Final exam: solutions

Exercise 1. Quiz. (26 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) (3 points) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $X : \Omega \to \mathbb{R}$ be a random variable with cumulative distribution function $F_X(x)$. Suppose $g : \mathbb{R} \to \mathbb{R}$ is a strictly decreasing and continuous function. Define $Y = g(X)$. What is the cumulative distribution function of Y?

Solution: For a strictly decreasing and continuous g, $Y \leq y \iff X \geq g^{-1}(y)$ Thus:

$$F_Y(y) = P(Y \leq y) = P(X \geq g^{-1}(y)) = 1 - F_X(g^{-1}(y)).$$

b) (3 points) Let X_1 and X_2 be two Gaussian random variables such that the random vector $X = (X_1, X_2)$ has a covariance matrix

$$\mathrm{Cov}(X) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Does this imply that X_1 and X_2 are independent?

Answer: No. Consider the following counter example. Let $X_1 \sim \mathcal{N}(0, 1)$ and Z be equiprobable on $\{-1, 1\}$ and independent of X_1. Let $X_2 = Z \cdot X_1$. Then

$$\mathrm{Cov}(X_1, X_2) = \mathbb{E}(X_1 X_2) - \mathbb{E}(X_1)\mathbb{E}(X_2) = \mathbb{E}(X_1 X_2) = \mathbb{E}(X_1 Z X_1) = \mathbb{E}(Z)\mathbb{E}(X_1^2) = 0.$$

Here we have that the random vector X is not a Gaussian random vector and so zero covariance does not imply independence.

c) (4 points) Let X, Y be integrable random variables on the measurable space $(\Omega, \mathcal{F}, P)$. Define $\mathcal{G} = \sigma(Y)$. Suppose that $\mathbb{E}[XY|\mathcal{G}] = aY^2 + bY$, where a and b are constants. Compute $\mathbb{E}[X]$ in terms of a, b, and $\mathbb{E}[Y]$.

Solution: We know that:

$$\mathbb{E}[XY|\mathcal{G}] = \mathbb{E}[X|\mathcal{G}]Y \text{ a.s. and} \quad (1)$$
$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|G]]. \quad (2)$$

Thus, using the linearity of expectation, we have that

$$\begin{aligned} \mathbb{E}[X] &= \mathbb{E}[\mathbb{E}[X|G]] \\ &= \mathbb{E}\left[\frac{\mathbb{E}[XY|G]}{Y}\right] \\ &= \mathbb{E}\left[\frac{aY^2 + bY}{Y}\right] \\ &= \mathbb{E}[aY + b] \\ &= a\mathbb{E}[Y] + b. \end{aligned}$$

d) (4 points) Let X, Y be i.i.d random variables on the measurable space $(\Omega, \mathcal{F}, \mathbb{P})$. Further, also suppose $X + Y$ and $X - Y$ are independent random variables. Is it true that $\phi_X(2t) = (\phi_X(t))^2|\phi_X(t)|^2$ for all $t \in \mathbb{R}$? Why or Why not?

Answer: Yes.
Since, X and Y are independent. We have

$$\begin{aligned}\phi_{(X+Y,X-Y)}(t_1, t_2) &= \mathbb{E}\left(e^{it_1(X+Y)+it_2(X-Y)}\right)\\ &= \mathbb{E}\left(e^{i(t_1+t_2)X+i(t_1-t_2)Y}\right)\\ &= \phi_X(t_1 + t_2)\phi_Y(t_1 - t_2).\end{aligned}$$

But, also recall that, we are given that $X + Y$ and $X - Y$ are independent as well. Thus,

$$\begin{aligned}\phi_{(X+Y,X-Y)}(t_1, t_2) &= \mathbb{E}\left(e^{it_1(X+Y)+it_2(X-Y)}\right)\\ &= \phi_{X+Y}(t_1)\phi_{X-Y}(t_2).\end{aligned}$$

Thus, we have that

$$\phi_X(t_1 + t_2)\phi_Y(t_1 - t_2) = \phi_{X+Y}(t_1)\phi_{X-Y}(t_2)$$

Substitute $t_1 = t_2 = t$,

$$\phi_X(2t) = (\phi_X(t))^3\phi_Y(-t)$$

again since X and Y are identical, we have $\phi_X(t) = \phi_Y(t)$ for all $t \in \mathbb{R}$. Thus,

$$\begin{aligned}\phi_X(2t) &= (\phi_X(t))^3\phi_X(-t)\\ &= (\phi_X(t))^2|\phi_X(t)|^2.\end{aligned}$$

e) (4 points) Let $M_0 = 0.4$ and define recursively

$$M_{n+1} = \begin{cases} M_n^2 & \text{with probability}\frac{1}{2} \\ 2M_n - M_n^2 & \text{with probability}\frac{1}{2}. \end{cases}$$

Is it true that the resulting martingale $(M_n, n \in \mathbb{N})$ converges almost surely to some random variable M_∞? Why or why not?

Answer: Yes. This is a special case of the example covered in lecture. You can show that the martingale is always bounded between zero and one, e.g. $M_n \in (0, 1)$. By MCT v1 it will converge to some M_∞. (In fact, here, M_∞ is distributed as Benouilli(0.6).)

A sequence of biased coins is flipped. The chance that the rth coin shows head is Θ_r, where Θ_r is a random variable taking values in $(0, 1)$. Let X_n be the number of heads after n flips.

f) (4 points) Does X_n obey the central limit theorem when Θ_r are independent and identically distributed?

Answer: Yes. Let Y_n be one if at time n the coin flip is heads. Then $X_n = \sum_{i=1}^n Y_n$ (i.e. X_n takes the role of S_n in our statement of the CLT in the notes). In this setting, we can show that all Y_n are iid and the probability that each coin shows heads is some constant p which depends on the distribution of Θ_r. Specifically,

$$\mathbb{P}(\{Y_n = 1\}) = \mathbb{E}(1_{\{Y_n=1\}}) = \mathbb{E}\left(\mathbb{E}\left(1_{\{Y_n=1\}}|\Theta_n\right)\right) = \mathbb{E}(\Theta_n)$$

But, since all the Θ_n are iid, the probability will be the same for all n.

g) (4 points) Does X_n obey the central limit theorem when $\Theta_r = \Theta$ for all r, where Θ is a random variable on $(0, 1)$?

Answer: No. Unless Θ is a constant random variable, the limiting distribution may not be a Gaussian. First, to get the intuition, think of Θ which is supported on $\{0, 1\}$. Then, for all $\omega \in \Omega$ the outcome of a sequence of $(Y_n, n \in \mathbb{N})$ will either be all zeros, or all ones. Likewise, $X_n = 0$ for all n or $X_n = n$ for all n. This clearly does not obey the CLT. In this problem, Θ is a random variable on $(0, 1)$, but the same logic still applies. Suppose it is supported on $\{\epsilon, 1 - \epsilon\}$. Then, for all $\omega \in \Omega$ the sequence $(Y_n, n \in \mathbb{N})$ will be either mostly made up of zeros, or mostly made up of ones and thus will not obey the CLT. (In this case, the limiting distribution will be actually be a mixture of Guassians with expectations at ϵ and $1 - \epsilon$).

Exercise 2. (26 points)

Let X and Y be random variables on common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The total variation distance is defined as

$$d_{TV}(X, Y) = \sup_{A \in \mathcal{B}(\mathbb{R})} |(\mathbb{P}(X \in A) - \mathbb{P}(Y \in A))| .$$

a) (5 points) Show that

$$\mathbb{P}(X = Y) \leq 1 - d_{TV}(X, Y).$$

Solution: Let $A \in \mathcal{F}$ be the set that achieves the supremum.

$$\begin{aligned}\mathbb{P}(\{X \neq Y\}) &\geq \mathbb{P}(\{X \in A, Y \notin A\}) = \mathbb{P}(\{X \in A\}) - \mathbb{P}(\{X \in A, Y \in A\}) \\ &\geq \mathbb{P}(\{X \in A\}) - \mathbb{P}(\{Y \in A\}) = d_{TV}(X, Y)\end{aligned}$$

This shows the desired result.

(Note, if no A that achieves the above supremum exists, we can take a set A that is arbitrarily close to achieving the supremum, and the same argument will hold with minor modification.)

b) (5 points) Let X and Y be discrete and supported on $\{0, 1, 2, \dots\}$ with pmfs p_X and p_Y. Show that

$$d_{TV}(X, Y) = \frac{1}{2} \sum_{m \geq 0} |p_X(m) - p_Y(m)|.$$

Solution: First, note that if $A \in \mathcal{F}$ achieves the supremum above, so does $\bar{A} \in \mathcal{F}$. Let $A = \{m \colon \mathbb{P}(\{X = m\}) > \mathbb{P}(\{Y = m\})\}$. We show that A achieves the supremum in the definition above. Take any other set $B \in \mathcal{F}$ and assume without loss of generality that $\mathbb{P}(\{X \in B\}) \geq \mathbb{P}(\{Y \in B\})$.

Then

$$\begin{aligned}
&\mathbb{P}(\{X \in B\}) - \mathbb{P}(\{Y \in B\}) \\
&= \mathbb{P}(\{X \in A \cap B\}) + \mathbb{P}(\{X \in \bar{A} \cap B\}) - (\mathbb{P}(\{Y \in A \cap B\}) + \mathbb{P}(\{Y \in \bar{A} \cap B\})) \\
&= \mathbb{P}(\{X \in A \cap B\}) - \mathbb{P}(\{Y \in A \cap B\}) + \mathbb{P}(\{X \in \bar{A} \cap B\}) - \mathbb{P}(\{Y \in \bar{A} \cap B\})) \\
&\leq \mathbb{P}(\{X \in A \cap B\}) - \mathbb{P}(\{Y \in A \cap B\}) \\
&\leq \mathbb{P}(\{X \in A\}) - \mathbb{P}(\{Y \in A\})
\end{aligned}$$

Finally,

$$\begin{aligned}
\frac{1}{2}\sum_{m \geq 0} |p_X(m) - p_Y(m)| &= \frac{1}{2}\sum_{m \in A} (p_X(m) - p_Y(m)) + \frac{1}{2}\sum_{m \in \bar{A}} (p_Y(m) - p_X(m)) \\
&= \frac{1}{2} d_{TV}(X, Y) + \frac{1}{2} d_{TV}(X, Y) = d_{TV}(X, Y)
\end{aligned}$$

c) (4 points) (Optimal coupling) Let $\Omega = \{1, \ldots k\}$, $\mathcal{F} = \mathcal{P}(\Omega)$, and $\mathbb{P}(\omega) = \frac{1}{k} \quad \forall \omega \in \Omega$. Suppose that X is Bernoulli$\left(\frac{k-1}{k}\right)$ and Y is Bernoulli$\left(\frac{2}{k}\right)$. What is $d_{TV}(X, Y)$?

Construct an explicit mappings $X : \Omega \to \{0, 1\}$ and $Y : \Omega \to \{0, 1\}$ so that the bound in part a) is satisfied with equality.

Solution: By using the formula in part b) we see that

$$d_{TV}(X, Y) = \frac{1}{2}\left(\left|\frac{k-1}{k} - \frac{2}{k}\right| + \left|\frac{1}{k} - \frac{k-2}{k}\right|\right) = \frac{k-3}{k}$$

One possible mapping that achieves this is

$$X(\omega) = \begin{cases} 1 & \omega \in \{1, 2, \ldots, k-1\} \\ -1 & \text{otherwise} \end{cases}$$

and

$$Y(\omega) = \begin{cases} 1 & \omega \in \{1, 2\} \\ -1 & \text{otherwise} \end{cases}$$

In this case $X(\omega) = Y(\omega)$ for $\omega \in \{1, 2, k\}$.

d) (4 points) Let $\mathbb{P}(\{X = 1\}) = \mathbb{P}(\{X = -1\}) = \frac{1}{2}$ and $Y \sim \mathcal{N}(0, 1)$. What is $d_{TV}(X, Y)$?

Solution: Letting $A = \{-1, 1\}$ we see that $d_{TV}(X, Y) = 1$, which is the maximum value possible. In other words, X and Y have essentially complementary supports (and this would be true for any pair of discrete and continuous random variables).

e) (4 points) Let $(X_n, n \geq 1)$ be a sequence of random variables and X be another random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. Show that if $\lim_{n\to\infty} d_{TV}(X_n, X) = 0$ then $X_n \underset{n\to\infty}{\overset{d}{\to}} X$.

Solution: For any t for which $F_Y(t)$ is continuous, we have

$$|F_{X_n}(t) - F_Y(t)| = |\mathbb{P}(\{X_n \leq t\}) - \mathbb{P}(\{Y \leq t\})| \leq \sup_{A \in \mathcal{B}(\mathbb{R})} |(\mathbb{P}(X_n \in A) - \mathbb{P}(Y \in A))| = d_{TV}(X_n, Y)$$

Thus, we see that $|F_{X_n}(t) - F_Y(t)| \to 0$, and thus X_n converges in distribution to Y.

f) (4 points) Is the converse true? That is, if $X_n \underset{n\to\infty}{\overset{d}{\to}} X$ then $\lim_{n\to\infty} d_{TV}(X_n, X) = 0$. If yes, prove the statement. If no, provide a counter example.

Solution: No, the converse is not true. Consider Z_n iid with $\mathbb{P}(\{Z_1 = 1\}) = \mathbb{P}(\{Z_1 = -1\}) = \frac{1}{2}$ and define $S_n = Z_1 + Z_2 + \cdots + Z_n$. The random variable $X_n = \frac{S_n}{\sqrt{n}}$ converges to $X \sim \mathcal{N}(0,1)$ by CLT. However, X_n is a discrete random variable for any finite n. By an argument analogous to part d) we see that $d_{TV}(X_n, X) = 1$ for all n.

Exercise 3. (22 points)

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $X : \Omega \to \mathbb{R}$ be an $\mathcal{F}$-measurable random variable on this space. For this problem, you may wish to recall Hölder's inequality from the mid-term exam:

$$\mathbb{E}(|XY|) \leq (\mathbb{E}(|X|^p))^{1/p}(\mathbb{E}(|Y|^q))^{1/q}$$

for $p, q \geq 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$.

a) (5 points) (Lyapunov's inequality) Show that for every $1 \leq r < s < \infty$ we have

$$\mathbb{E}(|X|^r))^{1/r} \leq \mathbb{E}(|X|^s))^{1/s}.$$

Solution: This inequality can be proved using either the Jensen's inequality or the Holder's inequality. Let's see both the proofs.

1. **Proof using Jensen's:**
   Apply Jensen's inequality to the random variable $Y := |X|^s$ with the concave function $f(Y) := Y^{r/s}$ (since $r/s < 1$). Let's have a detailed look:

$$\begin{aligned} \mathbb{E}(f(Y)) &\leq f(\mathbb{E}(Y)) \\ \mathbb{E}(|X|^r) &\leq f(\mathbb{E}(|X|^s)) \\ \mathbb{E}(|X|^r) &\leq (\mathbb{E}(|X|^s))^{r/s} \\ (\mathbb{E}(|X|^r))^{1/r} &\leq (\mathbb{E}(|X|^s))^{1/s} \end{aligned}$$

2. **Proof using Holder's:**
   Apply Holder's inequality with $Y := 1$, $X := |Z|^r$ (where Z be a dummy random variable on $(\Omega, \mathcal{F}, \mathbb{P})$) and $p := s/r$ (which implies $q = s/(s-r)$). Thus, we have that:

$$\begin{aligned} \mathbb{E}(|X|) &\leq (\mathbb{E}(|X|^p))^{1/p} \\ \mathbb{E}(|Z|^r) &\leq (\mathbb{E}(|Z|^{pr}))^{1/p} \\ \mathbb{E}(|Z|^r) &\leq (\mathbb{E}(|Z|^s))^{r/s} \\ (\mathbb{E}(|Z|^r))^{1/r} &\leq (\mathbb{E}(|Z|^s))^{1/s} \end{aligned}$$

Recall that $X_n \underset{n\to\infty}{\overset{L^r}{\to}} X$ if $\mathbb{E}(|X_n^r|) < \infty$ for all n and

$$\mathbb{E}(|X_n - X|^r) \underset{n\to\infty}{\to} 0.$$

b) (5 points) Assume $X_n \underset{n\to\infty}{\overset{L^s}{\to}} X$. Show that $X_n \underset{n\to\infty}{\overset{L^r}{\to}} X$ for every $1 \le r < s < \infty$.

Solution: Given $X_n \underset{n\to\infty}{\overset{L^s}{\to}} X$, to show that $X_n \underset{n\to\infty}{\overset{L^r}{\to}} X$, we need to prove two conditions i.e., $\mathbb{E}(|X_n^r|) < \infty$ for all n and $\mathbb{E}(|X_n - X|^r) \underset{n\to\infty}{\to} 0$.
Note that the first condition follows from the Exercise 4 (Part A) from the mid-term exam (see solutions for more details). The second part too, was a part of the graded homework problem from the problem set 9. (See Homework 9 for the full proof). It follows from Jensen's inequality; you could also use the result from part (a), alternatively.

c) (5 points) (Minkowski's inequality) Show that for every $r \ge 1$ and any two $\mathcal{F}$-measurable random variables X and Y such that $\mathbb{E}(|X|^r), \mathbb{E}(|Y|^r) < \infty$, we have

$$(\mathbb{E}(|X+Y|^r))^{1/r} \le (\mathbb{E}(|X|^r))^{1/r} + (\mathbb{E}(|Y|^r))^{1/r}.$$

Hint: Start by writing $|X+Y|^r = |X|X+Y|^{r-1} + Y|X+Y|^{r-1}|$. *Then, apply Hölder's inequality .*

Solution: From the hint, we have the following:

$$|X+Y|^r \le |X||X+Y|^{r-1} + |Y||X+Y|^{r-1}$$

Therefore,

$$\mathbb{E}(|X+Y|^r) \le \mathbb{E}(|X||X+Y|^{r-1}) + \mathbb{E}(|Y||X+Y|^{r-1})$$

Now, on applying Holder's inequality on both the terms on the R.H.S of the previous equation we have that

$$\begin{aligned}
\mathbb{E}(|X+Y|^r) &\le ((\mathbb{E}(|X|^r))^{1/r} + (\mathbb{E}(|Y|^r))^{1/r})(\mathbb{E}(|X+Y|^r))^{(r-1)/r} \\
(\mathbb{E}(|X+Y|^r))^{1-\frac{r-1}{r}} &\le ((\mathbb{E}(|X|^r))^{1/r} + (\mathbb{E}(|Y|^r))^{1/r}) \\
(\mathbb{E}(|X+Y|^r))^{1/r} &\le ((\mathbb{E}(|X|^r))^{1/r} + (\mathbb{E}(|Y|^r))^{1/r})
\end{aligned}$$

This proves Minkowski's inequality.

d) (7 points) Assume that $X_n \underset{n\to\infty}{\overset{L^r}{\to}} X$ and $Y_n \underset{n\to\infty}{\overset{L^r}{\to}} Y$ for some $r \ge 1$. Show that $X_n + Y_n \underset{n\to\infty}{\overset{L^r}{\to}} X + Y$.

Solution: We have that

$$\lim_{n\to\infty} (\mathbb{E}(|X_n - X|)^r) = 0 \text{ and } \lim_{n\to\infty} (\mathbb{E}(|Y_n - Y|)^r) = 0$$

and so

$$\lim_{n\to\infty} (\mathbb{E}(|X_n - X|)^r)^{1/r} = 0 \text{ and } \lim_{n\to\infty} (\mathbb{E}(|Y_n - Y|)^r)^{1/r} = 0$$

since $f(x) = x^{1/r}$ is a continuous function in the neighborhood of zero. Then, by Minkowski's inequality

$$\begin{aligned}
\mathbb{E}(|(X_n + Y_n) - (X+Y)|)^r)^{1/r} &= \mathbb{E}(|(X_n - X) + (Y_n - Y)|^r)^{1/r} \\
&\le \mathbb{E}(|X_n - X|^r)^{1/r} + \mathbb{E}(|Y_n - Y|^r)^{1/r}.
\end{aligned}$$

So

$$\lim_{n\to\infty} (\mathbb{E}(|(X_n + Y_n) - (X + Y|)^r)^{1/r} = 0$$

which implies that

$$\lim_{n\to\infty} \mathbb{E}(|(X_n + Y_n) - (X + Y|)^r = 0,$$

and this proves the result.

Exercise 4. (26 points)

Let $\{X_n, n \geq 1\}$ be an i.i.d. sequence such that $\mathbb{E}(|X_1|) < \infty$. Let $\{\mathcal{F}_n, n \in \mathbb{N}\}$ be the natural filtration and τ be a stopping time with respect to this filtration.

a) (Wald's) Given $\mathbb{E}(\tau) < \infty$, show that

$$\mathbb{E}\left(\sum_{n=1}^{\tau} X_n\right) = \mathbb{E}(\tau)\mathbb{E}(X_1).$$

Solution: We start by noting

$$\sum_{n=1}^{\tau} X_n = \sum_{n=1}^{\infty} X_n \mathbb{I}\{\tau > n-1\},$$

where $\mathbb{I}\{\tau > n-1\}$ is the indicator random variable. This holds since $\mathbb{I}\{\tau > n-1\} \in \mathcal{F}_{n-1}$.

We take the sum out of the expectation, use the independence of $\mathbb{I}\{\tau > n-1\}$ and X_n, as well as the fact that $\mathbb{E}(X_n) = \mathbb{E}(X_1)$, $\forall n \in \mathbb{N}$:

$$\begin{aligned}
\mathbb{E}\left(\sum_{n=1}^{\tau} X_n\right) &= \mathbb{E}\left(\sum_{n=1}^{\infty} X_n \mathbb{I}\{\tau > n-1\}\right) \\
&= \sum_{n=1}^{\infty} \mathbb{E}\left(X_n \mathbb{I}\{\tau > n-1\}\right) \\
&= \sum_{n=1}^{\infty} \mathbb{E}\left(X_n\right) \mathbb{P}(\{\tau > n-1\}) \\
&= \mathbb{E}(X_1) \sum_{n=1}^{\infty} \mathbb{P}(\{\tau > n-1\}) \\
&= \mathbb{E}(X_1) \sum_{n=0}^{\infty} \mathbb{P}(\{\tau > n\}) \\
&= \mathbb{E}(X_1)\mathbb{E}(\tau).
\end{aligned}$$

For the remaining parts of the problem, assume that $\{X_n\}$ is uniformly distributed over the discrete alphabet $\{1, 2, \ldots, k\}$.

b) Let $S_0 = 0$ and $S_n = \sum_{i=1}^{n} X_i$ for $n \geq 1$. Is the process $\{S_n, n \in \mathbb{N}\}$ a martingale with respect to $\{\mathcal{F}_n, n \in \mathbb{N}\}$? Why?

If it is a martingale, set $M_n = S_n$ for all $n \in \mathbb{N}$. If not, find the unique predictable and increasing process $\{A_n, n \in \mathbb{N}\}$ such that the process $\{M_n = S_n - A_n, n \in \mathbb{N}\}$ is a martingale with respect to $\{\mathcal{F}_n, n \in \mathbb{N}\}$.

Solution: We check the conditions:

$$\begin{aligned}
\mathbb{E}(|S_n|) &\leq nk < +\infty, \forall n \in \mathbb{N} \\
\mathbb{E}(S_{n+1}|\mathcal{F}_n) &= \mathbb{E}(S_n + X_{n+1}|\mathcal{F}_n) \\
&= S_n + \mathbb{E}(X_{n+1}|\mathcal{F}_n) \\
&= S_n + \mathbb{E}(X_{n+1}) \\
&= S_n + \frac{k+1}{2} \neq S_n
\end{aligned}$$

Since one of the conditions does not hold, S_n is not a martingale.

Using Doob's decomposition theorem, we set $A_0 = 0$ and $A_{n+1} = A_n + \mathbb{E}(S_{n+1}|\mathcal{F}_n) - S_n$. We can further calculate:

$$\begin{aligned}
A_{n+1} &= A_n + \mathbb{E}(X_{n+1}) \\
&= A_n + \frac{k+1}{2} \\
&= A_{n-1} + 2\frac{k+1}{2} \\
&= (n+1)\frac{k+1}{2}.
\end{aligned}$$

Then, $\{M_n = S_n - A_n, n \in \mathbb{N}\}$ is a martingale.

c) Let $\tau_a = \min\{n \geq 1 : X_n = a\}$ for a fixed $a \in \{1, 2, \ldots, k\}$. Is τ_a a stopping time? Explain why.

Solution: Yes, since $\{\tau_a = n\} = \{X_i \neq a, \forall i \in \{1, 2, \ldots, n-1\} \text{ and } X_n = a\} \in \mathcal{F}_n, \forall n \in \mathbb{N}$.

d) Find $\mathbb{E}(\tau_a)$. Does it depend on a?

Solution: No, it does not depend on a. We first calculate $\mathbb{P}(\{\tau_a + n\})$:

$$\begin{aligned}
\mathbb{P}(\{\tau_a + n\}) &= \mathbb{P}(\{X_i \neq a, \forall i \in \{1, 2, \ldots, n-1\} \text{ and } X_n = a\}) \\
&= \mathbb{P}(\{X_1 \neq a\})\mathbb{P}(\{X_2 \neq a\}) \ldots \mathbb{P}(\{X_{n-1} \neq a\})\mathbb{P}(\{X_n = a\}) && \text{(independence)} \\
&= \left(\frac{k-1}{k}\right)^{n-1} \frac{1}{k}. && \text{(uniformity)}
\end{aligned}$$

Then, using the formula $\sum\limits_{n=1}^{\infty} na^n = a/(1-a)^2$, $\forall a < 1$:

$$\begin{aligned}
\mathbb{E}(\tau_a) &= \sum_{n=1}^{\infty} n\mathbb{P}(\{\tau_a + n\}) \\
&= \frac{1}{k-1}\sum_{n=1}^{\infty} n\left(\frac{k-1}{k}\right)^n \\
&= \frac{1}{k-1}\frac{\frac{k-1}{k}}{(1-\frac{k-1}{k})^2} \\
&= k.
\end{aligned}$$

e) Using the previous parts, calculate $\mathbb{E}(S_{\tau_a})$ and $\mathbb{E}(M_{\tau_a})$.

Solution: Since $\mathbb{E}(\tau_a) = k < \infty$, we can use the result in part a).

$$\begin{aligned}
\mathbb{E}(S_{\tau_a}) &= \mathbb{E}\left(\sum_{n=1}^{\tau} X_n\right) \\
&= \mathbb{E}(\tau_a)\mathbb{E}(X_1) \\
&= k\frac{k+1}{2}
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}(M_{\tau_a}) &= \mathbb{E}(S_{\tau_a} - A_{\tau_a}) \\
&= \mathbb{E}\left(\sum_{n=1}^{\tau} X_n - \sum_{n=1}^{\tau}\mathbb{E}(X_n)\right) \\
&= \mathbb{E}(\tau_a)\mathbb{E}(X_1 - \mathbb{E}(X_1)) \\
&= 0.
\end{aligned}$$

f) Let $\tau_b = \inf\{n \geq 1 : |M_n| \geq b\}$ for a fixed $b \in \{k, k+1, \ldots, 3k\}$. Is τ_b a stopping time? Calculate $\mathbb{E}(M_{\tau_b})$.

Solution: Yes, τ_b is a stooping time since $\{\tau_b = n\} \in \mathcal{F}_n$. Here, the optional stopping theorem, version 3 holds since $|M_{n+1} - M_n| \leq k$ and the stopped martingale with respect to τ_b is bounded. Thus,

$$\mathbb{E}(M_{\tau_b}) = \mathbb{E}(M_0) = 0.$$