Part II — Probability and Measure

Based on lectures by Dr Sarkar

Michaelmas 2023

Contents

§0 Holes in classical theory

Analysis

1. What is the "volume" of a subset of $\mathbb{R}^d$.
2. Integration (Riemann Integration has holes)

- $\{f_n\}$ a sequence of continuous functions on $[0,1]$ s.t.
  - $0 \leq f_n(x) \leq 1 \; \forall \; x \in [0,1]$.
  - $f_n(x)$ is monotonically decreasing on $n \to \infty$, i.e. $f_n(x) \geq f_{n+1}(x) \; \forall \; x/$

  So, $\lim_{n\to\infty} f_n(x)$ exists. But f is not Riemann integrable . We want a theory of integration s.t. f is integrable and $\lim_{n\to\infty} \int_0^1 f_n(x) \, \mathrm{d}x = \int_0^1 f(x) \, \mathrm{d}x$.

3. $L^1 = ()$ If $f \in L^1$ is f Riemann integrable? Will have to change the definition of integral. L^2 a hilbert space

Probability

1. Discrete probability has its limitations,
   - Toss a unbiased coin 5 times. What is the probability if getting 3 heads?
   - Take an infinite sequence of coin tosses ($E = \{0,1\}^{\mathbb{N}}$ which is uncountable) and an event A that depends on that infinite sequence. How do you define $\mathbb{P}(A)$? E.g. $X_i \sim$ Ber $\left(\frac{1}{2}\right)$ and $A = \frac{\sum_{i=1}^n X_i}{n}$, the average number of heads. By strong law of large numbers $\mathbb{P}\left(\frac{\sum_{i=1}^n X_i}{n} \to \frac{1}{2}\right) = 1$.
   - How to draw a point uniformly at random from $[0,1]$? $U \sim U[0,1]$. Probability needs axioms to be made rigorous.
2. Define Expectation for a r.v.. Also would want the following if $0 \leq X_n \leq 1$ and $X_n \downarrow X$ then $\mathbb{E}X_n \to \mathbb{E}X$.

§2 Measurable Functions

§2.1 Definition

Definition 2.1 (Measurable)

Let $(E, \mathcal{E}), (G, \mathcal{G})$ be measurable spaces. A function $f : E \to G$ is called **measurable** if $f^{-1}(A) \in \mathcal{E} \ \forall \ A \in \mathcal{G}$, where $f^{-1}(A)$ is the preimage of A under f i.e. $f^{-1}(A) = \{x \in E : f(x) \in A\}$.

If $G = \mathbb{R}$ and $\mathcal{G} = \mathcal{B}$, we can just say that $f : (E, \mathcal{E}) \to G$ is measurable. Moreover, if E is a topological space and $\mathcal{E} = \mathcal{B}(E)$, we say f is Borel measurable.

Note that preimages f^{-1} commute with many set operations such as intersection, union, and complement. This implies that $\{f^{-1}(A) \mid A \in \mathcal{G}\}$ is a σ-algebra over E, and likewise, $\{A : f^{-1}(A) \in \mathcal{E}\}$ is a σ-algebra over G. Hence, if $\mathcal{A}$ is a collection of subsets s.t. $G \supset \sigma(\mathcal{A})$ then if $f^{-1}(A) \in \mathcal{E}$ for all $A \in \mathcal{A}$, the class $\{A : f^{-1} \in \mathcal{E}\}$ is a σ-algebra that contains $\mathcal{A}$ and so $\sigma(\mathcal{A})$. So f is measurable.

If $f : (E, \mathcal{E}) \to \mathbb{R}$, the collection $\mathcal{A} = \{(-\infty, y] : y \in \mathbb{R}\}$ generates $\mathcal{B}$ (Sheet 1). Hence f is Borel measurable iff $f^{-1}((-\infty, y]) = \{x \in E : f(x) \leq y\} \in \mathcal{E}$ for all $y \in \mathbb{R}$.

If E is a topological space and $\mathcal{E} = \mathcal{B}(E)$, then if $f : E \to \mathbb{R}$ is continuous, the preimages of open sets B are open, and hence Borel sets. The open sets in $\mathbb{R}$ generate the σ-algebra $\mathcal{B}$. Hence, continuous functions to the real line are measurable.

Example 2.1

Consider the indicator function 1_A of a set $A \subset E$. $1_A^{-1}(1) = A$ and $1_A^{-1}(0) = A^c$ hence measurable iff $A \in \mathcal{E}$.

Example 2.2

The composition of measurable functions is measurable. Note that given a collection of maps $\{f_i : E \to (G, \mathcal{G}) \mid i \in I\}$, we can make them all measurable by taking $\mathcal{E}$ to be a large enough σ-algebra, for instance $\sigma\left(\left\{f_i^{-1}(A) : A \in \mathcal{G}, i \in I\right\}\right)$ called the σ-algebra generated by $\{f_i\}_{i \in I}$.

Proposition 2.1

If $f_1, f_2, \ldots$ are measurable $\mathbb{R}$-valued. Then $f_1 + f_2$, $f_1 f_2$, $\inf_n f_n$, $\sup_n f_n$, $\liminf f_n$, $\limsup f_n$ are all measurable.

Proof. See Sheet 1. $\square$

§2.2 Monotone class theorem

Theorem 2.1 (Monotone class theorem)

Let $(E, \mathcal{E})$ be a measurable space and $\mathcal{A}$ be a π-system that generates the σ-algebra $\mathcal{E}$. Let $\mathcal{V}$ be a vector space of bounded maps from E to $\mathbb{R}$ s.t.

1. $1_E \in \mathcal{V}$;
2. $1_A \in \mathcal{V}$ for all $A \in \mathcal{A}$;
3. if f is bounded and $f_n \in \mathcal{V}$ are nonnegative functions that form an increasing sequence that converge pointwise to f on E, then $f \in \mathcal{V}$.

Then $\mathcal{V}$ contains all bounded measurable functions $f\colon E \to \mathbb{R}$.

Proof. Define $\mathcal{D} = \{A \in \mathcal{E} : 1_A \in \mathcal{V}\}$. Then $\mathcal{D}$ is a d-system as $1_E \in \mathcal{V}$ and for $A \subseteq B$, $1_{B \setminus A} = 1_B - 1_A \in \mathcal{V}$ as $\mathcal{V}$ a vector space so $B \setminus A \in \mathcal{D}$.
If $A_n \in \mathcal{D}$ increases to A, we have 1_{A_n} increases pointwise to 1_A, which lies in $\mathcal{V}$ by the (3.) so $A \in \mathcal{D}$.

$\mathcal{D}$ contains $\mathcal{A}$ by (2.), as well as E itself. So by Dynkin's lemma $\mathcal{D}$ contains $\sigma(\mathcal{A}) = \mathcal{E}$ so $\mathcal{E} = \mathcal{D}$ i.e. $1_A \in V \ \forall\ A \in \mathcal{E}$.

Since V a vector space it contains all finite linear combinations of indicators of measurable sets. Let $f\colon E \to \mathbb{R}$ be a bounded measurable function, which we will assume at first is nonnegative. We define

$$\begin{aligned} f_n(x) &= 2^{-n} \lfloor 2^n f(x) \rfloor \\ &= 2^{-n} \sum_{j=0}^{\infty} 1_{A_{n,j}}(x) \\ A_{n,j} &= \{2^n f(x) \in [j, j+1)\} \\ &= f^{-1}\left(\left[\frac{j}{2^n}, \frac{j+1}{2^n}\right)\right) \in \mathcal{E}. \end{aligned}$$

As f is bounded we do not need an infinite sum but only a finite one. Then $f_n \le f \le f_n + 2^{-n}$. Hence $|f_n - f| \le 2^{-n} \to 0$ and $f_n \uparrow f$.

So $0 \le f_n \uparrow f, f_n \in \mathcal{V}$ and f is bounded non-negative so $f \in \mathcal{V}$ by (3.).

Finally, for any f bounded and measurable, $f = f^{+}$[a] $- f^{-}$[b]. f^+, f^- are bounded, nonnegative and measurable, so in $\mathcal{V}$ and $\mathcal{V}$ a vector space thus $f \in \mathcal{V}$. □

[a] $\max(f, 0)$
[b] $\max(-f, 0)$

§2.3 Image measures

Definition 2.2 (Image Measure)

Let $f\colon (E,\mathcal{E}) \to (G,\mathcal{G})$ be a measurable function and μ a measure on $(E,\mathcal{E})$. Then the **image measure** $\nu = \mu \circ f^{-1}$ is obtained from assigning $\nu(A) = \mu(f^{-1}(A))$ for all $A \in \mathcal{G}$.

Remark 15. This is well defined as $f^{-1}(A) \in \mathcal{E}$ as f measurable. ν is countably additive because the preimage satisfies set operations and μ countably additive (See Sheet 1).

Starting from the Lebesgue measure, we can get all probability measures (in fact we can get all Radon measures) in this way.

Definition 2.3 (Right-Continuous)

A function f is **right-continuous** if $x_n \downarrow x \implies f(x_n) \to f(x)$.

Lemma 2.1

Let $g\colon \mathbb{R} \to \mathbb{R}$ be a non-constant, increasing, right-continuous function, and set $g(\pm\infty) = \lim_{z\to\pm\infty} g(z)$. On $I = (g(-\infty), g(+\infty))$ we define the **generalised inverse** $f : I \to \mathbb{R}$ by

$$f(x) = \inf\{y \in \mathbb{R} : g(y) \geq x\}.$$

Then f is increasing, left-continuous, and $f(x) \leq y$ iff $x \leq g(y)$ for all $x \in I, y \in \mathbb{R}$.

Remark 16. f and g form a Galois connection.

Proof. Fix $x \in I$.
Let $J_x = \{y \in \mathbb{R} : g(y) \geq x\}$. Since $x > g(-\infty)$, J_x is nonempty and bounded below. Hence $f(x)$ is a well-defined real number.
If $y \in J_x$, then $y' \geq y$ implies $y' \in J_x$ since g is increasing. Since g is right-continuous, if $y_n \downarrow y$, and all $y_n \in J_x$, then $g(y) = \lim_n g(y_n) \geq x$ so $y \in J_x$.
So $J_x = [f(x), \infty)$. Hence $f(x) \leq y \iff x \leq g(y)$ as required.

If $x \leq x'$, we have $J_x \supseteq J_{x'}$ (as $y \in J_x \impliedby y \in J'_x$), i.e. $[f(x),\infty) \supseteq [f(x'),\infty)$ so $f(x) \leq f(x')$.
Similarly, if $x_n \uparrow x$, we have $J_x = \bigcap_n J_{x_n}$ [a] so $[f(x),\infty) = \bigcap_n [f(x_n),\infty)$ so $f(x_n) \to f(x)$ as $x_n \to x$. □

[a] As $y \in \bigcap_n J_{x_n} \iff g(y) \geq x_n \;\forall\, n \iff g(y) \geq x \iff y \in J_x$.

Theorem 2.2

Let $g\colon \mathbb{R} \to \mathbb{R}$ as in the previous lemma. Then $\exists$ a unique Radon measure μ_g on $\mathbb{R}$ such that $\mu_g((a,b]) = g(b) - g(a)$ for all $a < b$. Further, all Radon measures on $\mathbb{R}$ can be obtained in this way.

Proof. Define I, f as in the previous lemma and λ the Lebesgue measure on I.

f is Borel measurable since $f^{-1}((-\infty, z]) = \{x \in I\colon f(x) \le z\} = \{x \in I\colon x \le g(z)\} = (-g(\infty), g(z)] \in \mathcal{B}$. As $\{(-\infty, z] : z \in \mathbb{R}\}$ generate $\mathcal{B}$, f measurable.

Therefore, the image measure $\mu_g = \lambda \circ f^{-1}$ exists on $\mathcal{B}$. Then for any $-\infty < a < b < \infty$, we have

$$\begin{aligned}
\mu_g((a,b]) &= \lambda\left(f^{-1}\left((a,b]\right)\right) \\
&= \lambda\left(\{x\colon a < f(x) \le f(b)\}\right) \\
&= \lambda\left(\{x\colon g(a) < x \le g(b)\}\right) \\
&= g(b) - g(a)
\end{aligned}$$

By the Uniqueness of extension for σ-finite measures, μ_g is uniquely defined.

Conversely, let ν be a Radon measure on $\mathbb{R}$. Define $g : \mathbb{R} \to \mathbb{R}$ as

$$g(y) = \begin{cases} \nu((0,y]) & \text{if } y \ge 0 \\ -\nu((y,0]) & \text{if } y < 0 \end{cases}$$

ν Radon tells us that g is finite. Easy to check g is right-continuous[a]. This is an increasing function in y, since ν is a measure. Finally, $\nu((a,b]) = g(b) - g(a)$ which can be seen by case analysis and additivity of the measure ν. By uniqueness as before, this characterises ν in its entirety. □

[a] For $y_n \downarrow y$ where $y \ge 0$, $(0, y_n] \downarrow (0, y]$ and then $\nu((0, y_n]) \downarrow \nu((0, y])$ by countably additivity. Similarly for $y < 0$.

Remark 17. Such image measures μ_g are called **Lebesgue–Stieltjes measures** associated with g, where g is the **Stieltjes distribution**.

Example 2.3

Fix $x \in \mathbb{R}$ and take $g = 1_{[x,\infty)}$. Then $\mu_g = \delta_x$ the *dirac measure at* x defined for all $A \in \mathcal{B}$ by

$$\delta_x(A) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases}$$

§2.4 Random variables

Definition 2.4 (Random Variable)

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and $(E, \mathcal{E})$ be a measurable space. If $X : \Omega \to E$ a measurable function then X is a **random variable** in E.

When $E = \mathbb{R}$ or $\mathbb{R}^d$ with the Borel σ-algebra, we simply call X a random variable or random vector.

Example 2.4

X models a "random" outcome of an experiment, e.g. when tossing a coin $\Omega = \{H, T\}, X = \#\text{ heads} : \Omega \to \{0, 1\}$.

Definition 2.5 (Distribution)

The **law** or **distribution** μ_X of a random variable X is given by the image measure $\mu_X = \mathbb{P} \circ X^{-1}$. It is a measure on $(E, \mathcal{E})$.

When $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B})$, μ_X is uniquely determined by its values on any π-system, we shall take $\{(-\infty, x] : x \in \mathbb{R}\}$ and

$$F_X(z) = \mu_X((-\infty, z]) = \mathbb{P}(X^{-1}(-\infty, z]) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq z\}) = \mathbb{P}(X \leq z)$$

The function F_x is called the **distribution function** of X, because it uniquely determines the distribution of X.

Using the properties of measures, we can show that any distribution function satisfies:

1. F_X is increasing;
2. F_X is right-continuous[3];
3. $F_X(-\infty) = \lim_{z \to -\infty} F_X(z) = \mu_X(\varnothing) = 0$;
4. $F_X(\infty) = \lim_{z \to \infty} F_X(z) = \mu_X(\mathbb{R}) = \mathbb{P}(\Omega) = 1$.

Given any function $F_X : \mathbb{R} \to [0, 1]$ satisfying each property, we can obtain a random variable X on $(\Omega, \mathcal{F}, \mathbb{P}) = ((0, 1), \mathcal{B}((0, 1)), \mu)$ by $X(\omega) = \inf\{x \mid \omega \leq f(x)\}$, and then F_X is the distribution function of X.

Definition 2.6

Consider a countable collection $(X_i \colon (\Omega, \mathcal{F}, \mathbb{P}) \to (E, \mathcal{E}))$ for $i \in I$. This collection of random variables is called *independent* if the σ-algebras $\sigma\Big(\Big\{X_i^{-1}(A) \colon A \in \mathcal{E}\Big\}\Big)$ are independent.

[3] $x_n \downarrow x \implies (-\infty, x_n] \downarrow (-\infty, x]$ hence by countable additivity of $\mathbb{P} \circ X^{-1}$.

For $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B})$ we show on an example sheet that this is equivalent to the condition

$$\mathbb{P}(X_1 \le x_1, \dots, X_n \le x_n) = \mathbb{P}(X_1 \le x_1) \dots \mathbb{P}(X_n \le x_n)$$

for all finite subsets $\{X_1, \dots, X_n\}$ of the X_i.

§2.5 Constructing independent random variables

We now construct an infinite sequence of independent random variables with prescribed distribution functions on $(\Omega, \mathcal{F}, \mathbb{P}) = ((0,1), \mathcal{B}, \mu)$ with μ the Lebesgue measure on $(0,1)$. We start with Bernoulli random variables.

Any $\omega \in (0,1)$ has a binary representation given by $(\omega_i) \in \{0,1\}^{\mathbb{N}}$, which is unique if we exclude infinitely long tails of zeroes from the binary representation. We can then define the *nth Rademacher function* $R_n(\omega) = \omega_n$ which extracts the nth bit from the binary expansion. Since each R_n can be given as the sum of 2^{n-1} indicator functions on measurable sets, they are measurable functions and are hence random variables. Their distribution is given by $\mathbb{P}(R_n = 1) = \frac{1}{2} = \mathbb{P}(R_n = 0)$, so we have constructed Bernoulli random variables with parameter $\frac{1}{2}$. We show they are independent. For a finite set $(x_i)_{i=1}^n$,

$$\mathbb{P}(R_1 = x_1, \dots, R_n = x_n) = 2^{-n} = \mathbb{P}(R_1 = x_1) \dots \mathbb{P}(R_n = x_n)$$

Therefore, the R_n are all independent, so countable sequences of independent random variables indeed exist. Now, take a bijection $m \colon \mathbb{N}^2 \to \mathbb{N}$ and define $Y_{nk} = R_{m(n,k)}$, which are independent random variables. We can now define $Y_n = \sum_k 2^{-k} Y_{nk}$. This converges for all $\omega \in \Omega$ since $|Y_{nk}| \le 1$, and these are still independent. We show the Y_n are uniform random variables, by showing the distribution coincides with the uniform distribution on the π-system of intervals $\left(\frac{i}{2^m}, \frac{i+1}{2^{m+1}}\right]$ for $i = 0, \dots, 2^m - 1$, which generates $\mathcal{B}$.

$$\mathbb{P}\left(Y_n \in \left(\frac{i}{2^m}, \frac{i+1}{2^m}\right]\right) = \mathbb{P}\left(\frac{i}{2^m} < \sum_k 2^{-k} Y_{nk} \le \frac{i+1}{2^n}\right) = 2^{-m} = \mu\left(\frac{i}{2^m}, \frac{i+1}{2^{m+1}}\right]$$

Hence $\mu_{Y_n} = \mu|_{(0,1)}$ by the uniqueness theorem, and so we have constructed an infinite sequence of independent uniform random variables Y_n. If F_n are probability distribution functions, taking the generalised inverse, we see that the $F_n^{-1}(Y_n)$ are independent and have distribution function F_n.

§2.6 Convergence of measurable functions

Definition 2.7

We say that a property defining a set $A \in \mathcal{E}$ holds μ-*almost everywhere* if $\mu(A^c) = 0$

for a measure μ on $\mathcal{E}$. If $\mu = \mathbb{P}$, we say a property holds $\mathbb{P}$-*almost surely* or *with probability one*, if $\mathbb{P}(A) = 1$.

Definition 2.8

If f_n and f are measurable functions on $(E, \mathcal{E}, \mu)$, we say f_n *converges to* f μ-*almost everywhere* if $\mu(\{x \in E \mid f_n(x) \nrightarrow f(x)\}) = 0$. We say f_n *converges to* f *in* μ-*measure* if for all $\varepsilon > 0$, $\mu(\{x \in E \mid |f_n(x) - f(x)| > \varepsilon\}) \to 0$ as $n \to \infty$. For random variables, we say $X_n \to X$ $\mathbb{P}$-*almost surely* or *in* $\mathbb{P}$-*probability*, written $X_n \to^p X$, respectively. If X_n, X take values in $\mathbb{R}$, we say $X_n \to X$ *in distribution*, written $X_n \to^d X$ if $\mathbb{P}(X_n \leq x) \to \mathbb{P}(X \leq x)$ at all points x for which the limit $x \mapsto \mathbb{P}(X \leq x)$ is continuous.

We can show that $X_n \to^p X \implies X^n \to^d X$.

Theorem 2.3

Let $f_n \colon (E, \mathcal{E}, \mu) \to \mathbb{R}$ be measurable functions. Then,

1. if $\mu(E) < \infty$, then $f_n \to 0$ almost everywhere implies that $f_n \to 0$ in measure;
2. if $f_n \to 0$ in measure, $f_{n_k} \to 0$ almost everywhere on some subsequence.

Proof. Let $\varepsilon > 0$.

$$\mu(|f_n| < \varepsilon) \geq \mu\left(\bigcap_{m \geq n} \{|f_m| \leq \varepsilon\}\right)$$

The sequence $\left(\bigcap_{m \geq n} \{|f_m| \leq \varepsilon\}\right)_n$ increases to $\bigcup_n \bigcap_{m \geq n} \{|f_m| \leq \varepsilon\}$. So by countable additivity,

$$\begin{aligned}
\mu\left(\bigcap_{m \geq n} \{|f_m| \leq \varepsilon\}\right) &\to \mu\left(\bigcup_n \bigcap_{m \geq n} \{|f_m| \leq \varepsilon\}\right) \\
&= \mu(|f_n| \leq \varepsilon \text{ eventually}) \\
&\geq \mu(|f_n| \to 0) = \mu(E)
\end{aligned}$$

Hence,

$$\liminf_n \mu(|f_n| \leq \varepsilon) \geq \mu(E) \implies \limsup_n \mu(|f_n| > \varepsilon) \leq 0 \implies \mu(|f_n| > \varepsilon) \to 0$$

For the second part, by hypothesis, we have

$$\mu\left(|f_n| > \frac{1}{k}\right) < \varepsilon$$

for sufficiently large n. So choosing $\varepsilon = \frac{1}{k^2}$, we see that along some subsequence n_k we have

$$\mu\left(|f_{n_k}| > \frac{1}{k}\right) \leq \frac{1}{k^2}$$

Hence,

$$\sum_k \mu\left(|f_{n_k}| > \frac{1}{n}\right) < \infty$$

So by the first Borel–Cantelli lemma, we have

$$\mu\left(|f_{n_k}| > \frac{1}{k} \text{ infinitely often}\right) = 0$$

so $f_{n_k} \to 0$ almost everywhere. □

Remark 18. Condition (i) is false if $\mu(E)$ is infinite: consider $f_n = 1_{(n,\infty)}$ on $(\mathbb{R}, \mathcal{B}, \mu)$, since $f_n \to 0$ almost everywhere but $\mu(f_n) = \infty$. Condition (ii) is false if we do not restrict to subsequences: consider independent events A_n such that $\mathbb{P}(A_n) = \frac{1}{n}$, then $1_{A_n} \to 0$ in probability since $\mathbb{P}(1_{A_n} > \varepsilon) = \mathbb{P}(A_n) = \frac{1}{n} \to 0$, but $\sum_n \mathbb{P}(A_n) = \infty$, and by the second Borel–Cantelli lemma, $\mathbb{P}(1_{A_n} > \varepsilon \text{ infinitely often}) = 1$, so $1_{A_n} \not\to 0$ almost surely.

Example 2.5

Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of independent exponential random variables distributed by $\mathbb{P}(X_1 \leq x) = 1 - e^{-x}$ for $x \geq 0$. Define $A_n = \{X_n \geq \alpha \log n\}$ where $\alpha > 0$, so $\mathbb{P}(A_n) = n^{-\alpha}$, and in particular, $\sum_n \mathbb{P}(A_n) < \infty$ if and only if $\alpha > 1$. By the Borel–Cantelli lemmas, we have for all $\varepsilon > 0$,

$$\mathbb{P}\left(\frac{X_n}{\log n} \geq 1 \text{ infinitely often}\right) = 1; \quad \mathbb{P}\left(\frac{X_n}{\log n} \geq 1 + \varepsilon \text{ infinitely often}\right) = 0$$

In other words, $\limsup_n \frac{X_n}{\log n} = 1$ almost surely.

§2.7 Kolmogorov's zero-one law

Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables. We can define $\mathcal{T}_n = \sigma(X_{n+1}, X_{n+2}, \dots)$. Let $\mathcal{T} = \bigcap_{n\in\mathbb{N}} \mathcal{T}_n$ be the *tail σ-algebra*, which contains all events in $\mathcal{F}$ that depend only on the limiting behaviour of (X_n).

Theorem 2.4

Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of independent random variables. Let $A \in \mathcal{T}$ be an event in the tail σ-algebra. Then $\mathbb{P}(A) = 1$ or $\mathbb{P}(A) = 0$. If $Y\colon (\Omega, \mathcal{T}) \to (\mathbb{R}, \mathcal{B})$ is measurable, it is constant almost surely.

Proof. Define $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$ to be the σ-algebra generated by the first n elements of (X_n). This is also generated by the π-system of sets $A = (X_1 \leq x_1, \dots, X_n \leq x_n)$ for any $x_i \in \mathbb{R}$. Note that the π-system of sets $B = (X_{n+1} \leq x_{n+1}, \dots, X_{n+k} \leq x_{n+k})$, for arbitrary $k \in \mathbb{N}$ and $x_i \in \mathbb{R}$, generates $\mathcal{T}_n$. By independence of the sequence, we see that $\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)$ for all such sets A, B, and so the σ-algebras $\mathcal{T}_n, \mathcal{F}_n$ generated by these π-systems are independent.

Let $\mathcal{F}_\infty = \sigma(X_1, X_2, \dots)$. Then, $\bigcup_n \mathcal{F}_n$ is a π-system that generates $\mathcal{F}_\infty$. If $A \in \bigcup_n \mathcal{F}_n$, we have $A \in \mathcal{F}_n$ for some n, so there exists $\overline{n}$ such that $B \in \mathcal{T}_{\overline{n}}$ is independent of A. In particular, $B \in \bigcap_n \mathcal{T}_n = \mathcal{T}$. By uniqueness, $\mathcal{F}_\infty$ is independent of $\mathcal{T}$.

Since $\mathcal{T} \subseteq \mathcal{F}_\infty$, if $A \in \mathcal{T}$, A is independent from A. So $\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)\,\mathbb{P}(A)$, so $\mathbb{P}(A)^2 - \mathbb{P}(A) = 0$ as required.

Finally, if $Y\colon (\Omega, \mathcal{T}) \to (\mathbb{R}, \mathcal{B})$, the preimages of $\{Y \leq y\}$ lie in $\mathcal{T}$, which give probability one or zero. Let $c = \inf\{y \mid F_Y(y) = 1\}$, so $Y = c$ almost surely. $\square$