E 703: Advanced Econometrics I
Solutions to Problem Set 1

Niccolò Lomys

October 13, 2016

Logistics

Before we start, here are some useful information.

Tutorials

- When: Thursdays, 13:45-15:15 and 15:30-17:00.
- Where: B6, 23-25, A3.02.

Niccolò Lomys

- Email: `niccolo.lomys@gess.uni-mannheim.de`.
- Office: L9, 7, 3rd floor, room 304.
- Office hours: Any time I am in the office.

Textbooks

There are many excellent textbooks that introduce probability theory from the measure-theoretic viewpoint. Some of them are already referenced in the Lecture Notes. I would add the following (and, of course, many other excellent treatments exist):

- Billingsley (1995), *Probability and Measure*. Wiley.
- Capiński and Kopp (2004), *Measure, Integral and Probability*. Springer.
- Gut (2005), *Probability: A Graduate Course*. Springer.
- Lévêque (2015), *Lecture Notes in Advanced Probability*. Manuscript available here.
- Ok (2016), *Probability with Economic Applications*. Manuscript available here.
- Pollard (2002), *A User's Guide to Measure Theoretic Probability*. Cambridge University Press.
- Rosenthal (2006), *A First Look at Rigorous Probability Theory*. World Scientific.

I personally enjoy the treatment of Ok (2016) very much. The manuscript is advanced, but accessible, crystal clear (as so is Ok's book on real analysis), and contains all the relevant real analysis background. In style, it particularly fits the needs of the economic theory student. Gut (2005) develops the theory with the ultimate aim of making a proper introduction to mathematical statistics, and proves results in good detail. Rosenthal (2006) is concise, accessible and well-written, with clearly proven results. A solution manual with answers to all even-numbered problems exists for this book.

Lévêque (2015) is roughly at the same level of our course. Lévêque's notes are short (but not simplistic) and accurate (but without excessive technicalities), and so they are a good read. Capiński and Kopp (2004) is particularly useful if you lack a solid real analysis background. It has solutions to all problems and good intuitive explanations. Billingsley (1995) is encyclopedic, but I do not find it too "user friendly" or particularly suited for self-study. Pollard (2002) offers an interesting and deep point of view, but the treatment is a bit idiosyncratic. I would not recommend it as the main reference unless you are taking a probability course with David Pollard.

For the mathematical statistics and the introduction to the linear regression model that we cover in this course, I find the following references useful (in addition to those listed in the Lecture Notes):

– Casella and Berger (2001), *Statistical Inference*. Duxbury.

– Hansen (2016), *Econometrics*. Manuscript available here.

– Rohatgi and Saleh (2001), *An Introduction to Probability and Statistics*. Wiley.

Grading Policy for Problem Sets

Each problem set will contain six exercises and will be graded out of 100 points. Two exercises will be starred, while the others will not. The two starred exercises will account for 60 points and will be graded (almost) as carefully as your exam will be. This means that you will receive an accurate feedback on them, so that you can get an idea about what we expect from you at the exam. However, this also means that points will be cut any time a step of your solution is not sufficiently motivated or your argument is loose. While I will be moderately tolerant at the beginning of the term, I will become increasingly less so over time, so that you can adjust before the final exam. The remaining four exercises will account for 10 points each and will be graded in a coarse way: 10 points if you provide a satisfactory answer with only minor flaws; 6 points if your solution contains one or more major problems; 2 points if you barely attempt to solve the exercise; 0 points if you do not answer at all.

Preliminaries

Throughout our exercise sessions, we will stick to the following conventions.

1. A *set* is a collection of objects we call elements. A *class* is a set of sets, and a *family* is a set of classes. Please, try to be consistent.

2. The set of *natural numbers* is $\mathbb{N} := \{1, 2, \dots\}$, i.e., we exclude zero.

3. We say that two nonempty sets A and B in some universal set U are *numerically equivalent*, or that A and B have the same *cardinality*, or that A and B have the same *cardinal number*, if there exists a bijection $f \colon A \to B$. In this case, we write $A \sim_{\text{card}} B$.

   Let $\mathbb{N}$ be the set of natural numbers, and set $\mathbb{N}_n := \{1, 2, 3, \dots, n\}$ for any $n \in \mathbb{N}$. For any $A \subseteq U$ we say:

   (a) A is *finite* if $A \sim_{\text{card}} \mathbb{N}_n$ for some natural number n;
   (b) A is *infinite* if A is not finite;
   (c) A is *countably infinite* if $A \sim_{\text{card}} \mathbb{N}$;
   (d) A is *countable* if A is either finite or countably infinite;
   (e) A is *uncountable* if A is not countable.

   The empty set is considered to be finite and its cardinal number is zero.

4. The symbols $\bigcup_{i=1}^{\infty}$ and $\bigcup_{i \in \mathbb{N}}$ are used interchangeably to denoted countably infinite unions. An analogous observation applies to countably infinite intersections.

5. The terms "algebra" and "field," as well as "σ-algebra" and "σ-field," are used interchangeably.

Algebras and σ-Algebras

Definition 1. *Let Ω be a nonempty set. A nonempty class $\mathcal{A}^*$ of subsets of Ω is called an **algebra** (or **field**) on Ω if*

(i) $\emptyset \in \mathcal{A}^*$*;*

(ii) $\Omega \setminus A \in \mathcal{A}^*$ *for all* $A \in \mathcal{A}^*$*;*

(iii) $A \cup B \in \mathcal{A}$ *for all* $A, B \in \mathcal{A}^*$*.*

*We say that $\mathcal{A}^*$ is a **finite algebra** on Ω if it is an algebra on Ω such that $|\mathcal{A}^*| < \infty$.*

In words, an algebra on a nonempty set Ω is a nonempty class of subsets of Ω that has the empty set as one of its elements and is closed under complementation and taking pairwise (and thus finite[1]) unions. A finite algebra on Ω is one that contains finitely many elements. Clearly, any algebra on a finite set is a finite algebra on that set.

Definition 2. *Let Ω be a nonempty set. A nonempty class $\mathcal{A}$ of subsets of Ω is called a **σ-algebra** (or **σ-field**) on Ω if*

(i) $\emptyset \in \mathcal{A}$*;*

(ii) $\Omega \setminus A \in \mathcal{A}$ *for all* $A \in \mathcal{A}$*;*

[1] *Quiz:* Why? In particular, why finite and not countably infinite unions?

(iii) $\bigcup_{n=1}^{\infty} \in \mathcal{A}$ *whenever* $A_n \in \mathcal{A}$ *for each* $n = 1, 2, \ldots$.

Any element of $\mathcal{A}$ *is called an* ***$\mathcal{A}$-measurable set*** *in* Ω. *If* $\mathcal{A}$ *is a* σ*-algebra on* Ω, *we refer to the pair* $(\Omega, \mathcal{A})$ *as a* ***measurable space***.

That is, a σ-algebra on Ω is a nonempty class of subsets of Ω that has the empty set as one of its elements and is closed under complementation and taking countably infinite unions. Note that there is no difference between an algebra and a σ-algebra when the ground set Ω under consideration is finite.[2]

Interpretation. We interpret the fundamental set Ω as the set of all possible outcomes (or realizations) ω of a given experiment whose result is uncertain. We call Ω the *sample space*. Given the sample space Ω, it is important to describe what information one has on the outcomes of the experiment. This notion of information is captured by the mathematical notion of σ-algebra. The sets belonging to a σ-algebra $\mathcal{A}$ on Ω are the events that one can decide on whether they happened or not, given the information $\mathcal{A}$. That is, A is an event if and only if $A \in \mathcal{A}$. If one knows the information $\mathcal{A}$, then one is able to tell which events of $\mathcal{A}$ (= subsets of Ω) the realization of the experiment ω belongs to. One may define many different σ-algebras on a given sample space, so what an "event" really is depends on the model one chooses to work with. We cannot, however, be completely arbitrary when specifying a model because the notion of σ-algebras imposes some restrictions. First, we need to be able to say that nothing happens, which requires $\emptyset \in \mathcal{A}$. Second, if A is an event, then we need to be able to talk about this event not occurring, that is, to deem the set $\Omega \setminus A$ also as an event. This requires $\mathcal{A}$ be closed under complementation. Finally, we wish to be able to talk about at least one of countably many events occurring, and this requires $\mathcal{A}$ be closed under taking countable unions. Besides, the last two properties warrant that we can view "countably many events occurring simultaneously" as an event as well.

Exercise 1 (Some Basic Properties of σ-Algebras)

Let $\mathcal{A}$ be a σ-algebra on a nonempty set Ω. Show that:

(a) $\Omega \in \mathcal{A}$;

(b) If $A_1, \ldots, A_n \in \mathcal{A}$ for some $n \in \mathbb{N}$, then $\bigcap_{k=1}^{n} A_k \in \mathcal{A}$;

(c) If $A, B \in \mathcal{A}$, then $A \setminus B \in \mathcal{A}$.

Solution

For any $A \subseteq \Omega$, define $A^c := \Omega \setminus A$.

(a) Since $\mathcal{A}$ is a σ-algebra on Ω, $\emptyset \in \mathcal{A}$ (by property (i) of σ-algebras) and $\emptyset^c \in \mathcal{A}$ (by property (ii) of σ-algebras). Observing that $\emptyset^c = \Omega$ completes the proof. ■

(b) We first show that $\mathcal{A}$ is closed under finite unions. That is,

$$\langle A_1, \ldots, A_n \in \mathcal{A} \text{ for some } n \in \mathbb{N} \rangle \Longrightarrow \bigcup_{k=1}^{n} A_k \in \mathcal{A}. \tag{1}$$

Let $A_1, \ldots, A_n \in \mathcal{A}$. For any natural number $m > n$, define $A_m := \emptyset$. By property (i) of σ-algebras, $A_m \in \mathcal{A}$ for any such m. By property (iii) of σ-algebras, $\bigcup_{k=1}^{\infty} A_k \in \mathcal{A}$. Since

$$\bigcup_{k=1}^{\infty} A_k = \left(\bigcup_{k=1}^{n} A_k \right) \cup \left(\bigcup_{l=n+1}^{\infty} A_l \right) = \bigcup_{k=1}^{n} A_k \cup \emptyset = \bigcup_{k=1}^{n} A_k,$$

[2] *Quiz:* Is there any redundant property in our definitions of algebra and σ-algebras? If so, which one and why?

the claim follows.

We now use the previous claim to show that

$$\langle A_1, \dots, A_n \in \mathcal{A} \text{ for some } n \in \mathbb{N} \rangle \Longrightarrow \bigcap_{k=1}^{n} A_k \in \mathcal{A}.$$

Let $A_1, \dots, A_n \in \mathcal{A}$. Property (ii) of σ-algebras gives $A_1^c, \dots, A_n^c \in \mathcal{A}$. By (1),

$$\bigcup_{k=1}^{n} A_k^c \in \mathcal{A},$$

and so, again by property (ii) of σ-algebras,

$$\left(\bigcup_{k=1}^{n} A_k^c \right)^c \in \mathcal{A}. \tag{2}$$

By De Morgan's laws,

$$\left(\bigcup_{k=1}^{n} A_k^c \right)^c = \bigcap_{k=1}^{n} \left(A_k^c \right)^c = \bigcap_{k=1}^{n} A_k. \tag{3}$$

The desired result follows from (2) and (3). ■

(c) First, note that $A \setminus B = A \cap B^c$. Since $\mathcal{A}$ is a σ-algebra and $B \in \mathcal{A}$, by property (ii) of σ-algebras we have $B^c \in \mathcal{A}$. As σ-algebras are closed under pairwise intersections (Exercise 1-(b)) and $A, B^c \in \mathcal{A}$, $A \cap B^c \in \mathcal{A}$. The desired result follows. ■

Remark. Algebras are closed under pairwise unions (by definition) and under pairwise intersections (by De Morgan's laws); by induction, we have that algebras are closed under finite unions and intersections. σ-algebras are closed under countably infinite unions (by definition) and under countably infinite intersections (by De Morgan's laws). By (1), σ-algebras are closed under pairwise unions, and so any σ-algebra on a nonempty set Ω is an algebra on Ω. It follows that σ-algebras are closed under countable unions and intersections, and that properties (a) and (c) in Exercise 1 hold for algebras as well. We will use these facts as a routine during the course.

Exercise 2

Solve the following problems.

(a) Let $\Omega := \{1, 2, 3\}$. Explicitly describe the family of all possible σ-algebras on Ω.

(b) Let $\mathbb{N}$ be the set of natural numbers and define

$$\mathcal{A} := \{A \subseteq \mathbb{N} : \min\{|A|, |\mathbb{N} \setminus A|\} < \infty\},$$

where $|A|$ denotes the cardinal number of $A \subseteq \mathbb{N}$. Is $\mathcal{A}$ an algebra on $\mathbb{N}$? A σ-algebra? Justify your answers.

(c) Let Ω be a nonempty set and define

$$\mathcal{A} := \{A \subseteq \Omega : \text{either } A \text{ or } \Omega \setminus A \text{ is countable}\}.$$

Is $\mathcal{A}$ an algebra on Ω? A σ-algebra? Justify your answers.

Solution

(a) The possible σ-algebras on $\Omega := \{1, 2, 3\}$ are:

$$\begin{aligned}
\mathcal{A}_1 &:= \{\emptyset, \Omega\}, \\
\mathcal{A}_2 &:= \{\emptyset, \{1\}, \{2, 3\}, \Omega\}, \\
\mathcal{A}_3 &:= \{\emptyset, \{2\}, \{1, 3\}, \Omega\}, \\
\mathcal{A}_4 &:= \{\emptyset, \{3\}, \{1, 2\}, \Omega\},
\end{aligned}$$

and
$$\mathcal{A}_5 := \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \Omega\}.$$

Hence, the family of all possible σ-algebras on Ω is

$$\{\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3, \mathcal{A}_4, \mathcal{A}_5\}.$$

Remark. How do we know we can stop looking for σ-algebras on Ω? Since all σ-algebras on Ω are subsets of $\mathcal{P}(\Omega) = \mathcal{A}_5$, we can start from $\mathcal{A}_5$ and check which of its proper subsets are σ-algebras on Ω. In our example, the cardinal number of Ω is just 3, so this is a simple task. In general, it can be shown that on a nonempty finite set Ω there are as many σ-algebras as partitions of Ω. The total number of partitions of a set with cardinal number n is the Bell number B_n. Bell numbers satisfy the recursion

$$B_{n+1} = \sum_{k=0}^{n} \binom{n}{k} B_k, \text{ with } B_0 = B_1 = 1.$$

The first several Bell numbers are $B_0 = 1$, $B_1 = 1$, $B_2 = 2$, $B_3 = 5$, $B_4 = 15$, $B_5 = 52$, and $B_6 = 203$. ■

(b) $\mathcal{A}$ is the class of all subsets A of $\mathbb{N}$ such that either A or $A^c := \mathbb{N} \setminus A$ is finite. We now show that $\mathcal{A}$ is an algebra, but not a σ-algebra, on $\mathbb{N}$. It is called the *cofinite algebra* on $\mathbb{N}$.

Claim 1. $\mathcal{A}$ is an algebra on $\mathbb{N}$.

Proof. First, since $\emptyset$ is finite, $\emptyset \in \mathcal{A}$. Second, let $A \in \mathcal{A}$; if A is finite, then $A^c \in \mathcal{A}$ because its complement $(A^c)^c = A$ is finite; if A^c is finite, then $A^c \in \mathcal{A}$ because it is finite. Third, let $A, B \in \mathcal{A}$. Then, we have four cases:

(i) A is finite and B is finite. Then, $A \cup B$ is finite (as the union of finite sets is finite), and so it is in $\mathcal{A}$.

(ii) A is finite and B^c is finite. Then, $(A \cup B)^c = A^c \cap B^c$ is finite (as $A^c \cap B^c \subseteq B^c$ and every subset of a finite set is finite), and therefore $A \cup B \in \mathcal{A}$.

(iii) A^c is finite and B is finite. Then, $(A \cup B)^c = A^c \cap B^c$ is finite (as $A^c \cap B^c \subseteq A^c$ and every subset of a finite set is finite), and therefore $A \cup B \in \mathcal{A}$.

(iv) A^c is finite and B^c is finite. Then, $(A \cup B)^c = A^c \cap B^c$ is finite (as $A^c \cap B^c \subseteq A^c$ and every subset of a finite set is finite), and therefore $A \cup B \in \mathcal{A}$.

This completes the proof that $\mathcal{A}$ is an algebra on $\mathbb{N}$.

Claim 2. $\mathcal{A}$ is not a σ-algebra on $\mathbb{N}$.

Proof. We have $\{2n\} \in \mathcal{A}$ for each $n \in \mathbb{N}$, but $\cup_{n \in \mathbb{N}} \{2n\} = \{2, 4, \dots\} \notin \mathcal{A}$ (as the set of even natural numbers is countably infinite, and so is the set of odd natural numbers, which is its complement in $\mathbb{N}$). So $\mathcal{A}$ is not closed under taking countably infinite unions. ■

(c) We show that $\mathcal{A}$ is a σ-algebra, and hence an algebra, on Ω. It is called the *cocountable σ-algebra* on Ω.

For any $A \subseteq \Omega$, let $A^c := \Omega \setminus A$. First, since $\emptyset$ is finite, it is countable, and so it is in $\mathcal{A}$. Second, let $A \in \mathcal{A}$; then, either A or A^c is countable, implying that either A^c or $(A^c)^c = A$ is countable; hence $A^c \in \mathcal{A}$. Third, suppose that $A_n \in \mathcal{A}$ for each $n \in \mathbb{N}$. There are two cases:

(i) A_n is countable for each $n \in \mathbb{N}$. In this case, $\bigcup_{n\in\mathbb{N}} A_n$ is countable, as the countable union of countable sets is countable, and therefore $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$.

(ii) A_j^c is countable for some $j \in \mathbb{N}$. In this case, $\left(\bigcup_{n\in\mathbb{N}} A_n\right)^c$ is countable and therefore $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$. To see that $\left(\bigcup_{n\in\mathbb{N}} A_n\right)^c$ is countable, note the following:

1. $$\left(\bigcup_{n\in\mathbb{N}} A_n\right)^c = A_j^c \cap \left(\bigcap_{n\in\mathbb{N}\setminus\{j\}} A_n^c\right) \subseteq A_j^c,$$
   where the equality follow by De Morgan's laws and the set inclusion by the properties of intersection;
2. Every subset of a countable set is countable. ■

Remark. Let $\mathcal{A}$ be the cocountable σ-algebra on a nonempty set Ω. We can show that $\mathcal{A} = \mathcal{P}(\Omega)$ if and only if Ω is countable. You can prove it as an exercise.

Probability Measures

Definition 3. *Let (Ω, P) be a measurable space. A function $P: \mathcal{A} \to [0, \infty)$ is said to be a* ***probability measure*** *on (Ω, P) if*

(a) $P(\Omega) = 1$;

(b) If $A_n \in \mathcal{A}$ for all $n \in \mathbb{N}$ and $A_n \cap A_m = \emptyset$ for all $m, n \in \mathbb{N}$ with $m \neq n$, then

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

Property (b) is called ***σ-additivity****. If P is a probability measure on $(\Omega, \mathcal{A})$, we refer to the triple $(\Omega, \mathcal{A}, P)$ as a* ***probability space****.*[3]

Exercise 3 (Properties of Probability Measures)

Let $(\Omega, \mathcal{A}, P)$ be a probability space. Prove the following statements.

(i) $P(\emptyset) = 0$.

(ii) *Finite additivity.* Let $n \in \mathbb{N}$. If $A_1, \ldots, A_n \in \mathcal{A}$ and $A_i \cap A_j = \emptyset$ for any $i, j \le n$ with $i \neq j$, then
$$P\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} P(A_k).$$

(iii) If $A \in \mathcal{A}$, then $P(\Omega \setminus A) = 1 - P(A)$.

(iv) *Monotonicity.* If $A, B \in \mathcal{A}$ and $A \subseteq B$, then $P(A) \le P(B)$.

[3] *Quiz:* How do we know that the infinite series appearing in property (b) is well-defined?

(v) *Bounded by* 1. $P(A) \leq 1$ for any $A \in \mathcal{A}$.

(vi) *Subtractivity.* If $A, B \in \mathcal{A}$,

$$P(B \setminus A) = P(B) - P(A \cap B) = P(A \cup B) - P(A).$$

(vii) *Poincaré-Sylvester.* If $A, B \in \mathcal{A}$ and $A \subseteq B$,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

(viii) *Continuity from below.* Let $\{A_n\}_{n=1}^{\infty}$ be a sequence of events in $\mathcal{A}$ such that $A_n \subseteq A_{n+1}$ for each n (in which case we say that $\{A_n\}_{n=1}^{\infty}$ is an increasing sequence). Then,

$$\lim_{n\to\infty} P(A_n) = P\left(\bigcup_{k=1}^{\infty} A_k\right).$$

(ix) *Continuity from above.* Let $\{A_n\}_{n=1}^{\infty}$ be a sequence of events in $\mathcal{A}$ such that $A_n \supseteq A_{n+1}$ for each n (in which case we say that $\{A_n\}_{n=1}^{\infty}$ is a decreasing sequence). Then,

$$\lim_{n\to\infty} P(A_n) = P\left(\bigcap_{k=1}^{\infty} A_k\right).$$

(x) *Sub-σ-additivity.* Let $\{A_n\}_{n=1}^{\infty}$ be a sequence of events in $\mathcal{A}$. Then,

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} P(A_n).$$

Solution

Let $(\Omega, \mathcal{A}, P)$ be a probability space.

(i) Set $A_1 := \Omega$ and $A_n := \emptyset$ for $n = 2, 3, \ldots$. Clearly, $\{A_n\}_{n=1}^{\infty}$ is a sequence of pairwise disjoint sets in $\mathcal{A}$. Thus, by σ-additivity,

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n). \tag{4}$$

Since $\bigcup_{n=1}^{\infty} A_n = \Omega$, (4) reads as

$$P(\Omega) = \sum_{n=1}^{\infty} P(A_n),$$

which is equivalent to

$$\sum_{n=2}^{\infty} P(\emptyset) = 0 \tag{5}$$

because $\sum_{n=1}^{\infty} P(A_n) = P(\Omega) + \sum_{n=2}^{\infty} P(\emptyset)$. The equation in (5) implies $P(\emptyset) = 0$ and concludes the proof. ■

(ii) Set $A_m := \emptyset$ for any natural number $m > n$. Clearly, $\{A_k\}_{k=1}^{\infty}$ is a sequence of pairwise disjoint sets in $\mathcal{A}$ such that $\bigcup_{k=1}^{\infty} A_k = \bigcup_{k=1}^{n} A_k$. Then,

$$\begin{aligned} P\left(\bigcup_{k=1}^{n} A_k\right) &= P\left(\bigcup_{k=1}^{\infty} A_k\right) \\ &= \sum_{k=1}^{\infty} P(A_k) \\ &= \sum_{k=1}^{n} P(A_k) + \sum_{m=n+1}^{\infty} P(\emptyset) \\ &= \sum_{k=1}^{n} P(A_k), \end{aligned}$$

where the second equality follows by σ-additivity, and the last equality holds true since $P(\emptyset) = 0$ by Exercise 3-(i). ■

(iii) Since $\Omega = A \cup (\Omega \setminus A)$, where A and $\Omega \setminus A$ are disjoint subsets of $\mathcal{A}$, finite additivity gives

$$P(\Omega) = P(A) + P(\Omega \setminus A).$$

Since $P(\Omega) = 1$, the claim follows. ■

(iv) If $A \subseteq B$, then $B = A \cup (B \setminus A)$, with A and $B \setminus A$ disjoint subsets of $\mathcal{A}$. Then, by finite additivity,

$$P(B) = P(A \cup (B \setminus A)) = P(A) + P(B \setminus A).$$

Since probability measures are non-negative, $P(B \setminus A) \geq 0$. The claim follows. ■

(v) For any $A \in \mathcal{A}$, $A \subseteq \Omega$. Then, by monotonicity (Exercise 3-(iv)),

$$P(A) \leq P(\Omega) = 1. \qquad ■$$

(vi) Since $B = (B \setminus A) \cup (A \cap B)$, with $B \setminus A$ and $A \cap B$ disjoint subsets of $\mathcal{A}$, finite additivity implies

$$P(B) = P((B \setminus A) \cup (A \cap B)) = P(B \setminus A) + P(A \cap B),$$

which rearranged gives $P(B \setminus A) = P(B) - P(A \cap B)$.

Since $A \cup B = A \cup (B \setminus A)$, with A and $B \setminus A$ disjoint subsets of $\mathcal{A}$, finite additivity implies

$$P(A \cup B) = P(A \cup (B \setminus A)) = P(A) + P(B \setminus A),$$

which rearranged gives $P(B \setminus A) = P(A \cup B) - P(A)$. ■

Remark. When $A \subseteq B$, subtractivity reads as $P(B \setminus A) = P(B) - P(A)$.

(vii) The set $A \cup B$ is equal to

$$(A \setminus (A \cap B)) \cup (B \setminus (A \cap B)) \cup (A \cap B)\,,$$

which is a union of pairwise disjoint sets in $\mathcal{A}$. By finite additivity,

$$P(A \cup B) = P(A \setminus (A \cap B)) + P(B \setminus (A \cap B)) + P((A \cap B)). \tag{6}$$

Since $A \cap B \subseteq A$ and $A \cap B \subseteq B$, subtractivity implies

$$P(A \setminus (A \cap B)) = P(A) - P(A \cap B) \tag{7}$$

and

$$P(B \setminus (A \cap B)) = P(B) - P(A \cap B). \tag{8}$$

The desired result follows by combining (6), (7) and (8). ■

(viii) Set $B_1 := A_1$ and $B_n := A_n \setminus A_{n-1}$ for $n = 2, 3, \ldots$, and note that $B_n \in \mathcal{A}$ for each n by Exercise 1-(c). By construction,

$$A_n = \bigcup_{k=1}^{n} B_k \tag{9}$$

for each n, and so

$$\bigcup_{k=1}^{\infty} A_k = \bigcup_{k=1}^{\infty} B_k. \tag{10}$$

Moreover, $B_i \cap B_j = \emptyset$ holds by construction for any distinct i and j, and so $\{B_k\}_{k=1}^{\infty}$ is a sequence of pairwise disjoint sets in $\mathcal{A}$. Then,

$$\begin{aligned}
P\left(\bigcup_{k=1}^{\infty} A_k\right) &= P\left(\bigcup_{k=1}^{\infty} B_k\right) \\
&= \sum_{k=1}^{\infty} P(B_k) \\
&= \lim_{n\to\infty} \sum_{k=1}^{n} P(B_k) \\
&= \lim_{n\to\infty} P\left(\bigcup_{k=1}^{n} B_k\right) \\
&= \lim_{n\to\infty} P(A_n),
\end{aligned}$$

where: the first equality holds by (10); the second equality holds by σ-additivity; the third equality holds by definition of infinite series; the fourth equality holds by finite additivity; the last equality holds by (9). ■

(ix) Since $\{A_n\}_{n=1}^{\infty}$ is a decreasing sequence of events in $\mathcal{A}$, $\{A_n^c\}_{n=1}^{\infty}$ is an increasing sequence of events in $\mathcal{A}$. Continuity from below implies that

$$\lim_{n\to\infty} P(A_n^c) = P\left(\bigcup_{k=1}^{\infty} A_k^c\right). \tag{11}$$

In view of Exercise 3-(iii),

$$\lim_{n\to\infty} P(A_n^c) = \lim_{n\to\infty} (1 - P(A_n)) = 1 - \lim_{n\to\infty} P(A_n), \tag{12}$$

and

$$P\left(\bigcup_{k=1}^{\infty} A_k^c\right) = 1 - P\left(\left(\bigcup_{k=1}^{\infty} A_k^c\right)^c\right) = 1 - P\left(\bigcap_{k=1}^{\infty} A_k\right), \tag{13}$$

where the last equality in (13) holds by De Morgan's laws. The desired result follows from (11), (12) and (13). ■

(x) Set $B_1 := A_1$ and $B_n := A_n \setminus \bigcup_{i=1}^{n-1} A_i$ for $n = 2, 3, \ldots$ Clearly:

(a) $B_n \in \mathcal{A}$ for each n;

(b) $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n$;

(c) $B_i \cap B_j = \emptyset$ for any $i \neq j$, and so $\{B_n\}_{n=1}^{\infty}$ is a sequence of pairwise disjoint sets;

(d) $B_n \subseteq A_n$ for each n.

(Take a moment to convince yourself about the previous statements.) Then,

$$\begin{aligned} P\left(\bigcup_{n=1}^{\infty} A_n\right) &= P\left(\bigcup_{n=1}^{\infty} B_n\right) \\ &= \sum_{n=1}^{\infty} P(B_n) \\ &\leq \sum_{n=1}^{\infty} P(A_n) \end{aligned}$$

where: the first equality holds by (b); the second equality holds by (a), (c) and σ-additivity; the inequality holds as, by (d) and monotonicity (Exercise 3-(iv)), $P(B_n) \leq P(A_n)$ for all n. ■

Exercise 4

Solve the following problems.

(a) Let $(\Omega, \mathcal{A}, P)$ be a probability space and $\{A_n\}_{n=1}^{\infty}$ a sequence of events in $\mathcal{A}$ such that $P(A_n) = 0$ for each n. Which value does $P\left(\bigcup_{n=1}^{\infty} A_n\right)$ take?

(b) Is the set function $P: \mathcal{P}(\mathbb{R}) \to \mathbb{R}$ defined by

$$P(A) := \begin{cases} 0 & \text{if } A \text{ is a finite set} \\ 1 & \text{else} \end{cases},$$

a probability measure on the measurable space $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$? Justify your answer.

Solution

(a) Since P is a probability measure, it takes values in $[0, \infty)$, and so

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) \geq 0. \tag{14}$$

By sub-σ-additivity,

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} P(A_n). \tag{15}$$

By assumption, $P(A_n) = 0$ for each n, and so $\sum_{n=1}^{\infty} P(A_n) = 0$. This fact, combined with with (15), gives

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) \leq 0. \tag{16}$$

Together, (14) and (16) imply

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = 0. \qquad ■$$

(b) P is not a probability measure on $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$ because it violates finite additivity (and hence σ-additivity). To see this, let $A_1 := \{1, 3, 5, \dots\}$ and $A_2 := \{2, 4, 6, \dots\}$. A_1 and A_2 are in $\mathcal{P}(\mathbb{R})$ and are not finite; thus,

$$P(A_1) + P(A_2) = 1 + 1 = 2.$$

Moreover, $A_1 \cup A_2 = \mathbb{N}$ is in $\mathcal{P}(\mathbb{R})$ and is not finite; hence,

$$P(A_1 \cup A_2) = 1.$$

Therefore,

$$P(A_1 \cup A_2) = 1 \neq 2 = P(A_1) + P(A_2). \tag{17}$$

Since A_1 and A_2 are also disjoint, (17) contradicts finite additivity. ■

Combinatorics, Laplace Experiments and Probabilities

Exercise 5

How many possible ways there exist to arrange 3 novels, 2 math books and 1 econ book in a bookshelf,

(a) If the order is arbitrary?

(b) If math books and novels are placed together?

(c) If only novels are to be placed together?

Solution

(a) If we arrange the $3 + 2 + 1 = 6$ books in an arbitrary order, we have $6!$ possible arrangements.

(b) Suppose the books of the same subject are all in a box, so that we have 3 boxes in total (one for each subject). The numbers of arbitrary arrangements of books in the same box are: $3!$ for novels, $2!$ for math books and $1!$ for the econ book. Thus, we have $3!2!1!$ arrangement of books ($3!$ arrangements of novels for each of the $2!$ arrangements of the math books for each of the $1!$ arrangements of the econ book) for each arrangement of the 3 boxes. As we have $3!$ arbitrary arrangements of the 3 boxes, the number of arrangements we are looking for is $3!3!2!1!$.

(c) Let's put the 3 novels in a box. In total, we have 4 objects: 3 books and 1 box. We can arrange these 4 objects in $4!$ ways. Since we can do this for each of the $3!$ arrangements of the novels in the box, the number of arrangements we are looking for is $4!3!$. ■

Exercise 6

A fair die is rolled for three times.

(a) What is the probability that no even number occurs?

(b) What is the probability to obtain an increasing sequence of numbers?

(c) What is the probability to obtain a strictly increasing sequence of numbers?

Briefly explain your reasoning.

Solution

The sample space for this experiment is $\Omega := \{1, 2, 3, 4, 5, 6\}^3$. Since $|\Omega| = 6^3 = 216$, the sample space is finite. We choose $\mathcal{P}(\Omega)$ as σ-field on Ω. The die is fair and rolls do not affect each other; hence, it is safe to assume that elementary events are equally likely, i.e.,

$$P(\{(\omega_1, \omega_2, \omega_3)\}) = \frac{1}{|\Omega|} \quad \text{for each} \quad (\omega_1, \omega_2, \omega_3) \in \Omega,$$

so that the underlying statistical experiment is a Laplace experiment.

(a) Let

$$A_1 := \{(\omega_1, \omega_2, \omega_3) \in \Omega : \omega_i \text{ is odd } \forall i \in \{1, 2, 3\}\}.$$

Note that $A_1 = \{1, 3, 5\}^3$, and so $|A_1| = 3^3 = 27$. Therefore,

$$P(A_1) = \frac{|A_1|}{|\Omega|} = \frac{27}{216} = \frac{1}{8},$$

where $P(A_1)$ is the Laplace probability of event A_1.

(b) Let

$$A_2 := \{(\omega_1, \omega_2, \omega_3) \in \Omega : \omega_1 \leq \omega_2 \leq \omega_3\}.$$

In this case order does not matter (e.g., $(1, 2, 3) \in A_2$, but $(2, 1, 3) \notin A_2$, so you do not want to count it twice) but replacement is allowed because the sequence has only to be weakly increasing (e.g., $(1, 1, 2) \in A_2$) Therefore (rearrangement and replacement),

$$|A_2| = \binom{6+3-1}{3} = \frac{8!}{3!\,(8-3)!} = 56.$$

Hence,

$$P(A_2) = \frac{|A_2|}{|\Omega|} = \frac{56}{216} = \frac{7}{27},$$

where $P(A_2)$ is the Laplace probability of event A_2.

(c) Let

$$A_3 := \{(\omega_1, \omega_2, \omega_3) \in \Omega : \omega_1 < \omega_2 < \omega_3\}.$$

In this case order does not matter and replacement is not allowed, as the sequence has to be strictly increasing. Therefore (rearrangement but not replacement),

$$|A_3| = \binom{6}{3} = \frac{6!}{3!\,(6-3)!} = 20.$$

Hence,

$$P(A_3) = \frac{|A_2|}{|\Omega|} = \frac{20}{216} = \frac{5}{54},$$

where $P(A_3)$ is the Laplace probability of event A_3. ∎

E 703: Advanced Econometrics I
Solutions to Problem Set 2

Niccolò Lomys

October 20, 2016

Exercise 7

Solve the following problems.

(a) How many functions are there from $\{1, 2, \ldots, k\}$ to $\{1, 2, \ldots, n\}$? How many of these are one-to-one if $k > n$? How many of these are one-to-one if $k \leq n$? Briefly explain your reasoning.

(b) A sports class with $2N$ students is grouped randomly into two teams, with N players in each team. What is the probability that two specific students (say, Bob and Tom) are in the same team? Briefly explain your reasoning.

Solution

(a) A function $f\colon\{1, 2, \ldots, k\} \to \{1, 2, \ldots, n\}$ is an assignment of exactly one integer in the codomain to each of the integers in the domain. In other words, such function is specified by the k-tuple $(f(1), \ldots, f(k))$. Since each $f(i)$ could be any of the n integers in $\{1, 2, \ldots, n\}$ for each integer i in $\{1, 2, \ldots, k\}$, there are $n \cdot n \cdots n$ (k times) $= n^k$ possible different assignments in total.

A function $f\colon\{1, 2, \ldots, k\} \to \{1, 2, \ldots, n\}$ is one-to-one if for each $i, j \in \{1, 2, \ldots, k\}$ with $i \neq j$ we have $f(i) \neq f(j)$. Thus, if $k > n$ there exists no one-to-one function from $\{1, 2, \ldots, k\}$ to $\{1, 2, \ldots, n\}$. If $k \leq n$, there are n possibilities for $f(1)$, $n - 1$ possibilities for $f(2)$ since the first integer, $f(1)$, cannot be used again, $n-2$ choices for $f(3)$ since the first two integers cannot be used again, and so on. In total, there are

$$n \cdot n - 1 \cdot n - 2 \cdot \ldots \cdot n - k + 1 = \frac{n!}{(n-k)!}$$

one-to-one functions from $\{1, 2, \ldots, k\}$ to $\{1, 2, \ldots, n\}$. Note that there are k terms in the last product. ■

(b) Since there is a finite number of pairs of opposing teams and students are assigned randomly to teams (so that pairs of opposing teams are equally likely), this is a Laplace experiment. Suppose Bob and Tom are assigned to the two teams in that order (first Bob, then Tom), with the rest of the students being assigned later. Whatever team Bob is assigned to, there are $N - 1$ free spots on that team and $2N - 1$ students yet to be assigned. The probability that Bob is given one of these spots is $(N-1)/(2N-1)$. Hence, the (Laplace) probability that Bob and Tom are in the same team is $(N-1)/(2N-1)$. ■

*Exercise 8

Solve the following problems.

(a) Let X be a metric space, and denote with $\mathcal{O}_X$ and $\mathcal{C}_X$ the class of all open and closed subsets of X, respectively. Is either $\mathcal{O}_X$ or $\mathcal{C}_X$ a field on X? A σ-field? How about $\mathcal{O}_X \cup \mathcal{C}_X$? Justify your answers. [*Note.* A positive answer needs a proof, while for negative answers a counterexample suffices.]

(b) Let $(\Omega, \mathcal{A})$ be a measurable space and pick any $\omega \in \Omega$. Consider the function $\delta_\omega : \mathcal{A} \to \{0, 1\}$ defined by

$$\delta_\omega(A) := \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{else} \end{cases} .$$

Show that δ_ω is a probability measure on $(\Omega, \mathcal{A})$. It is called the *Dirac (probability) measure* on $(\Omega, \mathcal{A})$.

(c) (Borel-Cantelli Lemma) Let $(\Omega, \mathcal{A}, P)$ be a probability space and $\{A_n\}_{n=1}^\infty$ a sequence of events in $\mathcal{A}$ such that $\sum_{n=1}^\infty P(A_n) < \infty$. For each n, define $B_n := \bigcup_{m \geq n} A_m$. Show that

$$P\left(\bigcap_{n=1}^\infty B_n\right) = 0.$$

Solution

(a) *Claim 1.* $\mathcal{O}_X$ is not a field, hence not a σ-field, on X.

Counterexample. Consider the metric space $(\mathbb{R}, d)$, where d is the Euclidean distance. The set $(0, +\infty)$ is open in $(\mathbb{R}, d)$, and so $(0, +\infty) \in \mathcal{O}_X$. However, $\mathbb{R} \setminus (0, +\infty) = (-\infty, 0]$ is not open[1] in $(\mathbb{R}, d)$, and so $\mathbb{R} \setminus (0, +\infty) \notin \mathcal{O}_X$.

Claim 2. $\mathcal{C}_X$ is not a field, hence not a σ-field, on X.

Counterexample. Consider the metric space $(\mathbb{R}, d)$, where d is the Euclidean distance. The set $(-\infty, 0]$ is closed in $(\mathbb{R}, d)$, and so $(-\infty, 0] \in \mathcal{C}_X$. However, $\mathbb{R} \setminus (-\infty, 0] = (0, +\infty)$ is not closed[2] in $(\mathbb{R}, d)$, and so $\mathbb{R} \setminus (-\infty, 0] \notin \mathcal{C}_X$.

Claim 3. $\mathcal{O}_X \cup \mathcal{C}_X$ is not a field, hence not a σ-field, on X.

Counterexample. Consider the metric space $(\mathbb{R}, d)$, where d is the Euclidean distance. The set $(0, 1)$ is open in $(\mathbb{R}, d)$, and so $(0, 1) \in \mathcal{O}_X \cup \mathcal{C}_X$. The set $\{1\}$ is closed in $(\mathbb{R}, d)$, and so $\{1\} \in \mathcal{O}_X \cup \mathcal{C}_X$. However, $(0, 1) \cup \{1\} = (0, 1]$ is neither open nor closed, and so $(0, 1) \cup \{1\} \notin \mathcal{O}_X \cup \mathcal{C}_X$. ■

(b) We show that the defining properties of probability measure (Definition 1.2 in the Lecture Notes) are satisfied by δ_ω.

(o) $\delta_\omega(A) \in \{0, 1\} \subseteq [0, \infty)$ for all $A \in \mathcal{A}$ by definition of δ_ω. Hence, $\delta_\omega : \mathcal{A} \to [0, \infty)$.

(i) Since $\omega \in \Omega$, $\delta_\omega(\Omega) = 1$.

(ii) Let $A_n \in \mathcal{A}$ for all $n = 1, 2, \ldots$, with $A_i \cap A_j = \emptyset$ for any $i \neq j$. We distinguish two cases:

[1] *Quiz:* Why?
[2] *Quiz:* Why?

1. $\omega \notin A_n$ for all $n = 1, 2, \ldots$.
   In this case, $\delta_\omega(A_n) = 0$ for all $n = 1, 2, \ldots$, and so

$$\sum_{n=1}^{\infty} \delta_\omega(A_n) = 0.$$

   Moreover, as $\omega \notin A_n$ for all $n = 1, 2, \ldots$, $\omega \notin \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$, so that

$$\delta_\omega\left(\bigcup_{n=1}^{\infty} A_n\right) = 0.$$

   Therefore,

$$\delta_\omega\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \delta_\omega(A_n).$$

2. $\omega \in A_k$ for exactly one k.[3]
   In this case, $\delta_\omega(A_k) = 1$ and $\delta_\omega(A_j) = 0$ for any $j \neq k$. Then,

$$\sum_{n=1}^{\infty} \delta_\omega(A_n) = \delta_\omega(A_k) + \sum_{j \neq k} \delta_\omega(A_j) = 1 + 0 = 1.$$

   Moreover, as $\omega \in A_k$, $\omega \in \bigcup_{n=1}^{\infty} A_n$, so that

$$\delta_\omega\left(\bigcup_{n=1}^{\infty} A_n\right) = 1.$$

   Therefore,

$$\delta_\omega\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \delta_\omega(A_n). \qquad \blacksquare$$

(c) Since $B_n \in \mathcal{A}$ and $B_n \supseteq B_{n+1}$ for all $n = 1, 2, \ldots$, $\{B_n\}_{n=1}^{\infty}$ is a decreasing sequence of events in $\mathcal{A}$. Then,

$$\begin{aligned}
P\left(\bigcap_{n=1}^{\infty} B_n\right) &= \lim_{n\to\infty} P(B_n) \\
&= \lim_{n\to\infty} P\left(\bigcup_{m=n}^{\infty} A_m\right) \\
&\leq \lim_{n\to\infty} \sum_{m=n}^{\infty} P(A_m) \\
&= \lim_{n\to\infty} \left[\sum_{m=1}^{\infty} P(A_m) - \sum_{m=1}^{n-1} P(A_m)\right] \qquad (1) \\
&= \sum_{m=1}^{\infty} P(A_m) - \lim_{n\to\infty} \sum_{m=1}^{n-1} P(A_m) \\
&= \sum_{m=1}^{\infty} P(A_m) - \sum_{m=1}^{\infty} P(A_m) \\
&= 0,
\end{aligned}$$

[3] Since the A_n's are pairwise disjoint, there exists at most one k for which $\omega \in A_k$.

where: the first equality holds by continuity from above of P, the second equality holds by definition of B_n, the inequality holds by sub-σ-additivity of P, and the last equality holds because the infinite series $\sum_{m=1}^{\infty} P(A_m)$ is convergent by assumption (i.e., $\sum_{m=1}^{\infty} P(A_m) < \infty$). Moreover, by non-negativity of P, we have

$$P\left(\bigcap_{n=1}^{\infty} B_n\right) \geq 0. \tag{2}$$

Together, (1) and (2) imply

$$P\left(\bigcap_{n=1}^{\infty} B_n\right) = 0,$$

as desired. ■

Exercise 9 (Final Exam - Spring 2012)

Let P be a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, where $\mathcal{B}(\mathbb{R})$ denotes the Borel σ-field on $\mathbb{R}$, and define

$$\mathcal{A} := \{A \in \mathcal{B}(\mathbb{R}) : P(A) \in \{0, 1\}\}.$$

Show that $\mathcal{A}$ is a σ-field on $\mathbb{R}$.

Solution

We show that the defining properties of σ-field (Definition 1.1 in the Lecture Notes) are satisfied by $\mathcal{A}$.

(o) $\mathcal{A} \subseteq \mathcal{B}(\mathbb{R}) \subseteq \mathcal{P}(\mathbb{R})$. Hence, $\mathcal{A}$ is a class of subsets of $\mathbb{R}$.

(i) Since P is a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, $P(\emptyset) = 0$. Hence, $\emptyset \in \mathcal{A}$.

(ii) Let $A \in \mathcal{A}$. Since $\mathcal{A} \subseteq \mathcal{B}(\mathbb{R})$ and $\mathcal{B}(\mathbb{R})$ is a σ-field, $A^c := \mathbb{R} \setminus A \in \mathcal{B}(\mathbb{R})$. Moreover, $P(A^c)$ is well defined because P is a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. We distinguish two cases:

   1. $P(A) = 0$. Then, $P(A^c) = 1 - P(A) = 1 - 0 = 1$, where the first equality follows by Theorem 1.3.(iii) in the Lecture Notes. Hence, $A^c \in \mathcal{A}$.
   2. $P(A) = 1$. Then, $P(A^c) = 1 - P(A) = 1 - 1 = 0$, where the first equality follows by Theorem 1.3.(iii) in the Lecture Notes. Hence, $A^c \in \mathcal{A}$.

(iii) Let $A_n \in \mathcal{A}$ for all $n = 1, 2, \ldots$. Clearly, $\bigcup_{n=1}^{\infty} A_n \in \mathcal{B}(\mathbb{R})$ and $P\left(\bigcup_{n=1}^{\infty} A_n\right)$ is well defined. We distinguish two cases:

   1. $P(A_n) = 0$ for all $n = 1, 2, \ldots$. By non-negativity and sub-σ-additivity of probability measure P,

   $$0 \leq P\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} P(A_n) = 0.$$

   Then, $P\left(\bigcup_{n=1}^{\infty} A_n\right) = 0$, showing that $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.

   2. $P(A_n) = 1$ for some n. Since $A_n \subseteq \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$, by monotonicity of probability measure P and the fact that P is bounded by 1,

   $$1 = P(A_n) \leq P\left(\bigcup_{n=1}^{\infty} A_n\right) \leq 1.$$

   Then, $P\left(\bigcup_{n=1}^{\infty} A_n\right) = 1$, showing that $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$. ■

*Exercise 10

Let Ω be a nonempty set and I a nonempty index set. Moreover, suppose that $\mathcal{A}_i$ is a σ-algebra on Ω for each $i \in I$.

(a) Show that $\bigcap_{i\in I} \mathcal{A}_i$ is a σ-algebra on Ω.

(b) Let $\mathcal{A}^*$ be a nonempty class of subsets of Ω. Show that the smallest σ-algebra on Ω generated by $\mathcal{A}^*$ is well-defined (i.e., show that it exists and that it is unique).

(c) Give an example to show that $\bigcup_{i\in I} \mathcal{A}_i$ need not be an algebra even if I is a finite set. Try to come up with an example which is as simple as possible.

(d) Now suppose that $I = \mathbb{N}$ and that $\mathcal{A}_i \subseteq \mathcal{A}_{i+1}$ for each $i \in \mathbb{N}$. Show that $\bigcup_{i\in\mathbb{N}} \mathcal{A}_i$ is an algebra on Ω, but it need not be a σ-algebra. [*Note.* A much stronger statement is actually true. Namely: if $\mathcal{A}_i \subset \mathcal{A}_{i+1}$ for each $i \in \mathbb{N}$, where the inclusion is now proper, then $\bigcup_{i\in\mathbb{N}} \mathcal{A}_i$ can *never* be a σ-algebra. You can try to establish this result, but it is not too trivial.]

Solution

(a) Let $\mathcal{A} := \bigcap_{i\in I} \mathcal{A}_i$. We show that the defining properties of σ-algebra (Definition 1.1 in the Lecture Notes) are satisfied by $\mathcal{A}$.

- (o) Clearly, $\mathcal{A}$ is a class of subsets of Ω.
- (i) Since $\mathcal{A}_i$ is a σ-algebra on Ω for all $i \in I$, $\emptyset \in \mathcal{A}_i$ for all $i \in I$. Then, $\emptyset \in \mathcal{A}$.
- (ii) Let $A \in \mathcal{A}$. Then, $A \in \mathcal{A}_i$ for all $i \in I$. Since $\mathcal{A}_i$ is a σ-algebra on Ω for all $i \in I$, $\Omega \setminus A \in \mathcal{A}_i$ for all $i \in I$. Then, $\Omega \setminus A \in \mathcal{A}$.
- (iii) Let $A_n \in \mathcal{A}$ for all $n = 1, 2, \ldots$. Then, for all $i \in I$, $A_n \in \mathcal{A}_i$ for $n = 1, 2, \ldots$. As $\mathcal{A}_i$ is a σ-algebra on Ω for all $i \in I$, $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}_i$ for all $i \in I$. Then, $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$. ■

(b) First, note that there is at least a σ-algebra on Ω, namely $\mathcal{P}(\Omega)$, which is sure to contain $\mathcal{A}^*$. That is, the collection

$$\Sigma := \{\mathcal{A} \subseteq \mathcal{P}(\Omega) : \mathcal{A} \text{ is a } \sigma\text{-algebra on } \Omega \text{ such that } \mathcal{A}^* \subseteq \mathcal{A}\}$$

is nonempty. Furthermore, if we intersect all members of this family we again get a σ-algebra on Ω (by Exercise 10-(a)) that contains $\mathcal{A}^*$ (by definitions of intersection and Σ). Obviously, this σ-algebra is a subset of any σ-algebra on Ω, say $\mathcal{C}$, that contains $\mathcal{A}^*$, as $\mathcal{C}$ is one of the intersected members. Thus, there is a unique smallest algebra that contains $\mathcal{A}^*$, so $\sigma(\mathcal{A}^*)$ is well-defined, and we have

$$\sigma(\mathcal{A}^*) = \bigcap_{\mathcal{A}\in\Sigma} \mathcal{A}. \qquad ■$$

Remark. Exercise 10-(b) proves Lemma 1.6 in the Lecture Notes.

(c) Let $\Omega := \{1, 2, 3\}$. It is easy to see that

$$\mathcal{A}_1 := \{\emptyset, \{1\}, \{2, 3\}, \Omega\}$$

and

$$\mathcal{A}_2 := \{\emptyset, \{2\}, \{1, 3\}, \Omega\}$$

are σ-algebras on Ω. However,

$$\mathcal{A} := \mathcal{A}_1 \cup \mathcal{A}_2 = \{\emptyset, \{1\}, \{2\}, \{1, 3\} \{2, 3\}, \Omega\}$$

is not an algebra on Ω because $\{1\}, \{2\} \in \mathcal{A}$, but $\{1\} \cup \{2\} = \{1, 2\} \notin \mathcal{A}$. ■

(d) Let $\mathcal{A} := \bigcup_{i\in\mathbb{N}} \mathcal{A}_i$. Note that, by definition of union, $\mathcal{A}_i \subseteq \mathcal{A}$ for all $n \in \mathbb{N}$. We show that the defining properties of algebra (Definition 1.9 in the Lecture Notes) are satisfied by $\mathcal{A}$.

   (o) Clearly, $\mathcal{A}$ is a class of subsets of Ω.

   (i) Fix some index $j \in \mathbb{N}$. Since $\mathcal{A}_j$ is a σ-algebra on Ω, $\emptyset \in \mathcal{A}_j$. Since $\mathcal{A}_j \subseteq \mathcal{A}$, $\emptyset \in \mathcal{A}$.

   (ii) Let $A \in \mathcal{A}$. Then, there exists some $j \in \mathbb{N}$ such that $A \in \mathcal{A}_j$. Since $\mathcal{A}_j$ is a σ-algebra on Ω, $\Omega \setminus A \in \mathcal{A}_j$. Since $\mathcal{A}_j \subseteq \mathcal{A}$, $\Omega \setminus A \in \mathcal{A}$.

   (iii) Let $A, B \in \mathcal{A}$. Then, there exists some $j, k \in \mathbb{N}$ such that $A \in \mathcal{A}_j$ and $B \in \mathcal{A}_k$. Set $m := \max\{j, k\}$. Since $\mathcal{A}_i \subseteq \mathcal{A}_{i+1}$ for each $i \in \mathbb{N}$, $A, B \in \mathcal{A}_m$. As $\mathcal{A}_m$ is a σ-field on Ω, $A \cup B \in \mathcal{A}_m$. Then, $A \cup B \in \mathcal{A}$ because $\mathcal{A}_m \subseteq \mathcal{A}$.

   The following counterexample shows that $\bigcup_{i\in\mathbb{N}} \mathcal{A}_i$ need not be a σ-algebra. Set $\Omega := \mathbb{N}$ and, for all $i \in \mathbb{N}$, let $\mathcal{A}_i$ be the σ-algebra on $\mathbb{N}$ generated by the singletons $\{k\}$ with $k \le i$. Clearly, $\mathcal{A}_i \subseteq \mathcal{A}_{i+1}$ for all $i \in \mathbb{N}$. However, $\bigcup_{i\in\mathbb{N}} \mathcal{A}_i$ is equal to the cofinite algebra on $\mathbb{N}$ (prove it!), which we know it is not a σ-algebra on $\mathbb{N}$ from Problem Set 1 (Exercise 2-(b)). ■

Exercise 11

Define $\mathcal{B}$ as the σ-field on $\mathbb{R}$ generated by the class of all open and bounded intervals (a, b) of the real line, with $-\infty < a < b < \infty$. Furthermore, define $\mathcal{B}^*$ as the σ-field on $\mathbb{R}$ generated by the class of all closed and bounded intervals $[a, b]$, with $-\infty < a < b < \infty$. Show that $\mathcal{B} = \mathcal{B}^*$. [*Hint.* You have to show that $\mathcal{B} \subseteq \mathcal{B}^*$ and $\mathcal{B}^* \subseteq \mathcal{B}$.]

Solution

Define

$$\mathcal{A} := \text{ set of all intervals of the form } (a, b), \text{ with } -\infty < a < b < \infty,$$

and

$$\mathcal{A}^* := \text{ set of all intervals of the form } [a, b], \text{ with } -\infty < a < b < \infty,$$

so that $\mathcal{B} := \sigma(\mathcal{A})$ and $\mathcal{B}^* := \sigma(\mathcal{A}^*)$. We show that $\mathcal{B} \subseteq \mathcal{B}^*$ and $\mathcal{B}^* \subseteq \mathcal{B}$, which imples $\mathcal{B} = \mathcal{B}^*$.

(i) $[\mathcal{B} \subseteq \mathcal{B}^*]$

   It suffices to show that $\mathcal{A} \subseteq \mathcal{B}^*$. This is so because, as $\mathcal{B}$ is the smallest σ-field on $\mathbb{R}$ containing $\mathcal{A}$, $\langle \mathcal{A} \subseteq \mathcal{B}^*, \mathcal{B}^* \ \sigma\text{-field on } \mathbb{R}\rangle \implies \mathcal{B} \subseteq \mathcal{B}^*$. Let $(a, b) \in \mathcal{A}$ and fix a natural number $k > 2/(b - a)$. Observe that

$$(a, b) = \bigcup_{n=k}^{\infty} \left[a + \frac{1}{n}, b - \frac{1}{n}\right].$$

   For each $n \ge k$, $[a + 1/n, b - 1/n] \in \mathcal{B}^*$ and so, since $\mathcal{B}^*$ is a σ-field,

$$\bigcup_{n=k}^{\infty} \left[a + \frac{1}{n}, b - \frac{1}{n}\right] \in \mathcal{B}^*.$$

   Hence, $(a, b) \in \mathcal{B}^*$. Since (a, b) was arbitrarily chosen in $\mathcal{A}$, the desired result follows.

(ii) $[\mathcal{B}^* \subseteq \mathcal{B}]$

   For the same argument as in (i), it suffices to show that $\mathcal{A}^* \subseteq \mathcal{B}$. Let $[a, b] \in \mathcal{A}^*$. Observe that

$$[a, b] = \bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, b + \frac{1}{n}\right).$$

   For each n, $(a - 1/n, b + 1/n) \in \mathcal{B}$, and so, since $\mathcal{B}$ is a σ-field,

$$\bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, b + \frac{1}{n}\right) \in \mathcal{B}.$$

Hence, $[a, b] \in \mathcal{B}$. Since $[a, b]$ was arbitrarily chosen in $\mathcal{A}^*$, the desired result follows. ■

Remark. Exercise 11 proves (part of) Theorem 1.8 in the Lecture Notes.

Exercise 12

Let $\Omega := \{1, 2, 3, 4\}$ and consider the function $p\colon \Omega \to [0, 1]$ defined by $p(\omega) := \omega/10$.

(a) Show that $P\colon \mathcal{P}(\Omega) \to [0, \infty)$, defined by $P(A) := \sum_{\omega \in A} p(\omega)$, defines a probability measure on $(\Omega, \mathcal{P}(\Omega))$.

(b) Determine the σ-field on Ω generated by $\mathcal{E} := \{A \in \mathcal{P}(\Omega) : P(A) = 1/2\}$.

Solution

(a) We show that the defining properties of probability measure are satisfied by P.

   (o) By definition of P, $P\colon \mathcal{P}(\Omega) \to [0, \infty)$ clearly holds true.

   (i) $P(\Omega) = \sum_{\omega \in \Omega} p(\omega) = 1/10 + 2/10 + 3/10 + 4/10 = 1$.

   (ii) Let $A_n \in \mathcal{P}(\Omega)$ for all $n = 1, 2, \ldots$, with $A_i \cap A_j = \emptyset$ for any $i \neq j$. Then,

$$
\begin{aligned}
P\left(\bigcup_{n=1}^{\infty} A_n\right) &= \sum_{\omega \in \bigcup_{n=1}^{\infty} A_n} p(\omega) \\
&= \sum_{\omega \in \Omega} p(\omega) \mathbb{1}_{\bigcup_{n=1}^{\infty} A_n}(\omega) \\
&= \sum_{\omega \in \Omega} p(\omega) \sum_{n=1}^{\infty} \mathbb{1}_{A_n}(\omega) \\
&= \sum_{n=1}^{\infty} \sum_{\omega \in \Omega} p(\omega) \mathbb{1}_{A_n}(\omega) \\
&= \sum_{n=1}^{\infty} \left[\sum_{\omega \in A_n} p(\omega) \right] \\
&= \sum_{n=1}^{\infty} P(A_n),
\end{aligned}
$$

   where the first and the last equality hold by definition of P, and the third equality holds because the set $A_1, A_2, \ldots$ are pairwise disjoint.

   Remark. Part (ii) also proves Lemma 1.4 in the Lecture Notes.

(b) Note that $\mathcal{E} = \{\{1, 4\}, \{2, 3\}\}$, and $\Omega \setminus \{1, 4\} = \{2, 3\}$. Then,

$$
\sigma(\mathcal{E}) = \{\emptyset, \{1, 4\}, \{2, 3\}, \Omega\}. \qquad ■
$$

E 703: Advanced Econometrics I
Solutions to Problem Set 3

Niccolò Lomys

October 27, 2016

*Exercise 13

Solve the following problems.

(a) Let P be a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and define the function $F\colon \mathbb{R} \to [0,1]$ as

$$F(b) := P\big((-\infty, b]\big) \quad \text{for all } b \in \mathbb{R}.$$

Show that:

(i) $P\big((a,b]\big) = F(b) - F(a)$ for all $a, b \in \mathbb{R}$ with $a < b$;

(ii) F is non-decreasing, i.e.,

$$\big\langle a, b \in \mathbb{R} \text{ with } a \leq b \big\rangle \Longrightarrow F(a) \leq F(b);$$

(iii) F is continuous from the right, i.e.,

$$\big\langle (b_n) \in \mathbb{R}^\infty \text{ with } b_n \downarrow b \big\rangle \Longrightarrow F(b_n) \to F(b);$$

(*Note.* $(b_n) \in \mathbb{R}^\infty$ simply means that $\{b_n\}_{n \in \mathbb{N}}$ is a sequence in $\mathbb{R}$.)

(iv) 1. $\lim_{x \to -\infty} F(x) = 0$;
2. $\lim_{x \to +\infty} F(x) = 1$.

(b) Let the function $f\colon \mathbb{R} \to \mathbb{R}$ be defined as

$$f(x) := \begin{cases} c(x+2) & \text{for } -2 \leq x < 0 \\ c(2-x) & \text{for } 0 \leq x < 2 \\ 0 & \text{else} \end{cases}.$$

Determine a real number c such that f is a probability density function.

Solution

(a) (i) Note that $(a,b] = (-\infty, b] \setminus (-\infty, a]$. Then,

$$P\big((a,b]\big) = P\big((-\infty, b] \setminus (-\infty, a]\big) = P\big((-\infty, b]\big) - P\big((-\infty, a]\big) = F(b) - F(a),$$

where the second equality holds by subtractivity of P (with $(-\infty, a] \subseteq (-\infty, b]$ as $a < b$), and the third equality holds by definition of F.

(ii) Let $a, b \in \mathbb{R}$, with $a \leq b$. Then, $(-\infty, a] \subseteq (-\infty, b]$. By monotonicity of P,

$$P\big((-\infty, a]\big) \leq P\big((-\infty, b]\big),$$

which is equivalent to

$$F(a) \leq F(b)$$

by definition of F.

(iii) Let $(b_n) \in \mathbb{R}^\infty$, with $b_n \downarrow b$. Since $b_{n+1} \leq b_n$, $(-\infty, b_{n+1}] \subseteq (-\infty, b_n]$ for $n = 1, 2, \ldots$. Then, by continuity from above of P,

$$\lim_{n\to\infty} P\big((-\infty, b_n]\big) = P\left(\bigcap_{n=1}^{\infty}(-\infty, b_n]\right) = P\big((-\infty, b]\big).$$

where the last equality holds because, as $b_n \to b$, $\bigcap_{n=1}^{\infty}(-\infty, b_n] = (-\infty, b]$. Observing that $P\big((-\infty, b_n]\big) = F(b_n)$ and $P\big((-\infty, b]\big) = F(b)$ by definition of F, the desired result follows.

(iv) 1. Let $(b_n) \in \mathbb{R}^\infty$, with $b_n \downarrow -\infty$. Then, $(b_n, +\infty) \subseteq (b_{n+1}, +\infty)$ for $n = 1, 2, \ldots$ and $\bigcup_{n=1}^{\infty}(b_n, +\infty) = \mathbb{R}$. By continuity from below of P,

$$\lim_{n\to\infty} P\big((b_n, +\infty)\big) = P(\mathbb{R}) = 1. \tag{1}$$

Since $(-\infty, b_n] = \mathbb{R} \setminus (b_n, +\infty)$, Theorem 1.3.(iii) in the Lecture Notes implies

$$P\big((-\infty, b_n]\big) = 1 - P\big((b_n, +\infty)\big). \tag{2}$$

Together, (1) and (2) give

$$\lim_{n\to\infty} P\big((-\infty, b_n]\big) = 0.$$

By definition of F, $P\big((-\infty, b_n]\big) = F(b_n)$. The desired result follows.

2. Let $(b_n) \in \mathbb{R}^\infty$, with $b_n \uparrow +\infty$. Then, $(-\infty, b_n] \subseteq (-\infty, b_{n+1}]$ for $n = 1, 2, \ldots$ and $\bigcup_{n=1}^{\infty}(-\infty, b_n] = \mathbb{R}$. By continuity from below of P,

$$\lim_{n\to\infty} P\big((-\infty, b_n]\big) = P(\mathbb{R}) = 1.$$

By definition of F, $P\big((-\infty, b_n]\big) = F(b_n)$. The desired result follows. ■

(b) The function f is continuous and so Riemann integrable on any interval of the form $[a, b]$, with $-\infty < a \leq b < \infty$. To be a probability density function, f has to satisfy

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1.$$

Note that

$$\begin{aligned}
\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x &= \lim_{n\to\infty}\int_{-n}^{-2} 0\,\mathrm{d}x + \int_{-2}^{0} c(x+2)\,\mathrm{d}x + \int_{0}^{2} c(2-x)\,\mathrm{d}x + \lim_{n\to\infty}\int_{0}^{n} 0\,\mathrm{d}x \\
&= 0 + c\int_{-2}^{0}(x+2)\,\mathrm{d}x + c\int_{0}^{2}(2-x)\,\mathrm{d}x + 0 \\
&= c\left[\frac{x^2}{2} + 2x\right]\Bigg|_{-2}^{0} + c\left[2x - \frac{x^2}{2}\right]\Bigg|_{0}^{2} \\
&= 2c + 2c \\
&= 4c.
\end{aligned}$$

Then,

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1 \iff c = \frac{1}{4}$$

Moreover, when $c = \frac{1}{4}$, $f(x) \geq 0$ for any $x \in \mathbb{R}$. Thus, we conclude that f is a probability density function for $c = \frac{1}{4}$. ■

*Exercise 14

Solve the following problems.

(a) Let $(\Omega_1, \mathcal{A}_1)$ and $(\Omega_2, \mathcal{A}_2)$ be measurable spaces, and $\mathcal{C}$ a class of subsets of Ω_2 such that $\sigma(\mathcal{C}) = \mathcal{A}_2$. Show that a function $f\colon \Omega_1 \to \Omega_2$ is $\mathcal{A}_1$-$\mathcal{A}_2$ measurable if and only if $f^{-1}(\mathcal{C}) \subseteq \mathcal{A}_1$. [*Note.* $f^{-1}(\mathcal{C}) := \{B \in \mathcal{P}(\Omega_1) : B = f^{-1}(C), C \in \mathcal{C}\}$.]

(b) Let $(\Omega, \mathcal{A}, P)$ be a probability space and X a real-valued random variable on $(\Omega, \mathcal{A}, P)$ such that $P(X > 0) > 0$. Show that there exists $\delta > 0$ such that $P(X \geq \delta) > 0$.

Solution

Let $\mathcal{F} \subseteq \mathcal{P}(\Omega_2)$. Define

$$f^{-1}(\mathcal{F}) := \left\{E \in \mathcal{P}(\Omega_1) : E = f^{-1}(F), F \in \mathcal{F}\right\}.$$

Hence, for $\mathcal{E} \subseteq \mathcal{P}(\Omega_1)$, $f^{-1}(\mathcal{F}) \subseteq \mathcal{E}$ means that $f^{-1}(F) \in \mathcal{E}$ for all $F \in \mathcal{F}$.

(a) *Necessity.*

Suppose that $f\colon \Omega_1 \to \Omega_2$ is $\mathcal{A}_1$-$\mathcal{A}_2$ measurable. Then, $f^{-1}(\mathcal{A}_2) \subseteq \mathcal{A}_1$. Since $\mathcal{C} \subseteq \mathcal{A}_2$, $f^{-1}(\mathcal{C}) \subseteq f^{-1}(\mathcal{A}_2)$. Hence, $f^{-1}(\mathcal{C}) \subseteq f^{-1}(\mathcal{A}_2) \subseteq \mathcal{A}_1$, and the desired result follows.

Sufficiency.

Suppose that $f^{-1}(C) \in \mathcal{A}_1$ for all $C \in \mathcal{C}$. We need to show that $f^{-1}(\mathcal{A}_2) \subseteq \mathcal{A}_1$. Define

$$\mathcal{B} := \left\{B \in \mathcal{P}(\Omega_2) : f^{-1}(B) \in \mathcal{A}_1\right\}$$

By assumption, $\mathcal{C} \subseteq \mathcal{B}$, which also implies that $\mathcal{B}$ is nonempty. First, we show that $\mathcal{B}$ is a σ-algebra on Ω_2. Let $B \in \mathcal{B}$. Then, $f^{-1}(B) \in \mathcal{A}_1$ and, since $\mathcal{A}_1$ is a σ-algebra on Ω_1, $\Omega_1 \setminus f^{-1}(B) \in \mathcal{A}_1$. As

$$f^{-1}(\Omega_2 \setminus B) = \Omega_1 \setminus f^{-1}(B),$$

it follows that $\Omega_2 \setminus B \in \mathcal{B}$. Now let $B_1, B_2, \dots \in \mathcal{B}$. Then, $f^{-1}(B_1), f^{-1}(B_2), \dots \in \mathcal{A}_1$ and, as $\mathcal{A}_1$ is a σ-algebra on Ω_1,

$$\bigcup_{n=1}^{\infty} f^{-1}(B_n) \in \mathcal{A}_1.$$

As inverse images behave well with respect to taking unions, we have

$$f^{-1}\left(\bigcup_{n=1}^{\infty} B_n\right) = \bigcup_{n=1}^{\infty} f^{-1}(B_n),$$

and so

$$\bigcup_{n=1}^{\infty} B_n \in \mathcal{B}.$$

We conclude that $\mathcal{B}$ is a σ-algebra on Ω_2, as we claimed.[1] Since $\mathcal{C} \subseteq \mathcal{B}$ and $\mathcal{B}$ is a σ-algebra on Ω_2, it follows that $\sigma(\mathcal{C}) = \mathcal{A}_2 \subseteq \mathcal{B}$, and so $f^{-1}(\mathcal{A}_2) \subseteq f^{-1}(\mathcal{B})$. Since $f^{-1}(\mathcal{B}) \subseteq \mathcal{A}_1$ by construction, the desired result follows. ■

Remark 1. Let $(\Omega, \mathcal{A})$ be a measurable space. In principle, to verify that a function $f: \Omega \to \mathbb{R}^k$ is a random variable on $(\Omega, \mathcal{A})$ we need to show that $f^{-1}(B) \in \mathcal{A}$ for *every* Borel subset B of $\mathbb{R}^k$. This is usually a too difficult task. Exercise 14-(a) tells us that there is a nice short-cut. If we can find a collection $\mathcal{C}$ of subsets of $\mathbb{R}^k$ that generates $\mathcal{B}(\mathbb{R}^k)$ and if we manage to verify that $f^{-1}(C) \in \mathcal{C}$ for every $C \in \mathcal{C}$, we may then conclude that f is an $\mathbb{R}^k$-valued random variable. For instance, if $f^{-1}(O) \in \mathcal{A}$ for every open subset O of $\mathbb{R}^k$, or $f^{-1}(C) \in \mathcal{A}$ for every closed subset C of $\mathbb{R}^k$, then f is an $\mathbb{R}^k$-valued random variable.[2]

Remark 2. (The Good Set Technique) Exercise 14-(a) is an application of the so called "good set technique," an extremely useful tool in measure theory.

The fact that there is often no way of giving an explicit description of a generated σ-algebra is a source of discomfort. In particular, this often makes it difficult to check whether or not all members of a given σ-algebra satisfy a property of interest. There is, however, a method of settling such problems in which we utilize the definition of the "σ-algebra generated by $\mathcal{C}$" directly.

Suppose we wish to verify that all members of a given σ-algebra $\mathcal{A}$ on a nonempty set Ω satisfy a certain property. Let us call any one member of $\mathcal{P}(\Omega)$ that satisfies this property a *good set*, and let

$$\mathcal{G} := \text{ the class of all good sets.}$$

The problem at hand is thus to show that $\mathcal{A} \subseteq \mathcal{G}$. Now suppose we know a bit more about $\mathcal{A}$, namely, we know that it is generated by a nonempty collection $\mathcal{C}$ of subsets of Ω. It is often easy to verify that all members of $\mathcal{C}$ are good sets (otherwise $\mathcal{A} \subseteq \mathcal{G}$ cannot be true anyway). So suppose we proved $\mathcal{C} \subseteq \mathcal{G}$. The point is that this is enough to conclude that all members of $\mathcal{A}$ are good sets, that is, $\mathcal{A} \subseteq \mathcal{G}$, provided that the class $\mathcal{G}$ of all good sets is a σ-algebra on Ω. In sum, "the good set technique" transforms the problem at hand into checking whether or not $\mathcal{G}$ is closed under complementation and taking countable unions.

(b) Set

$$A := \{X > 0\} := \{\omega \in \Omega : X(\omega) > 0\} := X^{-1}\big((0, +\infty)\big)$$

and

$$A_n := \left\{X \geq \frac{1}{n}\right\} := \left\{\omega \in \Omega : X(\omega) \geq \frac{1}{n}\right\} := X^{-1}\left(\left[\frac{1}{n}, +\infty\right)\right) \qquad \text{for all } n \in \mathbb{N}.$$

Since X is a real-valued random variable on $(\Omega, \mathcal{A}, P)$, $A, A_n \in \mathcal{A}$ for all $n \in \mathbb{N}$. Moreover, $A_n \subseteq A_{n+1}$ for all $n \in \mathbb{N}$ holds by construction. Therefore, $\{A_n\}_{n=1}^{\infty}$ is an increasing sequence of events in $\mathcal{A}$. By continuity from below of P, we have

$$\lim_{n\to\infty} P(A_n) = P\left(\bigcup_{n=1}^{\infty} A_n\right). \tag{3}$$

In addition, observe that

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} X^{-1}\left(\left[\frac{1}{n}, +\infty\right)\right) = X^{-1}\left(\bigcup_{n=1}^{\infty}\left[\frac{1}{n}, +\infty\right)\right) = X^{-1}\big((0, +\infty)\big) = A. \tag{4}$$

[1] *Quiz:* Why don't we need to show that $\emptyset \in \mathcal{B}$?
[2] *Quiz:* You see why, right?

Together, (3) and (4) imply that

$$\lim_{n\to\infty} P(A_n) = P(A). \tag{5}$$

As $P(A) > 0$ by assumption, (5) implies that there exists $N \in \mathbb{N}$ such that

$$P(A_n) > 0 \qquad \text{for all } n \in \mathbb{N} \text{ with } n \geq N.$$

In particular, we have $P(A_N) > 0$. Set $\delta := \frac{1}{N}$ to obtain the desired result. ■

Exercise 15

For a probability measure P on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ suppose that its cumulative distribution function F is strictly increasing and continuous. Define on $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P)$ the random variable X as $X(x) := F(x)$ for all $x \in \mathbb{R}$. Show that P^X is the uniform distribution on $[0,1]$ (i.e., a density of P^X is equal to 1 on $[0,1]$ and vanishes outside this interval).

Solution

Let F_X be the cumulative distribution function of the probability measure P^X. For any $b \in (0,1)$, we have

$$\begin{aligned}
F_X(b) &= P^X\big((-\infty, b]\big) && (F_X \text{ is the cdf of } P^X)\\
&= P\big(\{x \in \mathbb{R} : X(x) \leq b\}\big) && (P^X \text{ is the distribution of } X)\\
&= P\big(\{x \in \mathbb{R} : F(x) \leq b\}\big) && (\text{definition of } X)\\
&= P\Big(\big\{x \in \mathbb{R} : F^{-1}(F(x)) \leq F^{-1}(b)\big\}\Big) && (F^{-1} \text{ strictly increasing and } F \text{ continuous})\\
&= P\Big(\big\{x \in \mathbb{R} : x \leq F^{-1}(b)\big\}\Big) && (F \text{ is strictly increasing})\\
&= P\Big(\big(-\infty, F^{-1}(b)\big)\Big) && \\
&= F\big(F^{-1}(b)\big) && (F \text{ is the cdf of P})\\
&= b. && (F \text{ is continuous})
\end{aligned}$$

Note: (i) F^{-1} is strictly increasing because it is the inverse function of a strictly increasing function; (ii) Continuity of F implies that $(0,1) \subseteq F(\mathbb{R})$, so that $F^{-1}(b)$ is well defined for any $b \in (0,1)$ and

$$P\big(\{x \in \mathbb{R} : F(x) \leq b\}\big) = P\big(\{x \in \mathbb{R} : P((-\infty, b] \leq b\}\big) = b.$$

Since $F(\mathbb{R}) \subseteq [0,1]$, for $b < 0$ we have

$$F_X(b) = P\big(\{x \in \mathbb{R} : F(x) \leq b\}\big) = P(\emptyset) = 0,$$

and for $b > 1$

$$F_X(b) = P\big(\{x \in \mathbb{R} : F(x) \leq b\}\big) = P(\mathbb{R}) = 1.$$

Finally, the previous observations, together with continuity of F, imply that $F(0) = 0$ and $F(1) = 1$.[3] Therefore, $F_X\colon \mathbb{R} \to [0,1]$ is given by

$$F_X(b) = \begin{cases} 0 & \text{if } b < 0 \\ b & \text{if } 0 \leq b \leq 1 \\ 1 & \text{if } b > 1 \end{cases} .$$

[3] *Quiz:* Can you prove it?

Now, let f_X be a probability density function of P^X. We know that the Riemann integrable function $f_X\colon \mathbb{R} \to [0, \infty)$ has to satisfy

$$F'_X(b) = f_X(b)$$

at any $b \in \mathbb{R}$ where F_X is continuously differentiable. Note that F_X is continuously differentiable in $\mathbb{R} \setminus \{0, 1\}$, with $F'_X(b) = 1$ if $b \in (0, 1)$ and $F'_X(b) = 0$ if $b \in \mathbb{R} \setminus [0, 1]$. Since f_X can take arbitrary finite values at finitely many elements in its domain, we conclude that a density of P^X is

$$f_X(b) = \begin{cases} 0 & \text{if } b < 0 \\ 1 & \text{if } 0 \le b \le 1 \\ 0 & \text{if } b > 1 \end{cases},$$

which gives the desired result. ■

Remark 1. (On the role of continuity of F) Suppose that 0 is the only point of discontinuity of F, with $\lim_{x \to 0^-} F(x) = 1/3$ and $F(0) = 2/3$. Find the distribution P^X.

Remark 2. One application of the result in the previous exercise is in the generation of random samples from a particular distribution. If it is required to generate an observation Y from a population with cdf F_Y, we only need to generate a uniform random number U, between 0 and 1, and solve for y in the equation $F_Y(y) = u$, where u denotes the realization of U. From the computational viewpoint, this method is often not too efficient, but it is generally applicable.

Exercise 16

Let T be an arbitrary index set, and suppose that $\{X_t\}_{t \in T}$ is a collection of independent real-valued random variables on a common probability space $(\Omega, \mathcal{A}, P)$. Moreover, suppose that $f_t\colon \mathbb{R} \to \mathbb{R}$, $t \in T$, are $\mathcal{B}(\mathbb{R})$-$\mathcal{B}(\mathbb{R})$ measurable functions. Show that $\{f_t(X_t)\}_{t \in T}$ is a collection of independent real-valued random variables on $(\Omega, \mathcal{A}, P)$.

Solution

For any $t \in T$, define $g_t := f_t \circ X_t$. First, we show that $\{g_t\}_{t \in T}$ is a collection of real-valued random variables on $(\Omega, \mathcal{A}, P)$. Let $B \in \mathcal{B}(\mathbb{R})$. Since f_t is $\mathcal{B}(\mathbb{R})$-$\mathcal{B}(\mathbb{R})$ measurable by assumption, $f_t^{-1}(B) \in \mathcal{B}(\mathbb{R})$. Since X_t is $\mathcal{A}$-$\mathcal{B}(\mathbb{R})$ measurable by assumption, $X_t^{-1}\big(f_t^{-1}(B)\big) \in \mathcal{A}$. Then,

$$g_t^{-1}(B) = X_t^{-1}\big(f_t^{-1}(B)\big) \in \mathcal{A}.$$

We now establish independence. Consider any nonempty and finite index set $I_0 \subseteq T$ and, for each $t \in I_0$, let $Q_t \in \mathcal{B}(\mathbb{R})$ be arbitrarily chosen. Then

$$\begin{aligned} P\left(\bigcap_{t \in I_0} g_t^{-1}(Q_t)\right) &= P\left(\bigcap_{t \in I_0} X_t^{-1}\big(f_t^{-1}(Q_t)\big)\right) \\ &= \prod_{t \in I_0} P\big(X_t^{-1}\big(f_t^{-1}(Q_t)\big)\big) \\ &= \prod_{t \in I_0} P\big(g_t^{-1}(Q_t)\big), \end{aligned}$$

where the second equality holds because $\{X_t\}_{t \in T}$ is a collection of independent real-valued random variables on $(\Omega, \mathcal{A}, P)$ and $f_t^{-1}(Q_t) \in \mathcal{B}(\mathbb{R})$ because each f_t is $\mathcal{B}(\mathbb{R})$-$\mathcal{B}(\mathbb{R})$ measurable. The desired result follows. ■

Exercise 17

Consider the measurable space $([0, 1], \mathcal{B}([0, 1]))$, where $\mathcal{B}([0, 1])$ is the Borel σ-field on $[0, 1]$, and let $P_{\mathcal{L}}$ be the Lebesgue probability measure on $([0, 1], \mathcal{B}([0, 1]))$. Let Q be the set of all rational numbers in $[0, 1]$ and I the set of all irrational numbers in $[0, 1]$.

(i) Which value does $P_{\mathcal{L}}(Q)$ take?

(ii) A friend of yours, say Claudio, used the following way to compute the Lebesgue probability measure of irrationals in $[0,1]$:
$$P_{\mathcal{L}}(I) = \sum_{x \in I} P_{\mathcal{L}}\big(\{x\}\big).$$
Since the Lebesgue probability measure of $\{x\}$ is zero, he concludes that $P_{\mathcal{L}}(I) = 0$. Is Claudio's calculation right? If yes, explain why; if not, explain why and find the correct $P_{\mathcal{L}}(I)$.

Solution

We need first to understand what the probability space $([0,1], \mathcal{B}([0,1]), P_{\mathcal{L}})$ is. Roughly speaking, the Lebesgue probability measure $P_{\mathcal{L}}$ on $([0,1], \mathcal{B}([0,1]))$ is the probability measure on $([0,1], \mathcal{B}([0,1]))$ which on the intervals coincides with their length.

One way to go is to notice that
$$\mathcal{B}([0,1]) = \{S \in \mathcal{P}([0,1]) : S \in \mathcal{B}(\mathbb{R})\}. \tag{6}$$
Then, we can define $P_{\mathcal{L}}$ as the restriction of λ_1 to $\mathcal{B}([0,1])$, where λ_1 is the Lebesgue measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. To proceed this way, we need to verify that the equality among classes in (6) holds true (see Ok (2016), Chapter B, Exercise 1.20), and that $P_{\mathcal{L}}$ defines a probability measure on $([0,1], \mathcal{B}([0,1]))$.

Alternatively, we might construct the probability space $([0,1], \mathcal{B}([0,1]), P_{\mathcal{L}})$. Define
$$\mathcal{A} := \text{ set of all intervals of the form } (a,b], \text{ with } 0 \le a \le b \le 1,$$
and
$$\mathcal{A}^* := \mathcal{A} \cup \{0\}.$$
Moreover, let $\mathcal{B}([0,1]) := \sigma\big(\mathcal{A}^*\big)$ be the Borel σ-field on $[0,1]$. Clearly, every singleton, and hence every countable, subset of $[0,1]$ belongs to $\mathcal{B}([0,1])$.[4]

On the semialgebra[5] $\mathcal{A}^*$, define the probability pre-measure $P_{\mathcal{L}} \colon \mathcal{A}^* \to [0,1]$ as
$$P_{\mathcal{L}}\big((a,b]\big) := b - a,$$
with $P_{\mathcal{L}}(\{0\}) := 0$ and $P_{\mathcal{L}}\big([0,1]\big) := 1$. It can be shown that $P_{\mathcal{L}}$ satisfies σ-additivity.[6] Thus, Carathéodory's Extension Theorem[7] tells us that $P_{\mathcal{L}}$ can be extended uniquely to a probability measure on $\mathcal{B}([0,1])$. This is the Lebesgue probability measure on $([0,1], \mathcal{B}([0,1]))$.[8]

The Lebesgue probability measure of a singleton set is zero. Indeed, for or any $x \in (0,1]$, we have
$$P_{\mathcal{L}}\big(\{x\}\big) = P_{\mathcal{L}}\left(\bigcap_{n=k}^{\infty}\left(x - \frac{1}{n}, x\right]\right) = \lim_{n\to\infty} P_{\mathcal{L}}\left(\left(x - \frac{1}{n}, x\right]\right) = \lim_{n\to\infty}\frac{1}{n} = 0, \tag{7}$$
where $k := \min\{m \in \mathbb{N} : m > 1/x\}$, while $\mathbb{P}_{\mathcal{L}}\big(\{0\}\big) = 0$ holds by definition of $P_{\mathcal{L}}$. Note that the second equality in (7) holds by continuity from above of $P_{\mathcal{L}}$.

We are now ready to answer the two questions.

[4] *Quiz:* Why? Make sure you are able to prove it.

[5] *Definition.* Let Ω be a nonempty set and $\mathcal{S}$ a class of subsets of Ω. We say that $\mathcal{S}$ is a *semialgebra* on Ω if (i) both $\emptyset$ and Ω belong to $\mathcal{S}$; (ii) $\mathcal{S}$ is closed under taking finite intersections, and (iii) for any $S \in \mathcal{S}$, the set $\Omega \setminus S$ can be written as the union of a collection of finitely many pairwise disjoint elements of S. *Remark.* An algebra on Ω is a semialgebra on Ω. *Quiz:* Is $\mathcal{A}^*$ an algebra? Make sure you understand why the set of all intervals of the form (a,b), with $0 \le a < b \le 1$ is not a semialgebra on $[0,1]$.

[6] See Ok (2016), Chapter B, pages 39-42, to get an idea.

[7] For a statement of Carathéodory's Extension Theorem, see Ok (2016), Chapter B, page 32. The statement in Ok (2016) is slightly more general than the one in the Lecture Notes, and it suits better the present framework.

[8] *Quiz:* Let $\tilde{\mathcal{A}} :=$ set of all intervals of the form (a,b), with $0 \le a < b \le 1$. Clearly, $\sigma\big(\tilde{\mathcal{A}}\big) = \mathcal{B}([0,1])$. Can you see why I did not use $\tilde{\mathcal{A}}$ to generate $\mathcal{B}([0,1])$?

(i) Since Q is a countably infinite set, it can be written as the countably infinite union of the pairwise disjoint sets in

$$\big\{\{q\} \in \mathcal{P}([0,1]) : q \in Q\big\}.$$

That is,

$$Q = \bigcup_{q \in Q} \{q\}.$$

Since $P_{\mathcal{L}}$ is a probability measure on $([0,1], \mathcal{B}([0,1]))$, by σ-additivity we have

$$P_{\mathcal{L}}(Q) = P_{\mathcal{L}}\left(\bigcup_{q \in Q} \{q\}\right) = \sum_{q \in Q} P_{\mathcal{L}}\big(\{q\}\big). \tag{8}$$

Observing that $P_{\mathcal{L}}\big(\{q\}\big) = 0$ for all $q \in Q$, we conclude from (8) that

$$P_{\mathcal{L}}(Q) = 0.$$

(ii) Since I can be written as the union of the pairwise disjoint sets in

$$\big\{\{x\} \in \mathcal{P}([0,1]) : x \in I\big\},$$

that is as

$$I = \bigcup_{x \in I} \{x\},$$

Claudio thinks that, by σ-additivity of $P_{\mathcal{L}}$,

$$P_{\mathcal{L}}(I) = \sum_{x \in I} P_{\mathcal{L}}\big(\{x\}\big).$$

Since

$$\sum_{x \in I} P_{\mathcal{L}}\big(\{x\}\big) := \sup\left\{\sum_{x \in \tilde{I}} P_{\mathcal{L}}\big(\{x\}\big) : \tilde{I} \subseteq I \text{ and } \big|\tilde{I}\big| < \infty\right\}$$

and the Lebesgue probability measure of a singleton set is zero, he concludes that $P_{\mathcal{L}}(I) = 0$. Claudio's reasoning is wrong because the set I is uncountable, while σ-additivity of probability measures only holds for countably infinite unions of pairwise disjoint sets. Since $I := [0,1] \setminus Q \in \mathcal{B}([0,1])$, we have

$$P_{\mathcal{L}}(I) = P_{\mathcal{L}}\big([0,1] \setminus Q\big) = P_{\mathcal{L}}\big([0,1]\big) - P_{\mathcal{L}}(Q) = 1 - 0 = 1,$$

where the second equality holds by Theorem 1.3.(iii) in the Lecture Notes, and the third equality holds because $P_{\mathcal{L}}([0,1]) := 1$ and $P_{\mathcal{L}}(Q) = 0$ by part (i). ■

Exercise 18

Solve the following problems.

(a) Let X be a geometrically distributed random variable with parameter $\pi \in (0,1)$, $X \sim Geo(\pi)$.

(i) Show that the cumulative distribution function is

$$F(x) = 1 - (1-\pi)^{\lfloor x \rfloor}, \quad x \in [1, \infty),$$

where $\lfloor x \rfloor$ is the largest integer smaller than or equal to x.

(ii) Show that

$$E[X] = \frac{1}{\pi}.$$

(b) Let $X_1, \ldots, X_n$ be i.i.d. random variables with $X_i \sim Exp(\lambda)$, $\lambda > 0$, for all i. Find the distribution of the random variable m_n defined as

$$m_n := \min_{1 \le i \le n} X_i.$$

Solution

(a) Let $X \sim Geo(\pi)$, with $\pi \in (0,1)$. The random variable X is discrete, with support $\mathcal{S}_X = \mathbb{N}$. Its distribution takes values $P^X(\{i\}) = (1-\pi)^{i-1}\pi$ at each $i \in \mathcal{S}_X$.

(i) Let $x \in \mathcal{S}_X$. We have

$$\begin{aligned} P^X(X > x) &= P^X(X \ge x+1) \\ &= \sum_{i=x+1}^{\infty} (1-\pi)^{i-1}\pi \\ &= \pi \sum_{i=x+1}^{\infty} (1-\pi)^{i-1} \\ &= \pi \left(\frac{(1-\pi)^x}{\pi} \right) \\ &= (1-\pi)^x \end{aligned}$$

Then, at any $x \in \mathcal{S}_X$,

$$F(x) = P^X(X \le x) = 1 - P^X(X > x) = 1 - (1-\pi)^x.$$

The desired result follows by noting that $P^X((a,b)) = 0$ if $(a,b) \cap \mathcal{S}_X = \emptyset$. ■

(ii) We have

$$\begin{aligned} E[X] &:= \sum_{i \in \mathcal{S}_X} P^X(\{i\}) \\ &= \sum_{i=1}^{\infty} i(1-\pi)^{i-1}\pi \\ &= \pi \sum_{i=1}^{\infty} i(1-\pi)^{i-1} \\ &= \pi \left[\sum_{i=1}^{\infty} \left(-\frac{d}{d\pi}(1-\pi)^i \right) \right] \\ &= \pi \left[\frac{d}{d\pi} \left(-\sum_{i=1}^{\infty} (1-\pi)^i \right) \right] \\ &= \pi \frac{d}{d\pi} \frac{1+\pi}{\pi} \\ &= \frac{1}{\pi}, \end{aligned}$$

where the interchange of summation and differentiation is justified because convergent power series converge uniformly on compact subsets of the set of points where they converge. ■

(b) Let P^{m_n} be the distribution of m_n and P^X the distribution of $X_1, \ldots, X_n$. Fix $t \in \mathbb{R}$. Note that

$$m_n > t \iff \langle X_1 > t \wedge X_2 > t \wedge \cdots \wedge X_n > t \rangle.$$

Therefore, for $t \geq 0$,

$$\begin{aligned} P^{m_n}(m_n > t) &= P^{m_n}(X_1 > t, X_2 > t, \ldots, X_n > t) \\ &= P^X(X_1 > t) P^X(X_2 > t) \cdots P^X(X_n > t) \\ &= e^{-\lambda t} e^{-\lambda t} \cdots e^{-\lambda t} \\ &= e^{-n\lambda t}, \end{aligned}$$

where the second equality holds because $X_1, \ldots, X_n$ are i.i.d. random variables, and the third equality holds because $X_i \sim Exp(\lambda)$ for $i = 1, 2, \ldots, n$, and $P^{m_n}(m_n > t) = 1$ if $t < 0$. Thus,

$$F_{m_n}(t) := P^{m_n}(m_n \leq t) = \begin{cases} 1 - e^{-n\lambda t} & \text{if } t \geq 0 \\ 0 & \text{if } t < 0 \end{cases}.$$

We conclude that $m_n \sim Exp(n\lambda)$. ■

E 703: Advanced Econometrics I
Solutions to Problem Set 4

Niccolò Lomys

November 2, 2016

Exercise 19

Let X be a continuous real-valued random variable on a probability space $(\Omega, \mathcal{A}, P)$ with pdf $f_X\colon \mathbb{R} \to [0, +\infty)$ defined by

$$f_X(x) := \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Find a pdf f_Y of the random variable Y defined as $Y := X^2$.

Solution

Let $g(X) := X$. In this case, $g'(x) = 2x$, which is > 0 for $x > 0$ and < 0 for $x < 0$, so g is not invertible on $\mathbb{R}$ and the density transformation formula cannot be applied directly. However, we can proceed as follows. Let F_Y be the cdf of Y. For $y > 0$,

$$F_Y(y) := P\big(Y \leq y\big) = P\big(-\sqrt{y} \leq X \leq \sqrt{y}\big) = F_X\big(\sqrt{y}\big) - F_X\big(-\sqrt{y}\big),$$

where F_X is the cdf of X. Moreover, Y is nonnegative by definition and so $F_Y(y) = 0$ for $y < 0$, Since the cdf of Y is continuously differentiable on $\mathbb{R} \setminus \{0\}$, a pdf $f_Y\colon \mathbb{R} \to [0, \infty)$ of Y satisfies $f_Y(y) = F_Y'(y)$ at any $y \in \mathbb{R} \setminus \{0\}$ and can take an arbitrary finite value at $y = 0$. Therefore,

$$f_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}} \left[f_X\big(\sqrt{y}\big) + f_X\big(-\sqrt{y}\big) \right] & \text{if } y > 0 \\ 0 & \text{if } y \leq 0 \end{cases},$$

where we set $f_Y(0) := 0$. It follows that

$$f_Y(y) = \begin{cases} \frac{1}{2\sqrt{\pi y}} e^{-y/2} & \text{if } y > 0 \\ 0 & \text{if } y \leq 0 \end{cases}. \qquad \blacksquare$$

Exercise 20

Let X and Y be real-valued random variables on some probability space $(\Omega, \mathcal{A}, P)$ with joint pdf $f_{XY}\colon \mathbb{R}^2 \to [0, \infty)$ defined as

$$f_{XY}(x, y) := \begin{cases} c\big(x^2 + xy\big) & \text{for } (x, y) \in [0, 1] \times [0, 1] \\ 0 & \text{otherwise} \end{cases},$$

where c is a positive real constant to be determined.

(i) Determine c such that f_{XY} is a pdf, and find the joint cdf F_{XY}.

(ii) Find marginal pdf's f_X and f_Y and the marginal cdf's F_X and F_Y.

(iii) Find $E(X)$ and $\mathrm{Var}(X)$.

(iv) Find the covariance between X and Y.

Solution

The function f_{XY} is piecewise-continuous, hence Riemann integrable. Therefore, the integrals below are well defined.

(i) We have

$$\begin{aligned}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{XY}(x,y)\,\mathrm{d}x\mathrm{d}y &= \int_0^1\int_0^1 c\big(x^2+xy\big)\,\mathrm{d}y\mathrm{d}x\\ &= c\int_0^1\Big(x^2+\frac{x}{2}\Big)\,\mathrm{d}x\\ &= c\Big(\frac{1}{3}+\frac{1}{4}\Big)\\ &= \frac{7c}{12}.\end{aligned}$$

Then,

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{XY}(x,y)\,\mathrm{d}x\mathrm{d}y = 1 \iff c=\frac{12}{7}$$

Moreover, when $c=\frac{12}{7}$, $f_{XY}(x,y)$ is non-negative. Thus, f_{XY} is a probability density function for $c=\frac{12}{7}$.

To find the joint cdf F_{XY}, consider first $(x,y)\in[0,1]\times[0,1]$. We have

$$\begin{aligned}F_{XY}(x,y) &:= P(X\le x, Y\le y)\\ &= \int_{-\infty}^{x}\int_{-\infty}^{y} f_{XY}(u,v)\,\mathrm{d}u\mathrm{d}v\\ &= \frac{12}{7}\int_0^x\int_0^y\big(u^2+uv\big)\,\mathrm{d}v\mathrm{d}u\\ &= \frac{12}{7}\int_0^x\Big(u^2y+\frac{uy^2}{u}\Big)\,\mathrm{d}u\\ &= \frac{12}{7}\Big(\frac{x^3y}{3}+\frac{x^2y^2}{4}\Big).\end{aligned}$$

Then, $F_{XY}\colon\mathbb{R}^2\to[0,1]$ is given by

$$F_{XY}(x,y)=\begin{cases}0 & \text{if } x<01 \text{ or } y<1\\ \frac{12}{7}\Big(\frac{x^3y}{3}+\frac{x^2y^2}{4}\Big) & \text{if } (x,y)\in[0,1]\times[0,1]\\ \frac{12}{7}\Big(\frac{x^3}{3}+\frac{x^2}{4}\Big) & \text{if } x\in[0,1] \text{ and } y>1\\ \frac{12}{7}\Big(\frac{y}{3}+\frac{y^2}{4}\Big) & \text{if } y\in[0,1] \text{ and } x>1\\ 1 & \text{if } x>1 \text{ and } y>1\end{cases}.$$

(ii) We first find the marginal pdf's. For $x,y\in[0,1]$, we have

$$\begin{aligned}f_X(x) &= \int_{-\infty}^{\infty} f_{XY}(x,y)\,\mathrm{d}y\\ &= \int_0^1\frac{12}{7}\big(x^2+xy\big)\,\mathrm{d}y\\ &= \frac{12}{7}\Big(x^2+\frac{x}{2}\Big),\end{aligned}$$

and, similarly,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y)\, \mathrm{d}x = \frac{12}{7}\left(\frac{1}{3} + \frac{y}{2}\right).$$

Therefore, a pdf $f_X : \mathbb{R} \to [0, \infty)$ of X is given by

$$f_X(x) := \begin{cases} \frac{12}{7}\left(x^2 + \frac{x}{2}\right) & \text{for } x \in [0, 1] \\ 0 & \text{otherwise} \end{cases},$$

and a pdf $f_Y : \mathbb{R} \to [0, \infty)$ of Y is given by

$$f_Y(y) := \begin{cases} \frac{12}{7}\left(\frac{1}{3} + \frac{y}{2}\right) & \text{for } y \in [0, 1] \\ 0 & \text{otherwise} \end{cases}.$$

We now find the marginal cdf's. For $x, y \in [0, 1]$, we have

$$F_X(x) = \int_{-\infty}^{x} f_X(x)\, \mathrm{d}x = \frac{12}{7}\left(\frac{x^3}{3} + \frac{x^2}{4}\right)$$

and

$$F_Y(y) = \int_{-\infty}^{y} f_Y(y)\, \mathrm{d}y = \frac{12}{7}\left(\frac{y}{3} + \frac{y^2}{4}\right).$$

Therefore, the cdf $F_X : \mathbb{R} \to [0, 1]$ of X is given by

$$F_X(x) := \begin{cases} 0 & \text{for } x < 1 \\ \frac{12}{7}\left(\frac{x^3}{3} + \frac{x^2}{4}\right) & \text{for } x \in [0, 1] \\ 1 & \text{for } x > 1 \end{cases},$$

and the cdf $F_Y : \mathbb{R} \to [0, 1]$ of Y is given by

$$F_Y(y) := \begin{cases} 0 & \text{for } y < 1 \\ \frac{12}{7}\left(\frac{y}{3} + \frac{y^2}{4}\right) & \text{for } y \in [0, 1] \\ 1 & \text{for } y > 1 \end{cases}.$$

(iii) We have

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x f_X(x)\, \mathrm{d}x \\ &= \frac{12}{7}\int_0^1 x\left(x^2 + \frac{x}{2}\right) \mathrm{d}x \\ &= \frac{12}{7}\left(\frac{1}{4} + \frac{1}{6}\right) \\ &= \frac{5}{7}, \end{aligned}$$

and

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f_X(x)\, \mathrm{d}x = \frac{39}{70}.$$

Hence,

$$\mathrm{Var}(x) = E(X^2) - (E(X))^2 = \frac{39}{70} - \frac{25}{49} = \frac{23}{490}.$$

(iv) We have

$$
\begin{aligned}
E(Y) &= \int_{-\infty}^{+\infty} y f_Y(y)\,\mathrm{d}y \\
&= \frac{12}{7}\int_0^1 y\left(\frac{1}{3}+\frac{x}{2}\right)\mathrm{d}y \\
&= \frac{4}{7},
\end{aligned}
$$

and

$$
\begin{aligned}
E(XY) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xy f_{XY}(x,y)\,\mathrm{d}x\mathrm{d}y \\
&= \frac{12}{7}\int_0^1\int_0^1 xy\big(x^2+xy\big)\,\mathrm{d}x\mathrm{d}y \\
&= \frac{17}{42}.
\end{aligned}
$$

Thus,

$$
Cov(X,Y) = E(XY) - E(X)E(Y) = \frac{17}{42} - \frac{5}{7}\frac{4}{7} = \frac{17}{42} - \frac{20}{49} = -\frac{1}{294}. \qquad \blacksquare
$$

Expectation: A Brief Review

Let X be a real-valued random variable on some probability space $(\Omega, \mathcal{A}, P)$. We introduce the expectation of X, denoted by $E[X]$, in three steps.

1. Discrete Random Variables

Let $X\colon \Omega \to \mathbb{R}$ be a discrete random variable with range

$$
\mathcal{R}_X := \{x \in \mathbb{R} : x = X(\omega) \text{ for some } \omega \in \Omega\}.
$$

(i) If $\mathcal{R}_X$ is finite, the expectation of X is defined as

$$
E[X] := \sum_{x\in\mathcal{R}_X} xP(X=x) = \sum_{x\in\mathcal{R}_X} xP\big(\{\omega\in\Omega : X(\omega)=x\}\big) = \sum_{x\in\mathcal{R}_X} xP^X\big(\{x\}\big),
$$

where:

– The first equality holds because

$$
P(X=x) := P\big(\{\omega\in\Omega : X(\omega)=x\}\big);
$$

– The second equality holds because

$$
P\big(\{\omega\in\Omega : X(\omega)=x\}\big) := P^X\big(\{x\}\big).
$$

Note that $E[X]$ is finite.

(ii) If $\mathcal{R}_X$ is countably infinite, we distinguish two cases:

1. If $\mathcal{R}_X \subseteq [0, +\infty)$, the expectation of X is defined as

$$E[X] := \sum_{x \in \mathcal{R}_X} xP(X = x) = \sum_{x \in \mathcal{R}_X} xP\big(\{\omega \in \Omega : X(\omega) = x\}\big) = \sum_{x \in \mathcal{R}_X} xP^X\big(\{x\}\big),$$

where:

  * The first equality holds because

$$P(X = x) := P\big(\{\omega \in \Omega : X(\omega) = x\}\big);$$

  * The second equality holds because

$$P\big(\{\omega \in \Omega : X(\omega) = x\}\big) := P^X\big(\{x\}\big).$$

Note that $E[X]$ can be either finite or infinite.

2. If $\mathcal{R}_X$ is arbitrary, the expectation of X is defined as

$$E[X] := E\big[X^+\big] - E\big[X^-\big],$$

provided that $E\big[X^+\big]$ and $E\big[X^-\big]$ are not both infinite. If $E\big[X^+\big] = +\infty = E\big[X^-\big]$, we say that the expectation of X does not exist. Note that $E[X]$ can be either finite or infinite.

If $E[X]$ exists and is finite, we write $X \in L^1_d(\Omega, \mathcal{A}, P)$.

2. Non-Negative Random Variables

Let $X\colon \Omega \to \mathbb{R}$ be a non-negative random variable (i.e., with values in $[0, +\infty)$). For each $n \in \mathbb{N}$ define

$$D_n(X)(\omega) := \frac{k}{n} \qquad \text{if} \quad \frac{k}{n} \leq X(\omega) < \frac{k+1}{n}$$

for $k \in \mathbb{N} \cup \{0\}$. For each n, $D_n(X)$ is a discrete non-negative random variable on $(\Omega, \mathcal{A}, P)$ with range $\mathcal{R}_{D_n(X)} \subseteq \mathbb{N} \cup \{0\}$. The expectation of X is defined as

$$E[X] := \lim_{n \to \infty} E\big[D_n(X)\big] = \lim_{n \to \infty} \sum_{k \in \mathcal{R}_{D_n(X)}} \frac{k}{n} P\left(D_n(X) = \frac{k}{n}\right).$$

Note that $E[X]$ can be either finite or infinite. If $E[X]$ exists and is finite, we write $X \in L^1_{nn}(\Omega, \mathcal{A}, P)$.

3. General Random Variables

Let $X\colon \Omega \to \mathbb{R}$ be an arbitrary random variable. The expectation of X is defined as

$$E[X] := E\big[X^+\big] - E\big[X^-\big],$$

provided that $E\big[X^+\big]$ and $E\big[X^-\big]$ are not both infinite. If $E\big[X^+\big] = +\infty = E\big[X^-\big]$, we say that the expectation of X does not exist. Note that $E[X]$ can be either finite or infinite. If $E[X]$ exists and is finite, we write $X \in L^1(\Omega, \mathcal{A}, P)$.

Remark 1. $X \in L^1(\Omega, \mathcal{A}, P)$ if and only if $X^+, X^- \in L^1_{nn}(\Omega, \mathcal{A}, P)$.

Remark 2. $L^1_{d,+}(\Omega, \mathcal{A}, P) \subseteq L^1_{nn}(\Omega, \mathcal{A}, P) \subseteq L^1(\Omega, \mathcal{A}, P)$, where

$$L^1_{d,+}(\Omega, \mathcal{A}, P) := \left\{X \in L^1_d(\Omega, \mathcal{A}, P) : X \geq 0\right\},$$

and $L^1_d(\Omega, \mathcal{A}, P) \subseteq L^1(\Omega, \mathcal{A}, P)$.

Remark 3. $X \in L^1(\Omega, \mathcal{A}, P)$ if and only if $|X| \in L^1(\Omega, \mathcal{A}, P)$.

Proof. Suppose $X \in L^1(\Omega, \mathcal{A}, P)$. Then, $X^+, X^- \in L^1(\Omega, \mathcal{A}, P)$. Since $|X| = X^+ + X^-$, by linearity of expectation we have $E[|X|] = E\left[X^+\right] + E\left[X^-\right] < +\infty$, and so $|X| \in L^1(\Omega, \mathcal{A}, P)$. Now suppose that $|X| \in L^1(\Omega, \mathcal{A}, P)$. Then, $E[X^+ + X^-] < \infty$. Since $E[X^+ + X^-] = E\left[X^+\right] + E\left[X^-\right]$ and both terms are non-negative, we have that they are also both finite.[1] Thus, $X \in L^1(\Omega, \mathcal{A}, P)$. ■

Remark 4. Together, the previous remarks say that

$$X \in L^1(\Omega, \mathcal{A}, P) \iff X^+, X^- \in L^1(\Omega, \mathcal{A}, P) \iff |X| \in L^1(\Omega, \mathcal{A}, P).$$

Similarly, we have

$$X \in L^1_d(\Omega, \mathcal{A}, P) \iff X^+, X^- \in L^1_d(\Omega, \mathcal{A}, P) \iff |X| \in L^1_d(\Omega, \mathcal{A}, P).$$

*Exercise 21

Let X, Y, Z be real-valued random variables on some probability space $(\Omega, \mathcal{A}, P)$, and suppose that $X, Y \in L^1(\Omega, \mathcal{A}, P)$. Prove the following statements.

(i) If $X \le Z \le Y$, then $Z \in L^1(\Omega, \mathcal{A}, P)$, and

$$E[X] \le E[Z] \le E[Y].$$

(ii) $|E[X]| \le E[|X|]$.

[*Note.* When you are asked to prove a statement from the Lecture Notes, you can use all the results that come before (but not after) that statement in the Lecture Notes.]

Solution

(i) *Step 1.* X, Y, Z discrete and non-negative random variables.

Assume that $X, Y \in L^1_d(\Omega, \mathcal{A}, P)$, and that

$$X(\omega) \le Z(\omega) \le Y(\omega) \qquad \text{for all } \omega \in \Omega. \tag{1}$$

We want to show that $Z \in L^1_d(\Omega, \mathcal{A}, P)$, and that

$$E[X] \le E[Z] \le E[Y].$$

Define

$$\Omega_y := \{\omega \in \Omega : Y(\omega) = y\} \qquad \text{for all } y \in \mathcal{R}_Y,$$

and

$$\Omega_z := \{\omega \in \Omega : Z(\omega) = z\} \qquad \text{for all } z \in \mathcal{R}_Z.$$

Note that $\{\Omega_y\}_{y \in \mathcal{R}_Y}$ and $\{\Omega_z\}_{z \in \mathcal{R}_Z}$ are countable partitions of Ω. Consider the finer partition of Ω, $\{\Omega_{yz}\}_{y \in \mathcal{R}_Y, z \in \mathcal{R}_Z}$, where

$$\Omega_{yz} := \Omega_y \cap \Omega_z \qquad \text{for all } y \in \mathcal{R}_Y \text{ and } z \in \mathcal{R}_Z.$$

[1] To say that $E[X^+ + X^-] = E\left[X^+\right] + E\left[X^-\right]$ I use that additivity of expectation of nonnegative random variables also holds if the corresponding expectations are infinite.

We have

$$
\begin{aligned}
E[Z] &:= \sum_{z \in \mathcal{R}_Z} zP(\Omega_z) \\
&= \sum_{z \in \mathcal{R}_Z} zP(\Omega_z \cap \Omega)
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{z \in \mathcal{R}_Z} zP\left(\Omega_z \cap \left(\bigcup_{y \in \mathcal{S}_Y} \Omega_y\right)\right) \\
&= \sum_{z \in \mathcal{R}_Z} zP\left(\bigcup_{y \in \mathcal{S}_Y} (\Omega_z \cap \Omega_y)\right) \\
&= \sum_{z \in \mathcal{R}_Z} zP\left(\bigcup_{y \in \mathcal{S}_Y} \Omega_{yz}\right) \\
&= \sum_{z \in \mathcal{R}_Z} z \sum_{y \in \mathcal{R}_Y} P(\Omega_{yz}) \\
&= \sum_{z \in \mathcal{R}_Z} \sum_{y \in \mathcal{R}_Y} zP(\Omega_{yz}) \\
&= \sum_{y \in \mathcal{R}_Y} \sum_{z \in \mathcal{R}_Z} zP(\Omega_{yz}),
\end{aligned}
$$

where: the fifth equality holds by σ-additivity of P; the last equality follows by the discrete version of Tonelli's Theorem (cf. Ok (2016), Appendix 1, page 9) and the fact that $zP(\Omega_{yz}) \geq 0$ for all y and z. Similarly, we find

$$E[Y] = \sum_{y \in \mathcal{R}_Y} \sum_{z \in \mathcal{R}_Z} yP(\Omega_{yz}).$$

Then,

$$E[Z] = \sum_{y \in \mathcal{R}_Y} \sum_{z \in \mathcal{R}_Z} zP(\Omega_{yz}) \leq \sum_{y \in \mathcal{R}_Y} \sum_{z \in \mathcal{R}_Z} yP(\Omega_{yz}) = E[Y],$$

where the inequality holds because (1) implies that $z \leq y$ whenever $\Omega_{yz} \neq \emptyset$. That is, we have

$$E[Z] \leq E[Y] < +\infty, \tag{2}$$

where the strict inequality holds because $Y \in L^1_d(\Omega, \mathcal{A}, P)$. Moreover, as the sum of non-negative terms is non-negative, we have

$$0 \leq E[Z]. \tag{3}$$

Together, (2) and (3) give that $Z \in L^1_d(\Omega, \mathcal{A}, P)$. That $E[X] \leq E[Z]$ is proven analogously.

Step 2. X, Y, Z discrete random variables.

Assume that $X, Y \in L^1_{nn}(\Omega, \mathcal{A}, P)$, and that

$$X(\omega) \leq Z(\omega) \leq Y(\omega) \qquad \text{for all } \omega \in \Omega. \tag{4}$$

We want to show that $Z \in L^1_{nn}(\Omega, \mathcal{A}, P)$, and that

$$E[X] \leq E[Z] \leq E[Y].$$

Let $D_n(X), D_n(Y), D_n(Z)$ be the discrete approximations of X, Y, Z on a $1/n$ grid, with $n \in \mathbb{N}$. From (4) and the definition of $D_n(X), D_n(Y), D_n(Z)$, we have

$$D_n(X)(\omega) \leq D_n(Z)(\omega) \leq D_n(Y)(\omega) \qquad \text{for all } \omega \in \Omega. \tag{5}$$

Step 1., together with (5), gives

$$E[D_n(X)] \leq E[D_n(Z)] \leq E[D_n(Y)] \qquad \text{for all } n \in \mathbb{N}.$$

Therefore, by the sandwich theorem for limits of sequences,

$$-\infty < E[X] := \lim_{n\to\infty} E[D_n(X)] \leq \lim_{n\to\infty} E[D_n(Z)] \leq \lim_{n\to\infty} E[D_n(Y)] := E[Y] < +\infty,$$

where the strict inequalities hold because $X, Y \in L^1_{nn}(\Omega, \mathcal{A}, P)$. The desired result follows noting that

$$E[Z] := \lim_{n\to\infty} E[D_n(Z)].$$

Step 3. X, Y, Z arbitrary random variables.
Assume that $X, Y \in L^1(\Omega, \mathcal{A}, P)$, and that

$$X(\omega) \leq Z(\omega) \leq Y(\omega) \qquad \text{for all } \omega \in \Omega. \tag{6}$$

We want to show that $Z \in L^1(\Omega, \mathcal{A}, P)$, and that

$$E[X] \leq E[Z] \leq E[Y].$$

From (6), we have

$$X^+(\omega) \leq Z^+(\omega) \leq Y^+(\omega) \qquad \text{for all } \omega \in \Omega. \tag{7}$$

and

$$Y^-(\omega) \leq Z^-(\omega) \leq X^-(\omega) \qquad \text{for all } \omega \in \Omega. \tag{8}$$

Step 2., together with (7) and (8), gives

$$E[X^+] \leq E[Z^+] \leq E[Y^+],$$

and

$$E[Y^-] \leq E[Z^-] \leq E[X^-].$$

Therefore,

$$-\infty < E[X] := E[X^+] - E[X^-] \leq E[Z^+] - E[Z^-] := E[Z],$$

and

$$E[Z] := E[Z^+] - E[Z^-] \leq E[Y^+] - E[Y^-] := E[Y] < +\infty,$$

where the strict inequalities hold because $X, Y \in L^1(\Omega, \mathcal{A}, P)$. The desired result follows. ∎

(ii) Since $X \in L^1(\Omega, \mathcal{A}, P)$, $X^+, X^-, |X| \in L^1(\Omega, \mathcal{A}, P)$ and $|E[X]|$ is a non-negative real number. Then, we have

$$\begin{aligned}
|E[X]| &= \left|E\left[X^+\right] - E\left[X^-\right]\right| \\
&\leq \left|E\left[X^+\right]\right| + \left|E\left[X^-\right]\right| \\
&= E\left[X^+\right] + E\left[X^-\right] \\
&= E\left[X^+ + X^-\right] \\
&= E[|X|],
\end{aligned}$$

where: (i) the first equality holds by definition of $E[X]$; (ii) the inequality holds by triangle inequality; (iii) the second equality holds because X^+ and X^- are non-negative, and so $E\big[X^+\big]$ and $E\big[X^-\big]$ are non-negative[2]; (iv) the third equality holds by linearity of expectation and $X^+, X^- \in L^1(\Omega, \mathcal{A}, P)$; (v) the last equality holds because $|X| = X^+ + X^-$. The desired result follows.

Exercise 22

Prove the following statements.

(i) Let X be a non-negative real-valued random variable on some probability space $(\Omega, \mathcal{A}, P)$. Then, $E[X] = 0$ if and only if $P(X = 0) = 1$.

(ii) Let X and Y be real-valued random variables on some probability space $(\Omega, \mathcal{A}, P)$, and suppose that $X, Y \in L^2(\Omega, \mathcal{A}, P)$. Then, $X, Y, X \cdot Y \in L^1(\Omega, \mathcal{A}, P)$.

Solution

(i) *Claim 1.* $P(X = 0) = 1 \Longrightarrow E[X] = 0$.

Step 1. X non-negative and discrete.

If $X(\omega) = 0$ for all $\omega \in \Omega$, then $\Omega_0 := \{\omega \in \Omega = X(\omega) = 0\} = \Omega$ and

$$E[X] := 0 \cdot P(\Omega_0) = 0 \cdot 1 = 0.$$

If $X = 0$ a.s., $P(\Omega_0) = 1$, and $P(\Omega_x) = 0$ for any $x \in \mathcal{R}_X \setminus \{0\}$, where $\Omega_x := \{\omega \in \Omega = X(\omega) = x\}$. Then,

$$E[X] := 0 \cdot P(\Omega_0) + \sum_{x \in \mathcal{R}_X \setminus \{0\}} x P(\Omega_x) = 0 \cdot 1 + 0 = 0.$$

Step 2. X non-negative.

For each $n \in \mathbb{N}$, we have $P(D_n(X) = 0) = 1$. Then, by step 1, $E\big[D_n(X)\big] = 0$, and so

$$E[X] := \lim_{n \to \infty} E\big[D_n(X)\big] = \lim_{n \to \infty} 0 = 0.$$

Claim 2. $E[X] = 0 \Longrightarrow P(X = 0) = 1$.

Step 1. X non-negative and discrete.

By assumption

$$\sum_{x \in \mathcal{R}_X} x P(\Omega_x) = 0.$$

The fact that the sum of non-negative terms can be equal to 0 if and only if all terms are equal to 0 forces one of x and $P(\Omega_x)$ to be equal to zero for every $x \in \mathcal{R}_X \setminus \{0\}$. In particular, we must have $P(\Omega_x) = 0$ for any non-zero x, which shows that $P(\Omega_0) = 1$.[3]

Step 2. X non-negative.

For any $n \in \mathbb{N}$, set

$$\Omega_n := \left\{\omega \in \Omega : X(\omega) \geq \frac{1}{n}\right\}.$$

[2] We will prove that $X \geq 0 \Longrightarrow E[X] \geq 0$ after Exercise 22.
[3] *Quiz:* You see why, right?

Then,

$$\frac{1}{n}\mathbb{1}_{\Omega_n}(\omega) \leq D_n(X)(\omega)\mathbb{1}_{\Omega_n}(\omega) \leq X(\omega),$$

so that, by monotonicity and linearity of expectation,

$$\frac{1}{n}P(\Omega_n) \leq E\big[D_n(X)\mathbb{1}_{\Omega_n}\big] \leq E[X] = 0,$$

which forces $P(\Omega_n) = 0$ for all $n \in \mathbb{N}$. The desired result follows (recall Exercise 14-(b) in Problem Set 3).

(ii) We only need to show that $X, Y \in L^2(\Omega, \mathcal{A}, P)$ implies $X \cdot Y \in L^1(\Omega, \mathcal{A}, P)$. Then, setting $Y(\omega) := 1$ ($X(\omega) := 1$) for all $\omega \in \Omega$ shows that $X \in L^2(\Omega, \mathcal{A}, P)$ ($Y \in L^2(\Omega, \mathcal{A}, P)$) implies $X \in L^1(\Omega, \mathcal{A}, P)$ ($Y \in L^1(\Omega, \mathcal{A}, P)$).

Suppose that $X, Y \in L^2(\Omega, \mathcal{A}, P)$. For any $\omega \in \Omega$, we have

$$0 \leq |X(\omega) \cdot Y(\omega)| \leq \frac{X^2(\omega)}{2} + \frac{Y^2(\omega)}{2}.$$

Then, monotonicity and linearity of the expectation give

$$E[0^v] \leq E[|X \cdot Y|] \leq \frac{1}{2}\Big[E\big[X^2\big] + E\big[Y^2\big]\Big],$$

where $0^v(\omega) := 0$ for all $\omega \in \Omega$. Since $E[0^v] = 0$ by Exercise 22-(i) and $E\big[X^2\big] + E\big[Y^2\big] < +\infty$ because $X^2, Y^2 \in L^1(\Omega, \mathcal{A}, P)$, we have that $|X \cdot Y| \in L^1(\Omega, \mathcal{A}, P)$, which is equivalent to $X \cdot Y \in L^1(\Omega, \mathcal{A}, P)$ from Remark 4. ■

Remark. Monotonicity of expectation and Exercise 22-(i) give

$$\big\langle X \in L^1(\Omega, \mathcal{A}, P) \wedge X \geq 0\big\rangle \Longrightarrow E[X] \geq 0.$$

Exercise 23

A pdf $f_X\colon \mathbb{R} \to [0, +\infty)$ of the standard Cauchy distribution is defined by

$$f_X(x) := \big[\pi\big(1 + x^2\big)\big]^{-1},$$

where π is a strictly positive real constant. Does the expectation of the Cauchy distribution exist? Justify your answer.

Solution

We show that $E\big[X^+\big] = +\infty = E\big[X^-\big]$, and thus the expectation of a Cauchy random variable does not exists.

Note that f is piecewise continuous, and so Riemann integrable. We have

$$\begin{aligned}
E\big[X^+\big] &= \int_{-\infty}^{+\infty} \max\{0, x\} \frac{1}{\pi\big(1 + x^2\big)}\,\mathrm{d}x \\
&= \int_{-\infty}^{0} \frac{0}{\pi\big(1 + x^2\big)}\,\mathrm{d}x + \int_{0}^{+\infty} \frac{x}{\pi\big(1 + x^2\big)}\,\mathrm{d}x \\
&= \frac{1}{2\pi}\ln\big(1 + x^2\big)\Big|_0^{+\infty} \\
&= +\infty - 0 \\
&= +\infty.
\end{aligned}$$

That $E\left[X^-\right] = +\infty$ is shown similarly. ■

*Exercise 24

Let $(X_n)_n$, $(Y_n)_n$, X and Y be $\mathbb{R}^k$-valued random variables on a common probability space $(\Omega, \mathcal{A}, P)$. Prove the following statements:

(i) $\left\langle X_n \xrightarrow{a.s.} X \wedge Y_n \xrightarrow{a.s.} Y \right\rangle \Longrightarrow X_n + Y_n \xrightarrow{a.s.} X + Y$;

(ii) $\left\langle X_n \xrightarrow{P} X \wedge Y_n \xrightarrow{P} Y \right\rangle \Longrightarrow X_n + Y_n \xrightarrow{P} X + Y$.

Solution

(i) Define

$$\Omega_{X+Y} := \left\{\omega \in \Omega : X_n(\omega) + Y_n(\omega) \underset{n\to\infty}{\longrightarrow} X(\omega) + Y(\omega)\right\}.$$

We need to show that $P\left(\Omega_{X+Y}\right) = 1$.

Set

$$\Omega_X := \left\{\omega \in \Omega : X_n(\omega) \underset{n\to\infty}{\longrightarrow} X(\omega)\right\}$$

and

$$\Omega_Y := \left\{\omega \in \Omega : Y_n(\omega) \underset{n\to\infty}{\longrightarrow} Y(\omega)\right\},$$

and note that $\Omega_X \cap \Omega_Y \subseteq \Omega_{X+Y}$, as

$$\left\langle X_n(\omega) \underset{n\to\infty}{\longrightarrow} X(\omega) \wedge Y_n \underset{n\to\infty}{\longrightarrow} Y \right\rangle \Longrightarrow X_n(\omega) + Y_n(\omega) \underset{n\to\infty}{\longrightarrow} X + Y.$$

Then, by monotonicity of P, the claim follows if $P\left(\Omega_X \cap \Omega_Y\right) = 1$ or, equivalently, if $P\left(\Omega \setminus \left(\Omega_X \cap \Omega_Y\right)\right) = 0$ (as $P\left(\Omega \setminus \left(\Omega_X \cap \Omega_Y\right)\right) = 1 - P\left(\Omega_X \cap \Omega_Y\right)$ by Theorem 1.3.(iii)).

Note that

$$\begin{aligned} 0 \leq P\left(\Omega \setminus \left(\Omega_X \cap \Omega_Y\right)\right) &= P\left(\left(\Omega \setminus \Omega_X\right) \cup \left(\Omega \setminus \Omega_Y\right)\right) \\ &\leq P\left(\Omega \setminus \Omega_X\right) + P\left(\Omega \setminus \Omega_Y\right) \\ &= \left(1 - P\left(\Omega_X\right)\right) + \left(1 - P\left(\Omega_Y\right)\right) \\ &= 0, \end{aligned}$$

where: the first inequality holds by non-negativity of P; the first equality holds by De Morgan's laws; the second inequality holds by sub-σ-additivity of P; the second equality holds by Theorem 1.3.(iii); the last equality holds because $X_n \xrightarrow{a.s.} X$ and $Y_n \xrightarrow{a.s.} Y$ by assumption, and so $P\left(\Omega_X\right) = P\left(\Omega_Y\right) = 1$. The desired result follows. ■

(ii) Let $\varepsilon > 0$. We need to show that

$$\lim_{n\to\infty} P\left(\left\{\omega \in \Omega : \left\|\left(X_n(\omega) + Y_n(\omega)\right) - \left(X(\omega) - Y(\omega)\right)\right\| > \varepsilon\right\}\right) = 0.$$

Fix $n \in \mathbb{N}$. By triangle inequality,

$$\left\|\left(X_n(\omega) + Y_n(\omega)\right) - \left(X(\omega) - Y(\omega)\right)\right\| \leq \left\|X_n(\omega) - X(\omega)\right\| + \left\|Y_n(\omega) - Y(\omega)\right\|$$

holds for any $\omega \in \Omega$. Therefore, by monotonicity of P,

$$\begin{aligned} &P\left(\left\{\omega \in \Omega : \left\|\left(X_n(\omega) + Y_n(\omega)\right) - \left(X(\omega) - Y(\omega)\right)\right\| > \varepsilon\right\}\right) \\ &\leq P\left(\left\{\omega \in \Omega : \left\|X_n(\omega) - X(\omega)\right\| + \left\|Y_n(\omega) - Y(\omega)\right\| > \varepsilon\right\}\right). \end{aligned} \tag{9}$$

Moreover,

$$\|X_n(\omega) - X(\omega)\| + \|Y_n(\omega) - Y(\omega)\| > \varepsilon \Longrightarrow \left\langle \|X_n(\omega) - X(\omega)\| > \frac{\varepsilon}{2} \ \vee \ \|Y_n(\omega) - Y(\omega)\| > \frac{\varepsilon}{2} \right\rangle,$$

and so

$$\{\omega \in \Omega : \|X_n(\omega) - X(\omega)\| + \|Y_n(\omega) - Y(\omega)\| > \varepsilon\}$$

is a subset of

$$\{\omega \in \Omega : \|X_n(\omega) - X(\omega)\| > \varepsilon/2\} \cup \{\omega \in \Omega : \|Y_n(\omega) - Y(\omega)\| > \varepsilon/2\}.$$

Therefore, by monotonicity and sub-σ-additivity of P,

$$\begin{aligned} &P\Big(\{\omega \in \Omega : \|X_n(\omega) - X(\omega)\| + \|Y_n(\omega) - Y(\omega)\| > \varepsilon\}\Big) \\ &\le P\Big(\{\omega \in \Omega : \|X_n(\omega) - X(\omega)\| > \varepsilon/2\}\Big) \\ &+ P\Big(\{\omega \in \Omega : \|Y_n(\omega) - Y(\omega)\| > \varepsilon/2\}\Big). \end{aligned} \tag{10}$$

Together with non-negativity of P, (9) and (10) imply

$$\begin{aligned} 0 &\le P\Big(\{\omega \in \Omega : \|\big(X_n(\omega) + Y_n(\omega)\big) - \big(X(\omega) - Y(\omega)\big)\| > \varepsilon\}\Big) \\ &\le P\Big(\{\omega \in \Omega : \|X_n(\omega) - X(\omega)\| > \varepsilon/2\}\Big) \\ &+ P\Big(\{\omega \in \Omega : \|Y_n(\omega) - Y(\omega)\| > \varepsilon/2\}\Big) \end{aligned} \tag{11}$$

for all $n \in \mathbb{N}$. By assumption, $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$, and so the right hand side of (11) converges to zero as $n \to \infty$. The desired result follows by the sandwich theorem for limits of sequences. ■

E 703: Advanced Econometrics I
Solutions to Problem Set 5

Niccolò Lomys

November 10, 2016

Exercise 25

Let X be a real-valued random variable on some probability space $(\Omega, \mathcal{A}, P)$, and let $a, b \in \mathbb{R}$ such that $a \leq b$.

(i) Let X be discrete. Show by definition: If $a \leq X \leq b$, then $X \in L^1_d(\Omega, \mathcal{A}, P)$ and $a \leq E[X] \leq b$. Use only statements up to page 55 in the Lecture Notes.

(ii) Show by definition: If $a \leq X \leq b$, then $X \in L^1(\Omega, \mathcal{A}, P)$ and $a \leq E[X] \leq b$. Use only statements up to page 60 in the Lecture Notes.

(iii) Suppose that $P(a \leq X \leq b) = 1$ and $E[X] = b$. Show that $P(X = b) = 1$.

Solution

(i) Since $-\infty < a \leq b < +\infty$, we only need to show $a \leq E[X] \leq b$. That $X \in L^1_d(\Omega, \mathcal{A}, P)$ follows. We distinguish three cases.

Case 1: $0 \leq a \leq b$.

Let $\mathcal{R}_X$ be the range of X. For all $x \in \mathcal{R}_X$, define $\Omega_x := \{\omega \in \Omega : X(\omega) = x\}$. Note that $\{\Omega_x\}_{x \in \mathcal{R}_X}$ is a countable partition of Ω. Then, we have

$$a = \sum_{x \in \mathcal{R}_X} aP(\Omega_x) \leq \sum_{x \in \mathcal{R}_X} xP(\Omega_x) \leq \sum_{x \in \mathcal{R}_X} bP(\Omega_x) = b, \tag{1}$$

where the two inequalities hold because

$$a \leq X(\omega) \leq b \text{ for all } \omega \in \Omega \iff a \leq x \leq b \text{ for all } x \in \mathcal{R}_X.$$

Since

$$E[X] := \sum_{x \in \mathcal{R}_X} xP(\Omega_x),$$

(1) gives

$$a \leq E[X] \leq b,$$

as desired.

Case 2: $a < 0 \leq b$.

Since $0 \leq X^+ \leq b$ and $0 \leq X^-$, by case 1 we have $E[X^+] \leq b$ and $0 \leq E[X^-]$. But then

$$E[X] := E[X^+] - E[X^-] \leq b. \tag{2}$$

Moreover, since $0 \leq X^- \leq -a$ and $0 \leq X^+$, by case 1 we have $E[X^-] \leq -a$ or, equivalently, $a \leq -E[X^-]$ and $0 \leq E[X^+]$. But then

$$a \leq -E[X^-] + E[X^+] = E[X^+] - E[X^-] := E[X]. \tag{3}$$

The desired result follows from (2) and (3).

Case 3: $a \leq b < 0$.

In this case, $X = -X^-$, $-X = X^-$, and so $E[X] := -E[X^-] = -E[-X]$. Note that $a \leq X \leq b < 0 \iff 0 < -b \leq -X \leq -a$. Then, by case 1,

$$-b \leq E[-X] \leq -a$$

or, equivalently,

$$a \leq E[-X] \leq b,$$

which gives the desired result. ■

(ii) Since $-\infty < a \leq b < +\infty$, we only need to show $a \leq E[X] \leq b$. That $X \in L^1(\Omega, \mathcal{A}, P)$ follows.

Let $D_n(X)$ be the discrete approximation of X on a $1/n$ grid, with $n \in \mathbb{N}$. For each n, we have

$$a - \frac{1}{n} \leq D_n(X) \leq b$$

by the assumption $a \leq X \leq b$ and construction of $D_n(X)$.[1] By part (i) of this exercise,

$$a - \frac{1}{n} \leq E[D_n(X)] \leq b$$

holds for all $n \in \mathbb{N}$, and so

$$a \leq \lim_{n\to\infty} E[D_n(X)] \leq b$$

by the sandwich theorem for limits of sequences (note that $\lim_{n\to\infty}(a - 1/n) = a$). Since $E[X] := \lim_{n\to\infty} E[D_n(X)]$, the desired result follows. ■

(iii) For any $\omega \in \Omega$, set $Z(\omega) := X(\omega) - b$. Our assumptions, together with linearity of expectation, give $P(Z \leq 0) = 1$ and $E(Z) = 0$. We need to show that $P(Z = 0) = 1$.

The claim follows if we show that $P(Z < 0) = 0$. For each $n \in \mathbb{N}$, set

$$\Omega_n := \left\{ \omega \in \Omega : Z(\omega) < -\frac{1}{n} \right\}.$$

Then,

$$Z(\omega) \leq \left(D_n(Z)(\omega) + \frac{1}{n} \right) \mathbb{1}_{\Omega_n}(\omega) \leq -\frac{1}{n} \mathbb{1}_{\Omega_n}(\omega),$$

for all $\omega \in \Omega$, so that, by monotonicity and linearity of expectation,

$$0 = E[Z] \leq E\left[\left(D_n(Z) + \frac{1}{n} \right) \mathbb{1}_{\Omega_n} \right] \leq -\frac{1}{n} P(\Omega_n),$$

which forces $P(\Omega_n) = 0$ for all $n \in \mathbb{N}$. The desired result follows (recall Exercise 14-(b) in Problem Set 3). ■

[1] Recall that $D_n(X) \leq X \leq D_n(X) + 1/n$ for all $n \in \mathbb{N}$.

Exercise 26

Let $(X_n)_n$ be a sequence of i.i.d. real-valued random variables on some probability space $(\Omega, \mathcal{A}, P)$ with unknown cumulative distribution function F_X. An estimator for F_X based on the random variables $X_1, \ldots, X_n$ is the so-called *empirical distribution function*, defined as

$$\widehat{F}_n(z) := \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}(X_k \leq z), \qquad -\infty < z < +\infty.$$

Calculate the mean and the variance of $\widehat{F}_n(z)$. Show that $\widehat{F}_n(z) \xrightarrow{P} F_X(z)$.

Solution

To begin, observe the following: since $(X_n)_n$ is a sequence of i.i.d. random variables, the indicator function is $\mathcal{B}(\mathbb{R})$-$\mathcal{B}(\mathbb{R})$ measurable and measurable transformations of independent random variables are independent (Exercise 16), $(\mathbb{1}(X_n \leq z))_n$ is a sequence of i.i.d. random variables on $(\Omega, \mathcal{A}, P)$.
Now, fix arbitrary $n \in \mathbb{N}$ and $z \in \mathbb{R}$. To compute the expectation of $\widehat{F}_n(z)$, note first that

$$E\big[\mathbb{1}(X_k \leq z)\big] = P(X_k \leq z) = F_X(z) \quad \text{ for all } k \in \mathbb{N}.$$

Since $E\big[\mathbb{1}(X_k \leq z)\big]$ is finite, $\mathbb{1}(X_k \leq z) \in L^1(\Omega, \mathcal{A}, P)$ for all k. Thus, by linearity of expectation we have

$$\begin{aligned} E\Big[\widehat{F}_n(z)\Big] &:= E\Big[\frac{1}{n} \sum_{k=1}^{n} \mathbb{1}(X_k \leq z)\Big] \\ &= \frac{1}{n} \sum_{k=1}^{n} E\big[\mathbb{1}(X_k \leq z)\big] \\ &= \frac{1}{n} \sum_{k=1}^{n} P(X_k \leq z) \\ &= \frac{1}{n} \sum_{k=1}^{n} F_X(z). \\ &= F_X(z). \end{aligned}$$

To compute the variance of $\widehat{F}_n(z)$, note first that $\big(\mathbb{1}(X_k \leq z)\big)^2 = \mathbb{1}(X_k \leq z)$, and so

$$E\Big[\big(\mathbb{1}(X_k \leq z)\big)^2\Big] = F_X(z) \quad \text{ for all } k \in \mathbb{N}.$$

Then

$$\begin{aligned} \operatorname{Var}\big(\mathbb{1}(X_k \leq z)\big) &= E\Big[\big(\mathbb{1}(X_k \leq z)\big)^2\Big] - \Big(E\big[\mathbb{1}(X_k \leq z)\big]\Big)^2 \\ &= F_X(z) - (F_X(z))^2 \\ &= F_X(z)(1 - F_X(z)), \end{aligned}$$

which is finite. Thus, we have

$$\begin{aligned} \operatorname{Var}\Big(\widehat{F}_n(z)\Big) &= \operatorname{Var}\Big(\frac{1}{n} \sum_{k=1}^{n} \mathbb{1}(X_k \leq z)\Big) \\ &= \frac{1}{n^2} \sum_{k=1}^{n} \operatorname{Var}\big(\mathbb{1}(X_k \leq z)\big) \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n^2}\sum_{k=1}^{n} F_X(z)(1-F_X(z)) \\
&= \frac{1}{n}F_X(z)(1-F_X(z)),
\end{aligned}
$$

where the second equality holds because the $\mathbb{1}(X_k \leq z)$'s are independent.
Finally, since $(\mathbb{1}(X_k \leq n))_n$ is a sequence of i.i.d. random variables with finite expectation $E\big[\mathbb{1}(X_k \leq z)\big] = F_X(z)$, by the weak law of large numbers (Theorem 2.7) we have

$$
\frac{1}{n}\sum_{k=1}^{n} \mathbb{1}(X_k \leq z) \xrightarrow{P} F_X(z),
$$

as desired. ■

*Exercise 27

Solve the following problems.

(a) Let $(X_n)_n$ and X be $\mathbb{R}^k$-valued random variables on a common probability space $(\Omega, \mathcal{A}, P)$. Show that $X_n \xrightarrow{P} X$ if and only if $\|X_n - X\| \xrightarrow{P} 0$.

(b) Let $(X_n)_n$ be a sequence of $\mathbb{R}^k$-valued random variables on a common probability space $(\Omega, \mathcal{A}, P)$, and let $a \in \mathbb{R}^k$. Show that $X_n \xrightarrow{P} a$ if and only if $X_n \xrightarrow{d} a$.

Solution

(a) For all $n \in \mathbb{N}$, $\omega \in \Omega$ and $\varepsilon > 0$, we have

$$
\big\|X_n(\omega) - X(\omega)\big\| > \varepsilon \iff \Big|\big\|X_n(\omega) - X(\omega)\big\| - 0\Big| > \varepsilon.
$$

The desired result immediately follows by definition of convergence in probability. ■

(b) By Theorem 2.15, $X_n \xrightarrow{P} a \Longrightarrow X_n \xrightarrow{d} a$. Thus, we only need to prove that

$$
X_n \xrightarrow{d} a \Longrightarrow X_n \xrightarrow{P} a.
$$

Thus, assume that $X_n \xrightarrow{d} a$. Let $i \in \{1, 2, \dots, k\}$ and $\varepsilon > 0$. Moreover, let $F_{X_{n,i}}$ be the marginal cdf of $X_{n,i}$, for all $n \in \mathbb{N}$, and F_{a_i} the marginal cdf of a_i. Since joint convergence in distribution implies marginal convergence in distribution, $F_{X_{n,i}}(x) \underset{n\to\infty}{\longrightarrow} F_{a_i}(x)$ at all continuity points x of F_{a_i}. Note that

$$
F_{a_i}(x) = \begin{cases} 0 & \text{if } x < a_i \\ 1 & \text{if } x \geq a_i \end{cases},
$$

which is continuous everywhere but at $x = a_i$. Then,

$$
\begin{aligned}
P\big(\{\omega \in \Omega : |X_{n,i}(\omega) - a_i| > \varepsilon\}\big) &= 1 - P\big(\{\omega \in \Omega : a_i - \varepsilon \leq X_{n,i}(\omega) \leq a_i + \varepsilon\}\big) \\
&= 1 - F_{X_{n,i}}(a_i + \varepsilon) + F_{X_{n,i}}(a_i - \varepsilon) \\
&- P\big(\{\omega \in \Omega : X_{n,i}(\omega) = a_i - \varepsilon\}\big) \\
&\leq 1 - F_{X_{n,i}}(a_i + \varepsilon) + F_{X_{n,i}}(a_i - \varepsilon) \\
&\underset{n\to\infty}{\longrightarrow} 1 - F_{a_i}(a_i + \varepsilon) + F_{a_i}(a_i - \varepsilon)
\end{aligned}
$$

$$
\begin{aligned}
&= 1 - 1 + 0 \\
&= 0,
\end{aligned}
$$

since F_{a_i} is continuous at $x = a_i \pm \varepsilon$, and so $F_{X_n}(a_i + \varepsilon) \underset{n\to\infty}{\longrightarrow} F_{a_i}(a_i + \varepsilon) = 1$, $F_{X_n}(a_i - \varepsilon) \underset{n\to\infty}{\longrightarrow} F_{a_i}(a_i - \varepsilon) = 0$. Apply Lemma 2.13 to obtain the desired result. ■

Exercise 28

Solve the following problems.

(a) Suppose that $X_1 = X_2 = \ldots$ are standard normally distributed random variables. Show that $X_n \xrightarrow{d} -X_1$, but $(X_n)_n$ does not converge in probability to $-X_1$.

(b) Let $(X_n)_n$ and $(Y_n)_n$ be two sequences of real-valued random variables on a common probability space $(\Omega, \mathcal{A}, P)$. Suppose that $X_n \xrightarrow{P} 0$ and $Y_n \xrightarrow{P} 0$. Show that

$$\max\{|X_n|, |Y_n|\} \xrightarrow{P} 0$$

by using the definition of convergence in probability.

Solution

(a) *Claim 1.* $X_n \xrightarrow{d} -X_1$.

For all $n \in \mathbb{N}$, let F_{X_n} be the cdf of X_n. By Theorem 2.14, it is enough to show that $F_{X_n}(x) \underset{n\to\infty}{\longrightarrow} F_{-X_1}(x)$ at all continuity points of F_{-X_1}. Let x be a continuity point of F_{-X_1}. Note that

$$
\begin{aligned}
F_{X_2}(x) &:= P(X_2 \leq x) \\
&= P(X_2 \geq -x) \\
&= P(X_1 \geq -x) \\
&= P(-X_1 \leq x) \\
&:= F_{-X_1}(x),
\end{aligned} \tag{4}
$$

where the first equality holds because X_2 is standard normally distributed and so its distribution is symmetric about zero, and the second equality holds because $X_1 = X_2$. Since $X_2 = X_3 = \ldots$, $F_{X_2} = F_{X_3} = \ldots$, and so $(F_{X_n}(x))_{n\geq 2}$ is a constant sequence. Therefore, $F_{X_n}(x) \underset{n\to\infty}{\longrightarrow} F_{-X_1}(x)$ trivially follows from (4). Since $x \in \mathbb{R}$ was arbitrarily chosen, the claim follows.

Claim 2. $(X_n)_n$ does not converge in probability to $-X_1$.

For $\varepsilon := 2 > 0$ we have

$$
\begin{aligned}
P(|X_2 - (-X_1)| > 2) &= P(|X_1 + X_2)| > 2) \\
&= P(|2X_2| > 2) \\
&= P(|X_2| > 1) \\
&\approx 2 \cdot 0.1587 \\
&\neq 0,
\end{aligned}
$$

where the second equality holds because $X_1 = X_2$. Since $X_2 = X_3 = \ldots$,

$$P(|X_n - (-X_1)| > 2) = 2 \cdot 0.1587 > 0$$

holds true for all $n \geq 2$, and so

$$P(|X_n - (-X_1)| > 2) \underset{n\to\infty}{\longrightarrow} 2 \cdot 0.1587 > 0.$$

The claim follows. ■

(b) Let $\varepsilon > 0$. We need to show that

$$\lim_{n\to\infty} P\big(\{\omega \in \Omega : \max\{|X_n(\omega)|, |Y_n(\omega)|\} > \varepsilon\}\big) = 0.$$

Note that

$$\{\omega \in \Omega : \max\{|X_n(\omega)|, |Y_n(\omega)|\} > \varepsilon\}$$

is equal to

$$\{\omega \in \Omega : |X_n(\omega)| > \varepsilon\} \cup \{\omega \in \Omega : |Y_n(\omega)|\} > \varepsilon\}.$$

for all $n \in \mathbb{N}$. Then, by non-negativity and sub-σ-additivity of P we have

$$\begin{aligned} 0 &\leq P\big(\{\omega \in \Omega : \max\{|X_n(\omega)|, |Y_n(\omega)|\} > \varepsilon\}\big) \\ &\leq P\big(\{\omega \in \Omega : |X_n(\omega)| > \varepsilon\}\big) + P\big(\{\omega \in \Omega : |Y_n(\omega)| > \varepsilon\}\big) \end{aligned} \tag{5}$$

for all $n \in \mathbb{N}$. By assumption, $X_n \xrightarrow{P} 0$ and $Y_n \xrightarrow{P} 0$, and so the right hand side of (5) converges to zero as $n \to \infty$. The desired result follows by the sandwich theorem for limits of sequences. ■

Exercise 29

Solve the following problems.

(a) Let $X \sim Po(\lambda)$, with $\lambda > 0$. Show that $E[X] = \lambda$.

(b) Let X be a Gamma-distributed random variables with parameters $\alpha > 0$ and $\beta > 0$; in signs $X \sim \Gamma(\alpha, \beta)$. That is, X is continuous with density $f_X : \mathbb{R} \to [0, +\infty)$ defined as

$$f_X(x) := \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \mathbb{1}_{(0,+\infty)}(x).$$

Show that $E[X] = \alpha/\beta$. [*Hint.* $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$ for any $\alpha > 0$.]

Solution

(a) Answer: $E[X] = \lambda$. Details are omitted.

(b) Answer: $E[X] = \alpha/\beta$. Details are omitted.

*Exercise 30

Solve the following problems.

(a) Let $(X_n)_n$, X and Y be $\mathbb{R}^k$-valued random variables on a common probability space $(\Omega, \mathcal{A}, P)$. Suppose that $X_n \xrightarrow{P} X$ and $X_n \xrightarrow{P} Y$. Show that $X = Y$ almost surely.

(b) Suppose that for $\mathbb{R}^k$-valued random variables $(X_n)_n$, $(Y_n)_n$, X and Y we have $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} Y$. Does this imply $X_n + Y_n \xrightarrow{d} X + Y$? Justify your answer.

Solution

(a) Since $\|X(\omega) - Y(\omega)\| = 0$ if and only if $X(\omega) = Y(\omega)$ for all $\omega \in \Omega$, the desired claim follows if we show that

$$P\Big(\big\{\omega \in \Omega : \|X(\omega) - Y(\omega)\| > 1/k\big\}\Big) = 0$$

for all $k \in \mathbb{N}$.

Fix an arbitrary $k \in \mathbb{N}$. For any $n \in \mathbb{N}$, triangle inequality gives

$$\|X(\omega) - Y(\omega)\| \leq \|X(\omega) - X_n(\omega)\| + \|Y(\omega) - X_n(\omega)\|$$

for any $\omega \in \Omega$. Therefore, by non-negativity, monotonicity and sub-σ-additivity of P,

$$\begin{aligned} 0 &\leq P\Big(\big\{\omega \in \Omega : \|X(\omega) - Y(\omega)\| > 1/k\big\}\Big) \\ &\leq P\Big(\big\{\omega \in \Omega : \|X(\omega) - X_n(\omega)\| > 1/2k\big\}\Big) \\ &+ P\Big(\big\{\omega \in \Omega : \|Y(\omega) - X_n(\omega)\| > 1/2k\big\}\Big) \end{aligned} \tag{6}$$

holds for all $n \in \mathbb{N}$. By assumption, $X_n \xrightarrow{P} X$ and $X_n \xrightarrow{P} Y$, and so the right hand side of (6) converges to zero as $n \to \infty$. The desired result follows by the sandwich theorem for limits of sequences. ■

(b) We show that the statement is false by providing a counterexample. For each $n \in \mathbb{N}$, let $X_n \sim N(0,1)$ and set $Y_n := X_n$. Trivially, we have $X_n \xrightarrow{d} X \sim N(0,1)$ and $Y_n \xrightarrow{d} -X$ (recall Exercise 28-(a)). However, $X_n + Y_n = 2X_n \sim N(0,4)$ for all n, which clearly does not converge in distribution to $X + (-X) = 0$. ■

E 703: Advanced Econometrics I
Solutions to Problem Set 6

Niccolò Lomys

November 17, 2016

*Exercise 31

Solve the following problems.

(a) Let X be a real-valued random variable, and Y_n, Z_n real-valued discrete random variables for all $n \in \mathbb{N}$. Assume that

$$|X - Y_n| \leq Z_n, \qquad \lim_{n\to\infty} E[Z_n] = 0, \qquad \text{and} \qquad \lim_{n\to\infty} E[Y_n] = a,$$

with $a \in \mathbb{R}$, hold for all n, Show that $X \in L^1(\Omega, \mathcal{A}, P)$, and $E[X] = a$.

(b) Let $(X_n)_n$ be a sequence of real-valued random variables, and $a \in \mathbb{R}$. Show that

$$\left\langle E[X_n] \underset{n\to\infty}{\longrightarrow} a \quad \text{and} \quad \mathrm{Var}(X_n) \underset{n\to\infty}{\longrightarrow} 0 \right\rangle \iff X_n \xrightarrow{L_2} a.$$

(c) Let $(X_n)_n$, $(Y_n)_n$, $(Z_n)_n$ and X be real-valued random variables on a common probability space $(\Omega, \mathcal{A}, P)$ such that $X_n \leq Z_n \leq Y_n$ for all $n \in \mathbb{N}$. Show that

$$\left\langle X_n \xrightarrow{P} X \quad \text{and} \quad Y_n \xrightarrow{P} X \right\rangle \Longrightarrow Z_n \xrightarrow{P} X.$$

(d) Let $(X_n)_n$, $(U_n)_n$, $(W_n)_n$ and X be $\mathbb{R}^k$-valued random variables on a common probability space $(\Omega, \mathcal{A}, P)$. Prove the following statements.

(i) $X_n = O_p(1) \iff \lim_{C\to+\infty} \limsup_{n\to\infty} P(\|X_n\| > C) = 0$

(ii) Suppose that $X_n = o_p(1)$, $U_n = O_p(1)$ and $W_n = O_p(1)$. Then,

1. $U_n + W_n = O_p(1)$;
2. $X_n U_n = o_p(1)$ (assume that $(X_n)_n$, $(U_n)_n$ are real-valued for simplicity).

[*Note*. If you are asked to prove a statement from the Lecture Notes, you can use all the results that come before (but not after) that statement in the Lecture Notes.]

Solution

(a) We show that $E[X] = a$. That $X \in L^1(\Omega, \mathcal{A}, P)$ follows because $-\infty < a < +\infty$.

As $|X - Y_n| \leq Z_n$,

$$-Z_n \leq X - Y_n \leq Z_n,$$

or, equivalently,

$$Y_n - Z_n \leq X \leq Y_n + Z_n \tag{1}$$

for all $n \in \mathbb{N}$. Let $D_n(X)$ be the discrete approximation of X on a $1/n$ grid, with $n \in \mathbb{N}$. Since $D_n(X) \leq X \leq D_n(X) + 1/n$, (1) gives

$$Y_n - Z_n - \frac{1}{n} \leq D_n(X) \leq Y_n + Z_n$$

for all $n \in \mathbb{N}$. Linearity and monotonicity of expectation for discrete random variables yields

$$E[Y_n] - E[Z_n] - \frac{1}{n} \leq E[D_n(X)] \leq E[Y_n] + E[Z_n]$$

for all sufficiently large n[1], and so, by the sandwich theorem for limits of sequences (and the algebra of limits),

$$\lim_{n\to\infty} E[Y_n] - \lim_{n\to\infty} E[Z_n] - \lim_{n\to\infty} \frac{1}{n} \leq \lim_{n\to\infty} E[D_n(X)] \leq \lim_{n\to\infty} E[Y_n] + \lim_{n\to\infty} E[Z_n]. \tag{2}$$

By our assumptions and the fact that $E[X] := \lim_{n\to\infty} E[D_n(X)]$, (2) reads as

$$a \leq E[X] \leq a,$$

thus completing the proof. ■

(b) [$\Longrightarrow$]

We need to show that

$$E\big[|X_n - a|^2\big] \underset{n\to\infty}{\longrightarrow} 0.$$

For sufficiently large n, we have

$$\begin{aligned} E\big[|X_n - a|^2\big] &= E\big[(X_n - a)^2\big] \\ &= E\big[X_n^2 + a^2 - 2aE[X_n]\big] \\ &= E[X_n^2] + a^2 - 2aE[X_n] \\ &= \mathrm{Var}(X_n) + \big(E[X_n]\big)^2 + a^2 - 2aE[X_n] \\ &= \mathrm{Var}(X_n) + \big(E[X_n] - a\big)^2 \\ &\underset{n\to\infty}{\longrightarrow} 0 + (a-a)^2 = 0, \end{aligned} \tag{3}$$

where: the third equality holds by linearity of expectation; the fourth equality holds because $\mathrm{Var}(X_n) = E[X_n^2] - \big(E[X_n]\big)^2$; convergence follows by our assumptions and continuity of the function $f : \mathbb{R} \to \mathbb{R}$, $x \mapsto (x-a)^2$. The desired result obtains.

[$\Longleftarrow$]

We need to show that

$$\mathrm{Var}(X_n) \underset{n\to\infty}{\longrightarrow} 0 \qquad \text{and} \qquad E[X_n] \underset{n\to\infty}{\longrightarrow} a.$$

From (3) we have

$$E\big[|X_n - a|^2\big] = \mathrm{Var}(X_n) + \big(E[X_n] - a\big)^2 \tag{4}$$

for sufficiently large n. Since $\mathrm{Var}(X_n) \geq 0$ and $\big(E[X_n] - a\big)^2 \geq 0$ for all $n \in \mathbb{N}$, we have from (4)

$$0 \leq \mathrm{Var}(X_n) \leq E\big[|X_n - a|^2\big]$$

[1]By assumption, $\lim_{n\to\infty} E[Z_n] = 0$ and $\lim_{n\to\infty} E[Y_n] = a$. Therefore, there exist natural numbers N_Z and N_Y such that $Z_n \in L^1_d(\Omega, \mathcal{A}, P)$ for all $n > N_Z$, and $Y_n \in L^1_d(\Omega, \mathcal{A}, P)$ for all $n > N_Y$, By "sufficiently large n" we mean $n > \max\{N_Z, N_Y\}$, so that both $E[Y_n]$ and $E[Z_n]$ are finite, and using monotonicity and linearity of expectation is justified.

and

$$0 \le \big(E[X_n] - a\big)^2 \le E\big[|X_n - a|^2\big]$$

for all sufficiently large n. As $E\big[|X_n - a|^2\big] \underset{n\to\infty}{\longrightarrow} 0$ by assumption, the sandwich theorem for limits of sequences gives

$$\operatorname{Var}(X_n) \underset{n\to\infty}{\longrightarrow} 0 \qquad \text{and} \qquad \big(E[X_n] - a\big)^2 \underset{n\to\infty}{\longrightarrow} 0,$$

that is,

$$\operatorname{Var}(X_n) \underset{n\to\infty}{\longrightarrow} 0 \qquad \text{and} \qquad E[X_n] \underset{n\to\infty}{\longrightarrow} a,$$

which proves the claim. ■

(c) Let $\varepsilon > 0$ be arbitrary. We need to show that

$$P\big(Z_n - X < -\varepsilon \ \vee\ Z_n - X > \varepsilon\big) \underset{n\to\infty}{\longrightarrow} 0.$$

Since $X_n \le Z_n \le Y_n$ for all $n \in \mathbb{N}$, we have

$$Z_n - X < -\varepsilon \Longrightarrow X_n - X < -\varepsilon,$$

and

$$Z_n - X > \varepsilon \Longrightarrow Y_n - X > \varepsilon.$$

Thus,

$$\begin{aligned} 0 \le P\big(Z_n - X < -\varepsilon \ \vee\ Z_n - X > \varepsilon\big) &\le P\big(X_n - X < -\varepsilon \ \vee\ Y_n - X > \varepsilon\big) \\ &\le P\big(X_n - X < -\varepsilon\big) + P\big(Y_n - X > \varepsilon\big), \end{aligned} \tag{5}$$

where: the first inequality holds by non-negativity of P; the second inequality holds by monotonicity of P; the third inequality holds by sub-σ-additivity of P. Since $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} X$,

$$P\big(X_n - X < -\varepsilon\big) \underset{n\to\infty}{\longrightarrow} 0 \qquad \text{and} \qquad P\big(Y_n - X > \varepsilon\big) \underset{n\to\infty}{\longrightarrow} 0. \tag{6}$$

The desired result follows from (5), (6) and the sandwich theorem for limits of sequences. ■

(d) (i) $[\Longrightarrow]$

We need to show that

$$\lim_{C\to+\infty} \limsup_{n\to\infty} P(\|X_n\| > C) = 0,$$

i.e., that for any $\varepsilon > 0$, there exists C_ε such that

$$\limsup_{n\to\infty} P(\|X_n\| > C) \le \varepsilon \qquad \text{for all } C > C_\varepsilon.$$

Fix $\varepsilon > 0$. Since $X_n = O_p(1)$, there exist C_ε and N_ε such that

$$P(\|X_n\| > C_\varepsilon) \le \varepsilon \qquad \text{for all } n \ge N_\varepsilon.$$

Hence,

$$\sup_{k\ge n} P(\|X_k\| > C_\varepsilon) \le \varepsilon$$

for all $n \ge N_\varepsilon$. Thus, by monotonicity of P, for any $C > C_\varepsilon$ we have

$$P(\|X_k\| > C) \le P(\|X_k\| > C_\varepsilon),$$

and so

$$\sup_{k\geq n} P(\|X_k\| > C) \leq \varepsilon$$

for all $n \geq N_\varepsilon$. It follows that

$$\limsup_{n\to\infty} P(\|X_n\| > C) \leq \varepsilon,$$

completing the proof.

[$\Longleftarrow$]

Fix $\varepsilon > 0$. We need to show that there exist $C_\varepsilon > 0$ and $N_\varepsilon > 0$ such that

$$P(\|X_n\| \leq C_\varepsilon) \geq 1-\varepsilon \qquad \text{for all } n > N_\varepsilon.$$

By assumption, $\lim_{C\to+\infty}\limsup_{n\to\infty} P(\|X_n\| > C) = 0$. Then, by Theorem 1.3.(iii), $\lim_{C\to+\infty}\limsup_{n\to\infty} P(\|X_n\| \leq C) = 1$. Therefore, there exists $\overline{C}_\varepsilon > 0$ such that for all $C > \overline{C}_\varepsilon$ we have

$$\limsup_{n\to\infty} P(\|X_n\| \leq C) \geq 1 - \frac{\varepsilon}{2}.$$

In particular, for $C_\varepsilon := \overline{C}_\varepsilon + 1$ we have

$$\limsup_{n\to\infty} P(\|X_n\| \leq C_\varepsilon) \geq 1 - \frac{\varepsilon}{2},$$

that is,

$$\lim_{n\to\infty}\sup_{k\geq n} P(\|X_k\| \leq C_\varepsilon) \geq 1 - \frac{\varepsilon}{2}.$$

But then, there exists $N_\varepsilon > 0$ such that, for all $n > N_\varepsilon$,

$$\sup_{k\geq n} P(\|X_k\| \leq C_\varepsilon) \geq 1-\varepsilon,$$

and so, for all $n > N_\varepsilon$,

$$P(\|X_n\| \leq C_\varepsilon) \geq 1-\varepsilon. \qquad \blacksquare$$

(ii) 1. Fix $\varepsilon > 0$. Since $U_n = O_p(1)$ and $W_n = O_p(1)$, there exist $C_{\varepsilon,U} > 0$, $C_{\varepsilon,W} > 0$ and $n_{\varepsilon,U}$, $n_{\varepsilon,W}$ such that

$$P(\|U_n\| > C_{\varepsilon,U}/2) \leq \varepsilon/2 \qquad \text{for all } n \geq n_{\varepsilon,U} \tag{7}$$

and

$$P(\|W_n\| > C_{\varepsilon,W}/2) \leq \varepsilon/2 \qquad \text{for all } n \geq n_{\varepsilon,W}. \tag{8}$$

Set $C_\varepsilon := \max\{C_{\varepsilon,U}, C_{\varepsilon,W}\}$. Then, for all $n \geq n_\varepsilon := \max\{n_{\varepsilon,U}, n_{\varepsilon,W}\}$, we have

$$\begin{aligned}
P(\|U_n + W_n\| > C_\varepsilon) &\leq P(\|U_n\| + \|W_n\| > C_\varepsilon)\\
&\leq P(\|U_n\| > C_\varepsilon/2) + P(\|W_n\| > C_\varepsilon/2)\\
&\leq P(\|U_n\| > C_{\varepsilon,U}/2) + P(\|W_n\| > C_{\varepsilon,W}/2)\\
&\leq \varepsilon/2 + \varepsilon/2\\
&= \varepsilon,
\end{aligned}$$

where: the first inequality holds by triangle inequality of the Euclidean norm and monotonicity of P; the second inequality holds by monotonicity and sub-σ-additivity of P; the third inequality holds by the fact that $C := \max\{C_{\varepsilon,U}, C_{\varepsilon,W}\}$ and monotonicity of P; the fourth inequality holds by (7) and (8) and the fact that $n \geq n_\varepsilon$. The claim follows. $\blacksquare$

2. Let $\varepsilon, \delta > 0$ be arbitrary. We want to show that there exists $N_{\varepsilon,\delta}$ such that for $n \in \mathbb{N}$, $n > N_{\varepsilon,\delta}$, we have
$$P(|X_n U_n| > \varepsilon) < \delta.$$
Since $U_n = O_p(1)$, there exist C_δ and N_δ such that
$$P(|U_n| > C_\delta) \leq \frac{\delta}{2} \tag{9}$$
for all $n > N_\delta$. Moreover, as $X_n = o_p(1)$, there exists $N_{\varepsilon/C_\delta,\delta}$ such that
$$P(|X_n| > \varepsilon/C_\delta) < \frac{\delta}{2} \tag{10}$$
for all $n > N_{\varepsilon/C_\delta,\delta}$. Observe that
$$|X_n U_n| > \varepsilon \Longrightarrow \big\langle |X_n| > \varepsilon/C_\delta \ \vee \ |U_n| > C_\delta \big\rangle. \tag{11}$$
Then, for $n > N_{\varepsilon,\delta} := \max\left\{N_{\varepsilon/C_\delta,\delta}, N_{\epsilon,\delta}\right\}$
$$\begin{aligned} P(|X_n U_n| > \varepsilon) &\leq P(|X_n| > \varepsilon/C_\delta \ \vee \ |U_n| > C_\delta) \\ &\leq P(|X_n| > \varepsilon/C_\delta) + P(|U_n| > C_\delta) \\ &< \delta/2 + \delta/2 \\ &= \delta, \end{aligned}$$
where: the first inequality holds by monotonicity of P and (11); the second inequality holds by sub-σ-additivity of P; the last inequality holds by (9) and (10) and the fact that $n > N_{\varepsilon,\delta}$. The desired result follows. ■

Exercise 32

Solve the following problems.

(a) Let $X \sim B(n, \pi)$ and $Y \sim Po(\lambda)$.

  (i) Find the characteristic functions of X and Y.

  (ii) Use the characteristic function of X to compute $E[X]$ and $\mathrm{Var}(X)$.

(b) Let $X \sim Po(\lambda), Y \sim Po(\mu)$, with $\lambda, \mu > 0$, and assume that X and Y are independent. Show that $(X + Y) \sim Po(\lambda + \mu)$.

(c) Let $X_n \sim B(n, \pi_n)$ for all $n \in \mathbb{N}$, $Y \sim Po(\lambda)$ and suppose that $\lim_{n\to\infty} n\pi_n = \lambda$. Show that $X_n \xrightarrow{d} Y$.

Solution

(a) (i) For all $t \in \mathbb{R}$, we have (omitting details)
$$\varphi_X(t) := E\left[e^{itX}\right] = \sum_{k=0}^{n} e^{itk}\binom{n}{k}\pi^k(1-\pi)^{n-k} = \left(\pi e^{it} + (1-\pi)\right)^n,$$
and
$$\varphi_Y(t) := E\left[e^{itY}\right] = \sum_{k=0}^{\infty} e^{itk}\frac{\lambda^k}{k!}e^{-\lambda} = e^{\lambda(e^{it}-1)}.$$

(ii) We have (omitting details)

$$E[X] = i^{-1}\varphi_X^{(1)}(0) = n\pi,$$
$$E[X^2] = i^{-2}\varphi_X^{(2)}(0) = n\pi((n-1)\pi + 1),$$

and so

$$\text{Var}(X) = n\pi((n-1)\pi + 1) - n^2\pi^2 = n\pi(1-\pi). \quad \blacksquare$$

(b) The characteristic functions of X and Y are $\varphi_X(t) = e^{\lambda(e^{it}-1)}$ and $\varphi_Y(t) = e^{\mu(e^{it}-1)}$. Since X and Y are independent,

$$\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t) = e^{\lambda(e^{it}-1)}e^{\mu(e^{it}-1)} = e^{(\lambda+\mu)(e^{it}-1)},$$

which is the characteristic function of a discrete random variable with Poisson distribution with parameter $\lambda + \mu$. Hence, $(X+Y) \sim Po(\lambda+\mu)$ by Lemma 1.43. $\blacksquare$

(c) By Levy's continuity theorem (Corollary 2.17 in the Lecture Notes), it is enough to show that

$$\varphi_{X_n}(t) \underset{n\to\infty}{\longrightarrow} \varphi_Y(t)$$

for all $t \in \mathbb{R}$. Let $t \in \mathbb{R}$ be arbitrary. We have

$$\begin{aligned}
\lim_{n\to\infty} \varphi_{X_n}(t) &= \lim_{n\to\infty} \left(\pi_n e^{it} + (1-\pi_n)\right)^n \\
&= \lim_{n\to\infty} \left(1 + \frac{n\pi_n(e^{it}-1)}{n}\right)^n \\
&= \lim_{n\to\infty} \exp\left(n \ln\left(1 + \frac{n\pi_n(e^{it}-1)}{n}\right)\right) \\
&= \exp\left(\lim_{n\to\infty} n \ln\left(1 + \frac{\lim_{n\to\infty} n\pi_n(e^{it}-1)}{\lim_{n\to\infty} n}\right)\right) \\
&= \exp\left(\lim_{n\to\infty} n \ln\left(1 + \frac{\lambda(e^{it}-1)}{\lim_{n\to\infty} n}\right)\right) \\
&= \exp\left(\lim_{n\to\infty} n \ln\left(1 + \frac{\lambda(e^{it}-1)}{\lim_{n\to\infty} n}\right)\right) \\
&= \exp\left(\lim_{n\to\infty} n \ln\left(1 + \frac{\lambda(e^{it}-1)}{\lim_{n\to\infty} n}\right)\right) \\
&= \lim_{n\to\infty} \exp\left(\ln\left(1 + \frac{\lambda(e^{it}-1)}{n}\right)\right) \\
&= \lim_{n\to\infty} \left(1 + \frac{\lambda(e^{it}-1)}{n}\right)^n \\
&= e^{\lambda(e^{it-1})} \\
&= \varphi_Y(t),
\end{aligned}$$

where we used the fact that, for a continuous function f, $\lim_{n\to\infty} f(x_n) = f(\lim_{n\to\infty} x_n)$, the assumption that $\lim_{n\to\infty} n\pi_n = \lambda$, and the fact that $\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n = e^x$ for any real number x. $\blacksquare$

Exercise 33

Let $(X_n)_n$ and $(Y_n)_n$ be sequences of $\mathbb{R}^k$-valued random variables such that $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} Y$. Suppose that X_n and Y_n are independent for all n and that X and Y are independent. Show that $X_n + Y_n \xrightarrow{d} X + Y$.

Solution

By Levy's continuity theorem (Corollary 2.17 in the Lecture Notes), the claim follows if we show that

$$\varphi_{X_n+Y_n}(t) \underset{n\to\infty}{\longrightarrow} \varphi_{X+Y}(t) \quad \text{for all } t \in \mathbb{R}^k.$$

Let $t \in \mathbb{R}^k$ be arbitrary. Since X_n and Y_n are independent for all n, and so are and X and Y, we have

$$\varphi_{X_n+Y_n}(t) = \varphi_{X_n}(t)\varphi_{Y_n}(t) \tag{12}$$

for all n, and

$$\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t) \tag{13}$$

(cf. Proposition 1.42 in the Lecture Notes). As $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} Y$, Levy's continuity theorem gives

$$\varphi_{X_n}(t) \underset{n\to\infty}{\longrightarrow} \varphi_X(t) \qquad \text{and} \qquad \varphi_{Y_n}(t) \underset{n\to\infty}{\longrightarrow} \varphi_Y(t),$$

and so, from the algebra of limits,

$$\varphi_{X_n}(t)\varphi_{Y_n}(t) \underset{n\to\infty}{\longrightarrow} \varphi_X(t)\varphi_Y(t). \tag{14}$$

The desired result follows by combining (12), (13) and (14). ■

*Exercise 34

Solve the following problems.

(a) Suppose that $(X_n)_n$ is a sequence of i.i.d. real-valued random variables with known expectation $-\infty < \mu < +\infty$, unknown variance $0 < \sigma^2 < +\infty$, and finite fourth moments. Apply the delta-method to derive the limit distribution of $\sqrt{n}(S_n - \sigma)$, where S_n is the following estimator of σ:

$$S_n := \sqrt{\frac{1}{n}\sum_{k=1}^{n}(X_k - \mu)^2}.$$

(b) Suppose that $(X_n)_n$ is a sequence of i.i.d. real-valued random variables with $E[X_1^4] < +\infty$. Define $Y_n := n^{-1}\sum_{k=1}^{n}|X_k|$, $Z_n := n^{-1}\sum_{k=1}^{n}X_k^2$, and $T_n := (Y_n, Z_n)'$. Show that $\sqrt{n}(T_n - \theta) \xrightarrow{d} N((0,0)', \Sigma)$ and identify θ and Σ in terms of moments of X_1.

(c) For $i \in \{1, \ldots, k\}$, let Z_i be i.i.d. with $Z_i \sim N(0,1)$, and define $Z := (Z_1, \ldots, Z_k)$. Let $\mu \in \mathbb{R}^k$ and A be a $k \times k$ real matrix. Compute the mean, the variance and the characteristic function of the random variable $X := \mu + AZ$.

Solution

(a) Set $Y_n := (X_n - \mu)^2$ for all n. Since $(X_n)_n$ is a sequence of i.i.d. random variables, so is $(Y_n)_n$. Moreover,

$$E[Y_n] := E\big[(X_n - \mu)^2\big] = \sigma^2,$$

and

$$\begin{aligned}\operatorname{Var}(Y_n) &:= E\Big[\big(Y_n - \sigma^2\big)^2\Big] \\ &= E\Big[\big((X_n - \mu)^2 - \sigma^2\big)^2\Big] \\ &= E\big[(X_n - \mu)^4\big] - \sigma^4,\end{aligned}$$

where we use linearity of expectation and the fact that σ^4 is finite. Set $S^2 := E\big[(X_n - \mu)^4\big] - \sigma^4$. As X_n has finite fourth moments, we have $0 < S^2 < +\infty$.[2] Thus, from the central limit theorem for i.i.d. real-valued random variables,

$$\frac{\sum_{i=1}^n Y_i - n\sigma^2}{\sqrt{n}} \xrightarrow{d} Z \sim N(0, S^2),$$

or, by setting $T_n := \frac{1}{n}\sum_{i=1}^n Y_i$,

$$\sqrt{n}\big(T_n - \sigma^2\big) \xrightarrow{d} Z \sim N(0, S^2).$$

The function $\phi\colon [0, +\infty) \to \mathbb{R}$, $\phi(x) := \sqrt{x}$, is differentiable at any $x > 0$, and so in particular at $x = \sigma^2$, with

$$\begin{aligned}\phi(T_n) &= S_n, \\ \phi(\sigma^2) &= \sigma,\end{aligned}$$

and
$$\big(\phi'(\sigma^2)\big)^2 = \frac{1}{4\sigma^2}.$$

Applying the delta-method we obtain

$$\sqrt{n}\big(\phi(T_n) - \phi(\sigma^2)\big) \xrightarrow{d} N\Big(0, S^2\big(\phi'(\sigma^2)\big)^2\Big),$$

that is,

$$\sqrt{n}(S_n - \sigma) \xrightarrow{d} N\left(0, \frac{E\big[(X_n - \mu)^4\big] - \sigma^4}{4\sigma^2}\right),$$

which answers the question. ■

(b) Set $S_n := (|X_n|, X_n^2)$ for all n. Since $(X_n)_n$ is a sequence of i.i.d. random variables, so is $(S_n)_n$. Define

$$\theta_1 := E[|X_1|], \tag{15}$$
$$\theta_2 := E[X_1^2], \tag{16}$$

and
$$\theta := (\theta_1, \theta_2)'. \tag{17}$$

Clearly, $E[S_n] = \theta$ for all n. Let Σ denote the covariance matrix of S_1. We have

$$E\Big[\big(|X_1| - E[|X_1|]\big)^2\Big] = E\big[|X_1|^2\big] - \theta_1^2 = \theta_2 - \theta_1^2,$$

[2] *Quiz:* Why $0 < S^2$?

$$E\Big[\big(X_1^2 - E[X_1^2]\big)^2\Big] = E[X_1^4] - \theta_2^2,$$

and

$$E\Big[\big(|X_1| - \theta_1\big)\big(X_1^2 - \theta_2\big)\Big] = E\big[|X_1|X_1^2\big] - \theta_1\theta_2 = E\big[|X_1|^3\big] - \theta_1\theta_2.$$

Since $E[X_1^4] < +\infty$ by assumption, all previous moments are finite, and using linearity of expectation is correct. Therefore, we have

$$\Sigma = \begin{bmatrix} \theta_2 - \theta_1^2 & E\big[|X_1|^3\big] - \theta_1\theta_2 \\ E\big[|X_1|^3\big] - \theta_1\theta_2 & E[X_1^4] - \theta_2^2 \end{bmatrix}, \tag{18}$$

which is finite. Then, by the multivariate central limit theorem,

$$\frac{1}{\sqrt{n}}\left(\sum_{i=1}^{n} S_i - n\theta\right) \xrightarrow{d} Z \sim N((0,0)', \Sigma).$$

Observing that

$$\frac{1}{\sqrt{n}}\left(\sum_{i=1}^{n} S_i - n\theta\right) = \sqrt{n}(T_n - \theta)$$

completes the proof. That θ and Σ are identified in terms of moments of X_1 follows from (15), (16), (17) and (18). ∎

(c) (Detailed sketch) Since the Z_i's are i.i.d. with $Z_i \sim N(0,1)$, we have $Z \sim N(0_k, I_k)$, where I_k ($= \mathrm{Var}(Z) = E\big[ZZ'\big]$) denotes the $k \times k$ identity real matrix. We have

$$E[X] := E[\mu + AZ] = \mu + AE[Z] = \mu + E[Z] = \mu + 0_k = \mu$$

by linearity of expectation, and

$$\begin{aligned} \mathrm{Var}(X) &:= \mathrm{Var}(\mu + AZ) \\ &:= E\Big[\big[\mu + AZ - E[\mu + AZ]\big]\big[\mu + AZ - E[\mu + AZ]\big]'\Big] \\ &= E\big[AZZ'A'\big] \\ &= AE\big[ZZ'\big]A' \\ &= AI_kA' \\ &= AA'. \end{aligned}$$

Moreover, for all $t \in \mathbb{R}^k$,

$$\varphi_Z(t) := E\big[e^{it'Z}\big] = e^{-t't/2}.$$

Since $\varphi_{\mu+AZ}(t) = e^{i\mu't}\varphi_Z(A't)$ (cf. Proposition 1.42 in the Lecture Notes), we get

$$\varphi_X(t) = \exp(i\mu't)\exp(-t'AA't/2) = \exp(i\mu't - t'AA't/2)$$

for all $t \in \mathbb{R}^k$. ∎

Exercise 35

Let X be a real-valued random variable, and let φ_X be its characteristic function. Show that φ_X is real if and only if $X \stackrel{d}{=} -X$.

Solution

Fix $t \in \mathbb{R}$. First, we show that

$$\overline{\varphi_X(t)} = \varphi_X(-t) = \varphi_{-X}(t). \tag{19}$$

To see this, we simply let the minus sign wander through the exponent:

$$\begin{aligned}
\overline{e^{ixt}} &= \overline{\cos xt + i \sin xt} \\
&= \cos xt - i \sin xt \\
&= \cos(x(-t)) + i \sin(x(-t)) && (= e^{ix(-t)}) \\
&= \cos((-x)t) + i \sin((-x)t) \\
&= e^{i(-x)t}.
\end{aligned}$$

Now, if φ_X is real, then $\overline{\varphi_X(t)} = \varphi_X(t)$. If follows from (19) that $\varphi_{-X}(t) = \varphi_X(t)$, which means that X and $-X$ have the same characteristic function, and, hence, by uniqueness (Lemma 1.43 in the Lecture Notes), the same distribution.

If, on the other hand, $X \stackrel{d}{=} -X$, then $\varphi_X(t) = \varphi_{-X}(t)$ (again by Lemma 1.43). Together with (19), this yields $\varphi_X(t) = \overline{\varphi_X(t)}$, that is, φ_X is real. ■

Exercise 36

Let $(X_n)_n$ be a sequence of i.i.d. random variables, with $X_n \sim Exp(\lambda)$. Define $\hat{\Lambda}_n := 1/\overline{X}_n$, with $\overline{X}_n := \frac{1}{n}\sum_{k=1}^n X_k$. Find σ^2 (as a function of λ) such that

$$\sqrt{n}(\hat{\Lambda}_n - \lambda) \xrightarrow{d} Z \sim N(0, \sigma^2).$$

Solution

For each i, we have

$$E[X_i] = \frac{1}{\lambda} \qquad \text{and} \qquad \operatorname{Var}(X_i) = \frac{1}{\lambda^2}.$$

By the central limit theorem for i.i.d. sequences of real-valued random variables,

$$\sqrt{n}\left(\overline{X}_n - \frac{1}{\lambda}\right) \xrightarrow{d} N\left(0, \frac{1}{\lambda^2}\right).$$

The function $\phi\colon (0, +\infty) \to \mathbb{R}$, $\phi(x) := 1/x$, is differentiable at all x on its domain, and so in particular at $x = 1/\lambda$, with

$$\begin{aligned}
\phi(\overline{X}_n) &= \hat{\Lambda}_n, \\
\phi(1/\lambda) &= \lambda,
\end{aligned}$$

and

$$\big(\phi'(1/\lambda)\big)^2 = \lambda^4.$$

Applying the delta-method we obtain

$$\sqrt{n}\big(\phi(\overline{X}_n) - \phi(1/\lambda)\big) \xrightarrow{d} N\left(0, \frac{1}{\lambda^2}\big(\phi'(1/\lambda)\big)^2\right),$$

that is,

$$\sqrt{n}\big(\hat{\Lambda}_n - \lambda\big) \xrightarrow{d} N(0, \lambda^2),$$

which answers the question, with $\sigma^2(\lambda) = \lambda^2$. ■

E 703: Advanced Econometrics I
Solutions to Problem Set 7

Niccolò Lomys

November 24, 2016

Exercise 37

Let Y be a real-valued random variable on some probability space $(\Omega, \mathcal{A}, P)$, with $Y \in L^2(\Omega, \mathcal{A}, P)$. Show that $\mu := E[Y]$ is the unique minimizer of the function $G: \mathbb{R} \to [0, +\infty)$ defined as $G(c) := E\big[(Y-c)^2\big]$.

Solution

Note that

$$\begin{aligned} G(c) &:= E\big[(Y-c)^2\big] \\ &= E[Y^2 + c^2 - 2cY] \\ &= E[Y^2] + c^2 - 2cE[Y], \end{aligned}$$

where the last equality holds by linearity of expectation and the fact that $Y \in L^2(\Omega, \mathcal{A}, P)$ and $c \in \mathbb{R}$. We need to solve the following minimization problem:

$$\min_{c \in \mathbb{R}} \left[E[Y^2] + c^2 - 2cE[Y]\right].$$

The objective function is twice continuously differentiable (in c) on $\mathbb{R}$. The first order condition is

$$\frac{\partial G(c)}{\partial c} = 0 \iff 2c - 2E[Y] = 0 \iff c = E[Y].$$

The second order condition is

$$\frac{\partial^2 G(c)}{\partial c^2} = 2 > 0,$$

which shows that the objective function is strictly convex on $\mathbb{R}$. Thus, $\mu := E[Y]$ is the unique minimizer of $G(c)$. ■

*Exercise 38

Solve the following problems.

(a) Find a sequence of random vectors $((X_n, Y_n)')_n$ and a random vector $(X, Y)'$ such that $X_n \xrightarrow{d} X$, $Y_n \xrightarrow{d} Y$, but *not* $((X_n, Y_n)')_n \xrightarrow{d} (X, Y)'$.

(b) Let $(X_n)_n$, with $X_n := (X_{n,1}, \dots, X_{n,k})$ for all n, be a sequence of $\mathbb{R}^k$-valued random variables, and $X := (X_1 \dots, X_k)$ an $\mathbb{R}^k$-valued random variable. Suppose that $X_n \xrightarrow{d} X$. Show that $X_{n,i} \xrightarrow{d} X_i$ for all $i \in \{1, \dots, k\}$.

Solution

(a) Let $X \sim N(0, 1)$, $Y \sim (0, 1)$, and assume that X and Y are independent. Then,

$$(X, Y)' \sim N((0, 0)', I_2),$$

where I_2 is the 2×2 real identity matrix. Now, for all n, let $X_n \sim N(0, 1)$, and set $Y_n := -X_n$, Clearly, $X_n \xrightarrow{d} X$, and $Y_n \xrightarrow{d} Y$ (cf. Exercise 28-(a)). However,

$$(X_n, Y_n)' \xrightarrow{d} (X, -X)' \sim N((0, 0)', \Sigma),$$

with

$$\Sigma = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

As $(X, -X)' \overset{d}{\neq} (X, Y)'$, $((X_n, Y_n)')_n$ does not converge in distribution to $(X, Y)'$. ■

(b) Consider an arbitrary $i \in \{1, \ldots, k\}$, and let $\boldsymbol{e}_i$ the i-th vector of the standard basis of $\mathbb{R}^k$ (i.e., $\boldsymbol{e}_i$ is the vector in $\mathbb{R}^k$ with a 1 in the i-th coordinate and 0's elsewhere). Since $X_n \xrightarrow{d} X$, by the Cramér-Wold device (Theorem 2.18 in the Lecture Notes) we have

$$\boldsymbol{e}_i' X_n \xrightarrow{d} \boldsymbol{e}_i' X$$

Since $\boldsymbol{e}_i' X_n = X_{n,i}$ for all n, and $\boldsymbol{e}_i' X = X_i$, the desired result follows. ■

Exercise 39

A real-valued random variable X is standard Cauchy distributed if it is continuous with density $f_X\colon \mathbb{R} \to [0, +\infty)$ defined by

$$f_X(x) := \frac{1}{\pi\,(1 + x^2)},$$

where π is a strictly positive real constant. The characteristic function of X is $\varphi_X(t) = e^{-|t|}$. Suppose that $X_1, \ldots, X_n$ are i.i.d. standard Cauchy distributed random variables. Show that $\overline{X}_n := n^{-1} \sum_{i=1}^n X_i$ is standard Cauchy distributed.

Solution

For all $t \in \mathbb{R}$, we have

$$\begin{aligned}
\varphi_{\overline{X}_n}(t) &:= \varphi_{\frac{1}{n}\sum_{i=1}^n X_i}(t) \\
&= \varphi_{\sum_{i=1}^n X_i}\left(\frac{t}{n}\right) \\
&= \prod_{i=1}^n \varphi_X\left(\frac{t}{n}\right) \\
&= \left[\varphi_X\left(\frac{t}{n}\right)\right]^n \\
&= \left[e^{-\frac{|t|}{n}}\right]^n \\
&= e^{-|t|} \\
&= \varphi_X(t),
\end{aligned}$$

where the first equality holds by Proposition 1.42 in the Lecture Notes, and the second equality holds by Proposition 1.42 and the fact that $X_1, \ldots, X_n$ are i.i.d. standard Cauchy random variables. Thus, $\varphi_{\overline{X}_n} = \varphi_X$, and so $\overline{X}_n$ is standard Cauchy distributed by Lemma 1.43 in the Lecture Notes. ■

Exercise 40

Solve the following problems.

(a) (Final Exam - Spring 2013) Suppose that Ω_1 and Ω_2 are two nonempty sets, and that $f\colon \Omega_1 \to \Omega_2$ is a function. Further, let $\mathcal{A}_2$ be a σ-field on Ω_2. Show that the class

$$\mathcal{A}_1 := \{f^{-1}(A) \in \mathcal{P}(\Omega_1) : A \in \mathcal{A}_2\}$$

is a σ-field on Ω_1. [*Note*. For each $A \in \mathcal{P}(\Omega_2)$, $f^{-1}(A) := \{\omega \in \Omega_1 : f(\omega) \in A\}$, i.e. $f^{-1}(A)$ denotes the inverse image of A under f.]

(b) Let X be a real-valued random variable on some probability space $(\Omega, \mathcal{A}, P)$. Show that X is independent of itself if and only if X is constant almost surely.

Solution

(a) We show that the defining properties of σ-field (Definition 1.1 in the Lecture Notes) are satisfied by $\mathcal{A}_1$.

(o) By definition, $\mathcal{A}_1$ is a class of subsets of Ω_1.

(i) Since $\mathcal{A}_2$ is a σ-field, $\emptyset \in \mathcal{A}_2$. By definition of $\mathcal{A}_1$, $f^{-1}(\emptyset) \in \mathcal{A}_1$. As

$$f^{-1}(\emptyset) := \{\omega \in \Omega_1 : f(\omega) \in \emptyset\} = \emptyset,$$

we have $\emptyset \in \mathcal{A}_1$.

(ii) Let $B \in \mathcal{A}_1$. Then, by definition of $\mathcal{A}_1$, there exists $A \in \mathcal{A}_2$ such that $B = f^{-1}(A)$. Since $\mathcal{A}_2$ is a σ-field on Ω_2, $\Omega_2 \setminus A \in \mathcal{A}_2$. By definition of $\mathcal{A}_1$, $f^{-1}(\Omega_2 \setminus A) \in \mathcal{A}_1$. As

$$f^{-1}(\Omega_2 \setminus A) = \Omega_1 \setminus f^{-1}(A) = \Omega_1 \setminus B,$$

we conclude that $\mathcal{A}_1$ is closed under complementation.

(iii) Let $B_1, B_2, \dots \in \mathcal{A}_1$. Then, by definition of $\mathcal{A}_1$, there exist $A_1, A_2, \dots \in \mathcal{A}_2$ such that $B_1 = f^{-1}(A_1), B_2 = f^{-1}(A_2), \ldots$. Since $\mathcal{A}_2$ is a σ-field, $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}_2$. By definition of $\mathcal{A}_1$, $f^{-1}\big(\bigcup_{n=1}^{\infty} A_n\big) \in \mathcal{A}_1$. As

$$f^{-1}\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \bigcup_{n=1}^{\infty} f^{-1}(A_n) = \bigcup_{n=1}^{\infty} B_n,$$

we conclude that $\mathcal{A}_1$ is closed under taking countably infinite unions. ■

Remark. Note that $\mathcal{A}_1$ is the σ-algebra on Ω_1 generated by $f\colon \Omega_1 \to \Omega_2$ (cf. Exercise 48).

(b) First, suppose that $P(X = c) = 1$ for some $c \in \mathbb{R}$. For any $A \in \mathcal{B}(\mathbb{R})$, we have

$$P(X \in A) = 1 \iff c \in A,$$

and therefore,

$$P(X \in A) = \mathbb{1}_A(c).$$

Then, for any $A_1, A_2 \in \mathcal{B}(\mathbb{R})$,

$$\begin{aligned} P(X \in A_1, X \in A_2) &= P(X \in (A_1 \cap A_2)) \\ &= \mathbb{1}_{A_1 \cap A_2}(c) \\ &= \mathbb{1}_{A_1}(c)\mathbb{1}_{A_2}(c) \\ &= P(X \in A_1)P(X \in A_2), \end{aligned}$$

which shows that X is independent of itself.

Now, suppose that X is independent of itself. Then, for any $x \in \mathbb{R}$,

$$P(X \le x) = P(X \le x, X \le x) = [P(X \le x)]^2.$$

This means that $P(X \le x)$ can only equal 0 or 1 for any $x \in \mathbb{R}$. Since $\lim_{x \to -\infty} P(X \le x) = 0$, $\lim_{x \to +\infty} P(X \le x) = 1$, and $P(X \le x)$ is non-decreasing in x, there must be a $c \in \mathbb{R}$ such that $P(X \le c) = 1$ and $P(X < c) = 0$, which shows that $P(X = c) = 1$. ■

*Exercise 41

Solve the following problems

(a) Let X be an $\mathbb{R}^k$-valued random variable on some probability space $(\Omega, \mathcal{A}, P)$ and c a real constant. Show by definition of conditional expectation that $E[c \mid X] = c$ almost surely.

(b) Let Y be a real-valued random variable on some probability space $(\Omega, \mathcal{A}, P)$, $Y \in L^2(\Omega, \mathcal{A}, P)$, and c a constant in $\mathbb{R}^k$. Show that $E[Y \mid c] = E[Y]$.

(c) Let Y be a real-valued random variable, X an $\mathbb{R}^k$-valued random variable, and Z an $\mathbb{R}^m$-valued random variable all defined on the same probability space $(\Omega, \mathcal{A}, P)$. Let $Y \in L^1(\Omega, \mathcal{A}, P)$. Show by definition of conditional expectation that:

   (i) $E[E[Y \mid X]] = E[Y]$;

   (ii) $E[E[Y \mid X] \mid X, Z] = E[Y \mid X]$ almost surely;

   (iii) $E[Y \mid X, f(X)] = E[Y \mid X]$ almost surely for any $\mathcal{B}(\mathbb{R}^k)$-$\mathcal{B}(\mathbb{R}^m)$ measurable function $f: \mathbb{R}^k \to \mathbb{R}^m$.

Solution

(a) Since $E[c] = c \in \mathbb{R}$, we have $c \in L^1(\Omega, \mathcal{A}, P)$, and the conditional expectation of c given X exists by Theorem 3.4 in the Lecture Notes. Let $g(X)$ be a version of $E[c \mid X]$. By definition, $g(X)$ satisfies

$$E[ch(X)] = E[g(X)h(X)],$$

or, equivalently, using linearity of expectation,

$$E[(c - g(X))h(X)] = 0 \tag{1}$$

for every bounded and Borel measurable function $h: \mathbb{R}^k \to \mathbb{R}$. By way of contradiction, suppose that $g(X) = c$ almost surely does not hold. Then, there exist $\varepsilon_1, \eta_1 > 0$ such that

$$P(c - g(X) \ge \varepsilon_1) \ge \eta_1 > 0,$$

or there exist $\varepsilon_2, \eta_2 > 0$ such that

$$P(g(X) - c \ge \varepsilon_2) \ge \eta_2 > 0,$$

or both. Assume without loss of generality that

$$P(c - g(X) \geq \varepsilon) \geq \eta > 0, \tag{2}$$

for some $\varepsilon, \eta > 0$, and define $h\colon \mathbb{R}^k \to \mathbb{R}$ as

$$h(X) := \mathbb{1}_{\{c-g(X)\geq\varepsilon\}}.$$

Clearly, h is bounded and Borel measurable. However,

$$\begin{aligned} E\big[(c-g(X))\mathbb{1}_{\{c-g(X)\geq\varepsilon\}}\big] &\geq E\big[\varepsilon\mathbb{1}_{\{c-g(X)\geq\varepsilon\}}\big] \\ &= \varepsilon E\big[\mathbb{1}_{\{c-g(X)\geq\varepsilon\}}\big] \\ &= \varepsilon P(c-g(X)\geq\varepsilon) \\ &\geq \varepsilon\eta \\ &> 0, \end{aligned} \tag{3}$$

where: the first inequality holds by monotonicity of expectation; the first equality holds by linearity of expectation; the second inequality holds by (2). Since (3) contradicts (1), the desired result follows. ■

(b) By Lemma 3.6, if the measurable function $g\colon \mathbb{R}^k \to \mathbb{R}$ minimizes $E\big[(Y-g(c))^2\big]$, then $g(c) = E[Y \mid c]$ almost surely. Since c is a constant in $\mathbb{R}^k$, $g(c)$ is a real constant for any measurable function $g\colon \mathbb{R}^k \to \mathbb{R}$, and so we can use the standard calculus approach to find $g(c)$. Set $a := g(c)$. We want to solve

$$\min_{a\in\mathbb{R}} E\big[(Y-a)^2\big].$$

By Exercise 37, we know that $a = E[Y]$ is the unique solution to the previous minimization problem. The desired result follows. ■

(c) (i) By definition, $E[Y \mid X]$ satisfies

$$E[Yh(X)] = E[E[Y \mid X]h(X)]$$

for any bounded and Borel measurable function $h\colon \mathbb{R}^k \to \mathbb{R}$. Define h as $h(x) := 1$ for all $x \in \mathbb{R}^k$. Clearly, h is bounded and Borel measurable. Then, we have

$$E[Y \cdot 1] = E[E[Y \mid X] \cdot 1],$$

that is

$$E[Y] = E[E[Y \mid X]],$$

which gives the desired result. ■

(ii) Let $g(X,Z)$ be a version of the conditional expectation of $E[Y \mid X]$ given X and Z. Then, we have

$$E[E[Y \mid X]h(X,Z)] = E[g(X,Z)h(X,Z)],$$

or, equivalently, using linearity of expectation,

$$E\big[\big(E[Y \mid X] - g(X,Z)\big)h(X,Z)\big] = 0 \tag{4}$$

for every bounded and Borel measurable function $h\colon \mathbb{R}^k \times \mathbb{R}^m \to \mathbb{R}$. By way of contradiction, suppose that $g(X,Z) = E[Y \mid X]$ almost surely does not hold. Then, there exist $\varepsilon_1, \eta_1 > 0$ such that

$$P(E[Y \mid X] - g(X,Z) \geq \varepsilon_1) \geq \eta_1 > 0,$$

or there exist $\varepsilon_2, \eta_2 > 0$ such that

$$P(g(X,Z) - E[Y \mid X] \geq \varepsilon_2) \geq \eta_2 > 0,$$

or both. Assume without loss of generality that

$$P(E[Y \mid X] - g(X,Z) \geq \varepsilon) \geq \eta > 0, \tag{5}$$

for some $\varepsilon, \eta > 0$, and define $h\colon \mathbb{R}^k \times \mathbb{R}^m \to \mathbb{R}$ as

$$h(X,Z) := \mathbb{1}_{\{E[Y|X]-g(X,Z)\geq\varepsilon\}}.$$

Clearly, h is bounded and Borel measurable. However,

$$\begin{aligned} E\big[\big(E[Y \mid X] - g(X,Z)\big)\mathbb{1}_{\{E[Y|X]-g(X,Z)\geq\varepsilon\}}\big] &\geq E\big[\varepsilon\mathbb{1}_{\{E[Y|X]-g(X,Z)\geq\varepsilon\}}\big] \\ &= \varepsilon E\big[\mathbb{1}_{\{E[Y|X]-g(X,Z)\geq\varepsilon\}}\big] \\ &= \varepsilon P(E[Y \mid X] - g(X,Z) \geq \varepsilon) \qquad (6) \\ &\geq \varepsilon\eta \\ &> 0, \end{aligned}$$

where: the first inequality holds by monotonicity of expectation; the first equality holds by linearity of expectation; the second inequality holds by (5). Since (6) contradicts (4), the desired result follows. ■

(iii) Let $g(X)$ be a version of the conditional expectation of Y given X. We want to show that $g(X)$ is also a version of the conditional expectation of Y given X and $f(X)$; that is,

$$E[Y\tilde{h}(X, f(X))] = E[g(X)\tilde{h}(X, f(X))]$$

for every bounded and Borel measurable function $\tilde{h}\colon \mathbb{R}^k \times \mathbb{R}^m \to \mathbb{R}$. Consider an arbitrary such $\tilde{h}$. Since $g(X)$ be a version of $E[Y \mid X]$, we have

$$E[Yh(X)] = E[g(X)h(X)],$$

for every bounded and Borel measurable function $h\colon \mathbb{R}^k \to \mathbb{R}$. In particular, define $h\colon \mathbb{R}^k \to \mathbb{R}$ as $h(x) = \tilde{h}(x, f(x))$ for all $x \in \mathbb{R}^k$. Note that such h is bounded because $\tilde{h}$ is so, and is measurable because the identity function and f are so. Then, we have we have

$$E[Y\tilde{h}(X, f(X))] := E[Yh(X)] = E[g(X)h(X)] := E[g(X)\tilde{h}(X, f(X))],$$

Since $\tilde{h}$, was chosen arbitrarily, the desired result follows. ■

Exercise 42

Let $(X,Y)'$ be a jointly continuous $\mathbb{R}^2$-valued random variable with joint density $f_{XY}\colon \mathbb{R}^2 \to [0, +\infty)$ defined as

$$f_{XY}(x,y) := \begin{cases} \frac{1}{y} & \text{if } 0 < x < y < 1 \\ 0 & \text{otherwise} \end{cases}.$$

(i) Determine the conditional density of X given Y. What is the cdf of X given $Y = y$?

(ii) Determine $E[Y]$, $E[X \mid Y]$, and $E[X^2 \mid Y]$.

Solution

(i) We briefly answer the question omitting details. We have

$$f_{X|Y}(x \mid y) := \frac{f_{XY}(x,y)}{f_Y(y)}$$

for any y such that $f_Y(y) > 0$. Note that for $0 < y < 1$ we have

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x,y)\,\mathrm{d}x = \int_0^y \frac{1}{y}\,\mathrm{d}x = \frac{1}{y}\int_0^y \mathrm{d}x = \frac{1}{y}x\Big|_{x=0}^{y} = 1,$$

and $f_Y(y) = 0$ otherwise. Therefore,

$$f_{X|Y}(x \mid y) = \begin{cases} \frac{1}{y} & \text{if } 0 < x < y < 1 \\ \text{any density} & \text{otherwise} \end{cases}.$$

Moreover, for $x \in (0, y)$ we have

$$F_{X|Y}(x \mid y) = \int_0^x \frac{1}{y}\,\mathrm{d}s = \frac{1}{y}s\Big|_{s=0}^{x} = \frac{x}{y},$$

$F_{X|Y}(x \mid y) = 0$ for $x \le 0$, and $F_{X\mid Y}(x \mid y) = 1$ for $x \ge y$.

(ii) We have

$$E[Y] = \int_{-\infty}^{+\infty} y f_Y(y)\,\mathrm{d}y = \int_0^1 y\,\mathrm{d}y = \frac{y^2}{2}\Big|_{y=0}^{1} = \frac{1}{2}.$$

Moreover,

$$E[X \mid Y = y] = \int_{-\infty}^{+\infty} x f_{X|Y}(x \mid y)\,\mathrm{d}x = \int_0^y x\frac{1}{y}\,\mathrm{d}x = \frac{x^2}{2y}\Big|_{x=0}^{y} = \frac{y}{2}.$$

Finally,

$$E[X^2 \mid Y = y] = \int_{-\infty}^{+\infty} x^2 f_{X|Y}(x \mid y)\,\mathrm{d}x = \int_0^y x^2\frac{1}{y}\,\mathrm{d}x = \frac{x^3}{3y}\Big|_{x=0}^{y} = \frac{y^2}{3}. \qquad \blacksquare$$

E 703: Advanced Econometrics I
Solutions to Problem Set 8

Niccolò Lomys

December 1, 2016

Exercise 43

Solve the following problems.

(a) Let Y_1, Y_2 be real-valued random variables and X an $\mathbb{R}^k$-valued random variable on the same probability space $(\Omega, \mathcal{A}, P)$, with $Y_1.Y_2 \in L^1(\Omega, \mathcal{A}, P)$, and let $\alpha_1, \alpha_2 \in \mathbb{R}$. Show that $E[\alpha_1 Y_1 + \alpha_2 Y_2 \mid X] = \alpha_1 E[Y_1 \mid X] + \alpha_2 E[Y_2 \mid X]$ almost surely.

(b) Let X and Y be real-valued random variables on some probability space $(\Omega, \mathcal{A}, P)$, with $Y \in L^2(\Omega, \mathcal{A}, P)$. Show that $P(|Y| \geq \delta \mid X) \leq \delta^{-2} E[Y^2 \mid X]$ almost surely for any $\delta > 0$.

Solution

(a) We want to show that

$$E[(\alpha_1 Y_1 + \alpha_2 Y_2) h(X)] = E\big[(\alpha_1 E[Y_1 \mid X] + \alpha_2 E[Y_2 \mid X]) h(X)\big]$$

for every bounded and Borel measurable function $h \colon \mathbb{R}^k \to \mathbb{R}$. Consider an arbitrary such h. We have

$$\begin{aligned}
E[(\alpha_1 Y_1 + \alpha_2 Y_2) h(X)] &= E[\alpha_1 Y_1 h(X) + \alpha_2 Y_2 h(X)] \\
&= \alpha_1 E[Y_1 h(X)] + \alpha_2 E[Y_2 h(X)] \\
&= \alpha_1 E\big[E[Y_1 \mid X] h(X)\big] + \alpha_2 E\big[E[Y_2 \mid X] h(X)\big] \\
&= E\big[\alpha_1 E[Y_1 \mid X] h(X) + \alpha_2 E[Y_2 \mid X] h(X)\big] \\
&= E\big[(\alpha_1 E[Y_1 \mid X] + \alpha_2 E[Y_2 \mid X]) h(X)\big],
\end{aligned}$$

where: the second and fourth equality holds by linearity of expectation and the fact that $\alpha_1, \alpha_2 \in \mathbb{R}$, and $Y_1, Y_2, h(X) \in L^1(\Omega, \mathcal{A}, P)$; the third equality holds because, by definition of conditional expectation of Y_k given X, $k = 1, 2$, we have $E[Y_k h(X)] = E[E[Y_k \mid X] h(X)]$ for the bounded and measurable function $h \colon \mathbb{R}^k \to \mathbb{R}$. Then,

$$E[\alpha_1 Y_1 + \alpha_2 Y_2 \mid X] = \alpha_1 E[Y_1 \mid X] + \alpha_2 E[Y_2 \mid X] \qquad \text{a.s.}$$

by uniqueness of conditional expectation. ■

(b) Note that

$$\delta^2 \mathbb{1}_{\{|Y| \geq \delta\}} \leq Y^2.$$

Then, by monotonicity and linearity of conditional expectation, as $Y \in L^2(\Omega, \mathcal{A}, P)$,

$$\delta^2 E\big[\mathbb{1}_{\{|Y| \geq \delta\}} \mid X\big] \leq E[Y^2 \mid X] \quad \text{a.s.,}$$

i.e.,

$$E\big[\mathbb{1}_{\{|Y|\geq\delta\}} \mid X\big] \leq \frac{E[Y^2 \mid X]}{\delta^2} \quad \text{a.s.}$$

Observing that

$$E\big[\mathbb{1}_{\{|Y|\geq\delta\}} \mid X\big] = P(|Y| \geq \delta \mid X),$$

the desired result follows. ■

*Exercise 44

Let Y be a real-valued random variable and X an $\mathbb{R}^k$-valued random variable on a common probability space $(\Omega, \mathcal{A}, P)$, with $Y \in L^2(\Omega, \mathcal{A}, P)$.

(a) Show that $\operatorname{Var}(Y \mid X) = E[Y^2 \mid X] - (E[Y \mid X])^2$ almost surely.

(b) Suppose that X and Y are independent. Show that $\operatorname{Var}(Y \mid X) = \operatorname{Var}(Y)$ almost surely.

(c) Show that $\operatorname{Var}(Y) = E[\operatorname{Var}(Y \mid X)] + \operatorname{Var}(E[Y \mid X])$.

Solution

(a) As $Y \in L^2(\Omega, \mathcal{A}, P)$, we have

$$\begin{aligned}
\operatorname{Var}(Y \mid X) &:= E\big[(Y - E[Y \mid X])^2 \mid X\big] \\
&= E\big[Y^2 - 2YE[Y \mid X] + (E[Y \mid X])^2 \mid X\big] \\
&= E[Y^2 \mid X] - 2E[YE[Y \mid X] \mid X] + E[(E[Y \mid X])^2 \mid X] \quad \text{a.s.} \\
&= E[Y^2 \mid X] - 2E[Y \mid X]E[Y \mid X] + E\big[(E[Y \mid X])^2 \mid X\big] \quad \text{a.s.} \\
&= E[Y^2 \mid X] - 2(E[Y \mid X])^2 + (E[Y \mid X])^2 \quad \text{a.s.} \\
&= E\big[Y^2 \mid X\big] - (E[Y \mid X])^2,
\end{aligned}$$

where: the second equality holds by linearity of conditional expectation; the third and fourth equalities hold by the law of iterated expectations (Theorem 3.10-(iv) in the Lecture Notes). The claim follows. ■

(b) As $Y \in L^2(\Omega, \mathcal{A}, P)$, we have

$$\begin{aligned}
\operatorname{Var}(Y \mid X) &= E[Y^2 \mid X] - (E[Y \mid X])^2 \quad \text{a.s.} \\
&= E[Y^2] - (E[Y])^2 \quad \text{a.s.} \\
&= \operatorname{Var}(Y),
\end{aligned}$$

where: the first equality holds by part (a); the second equality holds by independence of X and Y and Proposition 3.11 in the Lecture Notes. The claim follows. ■

(c) As $Y \in L^2(\Omega, \mathcal{A}, P)$, we have

$$\begin{aligned}
E[\operatorname{Var}(Y \mid X)] + \operatorname{Var}(E[Y \mid X]) &= E\big[E[Y^2 \mid X] - (E[Y \mid X])^2\big] \\
&\quad + E\big[(E[Y \mid X])^2\big] - \big(E[E[Y \mid X]]\big)^2 \\
&= E[E[Y^2 \mid X]] - E\big[(E[Y \mid X])^2\big] \\
&\quad + E\big[(E[Y \mid X])^2\big] - (E[Y])^2 \\
&= E[Y^2] - (E[Y])^2 \\
&= \operatorname{Var}(Y),
\end{aligned}$$

where: the first equality follows by part (a); the second equality holds by linearity of conditional expectations and the law of iterated expectations; the third equality holds by the law of iterated expectations. ■

Exercise 45

Let $X \sim U(-1, 1)$, and define $Y := X^2$.

(i) Is $(X, Y)'$ jointly discrete or jointly continuous?

(ii) Determine $E[Y \mid X]$ by using Lemma 3.6 in the Lecture Notes.

(iii) Determine $E[X \mid Y]$.

Solution

(i) Note that the range of $(X, Y)'$, that we denote with $R\big((X, Y)'\big)$, always lies on the curve $y = x^2$, with $-1 \le x \le 1$. The graph of this function is a subset of $[-1, 1] \times [0, 1]$ of zero (Lebesgue) measure. That is, if $G := \big\{(x, y) \in [-1, 1] \times [0, 1] : y = x^2\big\}$ denotes the graph of this function, $\mu_{\mathcal{L}}(G) = 0$, where $\mu_{\mathcal{L}}$ is the Lebesgue measure on $([-1, 1] \times [0, 1], \mathcal{B}([-1, 1] \times [0, 1]))$. Suppose $(X, Y)'$ is jointly continuous. Then, $(X, Y)'$ has a density, i.e., there exists a function $f_{XY} : \mathbb{R}^2 \to [0, +\infty)$ such that

$$P\big(R((X, Y)') \in G\big) = \int\int_G f_{XY}(x, y)\, \mathrm{d}\mu_{\mathcal{L}}(x, y).$$

But the integral of a function over a set of zero Lebesgue measure is always zero, contradicting that $P((X, Y)' \in G) = 1$. Thus, $(X, Y)'$ cannot have a density, and so is not jointly continuous. Moreover, as the range of $(X, Y)'$ is uncountable, $(X, Y)'$ cannot be jointly discrete either.

(ii) By Lemma 3.6 in the Lecture Notes, if we find a Borel measurable function $g : \mathbb{R} \to \mathbb{R}$ that minimizes

$$E\big[(Y - g(X))^2\big], \tag{1}$$

then $g(X) = E[Y \mid X]$. As $Y := X^2$, minimizing (1) with respect to $g(X)$ is equivalent to minimize

$$E\big[(X^2 - g(X))^2\big].$$

with respect to $g(X)$. Since $(X^2 - g(X))^2 \ge 0$, by monotonicity of expectation we have

$$E\big[(X^2 - g(X))^2\big] \ge 0$$

for any Borel measurable $g : \mathbb{R} \to \mathbb{R}$, and

$$E\big[(X^2 - g(X))^2\big] = 0$$

if $g(X) = X^2$. Note that $g(X) = X^2$ is Borel measurable.[1] Thus, we conclude that $E[Y \mid X] = X^2 := Y$ almost surely.

(iii) Since $Y := X^2$, and $X \sim U(-1.1)$, for any $y \in [0, 1]$ we have

$$\begin{aligned} E[X \mid Y = y] &= E[X \mid X^2 = y] \\ &= -\sqrt{y}P(X < 0) + \sqrt{y}P(X \ge 0) \\ &= -\sqrt{y}\frac{1}{2} + \sqrt{y}\frac{1}{2} \\ &= 0. \end{aligned}$$

Then, $E[X \mid Y] = 0$. ■

[1] *Quiz:* Why?

Exercise 46

Let $X_1, \ldots, X_n$ be i.i.d. real-valued random variables, with $E[X_i] = \mu \in \mathbb{R}$ and $\mathrm{Var}(X_i) = \sigma^2 \in \mathbb{R}$. Let $\overline{X}_n := n^{-1} \sum_{i=1}^n X_i$ be the sample mean. An estimator for σ^2 is the sample variance

$$S_n^2 := \frac{1}{n-1} \sum_{i=1}^n \left(X_i - \overline{X}_n\right)^2.$$

Show that S_n^2 is unbiased for σ^2.

Solution

We have

$$\begin{aligned} E\left[S_n^2\right] &:= E\left[\frac{1}{n-1} \sum_{i=1}^n \left(X_i - \overline{X}_n\right)^2\right] \\ &= \frac{1}{n-1} E\left[\sum_{i=1}^n X_i^2 + \sum_{i=1}^n \overline{X}_n^2 - 2\overline{X}_n \sum_{i=1}^n X_i\right] \\ &= \frac{n}{n-1} E\left[X_i^2\right] + \frac{1}{n-1}\left(nE\left[\overline{X}_n^2\right] - 2nE\left[\overline{X}_n^2\right]\right) \\ &= \frac{n}{n-1}\left(E\left[X_i^2\right] - E\left[\overline{X}_n^2\right]\right). \end{aligned} \tag{2}$$

Moreover,

$$\mathrm{Var}\left(\overline{X}_n\right) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2} n \mathrm{Var}(X_i) = \frac{\sigma^2}{n} = \frac{E\left[X_i^2\right] - \mu^2}{n},$$

where the third equality holds because $X_1 \cdots, X_n$ are i.i.d. Since $\mathrm{Var}\left(\overline{X}_n\right) = E\left[\overline{X}_n^2\right] - \left(E\left[\overline{X}_n\right]\right)^2$,

$$E\left[\overline{X}_n^2\right] = \mathrm{Var}\left(\overline{X}_n\right) + \left(E\left[\overline{X}_n\right]\right)^2 = \frac{E\left[X_i^2\right] - \mu^2}{n} + \mu^2 = \frac{E\left[X_i^2\right] + (n-1)\mu^2}{n}, \tag{3}$$

where we used that $E\left[\overline{X}_n\right] = \mu$. Plugging (3) into (2), we have

$$\begin{aligned} E\left[S_n^2\right] &= \frac{n}{n-1}\left(E\left[X_i^2\right] - E\left[\overline{X}_n^2\right]\right) \\ &= \frac{n}{n-1}\left(E\left[X_i^2\right] - \frac{E\left[X_i^2\right] - (n-1)\mu^2}{n}\right) \\ &= \frac{n}{n-1}\left(\frac{(n-1)\left(E\left[X_i^2\right] - \mu^2\right)}{n}\right) \\ &= \frac{n}{n-1}\frac{(n-1)\sigma^2}{n} \\ &= \sigma^2, \end{aligned}$$

which shows that S_n^2 is unbiased for σ^2. ■

*Exercise 47

Solve the following problems.

(a) Let $(X_n)_n$ be a sequence of i.i.d. random variables, with $X_n \sim Exp(\lambda)$. By using Lemma 4.7 in the Lecture Notes, construct an asymptotic $(1-\alpha)$-confidence interval for λ.

(b) Let $(X_n)_n$ be a sequence of i.i.d. binomially distributed random variables with parameters $n \in \mathbb{N}$ and $p \in (0,1)$. Determine the method of moments estimators for the parameters n and p.

Solution

(a) Define $\hat{T}_n := 1/\overline{X}_n$, with $\overline{X}_n := \frac{1}{n}\sum_{k=1}^n X_k$. By Exercise 36, we know that

$$\sqrt{n}(\hat{T}_n - \lambda) \xrightarrow{d} Z \sim N(0, \lambda^2).$$

Since $\mathrm{Var}(X_1) = 1/\lambda^2$, and the sample variance $\hat{S}_n^2 := (n-1)^{-1}\sum_{i=1}^n \left(X_i - \overline{X}_n\right)^2$ is a consistent estimator for $\mathrm{Var}(X_1)$, by the continuous mapping theorem for convergence in probability we have that $1/\hat{S}_n$ is a consistent estimator for λ. By Lemma 4.7 in the Lecture Notes, an asymptotic $(1-\alpha)$-confidence interval for λ is

$$\left[\hat{T}_n - z_{1-\alpha/2}\frac{1/\hat{S}_n}{\sqrt{n}}, \hat{T}_n + z_{1-\alpha/2}\frac{1/\hat{S}_n}{\sqrt{n}}\right],$$

where

$$\hat{T}_n := \frac{1}{\overline{X}_n} \qquad \text{and} \qquad \hat{S}_n := \sqrt{\frac{\sum_{i=1}^n \left(X_i - \overline{X}_n\right)^2}{n-1}},$$

and $z_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of $Z \sim N(0,1)$. ■

(b) Since $m_1 := E[X_1] = np$, and $m_2 := E[X_1^2] = np(1-p) + n^2p^2$, solving

$$\begin{cases} m_1 = np \\ m_2 = np(1-p) + n^2p^2 \end{cases}$$

for n and p we find

$$n = \frac{m_1^2}{m_1 - m_2 + m_1^2} \qquad \text{and} \qquad p = \frac{m_1 - m_2 + m_1^2}{m_1}.$$

Let $\hat{m}_1 := n^{-1}\sum_{i=1}^n X_i$ and $\hat{m}_2 = n^{-1}\sum_{i=1}^n X_i^2$ be the empirical moments, which are consistent estimators of the respective population moments by the weak law of large numbers and the continuous mapping theorem for convergence in probability. The method of moments estimators for n and p, denoted as $\hat{n}$ and $\hat{p}$, are given by

$$\hat{n} = \frac{\hat{m}_1^2}{\hat{m}_1 - \hat{m}_2 + \hat{m}_1^2} \qquad \text{and} \qquad \hat{p} = \frac{\hat{m}_1 - \hat{m}_2 + \hat{m}_1^2}{\hat{m}_1}. \qquad ■$$

Exercise 48

Solve the following problems.

(a) Let $(\Omega, \mathcal{A})$ be a measurable space and X a real-valued random variable on $(\Omega, \mathcal{A})$. The σ-*field generated by* X, denoted as $\sigma(X)$, is the smallest σ-field Σ on Ω such that X is Σ-$\mathcal{B}(\mathbb{R})$ measurable.

(i) Suppose that X is a constant random variable on $(\Omega, \mathcal{A})$ (i.e., $X(\omega) := c \in \mathbb{R}$ for all $\omega \in \Omega$). What is $\sigma(X)$?

(ii) Let $A \in \mathcal{A}$. What is $\sigma(\mathbb{1}_A)$, where $\mathbb{1}_A$ is the indicator function of A?

(iii) Let $\Omega := \{1, \dots, 6\}$, $\mathcal{A} := \mathcal{P}(\Omega)$, and X such that $X(\omega) := \omega$ for all $\omega \in \Omega$. What is $\sigma(X)$?

(iv) Suppose that X is a real-valued random variable on $(\Omega, \mathcal{A})$. Show that

$$\sigma(X) = \sigma\big(\{X^{-1}(C) \in \mathcal{P}(\Omega) : C \in \mathcal{C}\}\big)$$

for any collection $\mathcal{C}$ of subsets of $\mathbb{R}$ such that $\sigma(\mathcal{C}) = \mathcal{B}(\mathbb{R})$. [*Hint.* Good set technique.]

(b) Let $(\Omega, \mathcal{A})$ be a measurable space, and suppose that X and Y are real-valued random variables on $(\Omega, \mathcal{A})$. Show that $X + Y$ and XY are (real-valued) random variables on $(\Omega, \mathcal{A})$.

Solution

(a) (i) We have $\sigma(X) = \{\emptyset, \Omega\}$.

(ii) We have $\sigma(X) = \{\emptyset, A, \Omega \setminus A, \Omega\}$.

(iii) We have $\sigma(X) = \mathcal{P}(\Omega)$.

(iv) Set

$$\mathcal{T} := \sigma\big(\{X^{-1}(C) \in \mathcal{P}(\Omega) : C \in \mathcal{C}\}\big).$$

We want to show that

$$\sigma(X) = \mathcal{T}.$$

[$\supseteq$] By Exercise 40-(a),

$$\sigma(X) = \{X^{-1}(B) \in \mathcal{P}(\Omega) : B \in \mathcal{B}(\mathbb{R})\}.$$

Since $\mathcal{C} \subseteq \mathcal{B}(\mathbb{R})$,

$$\{X^{-1}(C) \in \mathcal{P}(\Omega) : C \in \mathcal{C}\} \subseteq \{X^{-1}(B) \in \mathcal{P}(\Omega) : B \in \mathcal{B}(\mathbb{R})\}.$$

That $\mathcal{T} \subseteq \sigma(X)$ immediately follows.

[$\subseteq$] Consider the following class of subsets of $\mathbb{R}$:

$$\mathcal{G} := \{S \in \mathcal{B}(\mathbb{R}) : X^{-1}(S) \in \mathcal{T}\}.$$

First, we show that $\mathcal{G}$ is a σ-field on $\mathbb{R}$ that contains $\mathcal{C}$. To see that $\mathcal{C} \subseteq \mathcal{G}$, consider an arbitrary $C \in \mathcal{C}$. Then, $X^{-1}(C) \in \mathcal{T}$ by definition of $\mathcal{T}$, and so $C \in \mathcal{G}$ by definition of $\mathcal{G}$. Since $\mathcal{C} \neq \emptyset$, it follows that $\mathcal{G} \neq \emptyset$. To see that $\mathcal{G}$ is closed under complementation, let $G \in \mathcal{G}$. Then, $X^{-1}(G) \in \mathcal{T}$ by definition of $\mathcal{G}$. Since $\mathcal{T}$ is a σ-field on Ω, $\Omega \setminus X^{-1}(G) = X^{-1}(\mathbb{R} \setminus G) \in \mathcal{T}$. But then, $\mathbb{R} \setminus G \in \mathcal{G}$ by definition of $\mathcal{G}$. To see that $\mathcal{G}$ is closed under taking countably infinite unions, let $G_1, G_2, \dots \in \mathcal{G}$. Then, $X^{-1}(G_1) \in \mathcal{T}, X^{-1}(G_2) \in \mathcal{T}, \dots$ by definition of $\mathcal{G}$. Since $\mathcal{T}$ is a σ-field,

$$\bigcup_{n=1}^{\infty} X^{-1}(G_n) = X^{-1}\bigg(\bigcup_{n=1}^{\infty} G_n\bigg) \in \mathcal{T}.$$

But then, $\bigcup_{n=1}^{\infty} G_n \in \mathcal{G}$ by definition of $\mathcal{G}$. We conclude that $\mathcal{G}$ is a σ-field on $\mathbb{R}$ that contains $\mathcal{C}$, as we claimed. Therefore, $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{C}) \subseteq \mathcal{G}$. Since $\mathcal{G} \subseteq \mathcal{B}(\mathbb{R})$ by definition of $\mathcal{G}$, we have $\mathcal{B}(\mathbb{R}) = \mathcal{G}$. That is, for any $B \in \mathcal{B}(\mathbb{R})$, $X^{-1}(B) \in \mathcal{T}$. But then, $\sigma(X) \subseteq \mathcal{T}$, as we wanted to show. ■

(b) We first show that $X+Y$ is $\mathcal{A}$-$\mathcal{B}(\mathbb{R})$ measurable. By Exercise 14-(a), it is enough to show that, for each $a \in \mathbb{R}$,

$$A := (X+Y)^{-1}((-\infty, a)) := \{\omega \in \Omega : X(\omega) + Y(\omega) < a\} \in \mathcal{A}.$$

Fix $a \in \mathbb{R}$. Suppose that all the rationals are arranged in a sequence $\{q_n\}_{n \in \mathbb{N}}$. Now, note that

$$A = \bigcup_{n \in \mathbb{N}} \{\omega \in \Omega : X(\omega) < q_n, Y(\omega) < a - q_n\}.$$

[We decomposed the half plane below the line $x + y = a$ into a countable union of bounded boxes $\{(x, y) \in \mathbb{R}^2 : x < q_n, y < a - q_n\}$.] Clearly,

$$\begin{aligned} \{\omega \in \Omega : X(\omega) < q_n, Y(\omega) < a - q_n\} &= \{\omega \in \Omega : X(\omega) < q_n\} \cap \{\omega \in \Omega : Y(\omega) < a - q_n)\} \\ &= X^{-1}((-\infty, q_n)) \cap Y^{-1}((-\infty, a - q_n)) \end{aligned}$$

is an element of $\mathcal{A}$ as an intersection of sets in $\mathcal{A}$ (X and Y are $\mathcal{A}$-$\mathcal{B}(\mathbb{R})$ measurable). Hence $A \in \mathcal{A}$ as a countable union of elements of $\mathcal{A}$.

Next, we show that XY is $\mathcal{A}$-$\mathcal{B}(\mathbb{R})$ measurable. Note that if Y is $\mathcal{A}$-$\mathcal{B}(\mathbb{R})$ measurable, then so is $-Y$ (prove it!). Hence, $X - Y = X + (-Y)$ is $\mathcal{A}$-$\mathcal{B}(\mathbb{R})$ measurable. Then as

$$XY = \frac{1}{4}\left\{(X+Y)^2 - (X-Y)^2\right\},$$

it will suffice to prove that the square of $\mathcal{A}$-$\mathcal{B}(\mathbb{R})$ measurable functions is $\mathcal{A}$-$\mathcal{B}(\mathbb{R})$ measurable. So let $Z\colon \Omega \to \mathbb{R}$ be an arbitrary $\mathcal{A}$-$\mathcal{B}(\mathbb{R})$ measurable function, fix an arbitrary $a \in \mathbb{R}$, and consider the set $\left\{\omega \in \Omega : Z^2(\omega) > a\right\}$. For $a < 0$, this set is $\Omega \in \mathcal{A}$, and for $a \geq 0$

$$\left\{\omega \in \Omega : Z^2(\omega) > a\right\} = \left\{\omega \in \Omega : Z(\omega) > \sqrt{a}\right\} \cup \left\{\omega \in \Omega : Z(\omega) < -\sqrt{a}\right\}.$$

Both sets on the right hand side are elements of $\mathcal{A}$, as Z is $\mathcal{A}$-$\mathcal{B}(\mathbb{R})$ measurable. Hence, we have shown that Z^2 is $\mathcal{A}$-$\mathcal{B}(\mathbb{R})$ measurable. Apply this with $Z := X+Y$ and $Z := X-Y$ respectively to conclude that XY is $\mathcal{A}$-$\mathcal{B}(\mathbb{R})$ measurable. It follows that cX is $\mathcal{A}$-$\mathcal{B}(\mathbb{R})$ measurable for any real constant c. ■

Remark. Exercise 48-(b) tells us that the class of real-valued random variables on some measurable space $(\Omega, \mathcal{A})$ is a vector space under (pointwise) addition.

Exercise 49 (Optional - Not Graded)

Solve the following problems.

(a) Construct a probability space $(\Omega, \mathcal{A}, P)$ and a sequence of random variables $(X_n)_n$ on $(\Omega, \mathcal{A}, P)$ with $X_n(\omega) \underset{n\to\infty}{\longrightarrow} 0$ for all $\omega \in \Omega$, but where $E[X_n]$ does not converge to zero. Which insight can you draw from this example?

(b) Let $(U_k)_{k \in \mathbb{N}}$ be a sequence of i.i.d. uniformly distributed random variables over $[0, 1]$, and define $Y_n := \left(\prod_{k=1}^n U_k\right)^{-1/n}$. Show that $\sqrt{n}(Y_n - e) \xrightarrow{d} N(0, e^2)$. [*Hint.* Set $X_k := -\ln U_k$ and use the delta method.]

Solution

(a) Set $\Omega := [0,1]$, $\mathcal{A} := \mathcal{B}([0,1])$, and let P be the unique probability measure on the measurable space $([0,1], \mathcal{B}([0,1]))$ induced by the pdf of the uniform distribution over $[0,1]$ (cf. Corollary 1.18 in the Lecture Notes). Consider the sequence of discrete real-valued random variables $(X_n)_{n\in\mathbb{N}}$ on $([0,1], \mathcal{B}([0,1]), P)$ where, for each n, X_n is defined as

$$X_n(\omega) := \begin{cases} n & \text{if } 0 < \omega \leq \frac{1}{n} \\ 0 & \text{otherwise} \end{cases} .$$

Clearly, $\lim_{n\to\infty} X_n(\omega) = 0$ for all $\omega \in [0,1]$, and therefore $E[\lim_{n\to\infty} X_n] = 0$. However, for each n we have

$$E[X_n] = 0 \cdot P(X_n = 0) + n \cdot P(X_n = n) = 0 \cdot \frac{n-1}{n} + n \cdot \frac{1}{n} = 1.$$

Therefore, as $E[X_n] = 1$ for all n, $\lim_{n\to\infty} E[X_n] = 1 \neq 0$.

This exercise shows that, in general, $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ for all $\omega \in \Omega$ is not sufficient for $\lim_{n\to\infty} E[X_n] = E[X]$. That is, we need additional assumptions to assure that $E[\lim_{n\to\infty} X_n] = \lim_{n\to\infty} E[X_n]$ (cf. Monotone Convergence Theorem and Dominated Convergence Theorem). ■

(b) Set $X_k := -\ln U_k$ for all $k \in \mathbb{N}$. Then,

$$\ln Y_n := \ln\left(\prod_{k=1}^{n} U_k\right)^{-1/n} = -\frac{1}{n}\sum_{i=1}^{n} \ln U_i := \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}_n,$$

and so $Y_n = \exp(\overline{X}_n)$. Since $(U_k)_{k\in\mathbb{N}}$ is sequence of i.i.d. random variables, so is $(X_k)_{k\in\mathbb{N}}$.[2] Moreover, we have

$$E[X_1] := E[-\ln U_1] = \int_0^1 -\ln u \, \mathrm{d}u = -u(\ln u - 1)\Big|_{u=0}^{1} = 1,$$

and

$$E[X_1^2] := E\big[(-\ln U_1)^2\big] = E[-2\ln U_1] = 2E[-\ln U_1] = 2,$$

and therefore,

$$\operatorname{Var}(X_1) = E[X_1^2] - (E[X_1])^2 = 2 - 1 = 1.$$

By the central limit theorem for i.i.d. sequences of random variables we have

$$\sqrt{n}(\overline{X}_n - 1) \xrightarrow{d} Z \sim N(0,1).$$

The function $\phi\colon \mathbb{R} \to \mathbb{R}$, $\phi(x) := e^x$, is differentiable at all x on its domain, and so in particular at $x = 1$, with

$$\begin{aligned} \phi(\overline{X}_n) &= \exp(\overline{X}_n), \\ \phi(1) &= e, \\ \text{and} \qquad \big(\phi'(1)\big)^2 &= e^2. \end{aligned}$$

Applying the delta-method we obtain

$$\sqrt{n}\big(\phi(\overline{X}_n) - \phi(1)\big) \xrightarrow{d} N\Big(0, \big(\phi'(1)\big)^2\Big),$$

that is,

$$\sqrt{n}\big(\exp(\overline{X}_n) - e\big) \xrightarrow{d} N(0, e^2).$$

As $\exp(\overline{X}_n) = Y_n$, the desired result follows. ■

---

[2] *Quiz:* Why?

E 703: Advanced Econometrics I
Solutions to Problem Set 9

Niccolò Lomys

December 8, 2016

Exercise 50

Solve the following problems.

(a) For a square matrix $C := (c_{i,j})_{1 \le i,j \le m}$, the *trace of* C is defined as $\operatorname{tr}(C) := \sum_{i=1}^{n} c_{i,i}$. Let $Z := (Z_{i,j})_{1 \le i,j \le m}$ be a random square matrix, i.e., $Z_{i,j}$ are real-valued random variables on a common probability space $(\Omega, \mathcal{A}, P)$. Assume that $Z_{i,j} \in L^1(\Omega, \mathcal{A}, P)$ for all $i, j = 1, \dots, m$. The expectation of Z is defined as $E[Z] := (E[Z_{i,j}])_{1 \le i,j \le m}$. Show that $E[\operatorname{tr}(Z)] = \operatorname{tr}(E[Z])$.

(b) Let $(Z_n)_n$, Z be random $m \times m$ matrices, and suppose that $Z_n \xrightarrow{P} Z$. Show that $\det(Z_n) \xrightarrow{P} \det(Z)$, where $\det(C)$ denotes the determinant of the square matrix C.

Solution

(a) We have

$$E[\operatorname{tr}(Z)] = E\left[\sum_{i=1}^{m} Z_{i,i}\right] = \sum_{i=1}^{m} E[Z_{i,i}] = \operatorname{tr}(E[Z]),$$

where the second equality holds by linearity of expectation and the fact that $Z_{i,i} \in L^1(\Omega, \mathcal{A}, P)$ for all $i = 1, \dots, m$. ■

(b) Endow the space $\mathbb{R}^m \times \mathbb{R}^m$ of real-valued $m \times m$ matrices with the Euclidean distance. As $\det\colon \mathbb{R}^{m \times m} \to \mathbb{R}$ is a continuous function, the claim follows by Lemma 2.12 in the Lecture Notes (continuous mapping theorem for convergence in probability). ■

Exercise 51

Solve the following problems.

(a) Let $F\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a continuous function. Moreover, assume that $f, g\colon \mathbb{R} \to \mathbb{R}$ are $\mathcal{B}(\mathbb{R})$-$\mathcal{B}(\mathbb{R})$ measurable functions. Show that $h\colon \mathbb{R}^2 \to \mathbb{R}$, defined as $h(x) := F(f(x), g(x))$ is $\mathcal{B}(\mathbb{R}^2)$-$\mathcal{B}(\mathbb{R})$ measurable.

(b) Let Y be a real-valued random variable, with $Y \in L_1(\Omega, \mathcal{A}, P)$, and let $n \in \mathbb{N}$. Show that

$$E\big[Y \mathbb{1}_{[n,\infty)}(Y)\big] \underset{n \to \infty}{\longrightarrow} 0.$$

(c) Let Y be a real-valued random variable, and X be an $\mathbb{R}^k$-valued random variable on a common probability space $(\Omega, \mathcal{A}, P)$. Suppose that $Y \in (\Omega, \mathcal{A}, P)$. Show that

$$|E[Y \mid X]| \le E[|Y| \mid X] \qquad \text{almost surely.}$$

(d) Let $X_1, \dots X_n$ be i.i.d. random variables with finite expectation, and let $S := \sum_{k=1}^n X_k$. Show that $E[X_1 \mid S] = S/n$ almost surely.

Solution

(a) By Exercise 14-(a), it is enough to show that

$$h^{-1}((a, +\infty)) \in \mathcal{B}(\mathbb{R}^2)$$

for any $a \in \mathbb{R}$. Fix $a \in \mathbb{R}$, and note that

$$h^{-1}((a, +\infty)) := \{x \in \mathbb{R}^2 : h(x) > a\} = \{x \in \mathbb{R}^2 : (f(x), g(x)) \in G_a\},$$

where

$$G_a := \{(u, v) \in \mathbb{R}^2 : F(u, v) > a\} := F^{-1}((a, +\infty)).$$

Suppose first that G_a is an open rectangle of the form

$$G_a = (a_1, b_1) \times (c_1, d_1)$$

for some real numbers a_1, b_1, c_1, d_1. Then,

$$\begin{aligned} h^{-1}((a, +\infty)) &= \{x \in \mathbb{R}^2 : (f(x), g(x)) \in G_a\} \\ &= \{x \in \mathbb{R}^2 : (f(x), g(x)) \in (a_1, b_1) \times (c_1, d_1)\} \\ &= \{x \in \mathbb{R}^2 : f(x) \in (a_1, b_1) \text{ and } g(x) \in (c_1, d_1)\} \\ &= \{x \in \mathbb{R}^2 : f(x) \in (a_1, b_1)\} \cap \{x \in \mathbb{R}^2 : g(x) \in (c_1, d_1)\} \\ &= f^{-1}((a_1, b_1)) \cap g^{-1}((c_1, d_1)). \end{aligned}$$

Since f and g are Borel measurable functions and $(a_1, b_1), (c_1, d_1) \in \mathcal{B}(\mathbb{R})$, $f^{-1}((a_1, b_1))$ and $g^{-1}((c_1, d_1))$ are in $\mathcal{B}(\mathbb{R}^2)$. Thus, $f^{-1}((a_1, b_1)) \cap g^{-1}((c_1, d_1)) \in \mathcal{B}(\mathbb{R}^2)$ because the σ-algebra $\mathcal{B}(\mathbb{R}^2)$ is closed under taking finite intersections. It follows that $h^{-1}((a, +\infty)) \in \mathcal{B}(\mathbb{R}^2)$. Now, suppose that the set G_a is not a rectangle. In this case, we decompose G_a into the countable union of rectangles. Since F is continuous, G_a is an open subset of $\mathbb{R}^2$ (endowed with the Euclidean distance). Hence, it can be written as

$$G_a = \bigcup_{n=1}^{\infty} R_n,$$

where R_n are open rectangles of the form $(a_n, b_n) \times (c_n, d_n)$. Therefore,

$$\begin{aligned} h^{-1}((a, +\infty)) &= \{x \in \mathbb{R}^2 : (f(x), g(x)) \in G_a\} \\ &= \left\{x \in \mathbb{R}^2 : (f(x), g(x)) \in \bigcup_{n=1}^{\infty} (a_n, b_n) \times (c_n, d_n)\right\} \\ &= \left\{x \in \mathbb{R}^2 : f(x) \in \bigcup_{n=1}^{\infty} (a_n, b_n) \text{ and } g(x) \in \bigcup_{n=1}^{\infty} (c_n, d_n)\right\} \\ &= \left\{x \in \mathbb{R}^2 : f(x) \in \bigcup_{n=1}^{\infty} (a_n, b_n)\right\} \cap \left\{x \in \mathbb{R}^2 : g(x) \in \bigcup_{n=1}^{\infty} (c_n, d_n)\right\} \\ &= f^{-1}\left(\bigcup_{n=1}^{\infty} (a_n, b_n)\right) \cap g^{-1}\left(\bigcup_{n=1}^{\infty} (c_n, d_n)\right) \end{aligned}$$

$$= \left(\bigcup_{n=1}^{\infty} f^{-1}((a_n, b_n)) \right) \cap \left(\bigcup_{n=1}^{\infty} g^{-1}((c_n, d_n)) \right)$$
$$= \bigcup_{n=1}^{\infty} \left(f^{-1}((a_n, b_n)) \cap g^{-1}((c_n, d_n)) \right),$$

where we used that inverse images behave well with respect to taking unions. As f and g are Borel measurable and $\mathcal{B}(\mathbb{R}^2)$ is closed under countable intersections and unions, we conclude that $h^{-1}((a, +\infty)) \in \mathcal{B}(\mathbb{R}^2)$. ■

(b) Omitted.

(c) Recall that $Y = Y^+ - Y^-$, and $|Y| = Y^+ + Y^-$. Then,

$$\begin{aligned} |E[Y \mid X]| &= |E[Y^+ - Y^- \mid X]| \\ &= |E[Y^+ \mid X] - [Y^- \mid X]| \quad \text{almost surely} \\ &\leq |E[Y^+ \mid X]| + |E[Y^- \mid X]| \\ &= E[Y^+ \mid X] + E[Y^- \mid X] \\ &= E[Y^+ + Y^- \mid X] \quad \text{almost surely} \\ &= E[|Y| \mid X], \end{aligned}$$

where: The second and fourth equality holds by linearity of conditional expectation, and the fact that $Y^+, Y^- \in L^1(\Omega, \mathcal{A}, P)$ as $Y \in L^1(\Omega, \mathcal{A}, P)$; the inequality follows by triangle inequality for absolute value; the third equality holds because Y^+ and Y^- are non-negative, and so are their conditional expectations. The claim follows. ■

(d) Since $X_1, \dots X_n$ are i.i.d. random variables,

$$E[X_1 \mid S] = E[X_2 \mid S] = \cdots = E[X_n \mid S],$$

and so

$$\sum_{k=1}^{n} E[X_k \mid S] = nE[X_1 \mid S]. \tag{1}$$

Moreover, as the X_k's have finite expectations,

$$\sum_{i=1}^{n} E[X_i \mid S] = E\left[\sum_{i=1}^{n} X_i \mid S \right] = E[S \mid S] = S \quad \text{a.s.}, \tag{2}$$

where: the first equality holds almost surely by linearity of conditional expectation; the second equality holds by definition of S; the third equality holds almost surely by the law of iterated expectations (Theorem 3.10.(iv) in the Lecture Notes). From (1) and (2), we have

$$nE[X_1 \mid S] = S,$$

and therefore

$$E[X_1 \mid S] = \frac{S}{n} \quad \text{a.s.}$$

The claim follows. ■

Exercise 52

Solve the following problems.

(a) Let $X \sim Exp(\lambda)$. Find the characteristic function of X, and use it to compute $E[X^k]$, for $k = 1, 2, 3, \ldots$.

(b) Let $(X_n)_n$ be a sequence of i.i.d. random variables, with $X_n \sim Exp(\lambda)$. Construct an asymptotic $(1-\alpha)$-confidence interval for λ based on

$$\hat{\lambda}_{2,n} := \left(\frac{2!}{\frac{1}{n}\sum_{i=1}^n X_i^2} \right)^{1/2}.$$

Solution

(a) Let $\varphi_X : \mathbb{R} \to \mathbb{C}$ be the characteristic function of X. For each $t \in \mathbb{R}$,

$$\varphi_X(t) := E\big[e^{itX}\big] = \int_{-\infty}^{+\infty} e^{itx} \lambda e^{-\lambda x} \mathbb{1}_{[0,+\infty)}(x)\mathrm{d}x = \frac{\lambda}{it-\lambda} e^{(it-\lambda)x}\Big|_{x=0}^{+\infty} = \frac{\lambda}{\lambda - it}.$$

For $k = 1, 2, 3, \ldots$, we know that

$$E[X^k] = i^{-k}\varphi_X^{(k)}(0).$$

By induction on k, it is easy to show (do it!) that

$$\varphi_X^{(k)}(t) = \frac{i^k k! \lambda}{(\lambda - it)^{k+1}},$$

and so

$$\varphi_X^{(k)}(0) = \frac{i^k k!}{\lambda^k}.$$

Therefore,

$$E[X^k] = i^{-k}\varphi_X^{(k)}(0) = i^{-k}\frac{i^k k!}{\lambda^k} = \frac{k!}{\lambda^k}$$

for $k = 1, 2, 3, \ldots$. ■

(b) For each i, we have (using part (a) of this exercise)

$$E[X_i^2] = \frac{2}{\lambda^2} \qquad \text{and} \qquad \mathrm{Var}(X_i^2) = E[X_i^4] - \left(E[X_i^2]\right)^2 = \frac{24}{\lambda^4} - \frac{4}{\lambda^4} = \frac{20}{\lambda^4}.$$

Since $(X_n)_n$ is a sequence of i.i.d. random variables, so is $(X_n^2)_n$.[1]. Then, by the central limit theorem for i.i.d. sequences of real-valued random variables,

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n X_i^2 - \frac{2}{\lambda^2} \right) \xrightarrow{d} N\left(0, \frac{20}{\lambda^4}\right).$$

The function $\phi : (0, +\infty) \to \mathbb{R}$, $\phi(x) := (2/x)^{1/2}$, is differentiable at all x on its domain, and so in particular at $x = 2/\lambda^2$, with

$$\phi\left(\frac{1}{n}\sum_{i=1}^n X_i^2 \right) = \hat{\lambda}_{2,n},$$

[1] *Quiz:* Why?

$$\phi(2/\lambda^2) = \lambda,$$

and
$$\left(\phi'(2/\lambda^2)\right)^2 = \frac{\lambda^6}{16}.$$

Applying the delta-method we obtain

$$\sqrt{n}\left(\phi\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) - \phi(2/\lambda^2)\right) \xrightarrow{d} N\left(0, \frac{20}{\lambda^4}\left(\phi'(2/\lambda^2)\right)^2\right),$$

that is,
$$\sqrt{n}(\hat{\lambda}_{2,n} - \lambda) \xrightarrow{d} N\left(0, \frac{5\lambda^2}{4}\right).$$

Since
$$E\left(\frac{2}{\frac{1}{n}\sum_{i=1}^{n} X_i^2}\right) = \lambda^2,$$

by the weak law of large numbers we have that

$$\frac{2}{\frac{1}{n}\sum_{i=1}^{n} X_i^2} \xrightarrow{P} \lambda^2,$$

and so, by the continuous mapping theorem for convergence in probability (observing that $f\colon [0, +\infty) \to \mathbb{R}$, $f(x) := \sqrt{5x/4}$, is a continuous function),

$$\frac{\sqrt{5}\hat{\lambda}_{2,n}}{2} \xrightarrow{P} \frac{\sqrt{5}\lambda}{2}.$$

Hence, by Lemma 4.7 in the Lecture Notes, an asymptotic $(1-\alpha)$-confidence interval for λ is

$$\left[\hat{\lambda}_{2,n} - z_{1-\alpha/2}\frac{\sqrt{5}\hat{\lambda}_{2,n}}{2\sqrt{n}}, \hat{\lambda}_{2,n} + z_{1-\alpha/2}\frac{\sqrt{5}\hat{\lambda}_{2,n}}{2\sqrt{n}}\right],$$

where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of $Z \sim N(0,1)$. ■

Exercise 53

Solve the following problems.

(a) Let $(X_n)_n$ be a sequence of i.i.d. log-normally distributed random variables with pdf $f\colon (0, +\infty) \to [0, +\infty)$ defined by

$$f(x) := \frac{1}{\sqrt{2\pi}\sigma x} \exp\left\{-\frac{(\log x - \mu)^2}{2\sigma^2}\right\},$$

where $\mu \in \mathbb{R}$ and $\sigma > 0$. Determine method of moments estimators for the parameters μ and σ^2. [*Hint.* The following holds true for all $k \in \mathbb{N}$: $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{ky - \frac{(y-\mu)^2}{2\sigma^2}}\, dy = e^{k\mu + \frac{k^2\sigma^2}{2}}$.]

(b) Let $X_1, \ldots, X_n$ be an i.i.d. sample of random variables, with X_i uniformly distributed over the closed interval $[0, \theta]$, i.e. $X_i \sim U([0, \theta])$. We want to estimate the unknown (boundary) parameter $\theta > 0$. Let $\overline{X}_n := n^{-1}\sum_{i=1}^{n} X_i$ be the sample mean.

(i) Show that $2\overline{X}_n$ is a consistent and unbiased estimator for θ. Determine its variance.

(ii) Determine $\hat{\theta}_{MLE}$, the maximum likelihood estimator for θ?

(iii) Determine a cdf of $\hat{\theta}_{MLE}$, its expectation and its variance.

Solution

(a) We have

$$
\begin{aligned}
m_1 := E[X_1] &= \int_0^{+\infty} x \frac{1}{\sqrt{2\pi}\sigma x} \exp\left\{-\frac{(\log x - \mu)^2}{2\sigma^2}\right\} \mathrm{d}x \\
&= \int_0^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(\log x - \mu)^2}{2\sigma^2}\right\} \mathrm{d}x \\
&= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^y e^{-\frac{(y-\mu)^2}{2\sigma^2}} \mathrm{d}y \\
&= e^{\mu+\sigma^2/2},
\end{aligned}
$$

where: the third equality holds via integration by substitution, with $y = \log x$; the fourth equality follows by the hint. Similarly,

$$
\begin{aligned}
m_2 := E[X_1^2] &= \int_0^{+\infty} x^2 \frac{1}{\sqrt{2\pi}\sigma x} \exp\left\{-\frac{(\log x - \mu)^2}{2\sigma^2}\right\} \mathrm{d}x \\
&= \int_0^{+\infty} \frac{x}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(\log x - \mu)^2}{2\sigma^2}\right\} \mathrm{d}x \\
&= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{2y} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \mathrm{d}y \\
&= e^{2(\mu+\sigma^2)}.
\end{aligned}
$$

Solving

$$
\begin{cases} m_1 = e^{\mu+\sigma^2/2} \\ m_2 = e^{2(\mu+\sigma^2)} \end{cases}
$$

for μ and σ^2, we obtain

$$
\mu = 2\log(m_1) - \frac{1}{2}\log(m_2) \qquad \text{and} \qquad \sigma^2 = \log(m_2) - 2\log(m_1).
$$

Let $\hat{m}_1 := n^{-1}\sum_{i=1}^n$ and $\hat{m}_2 = n^{-1}\sum_{i=1}^n X_i^2$ be the empirical moments, which are consistent estimators of the respective population moments by the weak law of large numbers and the continuous mapping theorem for convergence in probability. The method of moments estimators for μ and σ^2, denoted as $\hat{\mu}$ and $\hat{\sigma}^2$, are given by

$$
\hat{\mu} = 2\log(\hat{m}_1) - \frac{1}{2}\log(\hat{m}_2) \qquad \text{and} \qquad \hat{\sigma}^2 = \log(\hat{m}_2) - 2\log(\hat{m}_1). \qquad \blacksquare
$$

(b) (i) Since the sample mean in an unbiased estimator for the population mean, we have $E[\overline{X}_n] = E[X_1] = \theta/2$. Using linearity of expectation, we deduce that $E[2\overline{X}_n] = 2E[\overline{X}_n] = \theta$, which shows that $2\overline{X}_n$ is an unbiased estimator for θ. Moreover, as $\overline{X}_n \xrightarrow{P} E[X_1] = \theta/2$, that $2\overline{X}_n$ is consistent for θ follows by the continuous mapping theorem for convergence in probability (noting that $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) := 2x$, is a continuous function). Finally, we have

$$
\operatorname{Var}(2\overline{X}_n) = 4\operatorname{Var}(\overline{X}_n) = \frac{4}{n}\operatorname{Var}(X_1) = \frac{4}{n} \cdot \frac{\theta^2}{12} = \frac{\theta^2}{3n}.
$$

(ii) First, note that a pdf of X_1 is $f(x \mid \theta) = \frac{1}{\theta}$ for $0 \leq x \leq \theta$, and $f(x \mid \theta) = 0$ otherwise. The likelihood function is

$$\begin{aligned} L(\theta \mid (X_1, \ldots, X_n)) &= \prod_{i=1}^{n} f(X_i \mid \theta) \\ &= \prod_{i=1}^{n} \frac{1}{\theta} \mathbb{1}(X_i \in [0, \theta]) \\ &= \frac{1}{\theta^n} \mathbb{1}(\max\{X_1, \ldots, X_n\} \leq \theta). \end{aligned}$$

The previous derivation shows that

$$L(\theta \mid (X_1, \ldots, X_n)) = 0 \quad \text{if} \quad \theta < \max\{X_1, \ldots, X_n\}$$

and

$$L(\theta \mid (X_1, \ldots, X_n)) = \frac{1}{\theta^n} \quad \text{if} \quad \theta \geq \max\{X_1, \ldots, X_n\}.$$

Therefore, as $\theta > 0$ and $\frac{1}{\theta^n}$ is a decreasing function of θ, we have $\hat{\theta}_{MLE} = \max\{X_1, \ldots, X_n\}$.[2]

(iii) We have

$$\hat{\theta}_{MLE} = \max\{X_1, \ldots, X_n\} \leq x \iff \big\langle X_1 \leq x \wedge X_2 \leq x \wedge \cdots \wedge X_n \leq x \big\rangle.$$

Since $X_1, \ldots, X_n \overset{i.i.d.}{\sim} U([0, \theta])$, it follows that a cdf of $\hat{\theta}_{MLE}$ is

$$F_{\hat{\theta}_{MLE}}(x) = P\big(\hat{\theta}_{MLE} \leq x\big) = \begin{cases} 0 & \text{if } x < 0 \\ (x/\theta)^n & \text{if } 0 \leq x \leq \theta \\ 1 & \text{if } x > \theta \end{cases}.$$

Differentiating the cdf with respect to x, we obtain the pdf of $\hat{\theta}_{MLE}$:

$$f_{\hat{\theta}_{MLE}}(x) = \begin{cases} n x^{n-1}/\theta^n & \text{if } 0 \leq x \leq \theta \\ 0 & \text{otherwise} \end{cases}.$$

Therefore,

$$E\big[\hat{\theta}_{MLE}\big] = \int_0^\theta x \frac{n x^{n-1}}{\theta^n} \mathrm{d}x = \frac{n}{n+1}\theta.$$

Analogously, $E\big[\hat{\theta}_{MLE}^2\big] = (n/(n+2))\theta^2$, which gives

$$\operatorname{Var}\big(E\big[\hat{\theta}_{MLE}\big]\big) = \frac{n\theta^2}{(n+1)^2(n+2)}. \qquad \blacksquare$$

Exercise 54

Let $(X_n)_n$ and $(Z_n)_n$ be two sequences of real-valued random variables on a common probability space $(\Omega, \mathcal{A}, P)$, and let c be a strictly positive real constant. Prove the following statements.

(i) If $Z_n \xrightarrow{P} +\infty$ and $X_n \xrightarrow{P} c$, then $X_n Z_n \xrightarrow{P} +\infty$.

[2] *Quiz:* Suppose that $X_i \sim U([0, \theta))$, i.e., X_i is uniformly distributed over the interval $[0, \theta)$. Does the maximum likelihood estimator for θ exist? Discuss.

(ii) If $Z_n \xrightarrow{P} +\infty$ and $X_n = O_P(1)$, then $X_n + Z_n \xrightarrow{P} +\infty$.

Solution

(i) Fix $C > 0$ and $\eta > 0$. We want to show that there exists $N_{C,\eta} > 0$ such that[3]

$$P\big(X_n Z_n > C\big) > 1 - \eta$$

for all positive integers $n > N_{C,\eta}$. Fix $\varepsilon > 0$. Since $X_n \xrightarrow{P} c$, there exists $N_{\varepsilon,\eta} > 0$ such that

$$P(X_n \leq c - \varepsilon) < \frac{\eta}{2} \tag{3}$$

for all positive integers $n > N_{\varepsilon,\eta}$. As $Z_n \xrightarrow{P} +\infty$, there exists $N_{C/(c-\varepsilon),\eta} > 0$ such that

$$P(Z_n \leq C/(c - \varepsilon)) < \frac{\eta}{2} \tag{4}$$

for all positive integers $n > N_{C/(c-\varepsilon),\eta}$. Moreover, note that

$$\left\langle X_n > c - \varepsilon \ \wedge \ Z_n > \frac{C}{c - \varepsilon} \right\rangle \Longrightarrow X_n Z_n > C. \tag{5}$$

Then, for all positive integers $n > N_{C,\eta} := \max\big\{N_{\varepsilon,\eta}, N_{C/(c-\varepsilon),\eta}\big\}$

$$\begin{aligned}
P\big(X_n Z_n > C\big) &\geq P\left(X_n > c - \varepsilon, Z_n > \frac{C}{c - \varepsilon}\right) \\
&\geq 1 - P\Big(X_n \leq c - \varepsilon\Big) - P\left(Z_n \leq \frac{C}{c - \varepsilon}\right) \\
&> 1 - \frac{\eta}{2} - \frac{\eta}{2} \\
&> 1 - \eta,
\end{aligned}$$

where: The first inequality holds by (5); the second inequality holds by De Morgan's laws, Theorem 1.3.(iii) in the Lecture Notes, and sub-σ-additivity of P; the third inequality holds by (3) and (4). The desired result follows. ■

(ii) Fix $C > 0$ and $\eta > 0$. We want to show that there exists $N_{C,\eta} > 0$ such that

$$P\big(X_n + Z_n > C\big) > 1 - \eta$$

for all natural numbers $n > N_{C,\eta}$. Since $X_n = O_P(1)$, there exist real numbers C_η and $N_\eta > 0$ such that

$$P(X_n < -C_\eta) < \frac{\eta}{2} \tag{6}$$

for all positive integers $n > N_\eta$. As $Z_n \xrightarrow{P} +\infty$, there exists $N_{C-C_\eta,\eta} > 0$ such that

$$P(Z_n \leq C - C_\eta) < \frac{\eta}{2} \tag{7}$$

[3] Since probability measures are bounded from above by 1, a limit exists if and only if "$\liminf = \limsup$," and "$\liminf \leq \limsup$" always holds true, we have

$$\liminf_{n\to\infty} P(X_x Z_n > C) = 1 \Longleftrightarrow \lim_{n\to\infty} P(X_n Z_n > C) = 1.$$

for all positive integers $n > N_{C-C_\eta,\eta}$. Moreover, note that

$$\left(X_n \geq -C_\eta \ \wedge \ Z_n > C - C_\eta\right) \Longrightarrow X_n + Z_n > C. \tag{8}$$

Then, for all natural numbers $n > N_{C,\eta} := \max\left\{N_\eta, N_{C-C_\eta,\eta}\right\} > 0$

$$\begin{aligned} P\left(X_n + Z_n > C\right) &\geq P\left(X_n \geq -C_\eta, Z_n > C - C_\eta\right) \\ &\geq 1 - P\left(X_n < -C_\eta\right) - P\left(Z_n \leq C - C_\eta\right) \\ &> 1 - \frac{\eta}{2} - \frac{\eta}{2} \\ &> 1 - \eta, \end{aligned}$$

where: The first inequality holds by (8); the second inequality holds by De Morgan's laws, Theorem 1.3.(iii) in the Lecture Notes, and sub-σ-additivity of P; the third inequality holds by (6) and (7). The desired result follows. ■

Exercise 55

Solve the following problems.

(a) A sample of size 1 is taken from a population distribution $Po(\lambda)$, where $\lambda > 0$. To test $H_0 : \lambda = 1$ against $H_1 : \lambda = 2$, consider the non-randomized test $\varphi(X) = 1$ if $X > 3$, and $\varphi(X) = 0$ if $X \leq 3$. Find the probabilities of type I and type II errors and the power of the test against $\lambda = 2$. If it is required to achieve a size equal to 0.05, how should one modify the test φ?

(b) A traditional medicament attains an effect in 50% of all cases. We examine the effect of a new medicament in a study with $n = 20$ test persons. The result is that in 15 cases the new medicament attains a positive effect. Consider the hypotheses H_0 : The new medicament is equally effective as the traditional one, against H_1 : The new medicament is more effective.

  (i) Construct a level-$\alpha = 0.05$ test. Would you reject the null hypothesis given the 15 positive effects?

  (ii) Determine (an expression for) the probability of type II error if the new medicament has a rate of 60% of positive effects.

  [*Hint.* For the cdf F of a $Bin(20, 1/2)$-distributed random variable, we have $F(14) \approx 0,979$, $F(13) \approx 0,942$.]

(c) Let $\overline{X}_n$ be the sample mean of an i.i.d. sample of size n from $N(\mu, 16)$. Find the smallest sample size n such that $(\overline{X}_n - 1, \overline{X}_n + 1)$ is a 0.90-confidence interval for μ.

(d) Let $X_1, X_2 \overset{i.i.d.}{\sim} U(\theta, \theta + 1)$. For testing $H_0 : \theta = 0$ versus $H_1 : \theta > 0$, we have two competing tests: $\varphi_1(X_1)$ rejects H_0 if $X_1 > 0.95$ (and does not otherwise); $\varphi_2(X_1, X_2)$ rejects H_0 if $X_1 + X_2 > C$ (and does not otherwise) for some real number $C \geq 1$.

  (i) Find the value of C so that φ_2 has the same size as φ_1.

  (ii) Calculate the power function of each test. Draw a well-labeled graph of each power function.

  (iii) Prove or disprove: φ_2 is a more powerful test that φ_1.

  (iv) Show how to get a test that has the same size but is more powerful than φ_2.

Solution

(a) Let $X \sim Po(\lambda)$ be our random sample. A type I error occurs when H_0 is true, but rejected. The probability of type I error (size) for test φ is

$$\begin{aligned} P_{\lambda=1}\big(\varphi \text{ rejects } H_0\big) &= P_{\lambda=1}(\varphi(X) = 1) \\ &= P_{\lambda=1}(X > 3) \\ &= 1 - P_{\lambda=1}(X \leq 3) \\ &= 1 - \sum_{k=0}^{3} \frac{1^k e^{-1}}{k!} \\ &= 1 - \frac{1}{e}\left(1 + \frac{1}{1} + \frac{1}{2} + \frac{1}{6}\right) \\ &\approx 0.019. \end{aligned}$$

A type II error occurs when H_1 is true, but rejected. The probability of type II error for test φ is

$$\begin{aligned} P_{\lambda=2}\big(\varphi \text{ rejects } H_1\big) &= P_{\lambda=2}(\varphi(X) = 0) \\ &= P_{\lambda=2}(X \leq 3) \\ &= \sum_{k=0}^{3} \frac{2^k e^{-2}}{k!} \\ &= \frac{1}{e^2}\left(1 + \frac{2}{1} + \frac{4}{2} + \frac{8}{6}\right) \\ &\approx 0.857. \end{aligned}$$

The power is the probability of rejecting H_0 when H_1 is true. For test φ, the power is

$$P_{\lambda=2}\big(\varphi \text{ rejects } H_0\big) = 1 - P_{\lambda=2}\big(\varphi \text{ rejects } H_1\big) \approx 1 - 0.857 = 0.143.$$

The size of a non-randomized test $\tilde{\varphi}$ such that $\tilde{\varphi}(X) = 1$ if $X > 2$, and $\varphi(X) = 0$ if $X \leq 2$ is approximately equal to 0.08, which is larger that 0.05. To achieve a size equal to 0.05, we might use a randomized test that always rejects H_0 if $X > 3$, and rejects H_0 with probability γ if $X = 3$, where γ solves

$$0.05 = \gamma P_{\lambda=1}(X = 3) + P_{\lambda=1}(X > 3).$$

The previous equation gives $\gamma \approx 0.51$. ■

(b) A binomial model is suitable. Denote $X_1, \ldots, X_n$ random variables which are i.i.d. with $X_i \sim Bin(1, p)$. The test statistic T is the number of positive effects in a sample of size $n = 20$. That is, $T := \sum_{i=1}^{20} X_i$, with $T \sim Bin(20, p)$. We test $H_0 : p = 1/2$ against $H_1 : p > 1/2$.

(i) We determine the critical value c for the non-randomized test

$$\varphi(X_1, \ldots, X_{20}) = \begin{cases} 0 & \text{if } T \leq c \\ 1 & \text{if } T > c \end{cases}.$$

For $\alpha = 0.05$, we find c as

$$\begin{aligned} c &:= \min\big\{k \in \mathbb{N} \cup \{0\} : P_{p=0.5}(T > k) \leq 0.05\big\} \\ &= \min\left\{k \in \mathbb{N} \cup \{0\} : \sum_{i=k}^{20} \binom{20}{i} (1/2)^{20} \leq 0.05\right\}, \end{aligned}$$

and deduce that $c = 14$ from the hint. For this sample, $T = 15 > 14 = c$, and so we reject the null hypothesis.

(ii) A type II error occurs when H_1 is true, but rejected. The probability of type II error for test φ is

$$\begin{aligned} P_{p=0.6}\big(\varphi \text{ rejects } H_1\big) &= 1 - P_{p=0.6}\big(\varphi \text{ does not reject } H_1\big) \\ &= 1 - E_{p=0.6}[\varphi(X_1, \dots, X_{20})] \\ &= 1 - P_{p=0.6}(\varphi(X_1, \dots, X_{20}) = 1) \\ &= 1 - P_{p=0.6}(T > 14). \qquad \blacksquare \end{aligned}$$

(c) We want to determine the smallest sample size n such that

$$P\big(\overline{X}_n - 1 < \mu < \overline{X}_n + 1\big) \geq 0.9.$$

As $\overline{X}_n \sim N(\mu, \sigma^2/n)$, we have $\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$. Let $z_{1-\alpha/2}$ denote the $1-\alpha/2$-quantile of $Z \sim N(0,1)$. Then,

$$P\left(-z_{1-\alpha/2} < \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} < z_{1-\alpha/2}\right) = 1 - \alpha,$$

from which we obtain

$$P\left(\overline{X}_n - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X}_n + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha.$$

For $\alpha = 0.1$, we have $z_{1-\alpha/2} \approx 1.645$. Then, as $\sigma = 4$, to determine n we solve

$$\overline{X}_n + 1.645\frac{4}{\sqrt{n}} = \overline{X}_n + 1,$$

for n or, equivalently,

$$1.645 \cdot 4 = \sqrt{n},$$

which gives $n = 43$. $\blacksquare$

(d) The density function of $Y := X_1 + X_2$ is given by[4]

$$f_Y(y;\theta) := \begin{cases} y - 2\theta & \text{if } 2\theta \leq y \leq 2\theta + 1 \\ 2\theta + 2 - y & \text{if } 2\theta + 1 < y \leq 2\theta + 2 \\ 0 & \text{otherwise} \end{cases}.$$

(i) A type I error occurs when H_0 is true, but rejected. The size of a test is the associated probability of type I error. The size of φ_1 is

$$P_{\theta=0}\big(\varphi_1 \text{ rejects } H_0\big) = P_{\theta=0}\big(\varphi_1(X_1) = 1\big) = P_{\theta=0}(X_1 > 0.95) = 1 - 0.95 = 0.05.$$

The size of φ_2 is

$$P_{\theta=0}\big(\varphi_2 \text{ rejects } H_0\big) = P_{\theta=0}(\varphi_2(X_1, X_2) = 1) = P_{\theta=0}(X_1 + X_2 > C).$$

[4]Let X_1 and X_2 be two independent real valued random variables with density functions f_{X_1} and f_{X_2}. Then, the sum $Y := X_1 + X_2$ is a random variable with density function f_Y, where f_Y is the convolution of f_{X_1} and f_{X_2}, i.e.,

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X_1}(y - x_2) f_{X_2}(x_2)\mathrm{d}x_2 = \int_{-\infty}^{+\infty} f_{X_2}(y - x_1) f_{X_1}(x_1)\mathrm{d}x_1.$$

For $1 \leq C \leq 2$, we have

$$
\begin{aligned}
P_{\theta=0}(X_1 + X_2 > C) &= P_{\theta=0}(Y > C) \\
&= \int_C^2 f_Y(y; \theta = 0)\mathrm{d}y \\
&= \int_C^2 (2 - y)\mathrm{d}y \\
&= \frac{(2 - C)^2}{2}.
\end{aligned}
$$

Now, solving

$$\frac{(2 - C)^2}{2} = 0.05$$

for C we obtain $C = 2 - \sqrt{0.1} \approx 1.68$.

(ii) The power function of a test φ is the map $\beta_\varphi \colon \Theta \to [0, 1]$ defined as $\beta_\varphi(\theta) := E_\theta[\varphi(X)]$. The power function for test φ_1 is

$$
\begin{aligned}
\beta_1(\theta) &:= E_\theta[\varphi_1(X_1)] \\
&= P_\theta(\varphi_1(X_1) = 1) \\
&= P_\theta(X_1 > 0.95) \\
&= \begin{cases} 0 & \text{if } \theta \leq -0.05 \\ \theta + 0.05 & \text{if } -0.05 < \theta \leq 0.95 \\ 1 & \text{if } 0.95 < \theta \end{cases} .
\end{aligned}
$$

The power function for test φ_2 is

$$
\begin{aligned}
\beta_2(\theta) &:= E_\theta[\varphi_2(X_1, X_2)] \\
&= P_\theta(\varphi_2(X_1, X_2) = 1) \\
&= P_\theta(X_1 + X_2 > C) \\
&= P_\theta(Y > C) \\
&= \begin{cases} 0 & \text{if } \theta \leq C/2 - 1 \\ (2\theta + 2 - C)^2/2 & \text{if } C/2 - 1 < \theta \leq (C - 1)/2 \\ 1 - (C - 2\theta)^2/2 & \text{if } (C - 1)/2 < \theta \leq C/2 \\ 1 & \text{if } C/2 < \theta \end{cases} .
\end{aligned}
$$

(iii) From the graph (draw it!) it is clear that φ_1 is more powerful for θ near 0, but φ_2 is more powerful for larger θ's. Thus, φ_2 is not uniformly more powerful than φ_1.

(iv) If either $X_1 \geq 1$ or $X_2 \geq 1$ (or both), we should reject H_0 because $P_{\theta=0}(X_i < 1) = 1$ for $i = 1, 2$. Now, consider a new test, φ_{new}, with the rejection region given by

$$\left\{(x_1, x_2) \in \mathbb{R}^2 : x_1 + x_2 > C\right\} \cup \left\{(x_1, x_2) \in \mathbb{R}^2 : x_1 > 1\right\} \cup \left\{(x_1, x_2) \in \mathbb{R}^2 : x_2 > 1\right\}$$

The first set is the rejection region for φ_2. The new test, φ_{new}, has the same size as φ_2 because the last two sets both have probability 0 if $\theta = 0$. But for $0 < \theta < C - 1$, the power function of φ_{new} is strictly larger than $\beta_2(\theta)$. Indeed, for $0 < \theta < C - 1$, φ_2 does not reject H_0 ($\varphi_2(X_1, X_2) = 0$) with positive probability despite either $X_1 > 1$ or $X_2 > 1$, while the new test rejects H_0 (i.e., $\varphi_{\text{new}}(X_1, X_2) = 1$) for the same realization of X_1 and X_2. That is, $P_\theta(\varphi_{\text{new}}(X_1, X_2) - \varphi_2(X_1, X_2) = 1) > 0$. If $C - 1 \leq \theta$, this test and φ_2 have the same power as, for those values of θ, $P_\theta(\varphi_{\text{new}}(X_1, X_2) - \varphi_2(X_1, X_2) = 1) = 0$. ■