Limit/large dev. thms. HW assignment 5.

1. Let T_1 denote the first time when the one dimensional simple symmetric random walk (X_n) reaches level 1 (see page 59 of the scanned lecture notes).

   (a) Show that $\mathbb{P}(T_1 > n) = \mathbb{P}(X_n = 0) + \mathbb{P}(X_n = 1)$.
   Hint: Use the reflection principle (see page 58).

   (b) Use the result of an earlier homework to show that $\lim_{n\to\infty} \frac{\mathbb{P}(T_1>n)}{\sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{n}}} = 1$.

   (c) Show that $\mathbb{E}(T_1) = +\infty$.

   Solution:

   (a) First note that if n is an even number then $\mathbb{P}(X_n = 1) = 0$ and $\mathbb{P}(X_n = 0) > 0$, but if n is an odd number then $\mathbb{P}(X_n = 0) = 0$ and $\mathbb{P}(X_n = 1) > 0$. However, we will consider the even and odd cases together in our proof.
   Recall that $\{T_1 \le n\} = \{M_n \ge 1\}$, where $M_n = \max\{X_0, X_1 \dots X_n\}$, thus

   $$\mathbb{P}(T_1 > n) = 1 - \mathbb{P}(T_1 \le n) = 1 - \mathbb{P}(M_n \ge 1) \overset{(*)}{=} 1 - (2\mathbb{P}(X_n > 1) + \mathbb{P}(X_n = 1)) \overset{(**)}{=}$$
   $$1 - \mathbb{P}(X_n \notin \{0,1\}) = \mathbb{P}(X_n \in \{0,1\}) = \mathbb{P}(X_n = 0) + \mathbb{P}(X_n = 1),$$

   where $(*)$ holds by the reflection principle, i.e., the lemma on the top of page 58 of lecture notes, and $(**)$ holds because

   $$2\mathbb{P}(X_n > 1) + \mathbb{P}(X_n = 1) \overset{(\Diamond)}{=} \mathbb{P}(X_n > 1) + \mathbb{P}(X_n < -1) + \mathbb{P}(X_n = 1) =$$
   $$\mathbb{P}(X_n \notin \{-1,0\}) \overset{(\Diamond)}{=} \mathbb{P}(X_n \notin \{0,1\}),$$

   where both equalities marked by $(\Diamond)$ hold because the random walk trajectory has the same law as its reflection with respect to the horizontal axis (since the increments of the walk have a symmetric distribution).

   (b) First note that if n is an even number then $\mathbb{P}(X_n = 1) = 0$, but if n is an odd number then $\mathbb{P}(X_n = 0) = 0$. Thus, if n is an even number then by the result of HW4.3, we have

   $$\mathbb{P}(X_n = 0) + \mathbb{P}(X_n = 1) = \mathbb{P}(X_n = 0) \approx \frac{2}{\sqrt{n}}\frac{1}{\sqrt{2\pi}}e^{-0^2/2} = \sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{n}}.$$

   If n is an odd number, then similarly we obtain

   $$\mathbb{P}(X_n = 0) + \mathbb{P}(X_n = 1) = \mathbb{P}(X_n = 1) \approx \frac{2}{\sqrt{n}}\frac{1}{\sqrt{2\pi}}e^{-(1/\sqrt{n})^2/2} \approx \frac{2}{\sqrt{n}}\frac{1}{\sqrt{2\pi}}e^{-0^2/2} = \sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{n}}.$$

   For further details of a similar calculation, see page 71 of the scanned lecture notes.

   (c) Here is a well-known formula for the expectation of a non-negative integer-valued random variable: $\mathbb{E}(T_1) = \sum_{n=0}^{\infty} \mathbb{P}(T_1 > n)$. Let's prove it. First note that $T_1 = \sum_{n=0}^{\infty} \mathbb{1}[T_1 > n]$, then we calculate

   $$\mathbb{E}(T_1) = \mathbb{E}\left(\sum_{n=0}^{\infty} \mathbb{1}[T_1 > n]\right) \overset{(*)}{=} \sum_{n=0}^{\infty} \mathbb{E}\left(\mathbb{1}[T_1 > n]\right) = \sum_{n=0}^{\infty} \mathbb{P}(T_1 > n),$$

   where in $(*)$ we used the linearity of expectation and the monotone convergence theorem. Now

   $$\mathbb{E}(T_1) = \sum_{n=0}^{\infty} \mathbb{P}(T_1 > n) = +\infty$$

   follows by the *limit comparison test* (see wikipedia): we have

   $$\lim_{n\to\infty} \frac{\mathbb{P}(T_1 > n)}{\sqrt{\frac{2}{\pi}\frac{1}{\sqrt{n}}}} = 1, \qquad \sum_{n=1}^{\infty} \sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{n}} = +\infty.$$

2. Let $y_{n,j} \in \mathbb{R}$, $j = 1, 2, \ldots, N_n$, $n = 1, 2, \ldots$ and assume

$$\lim_{n\to\infty} \max_{1\le j\le N_n} |y_{n,j}| = 0, \qquad \sup_n \sum_{j=1}^{N_n} |y_{n,j}| < \infty, \qquad \lim_{n\to\infty} \sum_{j=1}^{N_n} y_{n,j} = y.$$

Prove

$$\lim_{n\to\infty} \prod_{j=1}^{N_n} (1 + y_{n,j}) = e^y.$$

Hint: Use the first order Taylor expansion of the logarithm function: if $|y| < \frac{1}{2}$ then $|\ln(1+y) - y| \le Cy^2$.

Solution:

Denote

$$m_n := \max_{1\le j\le N_n} |y_{n,j}|, \qquad a_n := \sum_{j=1}^{N_n} |y_{n,j}|, \qquad s_n := \sum_{j=1}^{N_n} y_{n,j},$$

and use the Taylor expansion with error estimate: There exists $C < \infty$ such that for $|y| < 1/2$, $|\ln(1+y) - y| < Cy^2$. Then

$$\left| \ln \prod_{j=1}^{N_n} (1 + y_{n,j}) - s_n \right| \le \sum_{j=1}^{N_n} C y_{n,j}^2 \le C \sum_{j=1}^{N_n} m_n |y_{n,j}| = C m_n a_n.$$

Since $m_n \to 0$, $\sup_n a_n < +\infty$, we have $m_n a_n \to 0$, moreover $s_n \to y$, so we get

$$\lim_{n\to\infty} \prod_{j=1}^{N_n} (1 + y_{n,j}) = e^y.$$

3. The classical birthday paradox is the fact that if we choose 23 people randomly, then with probability at least 1/2 there will be at least two of them who celebrate their birthdays on the same day of the year. This fact can be viewed as the $n = 365$ case of the following limit theorem.

Let us fix $n \in \mathbb{N}$ and let $X_{n,j}$, $j = 1, 2, \dots$ be i.i.d. random variables uniformly distributed on $\{1, 2, \dots, n\}$. Define

$$T_n := \min\{\, k \,:\, \exists j < k,\ X_{n,j} = X_{n,k} \,\}.$$

In plain words: T_n is the index j when the first coincidence of the values is observed. Note that by the pigeonhole principle, we have $\mathbb{P}(T_n \le n + 1) = 1$.

Prove that $T_n/\sqrt{n}$ converges weakly as $n \to \infty$ and identify the limiting distribution. More specifically, prove that for any $x \ge 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{T_n}{\sqrt{n}} > x\right) = e^{-x^2/2}.$$

Hint: Use the result of exercise 2.

Solution: First note that

$$\mathbb{P}\left(T_n = j \,|\, T_n > j - 1\right) = \frac{j-1}{n}, \tag{1}$$

as we now explain. T_n is the time of the first collision, so $T_n > j - 1$ means that the values of $X_{n,1}, \dots, X_{n,j-1}$ are all different, so the probability that $X_{n,j}$ collides with one of $X_{n,1}, \dots, X_{n,j-1}$ is $\frac{j-1}{n}$, i.e., (1) holds. By iterative conditioning, from (1) we obtain

$$\mathbb{P}\left(T_n > k\right) = \prod_{j=1}^{k-1}\left(1 - \frac{j}{n}\right).$$

Then

$$\mathbb{P}\left(\frac{T_n}{\sqrt{n}} > x\right) = \prod_{j=1}^{\lfloor x\sqrt{n}\rfloor}\left(1 - \frac{j}{n}\right).$$

Having fixed $x \in \mathbb{R}_+$, we want to apply exercise 2 with $N_n = \lfloor x\sqrt{n}\rfloor$ and $y_{n,j} = -j/n$.

$$\begin{aligned}
\lim_{n\to\infty} \max_{1\le j\le N_n} |y_{n,j}| &= \lim_{n\to\infty} \max_{1\le j\le \lfloor x\sqrt{n}\rfloor} \frac{j}{n} = \lim_{n\to\infty} \frac{\lfloor x\sqrt{n}\rfloor}{n} = 0,\\
\sup_n \sum_{j=1}^{N_n} |y_{n,j}| &= \sup_n \sum_{j=1}^{\lfloor x\sqrt{n}\rfloor} \frac{j}{n} \overset{(*)}{<} +\infty,\\
\lim_{n\to\infty} \sum_{j=1}^{N_n} y_{n,j} = \lim_{n\to\infty} \sum_{j=1}^{\lfloor x\sqrt{n}\rfloor} \frac{-j}{n} &= \lim_{n\to\infty} -\frac{(\lfloor x\sqrt{n}\rfloor)(\lfloor x\sqrt{n}\rfloor + 1)}{2n} = -\frac{x^2}{2},
\end{aligned}$$

where in the equation marked by $(*)$ we used that a convergent sequence has a finite supremum. We obtain that for any $x \ge 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{T_n}{\sqrt{n}} > x\right) = e^{-x^2/2}.$$

This means that $T_n/\sqrt{n}$ converges in distribution as $n \to \infty$ to a r.v. with c.d.f.

$$F(x) = \begin{cases} 1 - e^{-x^2/2} & \text{if } x \ge 0,\\ 0 & \text{if } x \le 0.\end{cases}$$

Fun fact: this distribution is called the *Rayleigh distribution*.