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 - Advanced Statistics
Instructors: Roberto Monte & Massimo Regoli
Problems on Conditional Expectation with Solution 2021-11-25

Problem 1 *Let $(\Omega, \mathcal{E}, \mathbf{P}) \equiv \Omega$ be a probability space and let $X, Y \in \mathcal{L}^2(\Omega; \mathbb{R})$. Assume that $\mathbf{E}\left[Y^2 \mid X\right] = X^2$ and $\mathbf{E}[Y \mid X] = X$. Show that $X = Y$ a.s. on Ω. Hint: is it true that $X = Y$ a.s. on Ω if and only if $\mathbf{E}\left[(X-Y)^2\right] = 0$?*

Solution. Recall that, since $(X-Y)^2 \geq 0$, thanks to the properties of the Lebesgue integral, we have

$$\mathbf{E}\left[(X-Y)^2\right] = 0 \Leftrightarrow X = Y \text{ a.s.on } \Omega.$$

Now, by virtues of the properties of the conditional expectation operator, we can write

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

On the other hand, since the random variable X is clearly measurable with respect to the σ-algebra $\sigma(X)$, we have

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

Therefore, the assumptions on $\mathbf{E}\left[Y^2 \mid X\right]$ and $\mathbf{E}[Y \mid X]$, allow us to conclude that

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

It follows

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

which implies the desired result.

Problem 2 *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 $N \subseteq \mathbb{N}$, let $\{F_n\}_{n \in N}$ be a complete system of mutually exclusive events of Ω and let $\mathcal{F}$ be the σ-algebra generated by $\{F_n\}_{n \in N}$. In symbols $\mathcal{F} \equiv \sigma(\{F_n\}_{n \in N})$. We know that a map $Y : \Omega \to \mathbb{R}$ is an $(\mathcal{F}, \mathcal{B}(\mathbb{R}))$-random variable if and only if*

$$Y(\omega) = \sum_{n \in N} y_n 1_{F_n}(\omega), \quad \forall \omega \in \Omega,$$

where $(y_n)_{n \in N}$ is a suitable sequence of real numbers.

Exercise 3 *Consider a random variable $X \in L^2(\Omega; \mathbb{R})$ and let $L^2(\Omega_{\mathcal{F}}; \mathbb{R})$ the space of all $(\mathcal{F}, \mathcal{B}(\mathbb{R}))$-random variables with finite second-order moment. Use the above claim to prove that*

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

As a consequence, after proving that $L^2(\Omega_{\mathcal{F}}; \mathbb{R})$ is a subspace of $L^2(\Omega; \mathbb{R})$, show that $\mathbf{E}[X \mid \mathcal{F}]$ is the orthogonal projection of X on $L^2(\Omega_{\mathcal{F}}; \mathbb{R})$.

Solution. The space $L^2(\Omega_{\mathcal{F}};\mathbb{R})$ of all $(\mathcal{F},\mathcal{B}(\mathbb{R}))$-random variables with finite second-order moment is a subspace of $L^2(\Omega;\mathbb{R})$ because fulfills the conditions for a subset of a Hilbert space to be a subspace of the Hilbert space. In fact, for all $X,Y\in L^2(\Omega_{\mathcal{F}};\mathbb{R})$, $(\mathcal{F},\mathcal{B}(\mathbb{R}))$-random variables with finite second-order moment, and all $\alpha,\beta\in\mathbb{R}$ the linear combination $\alpha X+\beta Y$ is also an $(\mathcal{F},\mathcal{B}(\mathbb{R}))$-random variable with finite second order moment, that is to say $L^2(\Omega_{\mathcal{F}};\mathbb{R})$ is closed with respect to the linear combination. In addition, if $(X_n)_{n\geq 1}$ is a sequence belonging to $L^2(\Omega_{\mathcal{F}};\mathbb{R})$ and such that $X_n\overset{L^2}{\to}X$, where $X\in L^2(\Omega;\mathbb{R})$, we have also $X\in L^2(\Omega_{\mathcal{F}};\mathbb{R})$, that is to say $L^2(\Omega_{\mathcal{F}};\mathbb{R})$ is a closed subset of $L^2(\Omega;\mathbb{R})$ in the topology induced by the norm.

Now, given $X\in L^2(\Omega_{\mathcal{F}};\mathbb{R})$, consider the functional $\Delta_X : L^2(\Omega_{\mathcal{F}};\mathbb{R})\to\mathbb{R}_+$ given by

$$\Delta_X(Y)\overset{\text{def}}{=}\mathbf{E}\left[(X-Y)^2\right],\quad \forall Y\in L^2(\Omega_{\mathcal{F}};\mathbb{R}).$$

Since in the case under consideration

$$Y\in L^2(\Omega_{\mathcal{F}};\mathbb{R})\Leftrightarrow Y(\omega)=\sum_{n\in N}y_n 1_{F_n}(\omega),\quad \forall\omega\in\Omega,$$

we can write

$$\Delta_X(Y)=\mathbf{E}\left[\left(X-\sum_{n\in N}y_n 1_{F_n}\right)^2\right]\equiv\Delta_X(y_1,\dots,y_n,\dots).$$

Hence,

$$\begin{aligned}\Delta_X(y_1,\dots,y_n,\dots)&=\mathbf{E}\left[X^2-2\sum_{n\in N}y_n X 1_{F_n}+\sum_{m,n\in N}y_m y_n 1_{F_m}1_{F_n}\right]\\&=\mathbf{E}\left[X^2\right]-2\sum_{n\in N}y_n\mathbf{E}\left[X1_{F_n}\right]+\sum_{m,n\in N}y_m y_n\mathbf{E}\left[1_{F_m}1_{F_n}\right].\end{aligned}$$

On the other hand,

$$1_{F_m}1_{F_n}=\begin{cases}1_{F_n} & \text{if } m=n\\ 1_{\varnothing} & \text{if } m\neq n\end{cases}.$$

Moreover,

$$\mathbf{E}\left[1_E\right]=\mathbf{P}(E),\quad \forall E\in\mathcal{E}$$

and

$$\mathbf{E}\left[X1_E\right]=\int_\Omega X1_E\,d\mathbf{P}=\int_E X\,d\mathbf{P},\quad \forall E\in\mathcal{E}.$$

Therefore,

$$\Delta_X(Y)=\mathbf{E}\left[X^2\right]-2\sum_{n\in N}y_n\int_{F_n}X d\mathbf{P}+\sum_{n\in N}y_n^2\mathbf{P}(F_n).$$

As a consequence,

$$\partial_{y_m}\Delta_X(y_1,\dots,y_n,\dots)=-2\int_{Fm}X d\mathbf{P}+2y_m\mathbf{P}(F_m),\quad \forall m\in N,$$

which implies

$$\partial_{y_m}\Delta_X(y_1,\dots,y_n,\dots)=0\Leftrightarrow y_m=\frac{1}{\mathbf{P}(F_m)}\int_{F_m}X d\mathbf{P}=\mathbf{E}\left[X\mid F_m\right],\quad \forall m\in N.$$

Thus, a candidate minimum Y for the functional $\Delta_X : L^2(\Omega_{\mathcal{F}};\mathbb{R}) \to \mathbb{R}_+$ takes the form

$$Y = \sum_{n \in N} \mathbf{E}[X \mid F_n] 1_{F_n} = \mathbf{E}[X \mid \mathcal{F}].$$

Now, we have

$$\partial^2_{y_m} \Delta_X (y_1, \dots, y_n, \dots) = \mathbf{P}(F_m) > 0$$

and the functional $\Delta_X : L^2(\Omega_{\mathcal{F}};\mathbb{R}) \to \mathbb{R}_+$ is known to be convex[1]. It then follow that

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

To complete the proof, it is sufficient to observe that in a Hilbert space the ortoghonal projection of a given vector onto a subspace determines the vector in the subspace of the minimum distance from the given vector.

Problem 4 *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.*

[1] To prove the convexity of the functional $\Delta_X : L^2(\Omega_{\mathcal{F}};\mathbb{R}) \to \mathbb{R}_+$, we may observe that thanks, to the Cauchy-Schwarz inequality and the convexity of the standard quadratic function $f(u) \stackrel{\text{def}}{=} u^2$, we have

$$\begin{aligned}
\Delta_X(\theta Y + (1-\theta) Z) &= \mathbf{E}\left[(X - (\theta Y + (1-\theta) Z))^2\right] \\
&= \mathbf{E}\left[(\theta(X-Y) + (1-\theta)(X-Z))^2\right] \\
&= \mathbf{E}\left[\theta^2 (X-Y)^2 + 2\theta(1-\theta)(X-Y)(X-Z) + (1-\theta)^2 (X-Z)^2\right] \\
&= \theta^2 \mathbf{E}\left[(X-Y)^2\right] + 2\theta(1-\theta)\mathbf{E}[(X-Y)(X-Z)] + (1-\theta)^2 \mathbf{E}\left[(X-Z)^2\right] \\
&\le \theta^2 \mathbf{E}\left[(X-Y)^2\right] + 2\theta(1-\theta)\left|\mathbf{E}[(X-Y)(X-Z)]\right| + (1-\theta)^2 \mathbf{E}\left[(X-Z)^2\right] \\
&\le \theta^2 \mathbf{E}\left[(X-Y)^2\right] + 2\theta(1-\theta)\mathbf{E}\left[(X-Y)^2\right]^{1/2} \mathbf{E}\left[(X-Z)^2\right]^{1/2} + (1-\theta)^2 \mathbf{E}\left[(X-Z)^2\right] \\
&= \left(\theta \mathbf{E}\left[(X-Y)^2\right]^{1/2} + (1-\theta)\mathbf{E}\left[(X-Z)^2\right]^{1/2}\right)^2 \\
&\le \theta \mathbf{E}\left[(X-Y)^2\right] + (1-\theta)\mathbf{E}\left[(X-Z)^2\right],
\end{aligned}$$

for every $\theta \in [0,1]$.

To show the convexity of the standard quadratic function, $f(u) \stackrel{\text{def}}{=} u^2$, we may observe that the inequality

$$(u-v)^2 \ge 0,$$

which holds true for every $u, v \in \mathbb{R}$, implies

$$-\theta(1-\theta)(u-v)^2 \le 0,$$

which holds true for every $u, v \in \mathbb{R}$ and $\theta \in [0,1]$. The latter can be rewritten as

$$-\theta(1-\theta)\left(u^2 - 2uv + v^2\right) \le 0$$

or equivalently

$$\theta^2 u^2 - \theta u^2 + 2\theta(1-\theta) uv + (1-\theta)^2 v^2 - (1-\theta) v^2.$$

This implies

$$\theta^2 u^2 + 2\theta(1-\theta) uv + (1-\theta)^2 v^2 \le \theta u^2 + (1-\theta) v^2.$$

Hence,

$$(\theta u + (1-\theta) v)^2 \le \theta u^2 + (1-\theta) v^2,$$

which proves the desired result.

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

   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. .

Problem 5 *Let* $(\Omega, \mathcal{E}, \mathbf{P}) \equiv \Omega$ *be a probability space and let* X *and* Y *be independent standard Gaussian 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* $\mathbf{E}[XY \mid U]$.

   Hint: 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 Gaussian random variables, X and Y are also jointly Gaussian, that is the random vector $(X, Y)^\intercal$ is Gaussian. By virtue of the equation

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

   it then follows that the vector $(U, V)^\intercal$ is Gaussian. Hence, its entries U and V are Gaussian. Now,

$$\mathbf{E}[U] = \mathbf{E}[X + Y] = \mathbf{E}[X] + \mathbf{E}[Y] = 0$$

   and

$$\mathbf{E}[V] = \mathbf{E}[X - Y] = \mathbf{E}[X] - \mathbf{E}[Y] = 0.$$

   Furthermore,

$$\mathbf{D}^2[U] = \mathbf{D}^2[X + Y] = \mathbf{D}^2[X] + \mathbf{D}^2[Y] = 2$$

   and

$$\mathbf{D}^2[V] = \mathbf{D}^2[X - Y] = \mathbf{D}^2[X] + \mathbf{D}^2[Y] = 2.$$

   We then have

$$U \sim V \sim N(0, 2).$$

2. We clearly have,

$$\mathbf{E}[U]\,\mathbf{E}[V] = 0$$

   Moreover,

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

   As a consequence,

$$Cov(U, V) = \mathbf{E}[UV] - \mathbf{E}[U]\,\mathbf{E}[V] = 0.$$

   On the other hand, the vector $(U, V)^\intercal$ is Gaussian. Thus, the zero correlation of U and Y implies the independence of U and V.

3. Note that we can write

$$\begin{pmatrix} X \\ U \end{pmatrix} = \begin{pmatrix} X \\ X+Y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}.$$

Similarly

$$\begin{pmatrix} X \\ V \end{pmatrix} = \begin{pmatrix} X \\ X-Y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}$$

Hence, the vectors $(X,U)^\intercal$ and $(X,V)^\intercal$ are Gaussian. As a consequence, thanks to the conditional expectation formula for the entries of a Gaussian vector, we can write

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

Similarly

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

The same argument implies

$$\mathbf{E}[Y \mid U] = -\frac{1}{2}U \quad \text{and} \quad \mathbf{E}[Y \mid V] = -\frac{1}{2}V.$$

Alternatively, thanks to the properties of the conditional expectation and independence of U and V we can write

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

and

$$\mathbf{E}[X \mid U] - \mathbf{E}[Y \mid U] = \mathbf{E}[X-Y \mid U] = \mathbf{E}[V \mid U] = \mathbf{E}[V] = 0$$

Solving for $\mathbf{E}[X \mid U] + \mathbf{E}[Y \mid U]$, we obtain

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

We can also write

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

and

$$\mathbf{E}[X \mid V] + \mathbf{E}[Y \mid V] = \mathbf{E}[X+Y \mid V] = \mathbf{E}[U \mid V] = \mathbf{E}[U] = 0.$$

It follows

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

Note that $\mathbf{E}[X \mid U]$ and $\mathbf{E}[X \mid V]$ are the linear regressions of X against U and V, respectively while $\mathbf{E}[Y \mid U]$ and $\mathbf{E}[Y \mid V]$ are the linear regressions of Y against U and V, respectively.

4. In the end, observe that we have

$$X = \frac{1}{2}(U+V) \quad \text{and} \quad Y = \frac{1}{2}(U-V).$$

Hence,

$$XY = \frac{1}{4}\left(U^2 - V^2\right).$$

It follows

$$\begin{aligned}\mathbf{E}[XY \mid U] &= \frac{1}{4}\mathbf{E}\left[U^2 - V^2 \mid U\right] \\ &= \frac{1}{4}\left(\mathbf{E}\left[U^2 \mid U\right] - \mathbf{E}\left[V^2 \mid U\right]\right) \\ &= \frac{1}{4}\left(U^2 - \mathbf{E}\left[V^2\right]\right) \\ &= \frac{1}{4}\left(U^2 - \mathbf{D}^2[V]\right) \\ &= \frac{1}{4}\left(U^2 - 2\right).\end{aligned}$$

Problem 6 *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*

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

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

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

Solution. .

Problem 7 *Let Z [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

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

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 $\mathbf{E}[R] = 0$ and thanks to the independence of X and R, which implies the independence of X^2 and R, we have

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

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

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

and

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

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

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

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

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

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

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

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,

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

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 8 *Let $(\Omega, \mathcal{E}, \mathbf{P}) \equiv \Omega$ a probability space, let X and Y be independent standard Rademacher random variables on Ω. Define $Z \overset{def}{=} X + Y$.*

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. .

Problem 9 *Let* $(\Omega, \mathcal{E}, \mathbf{P}) \equiv \Omega$ *a probability space, let* X *and* Y *be independent standard Bernoulli random variables on* Ω*. Define* $Z \overset{def}{=} X + Y$*.*

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 Bernoulli random variables, can you compute* $\mathbf{E}\left[(X+Y)^2 \mid Z\right]$ *and* $\mathbf{E}[XY \mid Z]$*?*

Solution. .