Midterm exam: solutions

Please pay attention to the presentation of your answers! **(2 points)**

Exercise 1. Quiz. (18 points)

Answer each yes/no question below (1 pt) and provide a short justification (proof or counter-example) for your answer (2 pts).

a) Let X, Y be two random variables defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathcal{G} = \sigma(X) \cap \sigma(Y)$ [fact: it can be shown that $\mathcal{G}$ is a σ-field]. Is it true that $\{X \le Y\} \in \mathcal{G}$?

Answer: No. Take for example $\Omega = \{1,2,3\}$, $X(\omega) = \omega$ and $Y(\omega) = 2$ for every $\omega \in \Omega$. Then $\mathcal{G} = \sigma(X) \cap \sigma(Y) = \sigma(Y) = \{\emptyset, \Omega\}$, but $\{X \le Y\} = \{\omega \in \Omega : X(\omega) \le Y(\omega)\} = \{1,2\} \notin \mathcal{G}$.

b) Let X, Y be two independent random variables defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Is it always true that $\sigma(X+Y) = \sigma(X,Y)$?

Answer: No. Take for example $\Omega = \{1,2\}^2$, $X(\omega) = \omega_1$ and $Y(\omega) = -\omega_2$. Then $\{X+Y = 0\} = \{(1,1),(2,2)\}$, and so $\sigma(X+Y) = \sigma(\{(1,1),(2,2)\}, \{(1,2)\}, \{(2,1)\}) \neq \sigma(X,Y) = \mathcal{P}(\Omega)$ (in addition, note that the fact that X and Y are independent does not play a role here).

c) Let X be a continuous random variable whose pdf p_X is a continuous function on $\mathbb{R}$. Let now $Y = X^2$. Is it always true that the pdf p_Y is also a continuous function on $\mathbb{R}$?

Answer: No. Take for example $X \sim \mathcal{N}(0,1)$, whose pdf $p_X(x) = \frac{1}{\sqrt{2\pi x}} \exp(-x^2/2)$ is continuous. Then $Y = X^2$ has pdf

$$p_Y(y) = \begin{cases} \frac{1}{\sqrt{2\pi y}} \exp(-y/2) & \text{if } y \ge 0 \\ 0 & \text{if } y < 0 \end{cases}$$

which is discontinuous in $y = 0$.

d) Let F be a generic cdf. Is it always true that the function $G : \mathbb{R} \to [0,1]$ defined as

$$G(t) = F(t^3 + 3t^2 + 3t + 1), \quad t \in \mathbb{R}$$

is also a cdf ?

Answer: Yes. Actually, the map $t \mapsto t^3 + 3t^2 + 3t + 1 = (t+1)^3$ is non-decreasing and going from $-\infty$ to $+\infty$, thus the properties of the cdf F are preserved for G.

e) Let X, Y, Z be three square-integrable random variables defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$, each with variance 2. Is it possible that $\text{Cov}(X,Y) = \text{Cov}(X,Z) = \text{Cov}(Y,Z) = -1$?

Answer: Yes, as we can check that the covariance matrix of the random vector (X,Y,Z) is positive semi-definite in this case:

$$2\,(c_1^2 + c_2^2 + c_3^2) - 2\,(c_1c_2 + c_1c_3 + c_2c_3) = (c_1 - c_2)^2 + (c_1 - c_3)^2 + (c_2 - c_3)^2 \ge 0$$

for any $c_1, c_2, c_3 \in \mathbb{R}$.

Other possibilities:
- take A, B, C i.i.d.$\sim \mathcal{N}(0,1)$ random variables and $X = A - B$, $Y = B - C$, $Z = C - A$.
- take X, Y with Var 2 and Cov -1; take now $Z = -X - Y$.

f) Let $(X_n,\, n \geq 1)$ and $(Y_n,\, n \geq 1)$ be two sequences of random variables that both converge in probability to the same random variable X. Is it always true that $X_n - Y_n \underset{n\to\infty}{\overset{\mathbb{P}}{\to}} 0$?

Answer: Yes. Indeed, we have for any $\varepsilon > 0$:

$$\begin{aligned}\mathbb{P}(\{|X_n - Y_n| \geq \varepsilon\}) &= \mathbb{P}(\{|X_n - X + X - Y_n| \geq \varepsilon\}) \leq \mathbb{P}(\{|X_n - X| \geq \varepsilon/2\} \cup \{|Y_n - X| \geq \varepsilon/2\}) \\ &\leq \mathbb{P}(\{|X_n - X| \geq \varepsilon/2\}) + \mathbb{P}(\{|Y_n - X| \geq \varepsilon/2\}) \underset{n\to\infty}{\to} 0\end{aligned}$$

by the assumptions made.

Exercise 2. (15 points)

Let X, Y be two i.i.d. $\mathcal{N}(0,1)$ random variables, and Z be independent of X, Y and such that $\mathbb{P}\{Z = +1\}) = \mathbb{P}\{Z = -1\}) = 1/2$.

a) $(X + Z\,Y, Y)$ is a continuous random vector: compute its joint pdf.

Answer: The computation gives:

$$\begin{aligned}\mathbb{P}(\{X + ZY \leq t, Y \leq s\}) &= \int_{-\infty}^{s} dy\, p_Y(y)\, \mathbb{P}(\{X + Zy \leq t\}) \\ &= \int_{-\infty}^{s} dy\, p_Y(y) \left(\frac{1}{2}\,\mathbb{P}(\{X \leq t - y\}) + \frac{1}{2}\,\mathbb{P}(\{X \leq t + y\})\right)\end{aligned}$$

so

$$p_{X+ZY,Y}(t,s) = \frac{d^2}{dsdt}\,\mathbb{P}(\{X + ZY \leq t, Y \leq s\}) = \frac{1}{2}\,p_Y(s)\,(p_X(t-s) + p_X(t+s))$$

b) Is it true $X + Z\,Y$ is a Gaussian random variable ? Justify.

Answer: Yes. Because Y $\mathcal{N}(0,1)$ and Z is independent of Y, $ZY \sim \mathcal{N}(0,1)$; then, the sum of two independent Gaussians random variables is also Gaussian.

c) Is it true $(X + Z\,Y, Y)$ is a Gaussian random vector ? Justify.

Answer: No. Consider the linear combination $X + ZY + Y = X + (1+Z)Y$. Conditioned on the value of Z, this random variable is $\mathcal{N}(0,1)$ or $\mathcal{N}(0,5)$ (as an explicit computation of the pdf also shows); certainly not a Gaussian.

d) Compute $\text{Cov}(X + ZY, Y)$.

Answer: As all random variables are centered, we get:

$$\text{Cov}(X + ZY, Y) = \mathbb{E}(XY + ZY^2) = \mathbb{E}(X)\,\mathbb{E}(Y) + \mathbb{E}(Z)\,\mathbb{E}(Y^2) = 0$$

e) Is it true that $X + ZY$ and Y are independent random variables ? Justify.

Answer: No. For example, $\mathbb{P}(\{X + ZY \geq 0\})\,\mathbb{P}(\{Y \geq 0\}) = 1/4$ by symmetry, but

$$\begin{aligned}\mathbb{P}(\{X + ZY \geq 0, Y \geq 0\}) &= \frac{1}{2}\,\mathbb{P}(\{X + Y \geq 0, Y \geq 0\}) + \frac{1}{2}\,\mathbb{P}(\{X - Y \geq 0, Y \geq 0\}) \\ &= \mathbb{P}(\{X \geq Y, Y \geq 0\}) + \frac{1}{2}\,\mathbb{P}(\{|X| \leq Y, Y \geq 0\}) > \frac{1}{4}\end{aligned}$$

Exercise 3. (15 + 3 points)

Hint for this exercise (not necessarily needed): For $x \in \mathbb{R}$, $\exp(x) = \lim_{n\to\infty}(1 + x/n)^n$.

Let X be a random variable whose characteristic function is given by $\phi_X(t) = \max(1 - |t|, 0)$ for $t \in \mathbb{R}$.

Fact: ϕ_X is a characteristic function: we do not ask you to prove it.

a) Is X a continuous random variable ?

Answer: Yes, as $\int_{\mathbb{R}} dt\,|\phi_X(t)| < +\infty$.

b) What is the value of $\mathbb{E}(|X|)$ and $\mathbb{E}(X^2)$?

Answer: $\mathbb{E}(|X|) = +\infty$, as ϕ_X is not differentiable in $t = 0$; therefore, $\mathbb{E}(X^2) = +\infty$ also.

Let now $(X_n,\, n \geq 1)$ be a sequence of i.i.d. random variables such that $X_n \sim X$ for every $n \geq 1$.

c) For $n \geq 1$, define $Y_n = \dfrac{X_1 + X_2 + \ldots + X_n}{n}$. Compute the characteristic function of Y_n.

Answer: By independence, we have

$$\phi_{Y_n}(t) = \phi_{X_1+\ldots+X_n}(t/n) = \phi_{X_1}(t/n)\cdots\phi_{X_n}(t/n) = (\phi_X(t/n))^n = (\max(1 - |t/n|, 0)^n$$

d) Let $n \geq 1$ be fixed. For what values of $t \in \mathbb{R}$ does it hold that $\phi_{Y_n}(t) = 0$?

Answer: $\phi_{Y_n}(t) = 0$ for $|t| \geq n$.

e) Does there exist $\mu \in \mathbb{R}$ such that $Y_n \underset{n\to\infty}{\to} \mu$ almost surely ? Justify.

Answer: No. Two possible justifications here: 1) $\mathbb{E}(|X_1|) = +\infty$ so by the (extension of the) strong law of large numbers, Y_n diverges a.s. 2) the characteristic function of Y_n converges to $\exp(-|t|)$ (cf. hint), which is the characteristic function of the Cauchy distribution. The limit of Y_n can therefore not be constant.

BONUS f*) Compute the distribution of X.

Answer: $p_X(x) = \dfrac{1 - \cos(x)}{\pi\, x^2} = \dfrac{1}{2\pi}\,\dfrac{\sin(x/2)^2}{(x/2)^2}$ [this computation is not trivial].

Exercise 4. (10 points)

Hint for this exercise: For $X \sim \mathcal{N}(0,1)$ and $t \geq 0$, it holds that $F_X(t) \geq 1 - \exp(-t^2/2)$.

Let $(\sigma_n,\, n \geq 1)$ be a sequence of positive numbers and $(Z_n,\, n \geq 1)$ be a sequence of independent random variables such that $Z_n \sim \mathcal{N}(0, \sigma_n^2)$. Let also $\mu \in \mathbb{R}$ and $X_n = \mu + Z_n$ for $n \geq 1$.

a) Show that if $\sigma_n \underset{n\to\infty}{\to} 0$, then $X_n \overset{\mathbb{P}}{\underset{n\to\infty}{\to}} \mu$.

Answer: In this case, we obtain by Chebyshev's inequality with $\varphi(x) = x^2$ that for any $\varepsilon > 0$:

$$\mathbb{P}(\{|X_n - \mu| \geq \varepsilon\}) = \mathbb{P}(\{|Z_n| \geq \varepsilon\}) = \frac{\mathbb{E}(Z_n^2)}{\varepsilon^2} = \frac{\sigma_n^2}{\varepsilon^2} \underset{n\to\infty}{\to} 0$$

therefore the conclusion.

b) Assume now that $\sigma_n = \frac{1}{\log(n+1)}$ for $n \geq 1$. Is it true in this case that $X_n \underset{n\to\infty}{\to} \mu$ almost surely ? If yes, prove it; if no, explain why.

Answer: Yes, indeed:

$$\mathbb{P}(\{|X_n - \mu| \geq \varepsilon\}) = \mathbb{P}(\{|Z_n| \geq \varepsilon\}) = \mathbb{P}\left(\left\{|Z| \geq \frac{\varepsilon}{\sigma_n}\right\}\right)$$

where $Z \sim \mathcal{N}(0,1)$. Now, by the symmetry of Z and by the hint:

$$\mathbb{P}\left(\left\{|Z| \geq \frac{\varepsilon}{\sigma_n}\right\}\right) = 2\,\mathbb{P}\left(\left\{Z \geq \frac{\varepsilon}{\sigma_n}\right\}\right) = 2\left(1 - F_Z\left(\frac{\varepsilon}{\sigma_n}\right)\right) \leq 2\,\exp(-\varepsilon^2/2\sigma_n^2)$$

As $\sigma_n = \frac{1}{\log(n+1)}$, the above probability decays more than polynomially to 0, so

$$\sum_{n\geq 1} \mathbb{P}(\{|X_n - \mu| \geq \varepsilon\}) < +\infty$$

which allows to conclude by the Borel-Cantelli lemma that $X_n \underset{n\to\infty}{\to} \mu$ almost surely.

c) Does any of the conclusions of parts a) and b) rely on the fact that the random variables Z_n are independent ? Explain.

Answer: No. The independence assumption is clearly not needed in the previous computations (for example, we could replace Z_n by $\sigma_n\, Z$ in all the above computations).