Limit/large dev. thms. HW assignment 4.

1. *The Fréchet distribution.*

   (a) Let $U_1, U_2, \dots$ denote i.i.d. random variables with UNI$[0,1]$ distribution. Let $\beta > 0$. Let

   $$M_n = \max\{U_1^{-\beta}, \dots, U_n^{-\beta}\}.$$

   Show that M_n/n^β converges in distribution as $n \to \infty$ by determining the cumulative distribution function (c.d.f.) $F(x)$ of the limiting distribution.

   (b) Show that if Y_1 and Y_2 are i.i.d. with the above c.d.f. $F(x)$ then $(Y_1 \vee Y_2)/2^\beta$ also has c.d.f. $F(x)$. *Instruction:* Use the explicit formula for F that you have obtained in sub-exercise (a), similarly to the top of page 44 of the scanned lecture notes.

   (c) Show that if Y_1 and Y_2 are i.i.d. with the above c.d.f. $F(x)$ then $(Y_1 \vee Y_2)/2^\beta$ also has c.d.f. $F(x)$. *Instruction:* Do not use the explicit form of F, but use the limit theorem (i.e., $M_n/n^\beta \implies Y_1$) that you have obtained in sub-exercise (a), similarly to the middle of page 44 of scanned.

Solution:

(a) First note that for any $x \geq 1$ we have $\mathbb{P}(U_i^{-\beta} \leq x) = \mathbb{P}(U_i \geq x^{-1/\beta}) = 1 - x^{-1/\beta}$, thus for any $x > 0$ and for any n big enough so that $xn^\beta \geq 1$ we have

$$\mathbb{P}(M_n/n^\beta \leq x) = \mathbb{P}(M_n \leq xn^\beta) = \mathbb{P}(U_1^{-\beta} \leq xn^\beta, \dots, U_n^{-\beta} \leq xn^\beta) =$$
$$\mathbb{P}(U_1^{-\beta} \leq xn^\beta) \dots \mathbb{P}(U_n^{-\beta} \leq xn^\beta) = (1 - (xn^\beta)^{-1/\beta})^n = \left(1 - \frac{x^{-1/\beta}}{n}\right)^n$$

Therefore $\lim_{n\to\infty} \mathbb{P}(M_n/n^\beta \leq x) = e^{-x^{-1/\beta}}$ for any $x \geq 0$.
If we define $F_n(x) = \mathbb{P}(M_n/n^\beta \leq x)$ then $F_n \Rightarrow F$, where

$$F(x) = \begin{cases} e^{-x^{-1/\beta}} & \text{if} \quad x > 0, \\ 0 & \text{if} \quad x \leq 0. \end{cases}$$

Remark: The probability distribution corresponding to the c.d.f. F is known as the *Fréchet distribution* in extreme value theory (c.f. page 43 of the scanned lecture notes).

(b)

$$\mathbb{P}\left((Y_1 \vee Y_2)/2^\beta \leq x\right) = \mathbb{P}\left(Y_1 \vee Y_2 \leq 2^\beta x\right) = \mathbb{P}(Y_1 \leq 2^\beta x)\mathbb{P}(Y_2 \leq 2^\beta x) = F^2(2^\beta x)$$

Thus we want to show that $F^2(2^\beta x) = F(x)$. This is clear if $x \leq 0$, since both sides are zero. On the other hand, if $x > 0$ then

$$F^2(2^\beta x) = (e^{-(2^\beta x)^{-1/\beta}})^2 = (e^{-\frac{1}{2}x^{-1/\beta}})^2 = e^{-x^{-1/\beta}} = F(x)$$

(c) Let $M_n^* = \max\{U_{n+1}^{-\beta}, \dots, U_{2n}^{-\beta}\}$. Then M_n and M_n^* are i.i.d., moreover $M_n/n^\beta \implies Y_1$ and $M_n^*/n^\beta \implies Y_2$, where Y_1 and Y_2 are i.i.d. with c.d.f. F by sub-exercise (a). Also, we have $M_{2n}/(2n)^\beta \implies Y_3$, where Y_3 has c.d.f. F, again by (a). But on the other hand hand

$$\frac{M_{2n}}{(2n)^\beta} = \frac{M_n \vee M_n^*}{(2n)^\beta} = \frac{1}{2^\beta}\left(\frac{M_n}{n^\beta} \vee \frac{M_n^*}{n^\beta}\right) \implies \frac{Y_1 \vee Y_2}{2^\beta}$$

Thus $Y_3 \sim (Y_1 \vee Y_2)/2^\beta$, thus $(Y_1 \vee Y_2)/2^\beta$ also has c.d.f. $F(x)$.
Remark: Essentially what we have shown here was that the Fréchet distribution is *max-stable.*
The *Gumbel distribution* is also max-stable, see page 44 of the scanned lecture notes.
For the precise definition of max-stability, see the lecture notes of the *Extreme value theory* course BMETE95MM16 or wikipedia.
More about stable distributions later.

2. The goal of this exercise is to deduce the central limit theorem (CLT) for Poisson distribution using the CLT for the sum of i.i.d. $EXP(1)$ random variables (proved in class on March 6).

   (a) Let $F_n : \mathbb{R} \to [0,1]$ and $F : \mathbb{R} \to [0,1]$ denote c.d.f.'s. Assume that F is continuous and $F_n \Rightarrow F$. Prove that for any convergent sequence $x_n \to x$ of real numbers we have $F_n(x_n) \to F(x)$.
   Hint: Use Slutsky.

   (b) Let $X_1, X_2, \dots$ denote i.i.d. random variables with EXP(1) distribution. We can think of X_i as the waiting time between the arrivals of consecutive earthquakes. Denote by $T_n = X_1 + \cdots + X_n$ the time of the n'th earthquake. We have already determined the p.d.f. of T_n in HW3.3(a). Deduce from this that the c.d.f. of T_n is

$$F_n(t) = \mathbb{P}(T_n \le t) = 1 - \sum_{k=0}^{n-1} e^{-t}\frac{t^k}{k!} \tag{1}$$

   (c) Denote by N_t the number of earthquakes during the time interval $[0,t]$. Show that the identity $\{T_n \le t\} = \{N_t \ge n\}$ holds and deduce from sub-exercise (b) that N_t has POI(t) distribution.

   (d) Use the fact that $\frac{T_n - n}{\sqrt{n}} \Rightarrow \mathcal{N}(0,1)$ as $n \to \infty$ to deduce that $\frac{N_t - t}{\sqrt{t}} \Rightarrow \mathcal{N}(0,1)$ as $t \to \infty$.
   Hint: You will have to use $\{T_n \le t\} = \{N_t \ge n\}$ as well as the result of sub-exercise (a).

Solution:

(a) Let X_n denote a random variable with c.d.f. F_n. Let X denote a random variable with c.d.f. F. Then $X_n \implies X$ as $n \to \infty$. Let $a_n := x_n - x$, thus $x + a_n = x$. Thus $a_n \to 0$ as $n \to \infty$. Let $Y_n := X_n - a_n$. We have $Y_n \implies X$ by Slutsky. Let G_n denote the c.d.f. of Y_n. Thus we have $G_n \implies F$. Since F is continuous, we have $G_n(x) \to F(x)$. But

$$G_n(x) = \mathbb{P}(Y_n \le x) = \mathbb{P}(X_n - a_n \le x) = \mathbb{P}(X_n \le x + a_n) = \mathbb{P}(X_n \le x_n) = F_n(x_n),$$

thus $F_n(x_n) \to F(x)$.

(b) The p.d.f. of T_n is $f_n(t) = e^{-t}\frac{t^{n-1}}{(n-1)!}$ if $t \ge 0$. In order to prove that F_n defined in (1) is indeed satisfies $F_n(t) = \int_0^t f_n(s)\,\mathrm{d}s$, we only need to check that $F_n(0) = 0$ and $F_n'(t) = f_n(t)$ if $t \ge 0, n \ge 1$. Indeed, we have $F_n(0) = 0$ and

$$F_n'(t) = -\sum_{k=0}^{n-1} \frac{\mathrm{d}}{\mathrm{d}t}\left(e^{-t}\frac{t^k}{k!}\right) = \sum_{k=0}^{n-1} e^{-t}\frac{t^k}{k!} - \sum_{k=0}^{n-2} e^{-t}\frac{t^k}{k!} = f_n(t)$$

(c) The events $\{T_n \le t\}$ and $\{N_t \ge n\}$ both mean that the n'th earthquake happened by time t.

$$\mathbb{P}(N_t = n) = \mathbb{P}(N_t \ge n) - \mathbb{P}(N_t \ge n+1) = \mathbb{P}(T_n \le t) - \mathbb{P}(T_{n+1} \le t) \overset{(1)}{=} e^{-t}\frac{t^n}{n!}, \quad \text{hence } N_t \sim \text{POI}(t).$$

(d) Let us fix some $x \in \mathbb{R}$. We want to show that $\lim_{t\to\infty} \mathbb{P}\left(\frac{N_t - t}{\sqrt{t}} \le x\right) = \Phi(x)$.

This is equivalent to showing that $\lim_{t\to\infty} \mathbb{P}\left(\frac{N_t - t}{\sqrt{t}} \ge x\right) = 1 - \Phi(x)$.

$$\left\{\frac{N_t - t}{\sqrt{t}} \ge x\right\} = \left\{N_t \ge t + \sqrt{t}x\right\} = \left\{N_t \ge \lceil t + \sqrt{t}x \rceil\right\} = \left\{T_{\lceil t+\sqrt{t}x\rceil} \le t\right\} =$$

$$\left\{\frac{T_{\lceil t+\sqrt{t}x\rceil} - \lceil t + \sqrt{t}x\rceil}{\sqrt{\lceil t + \sqrt{t}x\rceil}} \le \frac{t - \lceil t + \sqrt{t}x\rceil}{\sqrt{\lceil t + \sqrt{t}x\rceil}}\right\} \tag{2}$$

Now observe that $\frac{T_n - n}{\sqrt{n}} \Rightarrow \mathcal{N}(0,1)$ as $n \to \infty$ implies that $\frac{T_{\lceil t+\sqrt{t}x\rceil} - \lceil t+\sqrt{t}x\rceil}{\sqrt{\lceil t+\sqrt{t}x\rceil}} \Rightarrow \mathcal{N}(0,1)$ as $t \to \infty$.

Also note that $\lim_{t\to\infty} \frac{t - \lceil t+\sqrt{t}x\rceil}{\sqrt{\lceil t+\sqrt{t}x\rceil}} = -x$, therefore

$$\lim_{t\to\infty} \mathbb{P}\left(\frac{N_t - t}{\sqrt{t}} \ge x\right) \overset{(2)}{=} \lim_{t\to\infty} \mathbb{P}\left(\frac{T_{\lceil t+\sqrt{t}x\rceil} - \lceil t + \sqrt{t}x\rceil}{\sqrt{\lceil t + \sqrt{t}x\rceil}} \le \frac{t - \lceil t + \sqrt{t}x\rceil}{\sqrt{\lceil t + \sqrt{t}x\rceil}}\right) \overset{(a)}{=} \Phi(-x) = 1 - \Phi(x).$$

3. *Local central limit theorem for* BIN$(n, \frac{1}{2})$

Let $X_1, X_2, \dots$ denote i.i.d. random variables, where $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = 0) = \frac{1}{2}$. Let $S_n = X_1 + \dots + X_n$, thus $S_n \sim \mathrm{BIN}(n, \frac{1}{2})$. In this exercise we write $a_n \approx b_n$ to denote that $\lim_{n\to\infty} a_n/b_n = 1$.

(a) Use Stirling's formula to show that if $(k(n))$ is an integer-valued sequence satisfying $k(n) \to \infty$ and $n - k(n) \to \infty$ then

$$\frac{\sqrt{n}}{2}\mathbb{P}\left(S_n = k(n)\right) \approx \frac{1}{\sqrt{2\pi}} \frac{1}{(2k(n)/n)^{k(n)+\frac{1}{2}} \cdot (2 - 2k(n)/n)^{(n-k(n))+\frac{1}{2}}}. \tag{3}$$

(b) Show that if $k(n) = \frac{n}{2} + \frac{\sqrt{n}}{2}z(n)$, where $(z(n))$ is a bounded real-valued sequence, then

$$(2k(n)/n)^{k(n)+\frac{1}{2}} \cdot (2 - 2k(n)/n)^{(n-k(n))+\frac{1}{2}} \approx e^{z(n)^2/2}. \tag{4}$$

(c) Prove the *local CLT* for S_n, i.e., show that for any $x \in \mathbb{R}$ we have

$$\lim_{n\to\infty} \frac{\sqrt{n}}{2}\mathbb{P}\left(S_n = \left\lfloor \frac{n}{2} + \frac{\sqrt{n}}{2}x \right\rfloor\right) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}. \tag{5}$$

Hint: There is a sequence $z(n)$ such that $k(n) = \left\lfloor \frac{n}{2} + \frac{\sqrt{n}}{2}x \right\rfloor = \frac{n}{2} + \frac{\sqrt{n}}{2}z(n)$ for all n.

Solution: The theorem (5) was first proved by Abraham de Moivre in 1718.

(a) Note that $k(n) \to \infty$ and $(n - k(n)) \to \infty$ as $n \to \infty$, so we can apply Stirling's formula to $k(n)!$ as well as $(n - k(n))!$ in the calculation below:

$$\frac{\sqrt{n}}{2}\mathbb{P}\left(S_n = k(n)\right) = \frac{\sqrt{n}}{2}\frac{n!}{k(n)!(n-k(n))!}2^{-n} \approx$$
$$\frac{\sqrt{n}}{2}\frac{\sqrt{2\pi}n^{n+\frac{1}{2}}e^{-n}}{\sqrt{2\pi}k(n)^{k(n)+\frac{1}{2}}e^{-k(n)} \cdot \sqrt{2\pi}(n-k(n))^{(n-k(n))+\frac{1}{2}}e^{-(n-k(n))}}2^{-n} =$$
$$\frac{1}{\sqrt{2\pi}}\frac{\sqrt{n}}{2}\frac{n^{n+\frac{1}{2}}}{k^{k(n)+\frac{1}{2}} \cdot (n-k(n))^{(n-k(n))+\frac{1}{2}}}2^{-n} = \frac{1}{\sqrt{2\pi}}\frac{(n/2)^{n+1}}{k(n)^{k(n)+\frac{1}{2}} \cdot (n-k(n))^{(n-k(n))+\frac{1}{2}}} =$$
$$\frac{1}{\sqrt{2\pi}}\frac{1}{(2k(n)/n)^{k(n)+\frac{1}{2}} \cdot (2-2k(n)/n)^{(n-k(n))+\frac{1}{2}}}.$$

(b) We will use that if $a_n \to 0$ and $b_n \to \infty$, moreover $(a_n b_n)$ is bounded, then $(1 + a_n)^{b_n} \approx e^{a_n b_n}$.

$$(2k(n)/n)^{k(n)+\frac{1}{2}} \cdot (2-2k(n)/n)^{(n-k(n))+\frac{1}{2}} = \left(1 + \frac{z(n)}{\sqrt{n}}\right)^{k(n)+\frac{1}{2}} \cdot \left(1 - \frac{z(n)}{\sqrt{n}}\right)^{(n-k(n))+\frac{1}{2}} \approx$$
$$\left(1 + \frac{z(n)}{\sqrt{n}}\right)^{k(n)} \cdot \left(1 - \frac{z(n)}{\sqrt{n}}\right)^{(n-k(n))} = \left(1 + \frac{z(n)}{\sqrt{n}}\right)^{\frac{n}{2}+\frac{\sqrt{n}}{2}z} \cdot \left(1 - \frac{z(n)}{\sqrt{n}}\right)^{\frac{n}{2}-\frac{\sqrt{n}}{2}z(n)} =$$
$$\left(1 - \frac{z(n)^2}{n}\right)^{\frac{n}{2}} \cdot \left(1 + \frac{z(n)}{\sqrt{n}}\right)^{\frac{\sqrt{n}}{2}z(n)} \cdot \left(1 - \frac{z(n)}{\sqrt{n}}\right)^{-\frac{\sqrt{n}}{2}z(n)} \approx e^{-z(n)^2/2} \cdot e^{z(n)^2/2} \cdot e^{z(n)^2/2} = e^{z(n)^2/2}.$$

(c) There is a sequence $z(n)$ such that $k(n) = \left\lfloor \frac{n}{2} + \frac{\sqrt{n}}{2}x \right\rfloor = \frac{n}{2} + \frac{\sqrt{n}}{2}z(n)$ for all n. We have $k(n) \to \infty$ and $(n - k(n)) \to \infty$, so (a) can be applied. Also, clearly, we have $|x - z(n)| \le \frac{2}{\sqrt{n}}$, thus $e^{-z(n)^2/2} \approx e^{-x^2/2}$ as $n \to \infty$. The sequence $z(n)$ converges, so it is bounded, therefore (b) can be applied. Thus we have

$$\frac{\sqrt{n}}{2}\mathbb{P}\left(S_n = \left\lfloor \frac{n}{2} + \frac{\sqrt{n}}{2}x \right\rfloor\right) \overset{(a),(b)}{\approx} \frac{1}{\sqrt{2\pi}}e^{-z(n)^2/2} \approx \frac{1}{\sqrt{2\pi}}e^{-x^2/2}, \qquad n \to \infty. \tag{6}$$

This completes the proof of (5).