-1120.

[3] Book, S. (1972). Large deviation probabilities for weighted sums. *Annals of Mathematical Statistics*, **43**, 1221-1234

[4] Chaganty, N.R., Sethuraman, J. (1993). Strong large deviation and local limit theorems. *Annals of Probability*, **21**, 1671–1690.

[5] Chaganty, N.R., Sethuraman, J. (1996). Multidimensional strong large deviation theorems. *Journal of Statistical Planning and Inference*, **55**(3), 265–280.

[6] Dembo, A., Zeitouni, O. (1998). *Large deviations techniques and applications.* Springer, New York. Second edition.

[7] Kendall, M.G., Stuart, A. (1979). *The Advanced Theory of Statistics*, Vol. II, MacMillan, New-York.

[8] Klotz, J. (1965). Alternative efficiencies for signed rank tests. *Annals of Mathematical Statistics*, **36**, 1759–1766.

[9] Sievers, G.L. (1969). On the probability of large deviations. *Annals of Mathematical Statistics*, **40**(3), 1908–1921.