Problem Set 2

Submission:
Thursday, 02/17/2022, until 1 PM, to be uploaded on the NYU Brightspace course homepage.

---

1. **Convergence in probability and in distribution** [4 Points]

   (a) Consider a sequence $(X_n)_{n\geq 1}$ of random variables with $\mathbf{P}[X_n = n^\alpha] = \frac{1}{n}$ and $\mathbf{P}[X_n = 0] = 1 - \frac{1}{n}$ for every $n \in \mathbb{N}$. For which $\alpha \in \mathbb{R}$ does one have $X_n \xrightarrow[n\to\infty]{\mathbf{P}} 0$, and for which does one have $X_n \xrightarrow[n\to\infty]{d} 0$? For which ones does one have $\mathbf{E}[X_n] \xrightarrow[n\to\infty]{} 0$?

   (b) Suppose that $(X_n)_{n\geq 1}$ is an i.i.d. sequence of $Ber(p)$-distributed random variables. Determine the quantity a in
   $$\frac{1}{n}\sum_{i=1}^n \exp(X_i) \xrightarrow[n\to\infty]{\mathbf{P}} a.$$

   (c) Suppose that $(X_n)_{n\geq 1}$ is an i.i.d. sequence of random variables with $\mathbf{E}[X_1] = 0$ and $\mathrm{Var}[X_1] = \sigma^2 \in (0,\infty)$. Argue in detail, why
   $$\frac{1}{\log(n)\sqrt{n}}\sum_{i=1}^n X_i \xrightarrow[n\to\infty]{\mathbf{P}} 0.$$

   Hint: Use the properties of convergence in probability / in distribution from the notes.

2. **Some applications of the central limit theorem** [4 Points]

   (a) Let $(X_n)_{n\geq 1}$ be a sequence of i.i.d. $\mathcal{U}(\{1,2,3,4\})$ distributed random variables. Consider the expression
   $$Z_n = \log\left(\frac{1}{n}\sum_{i=1}^n X_i\right).$$
   Find the approximate distribution of Z_n for large n.
   Hint: Use the δ-method.

   (b) Suppose that the number of goals in a soccer match is Poisson-distributed with mean 3. Assume also that during a season, there are $n = 300$ matches. Use the central limit theorem to find the approximate probability for the event that during a season, there are at least 860 goals, but less than 930 goals in total. You may assume that the numbers of goals in different matches are independent.

3. **Asymptotic normality of the t-statistics** [4 Points]

   In this problem, we study the t-statistics, which is defined as follows: Let $X_1, \dots, X_n$ be i.i.d. real random variables with $\mathbf{E}[X_1] = \mu$, and $\mathrm{Var}[X_1] = \sigma^2 \in (0,\infty)$. The t-statistics is
   $$T_{n-1} = \frac{\overline{X}_n - \mu}{S_n/\sqrt{n}}, \qquad \text{where } \overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i,\ S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2.$$
   The goal of this problem is to show step-by-step that $T_{n-1} \xrightarrow[n\to\infty]{d} \mathcal{N}(0,1)$.

(a) First argue that one can write $\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$ as $\sum_{i=1}^{n} X_i^2 - (\overline{X}_n)^2$.

(b) Show that $\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$ converges in probability to σ^2.
Hint: Apply the weak law of large numbers to both $\frac{1}{n}\sum_{i=1}^{n} X_i^2$ and to $\overline{X}_n$. Then use the continuous mapping theorem and the fact that if $Y_n \xrightarrow[n\to\infty]{\mathbf{P}} Y$ and $Z_n \xrightarrow[n\to\infty]{\mathbf{P}} Z$, also $Y_n + Z_n \xrightarrow[n\to\infty]{\mathbf{P}} Y + Z$.

(c) Conclude from (b) that $S_n^2 \xrightarrow[n\to\infty]{\mathbf{P}} \sigma^2$. Use the continuous mapping theorem *again* to argue that $\frac{1}{S_n} \xrightarrow[n\to\infty]{\mathbf{P}} \frac{1}{\sigma}$.

(d) Finally use Slutsky's theorem and the central limit theorem to conclude that $T_{n-1} \xrightarrow[n\to\infty]{d} \mathcal{N}(0,1)$.
Hint: Note that $T_{n-1} = \sqrt{n}\frac{\overline{X}_n - \mu}{\sigma} \cdot \frac{\sigma}{S_n}$.

4. **(R exercise) Simulating the central limit theorem and law of large numbers** [4 Points]

This problem illustrates the central limit theorem and the law of large numbers using simulations of random variables. You can use the software R, which can be found via *http://cran.r-project.org/*.

(a) Consider first i.i.d. random variables $X_1, ..., X_n$ with $X_1 \sim Pois(\lambda)$ with $\lambda > 0$.

(i) Formulate the law of large numbers and the central limit theorem explicitly for this situation.

(ii) To illustrate part (i), simulate $M = 5000$ times the expression in the central limit theorem that converges towards the standard normal distribution $\mathcal{N}(0,1)$, for the fixed parameter $\lambda = \frac{1}{2}$, but for different n, namely $n \in \{5, 10, 100, 1000\}$. Plot the respective histogram and the probability density function of $\mathcal{N}(0,1)$. What do you observe?

(iii) Now plot a histogram of the average $\overline{X}_n$ for $n \in \{5, 10, 100, 1000\}$ and $M = 5000$ repetitions. What do you observe?

(b) A random variable X is (standard) *Cauchy-distributed* if its law has density $f_X(x) = \frac{1}{\pi(1+x^2)}$. Repeat the simulations from (a), part (iii) but with i.i.d. Cauchy-distributed random variables $X_1, ..., X_n$. What do you observe? How can this difference be explained?

Hints:

- `rpois()` generates a Poisson-distributed random variable with a chosen parameter λ.
- `rcauchy()` generates a Cauchy-distributed random variable.
- `pnorm()` gives the values of the cumulative distribution function, and `dnorm()` the values of the probability density function of a normal distribution with chosen parameters μ and σ (not σ^2).
- `hist()` plots a histogram. Use `freq =FALSE` for a normalized histogram.
- To compare the histogram to the density of a normal distribution, use the command `curve()` in the form `curve(dnorm(x,...), add=TRUE)`.
- `mean()` calculates the mean of a vector, `sqrt()` the square-root of a number.
- With `function()` you can define a new function in R.
- You can find details for every R-command by putting a ? in front for the respective function (for instance `?hist`).