Final exam: solutions

Exercise 1. Quiz. (25 points) Answer each short question below. For yes/no questions explicitly say if the statement is true of false and provide a brief justification (proof or counter-example) for your answer. For other questions, provide the result of you computation, as well as a brief justification for your answer.

a) Let $\Omega = \{1, 2, \ldots, 6\}$ and $\mathcal{A} = \{\{1, 2, 3\}, \{1, 3, 5\}\}$. Let $\mathcal{F} = \sigma(\mathcal{A})$ be the σ-field generated by $\mathcal{A}$. What are the atoms of $\mathcal{F}$?

Solution: The atoms of $\sigma(\mathcal{A})$ are $\{\{2\}, \{5\}, \{1, 3\}, \{4, 6\}\}$. Indeed, you can check that each of these sets could be obtained with unions and intersections of the following sets $\{\emptyset, \{1, 2, 3\}, \{1, 3, 5\}, \Omega\}$. Thus, they must be in $\sigma(\mathcal{A})$. On the other hand, any smaller sets (such as $\{1\}$, $\{3\}$, $\{4\}$, or $\{6\}$) could not be obtained in this way. And so, the smallest σ-field containing $\mathcal{A}$ will not contain them.

b) Let $\Omega = [0, 1]^2$, $\mathcal{F} = \mathcal{B}([0, 1]^2)$, and $\mathbb{P}$ be the probability measure on $(\Omega, \mathcal{F})$ defined as

$$\mathbb{P}\left(]a, b[\times]c, d[\right) = (b - a) \cdot (d - c), \text{ for } 0 \leq a < b \leq 1 \text{ and } 0 \leq c < d \leq 1$$

which can be extended uniquely to all Borel sets in $\mathcal{B}([0, 1]^2)$, according to Caratheodory's extension theorem. Let us now consider the following random variable defined on $(\Omega, \mathcal{F}, \mathbb{P})$:

$$X(\omega_1, \omega_2) = \frac{\omega_1 - \omega_2}{2}.$$

Compute the cdf F_X of X.

Solution: First, note that the range of the random variable X is $\left[-\frac{1}{2}, \frac{1}{2}\right]$. Thus, the CDF $F_X(t) = 0$ for $t < -\frac{1}{2}$ and $F_X(t) = 1$ for $t \geq \frac{1}{2}$.

Now, for $t \in \left[-\frac{1}{2}, \frac{1}{2}\right]$, we have:

$$\begin{aligned} F_X(t) = \mu_X((-\infty, t]) &= \mathbb{P}(\{(\omega_1, \omega_2) \in [0, 1]^2 : X(\omega_1, \omega_2) \leq t\}) \\ &= \mathbb{P}(\{(\omega_1, \omega_2) \in [0, 1]^2 : \omega_1 - \omega_2 \leq 2t\}) \end{aligned}$$

Note that the area $\{(\omega_1, \omega_2) \in [0, 1]^2 : \omega_1 - \omega_2 \leq 2t\}$ represents different shapes in $[0, 1] \times [0, 1]$ for positive and negative values of $2t$. Thus, we divide our analysis into two cases:

Case 1: $-\frac{1}{2} < t \leq 0$:

The area $\{(\omega_1, \omega_2) \in [0, 1]^2 : \omega_1 - \omega_2 \leq 2t\}$ represents a right-angled triangle (Δ_1) is an element of the sigma field $\mathcal{F} = \mathcal{B}([0, 1]^2)$. Thus, the probability measure $\mathbb{P}(\Delta_1)$ is given by its area. Thus,

$$F_X(t) = Area(\Delta_1) = \frac{1}{2}(1 + 2t)(1 + 2t)$$

Case 2: $0 < t \leq \frac{1}{2}$:

The area $\{(\omega_1, \omega_2) \in [0,1]^2 : \omega_1 - \omega_2 \leq 2t\}$ represents a pentagon (Δ_2) in this case which is again an element of the sigma field $\mathcal{F} = \mathcal{B}([0,1]^2)$. Thus, the probability measure $\mathbb{P}(\Delta_2)$ is given by its area which can be easily computed as:

$$F_X(t) = Area(\Delta_2) = 1 - \frac{1}{2}(1-2t)(1-2t)$$

Thus, the CDF of the random variable X is the following:

$$F_X(t) = \begin{cases} 0 & \text{if } t \leq -\frac{1}{2}, \\ \frac{1}{2}(1+2t)^2 & \text{if } -\frac{1}{2} < t \leq 0 \\ 1 - \frac{1}{2}(1-2t)^2 & \text{if } 0 < t \leq \frac{1}{2} \\ 1 & \text{if } t > \frac{1}{2} \end{cases}$$

c) Let X be a random variable supported on $\{0, 1\}$ with $\mathbb{P}(\{X = 1\}) = \mathbb{P}(\{X = -1\}) = \frac{1}{2}$. Let $Z \sim \mathcal{N}(0,1)$ and assume that X and Z are independent. Then, is (XZ, Z) a Gaussian random vector?

Answer: No.

Consider the distribution of the random variable $XZ+Z$. We have that $XZ+Z = 0$ with probability $\frac{1}{2}$. Thus, this is not a continuous distribution and therefore it is not a Gaussian random variable. Recall that the sum of the components of a Guassian random vector has Gaussian distribution. Therefore, this is not a Gaussian random vector.

Lets compute the CDF (distribution) of the random variable XZ.

$$\begin{aligned} \mathbb{P}(\{XZ \leq t\}) &= \mathbb{P}(\{XZ \leq t\} | \{X = -1\}).\mathbb{P}(\{X = -1\}) + \mathbb{P}(\{XZ \leq t\} | \{X = 1\}).\mathbb{P}(\{X = 1\}) \\ &= \frac{1}{2}\mathbb{P}(\{Z \geq -t\}) + \frac{1}{2}\mathbb{P}(\{Z \leq t\}) = \mathbb{P}(\{Z \leq t\}) \end{aligned}$$

We see here that both X and XZ are continuous random variable. However, we see that for a diagonal line in $\mathbb{R}^2$ (which has Lebesgue measure 0 (i.e., $|\Delta| = 0$),

$$\mathbb{P}(\{(XZ, Z) \in \Delta\}) = \mathbb{P}(\{XZ + Z = 0\}) = \mathbb{P}(\{X(1+Z) = 0\}) = \mathbb{P}(\{Z = -1\}) = \frac{1}{2}.$$

Thus, (XZ, Z) is a not a continuous random vector.

(See Example 6.2 and Example 6.7 from the lecture notes for details.)

d) Let X and Z be as in part (c). Then, is (XZ, Z) a continuous random vector?

Answer: No. See solution above.

e) Let X and Y be integrable random variables. If $Y = g(X)$ for some measurable function $g : \mathbb{R} \to \mathbb{R}$, then is it true that $\mathbb{E}(X|Y) = h(X)$ for some function $h : \mathbb{R} \to \mathbb{R}$?

Answer: Yes. In fact, $\mathbb{E}(X|Y) = f(Y)$ for some measurable function $f : \mathbb{R} \to \mathbb{R}$. Hence, since $Y = g(X)$, we have that $\mathbb{E}(X|Y) = f(Y) = f(g(X)) = h(X)$ for $h = f \circ h$.

f) Let X and Y be two independent Bernoulli random variables with parameter $0 \leq p \leq 1$ Let Z be defined as

$$Z = \begin{cases} 1, & \text{if } X + Y = 0, \\ 0, & \text{otherwise.} \end{cases}$$

Are $\mathbb{E}(X|Z)$ and $\mathbb{E}(Y|Z)$ independent?

Answer: No (in general).

Yes if $p = 0$ or 1, no otherwise. In fact, note that $\mathbb{E}(X|Z) = f(Z)$ and $\mathbb{E}(Y|Z) = g(Z)$. Furthermore, by symmetry of the problem, we must have $f(Z) = g(Z)$, that is, $\mathbb{E}(X|Z)$ and $\mathbb{E}(Y|Z)$ are actually the same random variable. Then, a random variable is independent of itself if and only if it is constant. In our case, this is true if and only if $\mathbb{E}(X|Z = 0) = \mathbb{E}(X|Z = 1)$, which, in turn, is true if and only if $p = 0$ or 1.

g) Let $(S_n, n \in \mathbb{N})$ be the simple symmetric random walk and let $(\mathcal{F}_n, n \in \mathbb{N})$ be its natural filtration. Define a random time

$$T = \inf\{n \colon S_n = S_{n-2}, n \geq 2\}.$$

Is T a stopping time?

Answer: Yes. This is a stopping time since

$$\{T = n\} = \{S_n = S_{n-2}\} \bigcap \left(\bigcap_{2 \leq k < n} S_k \neq S_{k-2} \right).$$

Since $\{S_n = S_{n-2}\} \in \mathcal{F}_n$ and $\{S_k = S_{k-2}\} \in \mathcal{F}_k \subset \mathcal{F}_n$, the event $\{T = n\} \in \mathcal{F}_n$.

Exercise 2. (15 points)

Let X and Y be random variables defined on common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Define

$$d(X, Y) = \mathbb{E}\left(\log_2 \left(1 + \frac{|X - Y|}{1 + |X - Y|} \right) \right).$$

a) First, we would like to confirm that $d(X, Y)$ is a distance metric. Show that $d(X, Y)$ satisfies the triangle inequality. That is, $d(X, Z) \leq d(X, Y) + d(Y, Z)$ for any X, Y, and Z.

Hint: the function $f(x) = \log_2(1 + x)$ is sub-additive, e.g. $f(x + y) \leq f(x) + f(y)$.

Solution: For all $x, y, z \in \mathbb{R}$ we have

$$\begin{aligned}
\log_2\left(1+\frac{|x-z|}{1+|x-z|}\right) &= \log_2\left(1+\frac{|x-y+y-z|}{1+|x-y+y-z|}\right) \\
&\le \log_2\left(1+\frac{|x-y|+|y-z|}{1+|x-y|+|y-z|}\right) \\
&\le \log_2\left(1+\frac{|x-y|}{1+|x-y|}+\frac{|y-z|}{1+|y-z|}\right) \\
&\le \log_2\left(1+\frac{|x-y|}{1+|x-y|}\right)+\log_2\left(1+\frac{|y-z|}{1+|y-z|}\right)
\end{aligned}$$

where the first inequality follows from the fact that $\log_2(1+x)$ is an increasing function in x and the last inequality follows from the hint. Now, since the inequality holds for $X(\omega), Y(\omega), Z(\omega)$ for every $\omega \in \Omega$, we can take the expectation of both sides to get the desired result.

Next, we would like to check if convergence with respect to $d(X, Y)$ is equivalent to convergence in probability (a distance metric with this property is sometimes called a Ky-Fan metric).

b) Let $(X_n, n \ge 1)$ be sequence of random variables and X be another random variable, all defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Show that if $X_n \underset{n\to\infty}{\overset{\mathbb{P}}{\to}} X$ then $\lim_{n\to\infty} d(X_n, X) = 0$.

Solution: Fix $\epsilon > 0$ and note that convergence in probability implies that

$$\lim_{n\to\infty} \mathbb{P}(\{|X_n - X| \ge \epsilon\}) = 0.$$

For simplicity, define $g(x,y) = \log_2\left(1+\frac{|x-y|}{1+|x-y|}\right)$. We can write

$$\begin{aligned}
d(X_n, X) &= \mathbb{E}\left(g(X_n, X)1_{|X_n-X|\ge\epsilon|}\right) + \mathbb{E}\left(g(X_n, X)1_{|X_n-X|<\epsilon}\right) \\
&\le \mathbb{E}\left(1_{|X_n-X|\ge\epsilon|}\right) + \log_2\left(1+\frac{\epsilon}{1+\epsilon}\right) \\
&= \mathbb{P}\left(\{|X_n-X|\ge\epsilon\}\right) + \log_2\left(1+\frac{\epsilon}{1+\epsilon}\right)
\end{aligned}$$

Therefore

$$\lim_{n\to\infty} d(X_n, X) \le \log_2\left(1+\frac{\epsilon}{1+\epsilon}\right).$$

Since this is true for any ϵ, we can further take a limit as ϵ goes to zero to get the desired result.

c) Is the converse true? That is, if $\lim_{n\to\infty} d(X_n, X) = 0$ then $X_n \underset{n\to\infty}{\overset{\mathbb{P}}{\to}} X$. If yes, prove the statement. If no, provide a counter example.

Solution: Yes, the converse is also true. Fix $\epsilon > 0$ and define $\nu = \log_2\left(1+\frac{\epsilon}{1+\epsilon}\right)$. Then

$$
\begin{aligned}
\mathbb{P}\left(\{|X_n - X| \geq \epsilon\}\right) &= \nu \cdot \frac{1}{\nu} \mathbb{E}\left(1_{|X_n - X| \geq \epsilon|}\right) \\
&\leq \frac{1}{\nu} \mathbb{E}\left(g(X_n, X) 1_{|X_n - X| \geq \epsilon}\right) \\
&\leq \frac{1}{\nu} d(X_n, X).
\end{aligned}
$$

Since for a fixed ϵ, ν is just a constant, we have that

$$
\lim_{n \to \infty} \mathbb{P}\left(\{|X_n - X| \geq \epsilon\}\right) = \frac{1}{\nu} \lim_{n \to \infty} d(X_n, X) = 0.
$$

Exercise 3. (25 points)

Recall that the moment-generating function of a random variable X is defined for every $t \in \mathbb{R}$ as

$$
M_X(t) = \mathbb{E}\left(e^{tX}\right).
$$

a) Show that if $X \sim \mathcal{N}(0, \sigma^2)$, then

$$
M_X(t) = \exp\left(\frac{1}{2} t^2 \sigma^2\right).
$$

Solution: For $X \sim \mathcal{N}(0, \sigma^2)$ we have

$$
\begin{aligned}
M_X(t) = \mathbb{E}(e^{tX}) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} e^{tx} e^{-\frac{x^2}{2\sigma^2}}\, dx \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{t^2\sigma^2}{2}} \int_{-\infty}^{+\infty} e^{-\frac{(x-\sigma^2 t)^2}{2\sigma^2}}\, dx \\
&= \exp\left(\frac{t^2\sigma^2}{2}\right).
\end{aligned}
$$

We now introduce the concept of *sub-gaussianity*. A random variable X is called sub-gaussian if, for every $t > 0$,

$$
M_X(t) \leq \exp\left(\frac{1}{2} t^2 \eta^2\right)
$$

for some $\eta \in \mathbb{R}^+$. (Note that η^2 need not be the variance of X!).

b) Show that if $X \sim \mathcal{U}([-a, a])$ for some $a > 0$, then X is sub-gaussian with $\eta = a$.

Hint: Recall that $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$.

Solution: For $X \sim \mathcal{U}([-a, a])$ we have

$$
M_X(t) = \mathbb{E}(e^{tX}) = \int_{-a}^{a} \frac{1}{2a} e^{tx}\, dx = \frac{1}{2at}(e^{ta} - e^{-ta}).
$$

Now note that, using the Taylor expansion of e^x given in the hint, we can write

$$\begin{aligned}
e^{ta} - e^{-ta} &= \sum_{n=0}^{\infty} \frac{(ta)^n}{n!} - \sum_{n=0}^{\infty} \frac{(-ta)^n}{n!} \\
&= \sum_{n=0}^{\infty} \frac{(ta)^{2n+1}}{(2n+1)!} \\
&\leq ta \sum_{n=0}^{\infty} \frac{(t^2a^2)^n}{2^n n!} \\
&= ta \exp\left(\frac{t^2a^2}{2}\right)
\end{aligned}$$

where the inequality is due to the fact that $(2n+1)! \geq 2^n n!$, and the last equality is due to the Taylor expansion of $\exp\left(\frac{t^2a^2}{2}\right)$. Hence, we conclude that

$$M_X(t) \leq \frac{1}{2} \exp\left(\frac{t^2a^2}{2}\right) \leq \exp\left(\frac{t^2a^2}{2}\right).$$

c) Show that if X is sub-gaussian for some $\eta \in \mathbb{R}^+$, then for every $t > 0$,

$$\mathbb{P}(|X| \geq t) \leq 2 \exp\left(-\frac{t^2}{2\eta^2}\right).$$

Solution: By the Chebyshev-Markov inequality with $\psi(x) = e^{sx}$, we have

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(e^{sX})}{e^{st}} \leq \exp\left(\frac{s^2\eta^2}{2} - st\right).$$

The optimal s (which can be found by taking the derivative of the right-hand side and putting it equal to 0) is $s = \frac{t}{\eta^2}$, which we can substitute into the equation to get

$$\mathbb{P}(X \geq t) \leq \exp\left(\frac{t^2}{2\eta^2}\right).$$

The same upper-bound can be obtained similarly for $\mathbb{P}(X \leq -t)$, proving the result.

d) Prove the following generalization of Hoeffding's inequality. Let $X_i, i \in \{1, 2, \ldots, n\}$ be independent random variables, where for each i, $X_i - \mathbb{E}(X_i)$ is sub-gaussian for some $\eta_i \in \mathbb{R}^+$. Let also $S_n = \sum_{i=1}^n X_i$. Show that for every $t > 0$,

$$\mathbb{P}(|S_n - \mathbb{E}(S_n)| \geq t) \leq 2 \exp\left(-\frac{t^2}{2\sum_{i=1}^n \eta_i^2}\right).$$

Solution: Note that, if Y_1 and Y_2 are two independent sub-gaussian random variables for some η_1 and η_2, then $Y_1 + Y_2$ is sub-gaussian with $\eta^2 = \eta_1^2 + \eta_2^2$. In fact,

$$M_{Y_1+Y_2}(t) = \mathbb{E}(e^{t(Y_1+Y_2)}) = \mathbb{E}(e^{tY_1})\mathbb{E}(e^{tY_2}) \leq \exp\left(\frac{t^2(\eta_1^2 + \eta_2^2)}{2}\right).$$

One can apply this result recursively to prove the same property for the sum of n independent random variables. Then, the required result follows directly from part 3 with $X = \sum_{i=1}^{n}(X_i - \mathbb{E}(X_i))$.

e) Let $X_i, i \in \{1, 2, \ldots, n\}$ be sub-gaussian random variables with the same $\eta \in \mathbb{R}^+$. Show that

$$\mathbb{E}\left(\max_i X_i\right) \leq \eta\sqrt{2\ln n}.$$

Hint: Start by rewriting $\mathbb{E}\left(\max_i X_i\right) = \frac{1}{t}\mathbb{E}\left(\ln\exp\left(t\max_i X_i\right)\right)$.

Solution: Using the hint, we have

$$\begin{aligned}
\mathbb{E}\left(\max_i X_i\right) &= \frac{1}{t}\mathbb{E}\left(\ln\exp\left(t\max_i X_i\right)\right) \\
&\leq \frac{1}{t}\ln\mathbb{E}\left(\exp\left(t\max_i X_i\right)\right) \\
&= \frac{1}{t}\ln\mathbb{E}\left(\max_i \exp\left(tX_i\right)\right) \\
&\leq \frac{1}{t}\ln\mathbb{E}\left(\sum_{i=1}^{n}\exp\left(tX_i\right)\right) \\
&= \frac{1}{t}\ln\left(\sum_{i=1}^{n}\mathbb{E}(\exp(tX_i))\right) \\
&\leq \frac{\ln n}{t} + \frac{\eta^2 t}{2}
\end{aligned}$$

where the first inequality follows from Jensen's inequality, and the last one is due to the fact that the n random variables are sub-gaussian with the same η. The optimal t (obtained once again by putting the derivative equal to 0) is $t = \frac{\sqrt{2\ln(n)}}{\eta}$. Substituing this value into the last equation gives

$$\mathbb{E}\left(\max_i X_i\right) \leq 2\eta\sqrt{\frac{\ln n}{2}} = \eta\sqrt{2\ln n}.$$

Exercise 4. (25 points)

a) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\{\mathcal{F}_n, n \in \mathbb{N}\}$ be a filtration on this space. Let $A \in \mathcal{F}$ and define $Y_n = \mathbb{E}\left(1_A | \mathcal{F}_n\right)$. Show that $(Y_n, n \in \mathbb{N})$ is a martingale with respect to the filtration $\{\mathcal{F}_n, n \in \mathbb{N}\}$.

Solution: $(Y_n, n \in \mathbb{N})$ is a special case of the Doob's martingale studied in class. The three properties could be immediately checked:

- $0 \leq Y_n \leq 1$ for all n, so Y_n is bounded, and therefore integrable for all n
- Y_n if $\mathcal{F}_n$- measurable by definition of conditional expectation
- $\mathbb{E}\left(Y_{n+1}|\mathcal{F}_n\right) = \mathbb{E}\left(\mathbb{E}\left(1_A|\mathcal{F}_{n+1}\right)|\mathcal{F}_n\right) = \mathbb{E}\left(1_A|\mathcal{F}_n\right) = Y_n$ where the second to last equality is the towering property of conditional expectation.

b) Is it true that

$$Y_n \to Y_\infty, \text{ a.s.}$$

for some random variable Y_∞? Why or why not? Could we say something about convergence in distribution to Y_∞?

Solution: Yes, $(Y_n, n \in \mathbb{N})$ is a bounded martingale. Therefore it satisfies the conditions of the martingale convergence theorem (v1) and converges almost surely to some Y_∞. Convergence almost surely implies convergence in distribution. So, this martingale also converges in distribution.

Next, we will use this martingale to prove Kolmogorov's zero-one law. Let $X_0, X_1, \ldots$ be independent random variables. Recall that the tail σ-field is

$$\mathcal{T} = \bigcap_{n=0}^{\infty} \mathcal{H}_n$$

where $\mathcal{H}_n = \sigma(X_n, X_{n+1}, \ldots)$ and assume $A \in \mathcal{T}$. Our goal will be to prove that $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$.

c) Let $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ and $\mathcal{F}_\infty$ be the smallest σ-field that contains every $\mathcal{F}_n$. A standard measure-theoretic argument could be used to show that $Y_\infty = \mathbb{E}\left(1_A | \mathcal{F}_\infty\right)$, but we will take it as a fact here.

Assume $Y_\infty = \mathbb{E}\left(1_A | \mathcal{F}_\infty\right)$. Show, furthermore, that for all $A \in \mathcal{T}$,

$$Y_\infty := \mathbb{E}\left(1_A | \mathcal{F}_\infty\right) = 1_A.$$

Solution: Since $A \in \mathcal{T}$ we have that

$$A \in \mathcal{H}_0 = \sigma(X_0, X_1, \ldots) = \bigcup_{n=0}^{\infty} \sigma(X_0, \ldots, X_n) = \bigcup_{n=0}^{\infty} \mathcal{F}_n \subset \mathcal{F}_\infty$$

Then

$$\mathbb{E}\left(1_A | \mathcal{F}_\infty\right) = 1_A.$$

by definition of conditional expectation and the fact that 1_A is $\mathcal{F}_\infty$-measurable.

d) Show that

$$Y_n := \mathbb{E}\left(1_A | \mathcal{F}_n\right) = \mathbb{P}(A).$$

Hint: How are the σ-fields $\mathcal{T}$ and $\mathcal{F}_n$ related to each other?

Solution: Recall from class that the σ-fields $\mathcal{T}$ and $\mathcal{F}_n$ are independent. This is because $\mathcal{H}_{n+1}$ and $\mathcal{F}_n$ are independent, and $\mathcal{T} \subset \mathcal{H}_{n+1}$. Then

$$\mathbb{E}\left(1_A | \mathcal{F}_n\right) = \mathbb{E}\left(1_A\right) = \mathbb{P}(A).$$

e) Combine the ingredients above to prove Kolmogorov's zero-one law.

Solution: By parts (a) and (b) we know that $(Y_n, n \in \mathbb{N})$ is a martingale that converges almost surely to Y_∞. By part (d) we know that $Y_n = \mathbb{P}(A)$ is a constant sequence of random variables. By part (c) we know that it converges to 1_A which can only take values zero or one. Therefore, there are two options. Either $1_A = 0$ a.s. or $1_A = 1$ a.s.. and, likewise, $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$.