Correction of Exercises: Week1

APM304 - Asymptotic Statistics (2025-2026)

Ruben Bueno & Romain Poggi

Sebastien Gadat

September 11, 2025

Exercise 1

Consider an experiment consisting of a sequence of Bernoulli trials, which is stopped as soon as r (> 1) successes are observed. Thus all the outcomes in the sample space consist of exactly r successes and $x \geq 0$ failures. We can define a random variable Y that counts the number of failures before the r successes are observed.

1. Derive the frequency (pmf) of Y.
2. Suppose that X and Y are independent Geometric random variables with frequency function

$$f(x) = \theta\ (1-\theta)^x\,, \quad x \in \{0, 1, 2, \ldots\}.$$

   Show that $Z = X + Y$ has a Negative Binomial distribution and identify the parameters of Z.
3. Extend the result of the previous part: if $X_1, \ldots, X_n$ are i.i.d. Geometric random variables, show that $S = \sum_{i=1}^n X_i$ has a Negative Binomial distribution and identify its parameters.

Solution.

1. We are given $Y \sim \text{Ber}\,(p)$ for $p \in (0,1)$. Given $y \in \mathbb{N}$, if $Y = y$, then our experiment consists in:

   - A sequence of length $y + r - 1$, where we have y failures (and $r - 1$ successes)
   - A success

   We necessarily have a success at the end, since otherwise the experiment would have stopped sooner.
   Let's note that the probability of getting y failures out of $y + r - 1$ trials is the probability of a random variable having a Binomial distribution, and we thus get...:

$$\mathbb{P}\,(Y = y) = \binom{y+r-1}{y} p^{r-1}\,(1-p)^y\, p = \binom{y+r-1}{r-1} p^r\,(1-p)^y$$

2. One can compute, given $z \in \mathbb{N}$:

$$\mathbb{P}\,(Z = z) = \sum_{x=0}^{z} \mathbb{P}\,(X = x)\,\mathbb{P}\,(Y = z - x) = \sum_{x=0}^{z} \theta\,(1-\theta)^x\,\theta\,(1-\theta)^{z-x}$$

$$= (z+1)\,\theta^2\,(1-\theta)^z = \binom{z+2-1}{z}\theta^2\,(1-\theta)^z$$

   One can then conclude that $Z \sim \text{NB}\,(2, \theta)$

3. Let's proceed the same way. For any $s \in \mathbb{N}$ let $S_n(s) = \{ \bar{\mathbf{x}} = (x_1, \ldots, x_n) \in \mathbb{N}^n \mid \sum_{i=1}^n x_i = s \}$.

$$\begin{aligned}
\mathbb{P}(S = s) &= \sum_{\bar{\mathbf{x}} \in S_n(s)} \prod_{i=1}^n \mathbb{P}(X_i = x_i) \\
&= \sum_{\bar{\mathbf{x}} \in S_n(s)} \theta^n (1-\theta)^{\sum_{i=1}^n x_i} \\
&= \sum_{\bar{\mathbf{x}} \in S_n(s)} \theta^n (1-\theta)^z \\
&= \binom{s+n-1}{s} \theta^n (1-\theta)^s
\end{aligned}$$

(Using the 'stars and bars theorem'),
One can then conclude that $S \sim \mathrm{NB}(n, \theta)$.

Exercise 2

Suppose that $X_1, X_2, \ldots$ are i.i.d. random variables with moment generating function $m(t) = \mathbb{E}\left(e^{tX_i}\right)$. Let N be a Poisson random variable (independent of the X_i's) with parameter λ and define the compound Poisson random variable

$$S = \sum_{i=1}^N X_i,$$

where $S = 0$ if $N = 0$.

1. Show that the moment generating function of S is
$$\mathbb{E}\left(e^{tS}\right) = \exp\left(\lambda\left(m(t) - 1\right)\right).$$
2. Suppose that the X_i's are Exponential with mean 1 (rate 1) and that $\lambda = 5$. Evaluate $\mathbb{P}(S > 5)$.

Solution.

1. For any $t \in \mathbb{R}$, we get using the **law of total expectation**:
$$\mathbb{E}\left(e^{tS}\right) = \sum_{n=0}^{+\infty} \mathbb{P}(N = n)\, \mathbb{E}\left(e^{tS} \middle| N = n\right)$$
By indepenance of the X_i's, we have for any $n \in \mathbb{N}$:
$$\mathbb{E}\left(e^{tS} \middle| N = n\right) = \mathbb{E}\left(e^{t\sum_{i=1}^n X_i} \middle| N = n\right) = \prod_{i=1}^n \mathbb{E}\left(e^{tX_i}\right) = m(t)^n$$
We get:
$$\mathbb{E}\left(e^{tS}\right) = \sum_{n=0}^{+\infty} \mathbb{P}(N = n) \prod_{i=1}^n \mathbb{E}\left(e^{tX_i}\right) = \sum_{n=0}^{N} e^{-\lambda} \frac{\lambda^n}{n!} m(t)^n$$
We then get:
$$\mathbb{E}\left(e^{tS}\right) = e^{-\lambda} e^{\lambda m(t)} = e^{\lambda(m(t)-1)}$$

2. Let $n \in \mathbb{N}$. We can show (using the MGF) that the sum of n random variables i.i.d $\mathrm{Exp}(\lambda)$ has distribution $\Gamma(n, 1)$. We deduce that the conditional distribution of S given $N = n$ is $\Gamma(n, 1)$. Let $s > 0$, we can then compute:

$$\begin{aligned}
\mathbb{P}(S > s) &= \sum_{n=1}^{+\infty} \mathbb{P}(N = n)\, \mathbb{P}(S > s \mid N = n) \\
&= \sum_{n=1}^{+\infty} \left(\frac{e^{-\lambda}\lambda^n}{n!}\right)\left(\int_s^{+\infty} \frac{y^{n-1}}{n!} e^{-y} dy\right) \\
&= \sum_{n=1}^{+\infty} \left(\frac{e^{-\lambda}\lambda^n}{(n!)^2}\right)\left(\int_s^{+\infty} y^{n-1} e^{-y} dy\right)
\end{aligned}$$

Now if we fix $s = 5$, however $(\int_s^{+\infty} y^{n-1}e^{-y}dy)$ can't be simplified.

$$\mathbb{P}(S > 5) = \sum_{n=1}^{+\infty} \left(\frac{e^{-\lambda}\lambda^n}{(n!)^2}\right)\left(\int_5^{+\infty} y^{n-1} e^{-y} dy\right)$$

Exercise 3

Consider the experiment where a coin is tossed an infinite number of times where the probability of heads on the k-th toss is $\left(\frac{1}{2}\right)^k$. Define X to be the number of heads observed in the experiment.

1. Show that the probability generating function (p.g.f.) of X is

$$p(t) = \prod_{k=1}^{\infty} \left(1 - \frac{1-t}{2^k}\right).$$

Hint: think of X as a sum of independent Bernoulli random variables.

2. Use the result of part (a) to evaluate $\mathbb{P}(X = x)$ for $x = 0, \ldots, 5$.

Solution.

1. For any $k \in \mathbb{N}^*$, let $X_k \sim \mathrm{Ber}\left(\frac{1}{2^k}\right)$. We note that the X_k's are independent random variables. We can model X as a sum of independent Bernouilli random varrialbes by $X = \sum_{k=1}^{+\infty} X_k$.

   For $t \in (0, 1)$, the p.g.f. is given by $p(t) = \mathbb{E}\left(t^X\right) = \mathbb{E}\left(\exp(\log(t))^X\right) = \mathbb{E}\left(e^{\log(t)X}\right)$. We remark that $p(t) = m(\log(t))$, where $m(t)$ is the moment generating function of X.

   For any $s > 0$, we have

$$m(s) = \mathbb{E}\left(\exp\left(s \sum k = 1^{+\infty} X_k\right)\right) = \prod k = 1^{+\infty} e^{sX_k}.$$

   We can compute for any $k \in \mathbb{N}$:

$$\mathbb{E}\left(e^{sX_k}\right) = \left(1 - \frac{1}{2^k}\right) + \frac{1}{2^k}e^s$$

   Therefore:

$$p(t) = \prod_{k=1}^{+\infty} \left(\left(1 - \frac{1}{2^k}\right) + \frac{1}{2^k}t\right) = \prod_{k=1}^{\infty} \left(1 - \frac{1-t}{2^k}\right).$$

2. Since X is an integer random variable, $p(t)$ can also be writen:

$$p(t) = \mathbb{E}\left(t^X\right) = \sum_{k=0}^{+\infty} t^k \, \mathbb{P}\left(X = k\right)$$

We can show with a simple induction, that for all $k \in \mathbb{N}$,

$$\frac{\mathrm{d}^k p}{\mathrm{d}t^k}(0) = k! \mathbb{P}\left(X = k\right),$$

from which we get $\mathbb{P}\left(X = k\right) = \frac{1}{k!} \frac{\mathrm{d}^k p}{\mathrm{d}t^k}(0)$. Let's now simplify $p(t)$.

Define $f(t) = \log\left(p(t)\right) = \sum_{k=1}^{+\infty} \log\left(1 - \frac{1-t}{2^k}\right)$, we thus have $p(t) = e^{f(t)}$, and thus:

- $\mathbb{P}\left(X = 0\right) = e^{f(0)}$
- $\mathbb{P}\left(X = 1\right) = f'(0) e^{f(0)}$
- etc...

Exercise 4

Consider the following method (known as the rejection method) for generating random variables with a density $f(x)$. Suppose that $\gamma(x)$ is a function such that $\gamma(x) \geq f(x)$ for all x, and

$$\int_{-\infty}^{\infty} \gamma(x) \, dx = \alpha < \infty.$$

Then $g(x) = \gamma(x)/\alpha$ is a probability density function. Suppose we generate a random variable X by the following algorithm:

I: Generate a random variable Y with density function $g(x)$.

II: Generate a random variable $U \sim \text{Unif}(0,1)$, independent of Y. If $U \leq \frac{f(Y)}{\gamma(Y)}$ then set $X = Y$; if $U > \frac{f(Y)}{\gamma(Y)}$ then repeat steps I and II.

1. Show that the generated random variable X has density $f(x)$.
2. Show that the number of *rejections* before X is generated has a Geometric distribution. Give an expression for the parameter of this distribution.
3. Show that the rejection method also works if we want to generate from a joint density $f(x)$. (In this case, $U \sim \text{Unif}(0,1)$ as before but now Y is a random vector with density $g(x)$.)

Solution.

1. For any $x \in \mathbb{R}$, we want to show that $\mathbb{P}\left(X \leq x\right) = \int_{-\infty}^{x} f(y) dy$.

   Let $x \in \mathbb{R}$ be fixed. Let us show that $\mathbb{P}\left(X \leq x\right) = \mathbb{P}\left(Y \leq x \mid U \leq \frac{f(Y)}{\gamma(Y)}\right) = \frac{\mathbb{P}\left(Y \leq x, U \leq \frac{f(Y)}{\gamma(Y)}\right)}{\mathbb{P}\left(U \leq \frac{f(Y)}{\gamma(Y)}\right)}$.

   Let's first compute $\mathbb{P}\left(U \leq \frac{f(Y)}{\gamma(Y)}\right)$:

$$\mathbb{P}\left(U \leq \frac{f(Y)}{\gamma(Y)}\right) = \int_{-\infty}^{+\infty} \mathbb{P}\left(U \leq \frac{f(y)}{\gamma(y)}\right) g(y) \, dy = \int_{-\infty}^{+\infty} \frac{f(y)}{\gamma(y)} \frac{\gamma(y)}{\alpha} dy = \frac{1}{\alpha} \int_{-\infty}^{+\infty} f(y) \, dy = \frac{1}{\alpha}.$$

Now let's compute $\mathbb{P}\left(Y \leq x, U \leq \frac{f(Y)}{\gamma(Y)}\right)$:

$$\mathbb{P}\left(Y \leq x, U \leq \frac{f(Y)}{\gamma(Y)}\right) = \int_{y=-\infty}^{x} \int_{u=-\infty}^{+\infty} g(y)\, \mathbf{1}_{\left\{u \leq \frac{f(y)}{\gamma(y)}\right\}}\, du\, dy = \frac{1}{\alpha} \int_{y=-\infty}^{x} f(y)\, dy$$

Thus: $\mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(y)\, dy$.

2. Let N be the random variable counting the number of *rejections* before X has been generated. Let's recall that the probability of acceptance is $\mathbb{P}\left(U \leq \frac{f(Y)}{\gamma(Y)}\right) = p = \frac{1}{\alpha}$.
   One can compute the following, given $k \in \mathbb{N}$:
   $$\mathbb{P}(N = k) = (1-p)^k p = \left(1 - \frac{1}{\alpha}\right)^k \frac{1}{\alpha}$$
   N then follows a **Geometric distribution** with probability of success $\frac{1}{\alpha}$.
3. The logic from Part 1 extends directly to the multivariate case where we generate a random vector $\mathbf{X} \in \mathbb{R}^d$ from a joint density $f(\mathbf{x})$.
   The setup is identical, except that Y is now a random vector $\mathbf{Y}$ with a joint proposal density $g(\mathbf{y}) = \gamma(\mathbf{y})/\alpha$, where $\alpha = \int_{\mathbb{R}^d} \gamma(\mathbf{y}) d\mathbf{y}$. All single integrals in the derivation are replaced by multi-dimensional integrals.
   The probability of accepting a proposed vector $\mathbf{Y}$ is still:
   $$p = \mathbb{P}\left(U \leq \frac{f(\mathbf{Y})}{\gamma(\mathbf{Y})}\right) = \int_{\mathbb{R}^d} \mathbb{P}\left(U \leq \frac{f(\mathbf{y})}{\gamma(\mathbf{y})}\right) g(\mathbf{y}) d\mathbf{y} = \int_{\mathbb{R}^d} \frac{f(\mathbf{y})}{\gamma(\mathbf{y})} \frac{\gamma(\mathbf{y})}{\alpha} d\mathbf{y} = \frac{1}{\alpha}.$$
   The joint cumulative distribution function (CDF) of the accepted vector $\mathbf{X}$ is $\mathbb{P}(\mathbf{X} \leq \mathbf{x})$, where the inequality is component-wise. Following the same conditional probability argument as in **1.**
   $$\mathbb{P}(\mathbf{X} \leq \mathbf{x}) = \frac{\mathbb{P}(\mathbf{Y} \leq \mathbf{x}, \text{Accept})}{p} = \frac{\frac{1}{\alpha} \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} f(y_1, \ldots, y_d) dy_1 \ldots dy_d}{\frac{1}{\alpha}} = F_{\mathbf{X}}(\mathbf{x}).$$
   By the fundamental theorem of calculus for multiple variables, differentiating the joint CDF with respect to $x_1, \ldots, x_d$ yields the joint density $f(\mathbf{x})$. Thus, the method is valid for joint densities.

Exercise 5

Suppose we want to generate random variables with a Cauchy distribution. We can generate independent random variables V and W where $\mathbb{P}(V = 1) = \mathbb{P}(V = -1) = \frac{1}{2}$ and W has density

$$g(x) = \frac{2}{\pi(1+x^2)} \quad \text{for } |x| \leq 1.$$

(The variable W can be generated using the rejection method.) Define $X = W^V$. Show that X has a Cauchy distribution.

We propose two ways to solve the exercise. The first solution uses the method of test functions. In practice, this method is almost always used to determine the distributions of non-discrete random variables (with the typical exception of cases like real random variables defined by minimums or maximums, etc.).

Theorem 1: Method of test functions

- **Case: Real-valued random variable with a density** Let Y be a real-valued random variable and let $f : \mathbb{R} \to \mathbb{R}$ be a nonnegative function. Suppose that for every bounded and continuous function $h : \mathbb{R} \to \mathbb{R}$ we have
$$\mathbb{E}\left[h(Y)\right] = \int_{\mathbb{R}} h(x)\, f(x)\, dx.$$
Then Y has a (Lebesgue) density and f is a version of the density of Y at point x.

- **Case: Random vector with a density in $\mathbb{R}^n$** Let $X = (X_1, \ldots, X_n)$ be an $\mathbb{R}^n$-valued random variable and let $f : \mathbb{R}^n \to \mathbb{R}$ be a nonnegative function. Suppose that for every bounded and continuous function $h : \mathbb{R}^n \to \mathbb{R}$ we have
$$\mathbb{E}\left[h(X)\right] = \int_{\mathbb{R}^n} h(x_1, \ldots, x_n)\, f(x_1, \ldots, x_n)\, dx_1 \cdots dx_n.$$
Then $X = (X_1, \ldots, X_n)$ has a (Lebesgue) density and $f(x_1, \ldots, x_n)$ is a version of the density of X at the point $(x_1, \ldots, x_n)$.

Solution. To prove that $X = W^V$ has the standard Cauchy density, $f(x) = \frac{1}{\pi(1+x^2)}$, we will use the property provided. We must show that for any bounded, continuous function h, the expectation $\mathbb{E}[h(X)]$ equals $\int_{-\infty}^{\infty} h(x) f(x) dx$.

First, we find the expectation of $h(X)$ by conditioning on V:

$$\mathbb{E}\left(h(X)\right) = \frac{1}{2}\mathbb{E}[h(W)] + \frac{1}{2}\mathbb{E}[h(1/W)]$$

Using the density of W, $g(w) = \frac{2}{\pi(1+w^2)}$ on $[-1, 1]$, we express this as a single integral:

$$\mathbb{E}\left(h(X)\right) = \frac{1}{2}\int_{w=-1}^{1} \left[h(w) + h(1/w)\right] g(w) dw = \frac{1}{\pi}\int_{w=-1}^{1} \frac{h(w) + h(1/w)}{1+w^2} dw$$

We can split this into two integrals. For the integral involving $h(1/w)$, we use the substitution $x = 1/w$, which means $dw = -dx/x^2$. This maps the domain $w \in [-1, 0) \cup (0, 1]$ to $x \in (-\infty, -1] \cup [1, \infty)$. The integral becomes:

$$\int_{w=-1}^{1} \frac{h(1/w)}{1+w^2} dw = \int_{|x|\geq 1} \frac{h(x)}{1+(1/x)^2}\frac{dx}{x^2} = \int_{|x|\geq 1} \frac{h(x)}{1+x^2} dx$$

Substituting this result back into our expression for $\mathbb{E}[h(X)]$ allows us to combine the integration domains:

$$\mathbb{E}\left(h(X)\right) = \frac{1}{\pi}\left[\int_{-1}^{1} \frac{h(x)}{1+x^2} dx + \int_{|x|\geq 1} \frac{h(x)}{1+x^2} dx\right] = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{h(x)}{1+x^2} dx$$

This is precisely the required form:

$$\mathbb{E}\left(h(X)\right) = \int_{x=-\infty}^{\infty} h(x)\left(\frac{1}{\pi(1+x^2)}\right) dx$$

Since this holds for any bounded and continuous function h, the property confirms that X has the density of a standard Cauchy distribution.

Solution. We want to prove, given $x \in \mathbb{R}^*$ (The case $x = 0$ is trivial):

$$\begin{aligned}\mathbb{P}(X \le x) = \mathbb{P}\left(W^V \le x\right) &= \mathbb{P}(W \le x \mid V = 1)\,\mathbb{P}(V = 1) + \mathbb{P}\left(W^{-1} \le x \mid V = -1\right)\mathbb{P}(V = -1)\\ &= \frac{\mathbb{P}(W \le x) + \mathbb{P}\left(\frac{1}{W} \le x\right)}{2}\end{aligned}$$

We can separate this into two cases:

- Case $x < 0$:
$$\mathbb{P}(X \le x) = \frac{\mathbb{P}(W \le x) + \mathbb{P}\left(W \ge \frac{1}{x}\right)}{2},$$
in that case, looking at the values of $P(W \le x)$ and $P(W \ge \frac{1}{x})$ when $|x|$ is strictly smaller, equal or strictly greater than 1, we obtain that X is $\text{Cauchy}\,(0,1)$.

- Case $x > 0$:
$$\begin{aligned}\mathbb{P}(X \le x) &= \frac{\mathbb{P}(W \le x) + \left(\mathbb{P}\left(W \ge \frac{1}{x} \mid W > 0\right)\mathbb{P}(W > 0) + \mathbb{P}\left(\frac{1}{W} \le x \mid W < 0\right)\mathbb{P}(W < 0)\right)}{2}\\ &= \frac{\mathbb{P}(W \le x) + \left(\mathbb{P}\left(W \ge \frac{1}{x} \mid W > 0\right)\mathbb{P}(W > 0) + \mathbb{P}(W < 0)\right)}{2}\\ &= \frac{\mathbb{P}(W \le x) + \frac{1}{2}\left(1 + 2\,\mathbb{P}\left(W \ge \frac{1}{x}\right)\right)}{2}\end{aligned}$$
(since the pdf of W is even). Looking at the values of $P(W \le x)$ and $P(W \ge \frac{1}{x})$ when $|x|$ is strictly smaller, equal or strictly greater than 1, we obtain that X is $\text{Cauchy}\,(0,1)$.

Exercise 6

Suppose that X and Y are independent Uniform random variables on $[0,1]$.

1. Find the density function of $X + Y$.
2. Find the density function of XY.

Solution.

1. Definie $Z = X + Y$. Since $X, Y \in [0,1]$, $Z \in [0,2]$. For $t \in [0,1]$:
$$\mathbb{P}(Z \le t) = \int_0^1 \int_0^1 \mathbf{1}_{\{x+y \le t\}}\,dx\,dy$$
Since $t \le 1$, $\mathbb{P}(Z \le t) = \frac{t^2}{2}$
Now $t \in [1,2]$
Define $X' = 1 - X$, and $Y' = 1 - Y$. One can notice that both random variables are iid, with $U([0,1])$ distribution.
We have:
$$\mathbb{P}(X + Y \le t) = \mathbb{P}(1 - X' + 1 - Y' \le t) = \mathbb{P}(X' + Y' \ge 2 - t) = 1 - \frac{(2-t)^2}{2}$$

To get the pdf, we just need to differentiate the cdf. Therefore,

$$f_{X+Y}(t) = \begin{cases} t & \text{if } 0 \leq t \leq 1 \\ 2-t & \text{if } 1 < t \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

2. Let $t \in [0,1]$:

$$\begin{aligned} \mathbb{P}(XY \leq t) &= \int_{x=0}^{1} \mathbb{P}(xY \leq t)\, dx \\ &= \int_{x=0}^{t} \mathbb{P}(xY \leq t)\, dx + \int_{x=t}^{1} \mathbb{P}(xY \leq t)\, dx \\ &= \int_{x=0}^{t} \int_{y=0}^{1} dy\, dx + \int_{x=t}^{1} \int_{y=0}^{\frac{t}{x}} dy\, dx \\ &= t - t\, \ln(t) \end{aligned}$$

We thus get $f(t) = -\ln(t)$.

Exercise 7

$g(x)$ is a convex function if for all $t \in [0,1]$, $\quad g(tx + (1-t)y) \leq t\, g(x) + (1-t)\, g(y)$

1. Let $\mathbf{X}$ be a random vector with well-defined expected value $\mathbb{E}(\mathbf{X})$. Show that $\mathbb{E}[g(\mathbf{X})] \geq g(\mathbb{E}(\mathbf{X}))$ for any convex function g.
2. Let $g(\mathbf{x}) = \max(x_1, \ldots, x_k)$. Show that g is a convex function and hence

$$\mathbb{E}[\max(X_1, \ldots, X_k)] \geq \max(\mathbb{E}(X_1), \ldots, \mathbb{E}(X_k))$$

for any random variables $X_1, \ldots, X_k$.

Solution.

1. If f is a convex function, then there exists a function u such that for all $(x, y) \in \mathbb{R}$,

$$f(x) \geq f(y) + u(y)(y-x)$$

(u is called a "sub-differential", and if f is differentiable, $u = f'$)

If we set $x = X$, and $y = \mathbb{E}(X)$, we get that:

$$f(X) \geq f(\mathbb{E}(X)) + u(\mathbb{E}(X))(X - \mathbb{E}(X))$$

Then taking the expectation of both sides of the inequality, we get

$$\begin{aligned} \mathbb{E}(f(X)) &\geq \mathbb{E}(f(\mathbb{E}(X)) + u(\mathbb{E}(X))(X - \mathbb{E}(X))) \\ &= \mathbb{E}(f(\mathbb{E}(X))) + u(\mathbb{E}(X))\, \mathbb{E}(X - \mathbb{E}(X)) \\ &= \mathbb{E}(f(X)) \end{aligned}$$

2. Let $k \in \mathbb{N}$. The result is immediate if we show the $\bar{\mathbf{x}} \in \mathbb{R}^n \mapsto \max(\bar{\mathbf{x}})$ is convex.
   Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, and $t \in [0,1]$. Since for all $i \in [\![0,k]\!]$, $x_i \le \max(\mathbf{x})$ and $y_i \le \max(\mathbf{y})$, we get that
   $$\max(t\mathbf{x} + (1-t)\mathbf{y}) \le t\max(\mathbf{x}) + (1-t)\max(\mathbf{y}).$$
   Thus, $\bar{\mathbf{x}} \in \mathbb{R}^n \mapsto \max(\bar{\mathbf{x}})$ is convex, and we conclude using **Jensen's inequality**.

Exercise 8

Suppose that X and Y are random variables such that both $\mathbb{E}\left(X^2\right)$ and $\mathbb{E}\left(Y^2\right)$ are finite. Define
$$g(t) = \mathbb{E}\left([Y + tX]^2\right).$$

1. Show that $g(t)$ is minimized at $t = -\dfrac{\mathbb{E}(XY)}{\mathbb{E}(X^2)}$.
2. Show that $[\mathbb{E}(XY)]^2 \le \mathbb{E}\left(X^2\right)\mathbb{E}\left(Y^2\right)$ (the Cauchy–Schwarz inequality).
3. Use part (b) to show that $|\text{Corr }(X,Y)| \le 1$.

Solution.

1. We have:
   $$\begin{aligned} g(t) &= \mathbb{E}\left((Y+tX)^2\right) = \mathbb{E}\left(Y^2 + 2tXY + t^2X^2\right) \\ &= t^2\mathbb{E}\left(X^2\right) + 2t\mathbb{E}(XY) + \mathbb{E}\left(Y^2\right) \end{aligned}$$
   Thus:
   $$g'(t) = 2t\mathbb{E}(X^2) + 2\mathbb{E}(XY)$$
   Then, since $g'(-\frac{\mathbb{E}(XY)}{\mathbb{E}(X^2)}) = 0$ and g is strictly convex, g is minimized at $t = -\dfrac{\mathbb{E}(XY)}{\mathbb{E}(X^2)}$.
2. We have for any t: $g(t) \ge 0$, so that
   $$\begin{aligned} 0 \le g\left(\frac{-\mathbb{E}(XY)}{\mathbb{E}(X^2)}\right) &= \mathbb{E}\left(X^2\right)\left(\frac{-\mathbb{E}(XY)}{\mathbb{E}(X^2)}\right)^2 + 2\,\mathbb{E}(XY)\,\frac{-\mathbb{E}(XY)}{\mathbb{E}(X^2)} + \mathbb{E}\left(Y^2\right) \\ &= \frac{-\mathbb{E}(XY)^2 + \mathbb{E}\left(X^2\right)\mathbb{E}\left(Y^2\right)}{\mathbb{E}(X^2)} \end{aligned}$$
   Rewriting this differently, we get the **Cauchy-Schwarz inequality.**
3. First, if $\mathbb{E}(X) = \mathbb{E}(Y) = 0$, this is a direct implication from **Cauchy-Schwarz inequality**.
   Now let's assume that either $\mathbb{E}(X) \neq 0$ or $\mathbb{E}(Y) \neq 0$ and define $X^* := \mathbb{E}(X) - X$, $Y^* := \mathbb{E}(Y) - Y$. Notice that since $\mathbb{E}(X)$ and $\mathbb{E}(Y)$ are constants, we thus have that
   - $\text{Var}(X) = \text{Var}(X^*)$,
   - $\text{Cov}(X,Y) = \text{Cov}(X^*,Y^*)$,
   - $\text{Var}(Y) = \text{Var}(Y^*)$,
   - $\mathbb{E}(X^*) = \mathbb{E}(Y^*) = 0$.

   Thus $\text{Corr}(X,Y) = \text{Corr}(X^*,Y^*)$**, using the previous result, we conclude the proof.**