17 Conditional expectation/distribution

Exercise 17.1. Let $X \in \mathrm{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{G} = \sigma(\{\mathsf{B}_i : i \in \mathsf{I}\})$, for $(\mathsf{B}_i)_{i\in\mathsf{I}}$ pairwise disjoint events and I a countable set. Show that $\mathbb{P}$-almost surely

$$\mathbb{E}[X \mid \mathcal{G}] = \sum_{i\in\mathsf{I}} \mathbb{1}_{\mathsf{B}_i} \mathbb{E}[X \mid \mathsf{B}_i] \ . \tag{169}$$

Hint: First show that $\mathcal{G} = \{\mathsf{A} \subset \Omega : \text{there exists } \mathsf{J} \subset \mathsf{I} \text{ with } \mathsf{A} = \sqcup_{j\in\mathsf{J}} \mathsf{B}_j\}$.

Exercise 17.2. Let $(X_k)_{k\in\mathbb{N}^*}$ be a sequence of i.i.d. random variables valued in $\mathbb{N}$ and integrable. Let $T_n = X_N$ where $N_n \sim \mathrm{U}(\{1, \ldots, n\})$ independent of $(X_k)_{k\in\mathbb{N}^*}$ for any $n \in \mathbb{N}^*$.

(1) Show that for any $i \in \mathbb{N}^*$:

$$\mathbb{P}(T_n = i | X_{1:n}) = \frac{i \sum_{k=1}^n \mathbb{1}\{X_k = i\}}{\sum_{k=1}^n X_k} \ , \tag{170}$$

where $X_{1:n} = (X_1, \ldots, X_n)$.

(2) Show that for any $i \in \mathbb{N}$, $\lim_{n\to+\infty} \mathbb{P}(T_n = i) = ip_i/m$ denoting $p_i = \mathbb{P}(X_1 = i)$ and $m = \mathbb{E}[X_1]$.

Exercise 17.3. Let $\mathsf{G} \subset \mathcal{F}$ be a sub-σ field on $(\Omega, \mathcal{F}, \mathbb{P})$ and X a non-negative random variable. Denote $\mathsf{A} = \{\omega : \mathbb{E}[X|\mathsf{G}](\omega) > 0\}$.

(1) Show that $X\mathbb{1}_{\mathsf{A}} = 0$ almost surely and deduce that $\{X > 0\} \subset \mathsf{A}$. Hint: do not forget that X is supposed to be non-negative.

(2) Show that A is the smallest set in G (for the inclusion and up to negligeable set) containing $\{X > 0\}$.

Exercise 17.4. Let X be an integrable real random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{G}, \mathcal{H}$ two sub σ-field of $\mathcal{F}$. Denote by $\mathcal{H} \vee \mathcal{G} = \sigma(\mathcal{H} \cup \mathcal{G})$. Show that if $\sigma(X) \vee \mathcal{H} = \sigma(\sigma(X) \cup \mathcal{H})$ is independent of $\mathcal{G}$, $\mathbb{E}[X| \mathcal{H} \vee \mathcal{G}] = \mathbb{E}[X| \mathcal{H}]$. [Hint : first take an element of $\mathcal{H} \vee \mathcal{G}$ of the form $\mathsf{A} \cup \mathsf{B}$. Use the $\pi - \lambda$-theorem to conclude.]

Exercise 17.5. Let $X = F(Y, Z)$ where Y and Z are two random vectors on $(\Omega, \mathcal{F}, \mathbb{P})$ valued in $\mathbb{R}^p$ and $\mathbb{R}^q$, respectively, and F is a Borel function. Let $\mathcal{G} \subset \mathcal{F}$ be a sub σ-field. Suppose moreover that Y is $\mathcal{G}$-measurable and Z is independent of $\mathcal{G}$, then the conditional distribution of X given $\mathcal{G}$ is given by

$$\mathbb{P}^{X|\mathcal{G}}(\omega, \mathsf{A}) = \mathbb{P}(F(Y(\omega), Z) \in \mathsf{A}) \quad \text{for all } \omega \in \Omega \text{ and } \mathsf{A} \in \mathcal{Y} \ .$$

[Hint : first show in the case where $F = \mathbb{1}_{\mathsf{A}} \times \mathbb{1}_{\mathsf{B}}$ that $\mathbb{E}[X| \mathcal{G}] = \hat{F}(Y)$ where, for all y, $\hat{F}(y) = \mathbb{E}[F(y, Z)]$. Deduce the conditional distribution of (Y, Z) given $\mathcal{G}$ and conclude.]

Exercise 17.6. An important application of the projection theorem in Hilbert spaces is the computation of the conditional mean for L^2 random variables. It also provides an easy way to compute the conditional distribution in a Gaussian context, where the following result holds.

Proposition 17.1. The Hilbert space of all $\mathbb{R}^p$-valued L^2 random variables is endowed with the scalar product

$$\langle U, V \rangle = \mathbb{E}\left[U^\top V\right]$$

In this context, $\mathrm{Span}(1, \mathbf{Y})$ is seen as the linear space in L^2 obtained by a linear transformation of the random variables 1 and $\mathbf{Y}$, that is, we have

$$\mathrm{Span}(1, \mathbf{Y}) = \left\{a + \mathbf{A}\mathbf{Y} \ : \ a \in \mathbb{R}^p, \ \mathbf{A} \in \mathbb{R}^{p\times q}\right\} \tag{171}$$

$$= \left\{b + \mathbf{A}(\mathbf{Y} - \mathbb{E}[\mathbf{Y}]) \ : \ b \in \mathbb{R}^p, \ \mathbf{A} \in \mathbb{R}^{p\times q}\right\} \ , \tag{172}$$

where we set $b = a - \mathbf{A}\mathbb{E}[\mathbf{Y}]$. Let $p, q \geq 1$. Let $\mathbf{X}$ and $\mathbf{Y}$ be two jointly Gaussian vectors, respectively valued in $\mathbb{R}^p$ and $\mathbb{R}^q$. Then the following assertions hold.

(i) If $\text{Cov}(\mathbf{Y})$ is invertible, then $\widehat{\mathbf{X}} := \text{proj}\left(\mathbf{X} | \text{Span}\left(1, \mathbf{Y}\right)\right)$ is given by

$$\widehat{\mathbf{X}} = \mathbb{E}\left[\mathbf{X}\right] + \text{Cov}\left(\mathbf{X}, \mathbf{Y}\right) \text{Cov}(\mathbf{Y})^{-1} \left(\mathbf{Y} - \mathbb{E}\left[\mathbf{Y}\right]\right),$$

and

$$\text{Cov}(\mathbf{X} - \widehat{\mathbf{X}}) = \text{Cov}(\mathbf{X}) - \text{Cov}\left(\mathbf{X}, \mathbf{Y}\right) \text{Cov}(\mathbf{Y})^{-1} \text{Cov}\left(\mathbf{Y}, \mathbf{X}\right),$$

where here $\text{Span}\left(\dots\right)$ is understood as the space of $\mathbb{R}^p$-valued L^2 random variables obtained by linear transformations of $\dots$ and $\text{proj}\left(\cdot | \dots\right)$ is understood as the projection onto this space seen as a (closed) subspace of the Hilbert space of all $\mathbb{R}^p$-valued L^2 random variables.

(ii) We have

$$\mathbb{E}\left[\mathbf{X} | \mathbf{Y}\right] = \text{proj}\left(\mathbf{X} | \text{Span}\left(1, \mathbf{Y}\right)\right).$$

(iii) Let $\widehat{\mathbf{X}} = \mathbb{E}\left[\mathbf{X} | \mathbf{Y}\right]$. Then

$$\text{Cov}(\mathbf{X} - \widehat{\mathbf{X}}) = \mathbb{E}\left[\mathbf{X}(\mathbf{X} - \widehat{\mathbf{X}})^T\right] = \mathbb{E}\left[(\mathbf{X} - \widehat{\mathbf{X}})\mathbf{X}^T\right]$$

and the conditional distribution of $\mathbf{X}$ given $\mathbf{Y}$ is given by $\mathrm{N}(\widehat{\mathbf{X}}, \text{Cov}(\mathbf{X} - \widehat{\mathbf{X}}))$.

Let $\mathbf{X}$ and $\mathbf{Y}$ be as in Proposition 17.1.

(1) Use the characterization of the orthogonal projection to prove Proposition 17.1(i).

(2) In order to prove Proposition 17.1(ii) and (iii), use properties of the conditional distribution and expectation.

17.1 Conditional distribution

Exercise 17.7. Let X and Y be two independent real random variables with distribution $\mathbf{Pn}(\lambda)$ et $\mathbf{Pn}(\mu)$ respectively. Denote $S = X + Y$.

(i) Give the distribution of S.

(ii) For any $s \in \mathbb{N}$ give the conditional distribution of X given S.

(iii) Give $\mathbb{E}[X|S]$.

(iv) Check that $\text{Var}\left(\mathbb{E}\left[X|S\right]\right) \leq \text{Var}\left(X\right)$.

Exercise 17.8. Let X and Y be two independent real random variables with distribution $\mathbf{Unif}([0, 1])$. Denote $D = X - Y$.

(i) Find the distribution of D.

(ii) For any $d \in \mathbb{R}$ find the conditional distribution of X given $D = d$.

(iii) Compute $\mathbb{E}[X|D]$.

(iv) Check that $\text{Var}\left(\mathbb{E}\left[X|D\right]\right) \leq \text{Var}\left(X\right)$.

Exercice 17.9. Let X and Y be two independent real random variables with distribution $\mathbf{Exp}(\lambda)$, $\lambda > 0$ Denote $S = X + Y$.

(i) Find the distribution of S.

(ii) For any $s \in \mathbb{R}$ find the conditional distribution of X given $S = s$.

(iii) Compute $\mathbb{E}[X|S]$.

(iv) Check that $\text{Var}\left(\mathbb{E}\left[X|S\right]\right) \leq \text{Var}\left(X\right)$.

Exercise 17.10. Let X and Y be two random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. We assume that X is valued in $\mathbb{N}$ and Y follows a exponential distribution with parameter 1 on $\mathbb{R}$. In addition, we assume that the conditional distribution of X given $Y = y$ is the Poisson distribution with parameter y. Give the distribution of (X, Y) and the conditional distribution of Y given $X = x$.

17.2 Solutions

Solution to Exercise 17.4

Let $\mathcal{H}$ be a σ-field, $\mathcal{H} \subset \mathcal{F}$ and assume that $\sigma(X) \vee \mathcal{H}$ is independent of $\mathcal{G}$. We want to show that $\mathbb{E}[X|\mathcal{H} \vee \mathcal{G}] = \mathbb{E}[X|\mathcal{H}]$. By **??**-**??**, we just need to prove that for all $\mathsf{A} \in \mathcal{H} \vee \mathcal{G}$, we have

$$\mathbb{E}[\mathbb{1}_{\mathsf{A}} X] = \mathbb{E}[\mathbb{1}_{\mathsf{A}} \mathbb{E}[X|\mathcal{H}]] \ . \tag{173}$$

We first consider A of the form $\mathsf{B} \cap \mathsf{C}$, with $\mathsf{B} \in \mathcal{H}$ and $\mathsf{C} \in \mathcal{G}$. Indeed, using the assumption, for such measurable set, we get since $1_{\mathsf{B}} X$ is $\sigma(X) \vee \mathcal{H}$-measurable, $\mathbb{1}_{\mathsf{B}} \mathbb{E}[X|\mathcal{H}]$ is $\mathcal{H}$-measurable,

$$\mathbb{E}[\mathbb{1}_{\mathsf{A}} X] = \mathbb{E}[\mathbb{1}_{\mathsf{C}} \mathbb{1}_{\mathsf{B}} X] = \mathbb{E}[\mathbb{1}_{\mathsf{C}}]\mathbb{E}[\mathbb{1}_{\mathsf{B}} X] = \mathbb{E}[\mathbb{1}_{\mathsf{C}}]\mathbb{E}[\mathbb{1}_{\mathsf{B}} \mathbb{E}[X|\mathcal{H}]] = \mathbb{E}[\mathbb{1}_{\mathsf{A}} \mathbb{E}[X|\mathcal{H}]] \ . \tag{174}$$

Now consider $\mathcal{E} \subset \mathcal{F}$ and $\mathcal{C} \subset \mathcal{F}$ defined by

$$\mathcal{E} = \{\mathsf{A} \in \mathcal{F} \,:\, \mathbb{E}[\mathbb{1}_{\mathsf{A}} X] = \mathbb{E}[\mathbb{1}_{\mathsf{A}} \mathbb{E}[X|\mathcal{H}]]\} \ , \quad \mathcal{C} = \{\mathsf{B} \cap \mathsf{C} \,:\, \mathsf{B} \in \mathcal{H} \text{ and } \mathsf{C} \in \mathcal{G}\} \ . \tag{175}$$

By (174) we get that $\mathcal{C} \subset \mathcal{E}$. It is straightforward to check that $\mathcal{C}$ is stable by finite intersection, contains Ω and $\sigma(\mathcal{C}) = \mathcal{H} \vee \mathcal{G}$. Therefore it is a π-system. Then we just need to show that $\mathcal{E}$ is a λ-system since by the π-λ theorem, it will imply that $\sigma(\mathcal{C}) = \mathcal{H} \vee \mathcal{G} \subset \mathcal{E}$.

Let $\mathsf{A} \in \mathcal{E}$. Using that $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|\mathcal{H}]]$, we get that $\mathsf{A}^c \in \mathcal{E}$. Consider now a sequence $(\mathsf{A}_n)_{n\in\mathbb{N}} \in \mathcal{E}^{\mathbb{N}}$ such that for all $n < p$, $\mathsf{A}_n \cap \mathsf{A}_p = \emptyset$. Then for all $N \in \mathbb{N}$,

$$\mathbb{E}\left[\mathbb{1}_{\cup_{k=0}^N \mathsf{A}_k} X\right] = \sum_{n=1}^{N} \mathbb{E}[\mathbb{1}_{\mathsf{A}_k} X] = \sum_{n=1}^{N} \mathbb{E}[\mathbb{1}_{\mathsf{A}_k} \mathbb{E}[X|\mathcal{H}]] = \mathbb{E}\left[\mathbb{1}_{\cup_{k=0}^N \mathsf{A}_k} \mathbb{E}[X|\mathcal{H}]\right] \ . \tag{176}$$

Setting $\mathsf{A} = \cup_{k=0}^N \mathsf{A}_k$ and using the dominated convergence theorem, we get

$$\mathbb{E}[\mathbb{1}_{\mathsf{A}} X] = \mathbb{E}[\mathbb{1}_{\mathsf{A}} \mathbb{E}[X|\mathcal{H}]] \ . \tag{177}$$

Therefore $\mathsf{A} \in \mathcal{E}$ and $\mathcal{E}$ is a λ-system. Back to Exercise 17.4

Solution to Exercise 17.5

We first show the result when F is the identity. Namely, for all $\mathsf{A} \in \mathcal{B}(\mathbb{R}^{p+q})$ and all $\omega \in \Omega$, we prove

$$\mathbb{P}^{(Y,Z)|\mathcal{G}}(\omega, \mathsf{A}) = \mathbb{P}((Y(\omega), Z) \in \mathsf{A}) = \int_\Omega \mathbb{1}_{\mathsf{A}}(Y(\omega), Z)\mathbb{P}(\mathrm{d}\tilde{\omega}) \ . \tag{178}$$

Consider first A of the form $\mathsf{A} = \mathsf{B} \times \mathsf{C}$, with $\mathsf{B} \in \mathcal{B}(\mathbb{R}^p)$ and $\mathsf{C} \in \mathcal{B}(\mathbb{R}^q)$. Then for all $\mathsf{D} \in \mathcal{G}$, we have since $\mathbb{1}_{\mathsf{D}} Y$ is $\mathcal{G}$ measurable and Z is independent of $\mathcal{G}$

$$\begin{aligned}\mathbb{E}[\mathbb{1}_{\mathsf{D}} \mathbb{1}_{\mathsf{A}}(Y,Z)] &= \mathbb{E}[\mathbb{1}_{\mathsf{D}} \mathbb{1}_{\mathsf{B}}(Y)\mathbb{1}_{\mathsf{C}}(Z)] = \mathbb{E}[\mathbb{1}_{\mathsf{D}} \mathbb{1}_{\mathsf{B}}(Y)]\mathbb{E}[\mathbb{1}_{\mathsf{C}}(Z)] \\ &= \mathbb{E}[\mathbb{1}_{\mathsf{D}} \mathbb{1}_{\mathsf{B}}(Y)]\mathbb{P}(Z \in \mathsf{C}) = \mathbb{E}[\mathbb{1}_{\mathsf{D}} \mathbb{1}_{\mathsf{B}}(Y)\mathbb{P}(Z \in \mathsf{C})] \ .\end{aligned}$$

Therefore, we have almost surely

$$\mathbb{P}^{(Y,Z)|\mathcal{G}}(\omega, \mathsf{A}) = \mathbb{E}[\mathbb{1}_{\mathsf{A}}(Y,Z)|\mathcal{G}](\omega) = \mathbb{1}_{\mathsf{B}}(Y(\omega))\mathbb{P}(Z \in \mathsf{C}) = \int_\Omega \mathbb{1}_{\mathsf{A}}(Y(\omega), Z(\tilde{\omega}))\mathbb{P}(\mathrm{d}\tilde{\omega}) \ . \tag{179}$$

Consider now the two set $\mathcal{E}$ and $\mathcal{C}$ contained in $\mathcal{F} = \mathcal{B}(\mathbb{R}^{p+q})$ defined by

$$\mathcal{E} = \left\{\mathsf{A} \in \mathcal{F} \,:\, \mathbb{P}^{(Y,Z)|\mathcal{G}}(\omega, \mathsf{A}) = \int_\Omega \mathbb{1}_{\mathsf{A}}(Y(\omega), Z(\tilde{\omega}))\mathbb{P}(\mathrm{d}\tilde{\omega}) \ , \omega\text{-almost surely}\right\}$$

$$\mathcal{C} = \{\mathsf{B} \cap \mathsf{C} \,:\, \mathsf{B} \in \mathcal{B}(\mathbb{R}^p) \text{ and } \mathsf{C} \in \mathcal{B}(\mathbb{R}^q)\} \ .$$

By (179), we get that $\mathcal{C} \subset \mathcal{E}$. It is straightforward to check that $\mathcal{C}$ is stable by finite intersection, contains Ω and $\sigma(\mathcal{C}) = \mathcal{H} \vee \mathcal{G}$. Therefore it is a π-system. Then we just need to show that $\mathcal{E}$ is a λ-system since by the π-λ theorem, it will imply that $\sigma(\mathcal{C}) = \mathcal{H} \vee \mathcal{G} \subset \mathcal{E}$.

Let $\mathsf{A} \in \mathcal{E}$, A it is clear by definition that $\mathsf{A}^c \in \mathcal{E}$. Consider now a sequence $(\mathsf{A}_n)_{n\in\mathbb{N}} \in \mathcal{C}^{\mathbb{N}}$ such that for all $n < p$, $\mathsf{A}_n \cap \mathsf{A}_p = \emptyset$. Then by defition for all $N \in \mathbb{N}$, we have almost surely

$$\mathbb{P}^{(Y,Z)|\mathcal{G}}\left(\omega, \bigcup_{k=0}^{N} \mathsf{A}_k\right) = \int_\Omega \mathbb{1}_{\bigcup_{k=0}^{N} \mathsf{A}_k}(Y(\omega), Z(\tilde{\omega}))\mathbb{P}(\mathrm{d}\tilde{\omega}) \ . \tag{180}$$

Therefore almost surely for all $N \in \mathbb{N}$ (note the difference here), we get that (180) holds. Setting $\mathsf{A} = \cup_{k=0}^{N} \mathsf{A}_k$ and using the monotone convergence theorem, we get

$$\mathbb{P}^{(Y,Z)|\mathcal{G}}(\omega, \mathsf{A}) = \int_\Omega \mathbb{1}_{\mathsf{A}}(Y(\omega), Z(\tilde{\omega}))\mathbb{P}(\mathrm{d}\tilde{\omega}) \ . \tag{181}$$

Then $\mathsf{A} \in \mathcal{E}$ and $\mathcal{E}$ is a λ-system. So we have shown (178).

Let now $F : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}^m$ be a Borel function and $X = F(Y, Z)$. Then for all $\mathsf{A} \in \mathcal{B}(\mathbb{R}^m)$ and $\mathsf{B} \in \mathcal{G}$, we have

$$\begin{aligned}\mathbb{E}\left[\mathbb{1}_{\mathsf{B}}\mathbb{1}_{\mathsf{A}}(X)\right] = \mathbb{E}\left[\mathbb{1}_{\mathsf{B}}\mathbb{1}_{F^{-1}(\mathsf{A})}(Y, Z)\right] &= \int_\Omega \mathbb{1}_{\mathsf{B}}(\omega) \int_\Omega \mathbb{1}_{F^{-1}(\mathsf{A})}(Y(\omega), Z(\tilde{\omega}))\mathbb{P}(\mathrm{d}\tilde{\omega})\mathbb{P}(\mathrm{d}\omega)\\ &= \int_\Omega \mathbb{1}_{\mathsf{B}}(\omega) \int_\Omega \mathbb{1}_{\mathsf{A}}(F(Y(\omega), Z(\tilde{\omega})))\mathbb{P}(\mathrm{d}\tilde{\omega})\mathbb{P}(\mathrm{d}\omega) \ .\end{aligned}$$

Therefore, we get the expected result for all $\mathsf{A} \in \mathcal{B}(\mathbb{R}^m)$ almost surely:

$$\mathbb{P}^{X|\mathcal{G}}(\omega, \mathsf{A}) = \mathbb{P}\left(F(Y(\omega), Z) \in \mathsf{A}\right) \ .$$

Back to Exercise 17.5

Solution to Exercise 17.6

1. Note that $\mathrm{Span}\,(1, \mathbf{Y})$ is a finite dimensional linear subspace of H, hence is closed. So by the characterization of the orthogonal projection, $\widehat{\mathbf{X}} := \mathrm{proj}\,(\mathbf{X}|\,\mathrm{Span}\,(1, \mathbf{Y}))$ is given by

$$\widehat{\mathbf{X}} = \hat{b} + \hat{\mathbf{A}}(\mathbf{Y} - \mathbb{E}\left[\mathbf{Y}\right]) \ ,$$

with $\hat{b} \in \mathbb{R}^p$, $\hat{\mathbf{A}} \in \mathbb{R}^{p\times q}$ such that

$$\left\langle \mathbf{X} - (\hat{b} + \hat{\mathbf{A}}(\mathbf{Y} - \mathbb{E}\left[\mathbf{Y}\right]), b + \mathbf{A}(\mathbf{Y} - \mathbb{E}\left[\mathbf{Y}\right]) \right\rangle = 0 \text{ for all } b \in \mathbb{R}^p, \ \mathbf{A} \in \mathbb{R}^{p\times q} \ ,$$

which is equivalent to the two conditions

$$\left\langle \mathbf{X} - \hat{b}, b \right\rangle = 0, \left\langle \mathbf{X} - \hat{\mathbf{A}}(\mathbf{Y} - \mathbb{E}\left[\mathbf{Y}\right]), \mathbf{A}(\mathbf{Y} - \mathbb{E}\left[\mathbf{Y}\right]) \right\rangle = 0 \text{ for all } b \in \mathbb{R}^p, \ \mathbf{A} \in \mathbb{R}^{p\times q} \ .$$

This clearly yields $\hat{b} = \mathbb{E}\left[\mathbf{X}\right]$ for the first condition and since

$$\mathbb{E}\left[(\mathbf{X} - \hat{\mathbf{A}}(\mathbf{Y} - \mathbb{E}\left[\mathbf{Y}\right])^T\mathbf{A}(\mathbf{Y} - \mathbb{E}\left[\mathbf{Y}\right])\right] = \mathrm{Trace}\left(\mathbf{A}\mathbb{E}\left[(\mathbf{Y} - \mathbb{E}\left[\mathbf{Y}\right])(\mathbf{X} - \hat{\mathbf{A}}(\mathbf{Y} - \mathbb{E}\left[\mathbf{Y}\right])^T\right]\right) \ ,$$

the second condition gives

$$\mathrm{Cov}\left(\mathbf{Y}, \mathbf{X} - \hat{\mathbf{A}}\mathbf{Y}\right) = \mathbb{E}\left[(\mathbf{Y} - \mathbb{E}\left[\mathbf{Y}\right])(\mathbf{X} - \hat{\mathbf{A}}(\mathbf{Y} - \mathbb{E}\left[\mathbf{Y}\right])^T\right] = 0 \ .$$

which yields $\hat{\mathbf{A}} = \text{Cov}\,(\mathbf{Y}, \mathbf{X})\,\text{Cov}(\mathbf{Y})^{-1}$. Hence, as a result,

$$\widehat{\mathbf{X}} = \mathbb{E}\,[\mathbf{X}] + \text{Cov}\,(\mathbf{X}, \mathbf{Y})\,\text{Cov}(\mathbf{Y})^{-1}\,(\mathbf{Y} - \mathbb{E}\,[\mathbf{Y}])\;.$$

Now observe that

$$\begin{aligned}
\text{Cov}(\mathbf{X} - \widehat{\mathbf{X}}) &= \text{Cov}(\mathbf{X} - \text{Cov}\,(\mathbf{X}, \mathbf{Y})\,\text{Cov}(\mathbf{Y})^{-1}\mathbf{Y}) &\quad (182)\\
&= \text{Cov}\left(\mathbf{X} - \text{Cov}\,(\mathbf{X}, \mathbf{Y})\,\text{Cov}(\mathbf{Y})^{-1}\mathbf{Y}, \mathbf{X}\right) &\quad (183)\\
&= \text{Cov}(\mathbf{X}) - \text{Cov}\,(\mathbf{X}, \mathbf{Y})\,\text{Cov}(\mathbf{Y})^{-1}\text{Cov}\,(\mathbf{Y}, \mathbf{X})\;. &\quad (184)
\end{aligned}$$

Hence we have (i) of Proposition 17.1.

2. Let us write

$$\mathbf{X} = \widehat{\mathbf{X}} + \left(\mathbf{X} - \widehat{\mathbf{X}}\right)\;,$$

and observe that since $(\mathbf{X}, \mathbf{Y})$ is Gaussian, so is $(\mathbf{Y}, \left(\mathbf{X} - \widehat{\mathbf{X}}\right))$, which is obtained by a linear transform of it. Moreover since $\text{Cov}\left(\mathbf{Y}, \mathbf{X} - \widehat{\mathbf{X}}\right) = 0$ by definition of $\widehat{\mathbf{X}}$, they are independent. In the above decomposition, $\widehat{\mathbf{X}}$ is $\sigma(\mathbf{Y})$-measurable and $(\mathbf{X} - \widehat{\mathbf{X}})$ is independent of $\sigma(\mathbf{Y})$. This immediately gives

$$\mathbb{E}\,[\mathbf{X}|\,\mathbf{Y}] = \widehat{\mathbf{X}} = \text{proj}\,(\mathbf{X}|\,\text{Span}\,(1, \mathbf{Y}))\;,$$

that is, (ii) of Proposition 17.1. Moreover, for all $\omega \in \Omega$ and all $\mathsf{A} \in \mathcal{B}(\mathbb{R}^p)$,

$$\mathbb{P}^{\mathbf{X}|\mathbf{Y}}(\mathbf{Y}(\omega), \mathsf{A}) = \int \mathbb{1}_{\mathsf{A}}(\widehat{\mathbf{X}}(\omega) + \mathbf{X}(\omega') - \widehat{\mathbf{X}}(\omega'))\,\mathbb{P}(\mathrm{d}\omega')\;.$$

But since $\mathbf{X} - \widehat{\mathbf{X}}$ is a linear transform of $(\mathbf{X}, \mathbf{Y})$, it is a Gaussian vector, moreover it has mean 0, hence, for all $\omega \in \Omega$, the random vector $\omega' \mapsto \widehat{\mathbf{X}}(\omega) + \mathbf{X}(\omega') - \widehat{\mathbf{X}}(\omega')$ is $\text{N}(\widehat{\mathbf{X}}, \text{Cov}(\mathbf{X} - \widehat{\mathbf{X}}))$. Hence we obtain (iii) of Proposition 17.1.

Back to Exercise 17.6

Solution to Exercise 17.7

(i) For $s \in \mathbb{N}$, since $\{S = s\} = \bigcup_{x=0}^{s}\{X = x, Y = s - x\}$ with X and Y independent,

$$\mathbb{P}(S = s) = \sum_{x=0}^{s}\mathbb{P}(X = x)\mathbb{P}(Y = s - x) = \frac{e^{-(\lambda+\mu)}}{s!}\sum_{x=0}^{s}\binom{s}{x}\lambda^x\mu^{s-x} = \frac{e^{-(\lambda+\mu)}}{s!}(\lambda+\mu)^s. \quad (185)$$

Hence $S \sim \mathbf{Pn}(\lambda + \mu)$.

(ii) For $s \in \mathbb{N}$ and $x \in \{0, \ldots, s\}$,

$$\mathbb{P}(X = x|S = s) = \frac{\mathbb{P}(X = s, Y = s - x)}{\mathbb{P}(S = s)} = \binom{s}{x}\frac{\lambda^x\mu^{s-x}}{(\lambda+\mu)^s} \quad (186)$$

so that the conditional law of X given S is the binomial law with parameters $(s, \frac{\lambda}{\lambda+\mu})$.

(iii) The expectation of a binomial random variable being equal to the product of its parameters one deduces that $\mathbb{E}[X|S] = \frac{\lambda}{\lambda+\mu}S$.

(iv) The variance of a Poisson random variable being equal to its parameter,

$$\mathrm{Var}\left(\mathbb{E}\left[X|S\right]\right) = \left(\frac{\lambda}{\lambda+\mu}\right)^2 \mathrm{Var}\left(S\right) = \frac{\lambda^2}{\lambda+\mu} = \frac{\lambda}{\lambda+\mu}\mathrm{Var}\left(X\right) \le \mathrm{Var}\left(X\right).$$

Back to Exercise 17.7

Solution to Exercise 17.8

(i) Since X and Y are independent the density of (X,Y) is the product of their respective marginal densities. Hence, for $\varphi : \mathbb{R}^2 \to \mathbb{R}$ (measurable) bounded, using the change of variable $z = x - y$ in the integral over y, one has

$$\mathbb{E}\left[\varphi(X,D)\right] = \mathbb{E}\left[\varphi(X,X-Y)\right] = \int_{\mathbb{R}^2} \varphi(x, s-y) 1_{\{0\le x\le 1\}} 1_{\{0\le y\le 1\}} dxdy \tag{187}$$

$$= \int_{\mathbb{R}} 1_{\{0\le x\le 1\}} \left(-\int_x^{x-1} \varphi(x,z)dz \right) dx \tag{188}$$

$$= \int_{\mathbb{R}^2} 1_{\{\max(0,z)\le x\le \min(1,z+1)\}} \varphi(x,z) dxdz. \tag{189}$$

Hence the density of (X,D) is $p_{(X,D)}(x,d) = 1_{\{\max(0,d)\le x\le \min(1,d+1)\}}$ and the marginal density of D

$$p_D(d) = \int_{\mathbb{R}} 1_{\{\max(0,d)\le x\le \min(1,d+1)\}} dx = \begin{cases} 0 \text{ if } d \in (-\infty,-1] \cup [1,+\infty) \\ \int_0^{d+1} dx = d+1 \text{ if } d \in (-1,0] \\ \int_d^1 dx = 1-d \text{ if } d \in (0,1) \end{cases}. \tag{190}$$

(ii) For $d \in (-1,0]$ (resp. $d \in (0,1)$) the conditional density $\frac{p_{(X,D)}(x,d)}{p_D(d)}$ of X given $D = d$ is equal to $\frac{1_{\{0\le x\le d+1\}}}{d+1}$ (resp. $\frac{1_{\{d\le x\le 1\}}}{1-d}$) so that, conditionally on $D = d$, X is uniformly distributed on $[0, d+1]$ (resp. $[d,1]$). Since $\mathbb{P}(D \in (-\infty,-1] \cup [1,+\infty)) = 0$, it is not meaningfull to consider $d \in (-\infty,-1] \cup [1,+\infty)$.

(iii) The expectation of a random variable uniformly distributed on an interval being equal to the middle of the interval, we deduce that $\mathbb{E}[X|D] = \frac{D+1}{2}$.

(iv) One has, using that X and Y are i.i.d.,

$$\mathrm{Var}\left(\mathbb{E}\left[X|D\right]\right) = \mathrm{Var}\left(\frac{D+1}{2}\right) = \frac{1}{4}\mathrm{Var}\left(X+Y\right) = \frac{\mathrm{Var}\left(X\right)}{2} \le \mathrm{Var}\left(X\right).$$

Back to Exercise 17.8

Solution to Exercise 17.10

Let us find first the law of (X,Y). Let $A \subset \mathbb{N}$ and $B \in \mathcal{B}(\mathbb{R}_+)$. Then using that Y is $\sigma(Y)$-measurable and the conditional distribution of X given Y is a Poisson distribution with parameter Y we get by Fubini's theorem

$$\mathbb{P}\left(X\in A, Y\in B\right) = \mathbb{E}\left[\mathbb{1}_A(X)\mathbb{1}_B(Y)\right] = \mathbb{E}\left[\mathbb{1}_B(Y)\,\mathbb{E}\left[\mathbb{1}_A(X)|\,Y\right]\right] \tag{191}$$

$$= \mathbb{E}\left[\mathbb{1}_B(Y)\,\mathbb{P}^{X|Y}(A)\right] = \mathbb{E}\left[\mathbb{1}_B(Y)\,\mathbb{P}^{X|Y}(A)\right] = \mathbb{E}\left[\mathbb{1}_B(Y)\,\mathrm{e}^{-Y}\sum_{x\in A} Y^x/(x!)\right] \tag{192}$$

$$= \sum_{x\in A}\mathbb{E}\left[\mathbb{1}_B(Y)\,\mathrm{e}^{-Y}Y^x/\,(x!)\right] = \sum_{x\in A}\int_B (y^x/\,(x!))\mathrm{e}^{-2y}\mathrm{d}y\,. \tag{193}$$

Therefore the law of (X, Y) has density with respect to $\mu \otimes \lambda$

$$(x, y) \mapsto (y^x / (x!))e^{-2y} , \tag{194}$$

where μ is the counting measure on $\mathbb{N}$ and λ is the Lebesgue measure on $\mathbb{R}_+$.

If we take the marginal with respect to X, we get that the distribution of X has for density

$$x \mapsto 2^{-x-1} , \tag{195}$$

therefore X follows a geometric distribution with parameter 1/2.

Finally, the conditional density of Y given X is given by for all $y \geq 0$ and $x \in \mathbb{N}$ by

$$2^{x+1}(y^x / (x!))e^{-2y} . \tag{196}$$

We recognize the Gamma distribution with parameters $x + 1$ and 1/2. Besides the conditional expectation of Y given X is therefore

$$\mathbb{E}[Y | X] = (X + 1)/2 . \tag{197}$$

Back to Exercise 17.10