Z. Wahrscheinlichkeitstheorie verw. Gebiete
59, 101–115 (1982)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

A Laplace Approximation for Sums of Independent Random Variables

Anders Martin-Löf
Dept. of Math. Statistics, Univ. of Stockholm, Box 6701, S-11385 Sweden

Introduction

In the well known Laplace approximation one considers integrals of the form $\int \varphi(x)\, e^{nf(x)}\, dx$, where φ and f are sufficiently nice real valued functions of $x \in R^1$, and n is a large positive number. It is based on the following argument: Suppose that $f(x)$ attains its maximum at $x=0$ e.g., and that $f(x)-f(0) \sim -A|x|^a$ for some $A, a>0$ when $|x|$ is small. Then $e^{nf(x)}$ will have a very sharp maximum at $x=0$, so the main contribution to the integral comes from a neighbourhood of 0, and we get:

$$\int \varphi(x)\, e^{nf(x)}\, dx \sim \int_{|x| \leqq n^{-1/2a}} \varphi(x)\, e^{nf(x)}\, dx \sim \varphi(0) \int_{|x| \leqq n^{-1/2a}} e^{nf(0)-nA|x|^a}\, dx$$
$$= \varphi(0)\, e^{nf(0)} (nA)^{-1/a} \int_{|y| \leqq n^{1/2a} A^{1/a}} e^{-|y|^a}\, dy$$
$$\sim \varphi(0)\, e^{nf(0)} (nA)^{-1/a} \int e^{-|y|^a}\, dy = \varphi(0)\, e^{nf(0)} (nA)^{-1/a}\, 2\Gamma(1+1/a).$$

The meaning of $\sim$ is that the ratio of the two expressions goes to 1 as n goes to ∞.

A complete proof is straightforward and is given e.g. in [4].

In some recent work on the asymptotics of an epidemic process [3] we met with the problem of deriving an analogous approximation for an expression of the form

$$E_n = E(e^{nf(S_n/n)}) = \int e^{nf(x/n)}\, F^{n*}(dx), \tag{1}$$

where $\{S_n\}$ are the partial sums of i.i.d random variables $\{X_i\}$ with a common distribution $F(x)$, $f(x)$ is a sufficiently nice function of $x \in R^1$, and $n \to \infty$. In order to understand how the analogous approximation of E_n is formed let us first consider the situation when X_1 has a density. Then under appropriate conditions the central limit theorem for large deviations tells us that $\varphi_n(x)$, the density of S_n/n, is approximatively given by the following formula:

$$\varphi_n(x) \sim (n/2\pi)^{1/2} (-h''(x))^{1/2}\, e^{nh(x)}. \tag{2}$$

0044-3719/82/0059/0101/$03.00

Here $h(x)$ is the entropy function of the distribution $F(x)$, which is defined below.

A proof of the above formula is given e.g. in [9]. Using it in the expression for E_n we can then make the Laplace approximation of the resulting integral as follows:

Suppose that $f(x)+h(x)$ has its maximum at $x=0$ e.g. and that

$$f(x)+h(x)-f(0)-h(0)\sim -A\,|x|^a \tag{3}$$

with $A, a>0$ as $|x|\to 0$, then we get

$$\begin{aligned} E_n &\sim (n/2\pi)^{1/2} \int (-h''(x))^{1/2}\, e^{n(f(x)+h(x))}\, dx \\ &\sim (2\pi)^{-1/2}\, 2\Gamma(1+1/a)\, A^{-1/a} (-h''(0))^{1/2}\, n^{1/2-1/a}\, e^{n(f(0)+h(0))}. \end{aligned} \tag{4}$$

When $f(x)$ is twice continuously differentiable near 0 then $2A=-f''(0)-h''(0)$, and $a=2$ if $A>0$, so in this case the formula reduces to

$$E_n \sim (1+f''(0)/h''(0))^{-1/2}\, e^{n(f(0)+h(0))}. \tag{5}$$

In the following we give a complete proof of the above asymptotic formulas with rather weak assumptions about $F(x)$ and $f(x)$. (Theorem 2 and 3 below.)

Concerning $f(x)$ it is enough to assume that it is Hölder continuous near the maximum at $x=0$ and that it can be approximated by piecewise linear functions as explained in (26).

Concerning $F(x)$ we have to assume that its generating function

$$e^{g(t)} = \int e^{tx}\, F(dx) = E(e^{tX_1}) \tag{6}$$

is defined in some open interval $D_g \subset R^1$, so that laws of large deviations can be established using saddlepoint methods. Then the entropy function $h(x)$, defined by $h(x)=\inf_t (g(t)-tx)$, is finite and very regular in some open interval D_h. It is assumed that $f(x)+h(x)$ attains its maximum at a unique point $\bar{x}$ in D_h, which we can take to be $\bar{x}=0$.

It turns out that it is not necessary to assume that $F(x)$ has a density. Either F is arithmetic, i.e. X_1 takes only values $\xi+m\cdot\delta$, where m is integer and $\delta>0$, and $\operatorname{Re} g(it)$ is periodic with period $2\pi/\delta$ or if it is not, then it is enough to assume that the so called Cramér condition

$$\sup_{|t|\geqq u} \operatorname{Re} g(it) < 0 \qquad \text{for any} \quad u>0 \tag{7}$$

is fulfilled.

Under these assumptions we can establish the formula (4) for any $A, a>0$. (In the arithmetic case the formula has to be modified when $0<a\leqq 1$.) (Theorem 2 below.)

When $a=2$, $A>0$ it turns out that (5) is true even without assuming the Cramér condition (7), and the arithmetic case need not be considered separately. (Theorem 3 below.)

We have treated explicitly only the case when $f(x)+h(x)$ attains its maximal value at only one point. If it attains it at a finite number of points (4) is of course replaced by a sum of similar terms, one for each maximum, and only those with the largest value of a will contribute as $n \to \infty$.

The formula (4) is an example of the use of "thermodynamic" probability estimates familiar in statistical mechanics. I.e. one considers a "macroscopic variable" S_n/n whose probability distribution is asymptotically determined by an entropy function through a formula like (2), and in the resulting formula (4) all information about the underlying distribution F is contained in h and its derivatives at the maximal point $\bar{x}$. The formula (4) ought to be useful in other statistical and physical applications as well.

(In [4] similar formulas are derived in a special situation when $f(x) = \alpha x^2$.)

The fact that S_n/n behaves as if it had a density $\varphi_n(x)$ proportional to $e^{nh(x)}$ so that $\varphi_n(x)\, e^{nf(x)}$ is proportional to $e^{-nA|x-\bar{x}|^a}$ near $\bar{x}$ can be expressed more precisely in the form of a central limit theorem as follows: Let a modified probability distribution for S_n be defined by $F_{n,f}(dx) = e^{nf(x/n)} F^{n*}(dx)/E_n$. Then for this distribution the following central limit theorem holds: (Theorem 4 below). The distribution of $n^{1/a}(S_n/n - \bar{x})$ defined by $F_{n,f}$ converges weakly to the one defined by the density

$$e^{-A|x|^a} A^{1/a}\, dx/2\Gamma(1+1/a) \qquad \text{as } n \to \infty \qquad \text{if } a > 1.$$

In particular when $a = 2$, $A > 0$ this limit distribution is Gaussian with variance $1/2A = -1/(f''(0) + h''(0))$.

The problem of deriving logarithmic estimates for the distribution of S_n/n saying that $\lim_{n\to\infty} n^{-1} \log P(S_n/n \in I) = \sup_{x \in I} h(x)$ has recently received much attention following the work of Lanford [7], where this formula is established for intervals I using an elegant subadditivity argument. A survey of such results and their statistical background is given in the recent article [2] by Bahadur and Zabell. In a much more general context Varadhan [11] has shown how to derive the corresponding logarithmic estimate

$$\lim_{n\to\infty} n^{-1} \log E_n = \sup_x (f(x) + h(x))$$

using the above logarithmic estimates for the distribution of S_n/n and arguments valid for distributions on more general metric spaces than R^1.

Our more accurate formula for E_n, valid in the sense of $\sim$, gives so to speak the accuracy of the central limit theorem, and it depends on the local behaviour of $f(x) + h(x)$ near the maximal point $\bar{x}$. Analogous formulas for $P(S_n/n \in I)$ have recently been derived by Höglund [6].

1. Preparations

Let us first recall the basic properties of $g(t)$ defined by (6) and its Legendre transform or entropy function defined by

$$h(x) = \inf_t (g(t) - tx), \tag{8}$$

which we will need to know in the following (c.f. [8, 10]).

Let D_g be the maximal open interval in which $g(t) < \infty$, and let $g(t) = +\infty$ outside of $\bar{D}_g$. D_g is assumed to be non empty. Then $g(u+iv)$ is analytic in the vertical strip defined by $u \in D_g$ and $g(t)$ is continuous when t approaches the endpoints of D_g. In particular $g(t)$ is infinitely differentiable when $t \in D_g$. Since $g''(t) > 0$ $g(t)$ is strictly convex, and the mapping $t \to x$ defined by $x = g'(t)$ is strictly increasing and infinitely differentiable. Hence it is $1-1$ from D_g onto an open interval D_h, and the inverse mapping is also strictly increasing and infinitely differentiable. When $x \in D_h$ and $x = g'(t)$ then $\inf_t$ in (8) is attained at t, so

$$h(x) = g(t) - tx, \qquad \text{when } x = g'(t). \tag{9}$$

From (8) follows that $h(x)$ is concave and closed (i.e. upper semicontinuous) and that $g(t)$ is determined from $h(x)$ by

$$g(t) = \sup_x (h(x) + tx). \tag{10}$$

From (9) it follows that $h(x)$ is infinitely differentiable and strictly concave in D_h because we have

$$h'(x) = -t, \tag{11}$$

so the inverse mapping of $x = g'(t)$ is given by $t = -h'(x)$. Therefore

$$\frac{dx}{dt} = g''(t) = -1/h''(x), \tag{12}$$

which shows that $h''(x) < 0$, and $h(x)$ is strictly concave in D_h. Outside of D_h it is linear or $-\infty$.

We first derive a useful upper bound for E_n defined in (1) in the special case when $f(x)$ is linear in an interval and $-\infty$ outside it:

Lemma 1. *For any interval I*

$$B_n = E(e^{tS_n}, S_n/n \in I) \leqq \exp n \sup_{x \in I} (h(x) + tx). \tag{13}$$

Proof. Take first I to be open. If $\sup_{x \in I} (h(x) + tx) = \sup_x (h(x) + tx) = g(t)$ then $B_n \leqq E(e^{tS_n}) = e^{ng(t)}$, and (13) is true.

If $-\infty < \sup_{x \in I} < \sup_x$ then e.g. the left endpoint l of I is finite, $h(x)$ is finite to the right of l, $\sup_{x \in I} = h(l) + tl$, and the right derivative $h'_+(l) + t \leqq 0$. (A convex function is continuous and has one sided derivatives from the side where it is finite if it is closed as h is.) $t_+ = -h'_+(l) > -\infty$ defines a supporting line to $h(x)$ at l, i.e. $h(x) \leqq h(l) - t_+(x - l)$ for all x. This means that $g(t_+) = \sup_x (h(x) + t_+ x) = h(l) + t_+ l$, and since $t(x-l) \leqq t_+(x-l)$ for $x \in I$ we have

$$B_n \leqq E(e^{nt_+(n^{-1}S_n - l) + ntl}) = e^{ng(t_+) - nt_+ l + ntl} = e^{n(h(l) + tl)},$$

and (13) is valid.

Finally, if $\sup_{x\in I}(h(x)+tx)=-\infty$ then $h(x)=-\infty$ for all $x\in I$, i.e. $\inf_t(g(t)-tx)=-\infty$ for all $x\in I$. This means that $I\cap D_h=\emptyset$, so I lies e.g. to the right of D_h. I.e. $g'(t)<x$ for any $t\in D_g$ and any $x\in I$, so $g(t)-tx$ is decreasing and

$$\inf_t(g(t)-tx)=\lim_{t\to+\infty}(g(t)-tx)=-\infty \qquad \text{for any } x\in I.$$

Hence $\lim_{t\to+\infty} E(e^{t(S_n-nx)})=0$ for any $x\in I$, so that $P(S_n/n\geqq x)=0$. If we now let x decrease to l we can conclude that $P(n^{-1}S_n>l)=0$. Hence $B_n=0$ and (13) is again valid.

If I is not open and e.g. $I=[l,r)$ then we consider the slightly larger interval $I'=(l-\varepsilon,r)$ and let $\varepsilon\to 0$. Because h is closed we have that $\lim_{\varepsilon\to 0}\sup_{|x-l|\leqq\varepsilon} h(x)\leqq h(l)$. This means that

$$\lim_{\varepsilon\to 0}\sup_{x\in I'}(h(x)+tx)=\sup_{x\in I}(h(x)+tx),$$

and (13) follows by letting $\varepsilon\to 0$ in (13'):

$$B_n\leqq B'_n\leqq \exp n\sup_{x\in I'}(h(x)+tx).$$

We next proceed to prove a local limit theorem for large deviations saying that if the Cramér condition (7) is fulfilled, then (2) is true for suitably defined smoothed density, which is defined as follows: Let $k(x)$ be the usual Fejer kernel $k(x)=\max(1-|x|,0)$ with Fourier transform $\hat{k}(t)=2(1-\cos t)/t^2$, and let for $c>0$ small $k_c(x)=c^{-1}k(x/c)$ be the corresponding approximate delta function with $\hat{k}_c(t)=\hat{k}(ct)$. Since $k(x)$ has compact support $\hat{k}(t)$ is an entire function. Define the smoothed density of S_n as follows:

$$\varphi_{n,c}(x)=E(k_c(S_n-nx))=\int k_c(y-nx)\,F^{n*}(dy). \tag{14}$$

Then the following asymptotic formula holds when $n\to\infty$, $c\to 0$:

Theorem 1. *If the Cramér condition* (7) *is fulfilled, then*

$$\varphi_{n,c}(x)=(2\pi n)^{-1/2}(-h''(x))^{1/2}e^{nh(x)}(1+\mathcal{O}(n^{-1/2}+c+e^{-n\gamma}/c)) \tag{15}$$

uniformly when x varies on a compact subset of D_h for some $\gamma>0$. In the arithmetic case the same asymptotic formula is valid for $P(S_n=nx)/\delta$ uniformly when $x=\xi+m\delta/n$ varies on compact subsets of D_h and m is an integer. (Excluding of course the error terms involving c.)

Proof. We use Parsevals relation and the usual change of the path of integration used in the saddlepoint method to express $\varphi_{n,c}(x)$ as follows:

$$\varphi_{n,c}(x)=\int e^{n(g(iu)-iux)}\hat{k}_c(u)\,du/2\pi.$$

If $x=g'(t)$ with $t\in D_g$ change the path of integration so that $iu\to t+iu$. (Here we use the fact that $g(t+iu)$ is analytic when $t\in D_g$ and that $\hat{k}_c(u)$ is entire.)

Then $g(t)-tx=h(x)$ by (9) so we get:

$$\begin{aligned}\varphi_{n,c}(x)&=\int e^{n(g(t+iu)-(t+iu)x)}\hat{k}_c(u-it)\,du/2\pi\\&=e^{nh(x)}\int e^{n(g(t+iu)-g(t)-iug'(t))}\hat{k}_c(u-it)\,du/2\pi.\end{aligned}\tag{16}$$

We now split the integral into three parts $I_1+I_2+I_3$ corresponding to the regions $|u|\leqq d_1/n^{1/2}$, $d_1/n^{1/2}<|u|\leqq d_2$, $d_2<|u|$ respectively, where d_1 is large, $d_1/n^{1/2}$ is small and $d_2>0$ as $n\to\infty$. Each part is estimated separately:

I_1: When $|u|$ is small

$$g(t+iu)-g(t)-iug'(t)=-u^2v/2+\mathcal{O}|u|^3,\tag{17}$$

where $v=g''(t)=-1/h''(x)>0$.

This holds uniformly when $t\in K_g=\text{compact}\subset D_g$ and $|u|\leqq d_1/n^{1/2}$. Replacing u by $u/n^{1/2}$ in I_1 we get

$$I_1=\int_{|u|\leqq d_1} e^{-vu^2/2}(1+\mathcal{O}(|u|^3/n^{1/2}))\,\hat{k}_c(u/n^{1/2}-it)\,du/n^{1/2}\,2\pi.$$

Let us put

$$\begin{aligned}I_0&=\int e^{-vu^2/2}\hat{k}_c(u/n^{1/2}-it)\,du/n^{1/2}\,2\pi\\&=\int e^{-nvu^2/2}\hat{k}_c(u-it)\,du/2\pi\\&=\int(2\pi nv)^{-1/2}e^{-x^2/2nv}e^{tx}k_c(x)\,dx.\end{aligned}$$

The last equation follows from Parsevals realtion. We see that $I_0=(2\pi nv)^{-1/2}(1+\mathcal{O}(c))$ as $c\to0$ uniformly when $n\to\infty$ and $t\in K_g$. It is easy to check that

$$|\hat{k}(u-it)|\leqq K/(1+u^2)\tag{18}$$

for some $K>0$ uniformly when $t\in K_g$.

Hence

$$\begin{aligned}|I_1-I_0|&=\mathcal{O}\Big(\int_{|u|\leqq d_1}|u|^3e^{-vu^2/2}\,du/n+\int_{|u|>d_1}e^{-vu^2/2}\,du/n^{1/2}\Big)\\&=\mathcal{O}((nv)^{-1/2}(n^{-1/2}+e^{-vd_1^2/2}),\end{aligned}$$

so we have

$$I_1=(2\pi nv)^{-1/2}(1+\mathcal{O}(c+n^{-1/2}+e^{-vd_1^2/2})).\tag{19}$$

I_2: From (17) we see that in I_2 we have $g(t+iu)-g(t)-iug'(t)\geqq-u^2v/4$ when $|u|\leqq d_2$ is small enough. Hence using (18) we have

$$I_2=\mathcal{O}\Big(\int_{|u|\geqq d_1}e^{-vu^2/4}\,du/n^{1/2}\Big)=\mathcal{O}((nv)^{-1/2}e^{-vd_1^2/4}).\tag{20}$$

I_3: To estimate I_3 we need the Cramér condition (7) and (18). If we can show that $\operatorname{Re} g(t+iu)-g(t)\leqq\gamma<0$ when $|u|\geqq d_2$ and $t\in K_g$ then we have

$$I_3=\mathcal{O}\Big(e^{-n\gamma}\int du/(1+(cu)^2)\Big)=\mathcal{O}(e^{-n\gamma}/c).\tag{21}$$

To see that the above bound for Re g holds we use the fact that

$$e^{g(t+iu)-g(t)}=\int e^{iux}\,e^{tx-g(t)}\,F(dx)=\int e^{iux}\,F_t(dx) \tag{22}$$

where $F_t(dx)$ is absolutely continuous with respect to $F(dx)$. In [1], Lemma 4 it is shown that the Cramér condition is preserved when the distribution F is changed to one which is absolutely continuous with respect to F. Hence $\sup_{|u|\geqq d_2} \operatorname{Re} g(t+iu)-g(t)<0$ for each t. But it is easy to check that $g(t+iu)-g(t)$ is continuous in t uniformly in u, so that $\sup_{t\in K_g}\sup_{|u|\geqq d_2} \operatorname{Re} g(t+iu)-g(t)\leqq \gamma<0$ also. If we now put $v^{1/2}d_1=n^{1/4}$ e.g. and add the expressions in (19), (20) and (21) we get the desired formula (15) because $v=-1/h''(x)>0$. The proof in the arithmetic case follows the same lines starting from the inversion formula for the individual atoms of the distribution of S_n:

$$\begin{aligned}P(S_n=nx)&=\delta\int_{|u|\leqq\pi/\delta} e^{n(g(iu)-ixu)}\,du/2\pi\\ &=\delta e^{nh(x)}\int_{|u|\leqq\pi/\delta} e^{n(g(t+iu)-g(t)-iug'(t))}\,du/2\pi\end{aligned} \tag{23}$$

where $x=\xi+m\delta/n$, m integer and $g'(t)=x$. In this case we take $d_2=\pi/\delta$ and the term I_3 is absent.

2. First Proof of the Laplace Approximation

We now show that Theorem 1 can be used to establish the asymptotic formula (4) in the way indicated in the introduction. We assume that $f(x)+h(x)$ attains its maximum at only one point, $x=0$ e.g., and that $0\in D_h$. Near this point we assume (3) and that $f(x)$ is Hölder continuous:

$$|f(x+y)-f(x)|\leqq A'|y|^{a'} \quad \text{for some } A', a'>0. \tag{24}$$

Consider now E_n defined by (1) and split it into the contribution from a neighbourhood of 0 and the rest:

$$\begin{aligned}E_{n,1}&=E(e^{nf(S_n/n)}I_{2d}(S_n/n))\\ E_{n,2}&=E_n-E_{n,1},\end{aligned} \tag{25}$$

where $I_d(x)$ is the indicator of the interval $[-d,d]$. Let us estimate $E_{n,2}$ first with the help of Lemma 1. To this end we make the following regularity assumption:

For any $d>0$ there is a piecewise linear function $f_d(x)$ with at most d^{-M} pieces for some $M>0$ such that

$$e^{f_d(x)-d}\leqq e^{f(x)}\leqq e^{f_d(x)+d} \quad \text{when } h(x)>-\infty. \tag{26}$$

$f(x)$ and $f_d(x)$ are allowed to take the value $-\infty$. (This is true e.g. if $f(x)$ is piecewise continuous and asymptotically linear at infinity.)

Take now $b>0$ such that $f(x)+h(x)\leqq f(0)+h(0)-3b$ when $|x|>d$ (this means that $b=\mathcal{O}(d^a)$ when d is small), and take $f_b(x)$ as in (26). In each interval J where f_b is linear we can use Lemma 1 to get the following bound:

$$\begin{aligned} E(e^{nf(S_n/n)}, S_n/n\in J) &\leqq E(e^{nf_b(S_n/n)+nb}, S_n/n\in J)\\ &\leqq \exp n \sup_{x\in J}(f_b(x)+h(x)+b)\\ &\leqq \exp n \sup_{x\in J}(f(x)+h(x)+2b)\\ &\leqq \exp n(f(0)+h(0)-b) \end{aligned}$$

if $|x|\geqq 2d$ in J. We hence get the following bound for $E_{n,2}$:

$$E_{n,2}\leqq e^{n(f(0)+h(0))}\cdot b^{-M}\cdot e^{-nb}. \tag{27}$$

(Intervals where $h(x)=-\infty$ do not contribute to $E_{n,2}$.)
Coming now to $E_{n,1}$ we consider the non arithmetic case first and want to use Theorem 1. Therefore we approximate the integrand by convolving it with $k_{c/n}$:

$$\begin{aligned} E_{n,3}&=E(k_{c/n}*(e^{nf}I_{2d})(S_n/n))\\ E_{n,4}&=E_{n,1}-E_{n,3}, \end{aligned} \tag{28}$$

$$\begin{aligned} E_{n,3}&=\iint (n/c)\,k((n/c)(x/n-y))\,e^{nf(y)}I_{2d}(y)\,dy\,F^{n*}(dx)\\ &=n\int e^{nf(y)}I_{2d}(y)\,\varphi_{n,c}(y)\,dx \end{aligned} \tag{29}$$

When we approximate $\varphi_{n,c}(y)$ by using Theorem 1 the main term is:

$$\begin{aligned} E_{n,5}&=(n/2\pi)^{1/2}\int_{|x|\leqq 2d}(-h''(x))^{1/2}\,e^{n(f(x)+h(x))}\,dx\\ &=(n/2\pi)^{1/2}(-h''(0))^{1/2}\,e^{n(f(0)+h(0))}\cdot\int_{|x|\leqq 2d}(1+\mathcal{O}(d))\,e^{-n|x|^a(A+o(1))}\,dx. \end{aligned}$$

We have $\int e^{-(nB)|x|^a}dx=(nB)^{-1/a}2\Gamma(1+1/a)$, and for the tail of this integral the following simple estimate holds:

Lemma 2.

$$c^{1/a}\int_d^\infty e^{-cx^a}\,dx\leqq e^{-cd^a}(2/a)(cd^a)^{(1/a)-1} \tag{30}$$

when $cd^a\geqq 2/a$.

Proof.

$$\begin{aligned} \int_d^\infty e^{-cx^a}\,dx&=\mathop{I}_d^\infty xe^{-cx^a}+\int_d^\infty acx^a e^{-cx^a}\,dx\\ &\geqq -de^{-cd^a}+acd^a\int_d^\infty e^{-cx^a}\,dx. \end{aligned}$$

Hence $(acd^a-1)\int_d^\infty e^{-cx^a}\,dx\leqq de^{-cd^a}$, and (30) follows because $acd^a-1\geqq acd^a/2$.

Using this estimate we see that the integral in $E_{n,5}$ can be bounded above and below by

$$(nB)^{-1/a}\, 2\Gamma(1+1/a)\,(1+\mathcal{O}(d)+\mathcal{O}(e^{-nB(2d)^a}(nB(2d)^a)^{(1/a)-1})),$$

where B can be made arbitrarily close to A by chosing d small. Hence, if $d\to 0$, but $nd^a\to\infty$ we have

$$E_{n,5}=(2\pi)^{-1/2}\, 2\Gamma(1+1/a)\, A^{-1/2}(-h''(0))^{1/2}\, n^{1/2-1/a}\, e^{n(f(0)+h(0))}(1+o(1)) \quad (31)$$

and the same holds for $E_{n,3}$ if the error terms in Theorem 1 go to zero, i.e. if $c\to 0$ and $e^{-\gamma n}/c\to 0$.

It now remains to bound the term $E_{n,4}$ in (28), and we split it into two pieces as follows:

$$\begin{aligned}E_{n,4}&=E(I_d(e^{nf}I_{2d}-k_{c/n}*(e^{nf}I_{2d}))(S_n/n))\\&\quad+E((1-I_d)(e^{nf}I_{2d}-k_{c/n}*(e^{nf}I_{2d}))(S_n/n))=E_{n,6}+E_{n,7}.\end{aligned}$$

In $E_{n,6}$ the argument of the integrand lies in the interval $|x|\leqq d$. If also $c/n<d$ then in the integral defining $k_{c/n}*(e^{nf}I_{2d})$ the argument of the last factor lies in the interval $|y|\leqq 2d$, so the integrand in $E_{n,6}$ is given by:

$$\begin{aligned}(e^{nf}-k_{c/n}*e^{nf})(x)&=\int(e^{nf(x)}-e^{nf(x-y)})\,k_{c/n}(y)\,dy\\&=e^{nf(x)}\int(1-e^{n(f(x-y)-f(x))})\,k_{c/n}(y)\,dy,\end{aligned}$$

so using (24) it is bounded by:

$$|(e^{nf}-k_{c/n}*e^{nf})(x)|\leqq e^{nf(x)}\,2A'n(c/n)^{a'} \quad (32)$$

when $|x|\leqq d$ and d is small.

Hence we have:

$$|E_{n,6}|\leqq 2A'n(c/n)^{a'}E_n \quad (33)$$

when d is small and $c/n<d$.

For $E_{n,7}$ we have the estimate

$$|E_{n,7}|\leqq E((1-I_d)\,e^{nf}(S_n/n))+E((1-I_d)\,k_{c/n}*(e^{nf}I_{2d})(S_n/n)).$$

The first term is bounded just as $E_{n,2}$ above by (27). In the second term $k_{c/n}*(e^{nf}I_{2d})(x)=\int k_{c/n}(y)\,e^{nf(x-y)}I_{2d}(x-y)\,dy$ is zero if $|x|>2d+c/n$ because $k_{c/n}$ has compact support. Hence in the second term it can be bounded using (24) and (32) by e.g. $2e^{nf(x)}$ when d is small, so the second term can be bounded just as the first, and

$$|E_{n,7}|\leqq 3b^{-M}e^{-nb}\,e^{n(f(0)+h(0))} \quad (34)$$

when d is small and $c/n<d$.

Collecting all our estimates (27), (31), (33) and (34) we get the desired formula for E_n

$$E_n=(2\pi)^{-1/2}\, 2\Gamma(1+1/a)\, A^{-1/2}(-h''(0))^{1/2}\, n^{1/2-1/a}\, e^{n(f(0)+h(0))}(1+o(1)) \quad (35)$$

if c and d can be chosen suitably. We need to have $d \to 0$, $nd^a \to \infty$, $c/n < d$, $n(c/n)^{a'} \to 0$, $e^{-\gamma n}/c \to 0$, $n^{1/a-1/2} b^{-M} e^{-nb} \to 0$ with $b = \mathcal{O}(d^a)$. This can easily be achieved if e.g. $d = n^{-1/2a}$ and $c/n = n^{-l}$ with l large enough.

It finally remains to estimate $E_{n,1}$ in the arithmetic case. Now we can use Theorem 1 to estimate the atoms of the distribution of S_n directly and get:

$$E_{n,1} = (2\pi n)^{-1/2} \delta \sum_{|x| \leqq d} (-h''(x))^{1/2} e^{n(f(x)+h(x))} (1 + \mathcal{O}(n^{-1/2})), \tag{36}$$

where the summation is taken over $x = \xi + m\delta/n$, m integer. In estimating the sum in (36) we use (3) and the fact that $h''(x)$ is continuous at $x = 0$. We have to consider the cases $a > 1$, $a = 1$, $a < 1$ separatively however.

When $a > 1$ the sum can be approximated from above and below by

$$(2\pi n)^{1/2} (-h''(0))^{1/2} e^{n(f(0)+h(0))} \sum_{|x| \leqq d} e^{-nB|x|^a} (\delta/n)(1 + \mathcal{O}(d))$$

with B arbitrarily close to A if d is small. The sum is a Riemann sum approximating the integral $\int_{|x| \leqq d} e^{-nB|x|^a} dx$, so it follows that (35) holds as above if d is chosen as before.

When $a \leqq 1$ the sum is dominated by its largest terms as follows: Let $x_0 \leqq 0 < x_1$ be those x-values in the sum which are closest to $x = 0$. For a tail of the sum we then have the estimate

$$\sum_{y \geqq x} e^{-nB|y|^a} \leqq e^{-nB|x|^a} + \int_x^\infty e^{-nB|y|^a} dy \leqq 2e^{-nB|x|^a}$$

as $n \to \infty$ using Lemma 2.

Hence $\sum_{y > x_1} e^{-nB|y|^a} \leqq 2e^{-nB|x_1 + \delta/n|^a}$ e.g., and since $x_1 = \theta\delta/n$ with $0 \leqq \theta \leqq 1$ this is negligible compared to $e^{-nB|x_1|^a}$ because

$$n(|\theta\delta/n + \delta/n|^a) - |\theta\delta/n|^a) = n^{1-a} \delta^a (|\theta+1|^a - |\theta|^a) > n^{1-a} \delta^a (2^a - 1).$$

We conclude that the two terms corresponding to x_0 and x_1 dominate in the sum, so we have:

$$E_n = (2\pi n)^{-1/2} (-h''(0))^{1/2} \delta (e^{n(f(x_0)+h(x_0))} + e^{n(f(x_1)+h(x_1))})(1 + o(1)) \tag{37}$$

when $0 < a < 1$.

When $a = 1$ all the terms contribute however:

$$f(x) + h(x) = f(x_1) + h(x_1) - A(x - x_1) + o(x_1 + x)$$

for $x = x_1 + m\delta/n$ gives

$$\sum_{x \geqq x_1} e^{n(f(x)+h(x))} = e^{n(f(x_1)+h(x_1))} \sum_0^\infty e^{-Am\delta}(1 + o(1))$$
$$= e^{n(f(x_1)+h(x_1))}/(1 - e^{-A\delta})(1 + o(1)),$$

so in this case we see that

$$E_n = (2\pi n)^{-1/2}(-h''(0))^{1/2}(\delta/(1-e^{-A\delta})) \cdot (e^{n(f(x_0)+h(x_0))} + e^{n(f(x_1)+h(x_1))})(1+o(1)) \tag{38}$$

and we have finally proved the following result:

Theorem 2. *If $f(x)+h(x)$ has a single maximum at $x=0 \in D_h$ and if* (3) *and* (24) *hold near $x=0$ and the regularity condition* (26) *holds for $f(x)$ and the Cramér condition* (7) *holds for the distribution $F(x)$ then the asymptotic formula* (35) *holds for E_n defined by* (1) *in the non arithmetic case. In the arithmetic case* (7) *does not hold but* (35) *is true if $a>1$, whereas* (37) *holds if $a<1$ and* (38) *holds if $a=1$.*

3. Second Proof of the Laplace Approximation

In this section we show that the asymptotic formula (5) for E_n holds when $f(x)$ is twice continuously differentiable near 0 and $2A = -(f''(0)+h''(0)) > 0$, $a=2$ without assuming that the Cramér condition (7) holds.

From (3) it follows that

$$t \equiv f'(0) = -h'(0), \quad \text{so } g'(t) = 0, \tag{39}$$

and we have with $B = f''(0)$

$$f(x) = f(0) + tx + Bx^2/2 + o(x^2). \tag{40}$$

Again we split E_n into $E_{n,1}+E_{n,2}$ with

$$E_{n,1} = E(e^{nf(S_n/n)}, |S_n/n| \leqq d) \tag{41}$$

and estimate $E_{n,2}$ as before by (27) with $b = O(d^2)$. Because of (40) we approximate $E_{n,1}$ by

$$E_{n,3} = E(e^{n(f(0)+t(S_n/n)+B(S_n/n)^2/2)}, |S_n/n| \leqq d) \tag{42}$$

and have

$$|E_{n,1} - E_{n,3}| \leqq |e^{nd^2 r(d)} - 1| E_{n,3} \tag{43}$$

where $r(d) \to 0$ when $d \to 0$.

In estimating $E_{n,3}$ we have to consider the case $B>0$ and $B \leqq 0$ separately. When $B \leqq 0$ we use the fact that

$$e^{-z^2/2} = E(e^{izU}) \tag{44}$$

where U is a standard Gaussian random variable independent of $\{X_i\}$.

Hence

$$\begin{aligned}E_{n,3} &= e^{nf(0)} E(e^{(t+icU)S_n}, |S_n/n| \leqq d)\\ &= e^{nf(0)}(E(e^{(t+icU)S_n}) - E(e^{(t+icU),S_n}, |S_n/n| > d))\\ &= E_{n,4} - E_{n,5}\end{aligned} \tag{45}$$

with $c \equiv (|B|/n)^{1/2} \to 0$ as $n \to \infty$.

Since $|e^{(t+icU)S_n}| = e^{tS_n}$ $E_{n,5}$ can be bounded using Lemma 1 by

$$|E_{n,5}| \leqq e^{nf(0)} (\exp n \sup_{x>d} (h(x)+tx) + \exp n \sup_{x<-d} (h(x)+tx)).$$

From (10) and the fact that $t = -h'(0)$ we see that

$$g(t) = \max_x (h(x)+tx) = h(0)+0. \tag{46}$$

Since h is concave it follows that $\sup_{x>d} = h(d)+td$ and $\sup_{x<-d} = h(-d)-td = h(0) + d^2 h''(0)/2 + O(d^3)$, so we have

$$|E_{n,5}| \leqq 2 e^{n(f(0)+h(0))} e^{-nd^2 h''(0)/4} \quad \text{e.g.} \tag{47}$$

when d is small enough.

Consider now $E_{n,4}$. Using (46) we see that

$$E_{n,4} = e^{n(f(0)+h(0))} E(e^{n(g(t+icU)-g(t))}). \tag{48}$$

Since $|e^{g(t+icU)-g(t)}| \leqq 1$ it is no problem to truncate at $|U| = D$, $D \to \infty$ in (48). We have $P(|U| > D) = O(e^{-D^2/2})$. When $|U| \leqq D$ we use the Taylor expansion of g and get (with $g'(t) = 0$, $v \equiv g''(t) = -1/h''(0)$):

$$n(g(t+icU)-g(t)) = n(-c^2 U^2 v/2 + O(cD)^3) = BvU^2/2 + O(D^3/n^{1/2}),$$

and

$$\begin{aligned}E_{n,6} &\equiv E(e^{n(g(t+icU)-g(t))})\\ &= E(e^{BvU^2}, |U| \leqq D)(1+O(D^3/n^{1/2})) + O(e^{-D^2/2})\\ &= E(e^{BvU^2/2})(1+O(D^3/n^{1/2})) + O(e^{-D^2/2})\\ &= (1-Bv)^{-1/2}(1+O(D^3/n^{1/2})) + O(e^{-D^2/2}).\end{aligned} \tag{49}$$

If we choose $D = n^{1/8}$ e.g. we see that

$$E_{n,6} = (1-Bv)^{-1/2}(1+o(1)). \tag{50}$$

Collecting all the estimates (27), (43), (47), (48), (50) and putting $d = D_1/n^{1/2}$ we see that we have

$$E_n = (1-Bv)^{-1/2} e^{n(f(0)+h(0))}(1+o(1)) \tag{51}$$

if we first choose D_1 large enough and then n so big that $r(d)$ in (43) is small enough.

When $B>0$ the above proof has to be modified somewhat. Now we use (44) with z imaginary, and (45) is changed into:

$$E_{n,3}=e^{nf(0)}E(e^{(t+cU)S_n}, |S_n/n|\leqq d) \tag{52}$$

with $c=(B/n)^{1/2}\to 0$ as $n\to\infty$.

Let $J=[u_-, u_+]$ be the interval such that $t+cJ$ corresponds to the interval $[-d, d]\subset D_h$ in the mapping $t\to x=g'(t)$, i.e. $u_\pm$ are determined by

$$g'(t+cu_\pm)=\pm d. \tag{53}$$

$0\in J$ because $g'(t)=0$. We want to compare the integral in (52) to

$$\begin{aligned}E_{n,7}&=E(e^{(t+cU)S_n},\ U\in J)=E(e^{ng(t+cU)}, U\in J)\\&=e^{nh(0)}E(e^{n(g(t+cU)-g(t))}, U\in J),\end{aligned} \tag{54}$$

and to this end we show that the two integrals

$$E_{n,8}=E(E^{(t+cU)S_n}, |S_n/n|>d, U\in J) \tag{55}$$

and

$$E_{n,9}=E(e^{(t+cU)S_n}, |S_n/n|\leqq d, U\notin J) \tag{56}$$

are negligible.

For a fixed value $U=u$ the integral in (55) e.g. can be bounded using Lemma 1 by

$$\begin{aligned}E(e^{(t+cu)S_n}, |S_n/n|>d)\leqq&\exp n\sup_{x>d}(h(x)+(t+cu)x)\\&+\exp n\sup_{x<-d}(h(x)+(t+cu)x).\end{aligned}$$

Since $u\in J$ in (55) $h'(x)+(t+cu)\neq 0$ when $|x|<d$, so the sup above are attained for $x=\pm d$ respectively, and since $t=-h'(0)$ we have

$$\begin{aligned}h(\pm d)\pm td\leqq h(0)-d^2/2\bar{v}\quad&\text{if}\\ \bar{v}=\sup_{u\in J}g''(t+cu)=\sup_{|x|\leqq d}&(-1/h''(x)).\end{aligned} \tag{57}$$

From this we see that

$$\begin{aligned}E_{n,8}&\leqq e^{nh(0)-nd^2/2\bar{v}}E(e^{ncdU}+e^{-ncdU})\\&=e^{nh(0)-nd^2/2\bar{v}}2e^{(ncd)^2/2}=2e^{nh(0)}e^{-nd^2(1-B\bar{v})/2\bar{v}}\end{aligned} \tag{58}$$

because $(nc)^2=nB$.

By the same argument we can bound $E_{n,9}$ as in (58).

Returning now to (54) we have

$$n(g(t+cU)-g(t))=nc^2U^2v/2+O(n(cd)^3)=BuU^2/2+O(d^3/n^{1/2})$$

when $U\in J$, so

$$E_{n,7}=e^{nh(0)}E(e^{BvU^2/2}, U\in J)(1+O(d^3/n^{1/2})).$$

From (53) we see that

$$cu_{\pm} = -h'(\pm d) - t = -h'(\pm d) + h'(0) = O(d),$$

so $u_{\pm} \to \pm\infty$ if $d/c = n^{1/2} d/B^{1/2} \to \infty$, and

$$\begin{aligned} E_{n,7} &= e^{nh(0)}(E(e^{BvU^2/2}) + O(d^3/n^{1/2}) + O(e^{-(1-Bv)u_{\pm}^2/2})) \\ &= e^{nh(0)}((1-Bv)^{-1/2} + O(d^3/n^{1/2}) + O(e^{-(1-Bv)u_{\pm}^2/2})). \end{aligned} \tag{59}$$

Collecting the estimates (27), (43), (52), (54), (58), (59) we see again that (51) is true if we choose $d = D_1/n^{1/2}$ with D_1 large enough and then choose n big enough as before. Let us sum up this result:

Theorem 3. *If $f(x)+h(x)$ has a single maximum at $x=0 \in D_h$, and if $f(x)$ is twice continuously differentiable near 0 with $-2A = f''(0)+h''(0) < 0$, so (3) holds with $a=2$, and if the regularity condition (26) holds then the asymptotic formula*

$$E_n = (1 + f''(0)/h''(0))^{-1/2}\, e^{n(f(0)+h(0))}(1+o(1)) \tag{60}$$

holds for E_n defined by (1) as $n \to \infty$.

4. The Central Limit Theorem for $n^{1/a}(S_n/n - \bar{x})$

In this section we show that a simple modification of the proofs of Theorem 1 and 2 shows that the limit distribution of $Y_n = n^{1/a}(S_n/n - \bar{x})$ described in the introduction is obtained as $n \to \infty$.

The modified distribution of S_n was defined by $F_{n,f}(dx) = e^{nf(x/n)} F^{n*}(dx)/E_n$, so the characteristic function of Y_n is given by: (we take $\bar{x}=0$)

$$\begin{aligned} E_{n,f}(e^{isY_n}) &= E(e^{nf(S_n/n) + isn^{1/a}(S_n/n)})/E(e^{nf(S_n/n)}) \\ &= E_n(s)/E_n \qquad \text{for } s \in R^1. \end{aligned} \tag{61}$$

By the continuity theorem for characteristic functions it is enough to show that $E_n(s)/E_n$ converges to $\int e^{isx - A|x|^a} A^{1/a} dx/2\Gamma(1+1/a)$ as $n \to \infty$ when $a > 1$.

$E_n(s)$ is obtained from E_n simply by perturbing $f(x)$ to $f_n(x) = f(x) + isn^{1/a-1}x$. If we note that since $a>1$ $f(x)$ and hence $f_n(x)$ uniformly in n satisfy a Lip. condition, $|f_n(x) - f_n(y)| \leqq A'|x-y|$, near 0 and that $|e^{nf_n(x)}| = e^{nf(x)}$ we can repeat all the steps of the proofs of Theorem 1 and 2 almost as before and get a corresponding asymptotic formula for $E_n(s)$: (27) still holds for $|E_{n,2}(s)|$. In $E_{n,5}$ we have $d = n^{-1/2a}$, and $n(f_n(x/n^{1/a}) + h(x/n^{1/a})) \to -A|x|^a + isx$, so by bounded convergence we have

$$E_{n,5}(s) = (n/2\pi)^{1/2}(-h''(0))^{1/2} e^{n(f(0)+h(0))} n^{-1/a} \int e^{isx - A|x|^a} dx (1+o(1)). \tag{62}$$

(33) and (34) hold just as before, so from (35) we see that

$$E_n(s)/E_n = \int e^{isx - A|x|^a} A^{1/a} dx/2\Gamma(1+1/a)(1+o(1)). \tag{63}$$

In the proof of Theorem 2 we have to add a term $isn^{-1/2}S_n$ in the exponent in (41) and (42). Since $c=(-B/n)^{1/2}$ the exponent in (45) changes to $(t+ic(U+s/(-B)^{1/2}))S_n$. The only change in the following is hence that U is changed into $U+s/(-B)^{1/2}$ (with $(-B)^{1/2}$ imaginary if $B>0$). The rest of the proof then goes as before, and the main term in (49) and (59) is changed into

$$E(e^{Bv(U+s/(-B)^{1/2})^2/2})=e^{-s^2v/2(1-Bv)}/(1-Bv)^{1/2},$$

and we have

$$E_n(s)/E_n=e^{-s^2v/2(1-Bv)}(1+o(1)), \tag{64}$$

so the limit distribution is Gaussian with variance $v/(1-Bv)=-1/(f''(o)+h''(0))$. Let us sum up this result:

Theorem 4. *Under the assumptions of Theorem 2 and 3 respectively, when S_n is given the distribution $F_{n,f}(dx)=e^{nf(x/n)}F^{n*}(dx)/E_n$ then when $a>1$ the distribution of $Y_n=n^{1/a}(S_n/n-\bar{x})$ converges weakly to the one defined by the density $e^{-A|x|^a}A^{1/a}dx/2\Gamma(1+1/a)$ as $n\to\infty$. When $a=2$, $A>0$ this is a Gaussian with variance*

$$1/2A=-1/(f''(0)+h''(0)).$$

References

1. Bahadur, R.R., Ranga Rao, R.: On Deviations of the Sample Mean. Ann. Math. Statist. **31**, 1015–1027 (1960)
2. Bahadur, R.R., Zabell, S.L.: Large Deviations of the Sample Mean in General Vector Spaces. Ann. Probability **7**, 587–621 (1979)
3. von Bahr, B., Martin-Löf, A.: Threshold Limit Theorems for Some Epidemic Processes. Adv. Appl. Probability **12**, 319–349 (1980)
4. Copson, E.T.: Asymptotic Expansions. Cambridge University Press 1965
5. Ellis, R.S., Newman, C.M.: Limit Theorems for Sums of Dependent Random Variables Occurring in Statistical Mechanics. Z. Wahrscheinlichkeitstheorie verw. Geb. **44**, 117–139 (1978)
6. Höglund, T.: A Unified Central Limit Theorem for Small and Large Deviations from the Mean. Z. Wahrscheinlichkeitstheorie verw. Geb. **49**, 105–117 (1979)
7. Lanford, O.E.: in Statistical Mechanics and Mathematical Problems, Lecture Notes in Physics no. **20**, Berlin-Heidelberg-New York: Springer 1971
8. Lucas, E.: Characteristic Functions. New York: Griffin & Co. (1960)
9. Martin-Löf, P.: Statistiska Modeller. Lecture Notes, Dept. of Mathematical Statistics University of Stockholm (1969)
10. Rockafellar, R.T.: Convex Analysis. Princeton Univ. Press (1970)
11. Varadhan, S.R.S.: Asymptotic Probabilities and Differential Equations. Comm. Pure Appl. Math. **19**, 261–286 (1966)

Received March 19, 1980; in revised form August 31, 1981