Kolmogorov's Maximal Inequality and Applications of SLLN

6.1 Kolmogorov's maximal inequality

Theorem 6.1 (Kolmogorov's Maximal Inequality). *Let $X_1, \dots, X_n$ be independent random variables with $\mathbb{E}\, X_i = 0, \mathbb{E}\, X_i^2 < \infty$ for all $1 \leqslant i \leqslant n$. Define $S_n = X_1 + \cdots + X_n$. Then*

$$\mathbb{P}\Big(\max_{1 \leqslant k \leqslant n} |S_k| \geqslant a) \leqslant \frac{\mathbb{E}\, S_n^2}{a^2}$$

for any $a > 0$.

Proof Sketch: Let $A_i = \{|S_1| < a, |S_2| < a, \dots, |S_{i-1}| < a, |S_i| \geqslant a\},\ i = 1, 2, \dots, n$. Then $A_i \cap A_j = \phi$ for all $i \neq j$. Notice that $\mathbb{P}\Big(\max_{1 \leqslant k \leqslant n} |S_k| \geqslant a\Big) = \sum_{i=1}^n \mathbb{P}(|S_i| \geqslant a, \max_{1 \leqslant k < i} |S_k| < a)$. Then

$$\begin{aligned}\mathbb{E}\, S_n^2 &\geqslant \sum_{i=1}^n \mathbb{E}\, S_n^2 \mathbb{1}_{A_i} = \sum_{i=1}^n \mathbb{E}(S_i^2 + 2S_i(S_n - S_i) + (S_n - S_i)^2)\mathbb{1}_{A_i} \\ &\geqslant \sum_{i=1}^n (a^2\, \mathbb{P}(A_i) + 2\, \mathbb{E}\, S_i \mathbb{1}_{A_i}(S_n - S_i)) = a^2\, \mathbb{P}\Big(\max_{1 \leqslant k \leqslant n} |S_k| \geqslant a\Big).\end{aligned}$$

In the last line we used the fact that $S_i \mathbb{1}_{A_i}$ is independent of $X_{i+1}, \dots, X_n$. ■

Theorem 6.2 (Basic $\mathbf{L^2}$-convergence). *Let $X_1,\ X_2,\ \cdots$ be independent r.v.'s with $\mathbb{E}(X_i) = 0$ and $\sum_{i=1}^\infty \mathrm{Var}(X_i) < \infty$. Then $S_n = \sum_{i=1}^n X_i$ converges in L^2 and a.s.*

Proof. To see L^2-convergence, we will prove that $(S_n)_{n \geqslant 1}$ is Cauchy in L^2. It is enough to show that $||S_n - S_m||_2 < \varepsilon$ for all $n, m \geqslant N(\varepsilon)$. Suppose $n > m$, then we obtain

$$\begin{aligned}||S_n - S_m||_2^2 = \mathbb{E}(S_n - S_m)^2 &= \mathbb{E}(X_{m+1} + X_{m+2} + \cdots + X_n)^2 \\ &= \mathrm{Var}(X_{m+1} + X_{m+2} + \cdots + X_n) \\ &= \mathrm{Var}(X_{m+1}) + \mathrm{Var}(X_{m+2}) + \cdots + \mathrm{Var}(X_n).\end{aligned}$$

For any $\varepsilon > 0$, there exists $N = N(\varepsilon)$ with $\sum_{i=N}^\infty \mathrm{Var}(X_i) < \varepsilon^2$, thus we have $||S_n - S_m||_2 < \varepsilon$ for all $n, m \geqslant N(\varepsilon)$ which implies that $(S_n)_{n \geqslant 1}$ is a Cauchy sequence in L^2 and therefore converges in L^2. Now, it remains to show *a.s.* convergence. In other words, we need to show $\mathbb{P}(S_n \text{ converges}) = 1$. Define $M_N = \max\{|S_n - S_m| : n, m \geqslant N\}$, then clearly M_N is decreasing. Indeed, it is enough to show that $M_N \xrightarrow{\mathbb{P}} 0$ as $N \to \infty$. Let $\varepsilon > 0$ be given. Then,

$$\begin{aligned}
\mathbb{P}(M_N > \varepsilon) = \mathbb{P}(\max_{n,m\geqslant N} |S_n - S_m| > \varepsilon) &\leqslant \mathbb{P}(\max_{k>N} |S_k - S_N| > \varepsilon/2) \\
&= \lim_{n\to\infty} \mathbb{P}(\max_{1\leqslant k\leqslant n} |S_{N+k} - S_N| > \varepsilon/2) \\
&\leqslant \lim_{n\to\infty} \frac{\mathrm{Var}(S_{N+n} - S_N)}{(\varepsilon/2)^2} \\
&= \lim_{n\to\infty} \left(\frac{2}{\varepsilon}\right)^2 \cdot \sum_{i=N+1}^{N+n} \mathrm{Var}(X_i) = \left(\frac{2}{\varepsilon}\right)^2 \cdot \sum_{i>N} \mathrm{Var}(X_i),
\end{aligned}$$

where the second inequality follows by Kolmogorov's Maximal inequality. Hence, $\lim_{N\to\infty} \mathbb{P}(M_N > \varepsilon) = 0$ or $M_N \xrightarrow{\mathbb{P}} 0$, thus we obtain $M_N \to 0$ *a.s.* Therefore, $\mathbb{P}((S_n)_{n\geqslant 1}$ is Cauchy$) = 1$. ■

Using basic L^2-convergence theorem, we can give a simple proof of SLLN under finite second moment. Let $X_1, X_2, \ldots$ be i.i.d. with $\mathbb{E}\, X_1 = 0$ and $\mathrm{Var}(X_1) < \infty$. By basic L^2-convergence theorem, we have $\sum_{k=1}^n X_k/k$ converges a.s as $n \to \infty$. Now, using Kronecker's lemma (stated and proved below), we have $n^{-1} \sum_{k=1}^n X_k \to 0$ a.s.

Lemma 6.3 (Kronecker's lemma). *Ler $a_n \uparrow \infty$ and $\sum_{i=1}^n \frac{x_i}{a_i}$ converges. Then*

$$\frac{1}{a_n} \sum_{i=1}^n x_i \to 0$$

Proof. We have $b_n := \sum_{i=1}^n \frac{x_i}{a_i} \to b,$ for some $b \in \mathbb{R}$, as $n \to \infty$. Moreover, $x_n = a_n(b_n - b_{n-1})$ so

$$\frac{1}{a_n} \sum_{i=1}^n x_i = \frac{1}{a_n} \sum_{i=1}^n (a_i b_i - a_i b_{i-1}) = \frac{1}{a_n}(a_n b_n + \sum_{i=1}^{n-1} (a_i - a_{i+1}) b_i) = b_n - \frac{1}{a_n} \sum_{i=1}^{n-1} (a_{i+1} - a_i) b_i.$$

Let $d_i = a_{i+1} - a_i$. We know that $b_n \to b$ and thus

$$\frac{\sum_{i=1}^{n-1} d_i b_i}{\sum_{i=1}^{n-1} d_i} - b = \frac{\sum_{i=1}^{n-1} d_i (b_i - b)}{\sum_{i=1}^{n-1} d_i} \to 0$$

as $n \to \infty$, since $d_i \geqslant 0$ and $\sum_{i\geqslant 1} d_i = \infty$. So this proves that $\frac{1}{a_n} \sum_{i=1}^n x_i \to 0$ completing the proof. ■

6.2 Applications of SLLN

Theorem 6.4 (Renewal Theorem). *Let $X_1, X_2, \ldots$ be a sequence of i.i.d. r.v.'s with $X_i \geqslant 0$, $\mathbb{E}\, X_1 = \mu > 0$. Then*

$$\frac{N_t}{t} \xrightarrow[t\to\infty]{a.s.} \frac{1}{\mu}$$

where $N_t = \sup\{n \mid X_1 + X_2 + \cdots + X_n \leqslant t\}$.

Proof Sketch: By SLLN, if $S_n = X_1 + X_2 + \cdots X_n$, then $\frac{S_n}{n} \xrightarrow{a.s.} \mu$ as $n \to \infty$. We observe that

- $N_t \uparrow \infty$ a.s. as $t \to \infty$. ($\{N_t \geqslant n\} = \{X_1 + X_2 + \cdots X_n \leqslant t\}$).
- $\mathbb{P}\left(\frac{S_n}{n} \xrightarrow{n\to\infty} \mu, N_t \xrightarrow{t\to\infty} \infty\right) = 1$ or $\mathbb{P}\left(\frac{S_{N_t}}{N_t} \xrightarrow{t\to\infty} \mu, \frac{S_{N_t+1}}{N_t+1} \xrightarrow{t\to\infty} \mu\right) = 1$.

By definition, we have $S_{N_t} \leqslant t < S_{N_t+1}$ so that

$$\frac{S_{N_t}}{N_t} \leqslant \frac{t}{N_t} < \frac{S_{N_t+1}}{N_t} = \frac{S_{N_t+1}}{N_t+1} \cdot \frac{N_t+1}{N_t}.$$

Both upper and lower bounds converges to μ a.s. and this completes the proof. ■

Definition 6.5 (Empirical Distribution). *Let X_1, X_2, $\cdots$ be a sequence of i.i.d. r.v.'s with distribution function F. Define*

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{X_i \leqslant x} \text{ for all } x \in \mathbb{R}.$$

F_n is called the empirical distribution function. By SLLN, $F_n(x) \xrightarrow{a.s.} F(x)$, for each $x \in \mathbb{R}$.

Theorem 6.6 (Glivenko-Cantelli Lemma). *We have*

$$||F_n - F||_\infty = \sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \xrightarrow{a.s.} 0.$$

Proof Sketch: Fix $k \geqslant 1$, define

$$x_{j,k} = F^{-1}\left(\frac{j}{k}\right) = \inf\left\{x : F(x) \geqslant \frac{j}{k}\right\}, \ 1 \leqslant j < k.$$

Define $x_{0,k} = -\infty, x_{k,k} = \infty$. By SLLN,

$$\mathbb{P}\left(F_n(x_{j,k}) \xrightarrow{a.s.} F(x_{j,k}) \ \forall \ 0 \leqslant j \leqslant k\right) = 1.$$

Similarly,

$$F_n(x-) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{X_i < x} \xrightarrow{a.s.} F(x-).$$

Thus

$$\mathbb{P}\left(F_n(x_{j,k}-) - F(x_{j,k}-) \to 0, \ F_n(x_{j,k}) - F(x_{j,k}) \to 0 \ \forall \ 0 \leqslant j \leqslant k\right) = 1.$$

Define

$$\Delta_k^{(n)} = \max_{0\leqslant j\leqslant k} \left\{|F_n(x_{j,k}) - F(x_{j,k})| \ , \ |F_n(x_{j,k}-) - F(x_{j,k}-)|\right\}.$$

Take x in $[x_{j,k}, x_{j+1,k})$. Then,

$$F_n(x_{j,k}) \leqslant F_n(x) \leqslant F_n(x_{j+1,k}-) \implies F_n(x_{j,k}) - F(x) \leqslant F_n(x) - F(x) \leqslant F_n(x_{j+1,k}-) - F(x).$$

Since

$$\begin{aligned}|F_n(x_{j+1,k}-) - F(x)| &\leqslant |F_n(x_{j+1,k}-) - F(x_{j+1,k}-)| + |F(x_{j+1,k}-) - F(x)| \\ &\leqslant \Delta_k^{(n)} + \frac{1}{k}\end{aligned}$$

and

$$|F_n(x_{j,k}) - F(x)| \leqslant |F_n(x_{j,k}) - F(x_{j,k})| + |F(x_{j,k}) - F(x)| \leqslant \Delta_k^{(n)} + \frac{1}{k},$$

we have $\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \leqslant \Delta_k^{(n)} + \frac{1}{k}$. Note that $\Delta_k^{(n)} \xrightarrow{a.s.} 0$. Thus

$$\begin{aligned} &\mathbb{P}\left(\limsup_n \; ||F_n - F||_\infty \leqslant \frac{1}{k}\right) = 1 \; \forall k \\ \implies &\mathbb{P}\left(\limsup_n ||F_n - F||_\infty \leqslant \frac{1}{k} \; \forall k\right) = 1, \\ \implies &\mathbb{P}\left(\limsup_n ||F_n - F||_\infty \; = \; 0\right) = 1. \end{aligned}$$

∎

Remark. See "Bootstrap" in statistics for applications of Glivenko-Cantelli Lemma.

6.3 Tail events and Kolmogorov's 3-Series Theorem

Definition 6.7 (Tail σ-field). *Let $X_1, X_2, \ldots$ be independent r.v.s. The σ-field*

$$\mathcal{T} = \bigcap_{k=1}^{\infty} \sigma(X_k, X_{k+1}, \ldots)$$

is called the tail σ-field.

Example 6.8. *We have $\{\lim X_n \text{ exists}\} \in \mathcal{T}$, $\{\limsup_n (X_1 + \ldots + X_n)/n \leqslant a\} \in \mathcal{T}$. However, the event $\{\limsup_n (X_1 + \ldots + X_n) \leqslant a\} \notin \mathcal{T}$.*

Theorem 6.9 (Kolmogorov's 0-1 Law). *For all $A \in \mathcal{T}$, $\mathbb{P}(A) \in \{0, 1\}$.*

Proof. We will prove that $\mathcal{T} \perp \mathcal{T}$ (the tail σ-field is independent of itself). Then for all $A \in \mathcal{T}$,

$$\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A) \cdot \mathbb{P}(A)$$

which implies $\mathbb{P}(A) \in \{0, 1\}$. The proof is in 4 steps.

Step 1. $\sigma(X_1, \ldots, X_{k-1}) \perp \sigma(X_{l+1}, X_{l+2}, \ldots)$ for all $l \geqslant k$, since X_i's are independent.

Step 2. In particular, $\sigma(X_1, \ldots, X_{k-1}) \perp \mathcal{T}$ for all $k \geqslant 2$.

Step 3. $\sigma(X_1, \ldots) \perp \mathcal{T}$.

Step 4. $\mathcal{T} \perp \mathcal{T}$.

Proof of step 1: Clearly $\sigma(X_1, \ldots, X_{k-1}) \perp \sigma(X_l, \ldots, X_{l+j})$ for all $l \geqslant k, j \geqslant 0$, which implies that

$$\sigma(X_1, \ldots, X_{k-1}) \perp \bigcup_{j=0}^{\infty} \sigma(X_l, \ldots, X_{l+j}).$$

By $\pi-\lambda$ theorem, $\sigma(X_1,\ldots,X_{k-1}) \perp \sigma(\cup_{j=0}^{\infty}\sigma(X_l,\ldots,X_{l+j})) = \sigma(X_l,X_{l+1},\ldots)$.

Proof of step 2: Note that $\mathcal{F} \perp \mathcal{G}$ implies that any sub σ-field of $\mathcal{F}$ is independent of $\mathcal{G}$. $(*)$

Clearly $\mathcal{T} \subseteq \sigma(X_k, X_{k+1},\ldots)$. So by $(*)$, $\sigma(X_1,\ldots,X_{k-1}) \perp \mathcal{T}$.

Proof of step 3: We have $\bigcup_{k\geqslant 1}\sigma(X_1,\ldots,X_{k-1}) \perp \mathcal{T}$. Thus by $\pi-\lambda$ theorem, $\sigma(X_1,\ldots) \perp \mathcal{T}$.

Proof of step 4: Clearly $\mathcal{T} \subseteq \sigma(X_1, X_2,\ldots)$. So by $(*)$, we have the proof. ■

Theorem 6.10 (Kolmogorov's 3-series Theorem). *Let $X_1, X_2, \ldots$ be independent. Fix $b > 0$. Consider the three deterministic series:*

$$\text{(i)} \sum_{i=1}^{n} \mathbb{P}(|X_i| > b), \qquad \text{(ii)} \sum_{i=1}^{n} \mathbb{E}(X_i \mathbb{1}_{|X_i|\leqslant b}), \qquad \text{(iii)} \sum_{i=1}^{n} \operatorname{Var}(X_i \mathbb{1}_{|X_i|\leqslant b}).$$

Then

$$\sum_{i=1}^{n} X_i \text{ converges a.s. if and only if } (i),(ii),(iii) \text{ converge.}$$

Proof. We will first prove the if part. Write $X_i = U_i + v_i + W_i$, where

$$U_i := X_i\mathbb{1}_{|X_i|>b},\ v_i := \mathbb{E}\, X_i\mathbb{1}_{|X_i|\leqslant b},\ \text{and } W_i := X_i\mathbb{1}_{|X_i|\leqslant b} - v_i.$$

By convergence of (ii), we have $\sum_{i\geqslant 1} v_i < \infty$. We will prove that both $\lim_{n\to\infty}\sum_{i=1}^{n} U_i$ and $\lim_{n\to\infty}\sum_{i=1}^{n} W_i$ exists a.s. By Borel-Cantelli Lemma and convergence of series (i), we have $\mathbb{P}(|X_i| > b \text{ i.o.}) = 0$ which implies that $\lim_{n\to\infty}\sum_{i=1}^{n} U_i$ exists a.s. Convergence of (iii) and Basic L^2 convergence implies $\lim_{n\to\infty}\sum_{i=1}^{n} W_i$ exists a.s.

For the only if part, we need to use Central Limit Theorem, which states that for a sequence of iid mean zero, variance one rvs $X_1, X_2, \ldots$, the sequence $n^{-1/2}S_n$ converges in distribution to $\mathrm{N}(0,1)$ and will be proved later. ■

Another application of Central Limit Theorem and truncation + subsequence technique is the following result, which will not be proved here. The crucial idea is that $S_n/\sqrt{n}$ converges in distribution to $\mathrm{N}(0,1)$ and for a rv $X \sim \mathrm{N}(0,1)$ we have $\mathbb{P}(X \geqslant t) = \mathbb{P}(X \leqslant -t) \leqslant e^{-t^2/2}, t > 0$.

Theorem 6.11 (Law of Iterated Logarithms (LIL)). *Let $X_1, X_2, \ldots$ be i.i.d. with mean 0,* $\operatorname{Var}(X_1) = 1$. *Define $S_n = \sum_{i=1}^{n} X_i$. Then*

$$\limsup_n \frac{S_n}{\sqrt{2n\log\log n}} = 1 \ a.s.$$
$$\liminf_n \frac{S_n}{\sqrt{2n\log\log n}} = -1 \ a.s.$$

A sequence $\{t_n\}_{n\geqslant 1}$ is **subadditive**, if $t_{m+n} \leqslant t_m + t_n$ for all $m, n \geqslant 1$. The following lemma for a deterministic sequence can be generalize to random sequences to give another proof of SLLN.

Theorem 6.12 (Subadditive limit theorem). *If $\{t_n\}_{n\geqslant 1}$ is subadditive, then*

$$\frac{t_n}{n} \to \inf_{m\geqslant 1} \frac{t_m}{m} \text{ as } n \to \infty.$$

Proof. Clearly, $\liminf_n t_n/n \geqslant \inf_{m\geqslant 1} t_m/m$. Note that, $t_{mn+k} \leqslant nt_m + t_k$ for $m, n, k \geqslant 1$. Thus

$$\frac{t_{mn+k}}{mn+k} \leqslant \frac{nt_m + t_k}{mn+k} = \frac{\frac{t_m}{m} + \frac{t_k}{mn}}{1 + \frac{k}{mn}}$$

Fix $m \geqslant 1$ and $0 \leqslant k < m$, then

$$\limsup_n \frac{t_{mn+k}}{mn+k} \leqslant \frac{t_m}{m}.$$

Thus,

$$\limsup_n \frac{t_n}{n} \leqslant \frac{t_m}{m}.$$

Since m is arbitrary, we have

$$\limsup_{n\to\infty} \frac{t_n}{n} \leqslant \inf_{m\geqslant 1} \frac{t_m}{m}.$$

■

Corollary 6.13. *Let $X_1, X_2, \ldots$ be i.i.d. r.v.s. Define $S_n = X_1 + X_2 + \cdots + X_n, n \geqslant 1$. Then,*

$$\kappa(a) := \lim_{n\to\infty} -\frac{1}{n} \log \mathbb{P}(S_n \geqslant a) \text{ exists and is non-negative (need not be finite).}$$

Proof. We use the fact that $\mathbb{P}(S_{n+m} \geqslant (n+m)a) \geqslant \mathbb{P}(S_m \geqslant ma, S_{n+m} - S_m \geqslant na) \geqslant \mathbb{P}(S_m \geqslant ma) \cdot \mathbb{P}(S_n \geqslant na)$ and thus the sequence $t_n := -\log \mathbb{P}(S_n \geqslant a) \geqslant 0$ is subadditive. ■

In the next lecture, we will study conditions under which $\kappa(a)$ is positive and finite. The random version of the subadditive limit theorem is stated below without proof.

Theorem 6.14 (Subadditive ergodic theorem). *Let $\{X_{m,n} \mid n > m \geqslant 0\}$ be a collection of rvs indexed by integers $n > m \geqslant 0$, such that*

$$X_{0,n} \leqslant X_{0,m} + X_{m,n} \text{ a.s. for } 0 < m < n.$$

Assume that

(a) *The joint distributions of $\{X_{m+1,m+k+1}, k \geqslant 1\}$ are the same as those of $\{X_{m,m+k}, k \geqslant 1\}$ for each $m \geqslant 0$,*

(b) *For each $k \geqslant 1$, $\{X_{nk,(n+1)k}, n \geqslant 1\}$ is a stationary process and*

(c) *For each n, $\mathbb{E}|X_{0,n}| < \infty$ and $\mathbb{E} X_{0,n} \geqslant -cn$ for some constant c.*

Then, $\frac{1}{n}X_{0,n}$ converges a.s. and in L^1 to a r.v. X and $\mathbb{E} X \in [-c, \infty)$. Moreover, if the processes in (b) are ergodic, then X is a constant a.s.

6.4 Large Deviation Principle

Let $X_1, X_2, \ldots$ be i.i.d. r.v.s with mean μ and $S_n := X_1 + X_2 + \cdots + X_n, n \geqslant 1$. By SLLN, $S_n/n \to 0$ a.s. and thus for any $a > \mu$, $\lim_{n\to\infty} \mathbb{P}(S_n/n \geqslant a) = 0$. By previous result, we know that $\kappa(a) := -\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n/n \geqslant a)$ exists and $\kappa(a) \in [0, \infty]$. One can also prove the following.

Exercise 6.15. *The following are equivalent: (a) $\kappa(a) = \infty$, (b) $\mathbb{P}(X_1 \geqslant a) = 0$, and (c) $\mathbb{P}(S_n \geqslant na) = 0$ for all $n \geqslant 1$.*

We will use Large Deviation Principle (LDP) to find $\kappa(\cdot)$. Assume that

$$m(\theta) := \mathbb{E}\, e^{\theta X_1} < \infty \text{ if } \theta \in \Theta \text{ and is infinite if } \theta \in \bar{\Theta}^c,$$

where Θ is an open interval containing 0. Using Hölder's inequality, we can prove that $\log m(\theta)$ is a convex function (and thus Θ must be an interval).

The Legendre transform of $\log m(\cdot)$ is defined as,

$$\ell(a) := \sup_{\theta \in \Theta} (a\theta - \log m(\theta)) \geqslant 0$$

which is also convex in a. If $m(\cdot)$ is differentiable in Θ, then $\ell(a) = a\theta^* - \log m(\theta^*)$ where θ^* satisfies $m'(\theta^*) = am(\theta^*)$.

Theorem 6.16. *Let $X_1, X_2, \ldots$ be i.i.d. r.v.s with mean μ and the function $\ell(\cdot)$ is defined as above. Then for $a > \mu$, we have*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n/n \geqslant a) = -\ell(a).$$

Same result holds for $\mathbb{P}(S_n/n \leqslant a)$ with $a < \mu$.

Proof. We will only prove the upper bound for $\mathbb{P}(S_n/n \geqslant a)$. The lower bound is bit complicated and involves change of measure. Wlog, we can assume that $\mu = 0$, o.w. work with $X_1 - \mu$. By Markov's inequality, we have

$$\mathbb{P}(S_n \geqslant an) = \mathbb{P}(e^{\theta S_n} \geqslant e^{\theta an}) \leqslant e^{-a\theta n}\, \mathbb{E}\, e^{\theta S_n} = e^{-a\theta n} \prod_{i=1}^{n} \mathbb{E}\, e^{\theta X_i} = e^{-a\theta n}(m(\theta))^n,$$

for $\theta \in [0,\infty) \cap \Theta$. This implies that, $\log \mathbb{P}(S_n/n \geqslant a) \leqslant -n(a\theta - \log m(\theta))$ for $\theta \in [0,\infty) \cap \Theta$. Hence

$$\frac{1}{n} \log \mathbb{P}(S_n/n \geqslant a) \leqslant - \sup_{\theta \in [0,\infty)\cap\Theta} (a\theta - \log m(\theta))$$

Note that, by Jensen's inequality we have $\log m(\theta) \geqslant 0$ for all $\theta \in \Theta$ and for $a > 0, \theta < 0$ we have $\theta a - \log m(\theta) \leqslant 0$. Thus, we have

$$\sup\nolimits_{\theta \in [0,\infty)\cap\Theta} (a\theta - \log m(\theta)) = \sup\nolimits_{\theta\in\Theta} (a\theta - \log m(\theta)) = \ell(a).$$

■

Example 6.17. *Let $X_1, X_2, \ldots$ be i.i.d. $N(0,1)$ r.v.s with mean 0, variance 1 and let $S_n := X_1 + X_2 + \cdots + X_n, n \geqslant 1$. Then*

$$m(\theta) = \mathbb{E}\, e^{\theta X_1} = e^{\theta^2/2} \text{ for } \theta \in \mathbb{R}.$$

Thus,

$$\ell(a) = \sup_{\theta\in\mathbb{R}} (a\theta - \log m(\theta)) = \sup_{\theta\in\mathbb{R}} (a\theta - \theta^2/2) = a^2/2$$

for $a \in \mathbb{R}$. Thus, for $a > 0$ we have

$$\frac{1}{n} \log \mathbb{P}(S_n/n \geqslant a) \to -a^2/2.$$

Example 6.18. *Let $X_1, X_2, \dots$ be i.i.d. Poisson(λ) r.v.s with mean $\lambda > 0$ and let $S_n := X_1 + X_2 + \cdots + X_n, n \geqslant 1$. Then*

$$m(\theta) = \mathbb{E}\, e^{\theta X_1} = e^{\lambda(e^\theta - 1)} \text{ for } \theta \in \mathbb{R}.$$

Thus,

$$\ell(a) = \sup_{\theta \in \mathbb{R}} (a\theta - \log m(\theta)) = \sup_{\theta \in \mathbb{R}} (a\theta - \lambda(e^\theta - 1)) = a \log(a/e\lambda) + \lambda$$

for $a > 0$. Thus, for $t > 1$ we have

$$\frac{1}{n} \log \mathbb{P}(S_n/n \geqslant t\lambda) \to -\lambda(t \log t - t + 1).$$