

Chapter 9

Large Deviation Probabilities for Sums of Independent Random Variables

Abstract The material presented in this chapter is unique to the present text. After an introductory discussion of the concept and importance of large deviation probabilities, Cramér’s condition is introduced and the main properties of the Cramér and Laplace transforms are discussed in Sect. 9.1. A separate subsection is devoted to an in-depth analysis of the key properties of the large deviation rate function, followed by Sect. 9.2 establishing the fundamental relationship between large deviation probabilities for sums of random variables and those for sums of their Cramér transforms, and discussing the probabilistic meaning of the rate function. Then the logarithmic Large Deviations Principle is established. Section 9.3 presents integro-local, integral and local theorems on the exact asymptotic behaviour of the large deviation probabilities in the so-called Cramér range of deviations. Section 9.4 is devoted to analysing various types of the asymptotic behaviours of the large deviation probabilities for deviations at the boundary of the Cramér range that emerge under different assumptions on the distributions of the random summands. In Sect. 9.5, the behaviour of the large deviation probabilities is found in the case of heavy-tailed distributions, namely, when the distributions tails are regularly varying at infinity. These results are used in Sect. 9.6 to find the asymptotics of the large deviation probabilities beyond the Cramér range of deviations, under special assumptions on the distribution tails of the summands.

Let ξ, ξ_1, ξ_2, \dots be a sequence of independent identically distributed random variables,

$$\mathbf{E}\xi_k = 0, \quad \mathbf{E}\xi_k^2 = \sigma^2 < \infty, \quad S_n = \sum_{k=1}^n \xi_k.$$

Suppose that we have to evaluate the probability $\mathbf{P}(S_n \geq x)$. If $x \sim v\sqrt{n}$ as $n \rightarrow \infty$, $v = \text{const}$, then by the integral limit theorem

$$\mathbf{P}(S_n \geq x) \sim 1 - \Phi\left(\frac{v}{\sigma}\right) \tag{9.0.1}$$

as $n \rightarrow \infty$. But if $x \gg \sqrt{n}$, then the integral limit theorem enables one only to conclude that $\mathbf{P}(S_n \geq x) \rightarrow 0$ as $n \rightarrow \infty$, which in fact contains no quantitative

information on the probability we are after. Essentially the same can happen for fixed but “relatively” large values of v/σ . For example, for $v/\sigma \geq 3$ and the values of n around 100, the relative accuracy of the approximation in (9.0.1) becomes, generally speaking, bad (the true value of the left-hand side can be several times greater or smaller than that of the right-hand side). Studying the asymptotic behaviour of $\mathbf{P}(S_n \geq x)$ for $x \gg \sqrt{n}$ as $n \rightarrow \infty$, which is not known to us yet, could fill these gaps. This problem is highly relevant since questions of just this kind arise in many problems of mathematical statistics, insurance theory, the theory of queueing systems, etc. For instance, in mathematical statistics, finding *small* probabilities of errors of the first and second kind of statistical tests when the sample size n is large leads to such problems (e.g. see [7]). In these problems, we have to find explicit functions $P(n, x)$ such that

$$\mathbf{P}(S_n \geq x) = P(n, x)(1 + o(1)) \quad (9.0.2)$$

as $n \rightarrow \infty$. Thus, unlike the case of normal approximation (9.0.1), here we are looking for approximations $P(n, x)$ with a *relatively* small error rather than an *absolutely* small error. If $P(n, x) \rightarrow 0$ in (9.0.2) as $n \rightarrow \infty$, then we will speak of the probabilities of *rare events*, or of the *probabilities of large deviations* of sums S_n . Deviations of the order \sqrt{n} are called *normal deviations*.

In order to study large deviation probabilities, we will need some notions and assertions.

9.1 Laplace’s and Cramér’s Transforms. The Rate Function

9.1.1 The Cramér Condition. Laplace’s and Cramér’s Transforms

In all the sections of this chapter, except for Sect. 9.5, the following *Cramér condition* will play an important role.

[C] *There exists a $\lambda \neq 0$ such that*

$$\mathbf{E}e^{\lambda\xi} = \int e^{\lambda y} \mathbf{F}(dy) < \infty. \quad (9.1.1)$$

We will say that the *right-side* (*left-side*) Cramér condition holds if $\lambda > 0$ ($\lambda < 0$) in (9.1.1). If (9.1.1) is valid for some negative and positive λ (i.e. in a neighbourhood of the point $\lambda = 0$), then we will say that the *two-sided Cramér’s condition* is satisfied.

The Cramér condition can be interpreted as characterising a fast (at least exponentially fast) rate of decay of the tails $F_{\pm}(t)$ of the distribution \mathbf{F} . If, for instance, we have (9.1.1) for $\lambda > 0$, then by Chebyshev’s inequality, for $t > 0$,

$$F_+(t) := \mathbf{P}(\xi \geq t) \leq e^{-\lambda t} \mathbf{E}e^{\lambda\xi},$$

i.e. $F_+(t)$ decreases at least exponentially fast. Conversely, if, for some $\mu > 0$, one has $F_+(t) \leq ce^{-\mu t}$, $t > 0$, then, for $\lambda \in (0, \mu)$,

$$\begin{aligned} \int_0^\infty e^{\lambda y} \mathbf{F}(dy) &= - \int_0^\infty e^{\lambda y} dF_+(y) = F_+(0) + \lambda \int_0^\infty e^{\lambda y} F_+(y) dy \\ &\leq F_+(0) + c\lambda \int_0^\infty e^{(\lambda-\mu)y} dy = F_+(0) + \frac{c\lambda}{\mu - \lambda} < \infty. \end{aligned}$$

Since the integral $\int_{-\infty}^0 e^{\lambda y} \mathbf{F}(dy)$ is finite for any $\lambda > 0$, we have $\mathbf{E}e^{\lambda\xi} < \infty$ for $\lambda \in (0, \mu)$.

The situation is similar for the left tail $F_-(t) := \mathbf{P}(\xi < -t)$ provided that (9.1.1) holds for some $\lambda < 0$.

Set

$$\lambda_+ := \sup\{\lambda : \mathbf{E}e^{\lambda\xi} < \infty\}, \quad \lambda_- := \inf\{\lambda : \mathbf{E}e^{\lambda\xi} < \infty\}.$$

Condition [C] is equivalent to $\lambda_+ > \lambda_-$. The right-side Cramér condition means that $\lambda_+ > 0$; the two-sided condition means that $\lambda_+ > 0 > \lambda_-$. Clearly, the ch.f. $\varphi(t) = \mathbf{E}e^{it\xi}$ is analytic in the complex plane in the strip $-\lambda_+ < \text{Im } t < -\lambda_-$. This follows from the differentiability of $\varphi(t)$ in this region of the complex plane, since the integral $\int |ye^{ity}| \mathbf{F}(dy)$ for the said values of $\text{Im } t$ converges uniformly in $\text{Re } t$.

Here and henceforth by the *Laplace transform* (Laplace–Stieltjes or Laplace–Lebesgue) of the distribution \mathbf{F} of the random variable ξ we shall mean the function

$$\psi(\lambda) := \mathbf{E}e^{\lambda\xi} = \varphi(-i\lambda),$$

which conflicts with Sect. 7.1.1 (and the terminology of mathematical analysis), according to which the term Laplace’s transform refers to the function $\mathbf{E}e^{-\lambda\xi} = \varphi(i\lambda)$. The reason for such a slight inconsistency in terminology (only the sign of the argument differs, this changes almost nothing) is our reluctance to introduce new notation or to complicate the old notation. Nowhere below will it cause confusion.¹

As well as condition [C], we will also assume that the random variable ξ is nondegenerate, i.e. $\xi \not\equiv \text{const}$ or, which is the same, $\text{Var } \xi > 0$.

The main properties of Laplace’s transform.

As was already noted in Sect. 7.1.1, Laplace’s transform, like the ch.f., uniquely characterises the distribution \mathbf{F} . Moreover, it has the following properties, which are similar to the corresponding properties of ch.f.s (see Sect. 7.1). Under obvious conventions of notation,

$$(\Psi 1) \quad \psi_{a+b\xi}(\lambda) = e^{\lambda a} \psi_\xi(b\lambda), \text{ if } a \text{ and } b \text{ are constant.}$$

¹In the literature, the function $\mathbf{E}e^{\lambda\xi}$ is sometimes called the “moment generating function”.

($\Psi 2$) If ξ_1, \dots, ξ_n are independent and $S_n = \sum_{j=1}^n \xi_j$, then

$$\psi_{S_n}(\lambda) = \prod_{j=1}^n \psi_{\xi_j}(\lambda).$$

($\Psi 3$) If $\mathbf{E}|\xi|^k < \infty$ and the right-side Cramér condition is satisfied then the function ψ_ξ is k -times right differentiable at the point $\lambda = 0$,

$$\psi_\xi^{(k)}(0) = \mathbf{E}\xi^k =: m_k$$

and, as $\lambda \downarrow 0$,

$$\psi_\xi(\lambda) = 1 + \sum_{j=1}^k \frac{\lambda^j}{j!} m_j + o(\lambda^k).$$

This also implies that, as $\lambda \downarrow 0$, the representation

$$\ln \psi_\xi(\lambda) = \sum_{j=1}^k \frac{\gamma_j \lambda^j}{j!} + o(\lambda^k), \quad (9.1.2)$$

holds, where γ_j are the so-called *semi-invariants* (or *cumulants*) of order j of the random variable ξ . One can easily verify that

$$\gamma_1 = m_1, \quad \gamma_2 = m_2^0 = \sigma^2, \quad \gamma_3 = m_3^0, \quad \dots, \quad (9.1.3)$$

where $m_k^0 = \mathbf{E}(\xi - m_1)^k$ is the central moment of order k .

Definition 9.1.1 Let condition [C] be met. The Cramér transform at the point λ of the distribution \mathbf{F} is the distribution²

$$\mathbf{F}_{(\lambda)}(dy) = \frac{e^{\lambda y} \mathbf{F}(dy)}{\psi(\lambda)}. \quad (9.1.4)$$

²In some publications the transform (9.1.4) is also called the *Esscher transform*. However, the systematic use of transform (9.1.4) for the study of large deviations was first done by Cramér.

If we study the probabilities of large deviations of sums of random variables using the inversion formula, similarly to what was done for normal deviations in Chap. 8, then we will necessarily come to employ the so-called *saddle-point method*, which consists of moving the contour of integration so that it passes through the so-called *saddle point*, at which the exponent in the integrand function, as we move along the imaginary axis, attains its minimum (and, along the real axis, attains its maximum; this explains the name “saddle point”). Cramér’s transform does essentially the same, making such a translation of the contour of integration even before applying the inversion formula, and reduces the large deviation problem to the normal deviation problem, where the inversion formula is not needed if we use the results of Chap. 8. It is this technique that we will follow in the present chapter.

Clearly, the distributions \mathbf{F} and $\mathbf{F}_{(\lambda)}$ are mutually absolutely continuous (see Sect. 3.5 of Appendix 3) with density

$$\frac{\mathbf{F}_{(\lambda)}(dy)}{\mathbf{F}(dy)} = \frac{e^{\lambda y}}{\psi(\lambda)}.$$

Denote a random variable with distribution $\mathbf{F}_{(\lambda)}$ by $\xi_{(\lambda)}$.

The Laplace transform of the distribution $\mathbf{F}_{(\lambda)}$ is obviously equal to

$$\mathbf{E}e^{\mu\xi_{(\lambda)}} = \frac{\psi(\lambda + \mu)}{\psi(\lambda)}. \quad (9.1.5)$$

Clearly,

$$\begin{aligned} \mathbf{E}\xi_{(\lambda)} &= \frac{\psi'(\lambda)}{\psi(\lambda)} = (\ln \psi(\lambda))', & \mathbf{E}\xi_{(\lambda)}^2 &= \frac{\psi''(\lambda)}{\psi(\lambda)}, \\ \text{Var}(\xi_{(\lambda)}) &= \frac{\psi''(\lambda)}{\psi(\lambda)} - \left(\frac{\psi'(\lambda)}{\psi(\lambda)}\right)^2 = (\ln \psi(\lambda))''. \end{aligned}$$

Since $\psi''(\lambda) > 0$ and $\text{Var}(\xi_{(\lambda)}) > 0$, the foregoing implies one more important property of the Laplace transform.

($\Psi 4$) *The functions $\psi(\lambda)$ and $\ln \psi(\lambda)$ are strictly convex, and*

$$\mathbf{E}\xi_{(\lambda)} = \frac{\psi'(\lambda)}{\psi(\lambda)}$$

strictly increases on (λ_-, λ_+) .

The analyticity of $\psi(\lambda)$ in the strip $\text{Re } \lambda \in (\lambda_-, \lambda_+)$ can be supplemented by the following “extended” continuity property on the segment $[\lambda_-, \lambda_+]$ (in the strip $\text{Re } \lambda \in [\lambda_-, \lambda_+]$).

($\Psi 5$) *The function $\psi(\lambda)$ is continuous “inside” $[\lambda_-, \lambda_+]$, i.e. $\psi(\lambda_{\pm} \mp 0) = \psi(\lambda_{\pm})$ (where the cases $\psi(\lambda_{\pm}) = \infty$ are not excluded).*

Outside the segment $[\lambda_-, \lambda_+]$ such continuity, generally speaking, does not hold as, for example, is the case when $\psi(\lambda_+) < \infty$ and $\psi(\lambda_+ + 0) = \infty$, which takes place, say, for the distribution \mathbf{F} with density $f(x) = cx^{-3}e^{-\lambda_+x}$ for $x \geq 1$, $c = \text{const}$.

9.1.2 The Large Deviation Rate Function

Under condition [C], the *large deviation rate function* will play the determining role in the description of asymptotics of probabilities $\mathbf{P}(S_n \geq x)$.

Definition 9.1.2 The large deviation rate function (or, for brevity, simply the rate function) A of a random variable ξ is defined by

$$A(\alpha) := \sup_{\lambda} (\alpha\lambda - \ln \psi(\lambda)). \quad (9.1.6)$$

The meaning of the name will become clear later. In classical analysis, the right-hand side of (9.1.6) is known as the *Legendre transform* of the function $\ln \psi(\lambda)$.

Consider the function $A(\alpha, \lambda) = \alpha\lambda - \ln \psi(\lambda)$ of the supremum appearing in (9.1.6). The function $-\ln \psi(\lambda)$ is strictly concave (see property ($\Psi 4$)), and hence so is the function $A(\alpha, \lambda)$ (note also that $A(\alpha, \lambda) = -\ln \psi_{\alpha}(\lambda)$, where $\psi_{\alpha}(\lambda) = e^{-\lambda\alpha} \psi(\lambda)$ is the Laplace transform of the distribution of the random variable $\xi - \alpha$ and, therefore, from the “qualitative point of view”, $A(\alpha, \lambda)$ possesses all the properties of the function $-\ln \psi(\lambda)$). The foregoing implies that there always exists a *unique* point $\lambda = \lambda(\alpha)$ (on the “extended” real line $[-\infty, \infty]$) at which the supremum in (9.1.6) is attained. As α grows, the value of $A(\alpha, \lambda)$ for $\lambda > 0$ increases (proportionally to λ), and for $\lambda < 0$ it decreases. Therefore, the graph of $A(\alpha, \lambda)$ as the function of λ will, roughly speaking, “roll over” to the right as α grows. This means that the maximum point $\lambda(\alpha)$ will also move to the right (or stay at the same place if $\lambda(\alpha) = \lambda_+$).

We now turn to more precise formulations. On the interval $[\lambda_-, \lambda_+]$, there exists the derivative (respectively, the right and the left derivative at the endpoints λ_{\pm})

$$A'_{\lambda}(\alpha, \lambda) = \alpha - \frac{\psi'(\lambda)}{\psi(\lambda)}. \quad (9.1.7)$$

The parameters

$$\alpha_{\pm} = \frac{\psi'(\lambda_{\pm} \mp 0)}{\psi(\lambda_{\pm} \mp 0)}, \quad \alpha_- < \alpha_+, \quad (9.1.8)$$

will play an important role in what follows. The value of α_+ determines the angle at which the curve $\ln \psi(\lambda)$ “sticks” into the point $(\lambda_+, \ln \psi(\lambda_+))$. The quantity α_- has a similar meaning. If $\alpha \in [\alpha_-, \alpha_+]$ then the equation $A'_{\lambda}(\alpha, \lambda) = 0$, or (see (9.1.7))

$$\frac{\psi'(\lambda)}{\psi(\lambda)} = \alpha, \quad (9.1.9)$$

always has a unique solution $\lambda(\alpha)$ on the segment $[\lambda_-, \lambda_+]$ (λ_{\pm} can be infinite). This solution $\lambda(\alpha)$, being the inverse of an analytical and strictly increasing function $\frac{\psi'(\lambda)}{\psi(\lambda)}$ on (λ_-, λ_+) (see (9.1.9)), is also analytical and strictly increasing on (α_-, α_+) ,

$$\lambda(\alpha) \uparrow \lambda_+ \quad \text{as } \alpha \uparrow \alpha_+; \quad \lambda(\alpha) \downarrow \lambda_- \quad \text{as } \alpha \downarrow \alpha_-. \quad (9.1.10)$$

The equalities

$$A(\alpha) = \alpha\lambda(\alpha) - \ln \psi(\lambda(\alpha)), \quad \frac{\psi'(\lambda(\alpha))}{\psi(\lambda(\alpha))} = \alpha \quad (9.1.11)$$

yield

$$\Lambda'(\alpha) = \lambda(\alpha) + \alpha\lambda'(\alpha) - \frac{\psi'(\lambda(\alpha))}{\psi(\lambda(\alpha))}\lambda'(\alpha) = \lambda(\alpha).$$

Recalling that

$$\frac{\psi'(0)}{\psi(0)} = m_1 = \mathbf{E}\xi, \quad 0 \in [\lambda_-, \lambda_+], \quad m_1 \in [\alpha_-, \alpha_+],$$

we obtain the following representation for the function Λ :

(A1) *If $\alpha_0 \in [\alpha_-, \alpha_+]$, $\alpha \in [\alpha_-, \alpha_+]$ then*

$$\Lambda(\alpha) = \Lambda(\alpha_0) + \int_{\alpha_0}^{\alpha} \lambda(v) dv. \tag{9.1.12}$$

Since $\lambda(m_1) = \Lambda(m_1) = 0$ (this follows from (9.1.9) and (9.1.11)), we obtain, in particular, for $\alpha_0 = m_1$, that

$$\Lambda(\alpha) = \int_{m_1}^{\alpha} \lambda(v) dv. \tag{9.1.13}$$

The functions $\lambda(\alpha)$ and $\Lambda(\alpha)$ are analytic on (α_-, α_+) .

Now consider what happens outside the segment $[\alpha_-, \alpha_+]$. Assume for definiteness that $\lambda_+ > 0$. We will study the behaviour of the functions $\lambda(\alpha)$ and $\Lambda(\alpha)$ near the point α_+ and for $\alpha > \alpha_+$. Similar results hold true in the vicinity of the point α_- in the case $\lambda_- < 0$.

First let $\lambda_+ = \infty$, i.e. the function $\ln \psi(\lambda)$ is analytic on the whole semiaxis $\lambda > 0$, and the tail $F_+(t)$ decays as $t \rightarrow \infty$ faster than any exponential function. Denote by

$$s_{\pm} = \pm \sup\{t : F_{\pm}(t) > 0\}$$

the boundaries of the support of \mathbf{F} . Without loss of generality, we will assume that

$$s_+ > 0, \quad s_- < 0. \tag{9.1.14}$$

This can always be achieved by shifting the random variable, similarly to our assuming, without loss of generality, $\mathbf{E}\xi = 0$ in many theorems of Chap. 8, where we used the fact that the problem of studying the distribution of S_n is “invariant” with respect to a shift. (We can also note that $\Lambda_{\xi-a}(\alpha - a) = \Lambda_{\xi}(\alpha)$, see property (A4) below, and that (9.1.14) always holds provided that $\mathbf{E}\xi = 0$.)

(A2) (i) *If $\lambda_+ = \infty$ then $\alpha_+ = s_+$.*

Hence, for $s_+ = \infty$, we always have $\alpha_+ = \infty$ and so for any $\alpha \geq \alpha_-$ we are dealing with the already considered “regular” case, where (9.1.12) and (9.1.13) hold true.

(ii) If $s_+ < \infty$ then $\lambda_+ = \infty$, $\alpha_+ = s_+$,

$$\Lambda(\alpha_+) = -\ln \mathbf{P}(\xi = s_+), \quad \Lambda(\alpha) = \infty \quad \text{for } \alpha > \alpha_+.$$

Similar assertions hold true for s_- , α_- , λ_- .

Proof (i) First let $s_+ < \infty$. Then the asymptotics of $\psi(\lambda)$ and $\psi'(\lambda)$ as $\lambda \rightarrow \infty$ is determined by the integrals in a neighbourhood of the point s_+ : for any fixed $\varepsilon > 0$,

$$\psi(\lambda) \sim \mathbf{E}(e^{\lambda\xi}; \xi > s_+ - \varepsilon), \quad \psi'(\lambda) \sim \mathbf{E}(\xi e^{\lambda\xi}; \xi > s_+ - \varepsilon)$$

as $\lambda \rightarrow \infty$. Hence

$$\alpha_+ = \lim_{\lambda \rightarrow \infty} \frac{\psi'(\lambda)}{\psi(\lambda)} = \lim_{\lambda \rightarrow \infty} \frac{\mathbf{E}(\xi e^{\lambda\xi}; \xi > s_+ - \varepsilon)}{\mathbf{E}(e^{\lambda\xi}; \xi > s_+ - \varepsilon)} = s_+.$$

If $s_+ = \infty$, then $\ln \psi(\lambda)$ grows as $\lambda \rightarrow \infty$ faster than any linear function and therefore the derivative $(\ln \psi(\lambda))'$ increases unboundedly, $\alpha_+ = \infty$.

(ii) The first two assertions are obvious. Further, let $p_+ = \mathbf{P}(\xi = s_+) > 0$. Then

$$\psi(\lambda) \sim p_+ e^{\lambda s_+},$$

$$\alpha\lambda - \ln \psi(\lambda) = \alpha\lambda - \ln p_+ - \lambda s_+ + o(1) = (\alpha - \alpha_+)\lambda - \ln p_+ + o(1)$$

as $\lambda \rightarrow \infty$. This and (9.1.11) imply that

$$\Lambda(\alpha) = \begin{cases} -\ln p_+ & \text{for } \alpha = \alpha_+, \\ \infty & \text{for } \alpha > \alpha_+. \end{cases}$$

If $p_+ = 0$, then the relation $\psi(\lambda) = o(e^{\lambda s_+})$ as $\lambda \rightarrow \infty$ similarly implies $\Lambda(\alpha_+) = \infty$. Property (A2) is proved. \square

Now let $0 < \lambda_+ < \infty$. If $\alpha_+ < \infty$, then necessarily $\psi(\lambda_+) < \infty$, $\psi(\lambda_+ + 0) = \infty$ and $\psi'(\lambda_+) < \infty$ (here we mean the left derivative). If we assume that $\psi(\lambda_+) = \infty$, then $\ln \psi(\lambda_+) = \infty$, $(\ln \psi(\lambda))' \rightarrow \infty$ as $\lambda \uparrow \lambda_+$ and $\alpha_+ = \infty$, which contradicts the assumption $\alpha_+ < \infty$. Since $\psi(\lambda) = \infty$ for $\lambda > \lambda_+$, the point $\lambda(\alpha)$, having reached the value λ_+ as α grows, will stop at that point. So, for $\alpha \geq \alpha_+$, we have

$$\lambda(\alpha) = \lambda_+, \quad \Lambda(\alpha) = \alpha\lambda_+ - \ln \psi(\lambda_+) = \Lambda(\alpha_+) + \lambda_+(\alpha - \alpha_+). \quad (9.1.15)$$

Thus, in this case, for $\alpha \geq \alpha_+$ the function $\lambda(\alpha)$ remains constant, while $\Lambda(\alpha)$ grows linearly. Relations (9.1.12) and (9.1.13) remain true.

If $\alpha_+ = \infty$, then $\alpha < \alpha_+$ for all finite $\alpha \geq \alpha_-$, and we again deal with the ‘‘regular’’ case that we considered earlier (see (9.1.12) and (9.1.13)). Since $\lambda(\alpha)$ does not decrease, these relations imply the convexity of $\Lambda(\alpha)$.

In summary, we can formulate the following property.

(A3) *The functions $\lambda(\alpha)$ and $\Lambda(\alpha)$ can only be discontinuous at the points s_{\pm} and under the condition $\mathbf{P}(\xi = s_{\pm}) > 0$. These points separate the domain (s_-, s_+) where the function Λ is finite and continuous (in the extended sense) from the domain $\alpha \notin [s_-, s_+]$ where $\Lambda(\alpha) = \infty$. In the domain $[s_-, s_+]$ the function Λ is convex. (If we define convexity in the “extended” sense, i.e. including infinite values as well, then Λ is convex on the entire real line.) The function Λ is analytic in the interval (α_-, α_+) . If $\lambda_+ < \infty$ and $\alpha_+ < \infty$, then on the half-line (α_+, ∞) the function $\Lambda(\alpha)$ is linear with slope λ_+ ; at the boundary point α_+ the continuity of the first derivatives persists. If $\lambda_+ = \infty$, then $\Lambda(\alpha) = \infty$ on (α_+, ∞) . The function $\Lambda(\alpha)$ possesses a similar property on $(-\infty, \alpha_-)$.*

If $\lambda_- = 0$, then $\alpha_- = m_1$ and $\lambda(\alpha) = \Lambda(\alpha) = 0$ for $\alpha \leq m_1$.

Indeed, since $\lambda(m_1) = 0$ and $\psi(\lambda) = \infty$ for $\lambda < \lambda_- = 0 = \lambda(m_1)$, as the value of α decreases to $\alpha_- = m_1$, the point $\lambda(\alpha)$, having reached the value 0, will stop, and $\lambda(\alpha) = 0$ for $\alpha \leq \alpha_- = m_1$. This and the first identity in (9.1.11) also imply that $\Lambda(\alpha) = 0$ for $\alpha \leq m_1$.

If $\lambda_- = \lambda_+ = 0$ (condition [C] is not met), then $\lambda(\alpha) = \Lambda(\alpha) \equiv 0$ for all α . This is obvious, since the value of the function under the sup sign in (9.1.6) equals $-\infty$ for all $\lambda \neq 0$. In this case the limit theorems presented in the forthcoming sections will be of little substance.

We will also need the following properties of the function Λ .

(A4) *Under obvious notational conventions, for independent random variables ξ and η , we have*

$$\Lambda_{\xi+\eta}(\alpha) = \sup_{\lambda} (\alpha\lambda - \ln \psi_{\xi}(\lambda) - \ln \psi_{\eta}(\lambda)) = \inf_{\gamma} (\Lambda_{\xi}(\gamma) + \Lambda_{\eta}(\alpha - \gamma)),$$

$$\Lambda_{c\xi+b}(\alpha) = \sup_{\lambda} (\alpha\lambda - \lambda b - \ln \psi_{\xi}(\lambda c)) = \Lambda_{\xi}\left(\frac{\alpha - b}{c}\right).$$

Clearly, \inf_{γ} in the former relation is attained at the point γ at which $\lambda_{\xi}(\gamma) = \lambda_{\eta}(\alpha - \gamma)$. If ξ and η are identically distributed then $\gamma = \alpha/2$ and therefore

$$\Lambda_{\xi+\eta}(\alpha) = \Lambda_{\xi}\left(\frac{\alpha}{2}\right) + \Lambda_{\eta}\left(\frac{\alpha}{2}\right) = 2\Lambda_{\xi}\left(\frac{\alpha}{2}\right).$$

(A5) *The function $\Lambda(\alpha)$ attains its minimal value 0 at the point $\alpha = \mathbf{E}\xi = m_1$. For definiteness, assume that $\alpha_+ > 0$. If $m_1 = 0$ and $\mathbf{E}|\xi^k| < \infty$, then*

$$\lambda(0) = \Lambda(0) = \Lambda'(0) = 0, \quad \Lambda''(0) = \frac{1}{\gamma_2}, \quad \Lambda'''(0) = -\frac{\gamma_3}{\gamma_2^2}, \quad \dots \quad (9.1.16)$$

(In the case $\alpha_- = 0$ the right derivatives are intended.) As $\alpha \downarrow 0$, one has the representation

$$\Lambda(\alpha) = \sum_{j=2}^k \frac{\Lambda^{(j)}(0)}{j!} \alpha^j + o(\alpha^k). \quad (9.1.17)$$

The semi-invariants γ_j were defined in (9.1.2) and (9.1.3).

If the two-sided Cramér condition is satisfied then the series expansion (9.1.17) of the function $\Lambda(\alpha)$ holds for $k = \infty$. This series is called the *Cramér series*.

Verifying properties (A4) and (A5) is not difficult, and is left to the reader.

(A6) The following inversion formula is valid: for $\lambda \in (\lambda_-, \lambda_+)$,

$$\ln \psi(\lambda) = \sup_{\alpha} (\alpha \lambda - \Lambda(\alpha)). \quad (9.1.18)$$

This means that the rate function uniquely determines the Laplace transform $\psi(\lambda)$ and hence the distribution \mathbf{F} as well. Formula (9.1.18) also means that subsequent double applications of the Legendre transform to the convex function $\ln \psi(\lambda)$ leads to the same original function.

Proof We denote by $T(\lambda)$ the right-hand side of (9.1.18) and show that $T(\lambda) = \ln \psi(\lambda)$ for $\lambda \in (\lambda_-, \lambda_+)$. If, in order to find the supremum in (9.1.18), we equate to zero the derivative in α of the function under the sup sign, then we will get the equation

$$\lambda = \Lambda'(\alpha) = \lambda(\alpha). \quad (9.1.19)$$

Since $\lambda(\alpha)$, $\alpha \in (\alpha_-, \alpha_+)$, is the function inverse to $(\ln \psi(\lambda))'$ (see (9.1.9)), for $\lambda \in (\lambda_-, \lambda_+)$ Eq. (9.1.19) clearly has the solution

$$\alpha = a(\lambda) := (\ln \psi(\lambda))'. \quad (9.1.20)$$

Taking into account the fact that $\lambda(a(\lambda)) \equiv \lambda$, we obtain

$$\begin{aligned} T(\lambda) &= \lambda a(\lambda) - \Lambda(a(\lambda)), \\ T'(\lambda) &= a(\lambda) + \lambda a'(\lambda) - \lambda(a(\lambda)) a'(\lambda) = a(\lambda). \end{aligned}$$

Since $a(0) = m_1$ and $T(0) = -\Lambda(m_1) = 0$, we have

$$T(\lambda) = \int_0^{\lambda} a(u) du = \ln \psi(\lambda). \quad (9.1.21)$$

The assertion is proved, and so is yet another inversion formula (the last equality in (9.1.21), which expresses $\ln \psi(\lambda)$ as the integral of the function $a(\lambda)$ inverse to $\lambda(\alpha)$). \square

(A7) The exponential Chebyshev inequality. For $\alpha \geq m_1$, we have

$$\mathbf{P}(S_n \geq \alpha n) \leq e^{-n\Lambda(\alpha)}.$$

Proof If $\alpha \geq m_1$, then $\lambda(\alpha) \geq 0$. For $\lambda = \lambda(\alpha) \geq 0$, we have

$$\begin{aligned}\psi^n(\lambda) &\geq \mathbf{E}(e^{\lambda S_n}; S_n \geq \alpha n) \geq e^{\lambda \alpha n} \mathbf{P}(S_n \geq \alpha n); \\ \mathbf{P}(S_n \geq \alpha n) &\leq e^{-\alpha n \lambda(\alpha) + n \ln \psi(\lambda(\alpha))} = e^{-n \Lambda(\alpha)}.\end{aligned}\quad \square$$

We now consider a few examples, where the values of λ_{\pm} , α_{\pm} , and the functions $\psi(\lambda)$, $\lambda(\alpha)$, $\Lambda(\alpha)$ can be calculated in an explicit form.

Example 9.1.1 If $\xi \in \Phi_{0,1}$, then

$$\psi(\lambda) = e^{\lambda^2/2}, \quad |\lambda_{\pm}| = |\alpha_{\pm}| = \infty, \quad \lambda(\alpha) = \alpha, \quad \Lambda(\alpha) = \frac{\alpha^2}{2}.$$

Example 9.1.2 For the Bernoulli scheme $\xi \in \mathbf{B}_p$, we have

$$\begin{aligned}\psi(\lambda) &= pe^{\lambda} + q, \quad |\lambda_{\pm}| = \infty, \quad \alpha_+ = 1, \quad \alpha_- = 0, \quad m_1 = \mathbf{E}\xi = p, \\ \lambda(\alpha) &= \ln \frac{\alpha(1-p)}{p(1-\alpha)}, \quad \Lambda(\alpha) = \alpha \ln \frac{\alpha}{p} + (1-\alpha) \ln \frac{1-\alpha}{1-p} \quad \text{for } \alpha \in (0, 1), \\ \Lambda(0) &= -\ln(1-p), \quad \Lambda(1) = -\ln p, \quad \Lambda(\alpha) = \infty \quad \text{for } \alpha \notin [0, 1].\end{aligned}$$

Thus the function $H(\alpha) = \Lambda(\alpha)$, which described large deviation probabilities for S_n in the local Theorem 5.2.1 for the Bernoulli scheme, is nothing else but the rate function. Below, in Sect. 9.3, we will obtain generalisations of Theorem 5.2.1 for arbitrary arithmetic distributions.

Example 9.1.3 For the exponential distribution Γ_{β} , we have

$$\begin{aligned}\psi(\lambda) &= \frac{\beta}{\beta - \lambda}, \quad \lambda_+ = \beta, \quad \lambda_- = -\infty, \quad \alpha_+ = \infty, \quad \alpha_- = 0, \quad m_1 = \frac{1}{\beta}, \\ \lambda(\alpha) &= \beta - \frac{1}{\alpha}, \quad \Lambda(\alpha) = \alpha\beta - 1 - \ln \alpha\beta \quad \text{for } \alpha > 0.\end{aligned}$$

Example 9.1.4 For the centred Poisson distribution with parameter β , we have

$$\begin{aligned}\psi(\lambda) &= \exp\{\beta[e^{\lambda} - 1 - \lambda]\}, \quad |\lambda_{\pm}| = \infty, \quad \alpha_- = -\beta, \quad \alpha_+ = \infty, \quad m_1 = 0, \\ \lambda(\alpha) &= \ln \frac{\beta + \alpha}{\beta}, \quad \Lambda(\alpha) = (\alpha + \beta) \ln \frac{\alpha + \beta}{\beta} - \alpha \quad \text{for } \alpha > -\beta.\end{aligned}$$

9.2 A Relationship Between Large Deviation Probabilities for Sums of Random Variables and Those for Sums of Their Cramér Transforms. The Probabilistic Meaning of the Rate Function

9.2.1 A Relationship Between Large Deviation Probabilities for Sums of Random Variables and Those for Sums of Their Cramér Transforms

Consider the Cramér transform of \mathbf{F} at the point $\lambda = \lambda(\alpha)$ for $\alpha \in [\alpha_-, \alpha_+]$ and introduce the notation $\xi^{(\alpha)} := \xi_{(\lambda(\alpha))}$,

$$S_n^{(\alpha)} := \sum_{i=1}^n \xi_i^{(\alpha)},$$

where $\xi_i^{(\alpha)}$ are independent copies of $\xi^{(\alpha)}$. The distribution $\mathbf{F}^{(\alpha)} := \mathbf{F}_{(\lambda(\alpha))}$ of the random variable $\xi^{(\alpha)}$ is called *the Cramér transform of \mathbf{F} with parameter α* . The random variables $\xi^{(\alpha)}$ are also called Cramér transforms, but of the original random variable ξ . The relationship between the distributions of S_n and $S_n^{(\alpha)}$ is established in the following assertion.

Theorem 9.2.1 *For $x = n\alpha$, $\alpha \in (\alpha_-, \alpha_+)$, and any $t > 0$, one has*

$$\mathbf{P}(S_n \in [x, x + t]) = e^{-n\Lambda(\alpha)} \int_0^t e^{-\lambda(\alpha)z} \mathbf{P}(S_n^{(\alpha)} - \alpha n \in dz). \quad (9.2.1)$$

Proof The Laplace transform of the distribution of the sum $S_n^{(\alpha)}$ is clearly equal to

$$\mathbf{E}e^{\mu S_n^{(\alpha)}} = \left[\frac{\psi(\mu + \lambda(\alpha))}{\psi(\lambda(\alpha))} \right]^n \quad (9.2.2)$$

(see (9.1.5)). On the other hand, consider the Cramér transform $(S_n)_{(\lambda(\alpha))}$ of S_n at the point $\lambda(\alpha)$. Applying (9.1.5) to the distribution of S_n , we obtain

$$\mathbf{E}e^{\mu (S_n)_{(\lambda(\alpha))}} = \frac{\psi^n(\mu + \lambda(\alpha))}{\psi^n(\lambda(\alpha))}.$$

Since this expression coincides with (9.2.2), *the Cramér transform of S_n at the point $\lambda(\alpha)$ coincides in distribution with the sum $S_n^{(\alpha)}$ of the transforms $\xi_i^{(\alpha)}$* . In other words,

$$\frac{\mathbf{P}(S_n \in dv) e^{\lambda(\alpha)v}}{\psi^n(\lambda(\alpha))} = \mathbf{P}(S_n^{(\alpha)} \in dv) \quad (9.2.3)$$

or, which is the same,

$$\mathbf{P}(S_n \in dv) = e^{-\lambda(\alpha)v + n \ln \psi(\lambda(\alpha))} \mathbf{P}(S_n^{(\alpha)} \in dv) = e^{-n\Lambda(\alpha) + \lambda(\alpha)(n\alpha - v)} \mathbf{P}(S_n^{(\alpha)} \in dv).$$

Integrating this equality in v from x to $x + t$, letting $x := n\alpha$ and making the change of variables $v - n\alpha = z$, we get

$$\begin{aligned} \mathbf{P}(S_n \in [x, x + t]) &= e^{-n\Lambda(\alpha)} \int_x^{x+t} e^{\lambda(\alpha)(n\alpha - v)} \mathbf{P}(S_n^{(\alpha)} \in dv) \\ &= e^{-n\Lambda(\alpha)} \int_0^t e^{-\lambda(\alpha)z} \mathbf{P}(S_n^{(\alpha)} - \alpha n \in dz). \end{aligned}$$

The theorem is proved. \square

Since for $\alpha \in [\alpha_-, \alpha_+]$ we have

$$\mathbf{E}\xi^{(\alpha)} = \frac{\psi'(\lambda(\alpha))}{\psi(\lambda(\alpha))} = \alpha$$

(see (9.1.11)), one has $\mathbf{E}(S_n^{(\alpha)} - \alpha n) = 0$ and so for $t \leq c\sqrt{n}$ we have probabilities of normal deviations of $S_n^{(\alpha)} - \alpha n$ on the right-hand side of (9.2.1). This allows us to reduce the problem on large deviations of S_n to the problem on normal deviations of $S_n^{(\alpha)}$. If $\alpha > \alpha_+$, then formula (9.2.1) is still rather useful, as will be shown in Sects. 9.4 and 9.5.

9.2.2 The Probabilistic Meaning of the Rate Function

In this section we will prove the following assertion, which clarifies the probabilistic meaning of the function $\Lambda(\alpha)$.

Denote by $\Delta[\alpha] := [\alpha, \alpha + \Delta)$ the interval of length Δ with the left end at the point α . The notation $\Delta_n[\alpha]$, where Δ_n depends on n , will have a similar meaning.

Theorem 9.2.2 *For each fixed α and all sequences Δ_n converging to 0 as $n \rightarrow \infty$ slowly enough, one has*

$$\Lambda(\alpha) = - \lim_{n \rightarrow \infty} \frac{1}{n} \ln \mathbf{P}\left(\frac{S_n}{n} \in \Delta_n[\alpha]\right). \quad (9.2.4)$$

This relation can also be written as

$$\mathbf{P}\left(\frac{S_n}{n} \in \Delta_n[\alpha]\right) = e^{-n\Lambda(\alpha) + o(n)}.$$

Proof of Theorem 9.2.2 First let $\alpha \in (\alpha_-, \alpha_+)$. Then

$$\mathbf{E}\xi^{(\alpha)} = \alpha, \quad \text{Var}\xi^{(\alpha)} = (\ln \psi(\lambda))''_{\lambda=\lambda(\alpha)} < \infty$$

and hence, as $n \rightarrow \infty$ and $\Delta_n \rightarrow 0$ slowly enough (e.g., for $\Delta_n \geq n^{-1/3}$), by the central limit theorem we have

$$\mathbf{P}(S_n^{(\alpha)} - \alpha n \in [0, \Delta_n n]) \rightarrow 1/2.$$

Therefore, by Theorem 9.2.1 for $t = \Delta_n n$, $x = \alpha n$ and by the mean value theorem,

$$\mathbf{P}(S_n \in [x, x + t]) = \left(\frac{1}{2} + o(1)\right) e^{-n\Lambda(\alpha) - \lambda(\alpha)\Delta_n n\theta}, \quad \theta \in (0, 1);$$

$$\frac{1}{n} \ln \mathbf{P}(S_n \in [x, x + t]) = -\Lambda(\alpha) - \lambda(\alpha)\theta \Delta_n + o(1) = -\Lambda(\alpha) + o(1)$$

as $n \rightarrow \infty$. This proves (9.2.4) for $\alpha \in (\alpha_-, \alpha_+)$.

The further proof is divided into three stages.

(1) *The upper bound in the general case.* Now let α be arbitrary and $|\lambda(\alpha)| < \infty$. By Theorem 9.2.1 for $t = n\Delta_n$, we have

$$\mathbf{P}\left(\frac{S_n}{n} \in \Delta_n[\alpha]\right) \leq \exp\{-n\Lambda(\alpha) + \max(|\lambda(0)|, |\lambda(\alpha)|)n\Delta_n\}.$$

If $\Delta_n \rightarrow 0$ then

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \ln \mathbf{P}\left(\frac{S_n}{n} \in \Delta_n[\alpha]\right) \leq -\Lambda(\alpha). \quad (9.2.5)$$

(This inequality can also be obtained from the exponential Chebyshev's inequality (A7).)

(2) *The lower bound in the general case.* Let $|\lambda(\alpha)| < \infty$ and $|s_{\pm}| = \infty$. Introduce “truncated” random variables $^{(N)}\xi$ with the distribution

$$\mathbf{P}(^{(N)}\xi \in B) = \frac{\mathbf{P}(\xi \in B; |\xi| < N)}{\mathbf{P}(|\xi| < N)} = \mathbf{P}(\xi \in B \mid |\xi| < N)$$

and endow all the symbols that correspond to $^{(N)}\xi$ with the left superscript (N) . Then clearly, for each λ ,

$$\mathbf{E}(e^{\lambda\xi}; |\xi| < N) \uparrow \psi(\lambda), \quad \mathbf{P}(|\xi| < N) \uparrow 1$$

as $N \rightarrow \infty$, so that

$$^{(N)}\psi(\lambda) = \frac{\mathbf{E}(e^{\lambda\xi}; |\xi| < N)}{\mathbf{P}(|\xi| < N)} \rightarrow \psi(\lambda).$$

The functions $^{(N)}\Lambda(\alpha)$ and $\Lambda(\alpha)$ are the upper bounds for the concave functions $\alpha\lambda - \ln^{(N)}\psi(\lambda)$ and $\alpha\lambda - \ln\psi(\lambda)$, respectively. Therefore for each α we also have convergence $^{(N)}\Lambda(\alpha) \rightarrow \Lambda(\alpha)$ as $N \rightarrow \infty$.

Further,

$$\begin{aligned} \mathbf{P}\left(\frac{S_n}{n} \in \Delta_n[\alpha]\right) &\geq \mathbf{P}\left(\frac{S_n}{n} \in \Delta_n[\alpha]; |\xi_j| < N, j = 1, \dots, N\right) \\ &= \mathbf{P}^n(|\xi| < N) \mathbf{P}\left(\frac{^{(N)}S_n}{n} \in \Delta_n[\alpha]\right). \end{aligned}$$

Since $s_{\pm} = \pm\infty$, one has $^{(N)}\alpha_{\pm} = \pm N$ and, for N large enough, we have $\alpha \in (^{(N)}\alpha_-, ^{(N)}\alpha_+)$. Hence we can apply the first part of the proof of the theorem by virtue of which, as $\Delta_n \rightarrow 0$,

$$\begin{aligned} \frac{1}{n} \ln \mathbf{P}\left(\frac{^{(N)}S_n}{n} \in \Delta_n[\alpha]\right) &= -^{(N)}\Lambda(\alpha) + o(1), \\ \frac{1}{n} \ln \mathbf{P}\left(\frac{S_n}{n} \in \Delta_n[\alpha]\right) &\geq -^{(N)}\Lambda(\alpha) + o(1) + \ln \mathbf{P}(|\xi| < N). \end{aligned}$$

The right-hand side of the last inequality can be made arbitrarily close to $-\Lambda(\alpha)$ by choosing a suitable N . Since the left-hand side of this inequality does not depend on N , we have

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \ln \mathbf{P}\left(\frac{S_n}{n} \in \Delta_n[\alpha]\right) \geq -\Lambda(\alpha). \tag{9.2.6}$$

Together with (9.2.5), this proves (9.2.4).

(3) It remains to remove the restrictions stated at the beginning of stages (1) and (2) of the proof, i.e. to consider the cases $|\lambda(\alpha)| = \infty$ and $\min |s_{\pm}| < \infty$. These two relations are connected with each other since, for instance, the equality $\lambda(\alpha) = \lambda_+ = \infty$ can only hold if $\alpha \geq \alpha_+ = s_+ < \infty$ (see property (A2)). For $\alpha > s_+$, relation (9.2.4) is evident, since $\mathbf{P}(S_n/n \in \Delta_n[\alpha]) = 0$ and $\Lambda(\alpha) = \infty$. For $\alpha = \alpha_+ = s_+$ and $p_+ = \mathbf{P}(\xi = s_+)$, we have, for any $\Delta > 0$,

$$\mathbf{P}\left(\frac{S_n}{n} \in \Delta[\alpha_+)\right) = \mathbf{P}(S_n = n\alpha_+) = p_+^n. \tag{9.2.7}$$

Since in this case $\Lambda(\alpha_+) = -\ln p_+$ (see (A2)), the equality (9.2.4) holds true.

The case $\lambda(\alpha) = \lambda_- = -\infty$ with $s_- > -\infty$ is considered in a similar way. However, due to the asymmetry of the interval $\Delta[\alpha]$ with respect to the point α , there are small differences. Instead of an equality in (9.2.7) we only have the inequality

$$\mathbf{P}\left(\frac{S_n}{n} \in \Delta_n[\alpha_-)\right) \geq \mathbf{P}(S_n = n\alpha_-) = p_-^n, \quad p_- = \mathbf{P}(\xi = \alpha_-). \tag{9.2.8}$$

Therefore we also have to use the exponential Chebyshev's inequality (see (A7)) applying it to $-S_n$ for $s_- = \alpha_- < 0$:

$$\mathbf{P}\left(\frac{S_n}{n} \in \Delta_n[\alpha_-]\right) \leq \mathbf{P}\left(\frac{S_n}{n} < \alpha_- + \Delta_n\right) \leq e^{-n\Lambda(\alpha_- + \Delta_n)}. \quad (9.2.9)$$

Relations (9.2.8), (9.2.9), the equality $\Lambda(\alpha_-) = -\ln p_-$, and the right continuity of $\Lambda(\alpha)$ at the point α_- imply (9.2.4) for $\alpha = \alpha_-$. The theorem is proved. \square

9.2.3 The Large Deviations Principle

It is not hard to derive from Theorem 9.2.2 a corollary on the asymptotics of the probabilities of S_n/n hitting an arbitrary Borel set. Denote by (B) and $[B]$ the interior and the closure of B , respectively ((B) is the union of all open intervals contained in B). Put

$$\Lambda(B) := \inf_{\alpha \in B} \Lambda(\alpha).$$

Theorem 9.2.3 *For any Borel set B , the following inequalities hold:*

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \ln \mathbf{P}\left(\frac{S_n}{n} \in B\right) \geq -\Lambda((B)), \quad (9.2.10)$$

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \ln \mathbf{P}\left(\frac{S_n}{n} \in B\right) \leq -\Lambda([B]). \quad (9.2.11)$$

If $\Lambda((B)) = \Lambda([B])$, then the following limit exists:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \ln \mathbf{P}\left(\frac{S_n}{n} \in B\right) = -\Lambda(B). \quad (9.2.12)$$

This assertion is called the *large deviation principle*. It is one of the so-called “rough” (“logarithmic”) limit theorems that describe the asymptotic behaviour of $\ln \mathbf{P}(S_n/n \in B)$. It is usually impossible to derive from this assertion the asymptotics of the probability $\mathbf{P}(S_n/n \in B)$ itself. (In the equality $\mathbf{P}(S_n/n \in B) = \exp\{-n\Lambda(B) + o(n)\}$, the term $o(n)$ may grow in absolute value.)

Proof Without losing generality, we can assume that $B \subset [s_-, s_+]$ (since $\Lambda(\alpha) = \infty$ outside that domain).

We first prove (9.2.10). Let $\alpha_{(B)}$ be such that

$$\Lambda((B)) \equiv \inf_{\alpha \in (B)} \Lambda(\alpha) = \Lambda(\alpha_{(B)})$$

(recall that $\Lambda(\alpha)$ is continuous on $[s_-, s_+]$). Then there exist a sequence of points α_k and a sequence of intervals $(\alpha_k - \delta_k, \alpha_k + \delta_k)$, where $\delta_k \rightarrow 0$, lying in (B) and

converging to the point $\alpha_{(B)}$, such that

$$\Lambda((B)) = \inf_k \Lambda((\alpha_k - \delta_k, \alpha_k + \delta_k)).$$

Here clearly

$$\inf_k \Lambda((\alpha_k - \delta_k, \alpha_k + \delta_k)) = \inf_k \Lambda(\alpha_k),$$

and for a given $\varepsilon > 0$, there exists a $k = K$ such that $\Lambda(\alpha_K) < \Lambda((B)) + \varepsilon$. Since $\Delta_n[\alpha_k] \subset (\alpha_k - \delta_k, \alpha_k + \delta_k)$ for large enough n (here $\Delta_n[\alpha_k]$ is from Theorem 9.2.2), we have by Theorem 9.2.2 that, as $n \rightarrow \infty$,

$$\begin{aligned} \frac{1}{n} \ln \mathbf{P}\left(\frac{S_n}{n} \in B\right) &\geq \frac{1}{n} \ln \mathbf{P}\left(\frac{S_n}{n} \in (B)\right) \\ &\geq \frac{1}{n} \ln \mathbf{P}\left(\frac{S_n}{n} \in (\alpha_K - \delta_K, \alpha_K + \delta_K)\right) \\ &\geq \frac{1}{n} \ln \mathbf{P}\left(\frac{S_n}{n} \in \Delta_n[\alpha_K]\right) \geq -\Lambda(\alpha_K) + o(1) \\ &\geq -\Lambda((B)) - \varepsilon + o(1). \end{aligned}$$

As the left-hand side of this inequality does not depend on ε , inequality (9.2.10) is proved.

We now prove inequality (9.2.11). Denote by $\alpha_{[B]}$ the point at which $\inf_{\alpha \in [B]} \Lambda(\alpha) = \Lambda(\alpha_{[B]})$ is attained (this point always belongs to $[B]$ since $[B]$ is closed). If $\Lambda(\alpha_{[B]}) = 0$, then the inequality is evident. Now let $\Lambda(\alpha_{[B]}) > 0$. By convexity of Λ the equation $\Lambda(\alpha) = \Lambda(\alpha_{[B]})$ can have a second solution $\alpha'_{[B]}$. Assume it exists and, for definiteness, $\alpha'_{[B]} < \alpha_{[B]}$. The relation $\Lambda([B]) = \Lambda(\alpha_{[B]})$ means that the set $[B]$ does not intersect with $(\alpha'_{[B]}, \alpha_{[B]})$ and

$$\mathbf{P}\left(\frac{S_n}{n} \in B\right) \leq \mathbf{P}\left(\frac{S_n}{n} \in [B]\right) \leq \mathbf{P}\left(\frac{S_n}{n} \leq \alpha'_{[B]}\right) + \mathbf{P}\left(\frac{S_n}{n} \geq \alpha_{[B]}\right). \quad (9.2.13)$$

Moreover, in this case $m_1 \in (\alpha'_{[B]}, \alpha_{[B]})$ and each of the probabilities on the right-hand side of (9.2.13) can be bounded using the exponential Chebyshev's inequality (see (A7)) by the value $e^{-n\Lambda(\alpha_{[B]})}$. This implies (9.2.11).

If the second solution $\alpha'_{[B]}$ does not exist, then one of the summands on the right-hand side of (9.2.13) equals zero, and we obtain the same result.

The second assertion of the theorem (Eq. (9.2.12)) is evident.

The theorem is proved. \square

Using Theorem 9.2.3, we can complement Theorem 9.2.2 with the following assertion.

Corollary 9.2.1 *The following limit always exists*

$$\lim_{\Delta \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n} \ln \mathbf{P} \left(\frac{S_n}{n} \in \Delta[\alpha] \right) = -\Lambda(\alpha). \quad (9.2.14)$$

Proof Take the set B in Theorem 9.2.3 to be the interval $B = \Delta[\alpha]$. If $\alpha \notin [s_-, s_+]$ then the assertion is obvious (since both sides of (9.2.14) are equal to $-\infty$). If $\alpha = s_{\pm}$ then (9.2.14) is already proved in (9.2.7), (9.2.8) and (9.2.9).

It remains to consider points $\alpha \in (s_-, s_+)$. For such α , the function $\Lambda(\alpha)$ is continuous and $\alpha + \Delta$ is also a point of continuity of Λ for Δ small enough, and hence

$$\Lambda((B)) = \Lambda([B]) \rightarrow \Lambda(\alpha)$$

as $\Delta \rightarrow 0$. Therefore by Theorem 9.2.3 the inner limit in (9.2.14) exists and converges to $-\Lambda(\alpha)$ as $\Delta \rightarrow 0$.

The corollary is proved. \square

Note that the assertions of Theorems 9.2.2 and 9.2.3 and their corollaries are “universal”—they contain no restrictions on the distribution \mathbf{F} .

9.3 Integro-Local, Integral and Local Theorems on Large Deviation Probabilities in the Cramér Range

9.3.1 Integro-Local and Integral Theorems

In this subsection, under the assumption that the Cramér condition $\lambda_+ > 0$ is met, we will find the asymptotics of probabilities $\mathbf{P}(S_n \in \Delta[x])$ for scaled deviations $\alpha = x/n$ from the so-called *Cramér* (or *regular*) *range*, i.e. for the range $\alpha \in (\alpha_-, \alpha_+)$ in which the rate function $\Lambda(\alpha)$ is analytic.

In the non-lattice case, in addition to the condition $\lambda_+ > 0$, we will assume without loss of generality that $\mathbf{E}\xi = 0$. In this case necessarily

$$\alpha_- \leq 0, \quad \alpha_+ = \frac{\psi'(\lambda_+)}{\psi(\lambda_+)} > 0, \quad \lambda(0) = 0.$$

The length Δ of the interval may depend on n in some cases. In such cases, we will write Δ_n instead of Δ , as we did earlier. The value

$$\sigma_{\alpha}^2 = \frac{\psi''(\lambda(\alpha))}{\psi(\lambda(\alpha))} - \alpha^2 \quad (9.3.1)$$

is clearly equal to $\text{Var}(\xi^{(\alpha)})$ (see (9.1.5) and the definition of $\xi^{(\alpha)}$ in Sect. 9.2).

Theorem 9.3.1 *Let $\lambda_+ > 0$, $\alpha \in [0, \alpha_+)$, ξ be a non-lattice random variable, $\mathbf{E}\xi = 0$ and $\mathbf{E}\xi^2 < \infty$. If $\Delta_n \rightarrow 0$ slowly enough as $n \rightarrow \infty$, then*

$$\mathbf{P}(S_n \in \Delta_n[x]) = \frac{\Delta_n}{\sigma_\alpha \sqrt{2\pi n}} e^{-n\Lambda(\alpha)} (1 + o(1)), \quad (9.3.2)$$

where $\alpha = x/n$, and, for each fixed $\alpha_1 \in (0, \alpha_+)$, the remainder term $o(1)$ is uniform in $\alpha \in [0, \alpha_1]$ for any fixed $\alpha_1 \in (0, \alpha_+)$.

A similar assertion is valid in the case when $\lambda_- < 0$ and $\alpha \in (\alpha_-, 0]$.

Proof The proof is based on Theorems 9.2.1 and 8.7.1A. Since the conditions of Theorem 9.2.1 are satisfied, we have

$$\mathbf{P}(S_n \in \Delta_n[x]) = e^{-n\Lambda(\alpha)} \int_0^{\Delta_n} e^{-\lambda(\alpha)z} \mathbf{P}(S_n^{(\alpha)} - \alpha n \in dz).$$

As $\lambda(\alpha) \leq \lambda(\alpha_+ - \varepsilon) < \infty$ and $\Delta_n \rightarrow 0$, one has $e^{-\lambda(\alpha)z} \rightarrow 1$ uniformly in $z \in \Delta_n[0]$ and hence, as $n \rightarrow \infty$,

$$\mathbf{P}(S_n \in \Delta_n[x]) = e^{-n\Lambda(\alpha)} \mathbf{P}(S_n^{(\alpha)} - \alpha n \in \Delta_n[0]) (1 + o(1)) \quad (9.3.3)$$

uniformly in $\alpha \in [0, \alpha_+ - \varepsilon]$.

We now show that Theorem 8.7.1A is applicable to the random variables $\xi^{(\alpha)} = \xi_{(\lambda(\alpha))}$. That $\sigma_\alpha = \sigma(\lambda(\alpha))$ is bounded away from 0 and from ∞ for $\alpha \in [0, \alpha_1]$ is evident. (The same is true of all the theorems in this section.) Therefore, it remains to verify whether conditions (a) and (b) of Theorem 8.7.1A are met for $\lambda = \lambda(\alpha) \in [0, \lambda_1]$, $\lambda_1 := \lambda(\alpha_1) < \lambda_+$ and $\varphi_{(\lambda)}(t) = \frac{\psi(\lambda+it)}{\psi(\lambda)}$ (see (9.1.5)). We have

$$\psi(\lambda + it) = \psi(\lambda) + it\psi'(\lambda) - \frac{t^2}{2}\psi''(\lambda) + o(t^2)$$

as $t \rightarrow 0$, where the remainder term is uniform in λ if the function $\psi''(\lambda + iu)$ is uniformly continuous in u . The required uniform continuity can easily be proved by imitating the corresponding result for ch.f.s (see property 4 in Sect. 7.1). This proves condition (a) in Theorem 8.7.1A with

$$a(\lambda) = \frac{\psi'(\lambda)}{\psi(\lambda)}, \quad m_2(\lambda) = \frac{\psi''(\lambda)}{\psi(\lambda)}.$$

Now we will verify condition (b) in Theorem 8.7.1A. Assume the contrary: there exists a sequence $\lambda_k \in [0, \lambda_1]$ such that

$$q_{\lambda_k} := \sup_{\theta_1 \leq |t| \leq \theta_2} \frac{|\psi(\lambda_k + it)|}{\psi(\lambda_k)} \rightarrow 1$$

as $k \rightarrow \infty$. By the uniform continuity of ψ in that domain, there exist points $t_k \in [\theta_1, \theta_2]$ such that, as $k \rightarrow \infty$,

$$\frac{\psi(\lambda_k + it_k)}{\psi(\lambda_k)} \rightarrow 1.$$

Since the region $\lambda \in [0, \lambda_1]$, $|t| \in [\theta_1, \theta_2]$ is compact, there exists a subsequence $(\lambda_{k'}, t_{k'}) \rightarrow (\lambda_0, t_0)$ as $k' \rightarrow \infty$. Again using the continuity of ψ , we obtain the equality

$$\frac{|\psi(\lambda_0 + it_0)|}{\psi(\lambda_0)} = 1, \quad (9.3.4)$$

which contradicts the non-latticeness of $\xi_{(\lambda_0)}$. Property (b) is proved.

Thus we can now apply Theorem 8.7.1A to the probability on the right-hand side of (9.3.3). Since $\mathbf{E}\xi^{(\alpha)} = \alpha$ and $\mathbf{E}(\xi^{(\alpha)})^2 = \frac{\psi''(\lambda(\alpha))}{\psi(\lambda(\alpha))}$, this yields

$$\begin{aligned} \mathbf{P}(S_n \in \Delta_n[x]) &= e^{-n\Lambda(\alpha)} \left(\frac{\Delta_n}{\sigma_\alpha \sqrt{n}} \phi(0) + o\left(\frac{1}{\sqrt{n}}\right) \right) \\ &= \frac{\Delta_n}{\sigma_\alpha \sqrt{2\pi n}} e^{-n\Lambda(\alpha)} (1 + o(1)) \end{aligned} \quad (9.3.5)$$

uniformly in $\alpha \in [0, \alpha_1]$ (or in $x \in [0, \alpha_1 n]$), where the values of

$$\sigma_\alpha^2 = \mathbf{E}(\xi^{(\alpha)} - \alpha)^2 = \frac{\psi''(\lambda(\alpha))}{\psi(\lambda(\alpha))} - \alpha^2$$

are bounded away from 0 and from ∞ . The theorem is proved. \square

From Theorem 9.3.1 we can now derive integro-local theorems and integral theorems for fixed or growing Δ . Since in the normal deviation range (when x is comparable with \sqrt{n}) we have already obtained such results, to simplify the exposition we will consider here large deviations only, when $x \gg \sqrt{n}$ or, which is the same, $\alpha = x/n \gg 1/\sqrt{n}$. To be more precise, we will assume that there exists a function $N(n) \rightarrow \infty$, $N(n) = o(\sqrt{n})$ as $n \rightarrow \infty$, such that $x \geq N(n)\sqrt{n}$ ($\alpha \geq N(n)/\sqrt{n}$).

Theorem 9.3.2 *Let $\lambda_+ > 0$, $\alpha \in [0, \alpha_+)$, ξ be non-lattice, $\mathbf{E}\xi = 0$ and $\mathbf{E}\xi^2 < \infty$. Then, for any $\Delta \geq \Delta_0 > 0$, $x \geq N(n) = o(\sqrt{n})$, $N(n) \rightarrow \infty$ as $n \rightarrow \infty$, one has*

$$\mathbf{P}(S_n \in \Delta[x]) = \frac{e^{-n\Lambda(\alpha)}}{\sigma_\alpha \lambda(\alpha) \sqrt{2\pi n}} (1 - e^{-\lambda(\alpha)\Delta}) (1 + o(1)), \quad (9.3.6)$$

$o(1)$ being uniform in $\alpha = x/n \in [N(n)/\sqrt{n}, \alpha_1]$ and $\Delta \geq \Delta_0$ for each fixed $\alpha_1 \in (0, \alpha_+)$.

In particular (for $\Delta = \infty$),

$$\mathbf{P}(S_n \geq x) = \frac{e^{-n\Lambda(\alpha)}}{\sigma_\alpha \lambda(\alpha) \sqrt{2\pi n}} (1 + o(1)). \quad (9.3.7)$$

Proof Partition the interval $\Delta[x]$ into subintervals $\Delta_n[x + k\Delta_n]$, $k = 0, \dots, \Delta/\Delta_n - 1$, where $\Delta_n \rightarrow 0$ and, for simplicity, we assume that $M = \Delta/\Delta_n$ is an integer. Then, by Theorem 9.2.1, as $\Delta_n \rightarrow 0$,

$$\begin{aligned} & \mathbf{P}(S_n \in \Delta_n[x + k\Delta_n]) \\ &= \mathbf{P}(S_n \in [x, x + (k+1)\Delta_n]) - \mathbf{P}(S_n \in [x, x + k\Delta_n]) \\ &= e^{-n\Lambda(\alpha)} \int_{k\Delta_n}^{(k+1)\Delta_n} e^{-\lambda(\alpha)z} \mathbf{P}(S_n^{(\alpha)} - \alpha n \in dz) \\ &= e^{-n\Lambda(\alpha) - \lambda(\alpha)k\Delta_n} \mathbf{P}(S_n^{(\alpha)} - \alpha n \in \Delta_n[k\Delta_n]) (1 + o(1)) \end{aligned} \quad (9.3.8)$$

uniformly in $\alpha \in [0, \alpha_1]$. Here, similarly to (9.3.5), by Theorem 8.7.1A we have

$$\mathbf{P}(S_n^{(\alpha)} - \alpha n \in \Delta_n[k\Delta_n]) = \frac{\Delta_n}{\sigma_\alpha \sqrt{n}} \phi\left(\frac{k\Delta_n}{\sigma_\alpha \sqrt{n}}\right) + o\left(\frac{1}{\sqrt{n}}\right) \quad (9.3.9)$$

uniformly in k and α . Since

$$\mathbf{P}(S_n \in \Delta[x]) = \sum_{k=0}^{M-1} \mathbf{P}(S_n \in \Delta_n[x + k\Delta]),$$

substituting the values (9.3.8) and (9.3.9) into the right-hand side of the last equality, we obtain

$$\begin{aligned} \mathbf{P}(S_n \in \Delta[x]) &= \frac{e^{-n\Lambda(\alpha)}}{\sigma_\alpha \sqrt{n}} \sum_{k=0}^{M-1} \Delta_n e^{-\lambda(\alpha)k\Delta_n} \left(\phi\left(\frac{k\Delta_n}{\sigma_\alpha \sqrt{n}}\right) + o(1) \right) \\ &= \frac{e^{-n\Lambda(\alpha)}}{\sigma_\alpha \sqrt{n}} \int_0^{\Delta - \Delta_n} e^{-\lambda(\alpha)z} \left(\phi\left(\frac{z}{\sigma_\alpha \sqrt{n}}\right) + o(1) \right) dz. \end{aligned} \quad (9.3.10)$$

After the variable change $\lambda(\alpha)z = u$, the right-hand side can be rewritten as

$$\frac{e^{-n\lambda(\alpha)}}{\sigma_\alpha \lambda(\alpha) \sqrt{n}} \int_0^{(\Delta - \Delta_n)\lambda(\alpha)} e^{-u} \left(\phi\left(\frac{u}{\sigma_\alpha \lambda(\alpha) \sqrt{n}}\right) + o(1) \right) du, \quad (9.3.11)$$

where the remainder term $o(1)$ is uniform in $\alpha \in [0, \alpha_1]$, $\Delta \geq \Delta_0$, and u from the integration range. Since $\lambda(\alpha) \sim \alpha/\sigma^2$ for small α (see (9.1.12) and (9.1.16)), for $\alpha \geq N(n)/\sqrt{n}$ we have

$$\lambda(\alpha) > \frac{N(n)}{\sigma^2 \sqrt{n}} (1 + o(1)), \quad \sigma_\alpha \lambda(\alpha) \sqrt{n} > \frac{\sigma_\alpha N(n)}{\sigma^2} \rightarrow \infty.$$

Therefore, for any fixed u , one has

$$\phi\left(\frac{u}{\sigma_\alpha \lambda(\alpha) \sqrt{n}}\right) \rightarrow \phi(0) = \frac{1}{\sqrt{2\pi}}.$$

Moreover, $\phi(v) \leq 1/\sqrt{2\pi}$ for all v . Hence, by (9.3.10) and (9.3.11),

$$\begin{aligned} \mathbf{P}(S_n \in \Delta[x]) &= \frac{e^{-n\Lambda(\alpha)}}{\sigma_\alpha \lambda(\alpha) \sqrt{2\pi n}} \int_0^{\lambda(\alpha)\Delta} e^{-u} du (1 + o(1)) \\ &= \frac{e^{-n\Lambda(\alpha)}}{\sigma_\alpha \lambda(\alpha) \sqrt{2\pi n}} (1 - e^{-\lambda(\alpha)\Delta}) (1 + o(1)) \end{aligned}$$

uniformly in $\alpha \in [0, \alpha_1]$ and $\Delta \geq \Delta_0$. Relation (9.3.7) clearly follows from (9.3.6) with $\Delta = \infty$. The theorem is proved. \square

Note that if $\mathbf{E}|\xi|^k < \infty$ (for $\lambda_+ > 0$ this is a restriction on the rate of decay of the left tails $\mathbf{P}(\xi < -t)$, $t > 0$), then expansion (9.1.17) is valid and, for deviations $x = o(n)$ ($\alpha = o(1)$) such that $n\alpha^k = x^k/n^{k-1} \leq c = \text{const}$, we can change the exponent $n\Lambda(\alpha)$ in (9.3.6) and (9.3.7) to

$$n\Lambda(\alpha) = n \sum_{j=2}^k \frac{\Lambda^{(j)}(0)}{j!} \alpha^j + o(n\alpha^k), \tag{9.3.12}$$

where $\Lambda^{(j)}(0)$ are found in (9.1.16). For $k = 3$, the foregoing implies the following.

Corollary 9.3.1 *Let $\lambda_+ > 0$, $\mathbf{E}|\xi|^3 < \infty$, ξ be non-lattice, $\mathbf{E}\xi = 0$, $\mathbf{E}\xi^2 = \sigma^2$, $x \gg \sqrt{n}$ and $x = o(n^{2/3})$ as $n \rightarrow \infty$. Then*

$$\mathbf{P}(S_n \geq x) \sim \frac{\sigma\sqrt{n}}{x\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2n\sigma^2}\right\} \sim \Phi\left(-\frac{x}{\sigma\sqrt{n}}\right). \tag{9.3.13}$$

In the last relation we used the symmetry of the standard normal law, i.e. the equality $1 - \Phi(t) = \Phi(-t)$. Assertion (9.3.13) shows that in the case $\lambda_+ > 0$ and $\mathbf{E}|\xi|^3 < \infty$ the asymptotic equivalence

$$\mathbf{P}(S_n \geq x) \sim \Phi\left(-\frac{x}{\sigma\sqrt{n}}\right)$$

persists outside the range of normal deviations as well, up to the values $x = o(n^{2/3})$. If $\mathbf{E}\xi^3 = 0$ and $\mathbf{E}\xi^4 < \infty$, then this equivalence holds true up to the values $x = o(n^{3/4})$. For larger x this equivalence, generally speaking, no longer holds.

Proof of Corollary 9.3.1 The first relation in (9.3.13) follows from Theorem 9.3.2 and (9.3.12). The second follows from the asymptotic equivalence

$$\int_x^\infty e^{-\frac{u^2}{2}} du \sim \frac{e^{-x^2/2}}{x},$$

which is easy to establish, using, for example, l'Hospital's rule. \square

9.3.2 Local Theorems

In this subsection we will obtain analogues of the local Theorems 8.7.2 and 8.7.3 for large deviations in the Cramér range. To simplify the exposition, we will formulate the theorem for densities, assuming that the following condition is satisfied:

[D] *The distribution \mathbf{F} has a bounded density $f(x)$ such that*

$$f(x) = e^{-\lambda_+ x + o(x)} \quad \text{as } x \rightarrow \infty, \text{ if } \lambda_+ < \infty; \tag{9.3.14}$$

$$f(x) \leq c e^{-\lambda x} \quad \text{for any fixed } \lambda > 0, \ c = c(\lambda), \text{ if } \lambda_+ = \infty. \tag{9.3.15}$$

Since inequalities of the form (9.3.14) and (9.3.15) always hold, by the exponential Chebyshev inequality, for the right tails

$$F_+(x) = \int_x^\infty f(u) du,$$

condition **[D]** is not too restrictive. It only eliminates sharp “bursts” of $f(x)$ as $x \rightarrow \infty$.

Denote by $f_n(x)$ the density of the distribution of S_n .

Theorem 9.3.3 *Let*

$$\mathbf{E}\xi = 0, \quad \mathbf{E}\xi^2 < \infty, \quad \lambda_+ > 0, \quad \alpha = \frac{x}{n} \in [0, \alpha_+),$$

and condition [D] be met. Then

$$f_n(x) = \frac{e^{-n\Lambda(\alpha)}}{\sigma_\alpha \sqrt{2\pi n}} (1 + o(1)),$$

where the remainder term $o(1)$ is uniform in $\alpha \in [0, \alpha_1]$ for any fixed $\alpha_1 \in (0, \alpha_+)$.

Proof The proof is based on Theorems 9.2.1 and 8.7.2A. Denote by $f_n^{(\alpha)}(x)$ the density of the distribution of $S_n^{(\alpha)}$. Relation (9.2.3) implies that, for $x = \alpha n$, $\alpha \in [\alpha_-, \alpha_+]$, we have

$$f_n(x) = e^{-\lambda(\alpha)x} \psi^n(\lambda(\alpha)) f_n^{(\alpha)}(x) = e^{-n\Lambda(\alpha)} f_n^{(\alpha)}(x). \tag{9.3.16}$$

Since $\mathbf{E}\xi^{(\alpha)} = \alpha$, we see that $\mathbf{E}(S_n^{(\alpha)} - x) = 0$ and the density value $f_n^{(\alpha)}(x)$ coincides with the density of the distribution of the sum $S_n^{(\alpha)} - \alpha n$ at the point 0. In order to use Theorems 8.7.1A and 8.7.2A, we have to verify conditions (a) and (b) for $\theta_2 = \infty$ in these theorems and also the uniform boundedness in $\alpha \in [0, \alpha_1]$ of

$$\int |\varphi_{(\lambda(\alpha))}(t)|^m dt \tag{9.3.17}$$

for some integer $m \geq 1$, where $\varphi_{(\lambda(\alpha))}$ is the ch.f. of $\xi^{(\alpha)}$ (the uniform version of condition (c) in Theorem 8.7.2). By condition **[D]** the density

$$f^{(\alpha)}(v) = \frac{e^{\lambda(\alpha)v} f(v)}{\psi(\lambda(\alpha))}$$

is bounded uniformly in $\alpha \in [0, \alpha_1]$ (for such α one has $\lambda(\alpha) \in [0, \lambda_1]$, $\lambda_1 = \lambda(\alpha_1) < \lambda_+$). Hence the integral

$$\int (f^{(\alpha)}(v))^2 dv$$

is also uniformly bounded, and so, by virtue of Parseval's identity (see Sect. 7.2), is the integral

$$\int |\varphi_{(\lambda(\alpha))}(t)|^2 dt.$$

This means that the required uniform boundedness of integral (9.3.17) is proved for $m = 2$.

Conditions (a) and (b) for $\theta_2 < \infty$ were verified in the proof of Theorem 9.3.1. It remains to extend the verification of condition (b) to the case $\theta_2 = \infty$. This can be done by following an argument very similar to the one used in the proof of Theorem 9.3.1 in the case of finite θ_2 . Let $\theta_2 = \infty$. If we assume that there exist sequences $\lambda_k \in [0, \lambda_{+,\varepsilon}]$ and $|t_k| \geq \theta_1$ such that

$$\frac{|\psi(\lambda_k + it_k)|}{\psi(\lambda_k)} \rightarrow 1,$$

then, by compactness of $[0, \lambda_{+,\varepsilon}]$, there will exist sequences $\lambda'_k \rightarrow \lambda_0 \in [0, \lambda_{+,\varepsilon}]$ and t'_k such that

$$\frac{|\psi(\lambda'_k + it'_k)|}{\psi(\lambda_0)} \rightarrow 1. \tag{9.3.18}$$

But by virtue of condition **[D]** the family of functions $\psi(\lambda + it)$, $t \in \mathbb{R}$, is equicontinuous in $\lambda \in [0, \lambda_{+,\varepsilon}]$. Therefore, along with (9.3.18), we also have convergence

$$\frac{|\psi(\lambda_0 + it'_k)|}{\psi(\lambda_0)} \rightarrow 1, \quad |t'_k| \geq \theta_1 > 0,$$

which contradicts the inequality

$$\sup_{|t| \geq \theta_1} \frac{|\psi(\lambda_0 + it)|}{\psi(\lambda_0)} < 1$$

that follows from the existence of density.

Thus property (b) is proved for $\theta_2 = \infty$, and we can use Theorem 8.7.2A, which implies that

$$f_n^{(\alpha)}(x) = \frac{1}{\sigma(\lambda(\alpha))\sqrt{2\pi n}} (1 + o(1)).$$

This, together with (9.3.16), proves Theorem 9.3.3. \square

Remark 9.3.1 We can see from the proof that, in Theorem 9.3.3, as a more general condition instead of condition [D] one could also consider the integrability of $\psi^m(\lambda + it)$ for any fixed $\lambda \in [0, \lambda_1]$, $\lambda_1 < \lambda_+$, or condition [D] imposed on S_m for some $m \geq 1$.

For arithmetic distributions we cannot assume without loss of generality that $m_1 = \mathbf{E}\xi = 0$, but that does not change much in the formulations of the assertions. If $\lambda_+ > 0$, then $\alpha_+ = \psi'(\lambda_+)/\psi(\lambda_+) > m_1$ and the scaled deviations $\alpha = x/n$ for the Cramér range must lie in the region $[m_1, \alpha_+)$.

Theorem 9.3.4 *Let $\lambda_+ > 0$, $\mathbf{E}\xi^2 < \infty$ and the distribution of ξ be arithmetic. Then, for integer x ,*

$$\mathbf{P}(S_n = x) = \frac{e^{-n\Lambda(\alpha)}}{\sigma_\alpha\sqrt{2\pi n}} (1 + o(1)),$$

where the remainder term $o(1)$ is uniform in $\alpha = x/n \in [m_1, \alpha_1]$ for any fixed $\alpha_1 \in (m_1, \alpha_+)$.

A similar assertion is valid in the case when $\lambda_- < 0$ and $\alpha \in (\alpha_-, m_1]$.

Proof The proof does not differ much from that of Theorem 9.3.1. By (9.2.3),

$$\mathbf{P}(S_n = x) = e^{-\lambda(\alpha)x} \psi^{-n}(\lambda(\alpha)) \mathbf{P}(S_n^{(\alpha)} = x) = e^{-n\Lambda(\alpha)} \mathbf{P}(S_n^{(\alpha)} = x),$$

where $\mathbf{E}\xi^{(\alpha)} = \alpha$ for $\alpha \in [m_1, \alpha_+)$. In order to compute $\mathbf{P}(S_n^{(\alpha)} = x)$ we have to use Theorem 8.7.3A. The verification of conditions (a) and (b) of Theorem 8.7.1A, which are assumed to hold in Theorem 8.7.3A, is done in the same way as in the proof of Theorem 9.3.1, the only difference being that relation (9.3.4) for $t_0 \in [\theta_1, \pi]$ will contradict the arithmeticity of the distribution of ξ . Since $a(\lambda(\alpha)) = \mathbf{E}\xi^{(\alpha)} = \alpha$, by Theorem 8.7.3A we have

$$\mathbf{P}(S_n^{(\alpha)} = x) = \frac{1}{\sigma_\alpha\sqrt{2\pi n}} (1 + o(1))$$

uniformly in $\alpha = x/n \in [m_1, \alpha_1]$. The theorem is proved. \square

9.4 Integro-Local Theorems at the Boundary of the Cramér Range

9.4.1 Introduction

In this section we again assume that Cramér’s condition $\lambda_+ > 0$ is met. If $\alpha_+ = \infty$ then the theorems of Sect. 9.3 describe the large deviation probabilities for any $\alpha = x/n$. But if $\alpha_+ < \infty$ then the approaches of Sect. 9.3 do not enable one to find the asymptotics of probabilities of large deviations of S_n for scaled deviations $\alpha = x/n$ in the vicinity of the point α_+ .

In this section we consider the case $\alpha_+ < \infty$. If in this case $\lambda_+ = \infty$, then, by property (A2)(i), we have $\alpha_+ = s_+ = \sup\{t : F_+(t) > 0\}$, and therefore the random variables ξ_k are bounded from above by the value α_+ , $\mathbf{P}(S_n \geq x) = 0$ for $\alpha = x/n > \alpha_+$. We will not consider this case in what follows. Thus we will study the case $\alpha_+ < \infty, \lambda_+ < \infty$.

In the present and the next sections, we will confine ourselves to considering integro-local theorems in the non-lattice case with $\Delta = \Delta_n \rightarrow 0$ since, as we saw in the previous section, local theorems differ from the integro-local theorems only in that they are simpler. As in Sect. 9.3, the integral theorems can be easily obtained from the integro-local theorems.

9.4.2 The Probabilities of Large Deviations of S_n in an $o(n)$ -Vicinity of the Point α_+ ; the Case $\psi''(\lambda_+) < \infty$

In this subsection we will study the asymptotics of $\mathbf{P}(S_n \in \Delta[x]), x = \alpha n$, when α lies in the vicinity of the point $\alpha_+ < \infty$ and, moreover, $\psi''(\lambda_+) < \infty$. (The case of distributions \mathbf{F} , for which $\lambda_+ < \infty, \alpha_+ < \infty$ and $\psi''(\lambda_+) < \infty$, will be illustrated later, in Lemma 9.4.1.) Under the above-mentioned conditions, the Cramér transform $\mathbf{F}_{(\lambda_+)}$ is well defined at the point λ_+ , and the random variable $\xi^{(\alpha_+)}$ with the distribution $\mathbf{F}_{(\lambda_+)}$ has mean α_+ and a finite variance:

$$\mathbf{E}\xi^{(\alpha_+)} = \frac{\psi'(\lambda_+)}{\psi(\lambda_+)} = \alpha_+, \quad \text{Var}(\xi^{(\alpha_+)}) = \sigma_{\alpha_+}^2 = \frac{\psi''(\lambda_+)}{\psi(\lambda_+)} - \alpha_+^2 \quad (9.4.1)$$

(cf. (9.3.1)).

Theorem 9.4.1 *Let ξ be a non-lattice random variable,*

$$\lambda_+ \in (0, \infty), \quad \psi''(\lambda_+) < \infty, \quad y = x - \alpha_+ n = o(n).$$

If $\Delta_n \rightarrow 0$ slowly enough as $n \rightarrow \infty$ then

$$\mathbf{P}(S_n \in \Delta_n[x]) = \frac{\Delta_n}{\sigma_{\alpha_+} \sqrt{2\pi n}} e^{-n\Lambda(\alpha_+) - \lambda_+ y} \left(\exp\left\{ -\frac{y^2}{\sigma_{\alpha_+}^2 n} \right\} + o(1) \right),$$

where

$$\alpha = \frac{x}{n}, \quad \sigma_{\alpha_+}^2 = \frac{\psi''(\lambda_+)}{\psi(\lambda_+)} - \alpha_+^2,$$

and the remainder term $o(1)$ is uniform in y .

Proof As in the proof of Theorem 9.3.1, we use the Cramér transform, but now at the fixed point λ_+ , so there will be no triangular array scheme when analysing the sums $S_n^{(\alpha_+)}$. In this case the following analogue of Theorem 9.2.1 holds true.

Theorem 9.2.1A *Let $\lambda_+ \in (0, \infty)$, $\alpha_+ < \infty$ and $y = x - n\alpha_+$. Then, for $x = n\alpha$ and any fixed $\Delta > 0$, the following representation is valid:*

$$\mathbf{P}(S_n \in \Delta[x]) = e^{-n\Lambda(\alpha_+) - \lambda_+ y} \int_0^\Delta e^{-\lambda_+ z} \mathbf{P}(S_n^{(\alpha_+)} - \alpha n \in dz). \quad (9.4.2)$$

Proof of Theorem 9.2.1A repeats that of Theorem 9.2.1 the only difference being that, as was already noted, the Cramér transform is now applied at the fixed point λ_+ which does not depend on $\alpha = x/n$. In this case, by (9.2.3),

$$\mathbf{P}(S_n \in dv) = e^{-\lambda_+ v + n \ln \psi(\lambda_+)} \mathbf{P}(S_n^{(\alpha_+)} \in dv) = e^{-n\Lambda(\alpha_+) + \lambda_+ (\alpha_+ n - v)} \mathbf{P}(S_n^{(\alpha_+)} \in dv).$$

Integrating this equality in v from x to $x + \Delta$, changing the variable $v = x + z$ ($x = n\alpha$), and noting that $\alpha_+ n - v = -y - z$, we obtain (9.4.2).

The theorem is proved. □

Let us return to the proof of Theorem 9.4.1. Assuming that $\Delta = \Delta_n \rightarrow 0$, we obtain, by Theorem 9.2.1A, that

$$\mathbf{P}(S_n \in \Delta_n[x]) = e^{-n\Lambda(\alpha_+) - \lambda_+ y} \mathbf{P}(S_n^{(\alpha_+)} - \alpha_+ n \in \Delta_n[y])(1 + o(1)). \quad (9.4.3)$$

By virtue of (9.4.1), we can apply Theorem 8.7.1 to evaluate the probability on the right-hand side of (9.4.3). This theorem implies that, as $\Delta_n \rightarrow 0$ slowly enough,

$$\begin{aligned} \mathbf{P}(S_n^{(\alpha_+)} - \alpha_+ n \in \Delta_n[y]) &= \frac{\Delta_n}{\sigma_{\alpha_+} \sqrt{n}} \phi\left(\frac{y}{\sigma_{\alpha_+} \sqrt{n}}\right) + o\left(\frac{1}{\sqrt{n}}\right) \\ &= \frac{\Delta_n}{\sigma_{\alpha_+} \sqrt{2\pi n}} \exp\left\{-\frac{y^2}{\sigma_{\alpha_+}^2 n}\right\} + o\left(\frac{1}{\sqrt{n}}\right) \end{aligned}$$

uniformly in y . This, together with (9.4.3), proves Theorem 9.4.1. □

9.4.3 The Class of Distributions \mathcal{ER} . The Probability of Large Deviations of S_n in an $o(n)$ -Vicinity of the Point α_+n for Distributions \mathbf{F} from the Class \mathcal{ER} in Case $\psi''(\lambda_+) = \infty$

When studying the asymptotics of $\mathbf{P}(S_n \geq \alpha n)$ (or $\mathbf{P}(S_n \in \Delta[\alpha n])$) in the case where $\psi''(\lambda_+) = \infty$ and α is in the vicinity of the point $\alpha_+ < \infty$, we have to impose additional conditions on the distribution \mathbf{F} similarly to what was done in Sect. 8.8 when studying convergence to stable laws.

To formulate these additional conditions it will be convenient to introduce certain classes of distributions. If $\lambda_+ < \infty$, then it is natural to represent the right tails $F_+(t)$ as

$$F_+(t) = e^{-\lambda_+ t} V(t), \quad (9.4.4)$$

where, by the exponential Chebyshev inequality, $V(t) = e^{o(t)}$ as $t \rightarrow \infty$.

Definition 9.4.1 We will say that the distribution \mathbf{F} of a random variable ξ (or the random variable ξ itself) belongs to the class \mathcal{R} if its right tail $F_+(t)$ is a regularly varying function, i.e. can be represented as

$$F_+(t) = t^{-\beta} L(t), \quad (9.4.5)$$

where L is a slowly varying function as $t \rightarrow \infty$ (see also Sect. 8.8 and Appendix 6).

We will say that the distribution \mathbf{F} (or the random variable ξ) belongs to the class \mathcal{ER} if, in the representation (9.4.4), the function V is regularly varying (which will also be denoted as $V \in \mathcal{R}$).

Distributions from the class \mathcal{R} have already appeared in Sect. 8.8.

The following assertion explains which distributions from \mathcal{ER} correspond to the cases $\alpha_+ = \infty$, $\alpha_+ < \infty$, $\psi''(\lambda_+) = \infty$ and $\psi''(\lambda_+) < \infty$.

Lemma 9.4.1 *Let $\mathbf{F} \in \mathcal{ER}$. For α_+ to be finite it is necessary and sufficient that*

$$\int_1^\infty t V(t) dt < \infty.$$

For $\psi''(\lambda_+)$ to be finite, it is necessary and sufficient that

$$\int_1^\infty t^2 V(t) dt < \infty.$$

The assertion of the lemma means that $\alpha_+ < \infty$ if $\beta > 2$ in the representation $V(t) = t^{-\beta} L(t)$, where L is an s.v.f. and $\alpha_+ = \infty$ if $\beta < 2$. For $\beta = 2$, the finiteness of α_+ is equivalent to the finiteness of $\int_1^\infty t^{-1} L(t) dt$. The same is true for the finiteness of $\psi''(\lambda_+)$.

Proof of Lemma 9.4.1 We first prove the assertion concerning α_+ . Since

$$\alpha_+ = \frac{\psi'(\lambda_+)}{\psi(\lambda_+)},$$

we have to estimate the values of $\psi'(\lambda_+)$ and $\psi(\lambda_+)$. The finiteness of $\psi'(\lambda_+)$ is equivalent to that of

$$-\int_1^\infty t e^{\lambda_+ t} dF_+(t) = \int_1^\infty t(\lambda_+ V(t) dt - dV(t)), \tag{9.4.6}$$

where, for $V(t) = o(1/t)$,

$$-\int_1^\infty t dV(t) = V(1) + \int_1^\infty V(t) dt.$$

Hence the finiteness of the integral on the left-hand side of (9.4.6) is equivalent to that of the sum

$$\lambda_+ \int_1^\infty t V(t) dt + \int_1^\infty V(t) dt$$

or, which is the same, to the finiteness of the integral $\int_1^\infty t V(t) dt$. Similarly we see that the finiteness of $\psi(\lambda_+)$ is equivalent to that of $\int_1^\infty V(t) dt$. This implies the assertion of the lemma in the case $\int_1^\infty V(t) dt < \infty$, where one has $V(t) = o(1/t)$. If $\int_1^\infty V(t) dt = \infty$, then $\psi(\lambda_+) = \infty$, $\ln \psi(\lambda) \rightarrow \infty$ as $\lambda \uparrow \lambda_+$ and hence $\alpha_+ = \lim_{\lambda \uparrow \lambda_+} (\ln \psi(\lambda))' = \infty$.

The assertion concerning $\psi''(\lambda_+)$ can be proved in exactly the same way. The lemma is proved. □

The lemma implies the following:

- (a) If $\beta < 2$ or $\beta = 2$ and $\int_1^\infty t^{-1} L(t) dt = \infty$, then $\alpha_+ = \infty$ and the theorems of the previous section are applicable to $\mathbf{P}(S_n \geq x)$.
- (b) If $\beta > 3$ or $\beta = 3$ and $\int_1^\infty t^{-1} L(t) dt < \infty$, then $\alpha_+ < \infty$, $\psi''(\lambda_+) < \infty$ and we can apply Theorem 9.4.1.

It remains to consider the case

- (c) $\beta \in [2, 3]$, where the integral $\int_1^\infty t^{-1} L(t) dt$ is finite for $\beta = 2$ and is infinite for $\beta = 3$.

It is obvious that in case (c) we have $\alpha_+ < \infty$ and $\psi''(\lambda_+) = \infty$.

Put

$$V_+(t) := \frac{\lambda_+ t V(t)}{\beta \psi(\lambda_+)}, \quad b(n) := V_+^{(-1)}\left(\frac{1}{n}\right),$$

where $V_+^{(-1)}(1/n)$ is the value of the function inverse to V_+ at the point $1/n$.

Theorem 9.4.2 *Let ξ be a non-lattice random variable, $\mathbf{F} \in \mathcal{ER}$ and condition (c) hold. If $\Delta_n \rightarrow 0$ slowly enough as $n \rightarrow \infty$, then, for $y = x - \alpha_+ n = o(n)$,*

$$\mathbf{P}(S_n \in \Delta_n[x]) = \frac{\Delta_n e^{-n\Lambda(\alpha_+) - \lambda_+ y}}{b(n)} \left(f^{(\beta-1,1)} \left(\frac{y}{b(n)} \right) + o(1) \right),$$

where $f^{(\beta-1,1)}$ is the density of the stable law $\mathbf{F}_{(\beta-1,1)}$ with parameters $\beta - 1, 1$, and the remainder term $o(1)$ is uniform in y .

We will see from the proof of the theorem that studying the probabilities of large deviations in the case where $\alpha_+ < \infty$ and $\psi''(\lambda_+) = \infty$ is basically impossible outside the class \mathcal{ER} , since it is impossible to find theorems on the limiting distribution of S_n in the case $\text{Var}(\xi) = \infty$ without the conditions $[\mathbf{R}_{\gamma,\rho}]$ of Sect. 8.8 being satisfied.

Proof of Theorem 9.4.2 Condition (c) implies that $\alpha_+ = \mathbf{E}\xi^{(\alpha_+)} < \infty$ and $\text{Var}(\xi^{(\alpha_+)}) = \infty$. We will use Theorem 9.2.1A. For $\Delta_n \rightarrow 0$ slowly enough we will obtain, as in the proof of Theorem 9.4.1, that relation (9.4.3) holds true. But now, in contrast to Theorem 9.4.1, in order to calculate the probability on the right-hand side of (9.4.3), we have to employ the integro-local Theorem 8.8.3 on convergence to a stable law. In our case, by the properties of r.v.f.s, one has

$$\begin{aligned} \mathbf{P}(\xi^{(\alpha_+)} \geq t) &= -\frac{1}{\psi(\lambda_+)} \int_t^\infty e^{\lambda_+ u} dF_+(u) = \frac{1}{\psi(\lambda_+)} \int_t^\infty (\lambda_+ V(u) du - dV(u)) \\ &= \frac{\lambda_+}{\beta \psi(\lambda_+)} t^{-\beta+1} L_+(t) \sim V_+(t), \end{aligned} \tag{9.4.7}$$

where $L_+(t) \sim L(t)$ is a slowly varying function. Moreover, the left tail of the distribution $\mathbf{F}^{(\alpha_+)}$ decays at least exponentially fast. By virtue of the results of Sect. 8.8, this means that, for $b(n) = V_+^{(-1)}(1/n)$, we have convergence of the distributions of $\frac{S_n^{(\alpha_+)} - \alpha_+ n}{b(n)}$ to the stable law $\mathbf{F}_{\beta-1,1}$ with parameters $\beta - 1 \in [1, 2]$ and 1. It remains to use representation (9.4.3) and Theorem 8.8.3 which implies that, provided $\Delta_n \rightarrow 0$ slowly enough, one has

$$\mathbf{P}(S_n^{(\alpha_+)} - \alpha_+ n \in \Delta_n[y]) = \frac{\Delta_n}{b(n)} f^{(\beta-1,1)} \left(\frac{y}{b(n)} \right) + o \left(\frac{1}{b(n)} \right)$$

uniformly in y . The theorem is proved. □

Theorem 9.4.2 concludes the study of probabilities of large deviations of S_n/n in the vicinity of the point α_+ for distributions from the class \mathcal{ER} .

9.4.4 On the Large Deviation Probabilities in the Range $\alpha > \alpha_+$ for Distributions from the Class \mathcal{ER}

Now assume that the deviations x of S_n are such that $\alpha = x/n > \alpha_+$, and $y = x - \alpha_+n$ grows fast enough (faster than \sqrt{n} under the conditions of Theorem 9.4.1 and faster than $b(n)$ under the conditions of Theorem 9.4.2). Then, for the probability

$$\mathbf{P}(S^{(\alpha_+)} - \alpha_+n \in \Delta_n[y]), \tag{9.4.8}$$

the deviations y (see representation (9.4.3)) will belong to the zone of large deviations, so applying Theorems 8.7.1 and 8.8.3 to evaluate such probabilities does not make much sense. Relation (9.4.7) implies that, in the case $\mathbf{F} \in \mathcal{ER}$, we have $\mathbf{F}^{(\alpha_+)} \in \mathcal{R}$. Therefore, we will know the asymptotics of the probability (9.4.8) (and hence also of the probability $\mathbf{P}(S_n \in \Delta_n[x])$, see (9.4.3)) if we obtain integro-local theorems for the probabilities of large deviations of the sums S_n , in the case where the summands belong to the class \mathcal{R} . Such theorems are also of independent interest in the present chapter, and the next section will be devoted to them. After that, in Sect. 9.6 we will return to the problem on large deviation probabilities in the class \mathcal{ER} mentioned in the title of this section.

9.5 Integral and Integro-Local Theorems on Large Deviation Probabilities for Sums S_n when the Cramér Condition Is not Met

If $\mathbf{E}\xi = 0$ and the right-side Cramér condition is not met ($\lambda_+ = 0$), then the rate function $\Lambda(\alpha)$ degenerates on the right semiaxis: $\Lambda(\alpha) = \lambda(\alpha) = 0$ for $\alpha \geq 0$, and the results of Sects. 9.1–9.4 on the probabilities of large deviations of S_n are of little substance. In this case, in order to find the asymptotics of $\mathbf{P}(S_n \geq x)$ and $\mathbf{P}(S_n \in \Delta[x])$, we need completely different approaches, while finding these asymptotics is only possible under additional conditions on the behaviour of the tail $F_+(t)$ of the distribution \mathbf{F} , similarly to what happened in Sect. 8.8 when studying convergence to stable laws.

The above-mentioned additional conditions consist of the assumption that the tail $F_+(t)$ behaves regularly enough. In this section we will assume that $F_+(t) = V(t) \in \mathcal{R}$, where \mathcal{R} is the class of regularly varying functions introduced in the previous section (see also Appendix 6). To make the exposition more homogeneous, we will confine ourselves to the case $\beta > 2$, $\text{Var}(\xi) < \infty$, where $-\beta$ is the power exponent in the function $V \in \mathcal{R}$ (see (9.4.5)). Studying the case $\beta \in [1, 2]$ ($\text{Var}(\xi) = \infty$) does not differ much from the exposition below, but it would significantly increase the volume of the exposition and complicate the text, and therefore is omitted. Results for the case $\beta \in (0, 2]$ can be found in [8, Chap. 3].

9.5.1 Integral Theorems

Integral theorems for probabilities of large deviations of S_n and maxima $\bar{S}_n = \max_{k \leq n} S_k$ in the case $\mathbf{E}\xi = 0$, $\text{Var}(\xi) < \infty$, $\mathbf{F} \in \mathcal{R}$, $\beta > 2$, follow immediately from the bounds obtained in Appendix 8. In particular, Corollaries A8.2.1 and A8.3.1 of Appendix 8 imply the following result.

Theorem 9.5.1 *Let $\mathbf{E}\xi = 0$, $\text{Var}(\xi) < \infty$, $\mathbf{F} \in \mathcal{R}$ and $\beta > 2$. Then, for $x \gg \sqrt{n \ln n}$,*

$$\mathbf{P}(\bar{S}_n \geq x) \sim \mathbf{P}(S_n \geq x) \sim nV(x). \quad (9.5.1)$$

Under an additional condition $[\mathbf{D}_0]$ to be introduced below, the assertion of this theorem will also follow from the integro-local Theorem 9.5.2 (see below).

Comparing Theorem 9.5.1 with the results of Sects. 9.2–9.4 shows that the nature of the large deviation probabilities is completely different here. Under the Cramér condition and for $\alpha = x/n \in (0, \alpha_+)$, the large deviations of S_n are, roughly speaking, “equally contributed to by all the summands” ξ_k , $k \leq n$. This is confirmed by the fact that, for a fixed α , the limiting conditional distribution of ξ_k , $k \leq n$, given that $S_n \in \Delta[x]$ (or $S_n \geq x$) for $x = \alpha n$, $\Delta = 1$, as $n \rightarrow \infty$ coincides with the distribution $\mathbf{F}^{(\alpha)}$ of the random variable $\xi^{(\alpha)}$. The reader can verify this himself/herself using Theorem 9.3.2. In other words, the conditions $\{S_n \in \Delta[x]\}$ (or $\{S_n \geq x\}$), $x = \alpha n$, change equally (from \mathbf{F} to $\mathbf{F}^{(\alpha)}$) the distributions of all the summands.

However, if the Cramér condition is not met, then under the conditions of Theorem 9.5.1 the large deviations of S_n are essentially due to one large (comparable with x) jump. This is seen from the fact that the value of $nV(x)$ on the right-hand side of (9.5.1) is nothing else but the main term of the asymptotics for $\mathbf{P}(\bar{\xi}_n \geq x)$, where $\bar{\xi}_n = \max_{k \leq n} \xi_k$. Indeed, if $nV(x) \rightarrow 0$ then

$$\mathbf{P}(\bar{\xi}_n < x) = (1 - V(x))^n = 1 - nV(x) + O((nV(x))^2),$$

$$\mathbf{P}(\bar{\xi}_n \geq x) = nV(x) + O((nV(x))^2) \sim nV(x).$$

In other words, the probabilities of large deviations of S_n , \bar{S}_n and $\bar{\xi}_n$ are asymptotically the same. The fact that the probabilities of the events $\{\xi_j \geq y\}$ for $y \sim x$ play the determining role in finding the asymptotics of $\mathbf{P}(S_n \geq x)$ can easily be discovered in the bounds from Appendix 8.

Thus, while the asymptotics of $\mathbf{P}(S_n \geq x)$ for $x = \alpha n \gg \sqrt{n}$ in the Cramér case is determined by “the whole distribution \mathbf{F} ” (as the rate function $\Lambda(\alpha)$ depends on the “the whole distribution \mathbf{F} ”), these asymptotics in the case $\mathbf{F} \in \mathcal{R}$ are determined by the right tail $F_+(t) = V(t)$ only and do not depend on the “remaining part” of the distribution \mathbf{F} (for the fixed value of $\mathbf{E}\xi = 0$).

9.5.2 Integro-Local Theorems

In this section we will study the asymptotics of $\mathbf{P}(S_n \in \Delta[x])$ in the case where

$$\mathbf{E}\xi = 0, \quad \text{Var}\xi^2 < \infty, \quad \mathbf{F} \in \mathcal{R}, \quad \beta > 2, \quad x \gg \sqrt{n \ln n}. \quad (9.5.2)$$

These asymptotics are of independent interest and are also useful, for example, in finding the asymptotics of integrals of type $\mathbf{E}(g(S_n); S_n \geq x)$ for $x \gg \sqrt{n \ln n}$ for a wide class of functions g . As was already noted (see Subsection 4.4), in the next section we will use the results from the present section to obtain integro-local theorems under the Cramér condition (for summands from the class \mathcal{ER}) for deviations outside the Cramér zone.

In order to obtain integro-local theorems in this section, we will need additional conditions. Besides condition $\mathbf{F} \in \mathcal{R}$, we will also assume that the following holds:

Condition [D₀] For each fixed Δ , as $t \rightarrow \infty$,

$$V(t) - V(t + \Delta) = v(t)(\Delta + o(1)), \quad v(t) = \frac{\beta V(t)}{t}.$$

It is clear that if the function $L(t)$ in representation (9.4.5) (or the function $V(t)$) is differentiable for t large enough and $L'(t) = o(L(t)/t)$ as $t \rightarrow \infty$ (all sufficiently smooth s.v.f.s possess this property; cf. e.g., polynomials of $\ln t$ etc.), then condition [D₀] will be satisfied, and the derivative $-V'(t) \sim v(t)$ will play the role of the function $v(t)$.

Theorem 9.5.2 Let conditions (9.5.2) and [D₀] be met. Then

$$\mathbf{P}(S_n \in \Delta[x]) = \Delta n v(x)(1 + o(1)), \quad v(x) = \frac{\beta V(x)}{x},$$

where the remainder term $o(1)$ is uniform in $x \geq N\sqrt{n \ln n}$ and $\Delta \in [\Delta_1, \Delta_2]$ for any fixed $\Delta_2 > \Delta_1 > 0$ and any fixed sequence $N \rightarrow \infty$.

Note that in Theorems 9.5.1 and 9.5.2 we do not assume that $n \rightarrow \infty$. The assumption that $x \rightarrow \infty$ is contained in (9.5.2).

Proof For $y < x$, introduce the events

$$G_n := \{S_n \in \Delta[x]\}, \quad B_j := \{\xi_j < y\}, \quad B := \bigcap_{j=1}^n B_j. \quad (9.5.3)$$

Then

$$\mathbf{P}(G_n) = \mathbf{P}(G_n B) + \mathbf{P}(G_n \bar{B}), \quad \bar{B} = \bigcup_{j=1}^n \bar{B}_j, \quad (9.5.4)$$

where

$$\sum_{j=1}^n \mathbf{P}(G_n \bar{B}_j) \geq \mathbf{P}(G_n \bar{B}) \geq \sum_{j=1}^n \mathbf{P}(G_n \bar{B}_j) - \sum_{i < j \leq n} \mathbf{P}(G_n \bar{B}_i \bar{B}_j) \quad (9.5.5)$$

(see property 8 in Sect. 9.2.2).

The proof is divided into three stages: the bounding of $\mathbf{P}(G_n B)$, that of $\mathbf{P}(G_n \bar{B}_i \bar{B}_j)$, $i \neq j$, and the evaluation of $\mathbf{P}(G_n \bar{B}_j)$.

(1) *A bound on $\mathbf{P}(G_n B)$.* We will make use of the rough inequality

$$\mathbf{P}(G_n B) \leq \mathbf{P}(S_n \geq x; B) \quad (9.5.6)$$

and Theorem A8.2.1 of Appendix 8 which implies that, for $x = ry$ with a fixed $r > 2$, any $\delta > 0$, and $x \geq N\sqrt{n \ln n}$, $N \rightarrow \infty$, we have

$$\mathbf{P}(S_n \geq x; B) \leq (nV(y))^{r-\delta}. \quad (9.5.7)$$

Here we can always choose r such that

$$(nV(x))^{r-\delta} \ll n\Delta v(x) \quad (9.5.8)$$

for $x \gg \sqrt{n}$. Indeed, putting $n := x^2$ and comparing the powers of x on the right-hand and left-hand sides of (9.5.8), we obtain that for (9.5.8) to hold it suffices to choose r such that

$$(2 - \beta)(r - \delta) < 1 - \beta,$$

which is equivalent, for $\beta > 2$, to the inequality.

$$r > \frac{\beta - 1}{\beta - 2}.$$

For such r , we will have that, by (9.5.6)–(9.5.8),

$$\mathbf{P}(G_n B) = o(n\Delta v(x)). \quad (9.5.9)$$

Since $r - \delta > 1$, we see that, for $n \ll x^2$, relations (9.5.8) and (9.5.9) will hold true all the more.

(2) *A bound for $\mathbf{P}(G_n \bar{B}_i \bar{B}_j)$.* It is sufficient to bound $\mathbf{P}(G_n \bar{B}_{n-1} \bar{B}_n)$. Set

$$\delta := \frac{1}{r} < \frac{1}{2}, \quad H_k := \{v : v < (1 - k\delta)x + \Delta\}, \quad k = 1, 2.$$

Then

$$\mathbf{P}(G_n \bar{B}_{n-1} \bar{B}_n) = \int_{H_2} \mathbf{P}(S_{n-2} \in dz)$$

$$\times \int_{H_1} \mathbf{P}(z + \xi \in dv, \xi \geq \delta x) \mathbf{P}(v + \xi \in \Delta[x], \xi \geq \delta x). \tag{9.5.10}$$

Since in the domain H_1 we have $x - v > \delta x - \Delta$, the last factor on the right-hand side of (9.5.10) has, by condition $[D_0]$, the form $\Delta v(x - v)(1 + o(1)) \leq c \Delta v(x)$ as $x \rightarrow \infty$, so the integral over H_1 in (9.5.10), for x large enough, does not exceed

$$c \Delta v(x) \mathbf{P}(z + \xi \in H_1; \xi \geq \delta x) \leq c \Delta v(x) V(\delta x).$$

The integral over the domain H_2 in (9.5.10) evidently allows a similar bound. Since $nV(x) \rightarrow 0$, we obtain that

$$\sum_{i < j \leq n} \mathbf{P}(G_n \bar{B}_i \bar{B}_j) \leq c_1 \Delta n^2 v(x) V(x) = o(\Delta n v(x)). \tag{9.5.11}$$

(3) *The evaluation of $\mathbf{P}(G_n \bar{B}_j)$ is based on the relation*

$$\begin{aligned} \mathbf{P}(G_n \bar{B}_n) &= \int_{H_1} \mathbf{P}(S_{n-1} \in dz) \mathbf{P}(\xi \in \Delta[x - z], \xi \geq \delta x) \\ &\leq \int_{H_1} \mathbf{P}(S_{n-1} \in dz) \mathbf{P}(\xi \in \Delta[x - z]) \\ &= \Delta \int_{H_1} \mathbf{P}(S_{n-1} \in dz) v(x - z) (1 + o(1)), \end{aligned} \tag{9.5.12}$$

which yields

$$\begin{aligned} \mathbf{P}(G_n \bar{B}_n) &\leq \Delta \mathbf{E}[v(x - S_{n-1}); S_{n-1} < (1 - \delta)x + \Delta] (1 + o(1)) \\ &= \Delta v(x) (1 + o(1)). \end{aligned} \tag{9.5.13}$$

The last relation is valid for $x \gg \sqrt{n}$, since, by Chebyshev's inequality, $\mathbf{E}[v(x - S_{n-1}); |S_{n-1}| \leq M\sqrt{n}] \sim v(x)$ as $M \rightarrow \infty$, $M\sqrt{n} = o(x)$ and, moreover, the following evident bounds hold:

$$\mathbf{E}[v(x - S_{n-1}); S_{n-1} \in (M\sqrt{n}, (1 - \delta)x + \Delta)] = o(v(x)),$$

$$\mathbf{E}[v(x - S_{n-1}); S_{n-1} \in (-\infty, -M\sqrt{n})] = o(v(x))$$

as $M \rightarrow \infty$.

Similarly, by (virtue of (9.5.12)) we get

$$\mathbf{P}(G_n \bar{B}_n) \geq \int_{-\infty}^{(1-\delta)x} \mathbf{P}(S_{n-1} \in dz) \mathbf{P}(\xi \in \Delta[x - z]) \sim \Delta v(x). \tag{9.5.14}$$

From (9.5.13) and (9.5.14) we obtain that

$$\mathbf{P}(G_n \overline{B}_n) = \Delta v(x)(1 + o(1)).$$

This, together with (9.5.4), (9.5.9) and (9.5.11), yields the representation

$$\mathbf{P}(G_n) = \Delta n v(x)(1 + o(1)).$$

The required uniformity of the term $o(1)$ clearly follows from the preceding argument. The theorem is proved. \square

Theorem 9.5.2 implies the following

Corollary 9.5.1 *Let the conditions of Theorem 9.5.2 be satisfied. Then there exists a fixed sequence Δ_N converging to zero slowly enough as $N \rightarrow \infty$ such that the assertion of Theorem 9.5.2 remains true when the segment $[\Delta_1, \Delta_2]$ is replaced in it with $[\Delta_N, \Delta_2]$.*

9.6 Integro-Local Theorems on the Probabilities of Large Deviations of S_n Outside the Cramér Range (Under the Cramér Condition)

We return to the case where the Cramér condition is met. In Sects. 9.3 and 9.4 we obtained integro-local theorems for deviations inside and on the boundary of the Cramér range. It remains to study the asymptotics of $\mathbf{P}(S_n \in \Delta[x])$ outside the Cramér range, i.e. for $\alpha = x/n > \alpha_+$. Preliminary observations concerning this problem were made in Sect. 9.4.4 where it was reduced to integro-local theorems for the sums S_n when Cramér's condition is not satisfied. Recall that in that case we had to restrict ourselves to considering distributions from the class \mathcal{ER} defined in Sect. 9.4.3 (see (9.4.4)).

Theorem 9.6.1 *Let $\mathbf{F} \in \mathcal{ER}$, $\beta > 3$, $\alpha = x/n > \alpha_+$ and $y = x - \alpha_+ n \gg \sqrt{n}$. Then there exists a fixed sequence Δ_N converging to zero slowly enough as $N \rightarrow \infty$, such that*

$$\begin{aligned} \mathbf{P}(S_n \in \Delta_N[x]) &= e^{-n\Lambda(\alpha_+) - \lambda_+ y} n \Delta_N v_+(y) (1 + o(1)) \\ &= e^{-n\Lambda(\alpha)} n \Delta_N v_+(y) (1 + o(1)), \end{aligned}$$

where $v_+(y) \equiv \lambda_+ V(y)/\psi(\lambda_+)$, the remainder term $o(1)$ is uniform in x and n such that $y \gg N\sqrt{n \ln n}$, N being an arbitrary fixed sequence tending to ∞ .

Proof By Theorem 9.2.1A there exists a sequence Δ_N converging to zero slowly enough such that (cf. (9.4.3))

$$\mathbf{P}(S_n \in \Delta_N[x]) = e^{-n\Lambda(\alpha_+) - \lambda_+ y} \mathbf{P}(S_n^{(\alpha_+)} - \alpha_+ n \in \Delta_N[y]). \quad (9.6.1)$$

Since by properties (A1) and (A2) the function $\Lambda(\alpha)$ is linear for $\alpha > \alpha_+$:

$$\Lambda(\alpha) = \Lambda(\alpha_+) + (\alpha - \alpha_+)\lambda_+,$$

the exponent in (9.6.1) can be rewritten as

$$-n\Lambda(\alpha_+) - \lambda_+y = -n\Lambda(\alpha).$$

The right tail of the distribution of $\xi^{(\alpha_+)}$ has the form (see (9.4.7))

$$\mathbf{P}(\xi^{(\alpha_+)} \geq t) = \frac{\lambda_+}{\psi(\lambda_+)} \int_t^\infty V(u) du + V(t).$$

By the properties of regularly varying functions (see Appendix 6),

$$V(t) - V(t - u) = o(V(t))$$

as $t \rightarrow \infty$ for any fixed u . This implies that condition [D₀] of Sect. 9.5 is satisfied for the distribution of $\xi^{(\alpha_+)}$.

This means that, in order to calculate the probability on the right-hand side of (9.6.1), we can use Theorem 9.5.2 and Corollary 9.5.1, by virtue of which, as $\Delta_N \rightarrow 0$ slowly enough,

$$\mathbf{P}(S_n^{(\alpha_+)} - \alpha_+n \in \Delta_N[y]) = n\Delta_N v_+(y)(1 + o(1)),$$

where the remainder term $o(1)$ is uniform in all x and n such that $y \gg N\sqrt{n \ln n}$, $N \rightarrow \infty$.

The theorem is proved. □

Since $\mathbf{P}(S_n \in \Delta_N[x])$ decreases exponentially fast as x (or y) grows (note the factor $e^{-\lambda_+y}$ in (9.6.1)), Theorem 9.6.1 immediately implies the following integral theorem.

Corollary 9.6.1 *Under the conditions of Theorem 9.6.1,*

$$\mathbf{P}(S_n \geq x) = e^{-n\Lambda(\alpha)} \frac{nV(y)}{\psi(\lambda_+)} (1 + o(1)).$$

Proof Represent the probability $\mathbf{P}(S_n \geq x)$ as the sum

$$\begin{aligned} \mathbf{P}(S_n \geq x) &= \sum_{k=0}^\infty \mathbf{P}(S_n \in \Delta_N[x + k\Delta_N]) \\ &\sim e^{-n\Lambda(\alpha)} \frac{n\lambda_+}{\psi(\lambda_+)} \sum_{k=0}^\infty \Delta_N V(y + \Delta_N k) e^{-\lambda_+\Delta_N k}. \end{aligned}$$

Here the series on the right-hand side is asymptotically equivalent, as $N \rightarrow \infty$, to the integral

$$V(y) \int_0^\infty e^{-\lambda_+ t} dt = \frac{V(y)}{\lambda_+}.$$

The corollary is proved. □

Note that a similar corollary (i.e. the integral theorem) can be obtained under the conditions of Theorem 9.4.2 as well.

In the range of deviations $\alpha = \frac{x}{n} > \alpha_+$, only the case $\mathbf{F} \in \mathcal{ER}$, $\beta \in [2, 3]$ (recall that $\alpha_+ = \infty$ for $\beta < 2$) has not been considered in this text. As we have already said, it could also be considered, but that would significantly increase the length and complexity of the exposition. Results dealing with this case can be found in [8]; one can also find there a more complete study of large deviation probabilities.