MATH 6480/STAT 9300/AMCS 6481, Fall 2023, Homework IX

1. Let $\{X_n\}$ be IID Poisson with mean 1, and take $a > 1$ find the limit

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\sum_{k=1}^{n} X_k \geq an\right)$$

<u>*solution*</u>

The moment generating function of Poiss(1) is

$$M(\theta) = \mathbb{E}e^{\theta X} = \sum_{x \geqslant 0} \mathbb{E}\left(e^{\theta x} \mid X = x\right) \mathbb{P}(X = x) = e^{\left(e^{\theta}-1\right)}$$

Given the upper bound which has been proved in class

$$\mathbb{P}(S_n \geq na) \leq e^{-n(a\theta - \kappa(\theta))}$$

we need to find the maximum of $a\theta - \kappa(\theta)$ to optimize the upper bound, i.e., the least upper bound. Let $\kappa(\theta) = \log M(\theta) = e^{\theta} - 1$ and $\kappa'(\theta) = e^{\theta}$, notice that $\kappa'(\theta)$ is increasing and continuous so the maximum will achieve at the solution of $d(a\theta - \kappa(\theta))/d\theta = 0$, that is, $a - \kappa'(\theta) = 0$
Solving $\kappa'(\theta_a) = a$, we have $\theta_a = \log a > 0$, then

$$\gamma(a) = -a\theta_a + \kappa(\theta_a) = -a \log + e^{\log a} - 1 = -a \log a + a - 1$$

By Lemma 2.7.1(Thm 1.1),

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\sum_{k=1}^{n} X_k \geq an\right) = -a \log a + a - 1$$

2. Let $\{X_n\}$ be IID uniform on $[0,1]$.
(a) Use large deviations to estimate $\mathbb{P}(\sum_{k=1}^n X_k \leq 1)$.
(b) Compute this probability exactly using combinatorial methods and compare to the result in part(a).

<u>*solution*</u> (a) Take $Y_k = 1 - X_k$, then $\sum_{k=1}^n Y_k = n - \sum_{k=1}^n X_k \Rightarrow \sum_{k=1}^n X_k = n - \sum_{k=1}^n Y_k$

$$\mathbb{P}\left(\sum_{k=1}^n X_k \leq 1\right) = \mathbb{P}\left(n - \sum_{k=1}^n Y_k \leq 1\right) = \mathbb{P}\left(\sum_{k=1}^n Y_k \geqslant n-1\right)$$

Markov's inequality gives, $\forall \theta > 0$

$$\begin{aligned}\mathbb{P}\left(S_n \geqslant n-1\right) = \mathbb{P}\left(e^{\theta S_n} \geqslant e^{\theta(n-1)}\right) &\leq \mathbb{E}\left[\frac{e^{\theta S_n}}{e^{\theta(n-1)}} 1\left(e^{\theta S_n} \geqslant e^{\theta(n-1)}\right)\right] \\ &\leq \mathbb{E} e^{\theta S_n} e^{-\theta(n-1)} = e^{-\theta(n-1)}\left(\mathbb{E} e^{\theta X_n}\right)^n = e^{-\theta(n-1)}[M(\theta)]^n\end{aligned}$$

where the mgf is $M(\theta) = \mathbb{E} e^{\theta x} = \int_0^1 e^{\theta x} dx = \frac{1}{\theta}\left(e^\theta - 1\right)$. Taking logs and multipling by $\frac{1}{n}$, we have

$$\frac{1}{n} \log \mathbb{P}\left(S_n \geqslant n-1\right) \leqslant \frac{-\theta(n-1)}{n} + \kappa(\theta) = -\theta\left(1 - \frac{1}{n}\right) + \kappa(\theta)$$

where $\kappa(\theta) = \log M(\theta) = \log \frac{e^\theta - 1}{\theta} = -\log \theta + \log\left(e^\theta - 1\right)$. Notice that when $n \to \infty$, the upper bound is decreasing and $\log\left(e^\theta - 1\right) \approx \theta$ when θ getting large. We can optimize the upper bound by taking $\kappa'(\theta) = 1 - \frac{1}{n}$, then

$$-\frac{1}{\theta} + \frac{e^\theta}{e^\theta - 1} = 1 - \frac{1}{n}, \text{ i.e.,} \frac{1}{\theta} + \frac{1}{e^\theta - 1} = \frac{1}{n}$$

Since n is large, θ need to be large as well. However, the term $\frac{1}{e^\theta - 1}$ is negligible compared to $\frac{1}{\theta}$ (as $e^\theta - 1 \gg \theta$), then the optimal $\theta = n$ approximately. Plugging it in gives the least upper bound:

$$-n + 1 - \log n + \log\left(e^n - 1\right) \approx 1 - \log n$$

since n is large. It gives the optimal upper bound

$$\frac{1}{n} \log \mathbb{P}\left(\sum_{k=1}^n Y_k \geqslant n-1\right) \leqslant 1 - \log n$$

Finally, there is

$$\mathbb{P}\left(\sum_{k=1}^n X_k \leqslant 1\right) = \mathbb{P}\left(\sum_{k=1}^n Y_k \geqslant n-1\right) \leqslant \frac{e^n}{n^n}$$

(b) Here are some observations: When $n = 1$, $\mathbb{P}\left(X_1 \leq 1\right) = 1$.
When $n = 2$, $\mathbb{P}\left(X_1 + X_2 \leq 1\right) = \frac{1}{2!} = \frac{1}{2}$, it's the area of half square

When $n = 3$, $\mathbb{P}(X_1 + X_2 + X_3 \le 1) = \frac{1}{3!} = \frac{1}{6}$, it's the volume of the solid bounded by $x = 0, y = 0, z = 0$ $x + y + z = 1$.
When $n = 4$, $\mathbb{P}(X_1 + X_2 + X_3 + X_4 \le 1) = \frac{1}{24} = \frac{1}{4!}$, it's the hypervolume of the comer of a unit 4-dimensional cube counts for $\frac{1}{24}$ of the hypervolume of $x^2 + y^2 + z^2 + w^2 = 1$.
Continuing this process to $\mathbb{P}(X_1 + \cdots + X_n \le 1)$, it takes $\frac{1}{n!}$ of 1 (total mass / hypervolume of a unit n-dimensional cube).
By induction, the base case is when $n = 1$. Assume $\mathbb{P}(S_n \le 1) = \frac{1^n}{n!}$, then

$$\begin{aligned}\mathbb{P}(S_{n+1} \le 1) &= \int_0^1 \mathbb{P}(S_n + X_{n+1} \le 1) \cdot 1 dx \\ &= \int_0^1 \mathbb{P}(S_n \leqslant 1 - x)\, dx \\ &= \int_0^1 \frac{(1-x)^n}{n!} dx = \frac{1}{(n+1)!}\end{aligned}$$

The exact probability is

$$\mathbb{P}(S_n \leqslant 1) = \frac{1}{n!}$$

Comparing to the result from (a), these two quantities asymptotically connected by the Stirling's formula $n! \sim \sqrt{2\pi n}(\frac{n}{e})^n$. More specific, $(\frac{e}{n})^n > \frac{1}{n!}\sqrt{2\pi n} e^{\frac{1}{12n+1}}$. The optimal upper bound obtained using the large deviation is still not precise enough.

3. Let $\{X_n\}$ be IID random variables with mean 0, $\mathbb{E}[X_n] = 0$ and $\mathbb{E}e^{\theta X_n} = \infty$ for all $\theta > 0$. Take any $a > 0$, show

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\sum_{k=1}^{n} X_k \geq an\right) = 0.$$

<u>*solution*</u>

The basic idea is similar to transformed distribution we used in class. However, instead of fully bounded interval $(na, n\nu]$, we only consider the lower bound here. .
Let $S_n = \sum_{k=1}^{n} X_k$. For any $a > 0$ and $\varepsilon > 0$, there is

$$\{S_{n-1} \geqslant -n\varepsilon\} \cap \{x_n \geqslant n(a+\varepsilon)\} \subseteq \{S_n \geqslant an\}$$
$$\Rightarrow \mathbb{P}\left(S_{n-1} \geqslant -n\varepsilon \text{ and } X_n \geqslant n(a+\varepsilon)\right) \leqslant \mathbb{P}\left(S_n \geqslant an\right)$$
$$\Rightarrow \text{(by indep.)}\mathbb{P}\left(S_{n-1} \geqslant -n\varepsilon\right)\mathbb{P}\left(X_n \geqslant n(a+\varepsilon)\right) \leq \mathbb{P}\left(S_n \geqslant an\right)$$
$$\Rightarrow \text{(log both sides)} \log \mathbb{P}\left(S_{n-1} \geqslant -n\varepsilon\right) + \log \mathbb{P}\left(X_n \geqslant n(a+\varepsilon)\right) \leq \log \mathbb{P}\left(S_n \geqslant an\right)$$
$$\Rightarrow \text{(multiplying by } \frac{1}{n}) \quad \frac{1}{n}\log \mathbb{P}\left(S_{n-1} \geqslant -n\varepsilon\right) + \frac{1}{n}\log \mathbb{P}\left(X_n \geqslant n(a+\varepsilon)\right) \leqslant \frac{1}{n}\log \mathbb{P}\left(S_n \geqslant an\right)$$

Claim 1: $\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(S_{n-1} \geqslant -n\varepsilon\right) = 0$
The WLLN gives

$$\lim_{n\to\infty} \mathbb{P}(|\frac{S_{n-1}}{n}| \leqslant \varepsilon) = 1 = \lim_{n\to\infty} \mathbb{P}(-n\varepsilon \leqslant S_{n-1} \leqslant n\varepsilon)$$

this implies

$$\lim_{n\to\infty} \mathbb{P}\left(S_{n-1} \geqslant -n\varepsilon\right) = 1$$

then there is

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(S_{n-1} \geqslant -n\varepsilon\right) = 0$$

Claim 2: $\lim = \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(X_n > n(a+k)\right) = 0.$
Lemma 2.7.1 implies that $\lim = \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(X_n > n(a+\varepsilon)\right)$ exists ≤ 0 $n \to \infty$. However, if the limit was < 0, there would be $\mathbb{E}e^{\theta x} < \infty$ for some $\theta > 0$, then the only possible condition is

$$\lim = \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(X_n > n(a+\varepsilon)\right) = 0$$

Claim 3: $\frac{1}{n} \log \mathbb{P}\left(S_n \geqslant a_n\right) \leqslant 0.$
Since $0 \leqslant \mathbb{P}\left(S_n \geqslant a_n\right) \leqslant 1$, the log probability is restricted to nonpositive values, i.e., $\log \mathbb{P}\left(S_n \geqslant a_n\right) \leqslant 0$, and of course $\frac{1}{n} \log \mathbb{P}\left(S_n \geqslant a_n\right) \leqslant 0$
The lower and upper bound of $\frac{1}{n} \log \mathbb{P}\left(S_n \geqslant an\right)$ are $\frac{1}{n} \log \mathbb{P}\left(S_{n-1} \geqslant -n\varepsilon\right) + \frac{1}{n} \log \mathbb{P}\left(X_n \geqslant n(a+\varepsilon)\right)$ and 0 respectively, and both of them converge to 0, indicating that

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(S_n \geq an\right) = 0$$