Midterm Exam - March 27, 2024, Limit thms. of probab., SOLUTION

1. Let $X_1, X_2, \dots$ denote i.i.d. random variables with distribution $\mathbb{P}(X_i = k) = \frac{2}{3^k}$, $k = 1, 2, 3, \dots$
   Let us define $S_n = X_1 + \cdots + X_n$.
   (a) Show that $\mathbb{P}(S_n = k) = \binom{k-1}{n-1}\frac{2^n}{3^k}$, $k = n, n+1, n+2, \dots$
   (b) Calculate $\lim_{n\to\infty} \frac{1}{n} \ln\left(\mathbb{P}(S_n = \lfloor nx \rfloor)\right)$, $x \in \mathbb{R}$.
   (c) Briefly explain how this relates to Cramér's theorem and one of the formulas from the *Formula sheet: large deviation rate functions, exponential tilting.*

   Solution:

   (a) X_i has optimistic geometric distribution GEO($\frac{2}{3}$). Thus if we consider a sequence of independent trials with success probability $\frac{2}{3}$ then X_1 is the number of trials until (and including) the first success, while S_n is the number of trials until (and including) the n'th success. The event $\{S_n = k\}$ occurs if and only if there were exactly $n-1$ successes among the first $k-1$ trials and the the k'th trial is successful. Thus $\mathbb{P}(S_n = k) = \binom{k-1}{n-1}\left(\frac{2}{3}\right)^{n-1} \cdot \left(\frac{1}{3}\right)^{(k-1)-(n-1)} \cdot \frac{2}{3} = \binom{k-1}{n-1} \cdot \frac{2^n}{3^k}$
   (b) If $x < 1$ then $\mathbb{P}(S_n = \lfloor nx \rfloor) = 0$. If $x \geq 1$ then we use the crude Stirling formula:

$$\mathbb{P}(S_n = \lfloor nx \rfloor) = \binom{\lfloor nx \rfloor - 1}{n-1}\frac{2^n}{3^{\lfloor nx \rfloor}} \approx \frac{\lfloor nx \rfloor!}{n!(\lfloor nx \rfloor - n)!} 2^n 3^{-nx} \approx$$
$$\frac{(nx)^{nx} e^{-nx}}{n^n e^{-n} (n(x-1))^{n(x-1)} e^{-n(x-1)}} 2^n 3^{-nx} = \frac{x^{nx}}{(x-1)^{n(x-1)}} 2^n 3^{-nx} = \left(\frac{x^x}{(x-1)^{x-1}} \frac{2}{3^x}\right)^n, \quad (1)$$

   thus $\lim_{n\to\infty} \frac{1}{n} \ln\left(\mathbb{P}(S_n = \lfloor nx \rfloor)\right) = x\ln(x) - (x-1)\ln(x-1) + \ln(2) - x\ln(3)$.
   (c) Recalling how we proved Cramér's theorem for binomial distribution (see page 5 of the scanned lecture notes) and recalling how we related the large deviation rate functions of BER(p) and GEO(p) distributions (see page 29-30 of scanned), and also by the heuristic meaning of Cramér's theorem (see page 24 of scanned), we expect $\lim_{n\to\infty} \frac{1}{n} \ln\left(\mathbb{P}(S_n = \lfloor nx \rfloor)\right) = -I(x)$, where $I(x)$ is the large deviation rate function of the GEO($\frac{2}{3}$) distribution, and this is indeed the case, since

$$I(x) = (x-1)\ln\left(\frac{x-1}{1/3}\right) - x\ln(x) - \ln(2/3).$$

2. Let $Z_1, Z_2, \dots$ denote i.i.d. random variables with p.d.f. $f(x) = xe^{-x}\mathbb{1}[x \geq 0]$. Let $M_n := \max\{Z_1, \dots, Z_n\}$. Let us define $c_n := \ln(n) + \ln(\ln(n))$. Let $Y_n := M_n - c_n$. Show that Y_n weakly converges as $n \to \infty$ and identify the limiting distribution.

   Solution: Using the setup of HW4.2(b), $f(x)$ is the p.d.f. of the time of the second earthquake, thus the corresponding c.d.f. is $F(x) = 1 - e^{-x}(1+x)$ if $x \geq 0$ (or one can also calculate $\int_0^x f(x)\,\mathrm{d}x = F(x)$ using integration by parts). For any $x \in \mathbb{R}$, we have $c_n + x \geq 0$ if n is large enough, and then we have

$$\mathbb{P}(Y_n \leq x) = \mathbb{P}(M_n \leq c_n + x) = \mathbb{P}(Z_i \leq c_n + x,\, i = 1, \dots, n) = F(c_n + x)^n =$$
$$\left(1 - e^{-\ln(n) - \ln(\ln(n)) - x}(1 + \ln(n) + \ln(\ln(n)) + x)\right)^n = \left(1 - \frac{e^{-x}}{n} \frac{1 + \ln(n) + \ln(\ln(n)) + x}{\ln(n)}\right)^n. \quad (2)$$

   Note that

$$\lim_{n\to\infty} \frac{1 + \ln(n) + \ln(\ln(n)) + x}{\ln(n)} = 1,$$

   thus

$$\lim_{n\to\infty} \mathbb{P}(Y_n \leq x) = \lim_{n\to\infty} \left(1 - \frac{e^{-x}}{n}\right)^n = \exp\left(-e^{-x}\right), \qquad x \in \mathbb{R}.$$

   Thus $Y_n \Rightarrow Y$, where $\mathbb{P}(Y \leq x) = \exp\left(-e^{-x}\right)$, i.e., Y has standard Gumbel distribution.