II Università di Roma, Tor Vergata
Dipartimento d'Ingegneria Civile e Ingegneria Informatica
LM in Ingegneria dell'Informazione e dell'Automazione
Complementi di Probabilità e Statistica
Homework - 2019-11-16

Problem 1 *Let $(\Omega, \mathcal{E}, \mathbf{P}) \equiv \Omega$ a probability space, let X and Y be independent standard Rademacher random variables*[1] *on Ω. Set $Z \stackrel{def}{=} X + Y$.*

1. *Compute $\mathbf{E}[X \mid Z]$ and $\mathbf{E}[Y \mid Z]$.*
2. *Are the random variables $\mathbf{E}[X \mid Z]$ and $\mathbf{E}[Y \mid Z]$ uncorrelated?*
3. *Are the random variables $\mathbf{E}[X \mid Z]$ and $\mathbf{E}[Y \mid Z]$ independent?*
4. *By using the properties of the conditional expectation, on account that you are dealing with standard Rademacher random variables, can you compute $\mathbf{E}\left[(X+Y)^2 \mid Z\right]$ and $\mathbf{E}[XY \mid Z]$?*

Solution. Since X and Y be independent standard Rademacher random variables, we have

$$(X+Y)(\omega) = \begin{cases} -2, & \text{if } \omega \in \{X=-1, Y=-1\}, \\ 0, & \text{if } \omega \in \{X=-1, Y=1\} \cup \{X=1, Y=-1\}, \\ 2, & \text{if } \omega \in \{X=1, Y=1\}. \end{cases}$$

That is to say,

$$X + Y = -2 \cdot 1_{\{X=-1, Y=-1\}} + 2 \cdot 1_{\{X=1, Y=1\}}$$

Furthermore,

$$\mathbf{P}(X+Y=-2) = \mathbf{P}(X=-1, Y=-1) = \mathbf{P}(X=-1)\mathbf{P}(Y=-1) = \frac{1}{4},$$

$$\mathbf{P}(X+Y=2) = \mathbf{P}(X=1, Y=1) = \mathbf{P}(X=1)\mathbf{P}(Y=1) = \frac{1}{4},$$

and

$$\begin{aligned} \mathbf{P}(X+Y=0) &= \mathbf{P}(\{X=-1, Y=1\} \cup \{X=1, Y=-1\}) \\ &= \mathbf{P}(X=-1, Y=1) + \mathbf{P}(X=1, Y=-1) \\ &= \mathbf{P}(X=-1)\mathbf{P}(Y=1) + \mathbf{P}(X=1)\mathbf{P}(Y=-1) \\ &= \frac{1}{2}. \end{aligned}$$

1. Since Z is a discrete random variable, to compute $\mathbf{E}[X \mid Z]$ we can apply the formula

$$\mathbf{E}[X \mid Z] = \mathbf{E}[X \mid Z=-2]\, 1_{\{Z=-2\}} + \mathbf{E}[X \mid Z=2]\, 1_{\{Z=2\}},$$

[1] A standard Rademacher random variable R is given by

$$R \stackrel{\text{def}}{=} \begin{cases} 1, & \mathbf{P}(R=1) = 1/2, \\ -1, & \mathbf{P}(R=-1) = 1/2. \end{cases}$$

where

$$\begin{aligned}\mathbf{E}[X \mid Z=-2] &= \frac{1}{\mathbf{P}(Z=-2)}\int_{\{Z=2\}} X d\mathbf{P} = 4\int_{\{X=-1,Y=-1\}} X d\mathbf{P} \\ &= -4\int_{\{X=-1,Y=-1\}} d\mathbf{P} = -4\mathbf{P}(X=-1,Y=-1) = -1.\end{aligned}$$

and

$$\begin{aligned}\mathbf{E}[X \mid Z=2] &= \frac{1}{\mathbf{P}(Z=2)}\int_{\{Z=2\}} X d\mathbf{P} = 4\int_{\{X=1,Y=1\}} X d\mathbf{P} \\ &= 4\int_{\{X=1,Y=1\}} d\mathbf{P} = 4\mathbf{P}(X=1,Y=1) = 1.\end{aligned}$$

It follows

$$\mathbf{E}[X \mid Z] = -1_{\{Z=-2\}} + 1_{\{Z=2\}} = \frac{1}{2}Z.$$

Similarly,

$$\mathbf{E}[Y \mid Z] = -1_{\{Z=-2\}} + 1_{\{Z=2\}} = \frac{1}{2}Z.$$

Another argument, based on the properties of the conditional expectation, is the following. Observe that

$$Z = \mathbf{E}[Z \mid Z] = \mathbf{E}[X+Y \mid Z] = \mathbf{E}[X \mid Z] + \mathbf{E}[Y \mid Z].$$

On the other hand, we know that

$$\mathbf{E}[X \mid Z] = g_X(Z) \quad \text{and} \quad \mathbf{E}[Y \mid Z] = g_Y(Z)$$

where $g_X : \mathbb{R} \to \mathbb{R}$ and $g_X : \mathbb{R} \to \mathbb{R}$ are suitable Borel functions. The structure of the function $g_X : \mathbb{R} \to \mathbb{R}$ [resp. $g_Y : \mathbb{R} \to \mathbb{R}$] depends on the joint distribution of X and Z [resp. Y and Z] and on the distribution of Z. However, in our case, it is not difficult to prove that

$$F_{X,Z}(u,z) = F_{Y,Z}(u,z),$$

for every $(u,z) \in \mathbb{R}^2$. In fact,

$$\begin{aligned}&F_{X,Z}(u,z) \\ &= \mathbf{P}(X \le u, Z \le z) \\ &= \mathbf{P}(X \le u, X+Y \le z) \\ &= \mathbf{P}(X \le u, X+Y \le z, X=1) + \mathbf{P}(X \le u, X+Y \le z, X=-1) \\ &= \mathbf{P}(X \le u, X+Y \le z \mid X=1)\mathbf{P}(X=1) + \mathbf{P}(X \le u, X+Y \le z \mid X=-1)\mathbf{P}(X=-1) \\ &= \frac{1}{2}(\mathbf{P}(X \le u, X+Y \le z \mid X=1) + \mathbf{P}(X \le u, X+Y \le z \mid X=-1)) \\ &= \frac{1}{2}(\mathbf{P}(1 \le u, 1+Y \le z \mid X=1) + \mathbf{P}(-1 \le u, -1+Y \le z \mid X=-1)) \\ &= \frac{1}{2}(\mathbf{P}(1 \le u, Y \le z-1 \mid X=1) + \mathbf{P}(-1 \le u, Y \le z+1 \mid X=-1)) \\ &= \begin{cases} 0, & \text{if } u < -1, \\ \frac{1}{2}\mathbf{P}(Y \le z+1 \mid X=-1) = \frac{1}{2}\mathbf{P}(Y \le z+1), & \text{if } -1 \le u < 1, \\ \frac{1}{2}(\mathbf{P}(Y \le z-1 \mid X=1) + \mathbf{P}(Y \le z+1 \mid X=-1)) = \frac{1}{2}(\mathbf{P}(Y \le z-1) + \mathbf{P}(Y \le z+1)), & \text{if } 1 \le u. \end{cases}\end{aligned}$$

Similarly,

$$\begin{aligned}&F_{Y,Z}(u,z) \\ &= \begin{cases} 0, & \text{if } u < -1, \\ \frac{1}{2}\mathbf{P}(X \le z+1 \mid Y=-1) = \frac{1}{2}\mathbf{P}(X \le z+1), & \text{if } -1 \le u < 1, \\ \frac{1}{2}(\mathbf{P}(X \le z-1 \mid Y=1) + \mathbf{P}(X \le z+1 \mid Y=-1)) = \frac{1}{2}(\mathbf{P}(X \le z-1) + \mathbf{P}(X \le z+1)), & \text{if } 1 \le u. \end{cases}\end{aligned}$$

Therefore, on account that X and Y have the same distribution, we obtain the desired result. As a consequence, we can asses that

$$g_X = g_Y,$$

which implies

$$\mathbf{E}[X \mid Z] = \mathbf{E}[Y \mid Z].$$

It then follows

$$2\mathbf{E}[X \mid Z] = 2\mathbf{E}[Y \mid Z] = Z,$$

which yields

$$\mathbf{E}[X \mid Z] = \mathbf{E}[Y \mid Z] = \frac{1}{2}Z,$$

as expected.

2. Thanks to what shown above, we have

$$\mathbf{E}[X \mid Z]\,\mathbf{E}[Y \mid Z] = \frac{1}{4}Z^2 \sim Dir\left(\frac{1}{4}\right).$$

Hence,

$$\mathbf{E}[\mathbf{E}[X \mid Z]\,\mathbf{E}[Y \mid Z]] = \frac{1}{4}.$$

On the other hand,

$$\mathbf{E}[\mathbf{E}[X \mid Z]] = \mathbf{E}[\mathbf{E}[Y \mid Z]] = \mathbf{E}\left[\frac{1}{2}Z\right] = \frac{1}{2}\mathbf{E}[Z] = 0.$$

Hence, $\mathbf{E}[X \mid Z]$ and $\mathbf{E}[Y \mid Z]$ are not uncorrelated.

3. Since $\mathbf{E}[X \mid Z]$ and $\mathbf{E}[Y \mid Z]$ are not not uncorrelated, they cannot be independent.

4. By virtue of the properties of the conditional expectation, we have

$$\mathbf{E}\left[(X+Y)^2 \mid Z\right] = \mathbf{E}\left[Z^2 \mid Z\right] = Z^2.$$

On the other hand,

$$\begin{aligned}\mathbf{E}\left[(X+Y)^2 \mid Z\right] &= \mathbf{E}\left[X^2 + 2XY + Y^2 \mid Z\right] \\ &= \mathbf{E}\left[X^2 \mid Z\right] + 2\mathbf{E}[XY \mid Z] + \mathbf{E}\left[Y^2 \mid Z\right].\end{aligned}$$

Now, since $X \sim Y \sim Rad\,(1/2)$, we have $X^2 \sim Y^2 \sim Dir\,(1)$. We then obtain

$$Z^2 = \mathbf{E}\left[(X+Y)^2 \mid Z\right] = \mathbf{E}[1 \mid Z] + 2\mathbf{E}[XY \mid Z] + \mathbf{E}[1 \mid Z] = 2 + 2\mathbf{E}[XY \mid Z].$$

The latter yields

$$\mathbf{E}[XY \mid Z] = \frac{1}{2}Z^2 - 1.$$

Problem 2 *Let X [resp. R] be a standard Gaussian [Rademacher] random variable on a probability space Ω. In symbols, $X \sim N(0,1)$ and $R \sim Rad\,(1/2)$. Assume that X and R are independent and define $Y \equiv R \cdot X$.*

1. Is the random variable Y Gaussian?

2. *Are the random variables X and Y uncorrelated? Are X and Y independent?*

3. *Are the random variables R and Y uncorrelated? Are R and Y independent?*

4. *Does the random vector $(X, Y)^\intercal$ have a bivariate Gaussian distribution? Hint: consider the possibility that $(X, Y)^\intercal$ has a bivariate Gaussian distribution; how the random variable $Z \equiv X + Y$ should be distributed?*

5. *Can you compute $\mathbf{E}[Y \mid X]$ and $\mathbf{E}[X \mid Y]$?*

Solution.

1. To prove that Y is Gaussian we show that

$$\mathbf{P}(Y \leq y) = \mathbf{P}(X \leq y), \tag{1}$$

for every $y \in \mathbb{R}$. To this, on account that $\{R = 1\}$, $\{R = -1\}$ constitute a partition of Ω, the random variables R and X are independent and X is symmetric about 0, we can write

$$\begin{aligned}
\mathbf{P}(Y \leq y) &= \mathbf{P}(RX \leq y) \\
&= \mathbf{P}(RX \leq y, R = 1) + \mathbf{P}(RX \leq y, R = -1) \\
&= \mathbf{P}(RX \leq y \mid R = 1)\,\mathbf{P}(R = 1) + \mathbf{P}(RX \leq y \mid R = -1)\,\mathbf{P}(R = -1) \\
&= \frac{1}{2}\left(\mathbf{P}(X \leq y, \mid R = 1) + \mathbf{P}(X \geq -y \mid R = -1)\right) \\
&= \frac{1}{2}\left(\mathbf{P}(X \leq y,) + \mathbf{P}(X \geq -y)\right) \\
&= \mathbf{P}(X \leq y),
\end{aligned}$$

for every $y \in \mathbb{R}$. This proves that $Y \sim X \sim N(0, 1)$.

2. Since $Y \equiv R \cdot X$, the intuition is that the observation of the values taken by X transmits information on the values taken by Y. That is X and Y are not independent. However, on account that $R^2 \sim Dirac(1)$, thanks to the independence of X and R, we have

$$\mathbf{E}[XY] = \mathbf{E}[XRX] = \mathbf{E}\left[R^2 X\right] = \mathbf{E}[X] = 0 = \mathbf{E}[X]\,\mathbf{E}[R].$$

This shows that X and Y are uncorrelated. On the other hand, since $X \sim N(0, 1)$, we have

$$\mathbf{E}\left[X^2 Y^2\right] = \mathbf{E}\left[X^2 R^2 X^2\right] = \mathbf{E}\left[X^4\right] = 3$$

and

$$\mathbf{E}\left[X^2\right]\mathbf{E}\left[Y^2\right] = \mathbf{E}\left[X^2\right]\mathbf{E}\left[R^2 X^2\right] = \mathbf{E}\left[X^2\right]\mathbf{E}\left[X^2\right] = \mathbf{E}\left[X^2\right]^2 = 1.$$

This shows that X^2 and Y^2 are not uncorrelated, which prevents that X^2 and Y^2 are not independent. Eventually, X and Y cannot be independent.

3. Still on account that $R^2 \sim Dirac(1)$, we have

$$\mathbf{E}[RY] = \mathbf{E}[RRX] = \mathbf{E}\left[R^2 X\right] = \mathbf{E}[X] = 0 = \mathbf{E}[X]\,\mathbf{E}[R].$$

This shows that R and Y are uncorrelated. On the other hand, since $Y \equiv R \cdot X \sim N(0, 1)$ the intuition is that the observation of the values taken by R transmits no information on the values taken by Y. Hence, the intuition is that R and Y are independent. To prove this, we show that

$$\mathbf{P}(R \leq r, Y \leq y) = \mathbf{P}(R \leq r)\,\mathbf{P}(Y \leq y), \tag{2}$$

for all $r, y \in \mathbb{R}$. In fact, still on account that $\{R = 1\}$, $\{R = -1\}$ constitute a partition of Ω, the random variables R and X are independent and X is symmetric about 0, we have

$$
\begin{aligned}
&\mathbf{P}(R \leq r, Y \leq y,) \\
&= \mathbf{P}(R \leq r, Y \leq y, R = 1) + \mathbf{P}(R \leq r, Y \leq y, R = -1) \\
&= \mathbf{P}(R \leq r, XR \leq y, R = 1) + \mathbf{P}(R \leq r, XR \leq y, R = -1) \\
&= \mathbf{P}(R \leq r, XR \leq y \mid R = 1)\, \mathbf{P}(R = 1) + \mathbf{P}(R \leq r, XR \leq y \mid R = -1)\, \mathbf{P}(R = -1) \\
&= \tfrac{1}{2}\left(\mathbf{P}(1 \leq r, X \leq y \mid R = 1) + \mathbf{P}(-1 \leq r, X \geq -y \mid R = -1)\right) \\
&= \begin{cases} 0 & \text{if } r < -1 \\ \frac{1}{2}\mathbf{P}(X \geq -y \mid R = -1) = \frac{1}{2}\mathbf{P}(X \geq -y) = \frac{1}{2}\mathbf{P}(X \leq y) & \text{if } -1 \leq r < 1 \\ \frac{1}{2}\left(\mathbf{P}(X \leq y \mid R = 1) + \mathbf{P}(X \geq -y \mid R = -1)\right) = \frac{1}{2}\left(\mathbf{P}(X \leq y) + \mathbf{P}(X \geq -y)\right) = \mathbf{P}(X \leq y) & \text{if } 1 \leq r \end{cases}
\end{aligned}
$$

On the other hand

$$
\mathbf{P}(R \leq r)\, \mathbf{P}(Y \leq y) = \begin{cases} 0 & \text{if } r < -1 \\ \frac{1}{2}\mathbf{P}(Y \leq y) = \frac{1}{2}\mathbf{P}(X \leq y) & \text{if } -1 \leq r < 1 \\ \mathbf{P}(Y \leq y) = \mathbf{P}(X \leq y) & \text{if } 1 \leq r \end{cases}
$$

Therefore, the random variables R and Y are independent.

4. If the random vector $(X, Y)^{\intercal}$ had a bivariate Gaussian distribution, the random variable $Z \equiv X + Y$ should be Gaussian distributed. On the other hand,

$$
Z = X + Y = X + RX = (R + 1)\, X.
$$

Hence,

$$
\begin{aligned}
F_Z(x) &= \mathbf{P}(Z \leq z) \\
&= \mathbf{P}(Z \leq z, R = 1) + \mathbf{P}(Z \leq z, R = -1) \\
&= \mathbf{P}(Z \leq z \mid R = 1)\, \mathbf{P}(R = 1) + \mathbf{P}(Z \leq z \mid R = -1)\, \mathbf{P}(R = -1) \\
&= \tfrac{1}{2}\left(\mathbf{P}((R + 1)\, X \leq z \mid R = 1) + \mathbf{P}((R + 1)\, X \leq z \mid R = -1)\right) \\
&= \tfrac{1}{2}\left(\mathbf{P}(2X \leq z \mid R = 1) + \mathbf{P}(0 \leq z \mid R = -1)\right).
\end{aligned}
$$

Now, we have that the events

$$
\{2X \leq z\} \quad \text{and} \quad \{R = 1\}
$$

are independent. Moreover,

$$
\begin{aligned}
\{0 \leq z\} = \Omega \quad &\text{if } z \geq 0 \\
\{0 \leq z\} = \varnothing \quad &\text{if } z < 0
\end{aligned}
$$

Hence,

$$
F_Z(x) = \begin{cases} \frac{1}{2}\mathbf{P}(2X \leq z) & \text{if } z < 0 \\ \frac{1}{2}\left(\mathbf{P}(2X \leq z) + 1\right) & \text{if } z \geq 0 \end{cases}
$$

in particular, if $z < 0$, we have

$$
F_Z(x) \leq \frac{1}{2}\mathbf{P}(2X \leq 0) = \frac{1}{4}
$$

and if $z \geq 0$

$$
F_Z(x) \geq \frac{1}{2}\left(\mathbf{P}(2X \leq 0) + 1\right) = \frac{1}{2}\left(\frac{1}{2} + 1\right) = \frac{3}{4},
$$

Hence, F_Z cannot not continuous at $z = 0$. This prevents Z to be Gaussian.

5. By virtue of what shown above and the properties of the conditional expectation, we have,

$$\mathbf{E}[Y \mid X] = \mathbf{E}[RX \mid X] = X\mathbf{E}[R \mid X] = X\mathbf{E}[R] = 0$$

and

$$\mathbf{E}[X \mid Y] = \mathbf{E}\left[XR^2 \mid Y\right] = \mathbf{E}[XRR \mid Y] = \mathbf{E}[YR \mid Y] = Y\mathbf{E}[R \mid Y] = Y\mathbf{E}[R] = 0.$$

Problem 3 *Let* $(\Omega, \mathcal{E}, \mathbf{P}) \equiv \Omega$ *be a probability space and let* $(\mathbb{R}, \mathcal{B}(\mathbb{R})) \equiv \mathbb{R}$ *be the Euclidean real line endowed with the Borel* σ*-algebra. Let* $X, Y \in L^2(\Omega; \mathbb{R})$.

1. *Prove in all details that* $\mathbf{E}[Y \mid X] = \mathbf{E}[Y]$ *a.e. on* Ω *implies* $Cov(X, Y) = 0$*, but* X *and* Y *may not be independent.*

2. *Prove in all details that* $Cov(X, Y) = 0$ *does not imply* $\mathbf{E}[Y \mid X] = \mathbf{E}[Y]$.

Exercise 4 *Hint: in the first case, to generate a suitable counterexample one may consider the random variables* $X \sim Ber(p)$*,* $Z \sim N(0,1)$*, independent of* X*, and* $Y = XZ$*. In the second case consider* $X \sim N(0,1)$ *and* $Y = X^2$.

Solution.

1. Under the assumption $\mathbf{E}[Y \mid X] = \mathbf{E}[Y]$ a.e. on Ω, by virtue of the properties of the conditional expectation operator, we can write

$$\mathbf{E}[XY] = \mathbf{E}[\mathbf{E}[XY \mid X]] = \mathbf{E}[X\mathbf{E}[Y \mid X]] = \mathbf{E}[X\mathbf{E}[Y]] = \mathbf{E}[X]\,\mathbf{E}[Y]$$

Therefore,

$$Cov(X, Y) = \mathbf{E}[XY] - \mathbf{E}[X]\,\mathbf{E}[Y] = 0.$$

Now, if we consider the random $X \sim Ber(p)$, $Z \sim N(0,1)$, independent of X, and $Y = XZ$, we have

$$\begin{aligned} Cov(X, Y) &= \mathbf{E}[XY] - \mathbf{E}[X]\,\mathbf{E}[Y] = \mathbf{E}\left[X^2 Z\right] - \mathbf{E}[X]\,\mathbf{E}[XZ] \\ &= \mathbf{E}\left[X^2\right]\mathbf{E}[Z] - \mathbf{E}[X]^2\,\mathbf{E}[Z] \\ &= 0. \end{aligned}$$

On the other hand, we have

$$\mathbf{P}(X \leq 0) = q,$$

and, on account that X and Z are independent,

$$\begin{aligned} \mathbf{P}(Y \leq 0) &= \mathbf{P}(XZ \leq 0) \\ &= \mathbf{P}(XZ \leq 0, X = 0) + \mathbf{P}(XZ \leq 0, X = 1) \\ &= \mathbf{P}(XZ \leq 0 \mid X = 0)\,\mathbf{P}(X = 0) + \mathbf{P}(XZ \leq 0 \mid X = 1)\,\mathbf{P}(X = 1) \\ &= \mathbf{P}(0 \leq 0 \mid X = 0)\,\mathbf{P}(X = 0) + \mathbf{P}(Z \leq 0 \mid X = 1)\,\mathbf{P}(X = 1) \\ &= \mathbf{P}(\Omega)\,\mathbf{P}(X = 0) + \mathbf{P}(Z \leq 0)\,\mathbf{P}(X = 1) \\ &= q + \frac{1}{2}p. \end{aligned}$$

Furthermore, the same arguments as above shows that

$$\begin{aligned}\mathbf{P}(X \leq 0, Y \leq 0) &= \mathbf{P}(X \leq 0, XZ \leq 0) \\ &= \mathbf{P}(X = 0, XZ \leq 0) \\ &= \mathbf{P}(XZ \leq 0 \mid X = 0)\,\mathbf{P}(X = 0) \\ &= q.\end{aligned}$$

Hence, we have

$$\mathbf{P}(X \leq 0)\,\mathbf{P}(Y \leq 0) = q\left(q + \frac{1}{2}p\right) \neq q = \mathbf{P}(X \leq 0, Y \leq 0)$$

which shows that X and Y are not be independent.

2. To show that $Cov(X, Y) = 0$ does not imply $\mathbf{E}[Y \mid X] = \mathbf{E}[Y]$, we consider $X \sim N(0,1)$ and $Y = X^2$. We have

$$\begin{aligned}Cov(X, Y) &= \mathbf{E}[XY] - \mathbf{E}[X]\,\mathbf{E}[Y] \\ &= \mathbf{E}\left[X^3\right] - \mathbf{E}[X]\,\mathbf{E}\left[X^2\right] \\ &= 0,\end{aligned}$$

but

$$\mathbf{E}[Y \mid X] = \mathbf{E}\left[X^2 \mid X\right] = X^2 \neq \mathbf{E}\left[X^2\right] = \mathbf{E}[Y].$$

Problem 5 *Let $(\Omega, \mathcal{E}, \mathbf{P}) \equiv \Omega$ be a probability space, let $\mathcal{F}$ be a sub-σ-algebra of $\mathcal{E}$, and let X, Y be real random variables on Ω with finite second order moment.*

1. *Show that*
$$\mathbf{E}\left[(X - \mathbf{E}[X \mid \mathcal{F}])^2\right] \leq \mathbf{E}\left[(X - \mathbf{E}[X])^2\right].$$

2. *Show that*
$$\mathbf{E}[XY \mid \mathcal{F}]^2 \leq \mathbf{E}\left[X^2 \mid \mathcal{F}\right]\mathbf{E}\left[Y^2 \mid \mathcal{F}\right]. \tag{3}$$

Solution.

1. In the space $\mathcal{L}^2(\Omega_{\mathcal{F}}; \mathbb{R})$ of the real $\mathcal{F}$-random variables having fnite moment of the second order the conditional expectation of X given $\mathcal{F}$ is characterized as

$$\mathbf{E}[X \mid \mathcal{F}] = \underset{Y \in \mathcal{L}^2(\Omega_{\mathcal{F}}; \mathbb{R})}{\arg\min}\ \mathbf{E}\left[(X - Y)^2\right].$$

This means that

$$\mathbf{E}\left[(X - \mathbf{E}[X \mid \mathcal{F}])^2\right] \leq \mathbf{E}\left[(X - Y)^2\right],$$

for every $Y \in \mathcal{L}^2(\Omega_{\mathcal{F}}; \mathbb{R})$. In particular, since the deterministic random variable $\mathbf{E}[X] \equiv \mathbf{E}[X]\ 1_{\Omega}$ is clearly in $\mathcal{L}^2(\Omega_{\mathcal{F}}; \mathbb{R})$, setting $Y \equiv \mathbf{E}[X]$ we obtain the desired inequality.

2. Note first that for all real random variables X, Y on Ω we have

$$|XY| \leq \frac{1}{2}\left(X^2 + Y^2\right).$$

Therefore, the assumption that X and Y have finite second moment implies that XY has finite first order moment. Hence, both the sides of (3) are well defined. Now, given any $z \in \mathbb{R}$, the random variable $X + zY$ has finite second order moment and, by virtue of the positivity of the conditional expectation operator, we have

$$\mathbf{E}\left[(X+zY)^2 \mid \mathcal{F}\right] \geq 0.$$

On the other hand, the linearity of the conditional expectation operator implies

$$\mathbf{E}\left[(X+zY)^2 \mid \mathcal{F}\right] = \mathbf{E}\left[X^2 \mid \mathcal{F}\right] + 2z\mathbf{E}\left[XY \mid \mathcal{F}\right] + z^2\mathbf{E}\left[Y^2 \mid \mathcal{F}\right].$$

As a consequence, we can write

$$\mathbf{E}\left[X^2 \mid \mathcal{F}\right] + 2z\mathbf{E}\left[XY \mid \mathcal{F}\right] + z^2\mathbf{E}\left[Y^2 \mid \mathcal{F}\right] \geq 0$$

for every $z \in \mathbb{R}$. It follows that

$$\Delta \equiv \mathbf{E}\left[XY \mid \mathcal{F}\right]^2 - \mathbf{E}\left[X^2 \mid \mathcal{F}\right]\mathbf{E}\left[Y^2 \mid \mathcal{F}\right] \leq 0,$$

which is the desired (3).

Problem 6 *Let $(\Omega, \mathcal{E}, \mathbf{P}) \equiv \Omega$ be a probability space and let $(X_n)_{n\geq 1}$ be a sequence of independent and identically distributed real random variables. Consider the sample sum*

$$Z_n \stackrel{def}{=} \sum_{k=1}^n X_k.$$

1. *Determine $\mathbf{E}\left[X_n \mid X_m\right]$ for every $m, n \geq 1$.*
2. *Determine $\mathbf{E}\left[Z_n \mid X_m\right]$ for every $n \geq 1$ and $m \leq n$.*
3. *Determine $\mathbf{E}\left[Z_n \mid Z_m\right]$ for every $n \geq 1$ and $m \leq n$.*
4. *Compute $\mathbf{E}\left[Z_n\right]$ and $\mathbf{D}\left[Z_n\right]$, for every $n \geq 1$.*
5. *Assume $X_n \sim Ber\,(p)$, for some $p \in (0,1)$, determine the distribution of Z_n.*
6. *Assume $X_n \sim N\left(\mu, \sigma^2\right)$, for some $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}_{++}$, determine the distribution of Z_n.*
7. *Assume $X \sim Exp\,(\lambda)$, for some $\lambda > 0$, determine the distribution of Z_n.*

Solution.

1. Since the random variables of the sequence $(X_n)_{n\geq 1}$ are independent, we have

$$\mathbf{E}\left[X_n \mid X_m\right] = \begin{cases} \mu & \text{if } n \neq m \\ X_n & \text{if } n = m \end{cases},$$

where μ is the value of the mean of the identically distributed random variables of the sequence $(X_n)_{n\geq 1}$.

2. By virtue of the linearity of the conditional expectation operator, we have

$$\begin{aligned}\mathbf{E}\left[Z_n \mid X_m\right] &= \mathbf{E}\left[\textstyle\sum_{k=1}^{n} X_k \mid X_m\right] = \textstyle\sum_{k=1}^{n} \mathbf{E}\left[X_k \mid X_m\right] \\ &= \textstyle\sum_{k=1,k\neq m}^{n} \mathbf{E}\left[X_k \mid X_m\right] + \mathbf{E}\left[X_m \mid X_m\right] \\ &= \textstyle\sum_{k=1,k\neq m}^{n} \mathbf{E}\left[X_k\right] + X_m \\ &= \textstyle\sum_{k=1,k\neq m}^{n} \mu + X_m \\ &= \mu \textstyle\sum_{k=1,k\neq m}^{n} 1 + X_m \\ &= (n-1)\,\mu + X_m,\end{aligned}$$

where μ is the value of the mean of the identically distributed random variables of the sequence $(X_n)_{n\geq 1}$.

3. In case $m = n$, we have trivially

$$\mathbf{E}\left[Z_n \mid Z_m\right] = \mathbf{E}\left[Z_n \mid Z_n\right] = Z_n.$$

In case $m < n$, by virtue of the linearity of the conditional expectation operator, we can write

$$\begin{aligned}\mathbf{E}\left[Z_n \mid Z_m\right] &= \mathbf{E}\left[\textstyle\sum_{k=1}^{n} X_k \mid \textstyle\sum_{k=1}^{m} X_k\right] \\ &= \mathbf{E}\left[\textstyle\sum_{k=1}^{m} X_k + \textstyle\sum_{k=m+1}^{n} X_k \mid \textstyle\sum_{k=1}^{m} X_k\right] \\ &= \mathbf{E}\left[\textstyle\sum_{k=1}^{m} X_k \mid \textstyle\sum_{k=1}^{m} X_k\right] + \mathbf{E}\left[\textstyle\sum_{k=m+1}^{n} X_k \mid \textstyle\sum_{k=1}^{m} X_k\right] \\ &= \mathbf{E}\left[Z_m \mid Z_m\right] \\ &= Z_m + \textstyle\sum_{k=m+1}^{n} \mathbf{E}\left[X_k \mid \textstyle\sum_{k=1}^{m} X_k\right].\end{aligned}$$

On the other hand the independence of the random variables of the sequence $(X_n)_{n\geq 1}$ implies that each X_n is independent of $f(X_1, \dots, X_{n-1})$ for every $n \geq 1$, where $f : \mathbb{R} \to \mathbb{R}_+$ is any Borel function. It follows that

$$\textstyle\sum_{k=m+1}^{n} \mathbf{E}\left[X_k \mid \textstyle\sum_{k=1}^{m} X_k\right] = \textstyle\sum_{k=m+1}^{n} \mathbf{E}\left[X_k\right] = \textstyle\sum_{k=m+1}^{n} \mu = \mu \textstyle\sum_{k=m+1}^{n} 1 = (n-m)\,\mu.$$

As a consequence,

$$\mathbf{E}\left[Z_n \mid Z_m\right] = Z_m + (n-m)\,\mu.$$

4. Writing σ^2 for the alue of the variance of the identically distributed random variables of the sequence $(X_n)_{n\geq 1}$, thanks to the linearity of the operator expectation and the linearity of the operator variance on independent random variables, we have

$$\mathbf{E}\left[Z_n\right] = \mathbf{E}\left[\textstyle\sum_{k=1}^{n} X_k\right] = \textstyle\sum_{k=1}^{n} \mathbf{E}\left[X_k\right] = \textstyle\sum_{k=1}^{n} \mu = \mu \textstyle\sum_{k=1}^{n} 1 = n\mu$$

and

$$\mathbf{D}^2\left[Z_n\right] = \mathbf{D}^2\left[\textstyle\sum_{k=1}^{n} X_k\right] = \textstyle\sum_{k=1}^{n} \mathbf{D}^2\left[X_k\right] = \textstyle\sum_{k=1}^{n} \sigma^2 = \sigma^2 \textstyle\sum_{k=1}^{n} 1 = n\sigma^2.$$

5. Under the assumption $X_n \sim Ber(p)$, for some $p \in (0,1)$, it is well known (see sum of independent and Bernoulli distributed random variables e.g. sec. Statistics on Simple Random Samples of Notes) that $Z_n \sim Bin(n,p)$.

6. Under the assumption $X_n \sim N(\mu, \sigma^2)$, for some $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}_{++}$, it is well known (see sum of independent and Gaussian distributed random variables e.g. sec. Statistics on Simple Random Samples of Notes) that $Z_n \sim N(n\mu, n\sigma^2)$.

7. Under the assumption $X \sim Exp(\lambda)$, for some $\lambda > 0$, it is well known (see sum of independent and exponentially distributed random variables e.g. sec. Statistics on Simple Random Samples of Notes) that $Z_n \sim \Gamma(n, \lambda)$.

Problem 7 *Let $(\Omega, \mathcal{E}, \mathbf{P}) \equiv \Omega$ be a probability space and let X and Y be independent standard normally distributed random variables on Ω. Set*

$$U \stackrel{def}{=} X + Y, \qquad V \stackrel{def}{=} X - Y.$$

1. *Compute the distributions of U and V.*
2. *Prove that U and V are independent.*
3. *Compute $\mathbf{E}[X \mid U]$, $\mathbf{E}[X \mid V]$, $\mathbf{E}[Y \mid U]$, $\mathbf{E}[Y \mid V]$.*
4. *Compute $\mathbf{E}[XY \mid U]$.*

Exercise 8 *Hint: First, concentrate your attention on the circumstance that X and Y are independent and standard normally distributed. Second, it might be useful to consider $\mathbf{E}\left[X^2 \mid U\right]$ and $\mathbf{E}\left[Y^2 \mid U\right]$.*

Solution.

1. Since X and Y are independent standard normally distributed random variables, by virtue of Proposition 631 of Notes U and V are normally distributed with mean

$$\mu_U = \mu_V = \mu_X + \mu_Y = 0$$

and variance

$$\sigma_U^2 = \sigma_V^2 = \sigma_X^2 + \sigma_Y^2 = 1.$$

2. We can write

$$\begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}.$$

Therefore we can apply Theorem 642 of Notes to obtain that the random variables U and V are jointly Gaussian distributed. Now, since we have

$$\begin{aligned} Cov(U, V) &= Cov(X + Y, X - Y) = Cov(X, X) - Cov(X, Y) + Cov(Y, X) - Cov(Y, Y) \\ &= \mathbf{D}^2[X] - Cov(X, Y) + Cov(X, Y) - \mathbf{D}^2[Y] \\ &= 0 \end{aligned}$$

We can apply Proposition 652 of Notes to obtain that the random variables U and V are independent.

3. Observe that, by virtue of the properties of the conditional expectation operator, we have

$$U = \mathbf{E}[U \mid U] = \mathbf{E}[X + Y \mid U] = \mathbf{E}[X \mid U] + \mathbf{E}[Y \mid U]$$

and

$$V = \mathbf{E}[U \mid U] = \mathbf{E}[X - Y \mid U] = \mathbf{E}[X \mid V] - \mathbf{E}[Y \mid V]$$

Therefore, to compute the desired conditional expectations it is clearly sufficient to compute $\mathbf{E}[X \mid U]$ and $\mathbf{E}[X \mid V]$. On the other hand, we have

$$\begin{pmatrix} X \\ U \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} X \\ V \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}.$$

This, still on account of Theorem 642 of Notes, proves that both the vectors (X, U) and (X, V) are Gaussian. It is then possible to apply Corollary 704 of Notes to obtain that

$$\mathbf{E}[X \mid U] = \mathbf{E}[X] + \frac{Cov(X,U)}{\mathbf{D}^2[U]}(U - \mathbf{E}[U]) = \frac{1}{2}U$$

and

$$\mathbf{E}[X \mid V] = \mathbf{E}[X] + \frac{Cov(X,V)}{\mathbf{D}^2[V]}(V - \mathbf{E}[V]) = \frac{1}{2}V$$

since

$$Cov(X,U) = Cov(X, X+Y) = Cov(X,X) + Cov(X,Y) = \mathbf{D}^2[X] = 1$$

and

$$Cov(X,V) = Cov(X, X-Y) = Cov(X,X) - Cov(X,Y) = \mathbf{D}^2[X] = 1$$

It follows that

$$\mathbf{E}[Y \mid U] = U - \mathbf{E}[X \mid U] = \frac{1}{2}U$$

and

$$\mathbf{E}[Y \mid V] = \mathbf{E}[X \mid V] - V = -\frac{1}{2}V.$$

4. To compute $\mathbf{E}[XY \mid U]$, observe that, by virtue of the properties of the conditional expectation operator, we can write

$$U^2 = \mathbf{E}\left[U^2 \mid U\right] = \mathbf{E}\left[(X+Y)^2 \mid U\right] = \mathbf{E}\left[X^2 \mid U\right] + 2\mathbf{E}[XY \mid U] + \mathbf{E}\left[Y^2 \mid U\right].$$

It follows

$$\mathbf{E}[XY \mid U] = U^2 - \frac{1}{2}\left(\mathbf{E}\left[X^2 \mid U\right] + \mathbf{E}\left[Y^2 \mid U\right]\right).$$

We already know that the vector (X, U) is Gaussian and likewise also the vector (Y, U) is Gaussian. Therefore, we can compute $\mathbf{E}\left[X^2 \mid U\right]$ and $\mathbf{E}\left[Y^2 \mid U\right]$ applying again Corollary 704. We obtain

$$\mathbf{E}\left[X^2 \mid U\right] = \mathbf{D}^2[X] - \frac{Cov(X,U)^2}{\mathbf{D}^2[U]} + \left(\mathbf{E}[X] + \frac{Cov(X,U)}{\mathbf{D}^2[U]}(U - \mathbf{E}[U])\right)^2 = \frac{1}{2}\left(1 + \frac{1}{2}U^2\right)$$

amd

$$\mathbf{E}\left[Y^2 \mid U\right] = \mathbf{D}^2[Y] - \frac{Cov(Y,U)^2}{\mathbf{D}^2[U]} + \left(\mathbf{E}[Y] + \frac{Cov(Y,U)}{\mathbf{D}^2[U]}(U - \mathbf{E}[U])\right)^2 = \frac{1}{2}\left(1 + \frac{1}{2}U^2\right)$$

since

$$Cov(Y,U) = Cov(Y, X+Y) = Cov(Y,X) + Cov(Y,Y) = \mathbf{D}^2[Y] = 1.$$

In the end

$$\begin{aligned}\mathbf{E}[XY \mid U] &= U^2 - \frac{1}{2}\left(\mathbf{E}\left[X^2 \mid U\right] + \mathbf{E}\left[Y^2 \mid U\right]\right) \\ &= U^2 - \frac{1}{2}\left(1 + \frac{1}{2}U^2\right) \\ &= \frac{1}{2}\left(\frac{3}{2}U^2 - 1\right).\end{aligned}$$

Problem 9 *Let N be a geometric random variable with success probability p, which models the first occurrence of success in n independent trials, and let $(X_n)_{n\geq 1}$ be a sequence of independent and normally distributed random variables with mean μ and variance σ^2, which are also independent of N. Study the conditional expectation*

$$\mathbf{E}\left[\sum_{k=1}^{N} X_k \mid N\right].$$

Use the properties of the conditional expectation to compute the expectation and the variance of the random sum

$$S_N \stackrel{def}{=} \sum_{k=1}^{N} X_k.$$

Solution. Since the random variables of the sequence $(X_n)_{n\geq 1}$ are independent and are also independent of N, which is geometrically distributed, we can write

$$\begin{aligned}
\mathbf{E}\left[\sum_{k=1}^{N} X_k \mid N\right] &= \sum_{n=1}^{\infty} \mathbf{E}\left[\sum_{k=1}^{N} X_k \mid N=n\right] 1_{\{N=n\}} \\
&= \sum_{n=1}^{\infty} \left(\frac{1}{\mathbf{P}(N=n)} \int_{\Omega} \left(\sum_{k=1}^{N} X_k\right) 1_{\{N=n\}}\, d\mathbf{P}\right) 1_{\{N=n\}} \\
&= \sum_{n=1}^{\infty} \left(\frac{1}{\mathbf{P}(N=n)} \int_{\{N=n\}} \left(\sum_{k=1}^{N} X_k\right) d\mathbf{P}\right) 1_{\{N=n\}} \\
&= \sum_{n=1}^{\infty} \left(\frac{1}{\mathbf{P}(N=n)} \int_{\{N=n\}} \left(\sum_{k=1}^{n} X_k\right) d\mathbf{P}\right) 1_{\{N=n\}} \\
&= \sum_{n=1}^{\infty} \left(\frac{1}{\mathbf{P}(N=n)} \int_{\Omega} \left(\sum_{k=1}^{n} X_k\right) 1_{\{N=n\}}\, d\mathbf{P}\right) 1_{\{N=n\}} \\
&= \sum_{n=1}^{\infty} \left(\frac{1}{\mathbf{P}(N=n)} \mathbf{E}\left[\left(\sum_{k=1}^{n} X_k\right) 1_{\{N=n\}}\right]\right) 1_{\{N=n\}} \\
&= \sum_{n=1}^{\infty} \left(\frac{1}{\mathbf{P}(N=n)} \left(\sum_{k=1}^{n} \mathbf{E}[X_k]\right) \mathbf{E}\left[1_{\{N=n\}}\right]\right) 1_{\{N=n_m\}} \\
&= \sum_{n=1}^{\infty} \left(\frac{1}{\mathbf{P}(N=n)} \left(\sum_{k=1}^{n} \mu\right) \mathbf{P}(N=n)\right) 1_{\{N=n_m\}} \\
&= \sum_{n=1}^{\infty} n\mu 1_{\{N=n_m\}} \\
&= \mu \sum_{n=1}^{\infty} n 1_{\{N=n\}} \\
&\stackrel{\mathbf{P}\text{-a.s.}}{=} \mu N.
\end{aligned}$$

Now, we can write

$$\mathbf{E}\left[\sum_{k=1}^{N} X_k\right] = \mathbf{E}\left[\mathbf{E}\left[\sum_{k=1}^{N} X_k \mid N\right]\right] = \mathbf{E}[\mu N] = \mu \mathbf{E}[N] = \frac{\mu}{p}.$$

and

$$\mathbf{D}^2\left[\sum_{k=1}^{N} X_k\right] = \mathbf{E}\left[\left(\sum_{k=1}^{N} X_k\right)^2\right] - \mathbf{E}\left[\sum_{k=1}^{N} X_k\right]^2 = \mathbf{E}\left[\mathbf{E}\left[\left(\sum_{k=1}^{N} X_k\right)^2 \mid N\right]\right] - \frac{\mu^2}{p^2}.$$

Thus, we are left with computing

$$\mathbf{E}\left[\left(\sum_{k=1}^{N} X_k\right)^2 \mid N\right].$$

A straightforward computation yields

$$\begin{aligned}
\mathbf{E}\left[\left(\sum_{k=1}^{N} X_k\right)^2 \mid N\right] &= \sum_{n=1}^{\infty} \mathbf{E}\left[\left(\sum_{k=1}^{N} X_k\right)^2 \mid N=n\right] 1_{\{N=n\}} \\
&= \sum_{n=1}^{\infty} \left(\frac{1}{\mathbf{P}\left(N=n\right)} \int_{\Omega} \left(\sum_{k=1}^{N} X_k\right)^2 1_{\{N=n\}}\, d\mathbf{P}\right) 1_{\{N=n\}} \\
&= \sum_{n=1}^{\infty} \left(\frac{1}{\mathbf{P}\left(N=n\right)} \int_{\{N=n\}} \left(\sum_{k=1}^{N} X_k\right)^2 d\mathbf{P}\right) 1_{\{N=n\}} \\
&= \sum_{n=1}^{\infty} \left(\frac{1}{\mathbf{P}\left(N=n\right)} \int_{\{N=n\}} \left(\sum_{k=1}^{n} X_k\right)^2 d\mathbf{P}\right) 1_{\{N=n\}} \\
&= \sum_{n=1}^{\infty} \left(\frac{1}{\mathbf{P}\left(N=n\right)} \int_{\Omega} \left(\sum_{k=1}^{n} X_k\right)^2 1_{\{N=n\}}\, d\mathbf{P}\right) 1_{\{N=n\}} \\
&= \sum_{n=1}^{\infty} \left(\frac{1}{\mathbf{P}\left(N=n\right)} \mathbf{E}\left[\left(\sum_{k=1}^{n} X_k\right)^2 1_{\{N=n\}}\right]\right) 1_{\{N=n\}} \\
&= \sum_{n=1}^{\infty} \left(\frac{1}{\mathbf{P}\left(N=n\right)} \mathbf{E}\left[\left(\sum_{k=1}^{n} X_k\right)^2\right] \mathbf{E}\left[1_{\{N=n\}}\right]\right) 1_{\{N=n\}} \\
&= \sum_{n=1}^{\infty} \left(\frac{1}{\mathbf{P}\left(N=n\right)} \mathbf{E}\left[\left(\sum_{k=1}^{n} X_k\right)^2\right] \mathbf{P}\left(N=n\right)\right) 1_{\{N=n\}} \\
&= \sum_{n=1}^{\infty} \mathbf{E}\left[\left(\sum_{k=1}^{n} X_k\right)^2\right] 1_{\{N=n\}},
\end{aligned}$$

where

$$
\begin{aligned}
\mathbf{E}\left[\left(\sum_{k=1}^{n} X_k\right)^2\right] &= \mathbf{E}\left[\sum_{k=1}^{n} X_k^2 + \sum_{k,\ell=1}^{n} X_k X_\ell\right] \\
&= \sum_{k=1}^{n} \mathbf{E}\left[X_k^2\right] + \sum_{k,\ell=1}^{n} \mathbf{E}\left[X_k\right] \mathbf{E}\left[X_k\right] \\
&= \sum_{k=1}^{n} \left(\mu^2 + \sigma^2\right) + \sum_{k,\ell=1}^{n} \mu^2 \\
&= \left(\mu^2 + \sigma^2\right) n + \mu^2 (n-1) n \\
&= \sigma^2 n + \mu^2 n^2.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\mathbf{E}\left[\left(\sum_{k=1}^{N} X_k\right)^2 \mid N\right] &= \sum_{n=1}^{\infty} \left(\sigma^2 n + \mu^2 n^2\right) 1_{\{N=n\}} \\
&= \sigma^2 \sum_{n=1}^{\infty} n 1_{\{N=n\}} + \mu^2 \sum_{n=1}^{\infty} n^2 1_{\{N=n\}} \\
&= \sigma^2 N + \mu^2 N^2.
\end{aligned}
$$

It then follows

$$
\begin{aligned}
\mathbf{E}\left[\sigma^2 N + \mu^2 N^2\right] &= \sigma^2 \mathbf{E}[N] + \mu^2 \mathbf{E}\left[N^2\right] \\
&= \frac{\sigma^2}{p} + \mu^2 \left(\mathbf{D}^2[N] + \mathbf{E}[N]^2\right) \\
&= \frac{\sigma^2}{p} + \mu^2 \left(\frac{2-p}{p^2}\right).
\end{aligned}
$$

In the end,

$$
\mathbf{D}^2\left[\sum_{k=1}^{N} X_k\right] = \frac{\sigma^2}{p} + \mu^2\left(\frac{2-p}{p^2}\right) - \frac{\mu^2}{p^2} = \frac{\sigma^2}{p} + \mu^2\left(\frac{1-p}{p^2}\right).
$$

.

Problem 10 *Let $(\Omega, \mathcal{E}, \mathbf{P}) \equiv \Omega$ be a probability space and let $X, Y \in \mathcal{L}^2(\Omega; \mathbb{R})$ such that*

$$
\mathbf{E}[Y \mid X] = X \quad \text{and} \quad \mathbf{E}\left[Y^2 \mid X\right] = X^2.
$$

Prove that $Y = X$, $\mathbf{P}$-a.s. on Ω.

Solution. We have $Y = X$, $\mathbf{P}$-a.s. on Ω if and only if there exists an event $E \in \mathcal{E}$ such that $\mathbf{P}(E) = 0$ and $Y(\omega) = X(\omega)$ for every $\omega \in \Omega - E$. By virtue of the properties of the Lebesgue integral, we have

$$
Y = X, \ \mathbf{P}\text{-a.s. on } \Omega \Leftrightarrow \int_{\Omega} (X - Y)^2 \, d\mathbf{P} = 0.
$$

On the other hand,

$$
\int_{\Omega} (X - Y)^2 \, d\mathbf{P} \equiv \mathbf{E}\left[(X - Y)^2\right].
$$

Hence, we evaluate

$$\mathbf{E}\left[(X-Y)^2\right] = \mathbf{E}\left[X^2 - 2XY + Y^2\right] = \mathbf{E}\left[X^2\right] - 2\mathbf{E}\left[XY\right] + \mathbf{E}\left[Y^2\right]. \tag{4}$$

Now, by virtue of the properties of the conditional expectation operator, under our assumptions, we have

$$\mathbf{E}\left[XY\right] = \mathbf{E}\left[\mathbf{E}\left[XY \mid X\right]\right] = \mathbf{E}\left[X\mathbf{E}\left[Y \mid X\right]\right] = \mathbf{E}\left[X^2\right] \tag{5}$$

and

$$\mathbf{E}\left[Y^2\right] = \mathbf{E}\left[\mathbf{E}\left[Y^2 \mid X\right]\right] = \mathbf{E}\left[X^2\right]. \tag{6}$$

Combining (4)-(6) it follows

$$\mathbf{E}\left[(X-Y)^2\right] = 0,$$

which yields the deired result.