Homework 7 (due Thursday, April 18)

Exercise 1. (birthday problem)
Let \((X_n, n \geq 1)\) be a sequence of i.i.d. random variables, each uniform on \(\{1, \ldots, N\}\). Let also
\[
T_N = \min\{n \geq 1 : X_n = X_m \text{ for some } m < n\}
\]
(notice that whatever happens, \(T_N \in \{2, \ldots, N + 1\}\)). Show directly that
\[
P\left(\frac{T_N}{\sqrt{N}} \leq t\right) \to_{N \to \infty} 1 - e^{-t^2/2}, \quad \forall t \geq 0
\]

NB: Approximations are allowed here!

Application: Use this to obtain a rough estimate of \(P(\{T_{365} \leq 22\})\) and \(P(\{T_{365} \leq 50\})\) (i.e. what is the probability that among 22 / 50 people, at least two have their birthday on the same day?)

Exercise 2. Let \((X_n, n \geq 1)\) be a sequence of i.i.d. \(\mathcal{E}(\lambda)\) random variables defined on a common probability space \((\Omega, \mathcal{F}, P)\), i.e., \(X_1\) admits the following pdf:
\[
p_{X_1}(x) = \begin{cases} \lambda \exp(-\lambda x), & x \geq 0 \\ 0, & x < 0 \end{cases}
\]
Let also \(S_n = X_1 + \ldots + X_n\). Using the large deviations principle, find a tight upper bound on
\[
P(\{S_n \geq nt\}), \quad \text{for } t > \mathbb{E}(X_1) = \frac{1}{\lambda}
\]

Exercise 3. Let \((X_n, n \geq 1)\) be a sequence of i.i.d. \(\mathcal{N}(0,1)\) random variables defined on a common probability space \((\Omega, \mathcal{F}, P)\). Let also \(S_n = X_1 + \ldots + X_n\). Find the exact value of
\[
P(\{S_n \geq nt\}), \quad \text{for } t > 0
\]

Exercise 4. For this exercise, you will need a generalization of the Cauchy-Schwartz inequality: Hölder’s inequality (written here in a slightly unusual form to help you with the exercise). This inequality says that if \(X, Y\) are two integrable random variables, then for every \(\alpha \in [0,1]\),
\[
\mathbb{E}\left(|X|^{\alpha} |Y|^{1-\alpha}\right) \leq \mathbb{E}(|X|^{\alpha}) \mathbb{E}(|Y|)^{1-\alpha}
\]

Preliminary: Show that for \(\alpha = 1/2\), this is nothing but Cauchy-Schwarz’ inequality.

Let now \(X\) be a random variable such that \(\mathbb{E}(\exp(sX)) < \infty\) for every \(s \in \mathbb{R}\).

a) Show that the function \(\Lambda(s) = \log(\mathbb{E}(\exp(sX)))\) is convex.

b) Show that the function \(\Lambda^*(t) = \sup_{s \in \mathbb{R}} (st - \Lambda(s))\) is also convex.
Coding Exercise 5. Let \((X_n, n \geq 1)\) be a sequence of i.i.d. random variables such that

\[ \mathbb{P}(\{X_1 = +1\}) = \mathbb{P}(\{X_1 = -1\}) = \frac{1}{2} \]

Let also \(S_n = X_1 + \ldots + X_n\) for \(n \geq 1\). For a fixed value of \(n\), draw on the same graph the following functions:

\[
f(t) = -\frac{1}{n} \log \mathbb{P}(\{S_n > nt\}) \\
g(t) = \Lambda^*(t) = \min_{s \in \mathbb{R}} (st - \Lambda(s)) \quad \text{where} \quad \Lambda(s) = \log \mathbb{E}(e^{sX_1}) \\
h(t) = t^2 / 2
\]

**NB:** On these plots, \(t \in [0, +1]\).

In order to draw the function \(f(t)\), you should use Monte-Carlo simulation, that is, draw i.i.d. samples \(X_1^{(m)}, \ldots, X_n^{(m)}\) for \(m = 1, \ldots, M\) (with \(M\) reasonably large) and approximate \(f(t)\) as

\[
f(t) \simeq -\frac{1}{n} \log \left( \frac{1}{M} \# \{1 \leq m \leq M : S_n^{(m)} > nt\} \right)
\]

As you will see, considering even moderate values of \(n\) requires considering quite large values of \(M\).