Introduction

As Dembo and Zeitouni point out in the introduction to their monograph on the subject [1], there is no real theory of large deviations, but a variety of tools that allow asymptotic analysis of small probability.

To give an idea of what kind of *large deviations* we are talinkg about, let us consider a sequence of independent identical distributed real valued random variables $X_1, X_2, ..., X_n$ with mean zero and unit variance. Let $\hat{S}_n = \frac{1}{n} \sum_{i=1}^n X_i$ the empirical sums. The weak law of large numbers says that for any $\delta > 0$,

$$\mathbb{P}(|\hat{S}_n| \geqslant \delta) \longrightarrow_{n \to \infty} 0 \tag{1}$$

The central limit theorem is a refinement that says

$$\mathbb{P}(\sqrt{n}\hat{S}_n \in [a, b]) \longrightarrow_{n \to \infty} \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}\mathrm{d}x.$$

In the case $X_j \sim \mathcal{N}(0, 1)$, we have $\hat{S}_n \sim N(0, 1/n)$, and we can compute explicitly

$$\mathbb{P}(|\hat{S}_n| \geqslant \delta) = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} e^{-x^2/2}d\,x.$$

therefore (**exercise**)

$$\frac{1}{n}\log \mathbb{P}(|\hat{S}_n| \geqslant \delta) \longrightarrow_{n \to \infty} -\frac{\delta^2}{2} \tag{2}$$

Equation (2) is an example of a large deviation statement. Roughly it says that asymptotically in $n \to +\infty$, $\mathbb{P}(|\hat{S}_n| \geqslant \delta) \simeq e^{-n\delta^2/2}$.

Cramér's Theorem in $\mathbb{R}$

Let $\{X_n\}$ sequence of i.i.d. random variables on $\mathbb{R}$ with common probability distribution $\alpha(\mathrm{d}x)$. We define the moment generating function

$$M(\lambda) = \mathbb{E}\big[e^{\lambda X_1}\big] = \int_{\mathbb{R}} e^{\lambda x}\alpha(\mathrm{d}x) \tag{3}$$

and let us assume that there exists $\lambda^* > 0$ such that $M(\lambda) < \infty$ if $|\lambda| < \lambda^*$. Notice that, since $|x| \leqslant \lambda^{-1}(e^{\lambda x} + e^{-\lambda x})$ for any $\lambda > 0$, this condition implies that X_1 is integrable and we denote $m = \mathbb{E}(X_1) \in \mathbb{R}$. It is easy to see that $m = M'(0)$. We are interested in the *logarithmic moment generating function*

$$\Lambda(\lambda) = \log \mathbb{E}\big[e^{\lambda X_1}\big] \tag{4}$$

By Jensen's inequality, we have $\Lambda(\lambda) \geqslant \lambda m > -\infty$. Let $\mathcal{D}_\Lambda = \{\lambda \in \mathbb{R} : \Lambda(\lambda) < +\infty\}$. Under our hypothesis, $0 \in \mathcal{D}_\Lambda^o$ (the interior of $\mathcal{D}_\Lambda$).

Lemma 1. *The function Λ is convex and continuously differentiable in $\mathcal{D}_\Lambda^o$, moreover*

$$\Lambda'(\lambda) = \frac{\mathbb{E}(X_1 e^{\lambda X_1})}{M(\lambda)} \qquad \lambda \in \mathcal{D}_\Lambda^o.$$

Proof. For any $\alpha \in [0,1]$, it follows by Hölder inequality

$$\mathbb{E}(e^{(\alpha \lambda_1 + (1-\alpha)\lambda_2)X_1}) \leqslant M(\lambda_1)^\alpha M(\lambda_2)^{1-\alpha}$$

and consequently

$$\Lambda(\alpha \lambda_1 + (1-\alpha)\lambda_2) \leqslant \alpha \Lambda(\lambda_1) + (1-\alpha)\Lambda(\lambda_2).$$

The function $f_\epsilon(x) = (e^{(\lambda+\epsilon)x} - e^{\lambda x})/\epsilon$ converges point-wise to $x\,e^{\lambda x}$, and $|f_\epsilon(x)| \leqslant e^{\lambda x}(e^{\delta|x|} - 1)/\delta \leqslant e^{\lambda x}(e^{\delta x} + e^{-\delta x})/\delta = h(x)$, for every $|\epsilon| \leqslant \delta$. For any $\lambda \in \mathcal{D}_\Lambda^o$, there exists a $\delta > 0$ small enough such that $\mathbb{E}(h(X_1)) \leqslant M(\lambda + \delta) + M(\lambda - \delta) < +\infty$. Then the result follows by the dominated convergence theorem. $\square$

Using the same argument one can prove that $\Lambda \in \mathcal{C}^\infty(\mathcal{D}_\Lambda^o)$. Computing the second derivative we obtain

$$\Lambda''(\lambda) = \frac{\mathbb{E}(X_1^2 e^{\lambda X_1})}{M(\lambda)} - \left(\frac{\mathbb{E}(X_1 e^{\lambda X_1})}{M(\lambda)}\right)^2 \geqslant 0$$

by Jensen inequality. Observe that $\Lambda''(0) = \text{Var}(X_1)$. To avoid the trivial deterministic case, we assume that $\text{Var}(X_1) > 0$. It follows that $\Lambda''(\lambda) > 0$ for any $\lambda \in \mathcal{D}_\Lambda^o$, i.e. $\Lambda(\cdot)$ is strictly convex.

We define the rate function as the Fenchel-Legendre transform of Λ

$$I(x) = \sup_{\lambda \in \mathbb{R}} \{\lambda x - \Lambda(\lambda)\} \tag{5}$$

It is immediate to see that I is convex (as supremum of linear functions) and that $I(x) \geqslant 0$. Furthermore we have that $I(m) = 0$. In fact by Jensen's inequality $M(\lambda) \geqslant e^{\lambda m}$ for any $\lambda \in \mathbb{R}$, so that $\lambda m - \Lambda(\lambda) \leqslant 0$ and it is equal to 0 for $\lambda = 0$. We conclude that $I(m) = 0$.

Consequently m is a minimum of the convex positive function $I(x)$. It follows that $I(x)$ is non-decreasing for $x \geqslant m$ and non-increasing for $x \leqslant m$.

Observe that if $x > m$ and $\lambda < 0$ then $\lambda x - \Lambda(\lambda) \leqslant \lambda m - \Lambda(\lambda)$ and that implies

$$I(x) = \sup_{\lambda \geqslant 0} \{\lambda x - \Lambda(\lambda)\} \qquad x > m \tag{6}$$

Similarly one obtains

$$I(x) = \sup_{\lambda \leqslant 0} \{\lambda x - \Lambda(\lambda)\} \qquad x < m \tag{7}$$

Here are other important properties of I:

Lemma 2. $I(x) \to +\infty$ as $|x| \to \infty$, and its level sets are compact.

Proof. If $x > m$, for any positive $\lambda_+ \in \mathcal{D}_\Lambda$, $I(x)/x \geqslant \lambda_+ - \Lambda(\lambda_+)/x$ and $\lim_{x \to +\infty} \Lambda(\lambda_+)/x = 0$, so we have $\lim_{x \to +\infty} I(x)/x \geqslant \lambda_+$. A similar argument for $x < m$ gives $\lim_{x \to -\infty} I(x)/|x| \geqslant |\lambda_-|$. Consequently the level sets $\{x \in \mathbb{R}: I(x) \leqslant a\}$ are bounded, and closed by continuity of I. $\square$

We denote $\mathcal{D}_I = \{x \in \mathbb{R}: I(x) < \infty\}$.

Lemma 3. We have $x \in \mathcal{D}_I^o$ iff $\mathbb{P}(X_1 > x) > 0$ and $\mathbb{P}(X_1 < x) > 0$. For any $x \in \mathcal{D}_I^o$ there exists a unique $\lambda_x \in \mathcal{D}_\Lambda^o$ such that $x = \Lambda'(\lambda_x)$ and $I(x) = \lambda_x x - \Lambda(\lambda_x)$.

Proof. Consider $F_x(\theta) = \theta x - \Lambda(\theta) = -\log\mathbb{E}[e^{\theta(X_1-x)}]$. For any $\varepsilon > 0$ sufficiently small we have $\mathbb{P}(X_1 > x+\varepsilon) > 0$ and $\mathbb{P}(X_1 < x-\varepsilon) > 0$. Assume $x > m$ and $\theta \geqslant 0$ then we can estimate

$$\log\mathbb{E}[e^{\theta(X_1-x-\varepsilon)}] \geqslant \log\mathbb{E}[e^{\theta(X_1-x-\varepsilon)}1_{X_1>x+\varepsilon}] \geqslant \log\mathbb{P}(X_1 > x+\varepsilon)$$

so we have $I(x+\varepsilon) = \sup_{\theta\geqslant 0} F_{x+\varepsilon}(\theta) < +\infty$. By monotonicity we have $I(y) < +\infty$ for any $m < y < x+\varepsilon$ so $x \in \mathcal{D}_I^o$. A similar reasoning works for $x < m$.

Assume now that $\mathbb{P}(X_1 > x) = 0$ then for all $\varepsilon > 0$ $\mathbb{P}(X_1 < x+\varepsilon) = 1$ and $\mathbb{E}[e^{\theta(X_1-x-\varepsilon)}] \to 0$ for $\theta \to +\infty$ which gives $F_{x+\varepsilon}(\theta) \to +\infty$ and then $I(x+\varepsilon) = +\infty$. Since this is true for every $\varepsilon > 0$ we conclude that $x \notin \mathcal{D}_I^o$.

For fixed $x \in \mathcal{D}_I$ the function F_x is $C^2(\mathbb{R})$ and since $x \in \mathcal{D}_I^o$ we have that it exists $\varepsilon > 0$ for which $x + \varepsilon \in \mathcal{D}_I^o$ and so $\theta(x+\varepsilon) - \Delta(\theta) \leqslant I(x+\varepsilon) < +\infty$ which gives $F_x(\theta) \leqslant I(x+\varepsilon) - \varepsilon\theta \to -\infty$ for $\theta \to +\infty$. Similarly $F_x(\theta) \leqslant I(x-\varepsilon) + \varepsilon\theta \to -\infty$ for $\theta \to -\infty$. Then F_x has a unique maximum at a finite $\theta = \lambda_x$ and $F_x'(\lambda_x) = 0$, $F_x''(\lambda_x) < 0$. It follows that $I(x) = \lambda_x x - \Lambda(\lambda_x)$ and that $x = \Lambda'(\lambda_x)$. $\qquad\square$

We are ready to prove the following theorem, which is a large deviation statement for the empirical mean of a sequence of iid variables.

Theorem 4. (Cramér) *For any Borel set $A \subset \mathbb{R}$,*

$$-\inf_{x\in A^o} I(x) \leqslant \liminf_{n\to\infty} \frac{1}{n}\log\mathbb{P}(\hat{S}_n \in A) \leqslant \limsup_{n\to\infty} \frac{1}{n}\log\mathbb{P}(\hat{S}_n \in A) \leqslant -\inf_{x\in\bar{A}} I(x)$$

were A^o is the interior of A and $\bar{A}$ is the closure of A.

Proof.

Upper bound

Let us start with A a closed interval of the form $J_x = [x, +\infty)$ and let $x > m$. Then the exponential Chebycheff's inequality gives for any $\lambda > 0$

$$\mathbb{P}(\hat{S}_n \geqslant x) \leqslant e^{-n\lambda x}\mathbb{E}[e^{\sum_{i=1}^n \lambda X_i}] = e^{-n\lambda x}M(\lambda)^n = e^{-n(\lambda x - \Lambda(\lambda))}$$

Since $\lambda > 0$ is arbitrary, we can optimize the bound and obtain for $x > m$

$$\mathbb{P}(\hat{S}_n \geqslant x) \leqslant \exp\left(-n\sup_{\lambda>0}\{\lambda x - \Lambda(\lambda)\}\right) = \exp(-nI(x)) \tag{8}$$

where we use (6) in the last equality. Similarly for $x < m$ we obtain

$$\mathbb{P}(\hat{S}_n \leqslant x) \leqslant \exp\left(-n\sup_{\lambda<0}\{\lambda x - \Lambda(\lambda)\}\right) = \exp(-nI(x)) \tag{9}$$

Consider now an arbitrary closed set $C \subset \mathbb{R}$. If $m \in C$, then $I_C = \inf_{x\in C} I(x) = 0$ and the upper bound is trivial. If $m \notin C$ let $x_- = \sup\{x \in C : x < m\}$ and $x_+ = \inf\{x \in C : x > m\}$ and observe that by closedness $x_\pm \in C$ and that $\{\hat{S}_n \in C\} \Rightarrow \{\hat{S}_n \geqslant x_+ \text{ or } \hat{S}_n \leqslant x_-\}$. We have also $x_- < m < x_+$ and from the monotonicity of $I(x)$ on $(-\infty, x_-]$ and $[x_+, +\infty)$, $I_C = \min(I(x_-), I(x_+))$. Consequently, using (8) and (9),

$$\mathbb{P}(\hat{S}_n \in C) \leqslant \mathbb{P}(\hat{S}_n \geqslant x_+) + \mathbb{P}(\hat{S}_n \leqslant x_-) \leqslant e^{-nI(x_+)} + e^{-nI(x_-)} \leqslant e^{-nI_C}$$

and

$$\frac{1}{n}\log \mathbb{P}(\hat{S}_n \in C) \leqslant -I_C + \frac{1}{n}\log 2 \qquad (10)$$

which concludes the upper bound.

Lower bound

Given an open set G, it is enough to prove that for any $x \in G$

$$\liminf_{n\to\infty} \frac{1}{n}\log \mathbb{P}(\hat{S}_n \in G) \geqslant -I(x).$$

To this end, it is enough to prove that for any x and any $\delta > 0$,

$$\liminf_{n\to\infty} \frac{1}{n}\log \mathbb{P}(\hat{S}_n \in B_{x,\delta}) \geqslant -I(x)$$

where $B_{x,\delta} = (x - \delta, x + \delta)$. Clearly it is enough to consider $x \in \mathbb{R}$ such that $I(x) < \infty$. Assume $x \in \mathcal{D}_I^o$. Then by Lemma 3 there exists a unique $\lambda_x \in \mathcal{D}_\Lambda^o$ such that $I(x) = \lambda_x x - \Lambda(\lambda_x)$ and $x = \Lambda'(\lambda_x)$. Let us define the probability law on $\mathbb{R}$

$$\alpha_{\lambda_x}(\mathrm{d}y) = e^{\lambda_x y - \Lambda(\lambda_x)}\alpha(\mathrm{d}y)$$

Notice that

$$\int y\,\alpha_{\lambda_x}(\mathrm{d}y) = \Lambda'(\lambda_x) = x$$

Assuming $x \geqslant m$, we have also that $\lambda_x \geqslant 0$. Let $A_{n,\delta} = \{(x_1, ..., x_n)\colon (x_1 + ... + x_n)/n \in B_{x,\delta}\} \subset \mathbb{R}^n$, then for $\delta_1 < \delta$

$$\mathbb{P}(\hat{S}_n \in B_{x,\delta}) \geqslant \int_{A_{n,\delta_1}} \alpha(\mathrm{d}x_1)\cdots\alpha(\mathrm{d}x_n)$$

$$= \int_{A_{n,\delta_1}} e^{-\lambda_x(x_1+...+x_n)-n\Lambda(\lambda_x)}\alpha_{\lambda_x}(\mathrm{d}x_1)...\alpha_{\lambda_x}(\mathrm{d}x_n)$$

$$\geqslant e^{-n\lambda_x(x+\delta_1)-n\Lambda(\lambda_x)} \int_{A_{n,\delta_1}} \alpha_{\lambda_x}(\mathrm{d}x_1)\cdots\alpha_{\lambda_x}(\mathrm{d}\,x_n)$$

If $x < m$, we have $\lambda_x < 0$, and in the last step of the above we will have $x - \delta_1$ instead of $x + \delta_1$. By the law of large numbers, for any $\delta_1 > 0$

$$\int_{A_{n,\delta_1}} \alpha_{\lambda_x}(\mathrm{d}x_1)\cdots\alpha_{\lambda_x}(\mathrm{d}\,x_n) \longrightarrow_{n\to\infty} 1$$

so that

$$\liminf_{n\to\infty} \frac{1}{n}\log \mathbb{P}(\hat{S}_n \in B_{x,\delta}) \geqslant -[\lambda_x(x+\delta_1) - \Lambda(\lambda_x)] = -I(x) - \lambda_x\delta.$$

For any $\delta_1 < \delta$ we thus have

$$\liminf_{n\to\infty} \frac{1}{n}\log \mathbb{P}(\hat{S}_n \in B_{x,\delta}) \geqslant \liminf_{n\to\infty} \frac{1}{n}\log \mathbb{P}(\hat{S}_n \in B_{x,\delta_1}) = -I(x) - \lambda_x\delta_1.$$

4

and since $\delta_1 < \delta$ is arbitrary, we can let $\delta \to 0$, and this finish the proof. The proof for an arbitrary x is completed by observing that if $x \notin \mathcal{D}_I^o$ then either $\mathbb{P}(X_1 > x) = 0$ or $\mathbb{P}(X_1 < x) = 0$. Assume that $\mathbb{P}(X_1 < x) = 0$ then $F_x(\lambda) = \lambda x - \Lambda(\lambda) = -\log \mathbb{E}[e^{\lambda(X_1 - x)}]$ is a decreasing function of λ and

$$I(x) = \sup_{\lambda < 0} \{\lambda x - \Lambda(\lambda)\} = -\lim_{\lambda \to -\infty} \log \mathbb{E}[e^{\lambda(X_1 - x)}]$$

But for any $\varepsilon > 0$

$$\log \mathbb{P}(X_1 = x) \leqslant \log \mathbb{E}[e^{\lambda(X_1 - x)}] \leqslant \log \mathbb{P}(X_1 < x + \varepsilon) + e^{\lambda \varepsilon} \log \mathbb{P}(X_1 \geqslant x + \varepsilon) \to \log \mathbb{P}(X_1 < x + \varepsilon)$$

and taking $\varepsilon \to 0$ we get $I(x) = -\log \mathbb{P}(X_1 = x)$. Then we have

$$\mathbb{P}(\hat{S}_n \in B_{x,\delta}) \geqslant \mathbb{P}(X_1 = x, ..., X_n = x) = \mathbb{P}(X_1 = x)^n$$

and then

$$\liminf_{n \to \infty} \frac{1}{n} \log \mathbb{P}(\hat{S}_n \in B_{x,\delta}) \geqslant \log \mathbb{P}(X_1 = x) = -I(x)$$

concluding the proof. $\qquad\square$

Remark 5. Notice that the proof contains the non-asymptotic bound (10), i.e.

$$\forall n \geqslant 1, \quad \mathbb{P}(\hat{S}_n \in C) \leqslant 2e^{-n \inf_{x \in C} I(x)} \tag{11}$$

also called Chernoff's bound.

Remark 6. The lower bound was obtained by using the change of variable in conjunction with the law of large numbers for the new probabilities. One can get better bound by using the central limit theorem, and obtain the following corollary

Corollary 7.

$$\begin{aligned} \lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}(\hat{S}_n \geqslant x) &= -I(x) \qquad \text{if } x > m \\ \lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}(\hat{S}_n \leqslant x) &= -I(x) \qquad \text{if } x < m \end{aligned} \tag{12}$$

Proof. By the central limit theorem

$$\int_{\{x_1 + ... + x_n / n \in [x, x + \delta_1)\}} \alpha_{\lambda_o}(\mathrm{d}x_1) \cdots \alpha_{\lambda_o}(\mathrm{d}x_n) \longrightarrow_{n \to \infty} \frac{1}{2}$$

So in the proof of the lower bound one can substitute $(x - \delta, x + \delta)$ with $[x, x + \delta)$. Since $\mathbb{P}(\hat{S}_n \geqslant x) \geqslant \mathbb{P}(\hat{S}_n \in [x, x + \delta))$ one obtains

$$\liminf_{n \to \infty} \frac{1}{n} \log \mathbb{P}(\hat{S}_n \geqslant x) \geqslant -I(x)$$

The upper bound follows from the one in theorem 4. $\qquad\square$

Example 8.

1. Let α be the gaussian distribution with density

$$\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x-m)^2/2\sigma^2}$$

   then $I(x) = (x-m)^2/2\sigma^2$. In this case one can compute it directly, since $\hat{S}_n - n\,m$ has law $\mathcal{N}(0, \sigma^2/n)$.

2. $\alpha = \frac{1}{2}(\delta_0 + \delta_1)$ (Bernoulli). Then $M(\lambda) = \frac{1}{2}(1 + e^\lambda)$ and

$$I(x) = x\log x + (1-x)\log(1-x) + \log 2 \qquad \text{if } x \in [0,1]$$

   and $I(x) = +\infty$ otherwise.

3. For the exponential law $\alpha(d\,x) = \beta e^{-\beta x}1_{x\geqslant 0}\,d\,x$, we have $M(\lambda) = \beta/(\beta - \lambda)$ for $-\infty < \lambda < \beta$, otherwise $M(\lambda) = +\infty$. Then

$$I(x) = \beta x - 1 - \log(\beta x) \qquad \text{if } x > 0$$

   and $I(x) = +\infty$ if $x \leqslant 0$.

4. If ξ in a random variable with law $\mathcal{N}(0, 1/\beta)$, then ξ^2 has law $\chi^2(1)$, i.e. a gamma law $\Gamma(1/2, \beta/2)$, which has density

$$\frac{\beta^{1/2}}{\sqrt{2}\Gamma(1/2)}x^{-1/2}e^{-\beta x}$$

   Its moment generating function is $M(\lambda) = (\beta/(\beta - 2\lambda))^{1/2}$ if $\lambda < \beta/2$, otherwise equal to $+\infty$. The rate function results

$$I(x) = \frac{1}{2}\{\beta x - \log(\beta x) - 1\} \qquad \text{if } x > 0$$

   and $+\infty$ if $x < 0$.

Long rare segments in random walks

Consider the random walk $S_0 = 0$, $S_n = \sum_{i=1}^n X_i$, $n \geqslant 1$, where $(X_n)_{n\geqslant 1}$ is a sequence of iid random variables taking values in $\mathbb{R}^d$, $d \geqslant 1$. Let R_m be the maximal size of the intervals in which the empirical mean of the Xs belongs to some fixed measurable set $A \subset \mathbb{R}^d$:

$$R_m = \max\{\ell : \exists\, 0 \leqslant \ell \leqslant k \leqslant m, (S_k - S_{k-\ell})/\ell \in A\}$$

and similarly, let T_r be the first time when the empirical mean over stretches of size at least r belongs to A:

$$T_r = \inf\{n : \exists\, r \leqslant \ell \leqslant n, (S_n - S_{n-\ell})/\ell \in A\}.$$

Clearly T_r is a stopping time for all $r \geqslant 0$. Note moreover that $\{R_m \geqslant r\} \Leftrightarrow \{T_r \leqslant m\}$. We want to prove the following result.

Theorem 9. *Assume that A is such that*

$$\lim_{n\to+\infty} -\frac{1}{n}\log\mathbb{P}(\hat{S}_n \in A) = I_A \tag{13}$$

6

exists. Then almost surely

$$\lim_{m\to+\infty}\frac{R_m}{\log m}=\lim_{r\to+\infty}\frac{r}{\log T_r}=\frac{1}{I_A}.$$

Proof. Let $C_{k,\ell}=\{(S_k-S_{k-\ell})/\ell\in A\}$, then

$$\mathbb{P}(T_r\leqslant m)\leqslant\sum_{k=0}^{m}\sum_{\ell=r}^{k}\mathbb{P}(C_{k,\ell})=\sum_{k=0}^{m}\sum_{\ell=r}^{k}\mathbb{P}(\hat{S}_\ell\in A)\leqslant m\sum_{\ell\geqslant r}\mathbb{P}(\hat{S}_\ell\in A).$$

Assume $I_A\in(0,+\infty)$ then for all $0<\varepsilon<I_A$ and r large enough, by (13) we have

$$\mathbb{P}(T_r\leqslant m)\leqslant m\sum_{\ell\geqslant r}e^{-\ell(I_A-\varepsilon)}\leqslant m\,e^{-r(I_A-\varepsilon)}C_\varepsilon$$

and letting $m=\lfloor e^{r(I_A-2\varepsilon)}\rfloor$ we get

$$\sum_{r\geqslant1}\mathbb{P}(T_r\leqslant e^{r(I_A-2\varepsilon)}-1)\leqslant\sum_{r\geqslant1}e^{r(I_A-2\varepsilon)}e^{-r(I_A-\varepsilon)}C_\varepsilon=\sum_{r\geqslant1}e^{-r\varepsilon}C_\varepsilon<+\infty$$

and by Borel-Cantelli

$$\liminf_{r\to+\infty}\frac{\log T_r}{r}\geqslant I_A-2\varepsilon\,,\qquad a.s.$$

Being ε arbitrary we obtain that the liminf is $\geqslant I_A$. If $I_A=+\infty$ the proof is complete. Otherwise to establish the reverse inequality we consider the probability of the event $\{T_r>m\}$ for large m. Now

$$\{T_r\leqslant m\}=\{R_m\geqslant r\}\supseteq\cup_{1\leqslant k<m/r}C_{kr,r}$$

and the family of events $(C_{kr,r})_{1\leqslant k\leqslant m/r}$ are independent so for all $\varepsilon>0$ and r large enough

$$\mathbb{P}(T_r>m)\leqslant1-\mathbb{P}(\cup_{1\leqslant k<m/r}C_{kr,r})=\mathbb{P}(\cap_{1\leqslant k<m/r}C^c_{kr,r})$$

$$=\prod_{1\leqslant k<m/r}(1-\mathbb{P}(C_{kr,r}))=(1-\mathbb{P}(\hat{S}_r\in A))^{\lfloor m/r\rfloor}\leqslant e^{-\lfloor m/r\rfloor\mathbb{P}(\hat{S}_r\in A)}$$

$$\leqslant e^{-\lfloor m/r\rfloor e^{-r(I_A+\varepsilon)}}$$

so choosing m such that $m=e^{r(I_A+2\varepsilon)}$ we get

$$\sum_{r\geqslant1}\mathbb{P}(T_r>e^{r(I_A+2\varepsilon)})\leqslant\sum_{r\geqslant1}\exp\big(-(e^{r(I_A+2\varepsilon)}/r-1)e^{-r(I_A+\varepsilon)}\big)\leqslant\sum_{r\geqslant1}\exp(1-e^{r\varepsilon}/r)<+\infty$$

and exploiting again the Borel-Cantelli lemma we obtain

$$\limsup_{r\to+\infty}\frac{\log T_r}{r}\leqslant I_A+2\varepsilon\quad a.a.$$

which finally allows us to conclude that

$$\lim_{r\to+\infty}\frac{\log T_r}{r}=I_A,\quad a.s.$$

7

and by the duality of the events $\{T_r \geqslant m\}$ and $\{R_m \leqslant r\}$ we obtain also the corresponding statements for R_m. $\qquad\square$

Remark 10. Condition (13) is a typical consequence of large deviation statements. If Λ is the logarithmic mgf of X_1 by Cramér's theorem the limit in (13) exists whenever

$$I_A = \inf_{x \in A^o} \Lambda^*(x) = \inf_{x \in \bar{A}} \Lambda^*(x).$$

Bibliography

[1] Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications*, volume 38 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition, 1998.

[2] R. Durrett. *Probability: Theory and Examples*. Duxbury Press, third edition, 2005.

[3] Ivar Ekeland and Roger Temam. *Convex analysis and variational problems*. North-Holland Publishing Co., Amsterdam, 1976. Translated from the French, Studies in Mathematics and its Applications, Vol. 1.