1

Large deviations: basic results

Introduction

In the analysis of a system with a large number of interacting components (at a microscopic level) it is of clear importance to find out about its collective, or macroscopic, behaviour. This is quite an old problem, going back to the origins of statistical mechanics, in the search for a mathematical characterization of 'equilibrium states' in thermodynamical systems. Though the problem is old, and the foundations of equilibrium statistical mechanics have been settled, the general question remains of interest, especially in the set-up of non-equilibrium systems. We could then take as the object of study a (non-stationary) time evolution with a large number (n) of components, where the initial condition and/or the dynamics present some randomness. One example of such a collective description is the so-called hydrodynamic limit. Passing by a space-time scale change (micro $\to$ macro) it allows, through a limiting procedure, the derivation of a reduced description in terms of macroscopic variables, such as density and temperature. Other limits, besides the hydrodynamic, may also appear in different situations, giving rise to macroscopic equations.

In all such cases the macroscopic equation indicates the *typical* behaviour in a *limiting situation* ($n \to +\infty$, and proper rescaling). Thus, it is essential to know something about:

(i) rates of convergence, i.e. how are the fluctuations of the macroscopic random fields (for example, the empirical density) around the prescribed value given by the macroscopic equation?

(ii) how to estimate the chance of observing something quite different than what is prescribed by the macroscopic equation. According to the prescription of the macroscopic equation these are 'rare events' and their probabilities will tend to zero, but *at which speed*?

In the above description we identify the three most basic limit theorems in classical probability: the macroscopic description corresponds to a 'law of large numbers'; the behaviour of the fluctuations, or 'moderate deviations', fits into the frame of a 'central limit theorem'; and the estimates of the probability of rare events constitute what are usually called 'large deviation principles'. The program for investigating the collective behaviour for evolutions given by Markov processes on $\{0, 1\}^{\mathbb{Z}^d}$ or $\mathbb{N}^{\mathbb{Z}^d}$, has grown since the 1980s (see [79]), and has taken definite forms for a class of them, cf. [78, 178, 283]. The situation is much less developed in the context of mechanical systems (see [283]).

The content of this book is closely related to questions such as (ii) above, and in particular to their connection with metastability, which will be discussed from Chapters 4 to 7. Perhaps we should say a few words on possible motivations for such estimates, bearing in mind the collective description of large systems. For example, if one wants to investigate the behaviour of the system at time scales longer than those for which the macroscopic equation is valid, then it is necessary to pay attention to such 'large fluctuations', since they will eventually occur. The ability to compare their probabilities becomes a crucial point in order to predict the long-term behaviour of the system. The classic example is a tunnelling event between two stable points of the macroscopic equation. Somehow, this comparison can be seen as a first step: one would believe that the large fluctuation should occur in the *least improbable* way. Nevertheless, carrying out this long time analysis may present (technical or serious) difficulties. One instance where this has been done quite completely is that in which the dynamics is, in some sense, already macroscopic; more precisely, it is obtained by the addition of a small external noise to a non-chaotic dynamical system. This is the object of Freidlin and Wentzell's theory [122], which will be studied in Chapters 2 and 5 of this book, also in connection with the phenomenon of metastability.

One should stress how closely related are the three mentioned problems: derivation of macroscopic equations/law of large numbers, fluctuations, and large deviations. Large deviation estimates yield stronger statements on the convergence of macroscopic density fields. On the other hand, a standard method for the derivation of large deviation estimates involves the validity of a large class of deterministic macroscopic limits (law of large numbers). A very important example of such a connection comes from equilibrium theory, through the possibility of applying large deviations to obtain the equivalence of ensembles, as pointed out in the fundamental articles of Ruelle [257], and Lanford [189], which have stimulated intense research. This goes far beyond the scope of this book, as for instance, the questions related to phase separation and surface large deviations. A brief discussion will appear in Chapter 3, with indications to recent research articles.

As a usual set-up for large deviations we could take a sequence of probability measures $(\mu_n)_{n\geq 1}$ on some metric space M, weakly converging to a Dirac point measure at some $m \in M$, in the sense that $\lim_{n\to+\infty} \int f\, d\mu_n = f(m)$ for all $f: M \to \mathbb{R}$ continuous and bounded. (In our previous discussion, μ_n should represent the law of an observable such as the empirical density, m representing a macrostate such as an equilibrium density.) The goal is to find the speed at which $\mu_n(A)$ tends to zero, when A is a fixed measurable set staying at positive distance from m. In particular, one wishes to detect whether a fast, exponential decay happens, in the sense that there exists $I(A) \in (0, +\infty]$ such that

$$\mu_n(A) \approx e^{-nI(A)}. \tag{1.1}$$

Throughout the text, $\approx$ denotes logarithmic equivalence, i.e. (1.1) means $n^{-1} \log \mu_n(A) \to -I(A)$ as $n \to \infty$. (*Notation*. $\log = \log_e$ everywhere in this text.)

Let us assume that (1.1) holds for a certain class of sets A; let A and B be two disjoint sets for which it holds. Since $\mu_n(A \cup B) = \mu_n(A) + \mu_n(B)$ it follows at once that $\mu_n(A \cup B) \approx e^{-n \min\{I(A), I(B)\}}$. This might suggest $I(A)$ of the form

$$I(A) = \inf_{x \in A} I(x), \tag{1.2}$$

for some point function I, which would then be called a 'rate function'. If so, we cannot expect (1.1) to hold for all measurable sets A; to see this, consider for example continuous measures, so that $\mu_n\{x\} = 0$ for all points $x \in M$. If (1.1) were true for such sets, this would force I to be identically $+\infty$, incompatible with (1.1) and (1.2) for $A = M$. This means that some restriction on the sets for which (1.1) holds is needed. This will be discussed in the next two sections, where a possible set-up will be presented.

It is natural to ask why one chooses the logarithmic equivalence $\approx$ instead of a sharper estimate like the usual equivalence ($a_n \sim b_n$ iff $a_n/b_n \to 1$). Significant advantages of the previous choice (allowing polynomial errors in (1.1)) include simplicity and a wide range of applicability. On the other hand, 'exact' results are essential in many applications though in this text we shall not pursue them.

Moreover, situations are expected to occur where $I(A)$ could vanish, meaning that the decay is less than exponential, and that (1.1) does not provide enough information. In such cases, one definitely needs a more precise asymptotics.

For a comparison with 'moderate deviations' (central limit theorems), let us take $M = \mathbb{R}^d$. According to the previous notation, these refer to the asymptotics of $\mu_n(A_n)$ where $A_n = m + \alpha_n A$, $\alpha_n \to 0$ suitably, and A is fixed.

1.1 Cramér–Chernoff theorem on $\mathbb{R}$

Let us start with the simplest situation: the microstates correspond to the results of n independent tosses of a fair coin, and μ_n represents the law of the proportion of 'heads' (the macroscopic observable). The microstates are thus uniformly distributed on $\mathcal{X}_n = \{(\omega_1, \ldots, \omega_n) : \omega_i = 1 \text{ or } \omega_i = 0, \forall\, i\} = \{0, 1\}^n$, i.e. all 2^n points have equal probabilities, and if $A \subseteq [0, 1]$ is a Borel set, then

$$\mu_n(A) = 2^{-n} \sum_{k:k/n \in A} \binom{n}{k}, \tag{1.3}$$

where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ if $k \in \{0, 1, \ldots, n\}$, $\binom{n}{k} = 0$, otherwise.

We know that the weak law of large numbers holds in this situation, i.e. for any $\varepsilon > 0$ we have $\lim_{n\to+\infty} \mu_n\,(1/2 - \varepsilon, 1/2 + \varepsilon) = 1$. Let us in fact check this, getting a stronger estimate. Let $0 < a < b < 1$ and consider the probability that the resulting proportion of 'heads' belongs to the interval $[a, b]$. From (1.3) we immediately have:

$$2^{-n} \max_{k:k/n \in [a,b]} \binom{n}{k} \le \mu_n[a, b] \le (n+1) 2^{-n} \max_{k:k/n \in [a,b]} \binom{n}{k}.$$

In particular, as $n \to +\infty$

$$\frac{1}{n} \log\, \mu_n[a, b] \sim -\log 2 + \frac{1}{n} \max_{k:k/n \in [a,b]} \log \binom{n}{k}. \tag{1.4}$$

Recall now the classical Stirling formula (see e.g. [115], p. 54):

$$\sqrt{2\pi}\, n^{n+\frac{1}{2}} e^{-n}\, e^{1/(12n+1)} \le n! \le \sqrt{2\pi}\, n^{n+\frac{1}{2}} e^{-n}\, e^{1/12n}, \tag{1.5}$$

or its weaker version (see e.g. [115], p. 52)

$$n! \sim \sqrt{2\pi}\, n^{n+\frac{1}{2}} e^{-n}. \tag{1.6}$$

Considering logarithmic equivalence let us now use the even weaker relation:

$$\log\, n! = n\, \log\, n - n + O(\log\, n), \tag{1.7}$$

where $a_n = O(b_n)$ means that a_n/b_n remains bounded in n, which follows at once from (1.6). Applying (1.7) to equation (1.4) and performing simple calculations we get:

$$\lim_{n\to\infty} \frac{1}{n} \log\, \mu_n[a, b] = -\inf_{x \in [a,b]} I(x), \tag{1.8}$$

where, for each $x \in (0, 1)$,

$$I(x) = \log\, 2 + x \log\, x + (1-x) \log\,(1-x). \tag{1.9}$$

It is easily seen that (1.8) extends for any $a < b$ with the convention that $0 \log 0 = 0$ and $I(x) = +\infty$ if $x \notin [0, 1]$. Thus, we have a large deviation estimate for μ_n in the previously announced frame, and (1.1) holds for all intervals with non-empty

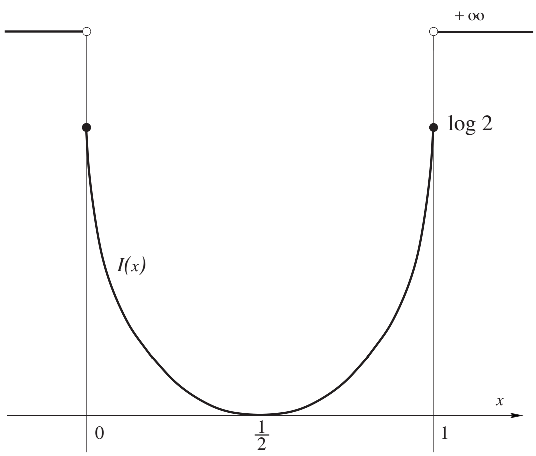

Figure 1.1

interior. The function I is continuous in the open interval $(0, 1)$, $I(x) = I(1 - x)$ for $0 \le x \le 1$, and it is strictly increasing in $(1/2, 1)$ (see Figure 1.1). In particular, $I(x) = 0$ iff $x = 1/2$, and for any $\varepsilon > 0$

$$\mu_n\big([0, 1] \setminus (1/2 - \varepsilon, 1/2 + \varepsilon)\big) \approx e^{-nI(1/2+\varepsilon)}.$$

The previous example, a very simple application of the Stirling formula, can be thought of as a particular case of the following: let $X_1, X_2, \ldots$ be independent and identically distributed (i.i.d.) real random variables on some probability space $(\Omega, \mathcal{A}, P)$. Let μ_n be the law of the sample average $\bar{X}_n := n^{-1} \sum_{i=1}^n X_i$. If X_1 is integrable and $m = EX_1$, the classical weak law of large numbers tells us that $\mu_n(\mathbb{R} \setminus (m - \varepsilon, m + \varepsilon)) = P(|\bar{X}_n - m| \ge \varepsilon)$ tends to zero as $n \to +\infty$, for any $\varepsilon > 0$. With proper conditions on the tails of the distribution of X_1, we again expect exponential decay of $\mu_n\big(\mathbb{R} \setminus (m - \varepsilon, m + \varepsilon)\big)$, for any given $\varepsilon > 0$, and we may hope to get something like (1.1) for a large class of sets A; the goal is to compute the rate, in the sense of logarithmic equivalence. This classical situation was treated by Cramér in 1937 (cf. [69]) for distributions with an absolutely continuous component, providing 'exact' asymptotics, and extended to the general case by Chernoff in 1952 [61] in the sense of logarithm equivalence. The result, as stated and proven by Chernoff, is the following.

Theorem 1.1 (Cramér–Chernoff) *Let $X_1, X_2, \ldots$ be i.i.d. real random variables with common law μ, and consider the sample average $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$. Let $\hat{\mu}$ denote the moment generating function of μ, i.e. $\hat{\mu}(\zeta) = E\, e^{\zeta X_1}$ for $\zeta \in \mathbb{R}$, and define*:

$$I_\mu(x) = \sup_{\zeta \in \mathbb{R}} (\zeta x - \log \hat{\mu}(\zeta)), \tag{1.10}$$

for $x \in \mathbb{R}$. *(*$\hat{\mu}$ *takes values on* $(0, +\infty]$ *and* I_μ *is* $[0, +\infty]$ *valued; we set* $\log(+\infty) = +\infty$.)
(*a*) *Upper bound. If* X_1 *is integrable and* $m = EX_1$*, then*:

$$\begin{aligned} &(i) \quad P(\bar{X}_n \ge x) \le e^{-nI_\mu(x)}, \quad \text{if } x \ge m; \\ &(ii) \quad P(\bar{X}_n \le x) \le e^{-nI_\mu(x)}, \quad \text{if } x \le m. \end{aligned} \tag{1.11}$$

(*b*) *Lower bound. For any* $x \in \mathbb{R}$:

$$(i) \quad \liminf_{n\to+\infty} \frac{1}{n} \log P(\bar{X}_n \ge x) \ge -I_\mu(x);$$

$$(ii) \quad \liminf_{n\to+\infty} \frac{1}{n} \log P(\bar{X}_n \le x) \ge -I_\mu(x).$$

Remark 1.2 With μ as above and letting ν denote the law of $aX_1 + b$, for $a, b \in \mathbb{R}$, $a \ne 0$, then $\hat{\nu}(\zeta) = e^{b\zeta}\hat{\mu}(a\zeta)$ and consequently $I_\nu(x) = I_\mu((x-b)/a)$, for all $x \in \mathbb{R}$.

Proof of the upper bound We shall now see that optimization over a class of exponential Markov inequalities gives us the upper estimate.

Notice first that Remark 1.2 reduces the proof to the case $x = 0 \le m$, all other cases being reduced to this by change of sign and considering $X_i' = X_i - x$.

If $\zeta \le 0$ we apply a Markov exponential inequality to write:

$$P(\bar{X}_n \le 0) = P(\zeta S_n \ge 0) \le P(e^{\zeta S_n} \ge 1) \le E\, e^{\zeta S_n} = (\hat{\mu}(\zeta))^n, \tag{1.12}$$

where $S_n = \sum_{i=1}^n X_i = n\bar{X}_n$. Thus we have:

$$P(\bar{X}_n \le 0) \le \big(\inf_{\zeta\le 0} \hat{\mu}(\zeta)\big)^n. \tag{1.13}$$

On the other hand, if $\zeta \ge 0$, the Jensen inequality yields

$$\hat{\mu}(\zeta) = E\, e^{\zeta X_1} \ge e^{\zeta m} \ge 1 = \hat{\mu}(0), \tag{1.14}$$

that is

$$m \ge 0 \Rightarrow \inf_{\zeta\le 0} \hat{\mu}(\zeta) = \inf_{\zeta\in\mathbb{R}} \hat{\mu}(\zeta) = e^{-I_\mu(0)}, \tag{1.15}$$

and the proof follows at once from (1.13). □

Remark 1.3 The previous argument with the Jensen inequality tells us that if $m = 0$ then $I_\mu(0) = -\log \hat{\mu}(0) = 0$ (if $m = 0$, (1.14) is applicable to any ζ), and by Remark 1.2 we conclude that $I_\mu(m) = 0$ whenever $m \in \mathbb{R}$.

The proof of part (b) of Theorem 1.1, as presented below, is exactly that given by Chernoff [61]. It does not fit so well in the general 'frame' to be used later; in the present discussion we wanted to stress the role of the Stirling formula and to give an idea of the historical development. We shall discuss this later.

Proof of the lower bound Again, by Remark 1.2, it suffices to prove

$$\liminf_{n\to+\infty} \frac{1}{n} \log P(\bar{X}_n \le 0) \ge -I_\mu(0). \tag{1.16}$$

We first observe that the statement becomes trivial if $P(X_1 \le 0) = 1$, in which case $n^{-1} \log P(\bar{X}_n \le 0) = 0 \ge -I_\mu(0)$. If $P(X_1 \ge 0) = 1$, we have $n^{-1} \log P(\bar{X}_n \le 0) = n^{-1} \log P(X_1 = 0, \ldots, X_n = 0) = \log P(X_1 = 0)$. On the other hand, the monotone convergence theorem implies that in this case $\hat{\mu}(\zeta)$ decreases to $P(X_1 = 0)$ as ζ decreases to $-\infty$, so that $\log P(X_1 = 0) = \inf_{\zeta\in\mathbb{R}} \log \hat{\mu}(\zeta) = -I_\mu(0)$, and (1.16) becomes trivial too.

We now assume that $P(X_1 < 0) > 0$ and $P(X_1 > 0) > 0$. The main ingredient for the proof is the following lemma, which gives a stronger form of the lower bound in the case of random variables taking finitely many values, usually called simple. As in the coin-tossing example, this proof is based on the Stirling formula. □

Lemma 1.4 *Assume there exist real numbers $x_1, \ldots, x_r$ such that*

$$\min_{1\le i\le r} x_i < 0 < \max_{1\le i\le r} x_i , \tag{1.17}$$

and that $p_i = P(X_1 = x_i) > 0 \quad (i = 1, \ldots, r)$. Let us set $\psi(\zeta) = \sum_{i=1}^r p_i\, e^{\zeta x_i}$ and $\chi = \inf_{\zeta\in\mathbb{R}} \psi(\zeta)$. Then, there exist $c > 0$ and $N \in \mathbb{N}$ such that for all $n \ge N$ we may find positive integer numbers $n_1, \ldots, n_r$ such that

$$\sum_{i=1}^r n_i = n,$$

$$\sum_{i=1}^r n_i x_i \le 0, \tag{1.18}$$

$$p(n_1, \ldots, n_r) := n! \prod_{i=1}^r \frac{p_i^{n_i}}{n_i!} \ge c\, n^{-(r-1)/2}\, \chi^n.$$

Before proving Lemma 1.4 let us examine its content. If $n = \sum_{i=1}^r n_i$ and $p_i = P(X_1 = x_i)$, then $p(n_1, \ldots, n_r)$, defined in (1.18), gives the probability that each x_i appears exactly n_i times among the observations $X_1, \ldots, X_n$. In particular, if $\sum_{i=1}^r n_i x_i \le 0$, then

$$p(n_1, \ldots, n_r) \le P(S_n \le 0) = P(\bar{X}_n \le 0).$$

Now, if X_1 takes values on the finite set $\{x_1, \ldots, x_r\}$, so that $p_1 + \cdots + p_r = 1$, we have $\psi(\zeta) = \hat{\mu}(\zeta)$ and $\log \chi = -I_\mu(0)$, and the lemma tells us: $P(\bar{X}_n \le 0) \ge c\, n^{-(r-1)/2}\, e^{-nI_\mu(0)}$, clearly stronger than the lower bound stated at Theorem 1.1 for this particular case. (The constant c appearing in the previous formula will depend on $r, p_1, \ldots, p_r, x_1, \ldots, x_r$.)

Proof of Lemma 1.4 If $n = \sum_{i=1}^r n_i$ and $n_1, \ldots, n_r$ are positive integers, we can write, using the Stirling formula (1.5):

$$p(n_1, \ldots, n_r) \geq \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n} \prod_{i=1}^r \frac{(p_i\, e)^{n_i}}{\sqrt{2\pi} n_i^{n_i+\frac{1}{2}} e^{\frac{1}{12n_i}}} \geq C(r)\, n^{-\frac{r-1}{2}} q(n_1, \ldots, n_r), \tag{1.19}$$

where $C(r) := e^{-\frac{r}{12}} (2\pi)^{-\frac{r-1}{2}}$ and, for $z_1, \ldots, z_r$ positive real numbers,

$$q(z_1, \ldots, z_r) := \prod_{i=1}^r \left(\frac{n\, p_i}{z_i}\right)^{z_i}. \tag{1.20}$$

Inequalities (1.17) imply that $\psi(\zeta) \to +\infty$ as $|\zeta| \to +\infty$. Thus we may take $\zeta_0 \in \mathbb{R}$ such that $\psi(\zeta_0) = \chi$, and since the function ψ is smooth, its derivative at the point ζ_0 vanishes. Taking $z_i = n\, p_i\, e^{\zeta_0 x_i}/\chi$ we have

$$\sum_{i=1}^r z_i = n, \quad \sum_{i=1}^r z_i\, x_i = 0 \quad \text{and} \quad q(z_1, \ldots, z_r) = \chi^n. \tag{1.21}$$

The numbers z_i might fail to be integers, but they are large, for large n, since $p_i > 0$ for each i. Assuming, without loss of generality, $x_1 \leq x_j$ for $j = 2, \ldots, r$, let:

$$n_j = [z_j], \qquad j = 2, \ldots, r,$$

$$n_1 = n - \sum_{i=2}^r n_j,$$

where $[x] := \max\{k \in \mathbb{N} : k \leq x\}$ denotes the integer part of x, if $x \geq 0$. Thus, $n_1, \ldots, n_r$ are now positive integers verifying the first two relations in (1.18). Moreover, as we shall now see, there exist $\tilde{c} = \tilde{c}\,(r, p_1, \ldots, p_r, x_1, \ldots, x_r) > 0$ and $N \in \mathbb{N}$ so that if $n \geq N$

$$q(n_1, \ldots, n_r) \geq \tilde{c}\, q(z_1, \ldots, z_r),$$

which will conclude the proof of the lemma. One way to see this consists (following Azencott [7], Chapter 1) in observing that if J is a bounded interval and $a > 0$ is fixed, we can find $c_1 = c(a, J) > 0$ so that

$$\left(\frac{u}{au+b}\right)^{au+b} \geq c_1 \left(\frac{u}{au}\right)^{au}, \tag{1.22}$$

for each $u \geq 1$, with $au + b \geq 1$, and $b \in J$. (Indeed, for u, b as above, we have $(\frac{u}{au+b})^{au+b}(\frac{au}{u})^{au} = (1 + \frac{b}{au})^{-au}(\frac{u}{au+b})^b \geq c(a, J)$ as one easily checks.)

For each fixed $i = 1, \ldots, r$ we apply this for $u_i = np_i$, $a_i u_i = z_i$, $a_i u_i + b = n_i$, so that $a_i = e^{\zeta_0 x_i}/\chi$, for each i, $J_i = [-1, 0]$ if $i = 2, \ldots, r$, $J_1 = [0, r]$, and take $\tilde{c}$ as the rth power of the smallest of such $c(a_i, J_i)$. □

Remark 1.5 As already noticed, the previous lemma gives a sharper estimate than (1.16) in the case of finitely valued random variables. Considering (1.19), to get (1.16) it would suffice to prove, for all large n, the existence of $n_1, \ldots, n_r$ verifying the first two equations in (1.18) and such that $\liminf_{n\to+\infty} n^{-1} \log q(n_1, \ldots, n_r) \geq \log \chi$. We shall come back to this in Section 1.3.

Proof of the lower bound (Continued) Assume $\mu(0, +\infty) > 0$ and $\mu(-\infty, 0) > 0$. If X_1 is a simple random variable, i.e. with values on a finite set, the result is contained in Lemma 1.4. Let us proceed to the general case.

Case 1. X_1 is a discrete random variable. In this case, let $x_1, x_2, \ldots$ be such that $p_i = P(X_1 = x_i) > 0$ for each i and $\sum_{i=1}^{\infty} P(X_1 = x_i) = 1$. By assumption we may take $r \geq 2$ such that

$$\min_{1\leq i\leq r} x_i < 0 < \max_{1\leq i\leq r} x_i \,. \tag{1.23}$$

Letting $\psi_k(\zeta) = \sum_{i=1}^{k} p_i\, e^{\zeta x_i}$ and using Lemma 1.4, for any $k \geq r$ we may take $c_k > 0$, $N_k \in \mathbb{N}$ such that if $n \geq N_k$, $n_1, \ldots, n_k$ are given by Lemma 1.4 then

$$P(\bar{X}_n \leq 0) \geq p(n_1, \ldots, n_k) \geq c_k\, n^{-\frac{k-1}{2}} \Big(\inf_{\zeta\in\mathbb{R}} \psi_k(\zeta) \Big)^n,$$

which implies:

$$\liminf_{n\to+\infty} \frac{1}{n} \log P(\bar{X}_n \leq 0) \geq \inf_{\zeta\in\mathbb{R}} \log \psi_k(\zeta),$$

for each $k \geq r$. It then suffices to check that

$$\lim_{k\to+\infty} \inf_{\zeta\in\mathbb{R}} \log \psi_k(\zeta) = \inf_{\zeta\in\mathbb{R}} \log \hat{\mu}(\zeta). \tag{1.24}$$

Since the sequence $0 < \psi_k(\zeta)$ increases to $\hat{\mu}(\zeta)$ for any given ζ, we have at once that

$$a := \lim_{k\to+\infty} \inf_{\zeta\in\mathbb{R}} \log \psi_k(\zeta) \leq \inf_{\zeta\in\mathbb{R}} \log \hat{\mu}(\zeta) \in (-\infty, 0].$$

It remains to check the reversed inequality and for this let

$$A_k := \{\zeta \in \mathbb{R} : \log \psi_k(\zeta) \leq a\}.$$

Due to the continuity of $\log \psi_k$ and to (1.23), A_k is a non-empty compact set, for each $k \geq r$. From the monotonicity of the sequence $\psi_k(\zeta)$ we have $A_r \supseteq A_{r+1} \supseteq \ldots$ and so their intersection is non-empty. But, if $\bar{\zeta} \in \cap_{k\geq r} A_k$, we must have $\log \hat{\mu}(\bar{\zeta}) = \lim_{k\to+\infty} \log \psi_k(\bar{\zeta}) \leq a$ proving (1.24) and thus concluding the proof of the lower bound for discrete random variables. (This last step corresponds to the Dini theorem applied to the functions $\log \psi_k(\cdot)$.)

General case. For each $k \in \mathbb{N}$ let

$$X_j^{(k)} = \sum_{i \in \mathbb{Z}} \frac{i}{k} \mathbf{1}_{\{X_j \in (\frac{i-1}{k}, \frac{i}{k}]\}},$$

where $\mathbf{1}_A$ denotes the indicator function of the set A ($\mathbf{1}_A(\omega) = 1$ if $\omega \in A$; $\mathbf{1}_A(\omega) = 0$ otherwise), and $\bar{X}_n^{(k)} = n^{-1} \sum_{j=1}^n X_j^{(k)}$. Since

$$X_j \le X_j^{(k)} \le X_j + 1/k,$$

we see that if $\mu_{(k)}$ is the law of $X_1^{(k)}$ and $\widehat{\mu_{(k)}}$ its moment generating function, we have:

$$\hat{\mu}(\zeta) \le \widehat{\mu_{(k)}}(\zeta) \le e^{\zeta/k}\, \hat{\mu}(\zeta), \qquad \text{if } \zeta \ge 0,$$

$$e^{\zeta/k}\, \hat{\mu}(\zeta) \le \widehat{\mu_{(k)}}(\zeta) \le \hat{\mu}(\zeta), \qquad \text{if } \zeta \le 0,$$

so that $\widehat{\mu_{(k)}}(\zeta) \ge e^{-|\zeta|/k}\, \hat{\mu}(\zeta)$ for each ζ. Since both $\mu(0, +\infty)$ and $\mu(-\infty, 0)$ are assumed to be positive, there exist a and b positive constants for which

$$\hat{\mu}(\zeta) \ge a\, e^{b|\zeta|},$$

and with the same argument (Dini theorem) leading to (1.24) applied now to the functions $f_k(\zeta) = e^{-|\zeta|/k}\, \hat{\mu}(\zeta)$, we conclude that given $\delta > 0$, there exists $k \ge 1$ such that

$$\inf_{\zeta \in \mathbb{R}} \log \widehat{\mu_{(k)}}(\zeta) \ge \inf_{\zeta \in \mathbb{R}} \log \hat{\mu}(\zeta) - \delta. \tag{1.25}$$

From Case 1 and (1.25) we have

$$\liminf_{n \to +\infty} \frac{1}{n} \log\, P(\bar{X}_n \le 0) \ge \liminf_{n \to +\infty} \frac{1}{n} \log\, P(\bar{X}_n^{(k)} \le 0) \ge -I_\mu(0) - \delta,$$

and the proof follows, since $\delta > 0$ can be made arbitrarily small. □

Theorem 1.1 says that if $A = (-\infty, x]$ with $x \le m$, or $A = [x, +\infty)$ with $x \ge m$, the sequence $n^{-1} \log \mu_n(A)$ converges to $-I_\mu(x)$, where μ_n denotes the law of $\bar{X}_n$. (If instead, $A = (-\infty, x]$ with $x > m$, or $A = [x, +\infty)$ with $x < m$, then $\log \mu_n(A)$ converges to 0, by the weak law of large numbers.) As we shall see, the non-negative function I_μ is convex and takes the mininum at m ($I_\mu(m) = 0$) so that the previous statement might be rephrased by saying that (1.1) and (1.2) hold for all such intervals, with $I = I_\mu$.

Except for the integrability in part (a), no further assumptions on the tails of the distribution of X_1 were imposed to derive Theorem 1.1. In particular, we could have the situation when $\hat{\mu}(\zeta) = +\infty$ for all $\zeta \neq 0$, i.e. none of the tails of the distribution decays exponentially. From the expression (1.10) it follows at once that in such cases the function I_μ will vanish identically (the expression on which we take the supremum will be zero for $\zeta = 0$ and $-\infty$ otherwise). One then asks about the information contained in Theorem 1.1 in such a case. The upper bound