Week 8: Conditional expectation

Submission of solutions. Feedback can be given on Exercise 1 and any other exercise from the Training exercises. If you want to hand in, do it so by Monday 13/11/2023 17:00 (online) following the instructions on the course website

https://metaphor.ethz.ch/x/2023/hs/401-3601-00L/

Please pay attention to the quality, the precision and the presentation of your mathematical writing.

***

All random variables are assumed to be defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

1 Exercise covered during the exercise class

The following exercise will be covered during the exercise class.

Exercise 1. The goal of this exercise is to compute some conditional expectations (please justify your computions).

(1) Let $(X_i)_{1\leq i\leq n}$ be i.i.d. integrable random variables and set $S_n = X_1 + X_2 + \cdots + X_n$.
   (a) Compute $\mathbb{E}[S_n \mid X_1]$.
   (b) Compute $\mathbb{E}[X_1 \mid S_n]$.

(2) Let X and Y be two independent Bernoulli distributed random variables with parameter $p \in [0,1]$. Set $Z = \mathbb{1}_{X+Y=0}$.
   (a) Compute $\mathbb{E}[X \mid Z]$ and $\mathbb{E}[Y \mid Z]$.
   (b) Are these two random variables independent? Justify your answer.

Solution:

(1) (a) By linearity we can write

$$\mathbb{E}[S_n \mid X_1] = \sum_{i=1}^{n} \mathbb{E}[X_i \mid X_1] = X_1 + (n-1)\mathbb{E}[X_1],$$

since $\mathbb{E}[X_1 \mid X_1] = X_1$ and $\mathbb{E}[X_i \mid X_1] = \mathbb{E}[X_i] = \mathbb{E}[X_1]$ for $2 \leq i \leq n$ bcause X_i and X_1 are independent.

(b) Also by linearity,

$$S_n = \mathbb{E}[S_n \mid S_n] = \sum_{i=1}^{n} \mathbb{E}[X_i \mid S_n].$$

Since the random variables $(X_i)_{1\leq i\leq n}$ play symmetric roles, all the conditional expectations

$\mathbb{E}[X_i \mid S_n]$ are equal for $1 \leq i \leq n$, so

$$\mathbb{E}[X_1 \mid S_n] = \frac{1}{n} S_n.$$

To justify this precisely, we can argue as follows. First we claim that for $1 \leq i \neq j \leq n$, (X_i, S_n) and (X_j, S_n) have the same law. Indeed, set $F(X_1, \ldots, X_n) = (X_i, X_1 + \cdots + X_n)$ and let π the permutation on $\{1, 2, \ldots, n\}$ that exchanges i and j. The random variables $(X_1, \ldots, X_n)$ and $(X_{\pi(1)}, \ldots, X_{\pi(n)})$ have same law (the law n i.i.d. random variables with same law as X_1), so that $F(X_1, \ldots, X_n)$ and $F(X_{\pi(1)}, \ldots, X_{\pi(n)})$ have same law. But $F(X_1, \ldots, X_n) = (X_i, S_n)$ and $F(X_{\pi(1)}, \ldots, X_{\pi(n)}) = (X_j, S_n)$ which entails the claim. As a consequence, if $f : \mathbb{R} \to \mathbb{R}$ is a measurable bounded function, we get $\mathbb{E}[X_i f(S_n)] = \mathbb{E}\left[X_j f(S_n)\right]$, so by summing we obtain

$$\mathbb{E}[X_1 f(S_n)] = \mathbb{E}\left[\frac{S_n}{n} f(S_n)\right],$$

which shows that $\mathbb{E}[X_1 \mid S_n] = \frac{1}{n} S_n$ (indeed, recall that by the Doob-Dynkin lemma every random variable Z measurable with respect to $\sigma(S_n)$ is of the form $f(S_n)$ with f measurable).

(2) (a) Since Z is a random variable taking values 0 and 1, we know that

$$\mathbb{E}[X \mid Z] = f(Z)$$

with $f(0) = \mathbb{E}[X \mid Z = 0]$ and $f(1) = \mathbb{E}[X \mid Z = 1]$. We compute these two quantities.
– When $Z = 1$, we have $X = Y = 0$, so $f(1) = 0$.
– When $Z = 0$, we have

$$\mathbb{E}[X \mid Z = 1] = \frac{\mathbb{E}[X \mathbb{1}_{X+Y>0}]}{\mathbb{P}(X + Y > 0)}.$$

To compute this quantity, write $\mathbb{E}[X \mathbb{1}_{X+Y>0}] = \mathbb{P}(X = 1, Y = 0) + \mathbb{P}(X = 1, Y = 1) = \mathbb{P}(X = 1) = p$ and $\mathbb{P}(X + Y > 0) = 1 - \mathbb{P}(X + Y = 0) = 1 - \mathbb{P}(X = 0)\, Y = 0 = 1 - (1-p)^2$. Thus

$$\mathbb{E}[X \mid Z = 1] = \frac{p}{1 - (1-p)^2} = \frac{1}{2-p}.$$

We conclude that a.s.

$$\mathbb{E}[X \mid Z] = \frac{1}{2-p}(1 - Z)$$

The exact same reasoning gives $\mathbb{E}[Y \mid Z] = \frac{1}{2-p}(1 - Z)$.

(b) Observe that a bounded real-valued random variable is independent from itself if and only if it is a constant. Indeed, it is clearly the case if it is a constant. If X is a bounded random variable independent from itself, then $\mathbb{E}[X \cdot X] = \mathbb{E}[X] \cdot \mathbb{E}[X]$, so that $\mathbb{E}[X^2] = \mathbb{E}[X]^2$. Thus $\mathsf{Var}(X - \mathbb{E}[X]) = \mathbb{E}[(X - \mathbb{E}[X])^2] = 0$, so that the nonnegative random variable $(X - \mathbb{E}[X])^2$ has 0 expectation. It is thus almost surely 0, which gives the desired result.

Remark. One can show more generally that any real-valued random variable is independent from itself if and only if it is a constant.

Since the two random variables $\mathbb{E}[X \mid Z]$ and $\mathbb{E}[Y \mid Z]$ are equal, they are independent if and only if Z is a.s. constant, that is if and only if $p = 0$ or $p = 1$.

Remark. In particular, independence is in general not preserved when taking conditional expectations.

□

2 Training exercises

Exercise 2. Let X be an integrable random variable and $\mathcal{A} \subset \mathcal{F}$ a σ-field. Let Y be an integrable $\mathcal{A}$-measurable random variable.

(1) Show that $Y = \mathbb{E}[X \mid \mathcal{A}]$ if and only if for every $A \in \mathcal{A}$, $\mathbb{E}[X\mathbb{1}_A] = \mathbb{E}[Y\mathbb{1}_A]$.

(2) Let $\mathcal{C} \subset \mathcal{A}$ be a generating π-system of $\mathcal{A}$ containing Ω. Using the Dynkin Lemma, show that $Y = \mathbb{E}[X \mid \mathcal{A}]$ if and only if for every $A \in \mathcal{C}$, $\mathbb{E}[X\mathbb{1}_A] = \mathbb{E}[Y\mathbb{1}_A]$.

Solution:

(1) The implication $\Longrightarrow$ is immediate by the definition of the conditional expectation by taking $Z = \mathbb{1}_A$, which is bounded and $\mathcal{A}$-measurable.

For the implication $\Longleftarrow$, we have to show that for every bounded $\mathcal{A}$-measurable random variable Z:

$$\mathbb{E}[ZX] = \mathbb{E}[ZY].$$

By linearity, it is enough to check that for every bounded $\mathcal{A}$-measurable nonnegative random variable Z:

$$\mathbb{E}[ZX] = \mathbb{E}[ZY].$$

To this end, we take a sequence $0 \leq Z_n \uparrow Z$ of simple functions. By assumption and linearity,

$$\mathbb{E}[Z_n X] = \mathbb{E}[Z_n Y].$$

We conclude that

$$\mathbb{E}[ZX] = \mathbb{E}[ZY]$$

by dominated convergence. Indeed, $Z_n X \to ZX$ with $|Z_n X| \leq |Z||X|$ integrable, and $Z_n Y \to ZY$ with $|Z_n Y| \leq |Z||Y|$ integrable.

(2) Similarly, the implication $\Longrightarrow$ is immediate. For the reverse implication, set

$$\mathcal{B} = \{A \in \mathcal{A} : \mathbb{E}[X\mathbb{1}_A] = \mathbb{E}[Y\mathbb{1}_A]\}.$$

We shall show that $\mathcal{B} = \mathcal{A}$ using the Dynkin Lemma, and the result will follow by (1).

To this end, we check that $\mathcal{B}$ is a Dynkin system:

- Since $\Omega \in \mathcal{C}$ we have $\Omega \in \mathcal{B}$.

- If $A \in \mathcal{B}$, then $\mathbb{E}[X\mathbb{1}_{A^c}] = 1 - \mathbb{E}[X\mathbb{1}_A] = 1 - \mathbb{E}[Y\mathbb{1}_A] = \mathbb{E}[Y\mathbb{1}_{A^c}]$ so that $A^c \in \mathcal{B}$.
- Les $(A_i)_{i\geq 1}$ be a pairwise disjoint sequence of elements of $\mathcal{B}$. Set $A = \bigcup_{i\geq 1} A_i$. Then, by the Fubini-Lebesgue theorem:

$$\mathbb{E}[X\mathbb{1}_A] = \mathbb{E}\left[\sum_{i=1}^{\infty} X\mathbb{1}_{A_i}\right] = \sum_{i=1}^{\infty}\mathbb{E}\left[X\mathbb{1}_{A_i}\right].$$

  Indeed, we have

$$\mathbb{E}\left[\sum_{i=1}^{\infty} |X\mathbb{1}_{A_i}|\right] = \mathbb{E}[|X|] < \infty.$$

  Thus

$$\mathbb{E}[X\mathbb{1}_A] = \sum_{i=1}^{\infty}\mathbb{E}\left[X\mathbb{1}_{A_i}\right] = \sum_{i=1}^{\infty}\mathbb{E}\left[Y\mathbb{1}_{A_i}\right] = \mathbb{E}\left[\sum_{i=1}^{\infty} Y\mathbb{1}_{A_i}\right] = \mathbb{E}[Y\mathbb{1}_A]$$

  so $A \in \mathcal{B}$.

Thus $\mathcal{B}$ contains the Dynkin generated by $\mathcal{C}$, which is the σ-field generated by $\mathcal{C}$ by the Dynkin Lemma. The latter σ-field is $\mathcal{A}$, which completes the proof.

□

Exercise 3. We consider a population in which there is a large number n of households. We model the size of the households by i.i.d. random variables $(X_i)_{1\leq i\leq n}$ on $\mathbb{N}^*$, with mean $m = \mathbb{E}[X_1] = \sum_{k\geq 1} kp_k < \infty$ where $p_k = \mathbb{P}(X_1 = k)$. Let T_n be the size of the household of an individual chosen uniformly at random in the population.

(1) Justify that for every integer $k \geq 1$ we have

$$\mathbb{P}(T_n = k \mid X_1, \ldots, X_n) = \frac{k\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{\{X_i=k\}}}{\frac{1}{n}\sum_{i=1}^{n}X_i}.$$

(2) Show that for every integer $k \geq 1$, $\mathbb{P}(T_n = k)$ converges to $\frac{k}{m}p_k$ as $n \to \infty$.

Solution:

The population has $X_1 + \cdots + X_n$ individuals. We model the choice of an individual at random by the choice, conditionally given $X_1, \ldots, X_n$, of an integer following the uniform distribution on the set $\{1, 2, \ldots, X_1 + \cdots + X_n\}$. For every $k \geq 1$, there are $N_k := \mathbb{1}_{\{X_1=k\}} + \cdots + \mathbb{1}_{\{X_n=k\}}$ households of size k which have in total kN_k individuals. Therefore, by definition of the uniform distribution:

$$\mathbb{P}(T_n = k \mid X_1, \ldots, X_n) = \frac{k\sum_{i=1}^{n}\mathbb{1}_{\{X_i=k\}}}{\sum_{i=1}^{n}X_i} = \frac{k\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{\{X_i=k\}}}{\frac{1}{n}\sum_{i=1}^{n}X_i}.$$

Hence, by the law of large numbers,

$$\mathbb{P}(T_n = k \mid X_1, \ldots, X_n) \quad \xrightarrow[n\to\infty]{a.s.} \quad \frac{k}{m}p_k.$$

Set $F(X_1, \ldots, X_n) = \mathbb{P}(T_n = k \mid X_1, \ldots, X_n)$, so that $\mathbb{P}(T = k) = \mathbb{E}[F(X_1, \ldots, X_n)]$ (more generally, recall that

$\mathbb{E}[U] = \mathbb{E}[\mathbb{E}[U \mid V]]$ for random variables U, V). We have just seen that $F(X_1, \ldots, X_n)$ converges almost surely to $\frac{k}{m} p_k$, and $F(X_1, \ldots, X_n)$, being a conditional probability, is dominated by 1. Hence, by the dominated convergence theorem,

$$\mathbb{P}(T_n = k) = \mathbb{E}[F(X_1, \ldots, X_n)] \quad \underset{n \to \infty}{\longrightarrow} \quad \frac{k}{m} p_k.$$

Interpretation. The limiting law of T_n is not the initial law, because large households are over-represented and small households are under-represented. This phenomenon is called *size-bias*. This is probably the most common sampling bias. The bias is all the more important as the size of the household differs from the average size m. □

Exercise 4. Let X, Y be two random variables with values in respectively E, F. Let $\mathcal{A}$ be a sub σ-field of $\mathcal{F}$. Assume that X is independent from $\mathcal{A}$ and that Y is $\mathcal{A}$-measurable. Show that for any measurable function $g : E \times F \to \mathbb{R}^+$, we have

$$\mathbb{E}[g(X, Y) \mid \mathcal{A}] = h(Y) \qquad \text{a.s.,} \qquad \text{where} \qquad h(y) = \mathbb{E}[g(X, y)].$$

Remark. In particular, this applies to $\mathbb{E}[g(X, Y) \mid Y]$ when X and Y are independent. Intuitively speaking, in this case, the conditional expectation is done by first computing the expectation "with respect to X" (i.e. by integrating with respect to the law of X) by considering Y as being "fixed", and then computing the expectation of the result "with respect to Y" (i.e. by integrating with respect to the law of Y).

Solution:

Since $h(Y)$ is $\sigma(Y)$-measurable, hence $\mathcal{A}$-measurable, to show that $\mathbb{E}[g(X, Y) \mid \mathcal{A}] = h(Y)$ almost surely, it suffices to check that for every non-negative random variable Z which is $\mathcal{A}$ measurable, we have

$$\mathbb{E}[g(X, Y)Z] = \mathbb{E}[h(Y)Z].$$

To this end, write, using Fubini's theorem,

$$\begin{aligned}
\mathbb{E}[g(X, Y)Z] &= \int_{E \times F \times \mathbb{R}^+} g(x, y) z \mathbb{P}_{(X,Y,Z)}(\mathrm{d}x \mathrm{d}y \mathrm{d}z) \\
&= \int_{E \times F \times \mathbb{R}^+} g(x, y) z \mathbb{P}_X(\mathrm{d}x) \mathbb{P}_{(Y,Z)}(\mathrm{d}y \mathrm{d}z) \quad \text{(because } X \text{ and } (Y, Z) \text{ are independent)} \\
&= \int_{F \times \mathbb{R}_+} z \left(\int_E g(x, y) \mathbb{P}_X(\mathrm{d}x) \right) \mathbb{P}_{(Y,Z)}(\mathrm{d}y \mathrm{d}z) \\
&= \int_{F \times \mathbb{R}_+} z h(y) \mathbb{P}_{(Y,Z)}(\mathrm{d}y \mathrm{d}z) \\
&= \mathbb{E}[h(Y)Z],
\end{aligned}$$

where we have used the fact that $h(y) = \int_E g(x, y) \mathbb{P}_X(\mathrm{d}x)$. □

3 More involved exercises (optional, will not be covered in the exercise class)

Exercise 5. Let $\mathcal{A} \subset \mathcal{F}$ be a σ-field, and X a nonnegative random variable. Show that $\{\mathbb{E}[X \mid \mathcal{A}] > 0\}$ is the smallest $\mathcal{A}$-measurable set (up to negligeable sets, that is events with 0 probability) containing $\{X > 0\}$.

Solution:
The random variable $\mathbb{E}[X \mid \mathcal{A}]$ is by definition $\mathcal{A}$-measurable, and $(0, +\infty)$ is a Borel set, hence $\{\mathbb{E}[X \mid \mathcal{A}] > 0\}$ is $\mathcal{A}$-mesurable. In addition, by definition of the conditional expectation,

$$\mathbb{E}\left[X \mathbb{1}_{\{\mathbb{E}[X|\mathcal{A}]=0\}}\right] = \mathbb{E}\left[\mathbb{E}[X|\mathcal{A}] \mathbb{1}_{\{\mathbb{E}[X|\mathcal{A})=0\}]}\right] = 0.$$

But $X\mathbb{1}_{\{\mathbb{E}[X|\mathcal{A}]=0\}} \geq 0$ almost surely, so $X\mathbb{1}_{\{\mathbb{E}[X|\mathcal{A}]=0\}} = 0$ almost surely. This means that

$$\{X > 0\} \subset \{\mathbb{E}[X \mid \mathcal{A}] > 0\}$$

up to a negligible set.

Let A be a $\mathcal{A}$-measurable set containing $\{X > 0\}$, i.e., on A^c we have $X = 0$ almost surely. Then, again by the definition of conditional expectation, we have

$$\mathbb{E}[\mathbb{E}[X \mid \mathcal{A}] \mathbb{1}_{A^c}] = \mathbb{E}[X \mathbb{1}_{A^c}] = 0.$$

Similarly $\mathbb{E}[X \mid \mathcal{A}] \geq 0$ hence on A^c we have $\mathbb{E}[X \mid \mathcal{A}] = 0$ almost surely, that is $\{\mathbb{E}[X \mid \mathcal{A}] > 0\} \subset A$ up to a negligible set. □

Exercise 6. Let $(X_i)_{i\geq 1}$ be a sequence of nonnegative real random variables and $(\mathcal{F}_i)_{i\geq 1}$ a sequence of σ-fields included in $\mathcal{F}$. Assume that $\mathbb{E}[X_i \mid \mathcal{F}_i]$ converges in probability to 0.

(1) Show that $(X_i)_{i\geq 1}$ converges in probability to 0.
(2) Show that the converse is false.

Solution:

(1) Fix $\varepsilon > 0$. For $i \geq 1$, denote by A_i the event $\{\mathbb{E}[X_i \mid \mathcal{F}_i] > \varepsilon^2/2\}$. Then, by assumption, $\mathbb{P}(A_i) \to 0$ as $i \to \infty$, hence $\mathbb{P}(A_i) \leq \varepsilon/2$ for i sufficiently large. In addition,

$$\mathbb{E}\left[X_i \mathbb{1}_{A_i^c}\right] = \mathbb{E}\left[\mathbb{E}[X_i \mid \mathcal{F}_i] \mathbb{1}_{A_i^c}\right] \leq \varepsilon^2/2,$$

so by Markov's inequality

$$\mathbb{P}\left(X_i \geq \varepsilon \text{ and } A_i^c\right) \leq \frac{1}{\varepsilon}\mathbb{E}\left[X_i \mathbb{1}_{A_i^c}\right] \leq \frac{\varepsilon}{2},$$

so that

$$\mathbb{P}(X_i > \varepsilon) < \mathbb{P}(A_i) + \mathbb{P}\left(X_i > \varepsilon \text{ and } A_i^c\right) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Alternative solution. Note that $X_i \to 0$ in probability if and only if $\max(X_i, 1) \to 0$ in probability. The random variables $\mathbb{E}[\max(X_i, 1) \mid \mathcal{F}_i]$ are bounded by 1 and tend to 0 in probability, so they tend to 0 in $\mathbb{L}^1$. Hence

$$\mathbb{E}[\max(X_i, 1)] = \mathbb{E}[\mathbb{E}[\max(X_i, 1) \mid \mathcal{F}_i]] \underset{n\to\infty}{\longrightarrow} 0.$$

We conclude that $\max(X_i, 1) \to 0$ in $\mathbb{L}^1$, and also in probability.

(2) It suffices to take $\mathcal{F}_i = \{\varnothing, \Omega\}$ and (X_i) a sequence converging in probability to 0 but not in $\mathbb{L}^1$ to 0.

□

Exercise 7. Let $\mathcal{A} \subset \mathcal{F}$ be a σ-field. We say that two random variables X and Y are independent conditionally given $\mathcal{A}$ if for every nonnegative measurable functions $f, g : \mathbb{R} \to \mathbb{R}_+$ we have

$$\mathbb{E}[f(X)g(Y) \mid \mathcal{A}] = \mathbb{E}[f(X) \mid \mathcal{A}]\mathbb{E}[g(Y) \mid \mathcal{A}]. \tag{1}$$

(1) What does this mean if $\mathcal{A} = \{\emptyset, \Omega\}$? If $\mathcal{A} = \mathcal{F}$?

(2) Show that the previous definition (1) is equivalent to (a) and that (1) is equivalent to (b):

(a) for every nonnegative $\mathcal{A}$-measurable random variable Z, for every nonnegative measurable functions $f, g : \mathbb{R} \to \mathbb{R}_+$, we have $\mathbb{E}[f(X)g(Y)Z] = \mathbb{E}[f(X)Z\mathbb{E}[g(Y) \mid \mathcal{A}]]$,

(b) for every nonnegative measurable function $g : \mathbb{R} \to \mathbb{R}_+$ we have $\mathbb{E}[g(Y) \mid \sigma(\mathcal{A}, \sigma(X))] = \mathbb{E}[g(Y) \mid \mathcal{A}]$.

Solution:

1. For $\mathcal{A} = \{\emptyset, \Omega\}$, we get $\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)]$ for every nonnegative measurable functions $f, g : \mathbb{R} \to \mathbb{R}_+$, which is equivalent to saying that X and Y are independent.

   If $\mathcal{A} = \mathcal{F}$, the equality is always true and we cannot say anything.

2. (a) Assume that for every nonnegative measurable functions $f, g : \mathbb{R} \to \mathbb{R}_+$, we have

$$\mathbb{E}[f(X)g(Y) \mid \mathcal{A}] = \mathbb{E}[f(X) \mid \mathcal{A}]\mathbb{E}[g(Y) \mid \mathcal{A}].$$

Let Z be a nonnegative $\mathcal{A}$-measurable random variable. Then

$$\begin{aligned} \mathbb{E}[f(X)g(Y)Z] &= \mathbb{E}[\mathbb{E}[f(X)g(Y) \mid \mathcal{A}]Z] \\ &= \mathbb{E}[\mathbb{E}[f(X) \mid \mathcal{A}]\mathbb{E}[g(Y) \mid \mathcal{A}]Z] \\ &= \mathbb{E}[f(X)\mathbb{E}[g(Y) \mid \mathcal{A}]Z] \end{aligned}$$

by the characteristic property of conditional expectations using the fact that $\mathbb{E}[g(Y) \mid \mathcal{A}]Z$ is a bounded non-negative $\mathcal{A}$-measurable random variable.

Now, conversely, assume that for every nonnegative $\mathcal{A}$-measurable random variable Z, for every nonnegative measurable functions $f, g : \mathbb{R} \to \mathbb{R}$, we have

$$\mathbb{E}[f(X)g(Y)Z] = \mathbb{E}[f(X)\mathbb{E}[g(Y) \mid \mathcal{A}]Z].$$

Since $Z\mathbb{E}[g(Y) \mid \mathcal{A}]$ is $\mathcal{A}$-mesurable, we have

$$\mathbb{E}[f(X)g(Y)Z] = \mathbb{E}[f(X)\mathbb{E}[g(Y) \mid \mathcal{A}]Z] = \mathbb{E}[\mathbb{E}[f(X) \mid \mathcal{A}]\mathbb{E}[g(Y) \mid \mathcal{A}]Z],$$

and by the definition of conditional expectation, since $\mathbb{E}[f(X) \mid \mathcal{A}]\mathbb{E}[g(Y) \mid \mathcal{A}]$ is $\mathcal{A}$-measurable, we get

$$\mathbb{E}[f(X) \mid \mathcal{A}]\mathbb{E}[g(Y) \mid \mathcal{A}] = \mathbb{E}[f(X)g(Y) \mid \mathcal{A}].$$

(b) Assume that for every nonnegative measurable function $g : \mathbb{R} \to \mathbb{R}_+$ we have

$$\mathbb{E}[g(Y) \mid \sigma(\sigma(X), \mathcal{A})] = \mathbb{E}[g(Y) \mid \mathcal{A}].$$

Then, for every nonnegative $\mathcal{A}$-measurable random variable Z, for every nonnegative measurable functions $f, g : \mathbb{R} \to \mathbb{R}_+$,

$$\begin{aligned} \mathbb{E}[f(X)Zg(Y)] &= \mathbb{E}[f(X)Z\mathbb{E}[g(Y) \mid \sigma(\sigma(X), \mathcal{A})]] \\ &= \mathbb{E}[f(X)Z\mathbb{E}[g(Y) \mid \mathcal{A}]] \end{aligned}$$

By a), this shows that X and Y are independent conditionally given $\mathcal{A}$.

Conversely, let $(\mathbb{H}, \mathcal{H})$ be the space in which X takes it values. The collection Π of subsets of Ω of the form $X^{-1}(H) \cap G$ with $H \in \mathcal{H}$ and $G \in \mathcal{A}$ generates $\sigma(\mathcal{A}, \sigma(X))$ and is stable by finite intersections. Using (a), assume that for every nonnegative $\mathcal{A}$-measurable random variable Z, for every nonnegative measurable functions $f, g : \mathbb{R} \to \mathbb{R}_+$, we have

$$\mathbb{E}[f(X)g(Y)Z] = \mathbb{E}[f(X)\mathbb{E}[g(Y) \mid \mathcal{A}]Z].$$

In particular, for every $H \in \mathcal{H}$ and $G \in \mathcal{A}$, we have

$$\mathbb{E}\Big[g(Y)\mathbb{1}_{X^{-1}(H)}\mathbb{1}_G\Big] = \mathbb{E}\Big[\mathbb{E}[g(Y) \mid \mathcal{A}]\,\mathbb{1}_{X^{-1}(H)}\mathbb{1}_G\Big].$$

By the Dynkin Lemma, it follows that for every $A \in \sigma(\sigma(X), \mathcal{A})$, we have

$$\mathbb{E}[g(Y)\mathbb{1}_A] = \mathbb{E}[\mathbb{E}[g(Y) \mid \mathcal{A}]\,\mathbb{1}_A].$$

By an application of the monotone convergence theorem, it follows that

$$\mathbb{E}[g(Y)Z] = \mathbb{E}[\mathbb{E}[g(Y) \mid \mathcal{A}]Z]$$

for every nonnegative $\sigma(\sigma(X), \mathcal{A})$-measurable random variable Z. Since $\mathbb{E}[g(Y) \mid \mathcal{A}]$ is $\sigma(\sigma(X), \mathcal{A})$-measurable, we conclude that

$$\mathbb{E}[g(Y) \mid \sigma(\sigma(X), \mathcal{A})] = \mathbb{E}[g(Y) \mid \mathcal{A}].$$

□

4 Fun exercise (optional, will not be covered in the exercise class)

Exercise 8. Imagine there are a 100 people in line to board a plane that seats 100. The first person in line, Alice, realizes she lost her boarding pass, so when she boards she decides to take a random seat instead. Every person that boards the plane after her will either take their "proper" seat, or if that seat is taken, a random seat instead.

What is the probability that the last person that boards will end up in their proper seat?

Solution:

The answer is 1/2!

Suppose whenever someone finds their seat taken, they politely evict the squatter and take their seat. In this case, the first passenger (Alice, who lost her boarding pass) keeps getting evicted (and choosing a new random seat) until, by the time everyone else has boarded, she has been forced by a process of elimination into her correct seat.

This process is the same as the original process except for the identities of the people in the seats, so the probability of the last boarder finding their seat occupied is the same.

When the last boarder boards, Alice is either in her own seat or in the last boarder's seat, which have both looked exactly the same (i.e. empty) to her up to now, so there is no way poor Alice could be more likely to choose one than the other. □