Advanced Probability Solutions
Autumn 2021, EPFL

John Ludovic Pirl

February 10, 2022

Abstract

Dear Students, I shall publish possible solutions to the exercises in this file. This is an "ongoing project", where solutions will be updated, complemented & corrected along the semester. You're encouraged to work with the electronic version of this file.

For the present course, and beyond the lecture notes, I have selected two of the "classics". Durret's "Probability, Theory and Examples" [1] is more famous than Jean-François le Gall's lecture notes at Ecole Normale "Intégration, Probabilités et Modèles Aléatoires" [4]. Both are excellent, but I would personally favor the latter, which is more compact and gives the whole spectrum of essential tools in an integrated, systematic manner.

As an advice for your future study of stochastic analysis: use Le Gall's (this author rocks!) book [3], which is self-contained and uncompromising. It is actually not easy to find a book that strikes the "right balance" between length and precision. The book by Karatzas & Shreve [2] is excellent too, but one easily gets lost in measure-theoretic considerations. Another advanced reference you may wish to use in your later study of Brownian motion and stochastic analysis is Revuz & Yor's "Continuous Martingales and Brownian Motion" [5]. Except for reading the literature, do the usual thing: type your questions on a search engine, check the Stack Exchange, use Wikipedia and other blogs.

Contents

1 Solutions - Sheet 1

1.1 Ex 1.1

For part (a), one could take fore example $\mathcal{F}_1 = \{\emptyset, \{1,2\}, \{3,4\}\}$ and $\mathcal{F}_2 = \{\emptyset, \{1,3\}, \{2,4\}\}$.

In part (b), taking $\mathcal{A} = \mathcal{F}_1 \cup \mathcal{F}_2$, we have that $\sigma(\mathcal{A}) = \mathcal{P}(\Omega)$. By setting (for example)

$$\mathbb{P}_1(\{1\}) = \mathbb{P}_1(\{2\}) = \mathbb{P}_1(\{3\}) = \mathbb{P}_1(\{4\}) = 1/4$$

and

$$\begin{aligned} \mathbb{P}_2(\{1\}) = \mathbb{P}_2(\{4\}) = 1/8 \\ \mathbb{P}_2(\{2\}) = \mathbb{P}_2(\{3\}) = 3/8, \end{aligned}$$

we have that $\mathbb{P}_1 \neq \mathbb{P}_2$.

Note however that if $\mathcal{A}$ were a π-system, these probabilities would then be equal.

1.2 Ex 1.2

- $B = A \sqcup B \backslash A$,the symbol $\sqcup$ denoting a **disjoint** union . Then $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B\backslash A) \geq \mathbb{P}(A)$.
- $\bigcup_{n=1}^{\infty} A_n = \bigsqcup_{n=1}^{\infty} (A_n \backslash \bigcup_{k=1}^{n-1} A_k)$. This immediately yields
$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}\left(A_n \backslash \bigcup_{k=1}^{n-1} A_k\right) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_n)$$
(term by term inclusion for the inequality).
- Let $C_n := A_n \backslash A_{n-1}$. $A_0 = \emptyset$ by convention. Then $\cup_n A_n = \cup_n C_n$ and the C_n are disjoint. Whence
$$\begin{aligned} \mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} C_n\right) &= \sum_{n \in \mathbb{N}} \mathbb{P}(C_n) \\ &= \lim_{N \to \infty} \uparrow \sum_{n=0}^{N} \mathbb{P}(C_n) \\ &= \lim_{N \to \infty} \uparrow \mathbb{P}(A_n) \end{aligned}$$
- Let $B_n = A_1 \backslash A_n$ for $n \geq 1$. Then $B_1 \subset B_2 \subset \dots B_n \subset B_{n+1} \subset \dots$. We have $A_1 = \bigcup_{n=1}^{\infty} B_n \sqcup \bigcap_{n=1}^{\infty} A_n$. As a consequence
$$\begin{aligned} \mathbb{P}(A_1) &= \mathbb{P}(\bigcup_{n=1}^{\infty} B_n) + \mathbb{P}(\bigcap_{n=1}^{\infty} A_n) \\ &= \lim_{n \to \infty} \uparrow \mathbb{P}(B_n) + \mathbb{P}(\bigcap_{n=1}^{\infty} A_n). \end{aligned}$$
So
$$\mathbb{P}(A_1) = \lim_{n \to \infty} \uparrow (\mathbb{P}(A_1) - \mathbb{P}(A_n)) + \mathbb{P}(\bigcap_{n=1}^{\infty} A_n),$$
which implies
$$\lim_{n \to \infty} \downarrow \mathbb{P}(A_n) = \mathbb{P}(\bigcap_{n=1}^{\infty} A_n)$$

1.3 Ex 1.3

$\mathbb{P}$ being a probability on $\mathcal{F}_0$, we have that $\mathbb{P}(\emptyset) = 0$ and $\mathbb{P}(\Omega) = 1$. So $\emptyset, \Omega \in \mathcal{F}_0$.

If $A \in \mathcal{F}_0$, we have $\mathbb{P}(A) \in \{0, 1\}$, so $\mathbb{P}(A^c) = 1 - \mathbb{P}(A) \in \{0, 1\}$. Hence $A^c \in \mathcal{F}_0$.

Let $(A_n)_{n\in\mathbb{N}}$ be a sequence of elements of $\mathcal{F}_0$. If $\mathbb{P}(A_n) = 0$ for all $n \in \mathbb{N}$, then

$$\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) \leq \sum_{n\in\mathbb{N}} \mathbb{P}(A_n) = 0,$$

implying $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{F}_0$. Similarly, if there is an n_0 such that $\mathbb{P}(A_{n_0}) = 1$, then

$$1 = \mathbb{P}(A_{n_0}) \leq \mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) \leq 1.$$

Whence $\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = 1$ and thus $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{F}_0$. $\square$

1.4 Ex 1.4

Several solutions have been suggested (see for instance 14). I'll explain here the one I think is most "natural", and which allows a mental representation of what is going on.

The idea is that of a "patchwork", where every piece of the patchwork (or rug, or cloth) lights up or remains turned off (so alternatively is labelled 0 or 1 if one prefers). Every configuration of 1's (or light-up pieces) would be one element of the sigma algebra.

One can start by grouping elements of the (finite) sigma-algebra by (distinct) pairs A_i, A_i^c where $i = 1, \ldots, m$. It is then immediate that $\mathcal{C} = \{\bigcap_{i=1}^m B_i \mid B_i \in \{A_i, A_i^c\}\} \subseteq \mathcal{F}$ is a disjoint covering of T. The empty set might appear several times, in which case you keep just one of them (which is of course the convention in set theory).

Now by construction, every element of the initial sigma algebra $\mathcal{G}$ can be expressed in a unique way as a disjoint union of elements of $\mathcal{C}$ (to prove mentally: elements of $\mathcal{C}$ are disjoint and cover T). An m-tuple of zeros and ones corresponds to one element (with the possibility of several tuples yielding the empty set). Uniqueness also follows immediately (try and explain why: the key is "disjoint and generating").

Note: Please be aware that my initial discussion of zeros and ones (for each piece of the patchwork) does not correspond to the m-tuple of zeros and ones discussed in the previous paragraph (in the first case, I am talking about elements of the partition, in the second, of a decision process of elements of the sigma algebra).

The converse part of the exercise should now be a triviality in light of the discussion above.

1.5 Ex 1.5

- Define $\mathcal{A}_1 = \{(a, b) \mid a, b \in \mathbb{Q}\}$. Immediately, $\sigma(\mathcal{A}_1) \subseteq B$. Since $\mathbf{Card}(\mathbb{N}\times\mathbb{N}) = \mathbf{Card}(\mathbb{N})$, $\mathcal{A}_1$ is countable. If $U \subseteq \mathbb{R}$ is open, then for all $x \in U$, there is $a, b \in \mathbb{Q}$ such that $x \in (a, b) \subseteq U$. So U can be written as a countable union of elements of $\mathcal{A}_1$. Whence $B \subseteq \sigma(\mathcal{A}_1)$. Thus $\sigma(\mathcal{A}_1) = B$.
- Define $\mathcal{A}_2 = \{[a, b] \mid a, b \in \mathbb{Q}$. Since $[a, b]^c$ is open, we immediately find that $\sigma(\mathcal{A}_2) \subseteq B$. Note again that $\mathcal{A}_2$ is countable. As above, for U open, $x \in U$, there is $a, b \in \mathbb{Q}$ such that $x \in [a, b] \subseteq U$. So U can be written as a countable union of elements of $\mathcal{A}_2$. Whence $B \subseteq \sigma(\mathcal{A}_2)$. We are done.

- $\mathcal{A}_3 := \{(-\infty, t) \mid t \in \mathbb{Q}\}$. Such a set $(-\infty, t)$ is open, so $\sigma(\mathcal{A}_3) \subseteq B$. Now for $t_1, t_2 \in \mathbb{Q}$, $t_1 < t_2$ the set $[t_1, t_2) = (-\infty, t_2) \backslash (-\infty, t_1)$, so $[t_1, t_2) \in \sigma(\mathcal{A}_3)$. Again, an open set $U \in \mathbb{R}$ can be covered by a countable union of sets $[t_1, t_2)$ with $t_1, t_2 \in \mathbb{Q}$. So $B \subseteq \sigma(\mathcal{A}_3)$ and thus $\sigma(\mathcal{A}_3) = B$.

- Same reasoning as in the list point, noting that $(\infty, t]$ is closed and operating a countable cover of the type $(t_1, t_2]$ with $t_1, t_2 \in \mathbb{Q}$.

1.6 Ex 1.6

Suppose that

$$\sum_{x \in [0,1]} p_x = 1, \quad \spadesuit$$

then for all $n \geq 1$ there exists a finite set $S_n \in [0, 1]$ such that

$$1 - \frac{1}{n} \leq \sum_{x \in [0,1]} p_x \leq 1.$$

Consider

$$x_0 \in \left(\bigcup_{n \geq 1} S_n \right)^c$$

with $p_x > 0$. There is an m such that $p_x > \frac{1}{m}$. Then

$$\sum_{x \in S_m \cup \{x_0\}} > 1 - \frac{1}{m} + \frac{1}{m} = 1,$$

contradicting $\spadesuit$. Note we are summing over a finite set. $\square$

2 Solutions - Sheet 2

2.1 Ex 2.1

This exercise is mostly a question of definition and playing around. Not the sexiest exercise either, but it helps you get your hands dirty.

We recall that $\mathcal{B}(\overline{\mathbb{R}}) = \sigma\left(\{[-\infty, a), a \in \mathbb{R}\} \cup \{(b, \infty], b \in \mathbb{R}\}\right)$

Claim 1. $\mathcal{B}(\mathbb{R}) \subseteq \mathbb{R} \cap \mathcal{B}(\overline{\mathbb{R}})$, the right-hand side denoting the trace sigma algebra of $\mathcal{B}(\overline{\mathbb{R}})$ on $\mathbb{R}$.

Proof. For $a, b \in \mathbb{R}$ with $a < b$ one has $(a, b) = [-\infty, b) \cap (a, \infty] \in \mathcal{B}(\overline{\mathbb{R}})$.

But $(a, b) \cap \mathbb{R} = (a, b) \in \mathcal{B}(\mathbb{R})$ and by exercise 1.5, we know all such sets generate $\mathcal{B}(\mathbb{R})$. So $\mathcal{B}(\mathbb{R}) \subseteq \mathcal{R} \cap \mathcal{B}(\overline{\mathbb{R}})$ since $\mathcal{B}(\overline{\mathbb{R}})$ is a sigma-algebra on $\mathbb{R}$ containing the generator $\{(a, b) \mid a < b; a, b \in \mathbb{R}\}$ of $\mathcal{B}(\mathbb{R})$. □

Claim 2. $\{-\infty\}$,$\{\infty\} \in \mathcal{B}(\overline{\mathbb{R}})$ (and thus $\mathbb{R} \in \mathcal{B}(\overline{\mathbb{R}})$).

Proof. $\{-\infty\} = \bigcap_{n\in\mathbb{N}}[-\infty, -n]$, $\{+\infty\} = \bigcap_{n\in\mathbb{N}}[n, +\infty]$ and $\mathbb{R} = \overline{\mathbb{R}} \backslash \{-\infty, +\infty\}$. □

Claim 3 $\mathcal{B}(\overline{\mathbb{R}}) = \mathcal{B}(\mathbb{R}) \sqcup (\mathcal{B}(\mathbb{R}) \cup \{-\infty\}) \sqcup (\mathcal{B}(\mathbb{R}) \cup \{+\infty\}) \sqcup (\mathcal{B}(\mathbb{R}) \cup \{-\infty, +\infty\})$.

Proof. Any set on the right-hand side is in $\mathcal{B}(\overline{\mathbb{R}})$ by our two previous claims and all these sets form a sigma algebra (verify! trivial). □

CONSEQUENCE: $\mathcal{B}(\overline{\mathbb{R}}) \cap \mathbb{R} \subseteq \mathcal{B}(\mathbb{R})$ and thus $\mathcal{B}(\overline{\mathbb{R}}) \cap \mathbb{R} = \mathcal{B}(\mathbb{R})$.

Now back to our exercise. $X : (\Omega, \mathcal{F}) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$.

" $\Rightarrow$ ":

- $\{-\infty\} \in \mathcal{B}(\overline{\mathbb{R}}) \Rightarrow X^{-1}(\{-\infty\}) \in \mathcal{F}$.
- $\{+\infty\} \in \mathcal{B}(\overline{\mathbb{R}}) \Rightarrow X^{-1}(\{+\infty\}) \in \mathcal{F}$.
- $A \in B(\mathbb{R}) \Rightarrow A \in \mathcal{B}(\overline{\mathbb{R}}) \Rightarrow X^{-1}(A) \in \mathcal{F}$

" $\Leftarrow$ ":

By our claim above, a set $A \in \mathcal{B}(\overline{\mathbb{R}})$ can be split into a set $\tilde{A} \in \mathcal{B}(\mathbb{R}) \subseteq \mathcal{B}(\overline{\mathbb{R}})$ with either, both or none of $\{+\infty\}$ or $\{-\infty\}$. Then $X^{-1}(A) = X^{-1}(\tilde{A}) \sqcup X^{-1}(\{-\infty\}) \sqcup X^{-1}(\{+\infty\}) \in \mathcal{F}$.

2.2 Ex 2.2

Note that $X^{-1}(\mathcal{G}) = \{X^{-1}(B) \mid B \in \mathcal{G}\}$ (set of all preimages of sets).

- $X^{-1}(\emptyset) = \emptyset$, so $\emptyset \in X^{-1}(\mathcal{G})$
- Let $A \in X^{-1}(\mathcal{G})$ then $A = X^{-1}(B)$ for some $B \in \mathcal{G}$. Then $A^c = (X^{-1}(B))^c = X^{-1}(B^c)$. $B^c \in \mathcal{G} \Rightarrow A^c \in X^{-1}(\mathcal{G})$.
- Suppose $A_n \in X^{-1}(\mathcal{G})$ for all $n \in \mathbb{N}$, then $A_n = X^{-1}(B_n)$ for some B_n for all n. Then $\bigcup A_n = \bigcup_{n\in\mathbb{N}} X^{-1}(B_n) = X^{-1}(\bigcup_{n\in\mathbb{N}} B_n)$. Since $\bigcup_{n\in\mathbb{N}} B_n \in \mathcal{G}$, we are done. □

2.3 Ex 2.3

- " $\Rightarrow$ " A projection $\pi_k : (X_1, \dots, X_n) \to X_k :: (x_1, \dots, x_n) \mapsto x_k$ is continuous, whence $\mathcal{B}(\mathbb{R}^n) - \mathcal{B}(\mathbb{R})$ measurable (see 2.2).
- " $\Leftarrow$ " Let $A_i \in \mathcal{B}(\mathbb{R}^n)$ for $i = 1, \dots, n$ and $(X_1, \dots, X_n) : (\Omega, \mathcal{F}) \to (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. Then $(X_1, \dots, X_n)^{-1}(A_1 \times \dots \times A_n) = X_1^{-1}(A_1) \cap \dots X_n^{-1}(A_n)$. This intersection is in $\mathcal{F}$ by hypothesis.
  Consequently $\mathcal{B}(\mathbb{R}^n) = \sigma(\{A_1 \times \dots \times A_n \mid A_1, \dots, A_n \in \mathcal{B}(\mathbb{R})\})$
  $\subseteq \{B \in \mathbb{R}^n \mid (X_1, \dots, X_n)^{-1}(B) \in \mathcal{F}\}$. So $(X_1, \dots, X_n)$ is $\mathcal{F} - \mathcal{B}(\mathbb{R}^n)$-measurable. □

2.4 Ex 2.4

(a) Consider $(\Omega, \mathcal{F}) \overset{X}{\to} (\mathbb{R}, \mathcal{B}(\mathbb{R})) \overset{f}{\to} (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, where X and f are measurable. Let $A \in \mathcal{B}(\mathbb{R})$. Then, f being measurable, we have $f^{-1} \in \mathcal{B}(\mathbb{R})$. X being measurable, $(f \circ X)^{-1}(A) = X^{-1}(f^{-1}(A)) \in \mathcal{F}$. So $f \circ X$ is measurable.

(b) By exercise 2.3, the function $(X, Y) : \Omega \to \mathbb{R}^n :: \omega \mapsto (X(\omega), Y(\omega))$ is measurable. We now use the following crucial fact, and prove it immediately hereafter: **a continuous function is measurable with respect to the sigma-algebras generated by the topologies at hand**.

Proof. Let $f : (U, \mathcal{U}) \to (V, \mathcal{V})$ be continuous, with $\mathcal{U}, \mathcal{V}$ topologies. In other words $f^{-1}(A) \in \sigma(U)$ for all $A \in \mathcal{V}$. Let $\Sigma = \{A \subseteq V \mid f^{-1}(A) \in \sigma(\mathcal{U})\}$. Σ is a sigma-algebra and by construction we have that $\mathcal{V} \subseteq \Sigma$, whence $\sigma(\mathcal{V}) \subseteq \Sigma$. So $f^{-1}(\sigma(\mathcal{V})) \subseteq \sigma(\mathcal{U})$. □

Now, we simply note that addition, subtraction and multiplication are continuous functions from $\mathbb{R}^2$ to $\mathbb{R}$. As a consequence, by composition and by point (a), the functions at hand are measurable (for instance: $+ \circ (X, Y)$).

(c) Division just requires no notice that $\backslash : \mathbb{R} \times \mathbb{R}\backslash\{0\} \to \mathbb{R} :: (x, y) \mapsto \frac{x}{y}$ is continuous too, whence measurable. As a consequence, by (a), $\frac{X}{Y} = \backslash \circ (X, Y)$ is a measurable function too.

2.5 Ex 2.5

Let $X_n : (\Omega, \mathcal{F}) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ be measurable functions, for all n. Let X denote the respective four functions studied. We immediately notice that if X_n converges point-wise surely (for an almost surely version, wait for a later discussion involving completeness of the underlying probability space), then it also converges in the lim inf or lim sup since, whence point-wise (surely) convergence implies measurability.

- The sets of the type $[-\infty, a)$ with $a \in \mathbb{R}$ generate the sigma algebra $\mathcal{B}(\overline{\mathbb{R}})$, so it's enough to prove that $X^{-1}([-\infty, a)) \in \mathcal{F}$ for all $a \in \mathbb{R}$. This however is clear because

  $$X^{-1}([-\infty, a)) = \{\omega \mid \inf_{n \in \mathbb{N}} X_n(\omega) < a\} = \bigcup_{n \in \mathbb{N}} \{\omega \mid X_n(\omega) < a\} \in \mathcal{F},$$

  since all elements on the right-hand side are in $\mathcal{F}$. So X is $\mathcal{F} - \mathcal{B}(\overline{\mathbb{R}})$ measurable.
- Exact same reasoning with (for example) sets of the type $(a, \infty]$ generating $\mathcal{B}(\overline{\mathbb{R}})$.
- Notice that

  $$X(\omega) = \liminf_{n \to \infty} X_n(\omega) = \sup_{n \geq 0} \left(\inf_{k \geq n} X_n(\omega) \right).$$

  By the two bullet points just discussed, the function on the r.h.s is measurable.
- Exact same reasoning with sup and inf reversed.

2.6 Ex 2.6

Grateful thanks to Wei Jiaye for reviewing and typing my solution.

This is a special case of a construction called **product measure** on a **product space** . We solve and calculate the easiest case, in which all sets Ω_i of $\Omega = \Omega_1 \times \cdots \times \Omega_n$ are finite and the σ-algebra is $\mathcal{P}(\Omega)$.
First, we show that $\mathbb{P}$ is a probability measure. Set $\mathbb{P}(\emptyset) = 0$.

- For disjoint sets A_n, we immediately have

$$\mathbb{P}(\coprod A_n) = \sum_{n \in \mathbb{N}} \mathbb{P}(A_n).$$

  To show that $\mathbb{P}(\Omega) = 1$, we calculate

$$\begin{aligned}\sum_{\omega_i \in \Omega_i, 1 \le i \le n} \mathbb{P}(\{(\omega_1, \ldots, \omega_n)\}) &= \sum_{\omega_i \in \Omega_i, 1 \le i \le n} \mathbb{P}_1(\{\omega_1\}) \cdots \mathbb{P}_n(\{\omega_n\}) \\ &= \left(\sum_{\omega_1 \in \Omega_1} \mathbb{P}_1(\{\omega_1\})\right)\left(\sum_{\omega_2 \in \Omega_2} \mathbb{P}_2(\{\omega_2\})\right) \cdots \left(\sum_{\omega_n \in \Omega_n} \mathbb{P}_n(\{\omega_n\})\right) \\ &= 1\end{aligned}$$

- Consider the projection

$$\pi_k : (\Omega, \mathcal{P}(\Omega), \mathbb{P}) \to \Omega_k, \text{where } \omega = (\omega_1, \ldots, \omega_n) \mapsto \omega_k,$$

  which has a **law** given by $\mathbb{P}_k$.

$$\mathbb{P}\left(\{\omega : \pi_k(\omega) = \omega_{k_0}\}\right) = \sum_{\omega_i \in \Omega_i, \omega_{k_0} \text{fixed}} \mathbb{P}_1(\{\omega_1\}) \cdots \mathbb{P}_k(\{\omega_{k_0}\}) \cdots \mathbb{P}_n(\{\omega_n\}) = \mathbb{P}_k(\{\omega_{k_0}\}.$$

  As a consequence, noting that $\tilde{X}_k = X_k \circ \pi_k$, for any $A \in \mathcal{B}(\mathbb{R})$ we have

$$\begin{aligned}\mathbb{P}(\tilde{X}_k(\omega) \in A) &= \mathbb{P}(X_k \circ \pi_k(\omega) \in A) \\ &= \mathbb{P}(\{\omega : \pi_k(\omega) \in X_k^{-1}(A)\}) \\ &= \mathbb{P}_k(X_k^{-1}(A)) \\ &= \mathbb{P}_k(X_k(\omega_k) \in A),\end{aligned}$$

  so $\tilde{X}_k$ has the same law as X_k.

- Finally we prove independence. If we can show that $\pi_1, \ldots, \pi_n$ are independent, then $X_1 \circ \pi_1, \ldots, X_n \circ \pi_n$ are independent too. We have

$$\begin{aligned}\mathbb{P}(\omega : \pi_1(\omega) \in A_1, \ldots, \pi_n(\omega) \in A_n) &= \mathbb{P}(A_1 \times \cdots \times A_n) \\ &= \mathbb{P}_1(A_1)\mathbb{P}_2(A_2) \cdots \mathbb{P}_n(A_n) \\ &= \mathbb{P}(\pi_1(\omega) \in A_1)\mathbb{P}(\pi_2(\omega) \in A_2) \cdots \mathbb{P}(\pi_n(\omega) \in A_n),\end{aligned}$$

  which completes the proof. □

2.7 Ex 2.7 - Challenge

See also section 14 for a more elegant solution of point (b) (in my opinion).

Solution 1. (a) Let $B_n = A_n$ if $\mathbb{P}[A_n] = 1$ and $B_n = A_n^c$ if $\mathbb{P}[A_n] = 0$. Define $E := \bigcap_n B_n$. Then if $A \supset E$ we have $\mathbb{P}[A] = 1$. Conversely consider $\mathcal{G}' := \{B \in \mathcal{G} : B \supset E \text{ or } B^c \supset E\}$. The set $\mathcal{G}'$ is a σ-algebra: Clearly $\varnothing \in \mathcal{G}'$ and if $B \in \mathcal{G}'$ then $B^c \in \mathcal{G}'$. Also if $(B_n)_n$ is a countable family of sets in $\mathcal{G}'$, then either one of B_n contains E, in which case so does $\bigcup_n B_n$. Otherwise $E \subset B_n^c$ for all n and thus $\Omega \setminus \bigcup_n B_n = \bigcap_n B_n^c \supset E$ so that $\bigcup_n B_n \in \mathcal{G}'$. Moreover $A_n \in \mathcal{G}'$ for all n so $\mathcal{G}' = \mathcal{G}$. In particular if $\mathbb{P}[A] = 1$ for some A, then $A \supset E$.

(b) For any Borel measurable $B \subset \mathbb{R}$ the event $Y \in B$ is also in $\sigma(X)$ and we have $\mathbb{P}[Y \in B] = \mathbb{P}[Y \in B]^2$, implying that $\mathbb{P}[Y \in B] \in \{0, 1\}$. Thus we may apply (a) with $\mathcal{G} = \sigma(Y)$. In particular there is some $E = \{Y \in E'\}$ with measurable $E' \subset \mathbb{R}$ such that $\mathbb{P}[Y \in A] = 1$ if and only if $A \supset E'$. But now if $x \in E'$, we must have $\mathbb{P}[Y = x] = 1$ (and $E = \{x\}$) since otherwise $\mathbb{P}[Y \neq x] = 1$ but $\mathbb{R} \setminus \{x\} \not\supset E'$.

(c) Let $\Omega = T = \mathbb{R}$ and $\mathcal{F} = \Sigma = \{A \subset \mathbb{R} : \text{either } A \text{ or } A^c \text{ is countable}\}$. Set $\mathbb{P}[A] = 0$ if A is countable and $\mathbb{P}[A] = 1$ otherwise. Clearly all events are pairwise independent. Moreover if we let $X(x) = Y(x) = x$ for all $x \in \Omega$ then Y is clearly X measurable but $\mathbb{P}[Y = x] = 0$ for all $x \in \mathbb{R}$.

3 Solutions - Sheet 3

3.1 Ex 3.1

- $\mathcal{G}$ is a λ-system (or "Dynkin"-system), so $\Omega \in \mathcal{G}$.
- $\mathcal{G}$ being a Dynkin-system, $A \in \mathcal{G} \Rightarrow A^c \in \mathcal{G}$.
- Let $A_1, A_2, \dots \in \mathcal{G}$, we now show that $\bigcup_{n\in\mathbb{N}} A_k \in \mathcal{G}$.

  Defining $B_k = A_k \backslash \bigcup_{j=1}^{k-1} A_j$, we immediately have that $\bigcup_{n\in\mathbb{N}} A_k = \bigsqcup_{n\in\mathbb{N}} B_k$.

  But $B_k = A_k \cap \left(\bigcap_{j=1}^{k-1} A_j^c\right) \in \mathcal{G}$, using that $\mathcal{G}$ is both a λ- and a $\pi-$system. Finally, $\mathcal{G}$ being a λ-system shows that $\bigsqcup_{n\in\mathbb{N}} B_k \in \mathcal{G}$. $\square$

3.2 Ex 3.2

$\mathcal{R} := \left\{\bigcup_{j=1}^{n} A_j \mid A_j \text{ 's are disjoint cylinder sets}, n \geq 1\right\}$

R is trivially non-empty, and we shall check that (1) $A \in \mathcal{R} \Rightarrow A^c \in \mathcal{R}$ and (2) $\mathcal{R}$ is closed by finite intersection.

1. If A is a cylindrical set, it can be written $A = X_1 \times X_2 \times \dots X_n \times \{0,1\}^{\mathbb{N}}$ for a certain n with $X_1, \dots, X_n \subseteq \{0,1\}$. Then by a dictionary-type logic, one has that

$$A^c = X_1^c \times \{0,1\}^{\mathbb{N}} \sqcup X_1 \times X_2^c \times \{0,1\}^{\mathbb{N}} \sqcup \dots X_1 \times X_{n-1} \times X_n^c \times \{0,1\}^{\mathbb{N}} \in \mathcal{R}.$$

2. Let A and B be two cylindrical sets with $A = X_1 \times X_2 \times \dots X_n \times \{0,1\}^{\mathbb{N}}$, $B = Y_1 \times Y_2 \times \dots Y_n \times \{0,1\}^{\mathbb{N}}$, with the X_i and Y_i subsets of $\{0,1\}$ as above. Then

$$A \cap B = (X_1 \cap Y_1) \times \dots \times (X_n \cap Y_n) \times \{0,1\}^{\mathbb{N}} \in \mathcal{R}.$$

   Note that if any of these intersections $X_1 \cap Y_i$ is the empty set, then $A \cap B$ is the empty set.

3. We check (1) for an element of $\mathcal{R}$. Consider to this end $A = A^1 \sqcup A^2 \dots \sqcup A^n$, where the A^k 's are disjoint cylinder sets. We have (by elementary set theory) that $A^c = (A^1)^c \cap \dots \cap (A^n)^c$ and by taking the complementary as above, one can write

$$(A^k)^c = C_1^k \sqcup \dots \sqcup C_{m_k}^k$$

   where the C_j^k are disjoint cylinder sets for fixed k and all $j = 1, \dots, m_k$. Going over to indicator notation we have that

$$1_{A^c} = 1_{(A^1)^c} \dots 1_{(A^n)^c} = (1_{C_1^1} + \dots + 1_{C_{m_1}^1}) \dots (1_{C_1^n} + \dots + 1_{C_{m_n}^n}) = \sum_{j_1=1}^{m_1} \dots \sum_{j_n=1}^{m_n} 1_{C_{j_1}^1} \dots 1_{C_{j_n}^n}$$

   $1_{C_{j_1}^1} \dots 1_{C_{j_n}^n}$ corresponds to the indicator of $C_{j_1}^1 \cap \dots \cap C_{j_n}^n$, which is a cylinder set (see above). All of them are disjoint. Thus $A^c \in \mathcal{R}$.

4. We check (2) for an element of $\mathcal{R}$. For this take $A = A^1 \sqcup \dots \sqcup A^n$ and $B = B^1 \sqcup \dots \sqcup B^m$ for disjoint cylinder sets $A^1, \dots, A^n$ and $B^1, \dots, B^m$. Then from expanding

$$1_{A\cap B} = (1_{A^1} + \dots + 1_{A^n})(1_{B^1} + \dots + 1_{B^m}),$$

   we can see that $A^i \cap B^j$ $(1 \leq i \leq n, 1 \leq j \leq m)$ are disjoint cylinder sets whose union is $A \cap B$ $\square$.

3.3 Ex 3.3

Grateful thanks to Wei Jiaye for reviewing and typing my solution.

First, since $\mathcal{R}$ is non-empty, there exists an $A \in \mathcal{R}$, and thus $A^c \in \mathcal{R}$. This implies that $T = A \sqcup A^c \in \mathcal{R}$ (hence $\emptyset = T^c \in \mathcal{R}$). Let $\mu : \mathbb{R} \to [0, \infty]$ be a countably additive map, i.e.,

$$\mu\left(\coprod_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} \mu(A_n), \text{ where } A_n \in \mathcal{R}\ \forall n \in \mathbb{N}, \text{ and } \coprod_{n\in\mathbb{N}} A_n \in \mathcal{R}$$

- For any $A \in \mathcal{P}(T)$, we have $A \subseteq T, T \in \mathcal{R}$, and $\mu(T) < \infty$, so $\mu^*(A) < \infty$.
- For any $A \subset B$ with $A, B \in \mathcal{P}(T)$, a cover of B is also a cover of A, so $\mu^*(A) \le \mu^*(B)$ is trivial.
- Let $A \in \mathcal{R}$, then $\mu^*(A) \le \mu(A)$ holds trivially. To show the other side, it suffices to show that for any $(A_n)_{n\in\mathbb{N}}$ such that $A \subseteq \bigcup_{n\in\mathbb{N}} A_n$, we have $\mu(A) \le \sum_{n\in\mathbb{N}} \mu(A_n)$. Indeed, let $B_n := A_n \cap A \in \mathcal{R}$, and $C_n := B_n \setminus \left(\bigcup_{j=1}^{n-1} B_j\right) \in \mathcal{R}$, so $(C_n)_{n\in\mathbb{N}}$ forms a disjoint cover of A. Since for any $n \in \mathbb{N}, C_n \subset A_n$, we have $\mu(C_n) \le \mu(A_n)$ (because $\mu(A_n \backslash C_n) \ge 0$, and using countable additivity again), the countable additivity of μ implies

$$\mu(A) = \sum_{n\in\mathbb{N}} \mu(C_n) \le \sum_{n\in\mathbb{N}} \mu(A_n).$$

- Fix $\varepsilon > 0$. For any countable family $(A_n)_{n\in\mathbb{N}}$ of subsets of T, there exists $(A_k^i)_{(i,k)\in\mathbb{N}\times\mathbb{N}}$ such that

$$A_k \subseteq \bigcup_{i\in\mathbb{N}} A_k^i, \quad \text{and} \quad \sum_{i\in\mathbb{N}} \mu(A_k^i) \le \mu^*(A_k) + \frac{\varepsilon}{2^k}, \forall\, k.$$

  Note that $\bigcup_{k\in\mathbb{N}} A_k \subseteq \bigcup_{(i,k)\in\mathbb{N}\times\mathbb{N}} A_k^i$, which implies

$$\mu^*\left(\bigcup_{k\in\mathbb{N}} A_k\right) \le \sum_{k\in\mathbb{N}} \sum_{i\in\mathbb{N}} \mu(A_k^i) \le \sum_{k\in\mathbb{N}} \left(\mu^*(A_k) + \frac{\varepsilon}{2^k}\right) = \sum_{k\in\mathbb{N}} \mu^*(A_k) + \varepsilon.$$

3.4 Ex 3.4

A pseudometric $\tilde{d}$ distinguishes itself from a metric $d : X \times X \to \mathbb{R}$ in that $d(x, y) = 0 \Rightarrow x = y$ whereas $\tilde{d}(x, y) = 0$ could happen for $x \neq y$ (so we don't have to prove the former condition).

Just to illustrate this behaviour, with the symmetric difference $A\Delta B = A\backslash B \sqcup B\backslash A$, we could imagine that $\mu^*(A\backslash B) = \mu^*(B\backslash A) = 0$ ($\Rightarrow \mu^*(A\Delta B) = 0$) with both sets being non-empty (and thus $A \neq B$). This happens all the time: think o the Lebesgue measure on $\mathbb{R}$ with the Borel sigma-algebra (implying $\mu^* = \mu$ on $\mathbb{B}(\mathbb{R})$). For example take $A = \mathbb{Q} \cup [-1, 1]$ and $B = \mathbb{Q} \cup [-1, 1]$. Then $A\Delta B = \{x \in \mathbb{Q} \mid |x| > 1\}$, $\mu(A\Delta B) = 0$.

- $d(A, A) = \mu^*(A\Delta A) = \mu^*(\emptyset) = 0$. The last equality comes from the fact we have assumed that $T \in \mathcal{R}$ (exercise 3), whence $\emptyset \in \mathcal{R}$. The last equality requires a bit of justification. by hypothesis $T \in \mathcal{R}$ whence $\emptyset \in \mathcal{R}$ too. By countable additivity, the equality follows.
- $A\Delta B = B\Delta A$ whence $d(A, B) = d(B, A)$.
- $A\backslash B \subseteq (A\backslash C) \cup (C\backslash B)$ and $B\backslash B \subseteq (B\backslash C) \cup (C\backslash A)$. Consequently, taking "unions on both sides" yields $A\Delta B \subseteq (A\Delta C) \cup (C\Delta B)$. So $\mu^*(-) \le \mu^*(-) + \mu^*(-) \Rightarrow d(A, B) \le d(A, C) + d(C, B)$. $\square$

3.5 Ex 3.5

See also chapter 14 for an original alternative solution.

Take $\Omega = (0,1)$ and the pi-system $\Pi = \{(0,t) \mid 0 < t < 1\}$. Define $\mu : \Pi \to [0,\infty] :: A \mapsto 1$.

μ is trivially seen to be countably additive (sigma additive), because no element of Π can be written as a union of more than one element of Π, namely itself. We also have that $\sigma(\Pi) = \mathcal{B}((0,1))$.

Should μ extend to a measure, we would have that

$$\mu(\emptyset) = \mu\left(\bigcap_{n\in\mathbb{N}} (0, \frac{1}{n})\right) \overset{\spadesuit}{=} \lim_{n\to\infty} \mu((0, \frac{1}{n})) = 1,$$

where the second last equality (continuity of measure) is true because one of the sets in the intersection has finite measure.

But then we have that

$$1 = \mu((0,1)) = \mu(\emptyset \sqcup (0,1)) = \mu(\emptyset) + \mu((0,1)) = 1 + 1 = 2,$$

a contradiction. $\square$

Note: Coming back to $\spadesuit$, here is a counter-example when the considered sets A_n satisfy $\mu(A_n) = \infty$ for all n and $A_{n+1} \subseteq A_n$. Take the Lebesgue measure on $\mathbb{R}$, then

$$\mu\left(\bigcap_{n\in\mathbb{N}} (n, \infty)\right) = \mu(\emptyset) = 0,$$

but $\mu((n,\infty)) = \infty$ for all n and of course $\lim_{n\to\infty} \mu((n,\infty)) = \infty$.

3.6 Ex 3.6 - Challenge

No solution provided. Your solutions are welcome!

3.7 Ex 3.7 - Challenge

No solution provided. Your solutions are welcome!

4 Solutions - Sheet 4

4.1 Ex 4.1

- $F_X(x) = \mathbb{P}(X \leq x)$. If $a \leq b$ then $\{X \leq a\} \subseteq \{X \leq b\}$. Taking probabilities on both sides reads $F_X(a) \leq F_X(b)$.
- Suppose à sequence $(a_n)_{n\in\mathbb{N}} \downarrow a$, then $\{X \leq a\} = \bigcap_{n\in\mathbb{N}}\{X \leq a_n\}$, so $F_X(a) = \mathbb{P}(-) = \lim_{n\to\infty} \mathbb{P}(X \leq a_n) = \lim_{n\to\mathbb{N}} F_X(a_n)$. $\square$

4.2 Ex 4.2

For part (a), we simply unfold the definitions and use independence:

$$\begin{aligned} F_Y(x) = \mathbb{P}(Y \leq x) &= \mathbb{P}(X_1 \leq x, \dots, X_n \leq x) \\ &= \mathbb{P}(\{X_1 \leq x\} \cap \dots \cap \{X_n \leq x\}) \\ &= \mathbb{P}(X_1 \leq x) \dots \mathbb{P}(X_n \leq x) \\ &= F_{X_1}(x) \dots F_{X_n}(x). \end{aligned}$$

For (b), we first see that $\mathbb{P}(\min(X_1, \dots, X_n) > a) = \mathbb{P}(X_1 > a, \dots, X_n > a) = \Pi_{k=1}^n \mathbb{P}(X_k > a)$ ♠, using independence.

- If $a > 0$ we have that ♠ $= \Pi_{k=1}^n \exp(-\lambda_k a) = \exp(-(\lambda_1 + \dots + \lambda_k)a)$. So $\mathbb{P}(\min(X_1, \dots, X_n) \leq a) = 1 - \exp(-(\lambda_1 + \dots + \lambda_k)a)$.
- If $a \leq 0$ then ♠ $= \Pi_{k=1}^n 1 = 1$, whence $\mathbb{P}(\min(X_1, \dots, X_n) \leq a) = 0$.

We thus conclude that $F_{\min(X_1,\dots,X_n)}(x)$ has the cumulative distribution function of an exponential of parameters $\lambda_1 + \dots + \lambda_k$.

Remark: the cumulative distribution function on $\mathbb{R}$ characterizes the law of the random variable. Indeed, it assigns probabilities to $(a, b] = F_X(b) - F_X(a)$, and we saw that $\sigma\{(a, b] \mid a < b\} = \mathcal{B}(\mathbb{R})$.

4.3 Ex 4.3

We have seen in the lecture that $\pi_i : \{0,1\}^{\mathbb{N}} \to \{0,1\} :: \omega = (\omega_1, \dots, \omega_n) \mapsto \omega_i$ is measurable, has uniform law on $\{0,1\}$ and that $\{\pi_i\}_{i\in\mathbb{N}}$ for a countable set of i.i.d random variables. We also saw that $X : \{0,1\}^{\mathbb{N}} \to [0,1] :: \omega = (\omega_1, \dots, \omega_n) \mapsto \sum_{n=1}^{\infty} \omega_n 2^{-n}$ has a uniform law on $[0,1]$♠. Finally, we have seen that for F a cumulative distribution function, defining $G(t) = \inf(\{x : F(x) \geq t\})$ and using $U \sim \text{Uniform}([0,1])$, the random variable $G(U)$ then has F as a c.d.f.

We basically have all the ingredients, except that we would need to have a countable copy of r.v. distributed as F. One way of proceeding would be to do a new countable product of measure spaces with a product measure (I have not thought about feasibility, but it's one approach). Another approach is to just work on the space we are given, namely $\{0,1\}^{\mathbb{N}}$. We partition (actually just countable disjoint subsets are needed) $\mathbb{N}$ into a countable disjoint union of sets. For example we could play with the primes:

- $I_2 = \{2^i \mid i \geq 1\}$
- $I_3 = \{3^i \mid i \geq 1\}$
- $I_5 = \{5^i \mid i \geq 1\}$ and so forth ...

For each of these sets, and since all of the π_i with $i \in I_k$ are i.i.d, we repeat the uniform image law construction above ♠ and compose with G. We are intuitively done, but still need the following to conclude:

Independent Grouping Theorem
Let $(\omega_i)_{i\in I}$ be independent sigma algebras indexed by **any** set I. Let $(I_j)_{j\in J}$ be a partition of I, i.e. $I = \sqcup_{j\in J} I_j$. Then $\{\sigma(\{\cup\sigma_i, i \in I_j\}) \mid j \in J\}$ form a set of independent sigma algebras indexed by J. □

4.4 Ex 4.4

We recall that

$$\limsup_{n\to\mathbb{N}} A_n = \bigcap_{n=1}^{\infty}\left(\bigcup_{k=n}^{\infty} A_k\right) = \{\omega \mid \exists \text{ an infinity of n with } \omega \in A_n\}$$

and

$$\liminf_{n\to\mathbb{N}} A_n = \bigcup_{n=1}^{\infty}\left(\bigcap_{k=n}^{\infty} A_k\right) = \{\omega \mid \exists\, n \text{ with } \omega \in A_k \text{ for all } k \geq n\}.$$

It's sufficient to show that $1 - \mathbb{P}(\limsup_{n\to\mathbb{N}} A_n) = 0$.

We calculate:

$$\begin{aligned}
1 - \mathbb{P}(\limsup_{n\to\infty} A_n) &= \mathbb{P}((\limsup_{n\to\infty} A_n)^c) \\
&= \mathbb{P}(\liminf_{n\to\infty} A_k^c) \\
&= \lim_{n\to\infty} \mathbb{P}(\cap_{k=n}^{\infty} A_k^c) \\
&= \lim_{n\to\infty} (\lim_{N\to\infty} \mathbb{P}(\cap_{k=n}^{N} A_k^c)) \\
&\overset{\spadesuit}{=} \lim_{n\to\infty} (\lim_{N\to\infty} \Pi_{k=n}^{N} \mathbb{P}(A_k^c)) \\
&= \lim_{n\to\infty} (\lim_{N\to\infty} \Pi_{k=n}^{\infty} 1 - \mathbb{P}(A_k)) \\
&\overset{\diamondsuit}{\leq} \lim_{n\to\infty} \Pi_{k=n}^{\infty} \exp(-\mathbb{P}(A_k)) \\
&= 0.
\end{aligned}$$

We have used independence for ♠, and the fact that (by convexity of the exponential) $1+x \leq \exp(x)$ for all $x \in \mathbb{R}$ in ♢ (and taking $x = -\mathbb{P}(A_k)$. The final equality just comes from the fact that for fixed k, the expression after the limit is always zero. □

4.5 Ex 4.5

We somehow want to use one of the Borel-Cantelli lemmas. The output resulting in a probability of 1 (the monkey will end up typing...) we have a strong hint at Borel-Cantelli 2.

Note: in probability, an assertion is considered almost surely.

So let's fabricate events that have positive probability, sum up to an infinite sum of probabilities (a strong hint that we can take events with the same probability, finding a pattern) and are independent.

Here is an example of solution (there are plenty...).

All piano notes have equal probability p of being hit (α keys, $\alpha p = 1$). $\mathbb{P}(\{X_n = k\}) = p$ for $1 \leq k \leq \alpha$ and X_n the random key struck at time n. We have that $(X_n)_{n\in\mathbb{N}}$ are i.i.d. The little prince has a finite length β of notes $(a_1, a_2, \dots, a_\beta)$ with $a_i \in \{1, \dots, d\}$ for all i. So $\mathbb{P}\left((X_n, X_{n+1}, \dots, X_{n+\alpha-1}) = (a_1, \dots, a_\beta)\right) = p^\beta > 0$.

Consider the independent events $A_k = \{(X_{k\beta}, \dots, X_{k\beta+(\beta-1)}) = (a_1, \dots, a_\beta)\}$ for $k \in \mathbb{N}$, with $\mathbb{P}(A_k) = p^\beta > 0$. Whence $\sum_{k\geq 0} \mathbb{P}(A_k)$
$= \infty$, so by Borel-Cantelli 2, we have that $\mathbb{P}(\limsup_{n\to\infty} A_n) = 1$. But we simply recall that

$$\limsup_{n\to\infty} A_n = \{\cap_{n=1}^\infty \cup_{k=n}^\infty A_n\} = \{\omega \mid \omega \in A_n \text{ for infinitely many n}\},$$

whence the partition is played (countably) infinitely many times on our slicing up of time into "intervals" of length β. $\square$

4.6 Ex 4.6

The underlying topic spanning the present exercise is the notion of a **complete probability space** (or measure space, for that matter). A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is complete if the relation $A \in \mathcal{F}, \mathbb{P}(A) = 0, B \subseteq A$ implies that $B \in \mathcal{F}$. Fundamentally, a random variable on a complete metric space that is modified on a (measurable) set of measure zero is still measurable. Here, we skillfully manage not to touch the topic, but you should know this is where you should look if you encounter such problems in the future.

First, suppose we have two random variables U and V. Then $\{U \neq V\}$ is measurable. Indeed,

$$\{U \neq V\} = \bigcup_{q\in\mathbb{Q}} \left((\{U < q\} \cap \{V > q\}) \cup (\{U > q\} \cap \{V < q\})\right) \spadesuit$$

is measurable!

For point (a), we now treat scalar multiplication (addition of a constant could be seen as a special case of addition of random variables) Let $X \sim Y$ (i.e. $X = Y$ in L^0, equivalently $\mathbb{P}(X = Y) = 1$. Then $\mathbb{P}(aX + b = aY + b) = \mathbb{P}(X = Y)$ for all a (even $a = 0$) and b. Whence $aX + b \sim aY + b$.

Now suppose $X \sim Y$ and $Z \sim W$. We want to show that $X + Z \sim Y + W$, showing that the equivalence class of the sum is well-defined. Indeed

$$\{X + Y \neq Y + W\} \subseteq \{X \neq Y\} \cup \{Z \neq W\},$$

so taking probabilities on each side, we get $\mathbb{P}(-) \leq \mathbb{P}(-) + \mathbb{P}(-) = 0$. $\square$

Now to point (b). Suppose $X_n \to X$ a.s. Then X is not necessarily measurable. However, we show there is an $\tilde{X}$ measurable such that $X_n \to \tilde{X}$ a.s.
Denote by $A := \{\omega \mid \lim_{n\to\infty} \text{ exists in } \mathbb{R}\cup\{-\infty, \infty\}\}$ is measurable. Indeed $A = \{\omega \mid \limsup X_n = \liminf X_n\}$. Indeed both random variables on the l.h.s and r.h.s are measurable (Serie 1) and A is then the complement of a set of the form $\spadesuit$.
Setting $\tilde{X} = (\lim X_n) \cdot \mathbb{1}_A + 0 \cdot \mathbb{1}_{A^c}$, we have a measurable function for which $X_n \to \tilde{X}$ almost surely. Of course, if the original function X is measurable, then $X \sim \tilde{X}$, as they will differ on a (measurable) set of measure zero. This is easily seen by noticing that these two random variables are equal except on the complement of the union of two sets of measure zero (unicity of the limit in L^0)

I am not so happy with the formulation of the last assertion, in the following sense: we really just want to see that this modified series X_n' converges almost surely to the **same** element of L^0, allowing us to now talk of convergence in the space of equivalence classes L^0. This however should be clear because the measurable set

$$B = \{\liminf X'_n = \limsup X'_n\} \cap \{\liminf X_n = \limsup X_n\}$$

has measure one. Indeed

$$\{\liminf X'_n = \limsup X'_n\} \supseteq \{\limsup X_n = \liminf X_n\} \setminus \left(\bigcup_{n\in\mathbb{N}} \{X_n \neq X'_n\} \right),$$

the right hand side having measure one (and symmetrically). So we're indeed talking about the same element of L^0.

4.7 Ex 4.7

- If $X \sim Y$ then $\mathbb{P}(\{|X-Y| > \epsilon\}) = 0$ for all ϵ, so $d_{KF}(X,Y) = 0$.
- Conversely, suppose that $d_{KF}(X,Y) = 0$ and $X \nsim Y$. Then

$$\{X \neq Y\} = \left(\bigcap_{n\in\mathbb{N}} \left\{ |X-Y| < \frac{1}{n} \right\} \right)^c = \bigcup_{n\in\mathbb{N}} \left\{ |X-Y| > \frac{1}{n} \right\},$$

so

$$0 < \mathbb{P}(\{X \neq Y\}) = \lim_{n\to\infty} \uparrow \mathbb{P}\left(\left\{ |X-Y| > \frac{1}{n} \right\} \right).$$

Then $\exists\, n_0, \epsilon$ such that for all $n \geq n_0 :\ \mathbb{P}\left(\left\{|X-Y| > \frac{1}{n}\right\}\right) > \epsilon > 0$. Taking $\frac{1}{n} < \epsilon$, we get for $0 < x \leq \frac{1}{n}$ that $\mathbb{P}\left(\{|X-Y| > x\}\right) > \epsilon > 0$. Whence $d_{KF}(X,Y) > 0$. A contradiction. Thus $X \nsim Y$.

- We will have to prove that $d(X,Y)$ does not depend on the representants X and Y in the equivalence classes of L^0. First we prove symmetry and the triangle inequality on random variables, then show the distance function is well-defined on equivalent classes.
- For X and Y random variables, $d_{KF}(X,Y) = d_{KF}(Y,X)$ is immediate.
- For X, Y, Z ($\mathbb{R}$-valued) random variables, we have $|X-Y| < |X-Z| + |Z-X|$ so

$$\begin{aligned} \mathbb{P}(\{|X-Y| > \epsilon_1 + \epsilon_2\}) &\leq \mathbb{P}(\{|X-Z| > \epsilon_1\} \cup \{|Y-Z| > \epsilon_2\}) \\ &\leq \mathbb{P}(\{|X-Z| > \epsilon_1\}) + \mathbb{P}(\{|Y-Z| > \epsilon_2\}). \end{aligned}$$

If the first element after the last inequality is $< \epsilon_1$ and the second $< \epsilon_2$, then $\mathbb{P}(\{|X-Y| > \epsilon_1 + \epsilon_2\}) < \epsilon_1 + \epsilon_2$. In other words, by taking infimums, we read $d_{KF}(X,Y) \leq d_{KF}(X,Z) \leq d_{KF}(Z,Y)$.

- We now have all elements in our hands to show that d_{KF} is well defined on L^0, thus concluding our proof that the Ky Fan metric is indeed a metric on L^0. Indeed, taking $X \sim \tilde{X}$ and $Y \sim \tilde{Y}$, we compute a classical triangular + inverse triangular argument.

$$\begin{aligned} |d(X,Y) - d(\tilde{X},\tilde{Y})| &\leq |d(X,Y) - d(X,\tilde{Y})| + |d(X,\tilde{Y}) - d(\tilde{X},\tilde{Y})| \\ &\leq d(Y,\tilde{Y}) + d(X,\tilde{X}) \\ &= 0. \end{aligned}$$

4.8 Ex 4.8 - Challenge

No solution provided. Your solutions are welcome!

5 Solutions - Sheet 5

5.1 Ex 5.1

We shall use the following:

Theorem. *Let X_n and X be random variables. Then $X_n \xrightarrow{\mathbb{P}} X$ if and only if for every subsequence X_{n_k} of X_n there is a further subsequence $X_{n_{k_l}}$ such that $X_{n_{k_l}} \xrightarrow{a.s.} X$.*

It will be used several times in the sequel.

- We consider $(X_n + Y_n)_{n\geq 0}$. Take any subsequence n_k. Then $\exists\, n_{k_l}$ such that $X_{n_{k_l}} \to X$ a.s. From this subsequence one can extract a further sequence $n_{k_{l_m}}$ such that $Y_{n_{k_{l_m}}} \to Y$ a.s. But then $X_{n_{k_{l_m}}} \to X$ a.s. as well. Hence $(X+Y)_{n_{k_{l_m}}} \to X+Y$ a.s. We conclude that $X_n + Y_n \xrightarrow{\mathbb{P}} X + Y$.
- The exact same argument gives us that $X_n Y_n \xrightarrow{\mathbb{P}} XY$.
- We also need $Y \neq 0$ a.s. We again use the same argument.
- The same argument is used again. Here, the continuity of g guarantees that the subsequences (and subsubsequences) of $(g(X_{n_k})_{k\geq 0}$ still converge almost surely to $g(X)$.

5.2 Ex 5.2

Consider $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1], \mathbb{B}([0,1]), \lambda)$ with λ the Lebesgue measure.

Define a sequence of indicator functions in the following way:

- $X_1 = \mathbb{1}_{[0,1]}$,
- $X_2 = \mathbb{1}_{[0,\frac{1}{2}]}$, $X_3 = \mathbb{1}_{[\frac{1}{2},1]}$,
- $X_4 = \mathbb{1}_{[0,\frac{1}{3}]}, X_5 = \mathbb{1}_{[\frac{1}{3},\frac{2}{3}]}, X_6 = \mathbb{1}_{[\frac{2}{3},1]}$,
- ...
- $X_{\frac{(n-1)n}{2}+1} = \mathbb{1}_{[0,\frac{1}{n}]}, \dots, X_{\frac{n(n+1)}{2}} = \mathbb{1}_{[1-\frac{1}{n},1]}$,
- ...

Each random variable Y on the n^{th} line satisfies $\mathbb{P}(|Y - 0| > \epsilon) = \frac{1}{n}$ for any $\epsilon < 1$. But then $X_n \xrightarrow{\mathbb{P}} 0_{[0,1]}$. We have $X_n \not\to 0_{[0,1]}$ almost surely. However (one know that every sequence converges in probability has a subsequence converging almost surely to the same random variable) we have, extracting the first random variable of each line that:

$$\left(X_{\frac{(k-1)k}{2}+1}\right)_{k\geq 1} \to 0_{[0,1]}, \text{ a.s.}$$

5.3 Ex 5.3

Notice that $Z \in \mathcal{S} \implies Z\mathbb{1}_A \in \mathcal{S}$ for any A in $\mathcal{F}$. $Z \in \mathcal{S} \implies |Z| \in \mathcal{S}$ and so $|Z|\mathbb{1}_{\{|Z|>\epsilon\}} \in \mathcal{S}$.

Now from

$$|Z| \geq |Z|\mathbb{1}_{\{|Z|>\epsilon\}} \geq \epsilon\mathbb{1}_{\{|Z|>\epsilon\}}$$

we deduce that $\mathbb{E}[|Z|] \geq \epsilon\mathbb{E}[\mathbb{1}_{\{|Z|>\epsilon\}}]$. This can be read as $\mathbb{P}(\{|Z| > \epsilon\}) \leq \frac{1}{\epsilon}\mathbb{E}[|Z|]$.

Note: the last relation is called **Markov's inequality** and is true for any $Z \in L^1$.

Let's suppose $\epsilon_n \downarrow d_{KF}(X,Y) := \inf\{\epsilon > 0 \mid \mathbb{P}(|X-Y| > \epsilon) \leq \epsilon\}$. Suppose $\sqrt{\mathbb{E}[|X-Y|] \leq \epsilon_0}$ for some ϵ_0. Then

$$\mathbb{P}(\{|X-Y| > \epsilon_0\}) \leq \frac{1}{\epsilon_0}\mathbb{E}[|X-Y|] = \frac{1}{\epsilon_0}\epsilon_0^2 \leq \epsilon_0,$$

whence $d_{KF}(X,Y) \leq \epsilon_0$.

We conclude by taking $\epsilon_n \downarrow \sqrt{\mathbb{E}[|X-Y|]}$.

5.4 Ex 5.4

$X = \sum_{k=1}^{n} \mathbb{1}_{\{k\}}$. But $\mathbb{P}(\{k\}) = \frac{1}{n}$ for all $k \in \{1, \cdots, n\}$ and since X is a simple random variable (a step function, if you prefer), we have by definition that $\mathbb{E}[X] = \sum_{k=1}^{n} X(k)$. $\square$

5.5 Ex 5.5

A few words before tackling this exercise. It certainly was very confusing to me, just to put you at ease. There are several concepts that might puzzle you.

First of all, you know the definition of $V = (L^1, ||\cdot||_1)$ from measure theory. You know that this vector space is a Banach space and that $\mathcal{S}$, the set of all step functions, is dense in V. You also know (obviously) that $L^1 \subsetneq L^0$ (at least for most probability spaces: think of Ω with finite cardinality as a counter-example).

You also know from functional analysis that two completions of a normed vector space are isomorphic.

Here, we are doing something different. You are now asked to forget about the measure-theoretic definition of L^1, and only remember $||L_1||$ as a norm on $\mathcal{S}$. The present exercise is meant to show that **IF** $\mathcal{S}$, can be completed in (L^0, d_{KF}) (a subset of which is then a **normed vector space**) by means of a continuous injection into the **metric** space (L^0, d_{KF}) , then this completion is unique. The course shows this can be done. Your knowledge from measure theory (the construction of the Lebesgue integral) shows that our construction coincides with what you have learnt in measure theory, because convergence in L^1 in the usual sense implies convergence in probability (and thus $\mathcal{S}$ is indeed continuously injected into L^0 endowed with the Ky-Fan metric.

Suppose now that $S \subseteq L^1, \tilde{L} \subseteq L^0$, with $(L^1, ||\cdot||)$ and $(\tilde{L}, ||\cdot||_{\tilde{L}})$ completions (as normed spaces) of $(\mathcal{S}, ||\cdot||_{L^1})$ in L^0. Notice we are slightly sloppy, as $\mathcal{S}$ is injected continuously into L^0 by two isometries $i_1 : \mathcal{S} \to L^1$ and $i_2 : \mathcal{S} \to \tilde{L}$. It might also confuse you that L^1 is used as a norm on $\mathcal{S}$ as well as on its image by i_1, but all shall be clear from the context.

Claim: $L^1 = \tilde{L}$ and $||\cdot||_{L^1} = ||\cdot||_{\tilde{L}}$.

Proof. Take $X \in \tilde{L}$. Then, since $\tilde{L}$ is a completion of $\mathcal{S}$, there is a sequence $X_n \to X$ in $\tilde{L}$ (i.e. $||X_n - X||_{\tilde{L}}) \to 0$). Teh sequence X_n is then also Cauchy in L^1 since $||X_n - X_m||_{L^1} = ||X_n - X_m||_{\tilde{L}}$, so it converges in L^1 to some $X' \in L^0$. But since the identity maps are continuous, $X_n \to X$ in probability and $X_n \to X'$ in probability as well (the Ky Fan metric models convergence in probability!). Hence $X = X'$. So $\tilde{L} \subseteq L^1$. Symmetrically, this argument yields $L^1 \subseteq \tilde{L}$. We have shown that $L^1 = \tilde{L}$.

The second claim follows immediately from continuity of norms:

$$||X'||_{L^1} = \lim_{n\to\infty} \mathbb{E}[|X_n|] = ||X'||_{\tilde{L}},$$

where $X_n \stackrel{L^1}{\to} X \iff X_n \stackrel{\tilde{L}}{\to} X$. $\square$

5.6 Ex 5.6 - Challenge

No solution provided. Your solutions are welcome!

6 Solutions - Sheet 6

6.1 Ex 6.1

First note that $\sum_{n\geq 1} Cn^{-1-\epsilon} < \infty$, since $1+\epsilon > 1$. $X = \sum_{n=1}^{\infty} n\mathbb{1}_{X=n}$, so by monotone convergence

$$\begin{aligned}
\mathbb{E}[X] &= \lim_{k\to\infty} \mathbb{E}[\sum_{n=1}^{k} n\mathbb{1}_{\{X=n\}}] \\
&= \sum_{n=1}^{\infty} n\mathbb{P}(\{X=n\}) \\
&= \sum_{n=1}^{\infty}\sum_{i=1}^{n} \mathbb{P}(\{X=n\}) \\
&= \sum_{i=1}^{\infty}\sum_{n=i}^{\infty} \mathbb{P}(\{X=n\}) \\
&= \sum_{i\geq 1} \mathbb{P}(\{X\geq i\}) \quad \square
\end{aligned}$$

6.2 Ex 6.2

We recall that the Ky Fan metric models convergence in probability, i.e.

$$d_{KF}(X_n, Y) \to 0 \iff X_n \xrightarrow{\mathbb{P}} Y.$$

Thus what we have to prove is that our present metric models convergence in probability.

Claim 1: $d_{L^0}(X,Y) := \mathbb{E}[|X-Y|\wedge 1]$ is a metric.

Proof. Symmetry is immediate. Obviously, $d_{L^0}(X,Y)\geq 0$. Then $d_{L^0}(X,Y) = 0 \iff |X-Y|\wedge 1 = 0$ a.s. $\iff |X-Y| = 0$ a.s., which proves we are not considering a pseudometric.

Now to the triangle inequality. We have that

$$\begin{aligned}
d_{L^0}(X,Y) &:= \mathbb{E}[|X-Z+Z-Y|\wedge 1] \\
&\leq \mathbb{E}[|X-Z|\wedge 1 + |Z-Y|\wedge 1] \\
&= d_{L^0}(X,Z) + d_{L^0}(Z,Y),
\end{aligned}$$

where in the first inequality we have used that $|(X-Z)+(Z-Y)|\wedge 1 \leq (|X-Z|+|Z-Y|)\wedge 1 \leq |X-Z|\wedge 1 + |Z-Y|\wedge 1$.

$\square$

Claim 2: $\mathbb{E}[|X_n - X|\wedge 1] \underset{n\to\infty}{\to} 0 \iff X_n \xrightarrow{\mathbb{P}} X$.

Proof. " $\Rightarrow$ ". If $\epsilon \in (0,1]$, we have that $\epsilon\cdot\mathbb{P}(\{|X_n - X| > \epsilon\}) \leq \mathbb{E}[|X_n - X|\wedge 1] \underset{n\to\infty}{\to} 0$. So $\mathbb{P}(\{|X_n - X| > \epsilon\} \underset{n\to\infty}{\to} 0$ if $\epsilon\in(0,1]$ (and a posteriori for all ϵ).

" $\Leftarrow$ ". $\mathbb{E}[|X_n - X|\wedge 1] \leq \mathbb{E}[(|X_n - X|\wedge 1)\cdot\mathbb{1}_{\{|X_n - X|\leq\epsilon\}}] + \mathbb{E}[|X_n - X|\wedge 1)\cdot\mathbb{1}_{\{|X_n-X|>\epsilon\}}]$. The first term is smaller than ϵ, the second smaller or equal to $\mathbb{P}(\{|X_n - X| < \epsilon\})$, which goes to zero as n goes to infinity. We are done. $\square$

6.3 Ex 6.3

For part (a), recall that a singleton $X \in L^1$ is uniformly integrable. So

$$\begin{aligned} t \cdot \mathbb{P}(\{|X| > t\}) &= t \cdot \mathbb{E}[\mathbb{1}_{\{|X|>t\}}] \\ &\leq \mathbb{E}[|X|\mathbb{1}_{\{|X|>t\}}] \\ &\overset{t\to\infty}{\longrightarrow} 0, \end{aligned}$$

using uniform integrability for the limit.

In exercise (b), we suppose w.l.o.g. that $X \geq 0$ and calculate:

$$\begin{aligned} \mathbb{E}[X] &\leq \mathbb{E}\left[\sum_{j=1}^{\infty} j \cdot \mathbb{1}_{\{X\in[j-1,j)\}}\right] \\ &= \lim_{N\to\infty} \sum_{j=1}^{N} j \cdot (\mathbb{P}(\{X < j\}) - \mathbb{P}(\{X < j-1\})) \\ &= \lim_{N\to\infty} \left(\sum_{j=1}^{N} j \cdot \mathbb{P}(\{X < j\}) - \sum_{j=0}^{N-1} (j+1) \cdot \mathbb{P}(\{X < j\})\right) \\ &= \lim_{N\to\infty} \left(N \cdot \mathbb{P}(\{X < N\}) - \sum_{j=0}^{N-1} \mathbb{P}(\{X < j\})\right) \\ &= \lim_{N\to\infty} \left(N - N \cdot \mathbb{P}(\{X \geq N\}) + \sum_{j=0}^{N-1} \mathbb{P}(\{X \geq j\}) - N\right) \\ &\leq \lim_{N\to\infty} \sum_{j=0}^{N-1} c \cdot j^{-1-\epsilon} \\ &< \infty, \end{aligned}$$

where for the second last inequality we introduced $c = \sup_{t\geq 0} t^{1+\epsilon} \cdot \mathbb{P}(\{X > t\})$ and the fact that by hypothesis, $N \cdot \mathbb{P}(\{X \geq N\}) \underset{N\to\infty}{\to} 0$.

For point (c), we shall use the "Cauchy condensation test": let $a_n \downarrow 0$, then

$$\sum a_n < \infty \iff \sum 2^n a_{2^n} < \infty.$$

Take $\mathbb{P}(\{|X| = k\}) = \frac{c}{k^2 \log(k)}$ (where c is just scaled so that $\mathbb{P}$ is a probability and ≥ 1).

$\mathbb{E}[X] = \sum_{k\geq 1} \frac{c}{k\log(k)}$, but by Cauchy's condensation criterion, we have that $\sum_{n\geq 1} \frac{2^n c}{2^n \log 2^n} \sim \sum_{n\geq 1} \frac{1}{n} = \infty$. So $\mathbb{E}[X] = \infty$.

However, $\mathbb{P}(\{X \geq k\}) = \sum_{j=k}^{\infty} \mathbb{P}(\{X = j\}) = \sum_{j=k}^{\infty} \frac{c}{j^2 \log(j)} \leq \sum_{j=k}^{\infty} \frac{c}{j^2 \log(k)}$. By using that $\int \frac{1}{x^2} = -\frac{1}{x}$ and playing smartly with Riemann sums, we bound the last term in our calculation by $\frac{\tilde{c}}{k\log(k)}$ for some constant $\tilde{c}$, whence $k \cdot \mathbb{P}(\{X \geq k\}) \leq \frac{\tilde{c}}{k\log(k)} \underset{k\to\infty}{\longrightarrow} 0 \quad \square$.

6.4 Ex 6.4

Let $(X_i)_{i\in I}$ a family of bounded random variables in L^1 (i.e. $\sup_{i\in I} ||X_i|| < \infty$). We must show the following two assertions (known as uniform integrability) are equivalent (note the second point assertion implies L^1-boundedness):

1. $\forall\, \epsilon > 0\; \exists\, \delta > 0\ :\ A \in \mathcal{F},\ \mathbb{P}(A) > 0\ \Rightarrow\ \forall\, i \in I,\ \mathbb{E}[|X_i|\mathbb{1}_A] < \epsilon$
2. $\lim_{a\to\infty} \sup_{i\in I} \mathbb{E}[|X_i|\mathbb{1}_{|X_i>a|}] = 0$.

$2 \Rightarrow 1$: Let $\epsilon > 0$. Choose $a > 0$ large enough so that

$$\sup_{i\in I} \mathbb{E}[\mathbb{1}_{\{|X_i|>a\}}] < \frac{\epsilon}{2}.$$

Then for all $i \in I$ and $A \in \mathcal{F}$ we have:

$$\begin{aligned}\mathbb{E}[|X_i|\mathbb{1}_A] &\leq \mathbb{E}[|X_i|\mathbb{1}_{A\cap\{|X_i|\leq a\}}] + \mathbb{E}[|X_i|\mathbb{1}_{\{|X_i|>a\}}] \\ &\leq a\mathbb{P}(A) + \frac{\epsilon}{2}.\end{aligned}$$

By choosing $\delta = \frac{\epsilon}{2a}$, we are done. □

$1 \Rightarrow 2$: Let $c = \sup_{i\in I} \mathbb{E}[|X_i|]$. By Markov's inequality, for any $a > 0$, we have

$$\forall\, i \in I\ :\ \mathbb{P}(\{|X_i| > a\}) \leq \frac{c}{a}.$$

Let $\epsilon > 0$ and choose $\delta > 0$ such that the left hand side of (1) is verified. Then if a is large enough so that $\frac{c}{a} < \delta$ we get:

$$\forall\, i \in I :\ \mathbb{E}[|X_i|\mathbb{1}_{\{|X_i|>a\}}] < \epsilon,$$

so we are done! □

6.5 Ex 6.5

A proof can be found in [1], namely this is the content of Theorem 4.6.3. In the proof of the theorem, a few references are made to other results, which are all referenced in the book (even though one result is left as an exercise). You might wish to skip this exercise on a first reading and come back to it later, or at least after having solved the other exercises of the exercise sheet.

6.6 Ex 6.6

Note that if $\sup_{i\in I} ||X_i||_{L^p} < \infty$ for $p > 1$, then we satisfy the assumptions of this theorem (from (b) to (a)). We will prove that (b) implies (a) under weaker conditions on ϕ: we drop convexity but retain all other conditions. We use the characterization of uniform integrability from 6.4.

First: for all $\epsilon > 0$ there is an x_ϵ such that $x \geq x_\epsilon \Rightarrow x \leq \epsilon\phi(x)$. Hence, by denoting $M := \sup_{i\in I} \mathbb{E}[\varphi(|X_i|)] < \infty$:

$$\begin{aligned}\sup_{i\in I} \mathbb{E}[|X_i|\mathbb{1}_{\{|X_i|\geq x_\epsilon\}}] &\leq \epsilon \sup_{i\in I} \mathbb{E}[\varphi(|X_i|)\mathbb{1}_{\{|X_i|\geq x_\epsilon\}}] \\ &\leq \epsilon \sup_{i\in I} \mathbb{E}[\varphi(|X_i|)] = \epsilon M.\end{aligned}$$

Whence $\lim_{a\to\infty} \mathbb{E}[|X_i|\mathbb{1}_{\{|X_i|>a\}}] = 0$. That is, $\{X_i\}_{i\in I}$ are u.i. □

7 Solutions - Sheet 7

7.1 Ex 7.1

We just show that λ **IS** the Lebesgue measure (denoted here by μ). Then in particular $\lambda([a,b]) = \mu([a,b])$ for $-\infty < a, b < \infty$.

Proof. Let $A \in \mathcal{B}(\mathbb{R})$, then

$$\begin{aligned}\lambda(A) &= \sum_{n\in\mathbb{Z}} \lambda_n(A \cap [n, n+1]) \\ &= \sum_{n\in\mathbb{Z}} \mu(A \cap [n, n+1]) \\ &= \sum_{n\in\mathbb{Z}} \mu(A \cap [n, n+1)) \\ &= \mu(A),\end{aligned}$$

since $\bigsqcup_{n\in\mathbb{Z}}[n, n+1) = \mathbb{R}$. □

7.2 Ex 7.2

See also Section 14 for an alternative solution.

Suppose first that two **finite** measures μ and ν of equal mass on a space $(\Omega, \mathcal{G})$ coincide on a pi-system $\Pi \subseteq \mathcal{G}$.

Claim: $\mu = \nu$ on $\sigma(\Pi)$.

Proof. We prove that $\mathcal{G}' = \{C \in \mathcal{G} \mid \mu(C) = \nu(C)\}$ is a sigma-algebra.

- $\mu(\emptyset) = 0 = \nu(\emptyset)$, so $\emptyset \in \mathcal{G}'$.
- If $C \in \mathcal{G}'$ then $\mu(C^c) = \mu(\Omega) - \mu(C) = \nu(\Omega) - \nu(C) = \nu(C^c)$.
- Suppose $(C_i)_{i\in\mathbb{N}} \in \mathcal{G}^{\mathbb{N}}$ and $\mu(C_i) = \nu(C_i)$ for all i and the C_i be disjoint. Then $\mu(\sqcup_{n\in\mathbb{N}} C_i) = \sum_{i\in\mathbb{N}} \mu(C_i) = \sum_{i\in\mathbb{N}} \nu(C_i) = \nu(\sqcup_{n\in\mathbb{N}} C_i)$. So $\sqcup_{i\in\mathbb{N}} C_i \in \mathcal{G}'$.

$\mathcal{G}'$ is thus a sigma-algebra. □

Now back to our exercise. Notice that for all n, $A_n \cap \Pi$ is a pi-system on A_n and $A_n \cap \Pi \subseteq \Pi$. Furthermore we know that $\mu(A_n) = \nu(A_n)$ for all n as well. We use the following fact (exercise for you: prove it):

$$\sigma|_A(A \cap \Pi) = A \cap \sigma(\Pi).$$

In words: "the sigma-algebra of the trace is the trace of the sigma-algebra".

Let's suppose $\sigma(\Pi) = \mathcal{G}$. Then $\sigma|_A(A_i \cap \Pi) = A \cap \sigma(\Pi) = A \cap \mathcal{G}$. We recall that $A_i \cap \Pi$ is a pi-system on A_i. So from the claim above: $\mu|_{A_i\cap\mathcal{G}} = \nu|_{A_i\cap\mathcal{G}}$ for each $i \in \mathbb{N}$.

We're in a position to conclude. We suppose w.l.o.g. (see below for an explanation) that $A_i \subseteq A_{i+1}$ for all i. Taking $B \in \mathcal{G}$,

$$\begin{aligned}\mu(B) &= \mu\big(\cup_{i\in\mathbb{N}}(B \cap A_i)\big) \\ &= \lim_{n\to\infty} \mu(B \cap A_i) \\ &= \lim_{n\to\infty} \nu(B \cap A_i) \\ &= \nu(B).\end{aligned}$$

The "without loss of generality" assumption remains to be seen. Again without loss of generality (why?), this boils down to showing that if A_i and A_j are two elements of the sequence that covers T, then μ and ν coincide on $\mathcal{G}|_{A_i \cup A_j}$. Taking $E \in \mathcal{G}$ we thus have that $\mu(E \cap A_i) = \nu(E \cap A_i)$ and $\mu(E \cap A_j) = \nu(E \cap A_j)$. We wish to prove that

$$\mu(E \cap (A_i \cup A_j)) = \nu(E \cap (A_i \cup A_j)).$$

We write $A_i \cup A_j = A_i^{'} \sqcup D \sqcup A_j^{'}$, where $D = A_i \cap A_j$, $A_i^{'} = A_1 \backslash D$ and $A_j^{'} = A_j \backslash D$. We unfold the definitions to conclude:

$$\begin{aligned}
\mu(E \cap (A_i \cup A_j)) &= \mu((E \cap A_i^{'}) \cup (E \cap D) \cup (E \cap A_j^{'})) \\
&= \mu(E \cap A_i^{'}) + \mu(E \cap D) + \mu(E \cap A_j^{'}) \\
&= \mu(E \cap A_i^{'} \cap A_i) + \mu(E \cap D \cap A_i) + \mu(E \cap A_j^{'} \cap A_j) \\
&= \nu(E \cap A_i^{'} \cap A_i) + \nu(E \cap D \cap A_i) + \nu(E \cap A_j^{'} \cap A_j) \\
&= \nu(E \cap A_i^{'}) + \nu(E \cap D) + \nu(E \cap A_j^{'}) \\
&= \nu((E \cap A_i^{'}) \cup (E \cap D) \cup (E \cap A_j^{'})) \\
&= \nu(E \cap (A_i \cup A_j)).
\end{aligned}$$

Note that the last step (w.l.o.g.) does not constitute the real heart of the argument, but rather a baroque variation. The important concept is really about working locally and going to the limit (the whole space). □

7.3 Ex 7.3

Solution 1: measure-theoretic spirit

By scaling we may assume that $a = 0$ and $b = 1$ so we are working on the probability space $([0,1], \mathcal{B}([0,1]), dx)$, where $\mathcal{B}$ is the Borel σ-algebra and dx denotes the Lebesgue measure. Define the simple functions

$$g_n(x) = \sum_{k=0}^{n-1} n \left(f\left(\frac{k+1}{n}\right) - f\left(\frac{k}{n}\right) \right) \mathbb{1}_{\left[\frac{k}{n}, \frac{k+1}{n}\right)}(x)$$

which are constant on every interval of the form $I_k = \left[\frac{k}{n}, \frac{k+1}{n}\right)$. Notice first that we have the telescoping sum

$$\int_0^1 g_n(x)\,dx = \sum_{k=0}^{n-1} n \left(f\left(\frac{k+1}{n}\right) - f\left(\frac{k}{n}\right) \right) \cdot \frac{1}{n} = \sum_{k=0}^{n-1} \left(f\left(\frac{k+1}{n}\right) - f\left(\frac{k}{n}\right) \right) = f(1) - f(0).$$

Notice then that when $x \in I_k$, then by the mean value theorem

$$g_n(x) = f'(\xi)$$

for some $\xi \in I_k$. In particular $|x - \xi| \le \frac{1}{n}$ and we have

$$|g_n(x) - f'(x)| = |f'(\xi) - f'(x)| \le \omega(|\xi - x|) \le \omega(1/n)$$

where ω is the modulus of continuity (see Prop. A.7. in the lecture notes if needed) of the uniformly continuous function f' (a continuous function on a compact interval is uniformly continuous). Notice that the right hand side in the inequality does not depend on the interval I_k, and thus in fact we see that $g_n \to f'$ in $L^\infty([0,1])$. As L^∞ is continuously embedded in L^1, we also have $g_n \to f'$ in $L^1([0,1])$ and in particular

$$\int f'(x)\,dx = \lim_{n\to\infty} \int g_n(x)\,dx = f(1) - f(0).$$

Solution 2: Riemann sums spirit

Recall that a continuous function f (here f') is Riemann integrable. Split the interval $[a,b]$ into intervals of length $\frac{b-a}{n}$, write a telescopic sum and use the mean value theorem:

$$\begin{aligned} f(b)-f(a) &= \sum_{k=0}^{n-1}\left(f\left(a+(b-a)\frac{k+1}{n}\right)-f\left(a+(b-a)\frac{k}{n}\right)\right) \\ &= \sum_{k=0}^{n-1}\frac{b-a}{n}f'(x_n^k), \end{aligned}$$

with $x_n^k \in \left(a+(b-a)\frac{k}{n}, a+(b-a)\frac{k+1}{n}\right)$.

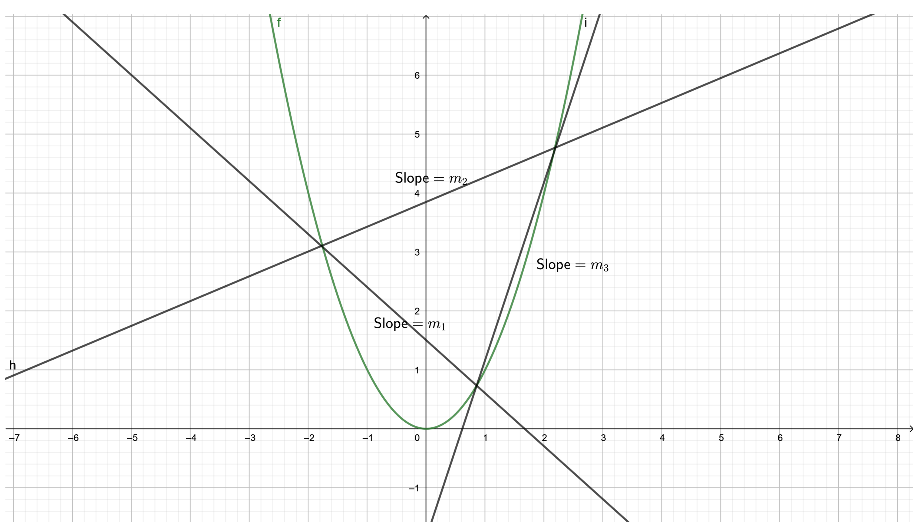
The last term can be identified as a Riemann sum. As n goes to infinity, it thus converges to $\int_a^b f'(x)\,dx$, f' being continuous.

7.4 Ex 7.4

Let $f:\mathbb{R}\to\mathbb{R}$ be convex. We set out to prove Jensen's inequality. We claim that for $s<t<u$ the relation

$$\frac{f(t)-f(s)}{t-s} \le \frac{f(u)-f(s)}{u-s} \le \frac{f(u)-f(t)}{u-t}$$

holds. Here is a graphical proof, where $m_1<m_2<m_3$ translates instantaneously into our desired relation. s, t and u are of course the x-coordinates of the three intersection points of the three lines.



As a consequence, fixing a point x, the following limits

$$R_x = \lim_{h\to 0^+}\frac{f(x+h)-f(x)}{h}$$

and

$$L_x = \lim_{h\to 0^-}\frac{f(x-h)-f(x)}{h}$$

exist, where R and L stand for left and right. Of course, $L_x \le R_x$ and

$$L_x = R_x \iff \text{f differentiable at x.}$$

As a consequence, we immediately have continuity of f (these reasonings lead to deduce interesting properties of differentiability for convex functions... think about playing around with reasonings such as "how many points of non-differentiability are there"... similar reasonings as those applied to the cumulative distribution function for instance).

Now consider the line of slope $m \in [L_x, R_x]$ at the point $(x_0, \varphi(x_0)$ on the graph of φ. Then $\varphi(x) \geq ax + b$ for all x in $\mathbb{R}$.

We have $\varphi(X) \geq aX + b$. Suppose $X \in L^1$. Consider the line above at the point $(\mathbb{E}[X], a\mathbb{E}[X] + b)$, noting that $a\mathbb{E}[X] + b = \varphi(\mathbb{E}[X])$. Then $\mathbb{E}[\varphi(X)] \geq a\mathbb{E}[X] + b = \varphi(\mathbb{E}[X])$ □.

Now to the last part of the exercise. Suppose $\mathbb{E}[X_+] < \infty$ and $\mathbb{E}[X_-] < \infty$. The case $\mathbb{E}[X_+] = -\infty$ and $\mathbb{E}[X_-] < \infty$ is treated similarly. 3 cases have to be analyzed:

$$\varphi(\infty) = \begin{cases} +\infty \\ 0 \\ -\infty. \end{cases}$$

- Suppose $\varphi(\infty) = \infty$. $\varphi(\mathbb{E}[X]) = \varphi(\infty) = \infty \overset{?}{\leq} \mathbb{E}[\varphi(X)]$. But then there is an $a > 0$ and a b such that $\varphi(X) \geq aX + b$. So $\mathbb{E}[\varphi(X)] \geq \mathbb{E}[aX + b] = \infty$. OK.

- Suppose $\varphi(\infty) = c \in \mathbb{R}$. Then $c = \varphi(\mathbb{E}[X]) \overset{?}{\leq} \mathbb{E}[\varphi(X)]$. But then there is a line $y = c$ such that $\varphi(x) \geq c$, whence $\varphi(X) \geq 0$ and so $\mathbb{E}[\varphi(X)] \geq c$. OK.

- Suppose $\varphi(\infty) = -\infty$. But then there is nothing to prove ($-\infty <$ anything), except that $\mathbb{E}[\varphi(X)]$ is well defined. We leave these details for the interested reader to solve. □

7.5 Ex 7.5

See also Section 14 for an alternative solution. Grateful thanks to Wei Jiaye for reviewing and typing my solution.

I'll prove a slightly more general statement: let $f : (X, \mu) \to \mathbb{R}^n$, where $f = (f_i)$, for $i = 1, \dots, n$ and $f_i \in L^1(\mu)$ for all $i = 1, \dots, n$. By definition,

$$\int f d\mu := \left(\int_X f_i d\mu \right)_{i=1}^n \in \mathbb{R}^n.$$

Theorem. $f \mapsto \|f\|$ is measurable and

$$\left\| \int_X f d\mu \right\| \leq \int_X \|f\| d\mu.$$

Proof. Every norm of $\mathbb{R}^n$ is continuous, thus $\|\cdot\|$ is measurable, so $\|f\|$ is measurable. Let $s : X \to \mathbb{R}^n$ be a (multi-dimensional) step function defined as

$$s = \sum_{i=1}^N s_i \mathbb{1}_{A_i},\ s_i \in \mathbb{R}^n,\ A_i \subset X \text{ disjoint, measurable sets for } 1 \leq i \leq N.$$

Then we have

$$\left\| \int_X s d\mu \right\| = \left\| \sum_{i=1}^N s_i \mu(A_i) \right\| \leq \sum_{i=1}^N \|s_i\| \mu(A_i) = \int_X \sum_{i=1}^N \|s_i\| \mathbb{1}_{A_i} d\mu = \int_X \|s\| d\mu,$$

taking into consideration that $s_i \mu(A_i) \in \mathbb{R}^n$. □

Since all norms in $\mathbb{R}^n$ are equivalent(they always are on finite-dimensional vector spaces), there exists $C \in (0, \infty)$ such that $\|v\| \leq C \max_{1\leq i\leq n} |v_i|, \forall v \in \mathbb{R}^n$. Let's choose $\varepsilon > 0$, for $s : X \to \mathbb{R}^n$ with

$$\max_{1\leq i\leq n} \int_X |(f-s)_i| d\mu \leq \frac{\varepsilon}{2C},$$

hence

$$\left\| \int_X (f-s) d\mu \right\| \leq C \max_{1\leq i\leq n} \left| \int_X (f-s)_i d\mu \right| \leq \frac{\varepsilon}{2}$$

and

$$\int_X \|f-s\| d\mu \leq C \max_{1\leq i\leq n} \int_X |(f-s)_i| d\mu \leq \frac{\varepsilon}{2}.$$

Whence for all $\varepsilon > 0$,

$$\begin{aligned} \left\| \int_X f d\mu \right\| &\leq \left\| \int_X (f-s) d\mu \right\| + \left\| \int_X s d\mu \right\| \\ &\leq \frac{\varepsilon}{2} + \int_X \|s\| d\mu \\ &\leq \frac{\varepsilon}{2} + \int_X \|s-f\| d\mu + \int_X \|f\| d\mu \\ &\leq \varepsilon + \int_X \|f\| d\mu. \end{aligned}$$

□

7.6 Ex 7.6

This exercise/type of reasoning will be important in stochastic calculus in the context of Girsanov's theorem, a procedure by which one is able (among other things) to eliminate the drift of certain random processes.

What is meant here is that $\mathbb{Q}(A) = \mathbb{E}[e^{\frac{1}{2}X - \frac{a^2}{2}} \mathbb{1}_A]$ for $A \in \mathcal{F}$ (on a space $(\Omega, \mathcal{F}, \mathbb{P})$).

Let $B \in \mathcal{B}(\mathbb{R})$, then

$$\begin{aligned} \mathbb{Q}(\{X \in B\}) &= \mathbb{E}[e^{\frac{1}{2}X - \frac{a^2}{2}} \mathbb{1}_{X^{-1}(B)}] \\ &= \int_B \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} e^{\frac{1}{2}x - \frac{a^2}{2}} \, dx \\ &= \int_B \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}(x-a)^2} \, dx. \end{aligned}$$

But this corresponds to the distribution of a $\mathcal{N}(0,1)$ random variable. The second equality should be seen conceptually as a composition of functions, namely

$$e^{\frac{1}{2}X(\omega)} \mathbb{1}_{X^{-1}(B)}(\omega) = \big(x \mapsto e^{\frac{1}{2}x} \mathbb{1}_B(x)\big) \circ \big(\omega \mapsto X(\omega)\big).$$

8 Solutions - Sheet 8

8.1 Ex 8.1

Note that our problem is symmetric in x and y.

Our strategy (classical proof) is the following:

(a) Prove this is true for $f = \mathbb{1}_A$, with $A = A_1 \times A_2,\ A_1 \in \mathcal{G}_1,\ A_2 \in \mathcal{G}_2$.

(b) Prove this is true for $f = \mathbb{1}_A$, with $A \in \mathcal{G}_1 \otimes \mathcal{G}_2$.

(c) Prove this is true for any f that is $\mathcal{G}_1 \otimes \mathcal{G}_2$-measurable ($f : \mathcal{G}_1 \otimes \mathcal{G}_2 \to \mathbb{R}$).

(b) $\implies$ (c): Any such f can be written as $f = \lim_{n\to\infty} \uparrow f_n$, where the f_n are step functions of the form $f_n = \sum_{i=1}^{\alpha(n)} a_i \mathbb{1}_{A_i^n}$ for $A_i^n \in \mathcal{G}_1 \otimes \mathcal{G}_2$. But then $y \mapsto f(x,y)$ is a pointwise limit of measurable functions $y \mapsto_n (x,y)$, for $n \geq 1$. Thus $y \mapsto f(x,y)$ is measurable.

(a) $\implies$ (b): Let $\mathcal{D} = \{A \in \mathcal{G}_1 \otimes \mathcal{G}_2 \mid\ y \mapsto \mathbb{1}_A(x,y) \text{ is measurable}\}$. If we prove that $\mathcal{D}$ is is a λ-system (which contains the π-system of cylinder sets), then by Dynkin's λ-π theorem, we obtain that $\mathcal{D} = \mathcal{G}_1 \otimes \mathcal{G}_2$. The (standard) proof goes as follows:

- $\Omega \in \mathcal{D}$ is obvious since $1 \mapsto 1$ is a measurable function,
- Let $A \in \mathcal{D}$. $\mathbb{1}_{A^c} = 1 - \mathbb{1}_A(x,y)$ is measurable in y, so $A^c \in \mathcal{D}$,
- Let $A_1 \subseteq A_2 \subseteq \ldots$, then $\mathbb{1}_{\{\cup_{n\geq 1} A_n\}}(x,y) = \sum_{n\geq 1} \mathbb{1}_{A_n}(x,y)$. A pointwise limit of measurable functions being measurable, we conclude that $\cup_{n\geq 1} A_n \in \mathcal{D}$.

Proof of (a): $y \mapsto \mathbb{1}_{A_1 \otimes A_2}(x,y) = \mathbb{1}_{A_1}(x)\mathbb{1}_{A_2}(y)$ is $\mathcal{G}$-measurable for a fixed x, since $\mathbb{1}_{A_2}(y)$ is equal to 1 if $y \in A_2$ and to 0 if $y \notin A_2$.

8.2 Ex 8.2

$$
\begin{aligned}
\mathbb{E}[F(X)] &= \int_0^\infty F(x)\,\mathbb{P}_X(dx) \\
&= \int_0^\infty \int_0^x F'(u)\,du\,\mathbb{P}(dx) \\
&= \int_0^\infty \int_0^\infty \mathbb{1}_{\{u<x\}}\,du\,\mathbb{P}_X(dx) \\
&= \int_0^\infty F'(u) \int_0^\infty \mathbb{1}_{\{u<x\}}\,\mathbb{P}_X(dx)\,du \\
&= \int_0^\infty F'(u)\,\mathbb{P}(\{X > u\})\,du,
\end{aligned}
$$

where in the second equality we used that $F(0) = 0$ and in the second last Fubini's theorem on $\mathbb{P} \otimes \mu$ (positive version, since $F'(u) \geq 0$).

The two particular cases are immediate.

8.3 Ex 8.3

$X_n : (\Omega, \mathcal{F}, \nu) \to \mathbb{R}$, where ν is σ-finite, $n \in \mathbb{N}$. We work on $(\Omega \times \mathbb{N}, \nu \otimes \mu)$. Our function is $F : \Omega \times \mathbb{N} \to \mathbb{R} :: (\omega, n) \mapsto X_n(\omega)$.

By Fubini, we have the equivalence

$$\int_{\Omega \times \mathbb{N}} |F| \; \nu \otimes \mu < 0 \iff \int_{\mathbb{N}} \int_{\Omega} |X_n(\omega)| \; \nu(d\omega) d\mu < \infty.$$

The right-hand side is nothing but $\sum_{n \in \mathbb{N}} \mathbb{E}[|X_n|]$, which by assumption is $< \infty$. And then, by Fubini again:

$$\sum_{n \in \mathbb{N}} = \int_{\Omega} \int_{n \in \mathbb{N}} X_n(\omega) \; d\mu d\nu = \mathbb{E}\left[\sum_{n \in \mathbb{N}} X_n\right]. \quad \square$$

8.4 Ex 8.4

For an enlightening counter-example, see 14.

For a technically more involved but much less intuitive counter-example, download the PDF file called “nofub.pdf” at the following index: `https://math.jhu.edu/~jmb/note`, an example on John Michael Boardman’s webpage at John Hopkins University. A “post-mortem” on what actually happened in this counter-example (an explanation) is offered.

8.5 Ex 8.5

This is the content of Theorems 9.2.1 and Corollary 9.2.3 (pp.108-111) of [4]. The fact I am not writing the proof here should not lead you to think this is not important. This result and method of proof are essential in probability theory, so please do tackle this exercise!

8.6 Ex 8.6 - Challenge

No solution provided. Your solutions are welcome!

9 Solutions - Sheet 9

9.1 Ex 9.1

Possible solution: Let $X \sim \mathcal{U}\big((-1,1)\big)$ a uniform random variable and $Y := X^2$.

$\mathbb{E}[X] = 0$; $\mathbb{E}[Y] = \int_{-1}^{1} u^2 \frac{1}{2}\, du = \big[\frac{u^3}{6}\big]_{-1}^{1} = \frac{1}{3}$; $\mathbb{E}[XY] = \int_{-1}^{1} u^3 \frac{1}{2}\, du = 0$.

As a consequence $\sigma(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X] \cdot \mathbb{E}[Y] = 0$. But X and Y are not independent.

To see this, consider for example

$$\mathbb{P}\big(\{X \in [0,0.5),\, Y \in [0,0.5)\}\big) = \mathbb{P}\big(\{X \in [0,0.5)\}\big) = 0.25$$

and

$$\mathbb{P}\big(\{X \in [0,0.5)\}\big) \cdot \mathbb{P}\big(\{Y \in [0,0.5)\}\big) = \frac{1}{4} \cdot \frac{2}{\sqrt{2}} = \frac{1}{2\sqrt{2}} \neq \frac{1}{4}.$$

9.2 Ex 9.2

That independence yields the stated property is immediate. The converse is a usual Dynkin argument: two probability measures coinciding on a generating system are equal. In this case the measurable space is $\mathbb{R}^n$ together with it's Borel sigma-algebra. The probability measure on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ is the law of $(X_1, \dots, X_n)$, in other words the push-forward $\mathbb{Q}$ on $\mathbb{R}^n$ of the probability measure $\mathbb{P}$ by $(X_1, \dots, X_n)$, denoted $(X_1, \dots, X_n)_*\mathbb{P}$. The stated property just says that $\mathbb{Q}$ is the product measure on $\mathbb{R}^n$ of $(X_1)_*\mathbb{P}, \dots, (X_n)_*\mathbb{P}$. But by Fubini, this is equivalent to $X_1, \dots, X_n$ being independent.

9.3 Ex 9.3

(a) $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]\mathbb{1}_\Omega] = \mathbb{E}[X\mathbb{1}_\Omega] = \mathbb{E}[X]$.

(c) ($X \in L^1$). We have that $\mathbb{E}[X\mathbb{1}_A] = \mathbb{E}[X\mathbb{1}_A]$ for all $A \in \mathcal{G}$. X being $\mathcal{G}$-measurable, we conclude that $\mathbb{E}[X|\mathcal{G}] = X$ by unicity of the conditional expectation.

(e) $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]\mathbb{1}_A] = \mathbb{E}[X\mathbb{1}_A] \geq 0$ for all $A \in \mathcal{G}$. Consider $B = \{\omega \mid \mathbb{E}[X|\mathcal{G}] < 0\}$. Then $\mathbb{E}\big[\mathbb{E}[X|\mathcal{G}]\ \mathbb{1}_B\big] = 0$, since this quantity is both ≥ 0 and ≤ 0. But then $\mathbb{E}[X|\mathcal{G}] \cdot \mathbb{1}_B = 0$, a.s. Whence $\mathbb{E}[X|\mathcal{G}] \geq 0$, a.s.

(f) $X - Y \geq 0$ so $\mathbb{E}[X - Y|\mathcal{G}] \geq 0$ by point (e). As a consequence $\mathbb{E}[X|\mathcal{G}] \geq \mathbb{E}[Y|\mathcal{G}]$.

(d) By using that X is independent of $\mathcal{G}$ (meaning that $\sigma(X)$ and $\mathcal{G}$ are independent under $\mathbb{P}$), we have that $\mathbb{E}[X\mathbb{1}_A] = \mathbb{E}[X] \cdot \mathbb{E}[\mathbb{1}_A] = \mathbb{E}[\mathbb{E}[X]\mathbb{1}_A]$. As a consequence, $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$.

(b) Since $\mathcal{H} \subset \mathcal{G}$,we have (by (c), $\mathbb{E}[X|\mathcal{G}]$ being $\mathcal{G}$-measurable) that $\mathbb{E}[\mathbb{E}[X|\mathcal{H}]|\mathcal{G}] = \mathbb{E}[X|H]$. Now for $A \in \mathcal{H}$, using $\mathcal{G}$-measurability of $\mathbb{E}[X|\mathcal{G}]$ and that $\mathcal{H} \subset \mathcal{G}$, we have

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathbb{1}_A] = \mathbb{E}[X\mathbb{1}_A] = \mathbb{E}[\mathbb{E}[X|\mathcal{H}]\mathbb{1}_A].$$

As a consequence, $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}]$.

(g) We use Jensen's inequality, which will be proven in exercise 10.1. For $\varphi(x) = |x|$ (convex), we have

$$|\mathbb{E}[X_n|\mathcal{G}] - \mathbb{E}[X|\mathcal{G}]| = \mathbb{E}[(X_n - X) \mid \mathcal{G}]| \overset{\text{Jensen}}{\leq} \mathbb{E}[|X_n - X| \mid \mathcal{G}].$$

Taking expectations and letting $n \to \infty$, we get that $\mathbb{E}[X_n|\mathcal{G}] \overset{L^1}{\to} \mathbb{E}[X|\mathcal{G}]$.

9.4 Ex 9.4

I cannot stress how important this small theorem/lemma is to probability theory, with several applications in this lecture. Let's call it the **Independent Conditioning Theorem**. This theorem comes up periodically in the theory of stochastic analysis, Markov chains or processes, among others. The proof can be found on page 150 of [4].

9.5 Ex 9.5

Part (a): we first prove the following:

Claim: If $\mathcal{F}_1$ and $\mathcal{F}_2$ are independent sigma-algebras, and X is an integrable random variable independent of $\mathcal{F}_2$, then the following holds:

$$\mathbb{E}[X|\mathcal{F}_1 \vee \mathcal{F}_2] = \mathbb{E}[X|\mathcal{F}_1].$$

Phrased in plain words, you could read this statement as "adding an independent sigma-algebra gives no additional information on the expected value of X". Example zero of this property is when $\mathcal{F}_1 = \{\emptyset, \Omega\}$, in which case this reads $\mathbb{E}[X|\mathcal{F}_2] = \mathbb{E}[X]$.

In our present exercise we have

$$\sigma(S_n, S_{n+1}, \ldots) = \sigma(S_n, X_{n+1}, X_{n+2}, \ldots) = \sigma(S_n, \sigma(X_{n+1}, X_{n_1}, \ldots))$$

using the fact that $X_{n+k} = S_{n+k} - S_{n+k-1}$. S_n and $\sigma(X_{n+1}, X_{n+2}, \ldots)$ are independent, and X_n is independent of $\sigma(X_{n+1}, \ldots)$, so the above-stated claim proves point (a).

Proof. Let $A \in \mathcal{F}_1$, $B \in \mathcal{F}_2$, then $X\mathbb{1}_A$ and $\mathbb{1}_B$ are independent, whence

$$\begin{aligned}\mathbb{E}[X\mathbb{1}_A\mathbb{1}_B] &= \mathbb{E}[X\mathbb{1}_A]\mathbb{E}[\mathbb{1}_B] \\ &= \mathbb{E}[\mathbb{E}[X|\mathcal{F}_1]\mathbb{1}_A]\mathbb{E}[\mathbb{1}_B] \\ &= \mathbb{E}[\mathbb{E}[X|\mathcal{F}_1]\mathbb{1}_A\mathbb{1}_B].\end{aligned}$$

We want to go from $\mathbb{1}_{A\cap B}$, with $A \in \mathcal{F}_1$, $B \in \mathcal{F}_2$ to $\mathbb{1}_C$ with $C \in \mathcal{F}_1 \vee \mathcal{F}_2$. We use Dynkin's $\Pi - \Lambda$ Theorem. Define $\Pi := \{A \cap B \mid A \in \mathcal{F}_1, B \in \mathcal{F}_2\}$. Then Π is a Π-system that generates $\mathcal{F}_1 \vee \mathcal{F}_2$ (take $A = \Omega$ or $B = \Omega$, alternatively). We now claim that

$$\Lambda := \big\{C \in \mathcal{F}_1 \vee \mathcal{F}_2 \mid \mathbb{E}[X\ \mathbb{1}_C]\ \mathbb{1}_C] = \mathbb{E}[\mathbb{E}[X|\mathcal{F}_1]\ \mathbb{1}_C]\big\}$$

is a Λ-system, which would complete our proof. We thus set out to prove this claim:

- $\Omega \in \Lambda$ is clear: take $A = B = \Omega$
- $(A \cap B)^c = (A^c \cap B^c) \sqcup (A^c \cap B) \sqcup (A \cap B^c)$. Whence by linearity of the expectation we have that $(A \cap B)^c \in \Lambda$.
- If $(A_i \cap B_i)_{i\in\mathbb{N}}$ are disjoint, $A_i \cap B_i \in \Lambda$, then by dominated convergence we have that $\sqcup_{i\in\mathbb{N}}(A_i \cap B_i) \in \Lambda$.

We are done!

∎

Part (b): Let $A \in \sigma(S_n)$, then let $B \in \mathcal{B}(\mathbb{R}^n)$ such that $A = (X_1 + \ldots + X_n)^{-1}(B)$. We have that

$$\begin{aligned}\mathbb{E}[X_i \mathbb{1}_A] &= \int \mathbb{1}_B(X_1 + \ldots + X_n) \cdot X_i \; d\mathbb{P} \\ &= \int \mathbb{1}_B(x_1 + \ldots + x_n) \cdot x_i \; d\mathbb{P} \\ &= \int \mathbb{1}_B(x_1 + \ldots + x_n) \cdot x_i \; d(X_* \mathbb{P})^{\bigotimes n},\end{aligned}$$

which is constant in i by Fubini! Whence we have $\mathbb{E}[X_i \mathbb{1}_A] = \mathbb{E}[\frac{S_n}{n} \mathbb{1}_A]$ for all i, $\frac{S_n}{n}$ is $\sigma(S_n)$-measurable, so $\mathbb{E}[X_i | S_n] = \frac{S_n}{n}$. □

10 Solutions - Sheet 10

10.1 Ex 10.1

The present proofs for Jensen's and Hölder's inequalities are based on Jason Swanson's notes, found at http://math.swansonsite.com/, under the tab "Other documents", "Conditional expectation for professionals". Other interesting lecture notes/documents are also available.

We recall, for Jensen's inequality, that both X and $\varphi(X)$ have to be integrable. Furthermore, for $\varphi : \mathbb{R} \to \mathbb{R}$ a convex function, the left-hand derivative

$$\varphi_-^{'}(c) = \lim_{h \downarrow 0} \frac{\varphi(c) - \varphi(c-h)}{h}$$

exists for all c and

$$\varphi(x) - \varphi(c) - (x-c)\varphi_-^{'}(c) \geq 0,$$

for all x and c.

Proof. (of Jensen's inequality). Let $Z = (X - \mathbb{E}[X|\mathcal{G}])\varphi_-^{'}(\mathbb{E}[X|\mathcal{G}])$, so that $\varphi(X) - \varphi(\mathbb{E}[X|\mathcal{G}]) - Z \geq 0$, a.s. This implies

$$0 \leq \mathbb{E}\big[(\varphi(X) - \varphi(\mathbb{E}[X|\mathcal{G}]) - Z)|\mathcal{G}\big] = \mathbb{E}[\varphi(X)|\mathcal{G}] - \varphi(\mathbb{E}[X|\mathcal{G}]) - \mathbb{E}[Z|\mathcal{G}].$$

It therefore suffices to show that $\mathbb{E}[Z|\mathcal{G}] = 0$. To see this, we calculate

$$\begin{aligned}
\mathbb{E}[Z|\mathcal{G}] &= \mathbb{E}\big[(X - \mathbb{E}[X|\mathcal{G}])\varphi_-^{'}(\mathbb{E}[X|\mathcal{G}])|\mathcal{G}\big] \\
&= \varphi_-^{'}(\mathbb{E}[X|\mathcal{G}])\mathbb{E}\big[X - \mathbb{E}[X|\mathcal{G}]|\mathcal{G}\big] \\
&= \varphi_-^{'}(\mathbb{E}[X|\mathcal{G}])(\mathbb{E}[X|\mathcal{G}] - \mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{G}]) \\
&= 0,
\end{aligned}$$

and we are done. □

We shall require a very basic fact of measure theory for the proof of Hölder's inquality. Recall that if U and V are $\mathcal{L}^1$-integrable on a space $(\Omega, \mathcal{H}, \mathbb{P})$ then:

- $\mathbb{E}[U\mathbb{1}_A] \leq \mathbb{E}[V\mathbb{1}_A]\ \forall A \in \mathcal{H} \Rightarrow U \leq V$, a.s.
- $\mathbb{E}[U\mathbb{1}_A] = \mathbb{E}[V\mathbb{1}_A]\ \forall A \in \mathcal{H} \Rightarrow U = V$, a.s.

The reader for whom this fact is not obvious should try and prove it.

Proof. (of Hölder's inequality). We first consider conjugate exponents $p, q \in (1, \infty)$. Note that by the ordinary Hölder inequality, XY is integrable, so that $\mathbb{E}[|XY|\,|\mathcal{G}]$ is well defined. Let $U = (\mathbb{E}[|X|^p\,|\mathcal{G}])^{\frac{1}{p}}$ and $V = (\mathbb{E}[|Y|^q\,|\mathcal{G}])^{\frac{1}{q}}$. Note that both U and V are $\mathcal{G}$-measurable. Observe that

$$\begin{aligned}
\mathbb{E}[|X|^p\mathbb{1}_{\{U=0\}}] &= \mathbb{E}\big[\mathbb{E}[|X|^p\mathbb{1}_{\{U=0\}}|\mathcal{G}]\big] \\
&= \mathbb{E}\big[\mathbb{1}_{\{U=0\}}\mathbb{E}[|X|^p|\mathcal{G}]\big] \\
&= \mathbb{E}[\mathbb{1}_{\{U=0\}}U^p] \\
&= 0.
\end{aligned}$$

Hence, $|X|^p\mathbb{1}_{\{U=0\}} = 0$ a.s., which implies

$$\mathbb{E}[|XY|\,|\mathcal{G}]\mathbb{1}_{\{U=0\}} = \mathbb{E}[|XY|\mathbb{1}_{\{U=0\}}|\mathcal{G}] = 0.$$

Similarly, $\mathbb{E}[|XY|\,|\mathcal{G}]\mathbb{1}_{\{U=0\}} = 0$. It therefore suffices to show that $\mathbb{E}[|XY|\,|\mathcal{G}]\mathbb{1}_H \le UV$, where $H = \{U > 0, V > 0\}$. For this, we use the result recalled above to prove that

$$\frac{\mathbb{E}[|XY|\,|\mathcal{G}]}{UV}\mathbb{1}_H \le 1 \text{ a.s.}$$

Note that the left-hand-side is defined to be zero on H^c. Let $A \in \mathcal{G}$ be arbitrary and define $G = H \cap A$. Then

$$\begin{aligned}
\mathbb{E}\left[\frac{\mathbb{E}[|XY|\,|\mathcal{G}]}{UV}\mathbb{1}_H\mathbb{1}_A\right] &= \mathbb{E}\left[\mathbb{E}[\frac{|XY|}{UV}\mathbb{1}_G|\mathcal{G}]\right] \\
&\le \mathbb{E}\left[\frac{|X|}{U}\mathbb{1}_G \cdot \frac{|Y|}{V}\mathbb{1}_G\right] \\
&\le \left(\mathbb{E}\left[\frac{|X|^p}{U^p}\mathbb{1}_G\right]\right)^{\frac{1}{p}} \cdot \left(\mathbb{E}\left[\frac{|Y|^q}{V^q}\mathbb{1}_G\right]\right)^{\frac{1}{q}} \\
&= \left(\mathbb{E}\left[\frac{\mathbb{E}[|X|^p|\mathcal{G}]}{U^p}\mathbb{1}_G\right]\right)^{\frac{1}{p}} \cdot \left(\mathbb{E}\left[\frac{\mathbb{E}[|Y|^q|\mathcal{G}]}{V^q}\mathbb{1}_G\right]\right)^{\frac{1}{q}} \\
&= \mathbb{E}[\mathbb{1}_G]^{\frac{1}{p}}\mathbb{E}[\mathbb{1}_G]^{\frac{1}{q}} \\
&= \mathbb{E}[\mathbb{1}_G] \\
&\le \mathbb{E}[\mathbb{1}_A].
\end{aligned}$$

Applying one more time the result recalled above, we conclude.

In the case where $p = 1$ and $q = \infty$ then $|XY| \le |X|\|Y\|_\infty$, so

$$\mathbb{E}[XY|\mathcal{G}] \le \|Y\|_\infty \mathbb{E}[|X||\mathcal{G}] = \mathbb{E}[\|Y\|_\infty|\mathcal{G}]\mathbb{E}[X|\mathcal{G}].$$

□

We shall now derive Minkowski's inequality from Hölder's!

Proof. (of Minkowski's inequality). $\frac{1}{p} + \frac{1}{q} = 1$, with $p, q \in [1, \infty]$.

$$|u+v|^p \le (|u| + |v|)^p \le 2^p \max/(|u|^p, |v|^p) \le 2^p(|u|^p + |v|^p),$$

so in particular if $u, v \in \mathcal{L}^p$ then $|u+v|^p \in \mathcal{L}^1$ or if you prefer $u + v \in \mathcal{L}^p$. Consider

$$|u+v|^p = |u+v| \cdot |u+v|^{p-1} \le |u| \cdot |u+v|^{p-1} + |v| \cdot |u+v|^{p-1}.$$

Taking expectations yields

$$\mathbb{E}[|u+v|^p|\mathcal{G}] \le \mathbb{E}[|u| \cdot |u+v|^{p-1}|\mathcal{G}] + \mathbb{E}[|v| \cdot |u+v|^{p-1}|\mathcal{G}].$$

- If $p = 1$, $|u+v| \le |u| + |v|$, we are done.
- If $p = \infty$, $\|u+v\|_\infty \le \|u\|_\infty + \|v\|_\infty$, we are done.
- If $p \in (1, \infty)$, we use Hölder's inequality, conditional version:

  $$\mathbb{E}[|u+v|^p|\mathcal{G}] \le \underbrace{\mathbb{E}[|u|^p|\mathcal{G}]^{\frac{1}{p}}}_{\spadesuit}\mathbb{E}[|u+v|^{(p-1)q}|\mathcal{G}]^{\frac{1}{q}} + \mathbb{E}[|v|^p|\mathcal{G}]^{\frac{1}{p}}\mathbb{E}[(u+v)^{(p-1)q}|\mathcal{G}]^{\frac{1}{p}}.$$

  Note that $(p-1)q = (p-1)(\frac{p}{p-1}) = q$, a miracle has happened! Now divide on both sides by $\spadesuit$ (if $\spadesuit = 0$, we have $0 \le 0$, O.K.). Then

  $$\mathbb{E}[|u+v|^p|\mathcal{G}]^{1-\frac{1}{q}} \le \mathbb{E}[|u|^p|\mathcal{G}]^{\frac{1}{p}} + \mathbb{E}[|v|^p|\mathcal{G}]^{\frac{1}{p}}.$$

  Since $1 - \frac{1}{q} = \frac{1}{p}$, we are done.

□

10.2 Ex 10.2

The function $x \mapsto x^p$ is convex for $p \geq 1$. we thus use Jensen's inequality (proven in 10.1). This reads

$$|\mathbb{E}[X|\mathcal{G}]|^p \leq \mathbb{E}[|X|^p|\mathcal{G}] \quad \text{a.s.}$$

Taking expectations reads $\mathbb{E}[|\mathbb{E}[X|\mathcal{G}]|^p] \leq \mathbb{E}[|X|^p]$, or $\|\mathbb{E}[X|\mathcal{G}]_p\| \leq \|X\|_p$ when taking taking the power $\frac{1}{p}$. □

10.3 Ex 10.3

Let $A \in \mathcal{G}$. Then

$$\mathbb{E}[\mathbb{1}_A(X - \mathbb{E}[X|\mathcal{G}])] = \mathbb{E}[\mathbb{1}_A X] - \mathbb{E}[\mathbb{1}_A \mathbb{E}[X|\mathcal{G}]] = 0,$$

by the very definition of conditional expectation. Suppose $Z \in \mathcal{L}^2(\Omega, \mathcal{G}, \mathbb{P})$, then we have $Z_n \overset{\mathcal{L}^2}{\to} Z$, where Z_n is a sequence of step functions of the form $Z_n = \sum_{i=1}^{n_k} a_i \mathbb{1}_{A_i^n}$ with $A_i^n \in \mathcal{G}$. From this, we have

$$\left|\mathbb{E}[(Z - Z_n)(X - \mathbb{E}[X|\mathcal{G}])]\right| \overset{C.S.}{\leq} \|Z - Z_n\|_{L^2} \|X - \mathbb{E}[X|\mathcal{G}]\|_{L^2}.$$

The first term on the r.h.s goes to zero as n goes to infinity. So for all $Z \in \mathcal{L}^2(\Omega, \mathcal{G}, \mathbb{P})$ we have that $\mathbb{E}[Z(X - \mathbb{E}[X|\mathcal{G}])] = 0$, where both random variables in the product are in $\mathcal{L}^2$. In other words, $X - \mathbb{E}[X|\mathcal{G}] \in \mathcal{L}^2(\Omega, \mathcal{G}, \mathbb{P})^\perp$, whence $\mathbb{E}[X|\mathcal{G}] = \text{Proj}_{\mathcal{L}^2(\Omega, \mathcal{G}, \mathbb{P})} X$, i.e. $\mathbb{E}[X|\mathcal{G}]$ minimizes $\mathbb{E}[|X - Y|^2]$ among all $Y \in \mathcal{L}^2$ with Y a $\mathcal{G}-$measurable function. □

A second approach, which does not use the Hilbert space structure of $\mathcal{L}^2$ (namely $\langle f|g \rangle = \mathbb{E}[fg]$). To this end, let Y be $\mathcal{G}$-measurable. Then

$$\begin{aligned}
\mathbb{E}[|X - Y|^2] &= \mathbb{E}[|X - \mathbb{E}[X|\mathcal{G}] + \mathbb{E}[X|\mathcal{G}] - Y|^2] \\
&= \mathbb{E}[|X - \mathbb{E}[X|\mathcal{G}]|^2 + |\mathbb{E}[X|\mathcal{G}] - Y|^2 + 2(X - \mathbb{E}[X|\mathcal{G}])(\mathbb{E}[X|\mathcal{G}] - Y)] \\
&= \mathbb{E}[|X - \mathbb{E}[X|\mathbb{G}]|^2] + \mathbb{E}[|\mathbb{E}[X|\mathcal{G}] - Y|^2] + 2\mathbb{E}[\mathbb{E}[(X - \mathbb{E}[X|\mathcal{G}])(\mathbb{E}[X|\mathcal{G}] - Y)|\mathcal{G}]] \\
&= \mathbb{E}[|X - \mathbb{E}[X|\mathbb{G}]|^2] + \mathbb{E}[|\mathbb{E}[X|\mathcal{G}] - Y|^2] + 2\mathbb{E}[(\mathbb{E}[X|\mathcal{G}] - Y)\mathbb{E}[X - \mathbb{E}[X|\mathcal{G}]|\mathcal{G}]] \\
&= \mathbb{E}[|X - \mathbb{E}[X|\mathcal{G}]|^2] + \mathbb{E}[|\mathbb{E}[X|\mathcal{G}] - Y|^2],
\end{aligned}$$

by the law of total expectation $\mathbb{E}[\cdot] = \mathbb{E}[\mathbb{E}[\cdot|\mathcal{G}]]$ and since $\mathbb{E}[X - \mathbb{E}[X|\mathcal{G}]|\mathcal{G}] = 0$. Thus

$$\mathbb{E}[|X - Y|^2] \geq \mathbb{E}[|X - \mathbb{E}[X|\mathcal{G}]|^2]$$

with equality if and only if $\mathbb{E}[X|\mathcal{G}] = Y$ a.s. □

Important Note: You should come back to this exercise in your future study of Gaussian spaces (for example in the context of linear regression theory), since in specific cases the conditional expectation will coincide with a linear combination of Gaussian vectors spanning (meaning the closure of the span) what is here denoted as $\mathcal{L}^2(\Omega, \mathcal{G}, \mathbb{P})$.

10.4 Ex 10.4

This is an essential result in probability theory and a special case of conditioning for random variables that possess a density function. You should try and give the proof and result an interpretation. The full proof is the content of page 151 in [4].

10.5 Ex 10.5

(a): We can proceed by direct computation:

$$\mathbb{E}[\exp(tX)] = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}\sigma} \exp\Big(-\frac{1}{2}\frac{x^2}{\sigma^2}\Big) \exp(tx)\ dx = \exp\Big(\frac{1}{2}\sigma^2 t^2\Big),$$

where we identified and used the fact that the density of $\mathcal{N}(\sigma^2 t, \sigma)$ integrates to 1.

Alternatively, we can proceed by using 7.6. If $X \sim \mathcal{N}(0, \sigma^2)$, then $\frac{X}{\sigma} \sim \mathcal{N}(0,1)$. By 7.6, the function $\exp\Big(a\frac{X}{\sigma} - \frac{a^2}{2}\Big)$ defines a probability measure, which implies it integrates to one. Taking $a = t\sigma$ yields $\int_{\mathbb{R}} \exp\Big(tX - \frac{t^2\sigma^2}{2}\Big)\ dx = 1$, which implies that $\int_{\mathbb{R}} \exp(tX) = \exp\Big(\frac{t^2\sigma^2}{2}\Big)$.

(b): The Chernoff bound is immediate (just an application of Markov's inequality with $\varphi(x) = \exp(tx)$. So $\mathbb{P}(\{X > \lambda\}) \leq \exp(-t\lambda)\exp\Big(\frac{t^2}{2}\sigma^2\Big)$ ♠ by (a). $\big(-t\lambda + \frac{t^2}{2}\sigma^2\big)' = -\lambda + t\sigma^2 = 0 \iff t = \frac{\lambda}{\sigma^2}$. As a consequence, ♠ $\leq \exp\Big(-\frac{1}{2}\frac{\lambda^2}{\sigma^2}\Big)$. □

10.6 Ex 10.6

Take logs in order to make the law of large numbers (LLN) appear!

$$Y_n = \big(\Pi_{k=1}^n X_k\big)^{\frac{1}{n}} \Rightarrow \log(Y_n) = \frac{1}{n}\big(\sum_{k=1}^n \log X_k\big).$$

Note: to avoid $\log(0) = -\infty$, equivalently, consider the uniform distribution on $(0,1]$ or on $(0,1)$. Anyway, we don't worry, because $\mu(\{0\}) = 0$. We have $\mathbb{E}[\log(X_k)] = \lim_{\epsilon\to 0}\int_\epsilon^1 \log(x)\ dx = -1$. Whence by the LLN, $\log(Y_n) \overset{n\to\infty}{\longrightarrow} -1$, a.s. Consequently, $\big(\Pi_{k=1}^n X_k\big)^{\frac{1}{n}} \overset{n\to\infty}{\longrightarrow} \exp(-1)$, a.s. □

10.7 Ex 10.7

Let $M \in \mathbb{R}$. Then $\mathbb{P}(\{\frac{S_n}{\sqrt{n}} > M\}) = \Phi(M)$, where Φ is the cumulative distribution function of a $\mathcal{N}(0,1)$ random variable.

$$\begin{aligned}\mathbb{P}\big(\{\frac{S_n}{\sqrt{n}} > M\}\big) &= \mathbb{E}\big[\mathbb{1}_{\{\liminf_{n\to\infty} \frac{S_n}{\sqrt{n}} > M\}}\big] \\ &= \mathbb{E}\big[\liminf_{n\to\infty} \frac{S_n}{\sqrt{n}} > M\big] \\ &\leq \limsup_{n\to\infty} \mathbb{E}[\mathbb{1}_{\{\frac{S_n}{\sqrt{n}\leq M}\}}] \\ &= \Phi(M) < 1,\end{aligned}$$

where in the last equality particular care was required to unfold and use the following chain of equivalences:

$$\mathbb{1}_{\{\cdot\}}(\omega) = 1 \iff \liminf_{n\to\infty} \frac{S_n}{\sqrt{n}}(\omega) > M \iff \exists\, k : \inf_{n\geq k} \frac{S_n}{\sqrt{n}}(\omega) > M \iff \liminf_{n\to\infty} \mathbb{1}_{\{\frac{S_n}{\sqrt{n}} > M\}}(\omega) = 1.$$

Now let $(X_i)_{i\in\mathbb{N}}$ be a sequence of random variables. We claim that $\liminf_{n\to\infty} X_n$ is measurable with respect to the tail sigma-algebra $\cap_{k=1}^\infty \sigma(\cup_{n\geq k}\sigma(X_n))$. Indeed, for $x \in \mathbb{R}$:

$$\{\liminf_{n\to\infty} X_n \leq x\} = \cap_{k\geq 1}\underbrace{(\cup_{n\geq k}\underbrace{\{X_n \leq x\}}_{\in\mathcal{F}_n})}_{\in\mathcal{F}_k} \in \cap_{k\geq 1}\mathcal{F}_k.$$

Whence, by Kolmogorov's 0-1 law, and since $\Phi(M) > 0$, we have that $\mathbb{P}(\{\liminf_{n\to\infty} \frac{S_n}{\sqrt{n}} = -\infty\}) = 1$ for all M. Consequently,

$$\liminf_{n\to\infty} \frac{S_n}{\sqrt{n}} = -\infty \quad \text{a.s.}$$

and

$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{n}} = -\underbrace{\liminf_{n\to\infty} -\frac{S_n}{\sqrt{n}}}_{\to -\infty \text{ a.s.}} = +\infty \quad \text{a.s.}$$

(convergence of the lim inf to $-\infty$ because $-X_1, -X_2, \ldots$ have normal distribution and are independent).

11 Solutions - Sheet 11

11.1 Ex 11.1

Define $\mathcal{F}_n = \sigma(X_i,\ i \geq n)$ then $\mathcal{F}_{n+1} \subseteq \mathcal{F}_n$ for all n. Define

$$\mathcal{G} := \bigcap_{n\geq 0} \mathcal{F}_n$$

as the tail σ-algebra.

By Kolmogorov's theorem: $A \in \mathcal{G} \Rightarrow \mathbb{P}(A) \in \{0,1\}$.

Let's start working in $\overline{\mathbb{R}}$ and consider

$$\limsup_{n\to\infty} X_n = \lim_{n\to\infty} \sup_{k\geq n} X_k.$$

$\sup_{k\geq n} X_k$ is $\mathcal{F}_{n_0}$-measurable, for $n \geq n_0$ with n_0 fixed. As a consequence $\limsup X_n$ is $\mathcal{F}_{n_0}$-measurable, for any $n_0 \in \mathcal{N}$. Whence $\limsup X_n$ is $\mathcal{G}$-measurable. Note that the exact same procedure/proof applies to $\liminf X_n$.

For $A \in \mathcal{B}(\overline{\mathbb{R}})$ we define $\Theta := \{\omega : \lim X_n \text{ exists in } A\}$. Then

$$\Theta = \{\liminf X_n = \limsup X_n\} \cap \{\limsup X_n \in A\} \cap \{\liminf X_n \in A\}.$$

The information provided by the last set is superfluous (already contained in the first two). But so $\Theta \in \mathcal{G}$, and we're in business for Kolmogorov's tail theorem!

- (a) Just take $A = \mathbb{R}$ or $A = \overline{\mathbb{R}}$ in the reasoning above. The next two points constitute an adaptation of the reasoning above; we shall use Θ again as a name for the considered sets.
- (b) Notice that for $A = \mathbb{R}$ or $A = \overline{\mathbb{R}}$ (not for any A, because of shifts),

  $$\Theta := \Big\{ \lim_{n\to\infty} \sum_{k=1}^{n} X_k \text{ exists in } A \Big\} = \Big\{ \lim_{n\to\infty} \sum_{k=k_0}^{n} X_k \text{ exists in } A \Big\}$$

  for any k_0. The first set in the equality is in $\mathcal{F}_1$ but the reformulated version is in $\mathcal{F}_{k_0}$. As a consequence, $\Theta \in \mathcal{F}_{k_0}$ for any $k_0 \in \mathbb{N}$. Thus $\Theta \in \mathcal{G}$. We are done.
- (c) Notice that for k_0 fixed, $\frac{1}{n}\sum_{i=1}^{k_0-1} X_k \overset{n\to\infty}{\longrightarrow} 0$. As a consequence, for $A \in \mathcal{B}(\overline{\mathbb{R}})$:

  $$\Theta := \Big\{ \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} X_k \text{ exists in } A \Big\} = \Big\{ \lim_{n\to\infty} \frac{1}{n} \sum_{k=k_0}^{n} X_k \text{ exists in } A \Big\}.$$

  The right-hand side is an element of $\mathcal{F}_{k_0}$, whence $\Theta \in \mathcal{F}_{k_0}$ for all $k_0 \in \mathbb{N}$. Thus $\Theta \in \mathcal{G}$. We are done.

11.2 Ex 11.2

Preliminary remarks:

- for $p > 1$ we get $\mathbb{E}\Big[\sum_{n=1}^{\infty} \frac{|X_n|}{n^p}\Big] = \sum_{n=1}^{\infty} \frac{1}{n^p} < \infty$,
- for $p = 0$, the series jumps constantly by -1 or $+1$, so divergence,
- remains the case $0 < p \leq 1$. It's not easy "just like that". But we have an atomic bomb the 3 series theorem of Kolmogorov!

We thus treat the last case, where $0 < p \le 1$.

- Choose $K = 1$, then $\sum_{n=1}^{\infty} \mathbb{P}\left[\left|\frac{X_n}{n^p}\right| > 1\right] = 0$.
- $Y_n = \frac{X_n}{n^p} \mathbb{1}_{\{|\frac{X_p}{n^p}| \le 1\}} = \frac{X_n}{n^p}$ thus $\sum_{n=1}^{\infty} \underbrace{\mathbb{E}[Y_n]}_{=0} = 0$.
- $\mathbb{E}[|Y_n|^2] = \mathbb{E}[\left(\frac{X_n}{n^p}\right)^2] = \frac{\mathbb{E}[X_n^2]}{n^{2p}} = \frac{1}{n^{2p}}$. But $\sum \frac{1}{n^{2p}} < \infty \iff p > \frac{1}{2}$.

So as a conclusion, $\sum_{n=1}^{\infty} \frac{X_n}{n^p}$ converges a.s. if and only if $p > \frac{1}{2}$. When it does **not** converge a.s., then $\{\sum_{k=1}^{n} \frac{X_k}{n^p}$ exists in $A \subseteq \overline{\mathbb{R}}\}$ (taking here $A = \mathbb{R}$ is in the tail σ-algebra $\bigcap_{n \ge 1} \sigma(X_k,\ k \ge n)$. We conlude then that $\sum_{k=1}^{\infty} \frac{X_k}{n^p}$ converges nowhere, a.s. □

11.3 Ex 11.3

Point (a). Convention (notational): $X \sim \mathrm{Law}(X_1)$.

$$
\begin{aligned}
\infty = \mathbb{E}[|X|] &= \int_0^{\infty} \mathbb{P}(\{|X| > t\})\, dt \\
&\le \int_0^{\infty} \mathbb{P}(\{|X| \ge t\})\, dt \\
&\le \sum_{n \in \mathbb{N}} \mathbb{P}(\{|X| \ge n\}) \\
&= \sum_{n \in \mathbb{N}} \mathbb{P}(\{|X| \ge n\})
\end{aligned}
$$

where in the last inequality we used that the function $t \mapsto \mathbb{P}(\{|X| \ge t\})$ is decreasing and in the last equality that the X_i are i.i.d. and have the same law as X. We also used exercise 8.2 in the first line.

By Borel-Cantelli II, $\mathbb{P}(\limsup_{n\to\infty}\{|X_n| \ge n\}) = 1$. In other wrods $|X_n| \ge n$ infinitely many times. But $S_n = \frac{S_n(n-1)}{n} + \frac{X_n}{n}$. If $S_n \to a \in \mathbb{R}$, then $\frac{X_n}{n} \overset{n\to\infty}{\longrightarrow} 0$. But this is not true as $\left|\frac{X_n}{n}\right| \ge 1$ infinitely many times. A contradiction. □

Point (b). Notation: $X_i = X_i^+ - X_i^-$ in the usual sense and $S_n^+ := \sum_{i=0}^{n} X_i^+,\ S_n^- := \sum_{i=0}^{n} X_i^-$ (so we don't exactly follow the usual convention for S_n^+ and S_n^-). So $\frac{S_n}{n} = \frac{S_n^+}{n} - \frac{S_n^-}{n}$. By the law of large numbers, $\frac{S_n^-}{n} \overset{a.s.}{\to} \mathbb{E}[X_1^-]$ as $n \to \infty$. Choose $K > 0$, then

$$
\begin{aligned}
\lim_{n\to\infty} \frac{S_n^+}{n} &\ge \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} X_k^+ \mathbb{1}_{\{X_k^+ \le K\}} \\
&\overset{a.s.}{=} \mathbb{E}\Big[X_1^+ \mathbb{1}_{\{X_1^+ \le K\}}\Big] \\
&\overset{K\to\infty}{\longrightarrow} \mathbb{E}[X_1^+] \quad \text{(dominated convergence)} \\
&= 0.
\end{aligned}
$$

Consequently, $\lim_{n\to\infty} \frac{S_n^+}{n} = \infty$, a.s. □

11.4 Ex 11.4

(Note: by extension, $I = (a, b)$ for all $a, b \in \mathbb{R},\ a < b$).

F increasing: $x \le y \Rightarrow F(x) \le F(y)$. Suppose x_0 is a point of discontinuity, then ($\iff$) $\lim_{x\downarrow x_0} F(x) = F(x^+) > F(x^-) = \lim_{x\uparrow x_0}$, whence there is a number $q_{x_0} \in \mathbb{Q}$ such that $F(x^-) < q < F(x^-)$. Let $\mathcal{D} = \{x_0 \in I \mid x_0$ a point of discontinuity of $F\}$. Then the map $\alpha : \mathcal{D} \to \mathbb{Q} :: x_0 \mapsto q_{x_0}$ is injective. Whence $\mathrm{Card}(\mathcal{D}) \le \mathrm{Card}(\mathbb{Q})$, so $\mathcal{D}$ is finite or infinite countable.

11.5 Ex 11.5

$S_n := \sum_{k=1}^n X_k$. We have $\mathbb{E}[S_n^2] = \mathbb{E}[\sum_{i=1}^n X_i^2] \geq 1$. In particular, since $(X_i)_{i=1}^n$ are independent, $X_i \neq 0$.

Taking into account that terms with odd powers cancel out and the fact that $|X_i| \leq 1$, we have that

$$\begin{aligned}\mathbb{E}[|S_n|^4] &= C_2^4 \sum_{1\leq i<j\leq n} \mathbb{E}[X_i^2] \cdot \mathbb{E}[X_j^2] + \sum_{i=1}^n \mathbb{E}[X_i^4] \\ &\leq C_2^4 \sum_{1\leq i<j\leq n} \mathbb{E}[X_i^2] \cdot \mathbb{E}[X_j^2] + \sum_{i=1}^n \mathbb{E}[X_i^2].\end{aligned}$$

We bound the first term:

$$\begin{aligned}C_2^4 \sum_{1\leq i<j\leq n} \mathbb{E}[X_i^2] \cdot \mathbb{E}[X_j^2] &\leq C_2^4 \sum_{1\leq i,j\leq n} \mathbb{E}[X_i^2] \cdot \mathbb{E}[X_j^2] \\ &= C_2^4 \Big(\sum_i \mathbb{E}[X_i^2]\Big) \cdot \Big(\sum_j \mathbb{E}[X_j^2\Big) \\ &= C_2^4 \mathbb{E}[S_n^2]\mathbb{E}[S_n^2] \\ &= C_2^4 \mathbb{E}[S_n^2]^2.\end{aligned}$$

As a consequence, $\mathbb{E}[|S_n|^4] \overset{\spadesuit}{\leq} \underbrace{(C_2^4 + 1)}_{7} \cdot \mathbb{E}[S_n^2]^2$.

Now we have

$$\begin{aligned}\mathbb{P}(|S_n|^2 \geq \frac{1}{2}\mathbb{E}[S_n^2]) \cdot \mathbb{E}[S_n^4] \\ \overset{\spadesuit}{\geq} \frac{1}{4} \cdot \frac{1}{7} \cdot \mathbb{E}[|S_n|^4],\end{aligned}$$

using Paley-Zygmund for the first inequality. As a consequence $\mathbb{P}(|S_n^2| \geq \frac{1}{2}) \geq \frac{1}{28}$, whence $\mathbb{P}(|S_n^2| \geq \frac{1}{28}) \geq \frac{1}{28}$ (i.e. taking $\epsilon = \frac{1}{28}$). $\square$

11.6 Ex 11.6

No solution provided. If any of you solved this technical exercise cleanly, I would be glad if you shared your solution with the class.

12 Solutions - Sheet 12

12.1 Ex 12.1

Suppose $(X_n)_n$ is not tight, then there is an $\epsilon > 0$ such that for all $K > 0$: $\mathbb{P}(\{|X_n| \geq K\}) \geq \epsilon$ for some n. Whence there exists $(X_{n_k})_{k\in\mathbb{N}}$ such that $\mathbb{P}(\{|X_{n_k}| \geq K\}) \geq \epsilon$ (notice k is varying).

We now claim that $(X_{n_k})_k$ has no subsequence that converges in law. Indeed, $\forall\ \lambda \geq 0$ and for any subsequence $X_{n_{k_j}}$, $\mathbb{P}(\{|X_{n_{k_j}}| \geq \lambda\}) \geq \epsilon$ for j large enough.

But then by the Portmanteau theorem, if $X_{n_{k_j}} \to X$, we would have that $\mathbb{P}(\{|X| \geq \lambda\}) \geq \epsilon$, for all $\lambda \geq 0$. But this is absurd: for any random variable $X : \Omega \to \mathbb{R}$, we have $\lim_{\lambda\to\infty} \mathbb{P}(\{|X| \geq \lambda\}) = 0$ (since $\cap_{n\in\mathbb{N}}\{|X| \geq n\} = \emptyset$).

Important comment (for your future study of probability theory):

In the context of convergence in law, let S be a polish space (separable complete metric space). Let $\mathcal{P}(S) = \{\mu \mid \mu \text{ a probability measure on } S\}$. Then one can endow $\mathcal{P}(S)$ with a metric d such that $d(X_n, X) \overset{n\to\infty}{\longrightarrow} 0 \iff X_n \overset{(\mathcal{L})}{\longrightarrow} X$. (we model convergence in law). In this "abstract set-up", what we have proved (which is called Prokhorov's theorem) reads:

Let $\mathcal{U} \subseteq \mathcal{P}(S)$, then $\mathcal{U}$ is tight $\iff$ $\mathcal{U}$ is relatively compact in $\mathcal{P}(S)$ ($\overline{\mathcal{U}}$ compact in $\mathcal{P}(S)$).

12.2 Ex 12.2

I shall use the lecture's numbering (points a to e). There, we showed that $(a) \iff (b) \iff (c)$. We shall prove that $(c) \iff (d)$, that $\big((c) \iff (d)\big) \Rightarrow (e)$ and that $(e) \Rightarrow (b)$, completing the proof of the Portmanteau theorem.

For the first equivalence, $(c) \iff (d)$, we have:

$$\begin{aligned}
&\liminf_{n\to\infty} \mathbb{P}(\{X_n \in U\}) \geq \mathbb{P}(\{X \in U\})\ \forall\, U \subseteq \mathbb{R} \text{ open} \\
&\iff \liminf_{n\to\infty} 1 - \mathbb{P}(\{X_n \in F\}) \geq 1 - \mathbb{P}(\{X \in F\})\ \forall\, F \subseteq \mathbb{R} \text{ closed} \\
&\iff \liminf_{n\to\infty} -\mathbb{P}(\{X_n \in f\}) \geq -\mathbb{P}(\{X \in F\})\ \forall\, F \subseteq \mathbb{R} \text{ closed} \\
&\iff -\limsup_{n\to\infty} \mathbb{P}(\{X_n \in f\}) \geq -\mathbb{P}(\{X \in F\})\ \forall\, F \subseteq \mathbb{R} \text{ closed.} \quad \square
\end{aligned}$$

Let's turn our attention to the second implication: $\big((c) \iff (d)\big) \Rightarrow (e)$. We recall that a set A is called a "continuity set" if $\mathbb{P}(\partial A) = 0$.

Let A be a continuity set, then $\mathring{A} \subseteq A \subseteq \overline{A}$ and $\mathbb{P}(\{X \in \mathring{A}\}) = 0$. By (c), we have

$$\liminf_{n\to\infty} \mathbb{P}(\{X_n \in \mathring{A}\}) \geq \mathbb{P}(\{X \in \mathring{A}\}) \iff \liminf_{n\to\infty} \mathbb{P}(\{X_n \in A\}) \geq \mathbb{P}(\{X \in \mathring{A}\}),$$

and by (d):

$$\limsup_{n\to\infty} \mathbb{P}(\{X_n \in \overline{A}\}) \leq \mathbb{P}(\{X \in \overline{A}\}) \iff \lim_{n\to\infty} \mathbb{P}(\{X_n \in A\}) \leq \mathbb{P}(X \in A).$$

Whence $\lim_{n\to\infty} \mathbb{P}(\{X_n \in A\}) = \mathbb{P}(\{X \in A\})$. $\square$

Now we finally prove that $(e) \Rightarrow (b)$. Recall from exercise 8.2 that for $X : \Omega \to \mathbb{R}_+$ a non-negative random variable, we have that

$$\mathbb{E}[X] = \int_0^\infty \mathbb{P}(\{X > t\})\, dt. \quad \heartsuit$$

We claim this equality is a special case of the following:

Proposition: Let μ be sigma-finite on $\mathbb{R}$ and $g : (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu) \to \overline{R}_+$ be measurable, then

$$\int_{\mathbb{R}} g \, d\mu = \int_0^\infty \mu(\{g > u\}) \, d\mu,$$

where μ is the Lebesgue measure.

Indeed, to retrieve ♡, take $g(x) = x$ and $\mu = X_*\mathbb{P}$; then our claim reads:

$$\begin{aligned}
\mathbb{E}[X] &= \int x \underbrace{\mu(dx)}_{X_*\mathbb{P}} \\
&= \int_{\mathbb{R}} g(x) \, \mu(dx) \\
&= \int_{\mathbb{R}} \mu(\{g > u\}) \, du,
\end{aligned}$$

and $\mu(\{g > u\}) = (X_*\mathbb{P})(\{g > u\}) = \mathbb{P}(\{X > u\})$. So this is indeed the relation displayed above.

We now set out to show our claim holds true.

Proof. We shall work on $\big(\mathbb{R} \times \overline{\mathbb{R}}_+, \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\overline{\mathbb{R}}_+), \mu \otimes \overline{\mu}\big)$, where $\overline{\mu}$ is the Lebesgue measure on $\overline{\mathbb{R}}$. Note the product measure is sigma-finite, which is a prerequisite to using Fubini's theorem. Let $A = \{(x, u) \in \mathbb{R} \times \overline{\mathbb{R}}_+$ such that $f(x) > u\}$. Then A can be rewritten as

$$A = \bigcup_{r \in \mathbb{Q}_+^*} \big(\{f > r\} \times [0, r]\big) \in \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\overline{\mathbb{R}}_+).$$

We have by Fubini that

$$\begin{aligned}
(\mu \otimes \overline{\mu})(A) &= \int_{\overline{R}_+} \mu(A_u) \, \overline{\mu}(du) \\
&= \int_{\mathbb{R}} \overline{\mu}(A_x) \, \mu(dx).
\end{aligned}$$

The last integral is just $\int_{\mathbb{R}} f(x) \, \mu(dx)$. On the other hand,

$$\int_{\overline{R}_+} \mu(A_u) \, \overline{\mu}(du) = \int_{\overline{\mathbb{R}}_+} \mu(\{f > u\}) \, \overline{\mu}(du) \overset{\overline{\mu}(\{-\infty, +\infty\})=0}{=} \int_{\mathbb{R}_+} \mu(\{f > u\}) \, \mu(du).$$

We are done. □

It remains to prove point (b). Let $h : \mathbb{R} \to \mathbb{R}_+$ be bounded continuous. We have

$$\int h(X_n) \, d\mathbb{P} = \int_{\mathbb{R}} (X_n)_*\mathbb{P}(\{h > u\}) \, \mu(du) = \int_0^{\|h\|_{\sup}} \big((X_n)_*\mathbb{P}\big)(h > u) \, \mu(du) \spadesuit.$$

h is continuous, so we have:

$$\overline{\{h > u\}} \subseteq \{h \geq u\} \Rightarrow \partial\{h > u\} \subseteq \{h = u\}.$$

Define

$$\mathcal{D} := \Big\{u \geq 0 : \; X_*\mathbb{P}(\{h = u\}) > 0\Big\} = \bigcup_{n \geq 1} \Big\{n \geq 0 : \; X_*\mathbb{P}(\{h = u\}) \geq \frac{1}{2}\Big\}.$$

Each set in the union is finite because the sets are disjoint 2 by 2. So $\mathcal{D}$ is countable. Whence $\mu(\mathcal{D}) = 0$. So from (e):

$$(X_n)_*\mathbb{P}(\{h > u\}) \overset{n\to\infty}{\longrightarrow} X_*\mathbb{P}(\{h > u\}),\ \mu\text{-a.s.},$$

whence by dominated convergence:

$$\spadesuit \overset{n\to\infty}{\longrightarrow} \int_0^{\|h\|_{\sup}} \mu(\{h > u\})\ \mu(du) = \int_{\mathbb{R}} h(u)\ d\mu.$$

Now for a general h, split h into $h = h_+ - h_-$. □

12.3 Ex 12.3

This exercise is called "Slutsky's theorem" in the literature and is an important result for statistics.

We have to show that

$$Y_n \overset{(\mathcal{L})}{\longrightarrow} c \Rightarrow Y_n \overset{(\mathbb{P})}{\longrightarrow} c.$$

The converse implication is, of course, true. We recall the following definition:

$$\mu_n \overset{(\mathcal{L})}{\to} \mu \text{ if for any bounded function } h : \mathbb{R} \to \mathbb{R} \text{ one has } \lim_{n\to\infty} \int h(x)\ d\mu_n(x) = \int h(x)\ d\mu(x).$$

Warning: In order for $Y_n \overset{(\mathbb{P})}{\longrightarrow} c$, we need all Y_n to be modelled on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$, or else this would not make sense (write convergence explicitly). For convergence in law, this assumption is not necessary, we are really just looking at the push-forward measure (the laws of the random variables at hand).

(a) We have, for h continuous bounded, that

$$\int h(Y_n)\ d\mathbb{P} \overset{n\to\infty}{\longrightarrow} \int h(c)\ d\mathbb{P}$$

(Convergence in law), and we should show that

$$\text{for all } \epsilon > 0 : \quad \mathbb{P}(\{|Y_n - c| \geq \epsilon\}) \overset{n\to\infty}{\longrightarrow} 0.$$

We compute (for a fixed $\epsilon > 0$):

$$\begin{aligned} 0 &\leq \mathbb{P}(\{|Y_n - c| \geq \epsilon\}) \\ &= \mathbb{E}[\mathbb{1}_{\{|Y_n - c| \geq \epsilon\}}] \\ &\leq \mathbb{E}[\alpha_\epsilon(Y_n)] \\ &\overset{n\to\infty}{\longrightarrow} \mathbb{E}[\alpha_\epsilon(0)] \\ &= 0, \end{aligned}$$

where the function $\alpha_\epsilon \geq \mathbb{1}_{\{|Y_n - \epsilon| \geq \epsilon\}}$ with $\alpha_\epsilon(x) \in [0,1]\ \forall\, x$ is defined as:

$$\alpha_\epsilon(x) := \begin{cases} 1 & \text{for } x \in (c-\epsilon, c+\epsilon)^c \\ 0 & \text{for } x = c \\ \in [0,1] & \text{elsewhere, so that the function is continuous.} \end{cases}$$

We are done.

(b) (X_n) is tight (since it converges in law to X): for any $\epsilon > 0$ there a K_ϵ such that $\mathbb{P}(\{|X_n| \geq K_\epsilon\}) \leq \epsilon$ for all $n \in \mathbb{N}$. $Y_n \xrightarrow{(\mathbb{P})} c$ by point (b) (i.e. for any $\epsilon > 0$: $\mathbb{P}(\{|Y_n - c| \geq \epsilon\}) \to 0$ as $n \to \infty$). We write

$$\Big|\int h(X_n, Y_n) - h(X, c)\, d\mathbb{P}\Big| \leq \underbrace{\Big|\int h(X_n, Y_n) - h(X_n, c)\, d\mathbb{P}\Big|}_{\alpha} + \underbrace{\Big|\int h(X_n, c) - h(X, c)\, d\mathbb{P}\Big|}_{\beta}$$

and deal with each term separately.

- As for β, we note that $h(\cdot, c)$ is continuous bounded; since $X_n \xrightarrow{(\mathcal{L})} X$, we have that $\beta \to 0$.
- The reasoning for α is a bit more involved. Decompose $\mathbb{R} \times \mathbb{R} = A \sqcup B \sqcup B$ as follows:
  - $A = K_\epsilon \times [c - \tilde{\epsilon}, c + \tilde{\epsilon}]$,
  - $B = K_\epsilon \times [c - \tilde{\epsilon}, c + \tilde{\epsilon}]^c$,
  - $C = K_\epsilon^c \times \mathbb{R}$,

  for some $\epsilon > 0$ and $\tilde{\epsilon} > 0$. Now decompose

$$\spadesuit = \mathbb{E}[\underbrace{|h(X_n, Y_n) - h(X_n, c)|}_{=:\aleph}]$$

  into three parts:

$$\mathbb{E}[|h(X_n, Y_n) - h(X_n, c)|] = \underbrace{\mathbb{E}[\aleph \mathbb{1}_A]}_{U} + \underbrace{\mathbb{E}[\aleph \mathbb{1}_B]}_{V} + \underbrace{\mathbb{E}[\aleph \mathbb{1}_C]}_{W}.$$

  - For fixed $\epsilon > 0$, using uniform continuity of (the continuous function) h (on the compact $K_\epsilon \times [c - \tilde{\epsilon}, c + \tilde{\epsilon}]$), we have that $U \to 0$ as $\tilde{\epsilon} \to 0$.
  - $V \leq \|h\|_\infty \mathbb{P}(\{|Y_n - c| > \tilde{\epsilon}\})$
  - $W \leq 2\epsilon \|h\|_\infty$.

  Putting all the bricks together yields

$$\spadesuit \leq U + 2\|h\|_\infty (\epsilon + \mathbb{P}(\{|Y_n - c| > \tilde{\epsilon}\})) \xrightarrow{\epsilon, \tilde{\epsilon}} 0$$

  (where we first take $\tilde{\epsilon}$ to 0 and then ϵ). We are done, $\alpha \to 0$. □

(c) We know that $(X_n, Y_n) \xrightarrow{(d)} (X, c)$ by (b). We're only interested in the distributional convergence of $X_n + Y_n$ and $X_n Y_n$, which is independent from the random variables "carrying" (by means of pushing the measures forward) the distributions at hand (!). Indeed convergence in law $X_n \xrightarrow{(\mathcal{L})} X$ reads

$$\int_{\mathbb{R}} f \, d((X_n)_* \mathbb{P}) \longrightarrow \int_{\mathbb{R}} f \, d(X_* \mathbb{P}) \quad \text{for all continous bounded function } f,$$

which only makes the laws $(X_n)_* \mathbb{P}$ and $X_* \mathbb{P}$ intervene. Let's choose $(X_n, Y_n)_{n \in \mathbb{N}}$ and (X, Y) in Skorokhod's representation theorem for vectors. Then $(X_n, Y_n) \xrightarrow{a.s.} (X, c)$ on $\mathbb{R} \times \mathbb{R}$ (and thus in law). But then (by continuity of + and *), we have $X_n Y_n \xrightarrow{a.s.} Xc$ and $X_n + Y_n \xrightarrow{a.s.} X + c$. We know that

$$\text{a.s. convergence} \Rightarrow \text{convergence in probability} \Rightarrow \text{convergence in law},$$

so we are done! □

12.4 Ex 12.4

What is meant is: let μ and ν be two σ-finite measures on $\mathbb{R}$ such that for all $a < b$:

$$\mu((a,b)) + \frac{1}{2}\mu(\{a,b\}) = \nu((a,b)) + \frac{1}{2}\nu(\{a,b\}),$$

then $\mu = \nu$.

Claim: a sigma-finite measure μ on $\mathbb{R}$ can only have a finite number of atoms of positive mass.

Proof. Suppose it is not so. We have a growing sequence of sets $A_n \subseteq A_{n_1}$ with $\cup_{n\in\mathbb{N}} A_n = \mathbb{R}$ and $\mu(A_n) < \infty$ by σ-finiteness. Denote $\mathcal{D} = \{x \in \mathbb{R} \mid \mu(x) > 0\}$. Suppose $\text{Card}(\mathcal{D}) > \text{Card}(\mathbb{N})$, then there is an A_n such that $\text{Card}(\mathcal{D} \cap A_n) > \text{Card}(\mathbb{N})$ (or else $\text{Card}(\mathcal{D}) = \text{Card}(\mathbb{N})$. But then $\infty > \mu(A) > \mu(\mathcal{D} \cap A_n) = \infty$ (the last equality by exercise 1.6, a contradiction. □

Now for all $a,\ b \notin \mathcal{D}$, we have $\mu((a,b)) = \nu((a,b))$ and $\overline{\mathbb{R}\backslash\mathcal{D}} = \mathbb{R}$, since $\mathcal{D}$ is countable. So one can choose $a_n \downarrow a$ with $a_n \in \mathbb{R}\backslash\mathbb{D}$ and $b_n \uparrow b$ with $b_n \in \mathbb{R}\backslash\mathbb{D}$ and then

$$\mu((a,b)) \overset{n\to\infty}{\longleftarrow} \mu(a_n, b_n) = \nu(a_n, b_n) \overset{n\to\infty}{\longrightarrow} \nu((a,b)).$$

□

12.5 Ex 12.5

This is called the "Dirichlet Integral". You can find five methods of proof on Wikipedia: `https://en.wikipedia.org/wiki/Dirichlet_integral`. These five methods are: the Laplace transform, double integration, differentiation under the integral sign (Feynman's trick), complex integration and the Dirichlet kernel. I suggest you read the proof based on complex integration.

13 Solutions - Sheet 13

13.1 Ex 13.1

- (a) $\varphi_X(0) = \mathbb{E}[\underbrace{e^{i\cdot 0\cdot X}}_{1}] = 1$.
- (b) $\varphi_{aX+b}(t) = \mathbb{E}[e^{it(aX+b)}] = e^{ibt}\mathbb{E}[e^{i(taX)}] = e^{ibt}\varphi_X(at)$.
- (c) $|\varphi_X(t)| = |\mathbb{E}[e^{itX}]| \leq \mathbb{E}[\underbrace{|e^{itX}|}_{1}] = 1$ for all $t \in \mathbb{R}$.
- (d) $\varphi_{-X}(t) = \mathbb{E}[e^{it(-X)}] = \mathbb{E}[\overline{e^{itX}}] = \overline{\mathbb{E}[e^{itX}]} = \overline{\varphi_X(t)}$. $\varphi_X(t) \in \mathbb{R}\ \forall\ t \in \mathbb{R} \iff \varphi_{-X}(t) = \varphi_X(t)\ \forall\ t \in \mathbb{R}$. And by using the fact we are dealing with characteristic functions, this is true if and only if $X \overset{(\mathcal{L})}{=} -X$.
- (e) $|\varphi_X(t+s) - \varphi_X(t)| \leq \mathbb{E}[\underbrace{|e^{i(t+s)X} - e^{itX}|}_{= \underbrace{|e^{itX}|}_{=1} \cdot |e^{isX}-1|}] = \mathbb{E}[\underbrace{|e^{isX} - 1|}_{\leq 2}] \to 0$, a.s. as $s \to 0$ by dominated convergence, independently of the chosen t! Thus we have uniform continuity.

13.2 Ex 13.2

(a) We calculate

$$\begin{aligned}\varphi_{X+Y}(t) &= \mathbb{E}[e^{it(X+Y}] \\ &= \mathbb{E}[e^{itX}e^{itY}] \\ &= \mathbb{E}[e^{itX}] \cdot \mathbb{E}[e^{itY}] \\ &= \varphi_X(t) \cdot \varphi_Y(t),\end{aligned}$$

using independence in the second equality.

(b) We start with a few measure-theoretic considerations. First, we recall that if X and Y are independent, then $(X,Y)_*\mathbb{P} = \mu_X \otimes \mu_Y$. Define

$$\mu_X * \mu_Y := (+)_*(\mu_X \otimes \mu_Y),$$

which is a measure on $\mathbb{R}$. ($+ : \mathbb{R} \to \mathbb{R} :: (x,y) \mapsto x+y$ is continuous, thus measurable). Then

$$\mu_X * \mu_Y(A) = \mathbb{E}_{\mu_X * \mu_Y}[\mathbb{1}_A] = \int_{\mathbb{R}^2} \mathbb{1}_A(x+y)\ \mu_X(dx) \otimes \mu_Y(dy).$$

Similarly,

$$\mu_X * \mu_Y(f) = \mathbb{E}_{\mu_X * \mu_Y}[f] = \int_{\mathbb{R}^2} f(x+y)\ \mu_X(dx) \otimes \mu_Y(dy).$$

Now to our problem. Let us calculate the characteristic function of $\mu_X * \mu_Y$:

$$\begin{aligned}\varphi_{\mu_X * \mu_Y}(t) &= \mathbb{E}_{\mu_X * \mu_Y}[e^{itz}] \\ &\overset{z=x+y}{=} \int_{\mathbb{R}^2} e^{it(x+y)}\mu_X(dx)\mu_Y(dy) \\ &\overset{Fubini}{=} \int_{\mathbb{R}} e^{itx}\ \mu_X(dx) \int_{\mathbb{R}} e^{ity}\ \mu_Y(dy) \\ &= \varphi_X(t) \cdot \varphi_Y(t) \\ &\overset{(a)}{=} \varphi_{X+Y}(t).\end{aligned}$$

So the law of $X+Y$ ist $\mu_X * \mu_Y$, by injectivity of the characteristic function on the set of probability measures.

(c)

$$\begin{aligned}
\mu_{X+Y}(A) &\overset{(b)}{=} \mu_X * \mu_Y(A) \\
&= \int_{\mathbb{R}^2} \mathbb{1}_A(x+y)\ \underbrace{\mu_X(dx) \otimes \mu_Y(dy)}_{p_X \cdot p_Y\ \mu\otimes\mu} \\
&= \int_{\mathbb{R}^2} \mathbb{1}_A(t)\ p_X(s) \cdot p_Y(t-s)\ \mu(ds) \otimes \mu(dt) \\
&\overset{Fubini}{=} \int_{\mathbb{R}} \mathbb{1}_A(t) \big(\int_{\mathbb{R}} p_X(s) \cdot p_Y(t-s)\ \mu(ds) \big)\ \mu(dt) \\
&= \int_{\mathbb{R}} \mathbb{1}_A(t) \big(\int_{\mathbb{R}} p_X(t-\tilde{s})\ p_Y(\tilde{s})\ \mu(d\tilde{s}) \big)\ \mu(dt),
\end{aligned}$$

where we denoted the Lebesgue measure by μ. In the third last equality we made the following change of variables: $x+y=t$ & $x=s$, which has determinant 1. In the very last equality and for fixed t, we made the change of variables $t-s=\tilde{s}$, which also has determinant 1. □

13.3 Ex 13.3

We shall deal with (b) first in order to recycle the result in (a). Other approaches are possible.

Part (b). Define

$$A_n := \{S_n > m\sqrt{n}\}$$

for some fixed number $m \in \mathbb{R}$. Then by the Portmanteau theorem,

$$\mathbb{P}(A_n) = \mathbb{P}\Big(\Big\{ \frac{S_n}{\sqrt{n}} \geq m \Big\} \Big) \overset{n\to\infty}{\longrightarrow} \int_m^\infty \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\ dx > 0\ \spadesuit,$$

since $\mu(\{y :\ \mathbb{1}_{\{x\geq m\}}(y) \text{ is discontinuous}\}) = 0$, with μ the Lebesgue measure on $\mathbb{R}$. Now

$$\begin{aligned}
\mathbb{P}(\limsup A_n) &= \mathbb{P}(\cap_{n\geq 1} \cup_{j\geq n} A_j) \\
&= \lim \downarrow \mathbb{P}(\cup_{j\geq n} A_j) \\
&\geq \limsup_{n\to\infty} \underbrace{\mathbb{P}(A_n)}_{\geq \sup_{j\geq n} \mathbb{P}(A_j)} \\
&\overset{\spadesuit}{>} 0.
\end{aligned}$$

Note: we also have that, for any $k \in \mathbb{N}$:

$$\big\{\omega : S_n > m\sqrt{n}\ \infty\text{-often},\ n \in \mathbb{N}\big\} = \big\{\omega : S_n > m\sqrt{n}\ \infty\text{-often},\ n \in \{k, k+1, k+2, \ldots\}\big\}.$$

In other words, rephrasing the calculation above: $\limsup A_n \in \mathcal{G}_k$ for all $k \in \mathbb{N}$, where $\mathcal{G}_k = \sigma(X_k, X_{k+1}, \ldots)$. Thus $\limsup A_n \in \cap_{n\in\mathbb{N}} \mathcal{G}_k$ - the tail sigma-algebra. As a consequence, by Kolmogorov, $\mathbb{P}(\limsup A_n) = 0$ or 1. But $\mathbb{P}(\limsup A_n) > 0$, so it is 1! We have proven $\limsup \frac{S_n}{\sqrt{n}} = +\infty$. Now we have that also $\limsup \frac{\tilde{S}_n}{\sqrt{n}} = \infty$ a.s., where $\tilde{S}_n = -X_1 - X_2 - \ldots - X_n$. Thus $-\liminf \frac{S_n}{\sqrt{n}} = \infty$, which is equivalent to $\liminf \frac{S_n}{\sqrt{n}} = -\infty$ a.s. □

Part (a). If $\frac{S_n}{\sqrt{n}} \to \mathbb{P}$, (then by the central limit theorem, $X \sim \mathcal{N}(0,1)$, since convergence in probability implies convergence in law). Then $\exists$ a sub-sequence n_k such that $\frac{S_{n_k}}{\sqrt{n_k}} \overset{a.s.}{\longrightarrow} X$. But applying part (b) to $(S_{n_k})_{k\in\mathbb{N}}$ yields a contradiction: indeed, $\frac{S_{n_k}}{\sqrt{n_k}}$ cannot converge to any random variable.

13.4 Ex 13.4

We proceed by induction (or iteration). $|a_1|,\ |a_2|\ |b_1|\ |b_2| \le 1$ for $a_1,\ a_2,\ b_1,\ b_2 \in \mathbf{C}$. Then

$$\begin{aligned}|a_1a_2 - b_1b_2| &= |(a_1-b_1)a_2 + \underbrace{b_1a_2 - b_1b_2}_{b_1(a_2-b_2)}| \\ &\le |a_1-b_1|\cdot|a_1| + |b_1|\cdot|a_1-b_2| \\ &\le |a_1-b_1| + |a_2-b_2|.\end{aligned}$$

Now for $n \ge 2$

$$\begin{aligned}|\Pi_{i=1}^n a_i - \Pi_{i=1}^n b_i| &= |a_n \cdot \underbrace{\Pi_{i=1}^{n-1} a_i}_{|\cdot|\le 1} - b_n \cdot \underbrace{\Pi_{i=1}^{n-1} a_i}_{|\cdot|\le 1}| \\ &\le |a_n - b_n| + |\Pi_{i=1}^{n-1} a_i - \Pi_{i=1}^{n-1} b_i| \\ &\overset{\text{induction}}{\le} |a_n - b_n| + \sum_{i=1}^{n-1} |a_i - b_i|.\end{aligned}$$

13.5 Ex 13.5

(a)

$$\begin{aligned}\varphi_{X_n}(t) &= \mathbb{E}[e^{itX_n}] \\ &= \int_{-n}^{n} \frac{e^{itx}}{2n}\, du \\ &= \frac{e^{itx}}{it2u}\Big|_{-n}^{n} \\ &= \frac{e^{itn} - e^{-itn}}{(it)2n} \\ &= \frac{i2\sin(tn)}{2tni} \\ &= \frac{\sin(tn)}{tn},\end{aligned}$$

where in the second equality we used that $(X_n)_*\mathbb{P} = \mathbb{1}_{[-n,n]}\frac{1}{2n}\, du$.

(b) Let t be fixed. For n large enough we have $|1 - \frac{t^2k^2}{6n^3}| \le 1$ for all $k = 1, \dots, n$. Also since $|\varphi_{X_k}(t)| \le 1$ for all t, we have by exercise 13.4 that

$$\begin{aligned}\Big|\varphi_{Z_n}(t) - \Pi_{k=1}^n\Big(1 - \frac{t^2k^2}{6n^3}\Big)\Big| &= \Big|\Pi_{k=1}^n \varphi_{X_k}\Big(\frac{t}{n^{\frac{3}{2}}}\Big) - \Pi_{k=1}^n\Big(1 - \frac{t^2k^2}{6n^3}\Big)\Big| \\ &\le \sum_{k=1}^n \Big|\varphi_{X_k}\Big(\frac{1}{n^{\frac{2}{3}}}\Big) - 1 + \frac{t^2k^2}{6n^3}\Big|.\end{aligned}$$

Now by (a) we have

$$\varphi_{X_k}\Big(\frac{t}{n^{\frac{3}{2}}}\Big) = \frac{\sin\Big(\frac{kt}{n^{\frac{3}{2}}}\Big)}{\frac{kt}{n^{\frac{3}{2}}}}.$$

The function $f(x) = \frac{\sin(x)}{x},\ f(0) = 1$, has the Taylor series

$$f(x) = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} x^{2k}$$

and hence there exists $C > 0$ such that $|f(x) - 1 + \frac{x^2}{6}| \leq Cx^4$ for all $x \in [-1,1]$. Thus for n large enough we have

$$\Big|\varphi_{X_k}(\frac{t}{n^{\frac{3}{2}}}) - 1 + \frac{t^2k^2}{6n^3}\Big| \leq C\frac{k^4t^4}{n^6}$$

and summing over k we get

$$\sum_{k=1}^{n} \frac{k^4t^5}{n^6} \leq \frac{t^5}{n^6} \cdot n \cdot n^4 = \frac{t^5}{n} \overset{n\to\infty}{\longrightarrow} 0.$$

This proves (b).

(c) We note that

$$\log\Big(\Pi_{k=1}^{n}\Big(1 - \frac{t^2k^2}{6n^3}\Big)\Big) = \sum_{k=1}^{n} \log\Big(1 - \frac{t^2k^2}{6n^3}\Big).$$

By using the Taylor series

$$\log(1-x) = -(x + \frac{x^2}{2} + \frac{x^3}{3} + \dots)$$

for $|x| < 1$, there exists $C > 1$ such that $|\log(1-x) + x| \leq Cx^2$ for all $x \in [-\frac{1}{2}, \frac{1}{2}]$. For n large enough we have $\frac{t^2k^2}{6n^3} \leq \frac{1}{2}$ and thus

$$\begin{aligned}\sum_{k=1}^{n} \log\Big(1 - \frac{t^2k^2}{6n^3}\Big) &= \sum_{k=1}^{n} \Big(\Big(\log\Big(1 - \frac{t^2k^2}{6n^3}\Big) + \frac{t^2k^2}{6n^3}\Big) - \frac{t^2k^2}{6n^3}\Big) \\ &= -\frac{t^2}{6n^3} \cdot \frac{n(n+1)(2n+1)}{6} + \sum_{k=1}^{n} \Big(\log\Big(1 - \frac{t^2k^2}{6n^3}\Big) + \frac{t^2k^2}{6n^3}\Big).\end{aligned}$$

The first term tends to $-\frac{t^2}{18}$ as $n \to \infty$, while the sum is in absolute value less than $C\sum_{k=1}^{n} \frac{t^4k^4}{6n^6} \lesssim \frac{1}{n} \to 0$. Thus $\sum_{k=1}^{n} \log\Big(1 - \frac{t^2k^2}{6n^3}\Big) \to -\frac{t^2}{18}$ and hence $\Pi_{k=1}^{\infty}(1 - \frac{t^2k^2}{6n^3}) \to e^{-\frac{t^2}{18}}$, which is the characteristic function of a $\mathcal{N}(0, \frac{1}{9})$.

14 Students' solutions

14.1 Jingeon An's solution to...

14.1.1 Ex 7.2

Let $\mu_n(A) = \frac{\mu(A \cap A_n)}{\mu(A_n)}$ and $\nu_n(A) = \frac{\nu(A \cap A_n)}{\nu(A_n)}$ for all $A \in \mathcal{G}$ if $\mu(A_n) = \nu(A_n) \neq 0$. Then μ_n and ν_n are probability measures and $\mu_n = \nu_n$ on Π, so $\mu_n = \nu_n$ by Dynkin's π-λ lemma.

Now let $\tilde{A_n} := A_n / \cup_{k=1}^{n-1} A_k$, so that $(\tilde{A_k})_k$ is a partition of Ω. In the upcoming calculation, we shall consider only those $n \in \mathbb{N}$ such that $\mu(A_i) = \nu(A_i) \neq 0$ (w.l.o.g. $\neq 0$ for all i ♠).

Then for any $A \in \mathcal{G}$:

$$\begin{aligned}
\mu(A) &= \sum_{n \geq 1} \mu(A \cap \tilde{A_n}) \\
&\overset{♠}{=} \sum_{n \geq 1} \frac{\mu(A \cap \tilde{A_n})}{\mu(A_n)} \cdot \mu(A_n) \\
&= \sum_{n \geq 1} \mu_n(A \cap \tilde{A_n}) \cdot \mu(A_n) \\
&= \sum_{n \geq 1} \nu_n(A \cap \tilde{A_n}) \cdot \nu(A_n) \\
&= \dots \\
&= \nu(A).
\end{aligned}$$

So $\mu = \nu$. $\square$

14.1.2 Ex 8.4

Consider the measure space $(\mathbb{N}^2, 2^{\mathbb{N}^2}, \mu \otimes \mu)$, where μ is the counting measure on $\mathbb{N}$. Let

$$f(n,m) := \begin{cases} -1 & \text{if } n = m-1 \\ 1 & \text{if } n = m+1 \\ 0 & \text{otherwise.} \end{cases}$$

Then obviously f is measurable (just consider the fact that the sigma-algebra is the power set), and

$$\int \int f(n,m)\, \mu(dn)\, \mu(dm) = 1 \neq -1 = \int \int f(n,m)\, \mu(dm)\, \mu(dn).$$

The condition that is not satisfied in Fubini's theorem is of course the integrability condition. Note that we have played with the fact that $\mathbb{N}^2$ has a corner (the origin)! This counter-example would fail on $\mathbb{Z}^2$.

14.2 Carl Johansson's solution to Ex 7.5

We begin by proving the Jensen inequality for vector-valued random variables $X : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}^n$.

Let $\varphi : \mathbb{R}^n \to \mathbb{R}$ be a convex function.

For any $x_0 \in \mathbb{R}^n$, there are two numbers $a_{x_0} \in \mathbb{R}^n$ and $b_{x_0} \in \mathbb{R}^n$ such that $\varphi(x_0) = \langle a_{x_0}, x_0 \rangle + b_{x_0}$ and $\varphi(y) \ \geq \langle a_y, x_0 \rangle + b_{x_0}$ for all $y \in \mathbb{R}^n$. Then, taking $x_0 = \mathbb{E}[X]$,

$$
\begin{aligned}
\varphi(\mathbb{E}[X]) &= \langle a_{x_0}, \mathbb{E}[X]\rangle + b_{x_0} \\
&= \mathbb{E}[\langle a_{x_0}, x_0\rangle + b_{x_0}] \\
&\leq \mathbb{E}[\varphi(X)] \\
&\leq \infty.
\end{aligned}
$$

Note that in the first inequality, it is implicit that $\mathbb{E}[\varphi(X)^-] < \infty$, which makes this operation legal.

Thus, since the function $\|\cdot\| : \mathbb{R}^n \to \mathbb{R} :: x \mapsto \|x\|$ is Borel measurable (because continuous), $\|X\|$ is measurable. $\|\cdot\|$ is also convex and therefore, using the result above: $\|\mathbb{E}[X]\| \leq \mathbb{E}[\|X\|]$. □

14.3 Jonas Papazoglou-Hennig's solution to...

14.3.1 Ex 1.4

I don't know about "interesting" or "innovative", but hoping for "working" :)

For the first part of the exercise, I propose a maximality argument:

Proof. Let $\mathcal{P}_n = \{\Pi \text{ is a partition of } T : \Pi \subseteq \mathcal{G}, |\Pi| = n\}$. T is finite and there is at least one non-empty $\mathcal{P}_n$, namely $\mathcal{P}_1 = \{\{T\}\}$. Any partition of T can contain at most $|T|$ elements, which would correspond to the finest partition into singletons of the set T. Hence we can find $1 \leq m \leq |T|$, which is the maximal index such that $\mathcal{P}_m \neq \emptyset$ and $\mathcal{P}_n = \emptyset$ for all $n > m$.
Assume that there existed $\Pi_1, \Pi_2 \in \mathcal{P}_m$ such that $\Pi_1 \neq \Pi_2$. Then there would be a partition set $A \in \Pi_2$, which is not in Π_1. But since Π_1 is a partition, there must be some $B \in \Pi_1$, such that $C = A \cap B \neq \emptyset$ and $D = B \setminus C \neq \emptyset$. Note that $C, D \in \mathcal{G}$. But setting

$$\tilde{\Pi} = \Pi_1 \cup \{C, D\} \setminus \{B\} \subseteq \mathcal{G}$$

forms a partition of T with cardinality $m+1$, contradicting maximality of m. Hence we have shown that $|\mathcal{P}_m| = 1$, i.e. there exists a unique maximal partition $\Pi^* \subseteq \mathcal{G}$.
Any element of $\mathcal{G}$ can be expressed through a finite union of elements in Π^*. Because if there were some set $E \in \mathcal{G}$, which could not be written as such, then E must properly intersect (i.e. $\emptyset \neq E \cup A \subset E$ is a proper subset) at least one element of Π^*, allowing us to use an analogous construction like before to create a larger partition of T contained in $\mathcal{G}$, which would contradict maximality of Π^*.
Conversely any non-maximal partition $\Pi \subseteq \mathcal{G}$ cannot form every set in $\mathcal{G}$ by taking unions. This is because if Π is not maximal, there must be some $F \in \Pi$ such that F is the disjoint union of at least two elements in Π^*, and clearly either one of these elements cannot be formed by union of sets in Π alone.
We therefore conclude, that the maximal partition Π^* is the unique partition of T contained in $\mathcal{G}$, which allows every set in $\mathcal{G}$ to be written as a union of elements within it. □

For the second part of the exercise, I use the *principle of good sets*:

Proof. Let $\{A_n\}_{k=1,\dots,n}$ be a partition of T as given. Take

$$\tilde{\mathcal{G}} = \{A : A \text{ is a union of some of the } A_1, ..., A_n\}.$$

Note that we allow the empty union, hence $\emptyset \in \tilde{\mathcal{G}}$. Also, for $A \in \tilde{\mathcal{G}}$, we know that

$$A = \bigcup_{k \in I} A_k,$$

for some $I \subseteq [n]$, therefore

$$A^c = \bigcup_{k \in [n] \setminus I} A_k,$$

which implies $A^c \in \tilde{\mathcal{G}}$. Finally, let $(B_n) \subseteq \tilde{\mathcal{G}}$ and $B = \bigcup_n B_n$. Then for each A_n there exists $I_n \subseteq [n]$ such that $B_n = \bigcup_{k \in I_n} A_k$ and thus

$$B = \bigcup_n B_n = \bigcup_n \bigcup_{k \in I_n} A_k,$$

which implies $B \in \tilde{\mathcal{G}}$.
We see that $\tilde{\mathcal{G}}$ is a σ-field and conclude $\tilde{\mathcal{G}} = \mathcal{G}$, proving that indeed every set in $\mathcal{G}$ may be written as a union of the given partition. By the previous exercise, we further note that the sets $\{A_k\}_{k=1,\dots,n}$ must be a maximal partition of T contained in $\mathcal{G}$, hence it must be the unique generating partition of $\mathcal{G}$. □

The reason for the name *principle of good sets* stems from the fact that we identify the system of all desirable sets (in the above case $\tilde{\mathcal{G}}$, the *good* sets) and can show some structure of this system to conclude that it is in fact the entire σ-field.

14.3.2 Ex 2.7 (b)

I propose a "0-1"-argument:

Proof. Y is X-measurable, therefore, for any $k \in \mathbb{R}$,

$$\{Y \leq k\}$$

is contained both in $\sigma(X)$ and $\sigma(Y)$, where these sigma-fields are further independent by assumption. Hence, we can write

$$\mathbb{P}[Y \leq k] = \mathbb{P}[\{Y \leq k\} \cap \{Y \leq k\}] = \mathbb{P}[Y \leq k]^2,$$

where we used a trivial decomposition of the event $\{Y \leq k\}$ and consider the first set in the decomposition to belong to $\sigma(X)$, the second to $\sigma(Y)$, and then invoke independence of the sigma-fields to write the probability of intersection as a product of probabilities. The equation immediately yields that

$$\mathbb{P}[Y \leq k] \in \{0, 1\}.$$

Note that in general, the distribution function $F_Y(k) := \mathbb{P}[Y \leq k]$ is increasing and right-continuous in k, moreover

$$\lim_{k \to -\infty} F_Y(k) = 0, \lim_{k \to \infty} F_Y(k) = 1.$$

We can infer that

$$S = \{k : F_Y(k) := \mathbb{P}[Y \leq k] = 1\}$$

is closed and $k^* = \inf S$ exists and is contained in S. Using that in this case, $\mathbb{P}[Y \leq k] \in \{0, 1\}$ for all k,

$$\begin{aligned}
\mathbb{P}[Y = k^*] &= \mathbb{P}\left[\bigcap_{n=1}^{\infty} \{k^* - 1/n \leq Y \leq k^* + 1/n\}\right] \\
&= \lim_{n \to \infty} \Big(\mathbb{P}[Y \leq k^* + 1/n] - \mathbb{P}[Y < k^* - 1/n]\Big) \\
&= 1 - 0 \\
&= 1,
\end{aligned}$$

where we used continuity of measure for the second equality. So, $Y = k^* \in \mathbb{R}$ a.s., as desired. □

14.4 Salim Benchelabi's solution to Ex 3.5

We set $\Omega = \{1, 2, 3\}$ and we take the π-system $= \{\{1\}, \{2\}, \emptyset, \Omega\}$. We define a mapping μ such that:

$$\mu(\{1\}) = \mu(\{2\}) = \mu(\Omega) = 1$$

$$\mu(\emptyset) = 0$$

To find a countable family $(A_n)_n$ of disjoint sets of S such that the union is in S, one has to set all A_n to $\emptyset$ except for one which can be any set of S. If one considers these possibilities, the countably additive property is indeed verified.

However the set $\{1, 2\}$ verifies $\{1, 2\} \in \sigma(S)$ and $\mu(\{1\}) + \mu(\{2\}) > \mu(\Omega) \geq \mu(\{1, 2\})$. Then μ is not a measure on $\sigma(S)$.

15 Past Exams

Dear students, here are the practice and final exams for the autumn semester 2020. Due to Covid, the exam was open book & remote. This year, the format will not be the same. It will be on campus and without any material.

15.1 December 2020 Practice Exam with Solutions

Question 15.1.1 (6 points)**.**

Let $(X_n)_{n=1}^\infty$, $(Y_n)_{n=1}^\infty$, X and Y be random variables. Are the following claims true or false? (Answer only "true" or "false", no need to justify your answers.)

(a) If E and F are two events with $\mathbb{P}[E] = 1$, then $\mathbb{P}[E \cap F] = \mathbb{P}[F]$.

(b) If $(X_n, Y_n) \xrightarrow{d} (X, Y)$, then $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} Y$.

(c) The set $\{\emptyset, \{1, 2\}, \{2, 3\}, \{1, 2, 3\}\}$ is a σ-algebra on $\{1, 2, 3\}$.

(d) The half-open intervals $[a, b)$, $a, b \in \mathbb{R}$, $a < b$, form a semi-algebra.

(e) The law of a random vector $(X_1, X_2, \dots, X_n)$ is determined by the c.d.f.s $F_{X_1}, \dots, F_{X_n}$.

(f) Two random variables X and Y are independent if and only if for any Borel sets $A, B \subset \mathbb{R}$ we have $\mathbb{P}[X \in A, Y \in B] = \mathbb{P}[X \in A]\mathbb{P}[Y \in B]$.

Grading: 1 point for every correct answer, -1 points for every wrong answer, 0 points for no answer. Minimum number of points for the whole question is 0.

Solution 15.1.1.

(a) True.

Since $E \cap F \subset F$, we have $\mathbb{P}[E \cap F] \leq \mathbb{P}[F]$. Note that $1 = \mathbb{P}[\Omega] = \mathbb{P}[E \cup E^c] = \mathbb{P}[E] + \mathbb{P}[E^c] = 1 + \mathbb{P}[E^c]$, so that $\mathbb{P}[E^c] = 0$. Since $F = (E \cap F) \uplus (F \cap E^c)$, we have $\mathbb{P}[F] = \mathbb{P}[E \cap F] + \mathbb{P}[E^c \cap F] \leq \mathbb{P}[E \cap F] + \mathbb{P}[E^c] = \mathbb{P}[E \cap F]$.

(b) True.

Let $h \colon \mathbb{R} \to \mathbb{R}$ be bounded and continuous. Then the function $g \colon \mathbb{R}^2 \to \mathbb{R}$ given by $g(x, y) = h(x)$ is also bounded and continuous and thus $\mathbb{E}[h(X_n)] = \mathbb{E}[g(X_n, Y_n)] \to \mathbb{E}[g(X, Y)] = \mathbb{E}[h(X)]$.

(c) False.

The complement of $\{1, 2\}$ is $\{3\}$ which does not belong to the set.

(d) False.

The empty set is not included in the given set. If we add the empty set, they do form a semi-algebra.

(e) False.

Let X and Y be two i.i.d. Bernoulli random variables which take the values 0 and 1 with equal probability. Then (X, X) and (X, Y) have same marginal laws but their joint laws differ.

(f) True.

By definition X and Y are independent if the generated σ-algebras $\sigma(X)$ and $\sigma(Y)$ are. But every event in $\sigma(X)$ is of the form $\{X \in A\}$ for some Borel set A (See Definition 1.22 and Exercise 1.23 in the lecture notes.)

Question 15.1.2 ($12 = 2 + 5 + 5$ points)**.**

Let $(X_n)_{n=1}^{\infty}$ be a sequence of centered i.i.d. Cauchy random variables with scale parameter 1, meaning that X_n has the probability density function $p(x) := \frac{1}{\pi(1+x^2)}$ ($x \in \mathbb{R}$). You may also freely use the fact that the characteristic function of X_n is given by $\varphi_{X_n}(t) := e^{-|t|}$ ($t \in \mathbb{R}$).

(a) Show that $A_n := \frac{1}{n}\sum_{k=1}^{n} X_k \overset{d}{=} X_1$.

(b) Show that $\mathbb{P}[\limsup_{n\to\infty} A_n > \lambda] > 0$ for all $\lambda \in \mathbb{R}$.

(c) Show that $\limsup_{n\to\infty} A_n = \infty$ almost surely.

Solution 15.1.2.

(a) Since the characteristic function determines the law of a random variable, it is enough to show that $\varphi_{A_n} = \varphi_{X_1}$. The characteristic function of $n^{-1}\sum_{k=1}^{n} X_k$ is given by

$$\prod_{k=1}^{n} \varphi_{X_k}(t/n) = \varphi_{X_1}(t/n)^n = e^{-|t|} = \varphi_{X_1}(t).$$

(b) It is enough to show that for all $\lambda \in \mathbb{R}$ we have $\mathbb{P}[\limsup_{n\to\infty} A_n \geq \lambda] > 0$ with non-strict inequality inside the probability. Note that $\{\limsup_{n\to\infty} A_n \geq \lambda\} \supset \limsup_{n\to\infty}\{A_n \geq \lambda\}$. Since the events $E_n = \bigcup_{k=n}^{\infty}\{A_k \geq \lambda\}$ are decreasing, we have

$$\begin{aligned}\mathbb{P}[\limsup_{n\to\infty} A_n \geq \lambda] &\geq \mathbb{P}[\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}\{A_k \geq \lambda\}] = \lim_{n\to\infty}\mathbb{P}[\bigcup_{k=n}^{\infty}\{A_k \geq \lambda\}] \geq \lim_{n\to\infty}\mathbb{P}[A_n \geq \lambda] \\ &= \mathbb{P}[X_1 \geq \lambda] = \int_{\lambda}^{\infty} \frac{dx}{\pi(1+x^2)} > 0.\end{aligned}$$

(c) Note that $\{\limsup_{n\to\infty} A_n = \infty\} = \bigcap_{k=1}^{\infty}\{\limsup_{n\to\infty} A_n \geq k\}$ so it is enough to show that $\mathbb{P}[\limsup_{n\to\infty} A_n \geq k] = 1$ for every fixed $k \geq 1$. By (b) and Kolmogorov's 0–1 law we are done if we show that $\limsup_{n\to\infty} A_n \geq k$ is a tail event. But note that for any $m \geq 1$ we have

$$\limsup_{n\to\infty} A_n = \limsup_{n\to\infty} \frac{1}{n}(\sum_{k=1}^{m} X_k + \sum_{k=m+1}^{n} X_k) = \limsup_{n\to\infty} \frac{1}{n}\sum_{k=m+1}^{n} X_k,$$

where the right hand side is measurable w.r.t. $\sigma\Big(\bigcup_{k=m+1}^{\infty}\sigma(X_k)\Big)$ and hence it follows that $\limsup_{n\to\infty} A_n$ is measurable with respect to the tail σ-algebra generated by the random variables $(X_n)_{n=1}^{\infty}$.

Question 15.1.3 (14 = 6 + 2 + 6 points)**.**

Assume that $(X_n)_{n=1}^{\infty}$ is a sequence of random variables converging in distribution to a random variable X. Let F_n and F be the c.d.f.s of X_n and X respectively.

(a) Show that if F is continuous, then

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \to 0 \tag{$*$}$$

(b) Give an example where F is not continuous and $(*)$ is not true.

(c) Assume that X_n has the density $f_n(x) = \left(1 - \frac{x}{n}\right)^{n-1} \mathbb{1}_{(0,n)}(x)$, $x \in \mathbb{R}$. Construct random variables $Y_n \overset{d}{=} X_n$ such that Y_n converge almost surely. You have to show convergence for your choice of Y_n.

Hint: You can use the fact that $\lim_{n\to\infty} n(x^{1/n} - 1) = \log(x)$.

Solution 15.1.3.

(a) Let us fix $\varepsilon > 0$ and try to show that $\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \le \varepsilon$ for large enough n. Pick $\lambda > 0$ so big that $\mathbb{P}[X \in [-\lambda, \lambda]] \ge 1 - \varepsilon/2$. Let also $n_0 \ge 1$ be so large that $|F_n(\pm\lambda) - F(\pm\lambda)| \le \varepsilon/2$ for all $n \ge n_0$. Then for any x with $x > \lambda$ and $n \ge n_0$ we have

$$|F_n(x) - F(x)| \le \begin{cases} 1 - F(\lambda), & \text{if } F_n(x) \ge F(x) \\ 1 - F_n(\lambda), & \text{if } F_n(x) < F(x) \end{cases}.$$

By the definition of λ we have $1 - F(\lambda) \le \varepsilon/2$ in the first case, while for the second case we have

$$1 - F_n(\lambda) \le 1 - F(\lambda) + |F(\lambda) - F_n(\lambda)| \le \varepsilon$$

for $n \ge n_0$. Similar computation works for $x < -\lambda$ as well.

Assume then that $x \in [-\lambda, \lambda]$. Since $[-\lambda, \lambda]$ is compact, F is uniformly continuous on it and we may pick a finite sequence $-\lambda = x_0 < x_1 < \cdots < x_N = \lambda$ such that $F(x_{k+1}) - F(x_k) \le \varepsilon/2$ for all $0 \le k \le N-1$. Choose now $n_1 \ge n_0$ so large that $|F_n(x_k) - F(x_k)| \le \varepsilon/2$ for all $0 \le k \le N-1$ and $n \ge n_1$. Then if $x_k \le x \le x_{k+1}$ and $F_n(x) \ge F(x)$, we have

$$F_n(x) - F(x) \le F_n(x_{k+1}) - F(x_k) \le |F_n(x_{k+1}) - F(x_{k+1})| + F(x_{k+1}) - F(x_k) \le \varepsilon.$$

Similarly if $F_n(x) \le F(x)$, then

$$F(x) - F_n(x) \le F(x_{k+1}) - F_n(x_k) \le F(x_{k+1}) - F(x_k) + |F(x_k) - F_n(x_k)| \le \varepsilon.$$

In either case $|F(x) - F_n(x)| \le \varepsilon$.

(b) Let $X = 0$ be the constant random variable at 0 and let $X_n = 1/n$. Then $F(x) = \mathbb{1}_{[0,\infty)}(x)$ and $F_n(x) = \mathbb{1}_{[1/n,\infty)}(x)$. Clearly $F(x) - F_n(x) = 1$ for $x \in [0, 1/n)$.

(c) Let's try to mimic the proof of Skorokhod's representation theorem. The c.d.f. of X_n is given by

$$F_n(x) = \int_0^x \left(1 - \frac{t}{n}\right)^{n-1} dt = \Big[_{t=0}^{x} - \left(1 - \frac{t}{n}\right)^n\Big] = 1 - \left(1 - \frac{x}{n}\right)^n$$

for $x \in [0, n]$. We can compute the inverse G_n of F_n as follows

$$F_n(G_n(y)) = y \Leftrightarrow 1 - \Big(1 - \frac{G_n(y)}{n})\Big)^n = y \Leftrightarrow (1-y)^{\frac{1}{n}} = 1 - \frac{G_n(y)}{n} \Leftrightarrow G_n(y) = n(1-(1-y)^{\frac{1}{n}}).$$

Let U be a uniform random variable on $[0, 1]$ and set $Y_n := G_n(U)$. Then

$$\mathbb{P}[Y_n \leq t] = \mathbb{P}[U \leq F_n(t)] = F_n(t),$$

so $Y_n \overset{d}{=} X_n$. Moreover, by the hint $G_n(U) \to -\log(1-U)$ almost surely.

Question 15.1.4 (8 = 4 + 4 points)**.**

Let $(\Omega_i, \mathcal{F}_i)_{i \in I}$ be measurable spaces and assume that $\mathcal{F}_i$ equals the power set σ-algebra $\mathcal{P}(\Omega_i)$ for all $i \in I$.

(a) Show that if I is finite and Ω_i is countable for all $i \in I$, then the product σ-algebra $\bigotimes_{i \in I} F_i$ equals $\mathcal{P}(\prod_{i \in I} \Omega_i)$.

(b) Give an example where I is countable and each Ω_i is finite but the claim in (a) does not hold.

Solution 15.1.4.

(a) The product σ-algebra is generated by sets of the form $\prod_{i \in I} A_i$ with $A_i \subset \Omega_i$. In particular it contains every singleton $\{\omega\}$ with $\omega \in \Omega := \prod_{i \in I} \Omega_i$. Since Ω is countable, also every subset of Ω is countable, and thus a countable union of such singletons.

(b) Let us choose $I = \mathbb{Z}$ and $\Omega_k := \{0, 1\}$ for all $k \in \mathbb{Z}$. Let $T\colon \{0, 1\}^{\mathbb{Z}} \to \{0, 1\}^{\mathbb{Z}}$ be the shift operator mapping $(\omega_n)_{n=-\infty}^{\infty} \mapsto (\omega_{n+1})_{n=-\infty}^{\infty}$. By Proposition 1.31 in the notes there does not exist T-invariant probability measure on $\mathcal{P}(\{0, 1\}^{\mathbb{Z}})$. However the countable product of uniform measures on $\{0, 1\}$ defined on the product σ-algebra is T-invariant, since it is T-invariant on the π-system of cylinder sets, and the set

$$\mathcal{G} := \{A \in \bigotimes_{k \in \mathbb{Z}} \mathcal{P}(\{0, 1\}) : \mu(A) = \mu(T^{-1}A)\}$$

of T-invariant subsets form a λ-system:

- Clearly $\emptyset \in \mathcal{G}$.
- If $A \in \mathcal{G}$ then $\mu(A^c) = 1 - \mu(A) = 1 - \mu(T^{-1}A) = \mu((T^{-1}A)^c) = \mu(T^{-1}A^c)$, so $A^c \in \mathcal{G}$.
- If $(A_n)_{n=1}^{\infty}$ are disjoint elements of $\mathcal{G}$, then

$$\mu(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n) = \sum_{n=1}^{\infty} \mu(T^{-1}A_n) = \mu(\bigcup_{n=1}^{\infty} T^{-1}A_n) = \mu(T^{-1} \bigcup_{n=1}^{\infty} A_n).$$

Question 15.1.5 (6 points)**.**

Either prove the following claim or give a counterexample: Two random variables X and Y are independent if and only if for all continuous and bounded f we have $\mathbb{E}[f(X)f(Y)] = \mathbb{E}[f(X)]\mathbb{E}[f(Y)]$.

Solution 15.1.5.

The claim is false. There are probably many ways to go about finding a counterexample but this might need a bit of creativity. (In retrospect I think this problem might have been a bit too hard as an exam question, although it's a cool puzzle.)
Here's one line of thought: Let us try to find a counterexample of a random vector (X, Y) on $\mathbb{R}^2$ with p.d.f. $p(x, y)$. Note that if (X, Y) satisfies the condition given in the problem statement, then so does (Y, X). It follows that if (X', Y') has the symmetric p.d.f.

$$\tilde{p}(x, y) = \frac{p(x, y) + p(y, x)}{2},$$

then (X', Y') satisfies

$$\mathbb{E}[f(X')f(Y')] = \mathbb{E}[f(X)f(Y)] = \mathbb{E}[f(X)]\mathbb{E}[f(Y)].$$

If we further assume that $X \stackrel{d}{=} Y$, then

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

and

$$\mathbb{E}[f(X')f(Y')] = \mathbb{E}[f(X')]\mathbb{E}[f(Y')].$$

This also works in the other direction: If (X', Y') satisfies the condition then so does (X, Y) (as long as $X \stackrel{d}{=} Y$). We also know that independence implies the condition, so let us try what happens if we assume that X' and Y' are indepedent, i.e. $\tilde{p}$ is of the form

$$\tilde{p}(x, y) = u(x)u(y)$$

for some p.d.f. u on $\mathbb{R}$. Our goal would then to be to *tilt* this symmetric density to obtain a nonindependent density p. Let's assume that u is any continuous p.d.f. with $u(0) > 0$. Then for small enough $\varepsilon > 0$ we may define

$$p(x, y) := u(x)u(y) + (x - y)h(x)h(y),$$

where h is a suitable function supported in $[-\varepsilon, \varepsilon]$. Then automatically we have $\tilde{p}(x, y) = u(x)u(y)$ and $\mathbb{E}[f(X')f(Y')] = \mathbb{E}[f(X')]\mathbb{E}[f(Y')]$ for every f. In order to have also $\mathbb{E}[f(X)f(Y)] = \mathbb{E}[f(X)]\mathbb{E}[f(Y)]$ we need to ensure that $X \stackrel{d}{=} Y$. This will follow if we have that

$$\int (x - y)h(x)h(y)\, dx = 0$$

for every fixed $y \in \mathbb{R}$, in which case both X and Y will have u as their p.d.f. Thus it is enough to pick h in such a way that

$$\int h(x)\, dx = 0 \quad \text{and} \quad \int xh(x)\, dx = 0.$$

For example $h(x) = \cos\left(\frac{\pi x}{\varepsilon}\right)\mathbb{1}_{[-\varepsilon,\varepsilon]}(x)$ works. Clearly then $p(x, y) \neq u(x)u(y)$ in a set of positive measure so X and Y are not independent.

One can also play around with similar ideas in discrete setting. For instance starting from a uniform law on $\{1,2,3\}$ we may construct the i.i.d. pair (X',Y') which is uniform on $\{1,2,3\}^2$. We would then like to consider

$$p(x,y) = \frac{1}{9} + \sigma(x,y)h(x)h(y)$$

where $\sigma(x,y)$ has to be chosen so that $\sigma(x,y) = -\sigma(y,x)$. For instance we can set $\sigma(x,y) = S_{x,y}$, where S is the matrix

$$\begin{pmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix}.$$

The condition for h will this time be

$$\sigma(x,0)h(0) + \sigma(x,1)h(1) + \sigma(x,2)h(2) = 0$$

for any fixed x and a similar condition for any fixed y. This actually implies that h is a constant, and we can choose that constant freely as long as $\frac{1}{9} - |h|^2$ is non-negative. For instance choosing $h = 1/\sqrt{9}$ gives the counterexample $p(x,y) = P_{x,y}$ where P is the matrix

$$\begin{pmatrix} \frac{1}{9} & \frac{2}{9} & 0 \\ 0 & \frac{1}{9} & \frac{2}{9} \\ \frac{2}{9} & 0 & \frac{1}{9} \end{pmatrix}.$$

Question 15.1.6 (6 points)**.**

Let (X,Y) be a uniformly distributed point in the unit disc in $\mathbb{R}^2$, meaning that the random vector (X,Y) has the p.d.f. $p(x,y) := \frac{1}{\pi}\mathbb{1}_{\{x^2+y^2\leq 1\}}$ w.r.t. the Lesbesgue measure.

(a) Compute the regular conditional distribution of X given $\sigma(Y)$.

(b) Next let us write (X,Y) in the polar form $(X,Y) = (R\cos(\theta), R\sin(\theta))$ with $0 \leq R \leq 1$ and $\theta \in [0,2\pi)$. Show that R and θ are independent.

Hint: You may use without proof the fact that

$$\int_{\mathbb{R}^2} f(x,y)\,dx\,dy = \int_0^{2\pi}\int_0^{\infty} f((r\cos(\theta), r\sin(\theta)))r\,dr\,d\theta$$

for any measurable $f \geq 0$.

Solution 15.1.6. (a) By Exercise 6 of Sheet 9 the r.c.d. μ of X given $\sigma(Y)$ a.s. equals

$$\mu(A,\omega) = \frac{\int_A p(x,Y(\omega))\,dx}{\int_{\mathbb{R}} p(x,Y(\omega))\,dx} = \frac{\int_A \mathbb{1}_{[-\sqrt{1-Y(\omega)^2},\sqrt{1-Y(\omega)^2}]}(x)\,dx}{2\sqrt{1-Y(\omega)^2}}$$

Thus μ is a uniform distribution on $[-\sqrt{1-Y}, \sqrt{1-Y}]$.

(b) It is enough to show that for any measurable $f,g \geq 0$ we have

$$\mathbb{E}[f(R)g(\theta)] = \mathbb{E}[f(R)]\mathbb{E}[g(\theta)].$$

Note that if $A(x,y) \in [0,2\pi)$ denotes the argument of (x,y), then

$$\begin{aligned}
\mathbb{E}[f(R)g(\theta)] &= \mathbb{E}[f(\sqrt{X^2+Y^2})g(A(X,Y))] \\
&= \frac{1}{\pi}\int_{\mathbb{R}^2} f(\sqrt{x^2+y^2})g(A(x,y))\mathbb{1}_{[0,1]}(x^2+y^2)\,dx\,dy \\
&= \frac{1}{\pi}\int_0^{2\pi}\int_0^1 f(r)g(\theta)r\,dr\,d\theta \\
&= \frac{1}{\pi}\left(\int_0^1 f(r)r\,dr\right)\left(\int_0^{2\pi} g(\theta)\,d\theta\right) \\
&= \left(\frac{1}{\pi}\int_0^{2\pi}\int_0^1 f(r)r\,dr\,d\theta\right)\left(\frac{1}{\pi}\int_0^{2\pi}\int_0^1 g(\theta)r\,dr\,d\theta\right) \\
&= \left(\frac{1}{\pi}\int_{\mathbb{R}^2} f(\sqrt{x^2+y^2})\mathbb{1}_{[0,1]}(x^2+y^2)\,dx\,dy\right)\left(\frac{1}{\pi}\int_{\mathbb{R}^2} g(A(x,y))\,dx\,dy\right) \\
&= \mathbb{E}[f(\sqrt{X^2+Y^2})]\mathbb{E}[g(A(X,Y))] \\
&= \mathbb{E}[f(R)]\mathbb{E}[g(\theta)].
\end{aligned}$$

Question 15.1.7 (10 = 4 + 3 + 2 + 1 points)**.**

Let $(X_n)_{n=1}^\infty$ be a sequence of i.i.d. standard normal random variables and set

$$Z_n := \prod_{k=1}^n e^{X_i - \frac{1}{2}}$$

for all $n \geq 1$.

(a) Show that there exists $p_0 \in (0,\infty)$ such that, as $n \to \infty$, Z_n converges to 0 in L^p for all $p \in [0,p_0)$ but *not* in L^p for $p \in [p_0,\infty]$.

(b) Examine whether $Z_n \overset{a.s.}{\to} 0$ a.s. as $n \to \infty$ or not.

(c) Examine whether $(Z_n)_{n=1}^\infty$ is uniformly integrable or not.

(d) Examine whether $(Z_n)_{n=1}^\infty$ is tight or not.

Solution 15.1.7.

(a) Using the formula $\mathbb{E}[e^{aX_k}] = e^{\frac{a^2}{2}}$ (Exercise 2 of Sheet 10) we compute for $p \in (0,\infty)$ that

$$\mathbb{E}[Z_n^p] = \prod_{k=1}^n \mathbb{E}[e^{pX_k - \frac{p}{2}}] = e^{\frac{p^2-p}{2}n}.$$

When $p < 1$ we see that this tends to 0, so $Z_n \to 0$ in L^p. Similarly when $p \geq 1$ the sequence Z_n cannot converge, since if it did, it would have to converge to 0 (since it converges to 0 in probability), but its norm doesn't converge to 0. Thus $p_0 = 1$ works.

(b) The sequence does indeed converge almost surely. To show this we may use Borel–Cantelli. Fix $p < 1$ and notice that for all $k \geq 1$ we have

$$\mathbb{P}[Z_n > k^{-1}] \leq \mathbb{E}[Z_n^p]k^p \leq e^{\frac{p(p-1)}{2}n}k^p$$

where the right hand side is summable over n. Thus for any $k \geq 1$ there almost surely exists $n_k \geq 1$ such that $Z_n \leq k^{-1}$ for $n \geq n_k$. Since there are countably many such k, we can almost surely find for all $k \geq 1$ such n_ks simultaneously, and hence Z_n converges to 0.

(c) Since Z_n converges in probability but not in L^1, it cannot be uniformly integrable.

(d) Since Z_n converges almost surely, it converges in law and hence the sequence is tight.

15.2 January 2021 Final Exam with Solutions

Note on the format: The January 2021 Exam was open book and remote due to COVID. Questions were randomly shuffled and for some of the exercises only a random subset of all possible questions appeared for a specific student. The exam lasted 4 hours, giving enough time for students to handle IT issues (scanning, submitting).

Question 15.2.1 (6 points).

Let $(X_n)_{n=1}^\infty$, $(Y_n)_{n=1}^\infty$, X and Y be random variables. Are the following claims true or false? (Answer only "true" or "false", no need to justify your answers.)

(a) If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{\mathbb{P}} Y$ then $(X_n, Y_n) \xrightarrow{d} (X, Y)$.

(b) If $X_n \xrightarrow{\mathbb{P}} X$, $\mathbb{E}[|X_n|] \leq 1$ for all $n \geq 1$, and $X \in L^1$, then $X_n \to X$ in L^1.

(c) If $X_n \to X$ in L^4, then $X_n \to X$ in L^2.

(d) If $(X_n)_{n=1}^\infty$ is a sequence of positive i.i.d. random variables with $\mathbb{E}[X_1] = \pi$, then

$$\frac{1}{n} \log\Big(\prod_{k=1}^n X_k\Big) \to \log(\pi)$$

almost surely as $n \to \infty$.

(e) If $(X_n)_{n=1}^\infty$ are i.i.d. random variables with $\mathbb{P}[X_1 = 1] = \mathbb{P}[X_1 = -1] = 1/2$, then $\frac{\sum_{k=1}^n X_k}{\sqrt{n}}$ converges in law to a standard normal random variable.

(f) If $(X_n)_{n=1}^\infty$ are i.i.d. random variables with $\mathbb{P}[X_1 = 1] = \mathbb{P}[X_1 = -1] = 1/2$, then $\cos\Big(\frac{\sum_{k=1}^n X_k}{n}\Big)$ converges almost surely to 1.

(g) If μ and ν are two probability measures defined on a measurable space $(\Omega, \mathcal{F})$ that agree on a semialgebra $\mathcal{A} \subset \mathcal{F}$ such that $\sigma(\mathcal{A}) = \mathcal{F}$, then $\mu = \nu$.

(h) If $\mathbb{E}[\sqrt{|X_n|}] \leq 4$ for all n, then the sequence $(X_n)_{n=1}^\infty$ is tight.

(i) There exists a σ-finite measure μ on the Borel σ-algebra of $\mathbb{R}$ such that $\mu(\{x\}) > 0$ for every $x \in \mathbb{R}$.

(j) If X is uniformly distributed on $[0, 1]$, then its characteristic function is integrable.

(k) If the sequence $(X_n)_{n=1}^\infty$ is tight, then $\sup_n \mathbb{E}[|X_n|] < \infty$.

(l) If for all $n \geq 1$ the random variable X_n is uniformly distributed in $\{1, 2, \ldots, n\}$, then X_n/n converges in law to a uniform random variable on $[0, 1]$.

Grading: 1 point for every correct answer, -1 points for every wrong answer, 0 points for no answer. Minimum number of points for the whole question is 0.

Solution 15.2.1.

(a) False.

(b) False.

(c) True.

(d) False.

(e) True.

(f) True.

(g) True.

(h) True.

(i) False.

(j) False.

(k) False.

(l) True.

Question 15.2.2 ($10 = 3 + 3 + 4$ points)**.**

Solve the following problems:

(a) Give an example of two probability measures μ and ν defined on the power set σ-algebra $\mathcal{F} := \mathcal{P}(\{1, 2, 3, 4\})$ and a subcollection $\mathcal{A} \subset \mathcal{F}$ with $\sigma(\mathcal{A}) = \mathcal{F}$ such that $\mu(A) = \nu(A)$ for every $A \in \mathcal{A}$ but $\mu \neq \nu$.

(b) Give a counterexample to the following claim: If μ is a countably additive function defined on a λ-system Λ on the sample space $\{1, 2, 3, 4\}$ such that $\mu(\{1, 2, 3, 4\}) = 1$, then μ extends uniquely to a probability measure on $\sigma(\Lambda)$.

(c) Let X be an $\mathbb{R}^d$-valued random variable. Show that there exists a sequence X_n of $\mathbb{R}^d$-valued random variables such that each X_n takes only finitely many values and $X_n \to X$ surely as $n \to \infty$.

(d) Show that there does not exist a probability measure μ on the Borel σ-algebra $\mathcal{B}$ of $\mathbb{R}$ such that $\mu(A + x) = \mu(A)$ for all $A \in \mathcal{B}$ and $x \in \mathbb{R}$. (Here $A + x := \{a + x : a \in A\}$.)

(e) Show that the Borel σ-algebra on $\mathbb{R}$ is generated by intervals of the form $[k2^{-n}, (k + 1)2^{-n}]$ with $n \geq 0$ and $k \in \mathbb{Z}$.

(f) Show that if $X_n \xrightarrow{\mathbb{P}} X$ and $(X_n^2)_{n=1}^{\infty}$ is a uniformly integrable sequence then $X_n \to X$ in L^2.

Solution 15.2.2.

(a) Take $\mathcal{A} = \{\{1, 2\}, \{1, 3\}\}$ and $\mu(\{k\}) = p_k$, $\nu(\{k\}) = q_k$ with $p_1 + p_2 = q_1 + q_2$ and $p_1 + p_3 = q_1 + q_3$. For instance $p_1 = p_2 = p_3 = p_4 = \frac{1}{4}$ and $q_1 = p_4 = \frac{1}{2}, q_2 = q_3 = 0$. Then $\mathcal{A}$ generates $\mathcal{P}(\{1, 2, 3, 4\})$ since it generates the singletons (we let the reader carry out the details).

(b) One counterexample is the following: $\Lambda = \{\emptyset, \{1, 2\}, \{1, 3\}, \{2, 4\}, \{3, 4\}, \{1, 2, 3, 4\}\}$, $\mu(\{1, 2\}) = \mu(\{3, 4\}) = \mu(\{1, 3\}) = \mu(\{2, 4\}) = \frac{1}{2}$. We can extend μ either as the uniform measure on $\{1, 2, 3, 4\}$, or as $\mu(\{1\}) = \mu(\{4\}) = \frac{1}{2}$ and $\mu(\{2\}) = \mu(\{3\}) = 0$.

(c) We showed during lectures that any $\mathbb{R}-$valued random variable Y admits a sequence $(Y_n)_{n=1}^\infty$ of simple random variables such that $Y_n \to Y$ surely. Let us denote $X = (X^{(1)}, \dots, X^{(d)})$. Then for each $X^{(k)}$ there exists a sequence of random variables $X_n^{(k)} \to X^{(k)}$ surely, and then $X_n := (X_n^{(1)}, X_n^{(2)}, \dots, X_n^{(d)}) \to X$ surely.

(d) Assume that such a probability measure exists. Then $1 = \mu(\mathbb{R}) = \mu(\cup_{k=1}^\infty [k, k+1)) = \sum_{k=1}^\infty \mu([k, k+1))$, so one of the $\mu([k, k+1))$ has to be nonzero. But then by translation invariance the sum is infinite, which is a contradiction.

(e) Let $\mathcal{A}$ be the collection of the intervals of the given form. We know that the Borel sigma-algebra is generated by open intervals, so it is enough to show that $(a, b) \in \sigma(\mathcal{A})$ for all $a < b$. In fact we claim that $(a, b) = \cup\big\{[\frac{k}{2^n}, \frac{k+1}{2^n}] : k \in \mathbb{Z}, n \geq 0, [\frac{k}{2^n}, \frac{k+1}{2^n}] \subset (a, b)\big\}$. Clearly the right hand side is contained in the left hand side. To show the other inclusion let $x \in (a, b)$. Then there exists $n \geq 0$ such that $x - 2^n > a$ and $x + 2^n < b$. Letting $k = \lfloor 2^n x \rfloor$ we have $\frac{k}{2^n} \leq x$ and $\frac{k+1}{2^n} \geq \frac{2^n x - 1 + 1}{2^n} = x$, so $x \in [\frac{k}{2^n}, \frac{k+1}{2^n}] \subset (a, b)$.

(f) First of all note that $X \in L^2$ since by Fatou's lemma $\mathbb{E}[X^2] \leq \liminf_{n\to\infty} \mathbb{E}[X_n^2] < \infty$. Note that we have that $|X_n - X|^2 \to 0$ in probability. It is enough to show that $|X_n - X|^2$ is uniformly integrable. Note that for any event A we have $\mathbb{E}[|X_n - X|^2 \mathbb{1}_A] \leq 2\mathbb{E}[X_n^2 \mathbb{1}_A] + 2\mathbb{E}[X^2 \mathbb{1}_A]$, and the right hand side tends to 0 as $\mathbb{P}(A) \to 0$, uniformly in n by the uniform integrability of $(X_n^2)_{n=1}^\infty$.

Question 15.2.3 ($10 = 3 + 2 + 3 + 2$ points)**.**

Let $(X_n)_{n=1}^\infty$ be a sequence of non-negative uniformly integrable random variables.

(a) Show that for any $\varepsilon > 0$ we have $\sum_{k=1}^n \mathbb{P}[X_k \geq \varepsilon n] \to 0$ as $n \to \infty$.

(b) Show that $n^{-1} \max_{1\leq k\leq n} X_k \to 0$ in probability as $n \to \infty$.

(c) Show that $n^{-1} \max_{1\leq k\leq n} X_k \to 0$ in L^1 as $n \to \infty$.

(d) Assume now that $(Y_n)_{n=1}^\infty$ is a sequence of identically distributed non-negative random variables in L^p for some $p \geq 1$. Show that

$$\lim_{n\to\infty} n^{-1/p} \mathbb{E}\Big[\max_{1\leq k\leq n} Y_k\Big] = 0.$$

Solution 15.2.3.

(a) Note that $\mathbb{P}(\{X_k \geq \epsilon n\}) = \mathbb{E}[\mathbb{1}_{\{X_k \geq \epsilon n\}}] \leq \mathbb{E}[\frac{X_k}{\epsilon n} \mathbb{1}_{\{X_k \geq \epsilon n\}}]$ and hence $\sum_{k=1}^n \mathbb{P}(\{X_k \geq \epsilon n\} \leq \sum_{k=1}^n \frac{1}{\epsilon n} \mathbb{E}[X_k \mathbb{1}_{\{X_k \geq \epsilon n\}}] \leq \frac{1}{\epsilon} \sup_{k\geq 1} \mathbb{E}[X_k \mathbb{1}_{\{X_k \geq \epsilon n\}}]$, where the right hand side goes to zero as $n \to \infty$ by uniform integrability.

(b) It is enough to show that for any $\epsilon > 0$ we have $\mathbb{P}(\{n^{-1} \max_{1\leq k\leq n} X_k \geq \epsilon\}) \to 0$ as $n \to \infty$. Note that $\mathbb{P}(\{n^{-1} \max_{1\leq k\leq n} X_k \geq \epsilon\}) = \mathbb{P}(\cup_{k=1}^n \{X_k \geq \epsilon n\}) \leq \sum_{k=1}^n \mathbb{P}(\{X_k \geq \epsilon n\}) \to 0$ by point (a).

(c) By (b) it is enough to show that $n^{-1} \max_{1\leq k\leq n} X_k$ is uniformly integrable. We note that for any $t > 0$ and $n \geq 1$ we have $\mathbb{E}[n^{-1} \max_{1\leq k\leq n} X_k \mathbb{1}_{\{n^{-1} \max_{1\leq k\leq n} X_k > t\}}] \leq \sum_{n=1}^n X_k \mathbb{1}_{\{n^{-1} X_k > t\}}$ and this further less than $\sup_{k\geq 1} \mathbb{E}[X_k \mathbb{1}_{\{X_k > nt\}}] \leq \sup_{k\geq 1} \mathbb{E}[X_k \mathbb{1}_{\{X_k > t\}}]$. The right hand side does not depend on n anymore and tends to 0 as $t \to \infty$.

(d) Since Y_n^p are identically distributed, $(Y_n^p)_{n=1}^\infty$ is a uniformly integrable sequence. By applying (c) we thus have $n^{-1}\mathbb{E}[\max_{1\le k\le n} Y_n^p] \to 0$ as $n \to \infty$, and by Jensen's inequality

$$(n^{-\frac{1}{p}}\mathbb{E}[\max_{1\le k\le n} Y_n])^p = n^{-1}(\mathbb{E}[\max_{1\le k\le n} Y_n])^p \le n^{-1}\mathbb{E}[\max_{1\le k\le n} Y_n^p]) \to 0,$$

so $(n^{-\frac{1}{p}}\mathbb{E}[\max_{1\le k\le n} Y_n]) \to 0$.

Question 15.2.4 ($12 = 1+3+1+2+3+2$ points)**.**

Let $(X_{n,k})_{n,k=1}^\infty$ be a family of i.i.d. random variables such that each $X_{n,k}$ has the p.d.f.

$$p(x) = \frac{3 \cdot \mathbb{1}_{[1,\infty)}(x)}{x^4}.$$

Denote $S_n := \sum_{k=1}^n X_{n,k}$ for all $n \ge 1$.

(a) Show that the expectation $\mu := \mathbb{E}[X_{1,1}]$ exists and compute its value.

(b) Show that $n^{-1}S_n \to \mu$ in L^2.

(c) Show that $n^{-1/2}(S_n - n\mu)$ converges in law to a normal random variable. What is its variance?

(d) Let $\tilde{X}_{n,k} := X_{n,k}\mathbb{1}_{\{X_{n,k}\le n\}}$ and $\tilde{S}_n := \sum_{k=1}^n \tilde{X}_{n,k}$. Show that almost surely $S_n = \tilde{S}_n$ for all large enough n.

(e) Prove that there exists $C > 0$ such that $\mathbb{E}[(n^{-1}\tilde{S}_n - \mathbb{E}[\tilde{X}_{n,1}])^4] \le Cn^{-2}$ for all $n \ge 1$.

(f) Show that $n^{-1}S_n \to \mu$ almost surely.

Solution 15.2.4.

(a) Since everything is non-negative we may simply compute $\mathbb{E}[X1, 1]$ $= \int_1^\infty x \cdot \frac{3}{x^4}\, dx = \frac{-3x^{-2}}{2}\Big|_1^\infty = \frac{3}{2} =: \mu < \infty$.

(b) We have $\mathbb{E}[|n^{-1}S_n - \mu|^2] = \mathbb{E}[(n^{-1}\sum_{k=1}^n X_{n,k} - \mu)^2] = \mathbb{E}[(n^{-1}\sum_{k=1}^n (X_{n,k} - \mu))^2]$ $= n^{-2}\sum_{j,k=1}^n \mathbb{E}[(X_{n,k} - \mu)(X_{n,j} - \mu)]$. Note that by independence the expectation is 0 if $j \ne k$. Hence this equals $n^{-2}\sum_{k=1}^n \mathbb{E}[(X_{n,k} - \mu)^2] = n^{-1}\mathbb{E}[(X_{1,1} - \mu)^2] = n^{-1}\sigma^2 \to 0$, where $\sigma^2 := \mathbb{E}[(X_{1,1} - \mu)^2] = \mathbb{E}[X_{1,1}^2] - \mu^2 = \int_1^\infty \frac{3}{x^2}\, dx - \mu^2 = 3 - \frac{9}{4} = \frac{3}{4}$.

(c) Note that S_n has the same distribution as $\sum_{k=1}^n X_{1,k}$, so $\frac{S_n - n\mu}{\sqrt{n}} \to \mathcal{N}(0, \sigma^2 = \frac{3}{4})$ by the Central Limit Theorem.

(d) By the Borel-Cantelli lemma it is enough to show that $\sum_{n=1}^\infty \sum_{k=1}^n \mathbb{P}(\{X_{n,k} \ge n\}) < \infty$. We have $\mathbb{P}(\{X_{n,k} \ge n\}) = \int_n^\infty \frac{3}{x^4}\, dx = \frac{-1}{x^3}\Big|_n^\infty$, so $\sum_{n=1}^\infty \sum_{k=1}^n \mathbb{P}(\{X_{n,k} \ge n\}) = \sum_{n=1}^\infty \frac{1}{n^2} < \infty$.

(e) We compute $\mathbb{E}[(n^{-1}\tilde{S}_n - \mathbb{E}[\tilde{X}_{n,k}])^4] = n^{-4}\sum_{i,j,k,l=1}^n \mathbb{E}[(\tilde{X}_{n,i} - \mathbb{E}[\tilde{X}_{n,i}]) \cdot (\tilde{X}_{n,j} - \mathbb{E}[\tilde{X}_{n,j}]) \cdot (\tilde{X}_{n,k} - \mathbb{E}[\tilde{X}_{n,k}]) \cdot (\tilde{X}_{n,l} - \mathbb{E}[\tilde{X}_{n,l}])]$. Note that by independence only terms where $i = j = k = l$ or $i = j \ne k = l$ (or permutations thereof) are nonzero. Hence this equals $n^{-3}\mathbb{E}[(\tilde{X}_{n,1} - \mathbb{E}[\tilde{X}_{n,1}])^4] + \frac{6C_2^n}{n^4}\mathbb{E}[(\tilde{X}_{n,1} - \mathbb{E}[\tilde{X}_{n,1}])^2]^2$. The second term is $\mathcal{O}(n^{-1})$ since $\mathbb{E}[(\tilde{X}_{n,1} - \mathbb{E}[\tilde{X}_{n,1}])^2] \le \mathbb{E}[X_{n,1}^2] = 3$. For the first term we note that $\mathbb{E}[(\tilde{X}_{n,1} - \mathbb{E}[\tilde{X}_{n,1}])^4] \le 16\mu^4 + \mathbb{E}[\tilde{X}_{n,1}^4] = 16\mu^4 + 16\int_1^n \frac{x^4 3}{x^4}\, dx = 16\mu^4 + 48n$, so the first term is also $\mathcal{O}(n)$.

(f) By (d) and since $\mathbb{E}[\tilde{X}_{n,1}] \to \mu$ as $n \to \infty$, it is engouh to show that $n^{-1}\tilde{S}_n - \mathbb{E}[\tilde{X}_{n,1}] \to 0$ almost surely. We will again use Borel-Cantelli. Note that for any $\epsilon > 0$ we have $\sum_{n=1}^{\infty} \mathbb{P}(\{|n^{-1}\tilde{S}_n - \mathbb{E}[\tilde{X}_{n,1}]| > \epsilon\}) \leq \sum_{n=1}^{\infty} \frac{\mathbb{E}[(n^{-1}\tilde{S}_n - \mathbb{E}[\tilde{X}_{n,1}])^4]}{\epsilon^4} \leq \frac{C}{\epsilon^2} \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty$. Thus by Borel-Cantelli $|n^{-1}\tilde{S}_n - \mathbb{E}[\tilde{X}_{n,1}]| \leq \epsilon$ for n large enough. Considering the countable sequence of epsilons $\epsilon = \frac{1}{m}$ and $m = 1, 2, \ldots$, we see that with full probability $|n^{-1}\tilde{S}_n - \mathbb{E}[\tilde{X}_{n,1}]| \to 0$ as $n \to \infty$.

Question 15.2.5 (8 points)**.**

Write a short essay (at most 500 words) on Carathéodory's extension theorem and product spaces. The essay should contain high level answers at least to the following questions:

- What is Carathéodory's extension theorem?
- Why is it important?
- How is it proven? (A brief summary of main ideas, no details!)
- How can it be applied in the construction of product measures?

Solution 15.2.5. We refer to your lecture notes.

16 Exercise Sheets

16.1 Exercise Sheet 1

16.2 Exercise Sheet 2

16.3 Exercise Sheet 3

16.4 Exercise Sheet 4

16.5 Exercise Sheet 5

16.6 Exercise Sheet 6

16.7 Exercise Sheet 7

16.8 Exercise Sheet 8

16.9 Exercise Sheet 9

16.10 Exercise Sheet 10

16.11 Exercise Sheet 11

16.12 Exercise Sheet 12

16.13 Exercise Sheet 13

References

[1] Rick Durrett. *Probability: Theory and Examples*. Ed. by Rick Durett. 2019 (cit. on pp. 1, 22).

[2] Ioannis Karatzas and Steven E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer Verlag, 1991 (cit. on p. 1).

[3] Jean-François Le Gall. *Brownian Motion, Martingales, and Stochastic Calculus*. Ed. by Sheldon Axler and Kenneth Ribet. Springer Verlag, 2010 (cit. on p. 1).

[4] Jean-François Le Gall. *Intégration, Probabilités et Processus Aléatoires*. Ed. by Jean-François Le Gall. 2005 (cit. on pp. 1, 29, 31, 35).

[5] Daniel Revuz and Marc Yor. *Continuous Martingales and Brownian Motion*. Springer Verlag, 1999 (cit. on p. 1).

Index