Probability Theory

Wilhelm Stannat

Technische Universität Darmstadt

Winter Term 2007/08

Second part - corrected version

This text is a summary of the lecture on Probability Theory held at the TU Darmstadt in Winter Term 2007/08.

Please email all misprints and mistakes to

stannat@mathematik.tu-darmstadt.de

Bibliography

1. Bauer, H., *Probability theory*, de Gruyter, 1996.
2. Bauer, H., *Maß- und Integrationstheorie*, de Gruyter, 1996.
3. Billingsley, P., *Probability and Measure*, Wiley, 1995.
4. Billingsley, P., *Convergence of probability measures*, Wiley, 1999.
5. Dudley, R.M., *Real analysis and probability*, Cambridge University Press, 2002.
6. Elstrodt, J., *Maß- und Integrationstheorie*, Springer, 2005.
7. Feller, W., *An introduction to probability theory and its applications*, Vol. 1 & 2, Wiley, 1950.
8. Halmos, P.R., *Measure Theory*, Springer, 1974.
9. Klenke, A., *Wahrscheinlichkeitstheorie*, Springer, 2006.
10. Shiryaev, A.N., *Probability*, Springer, 1996.

1 Basic Notions

9 Distribution of random variables

Let $(\Omega, \mathcal{A}, P)$ be a probability space, and $X : \Omega \to \bar{\mathbb{R}}$ be a r.v.

Let μ be the *distribution of X (under P)*, i.e., $\mu(A) = P[X \in A]$ for all $A \in \mathcal{B}(\bar{\mathbb{R}})$.

Assume that $P[X \in \mathbb{R}] = 1$ (in particular, X P-a.s. finite, and μ is a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Definition 9.1. The function $F : \mathbb{R} \to [0,1]$, defined by

$$F(b) := P[X \leqslant b] = \mu\big(]-\infty, b]\big), \quad b \in \mathbb{R}, \tag{1.1}$$

is called the *distribution function* of X resp. μ.

Proposition 9.2. *(i) F is monotone increasing:* $a \leqslant b \quad \Rightarrow \quad F(a) \leqslant F(b)$

right continuous: $F(a) = \lim_{b \searrow a} F(b)$

normalized: $\lim_{a \searrow -\infty} F(a) = 0, \quad \lim_{b \nearrow +\infty} F(b) = 1.$

(ii) To any such function there exists a unique probability measure μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *with* (1.10).

Proof. (i) Monotonicity is obvious.

Right continuity: if $b \searrow a$ then $]-\infty, b] \searrow]-\infty, a]$, hence by continuity of μ from above (vgl. Proposition 1.9):

$$F(a) = \mu\big(]-\infty, a]\big) \overset{1.9}{=} \lim_{b \searrow a} \mu\big(]-\infty, b]\big) = \lim_{b \searrow a} F(b).$$

Similarly, $]-\infty, a] \searrow \emptyset$ if $a \searrow -\infty$ (resp. $]-\infty, b] \nearrow \mathbb{R}$ if $b \nearrow \infty$), and thus

$$\lim_{a \searrow -\infty} F(a) = \lim_{a \searrow -\infty} \mu\big(]-\infty, a]\big) = 0$$

(resp. $\lim_{b \nearrow \infty} F(b) = \lim_{b \nearrow \infty} \mu\big(]-\infty, b]\big) = 1$).

(ii) *Existence:* Let λ be the Lebesgue measure on $]0,1[$. Define the "inverse function" G of $F : \mathbb{R} \to [0,1]$ by

$$G :]0,1[\to \mathbb{R}$$
$$G(y) := \inf\big\{x \in \mathbb{R} \;\big|\; F(x) > y\big\}.$$

Note that $y < F(x) \quad \Rightarrow \quad G(y) \leqslant x$ implies

$$\left]0, F(x)\right[\subset \{G \leqslant x\}$$

and $G(y) \leqslant x \quad \Rightarrow \quad \exists\, x_n \searrow x$ with $F(x_n) > y$, hence $F(x) \geqslant y$, so that

$$\{G \leqslant x\} \subset \left]0, F(x)\right].$$

Combining both inclusions we obtain that

$$\left]0, F(x)\right[\subset \{G \leqslant x\} \subset \left]0, F(x)\right].$$

so that G is measurable.

Let $\mu := G(\lambda) = \lambda \circ G^{-1}$ (probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$). Then

$$\mu(]-\infty, x]) = \lambda(\{G \leqslant x\}) = \lambda\Big(\left]0, F(x)\right]\Big) = F(x) \qquad \forall\, x \in \mathbb{R}.$$

Uniqueness: later. □

Remark 9.3. *(i) Let Y be a r.v. with uniform distribution on $[0,1]$, then $X = G(Y)$ has distribution μ. In particular: simulating the uniform distribution on $[0,1]$ gives by transformation with G a simulation of μ.*

(ii) Some authors define the distribution function F by $F(x) := \mu(]-\infty, x[)$. In this case F is left continuous, not right continuous.

Remark 9.4. *(i) Let F be a distribution function and let $x \in \mathbb{R}$: Then*

$$F(x) - F(x-) = \lim_{n \nearrow \infty} \mu\left(\left]x - \frac{1}{n}, x\right]\right) = \mu(\{x\})$$

is called the step height of F in x. *In particular:*

$$F \text{ continuous} \quad \Leftrightarrow \quad \forall\, x \in \mathbb{R} : \mu(\{x\}) = 0 \quad \text{“}\mu \text{ is continuous”}.$$

(ii) Let F be monotone increasing and bounded, then F has at most countable many points of discontinuity.

Definition 9.5. (i) F (resp. μ) is called *discrete*, if there exists a countable set $S \subset \mathbb{R}$ with $\mu(S) = 1$. In this case, μ is uniquenely determined by the weights $\mu(\{x\})$, $x \in S$, and F is a *step function* of the following type:

$$F(x) = \sum_{\substack{y \in S, \\ y \leqslant x}} \mu(\{y\}).$$

(ii) F (resp. μ) is called *absolutely continuous*, if there exists a measurbale function $f \geqslant 0$ (called the "*density*"), such that

$$F(x) = \int_{-\infty}^{x} f(t)\, \mathrm{d}t, \tag{1.2}$$

resp., for all $A \in \mathcal{B}(\mathbb{R})$:

$$\mu(A) = \int_A f(t)\,\mathrm{d}t = \int_{-\infty}^{\infty} 1_A \cdot f\,\mathrm{d}t. \tag{1.3}$$

In particular $\displaystyle\int_{-\infty}^{+\infty} f(t)\,\mathrm{d}t = 1.$

Remark 9.6. *(i) Every measurable function $f \geqslant 0$ with $\int_{-\infty}^{+\infty} f(t)\,\mathrm{d}t = 1$ defines a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ by $A \mapsto \int_A f(t)\,\mathrm{d}t$.*

(ii) In the previous definition "(1.11)⇒(1.12)", because $A \mapsto \int_A f(t)\,\mathrm{d}t$ defines a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with distribution function F. Uniqueness in 9.2 implies the assertion.

Example 9.7. (i) **Uniform distribution on $[a, b]$.** Let $f := \frac{1}{b-a} \cdot 1_{[a,b]}$. The associated distribution function is given by

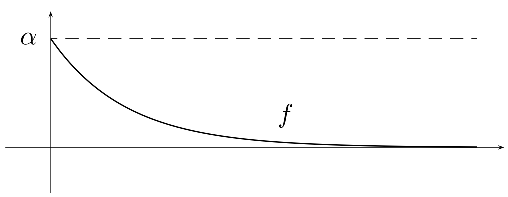
$$F(x) := \begin{cases} 0 & \text{if } x \leqslant a \\ \frac{1}{b-a} \cdot (x-a) & \text{if } x \in [a,b] \\ 1 & \text{if } x \geqslant b. \end{cases}$$

(continuous analogue to the dicrete uniform distribution on a finite set)

(ii) **Exponential distribution with parameter $\alpha > 0$.**

$$f(x) := \begin{cases} \alpha e^{-\alpha x} & \text{if } x \geqslant 0 \\ 0 & \text{if } x < 0, \end{cases}$$

$$F(x) := \begin{cases} 1 - e^{-\alpha x} & \text{if } x \geqslant 0 \\ 0 & \text{if } x < 0. \end{cases}$$



(continuous analogue of the geometric distribution)

$$\int_k^{k+1} f(x)\,\mathrm{d}x = F(k+1) - F(k) = e^{-\alpha k}(1 - e^{-\alpha}) = (1-p)^k p \text{ with } p = 1 - e^{-\alpha}.$$

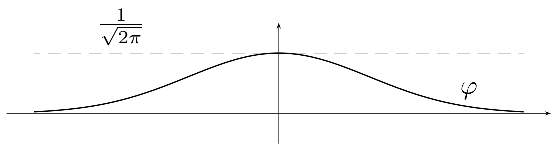
(iii) **Normal distribution $N(m, \sigma^2)$, $m \in \mathbb{R}$, $\sigma^2 > 0$**

$$f_{m,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

The associated distribution function is given by

$$F_{m,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \int_{-\infty}^{x} e^{-\frac{(y-m)^2}{2\sigma^2}}\, \mathrm{d}y$$
$$\overset{z=\frac{y-m}{\sigma}}{=} \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{\frac{x-m}{\sigma}} e^{-\frac{z^2}{2}}\, \mathrm{d}z = F_{0,1}\left(\frac{x-m}{\sigma}\right).$$

$\Phi := F_{0,1}$ is called the distribution function of the *standard normal distribution* $N(0,1)$.



The expectation $E[X]$ (or more general $E[h(X)]$) can be calculated with the help of the distribution μ of X:

Proposition 9.8. *Let $h \geqslant 0$ be measurable, then*

$$\mathbb{E}\big[h(X)\big] = \int_{-\infty}^{+\infty} h(x)\, \mu(\mathrm{d}x)$$
$$= \begin{cases} \int_{-\infty}^{+\infty} h(x) \cdot f(x)\, \mathrm{d}x & \text{if } \mu \text{ absolutely continuous with density } f \\ \sum_{x \in S} h(x) \cdot \mu(\{x\}) & \text{if } \mu \text{ discrete, } \mu(S) = 1 \text{ and } S \text{ countable.} \end{cases}$$

Proof. See exercises. □

Example 9.9. Let X be $N(m,\sigma^2)$-distributed. Then

$$\mathbb{E}[X] = \int x \cdot f_{m,\sigma^2}(x)\, \mathrm{d}x = m + \underbrace{\int (x-m) \cdot f_{m,\sigma^2}(x)\, \mathrm{d}x}_{=0} = m\,.$$

The p^{th} *central moment of* X is given by

$$\mathbb{E}\big[|X-m|^p\big] = \int |x-m|^p \cdot f_{m,\sigma^2}(x)\, \mathrm{d}x,$$
$$= \int |x|^p \cdot f_{0,\sigma^2}(x)\, \mathrm{d}x.$$
$$= 2\int_0^{\infty} x^p \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{x^2}{2\sigma^2}}\, \mathrm{d}x,$$
$$\underbrace{=}_{y=\frac{x^2}{2\sigma^2}} \frac{1}{\sqrt{\pi}} \cdot 2^{\frac{p}{2}} \cdot \sigma^p \underbrace{\int_0^{\infty} y^{\frac{p+1}{2}-1} \cdot e^{-y}\, \mathrm{d}y}_{=\Gamma(\frac{p+1}{2})}.$$

In particular:

$$p=1:\mathbb{E}\big[|X-m|\big]=\sigma\cdot\sqrt{\frac{2}{\pi}}$$
$$p=2:\mathbb{E}\big[|X-m|^2\big]=\sigma^2$$
$$p=3:\mathbb{E}\big[|X-m|^3\big]=2^{\frac{3}{2}}\cdot\frac{\sigma^3}{\sqrt{\pi}}$$
$$p=4:\mathbb{E}\big[|X-m|^4\big]=3\sigma^4.$$

10 Weak convergence of probability measures

Let S be a topological space and $\mathcal{S}$ be the Borel σ-algebra on S.

Let μ, μ_n, $n\in\mathbb{N}$, be probability measures on $(S,\mathcal{S})$.

What is a reasonable notion of convergence of the sequence μ_n towards μ? The notion of "pointwise convergence" in the sense that $\mu_n(A)\xrightarrow{n\to\infty}\mu(A)$ for all $A\in\mathcal{S}$ is too strong for many applications.

Definition 10.1. Let μ and μ_n, $n\in\mathbb{N}$, be probability measures on $(S,\mathcal{S})$. The sequence (μ_n) converges to μ *weakly* if for all $f\in C_b(S)$ (= the space of bounded continuous functions on S) it follows that

$$\int f\,\mathrm{d}\mu_n\xrightarrow{n\to\infty}\int f\,\mathrm{d}\mu.$$

Example 10.2. (i) $x_n\xrightarrow{n\to\infty}x$ in S implies $\delta_{x_n}\xrightarrow{n\to\infty}\delta_x$ weakly.

(ii) Let $S:=\mathbb{R}^1$ and $\mu_n:=N\big(0,\frac{1}{n}\big)$. Then $\mu_n\to\delta_0$ weakly, since for all $f\in\mathbb{C}_b(\mathbb{R})$

$$\begin{aligned}\int f\,\mathrm{d}\mu_n&=\int f(x)\cdot\frac{1}{\sqrt{2\pi\frac{1}{n}}}\cdot e^{-\frac{x^2}{2\cdot\frac{1}{n}}}\,\mathrm{d}x\\ &\overset{x=\frac{y}{\sqrt{n}}}{=}\int f\Big(\frac{y}{\sqrt{n}}\Big)\cdot\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{y^2}{2}}\,\mathrm{d}y\\ &\xrightarrow[\text{Lebesgue}]{n\to\infty}f(0)\quad=\quad\int f\,\mathrm{d}\delta_0.\end{aligned}$$

Proposition 10.3 (Portemanteau-Theorem)**.** *Let S be a metric space with metric d. Then the following statements are equivalent:*

(i) $\mu_n\to\mu$ *weakly*

(ii) $\int f\,\mathrm{d}\mu_n\xrightarrow{n\to\infty}\int f\,\mathrm{d}\mu$ *for all f bounded and uniformly continuous (w.r.t. d)*

(iii) $\limsup_{n\to\infty}\mu_n(F)\leqslant\mu(F)$ *for all $F\subset S$ closed*

(iv) $\liminf_{n\to\infty}\mu_n(G)\geqslant\mu(G)$ *for all $G\subset S$ open*

(v) $\lim_{n\to\infty}\mu_n(A)=\mu(A)$ *for all* $A\in\mathcal{S}$ *with* $\mu(\bar{A}\setminus\mathring{A})=0$.

Proof. **(iii)⇔(iv):** Obvious by considering the complement.

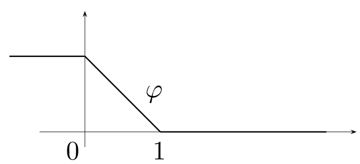
(i)⇒(ii): Trivial.

(ii)⇒(iii): Let $F\subset S$ be closed, let

$$G_m := \left\{ x\in S \,\middle|\, d(x,F)<\frac{1}{m}\right\},\quad m\in\mathbb{N} \qquad \text{open!}$$

Then $G_m \searrow F$, hence $\mu(G_m)\searrow\mu(F)$.
If $\varepsilon>0$ there exists some $m\in\mathbb{N}$ mit $\mu(G_m)<\mu(F)+\varepsilon$. Define

$$\varphi(x) := \begin{cases} 1 & \text{if } x\leqslant 0\\ 1-x & \text{if } x\in[0,1]\\ 0 & \text{if } x\geqslant 1.\end{cases}$$



and let $f:=\varphi\big(m\cdot d(\,\cdot\,,F)\big)$.
f is Lipschitz, in particular uniformly continuous, $f=0$ on G_m^c and $f=1$ on F, and thus

$$\begin{aligned}\limsup_{n\to\infty}\mu_n(F)\leqslant\limsup_{n\to\infty}\int f\,d\mu_n &\overset{\text{(ii)}}{=}\int f\,d\mu\\ &\leqslant \mu(G_m)<\mu(F)+\varepsilon.\end{aligned}$$

(iii)⇒(v): Let A be such that $\mu(\bar{A}\setminus\mathring{A})=0$. Then

$$\begin{aligned}\mu(A)=\mu(\mathring{A}) &\overset{\text{(iv)}}{\leqslant}\liminf_{n\to\infty}\mu_n(\mathring{A})\leqslant\liminf_{n\to\infty}\mu_n(A)\leqslant\limsup_{n\to\infty}\mu_n(A)\\ &\leqslant\limsup_{n\to\infty}\mu_n(\bar{A})\overset{\text{(iii)}}{\leqslant}\mu(\bar{A})=\mu(A).\end{aligned}$$

(v)⇒(iii): Let $F\subset S$ be closed. For all $\delta>0$ we have that

$$\partial\big\{d(\,\cdot\,,F)\geqslant\delta\big\}\subset\big\{d(\,\cdot\,,F)=\delta\big\}.$$

Note The set

$$D:=\left\{\delta>0 \,\middle|\, \mu\big(\{d(\,\cdot\,,F)=\delta\}\big)>0\right\}$$

is countable, since for all n the set

$$D_n := \left\{ \delta > 0 \;\middle|\; \mu\big(\underbrace{\{d(\,\cdot\,, F) = \delta\}}_{\text{disjoint!}}\big) > \frac{1}{n} \right\}$$

is finite for any $n \in \mathbb{N}$. In particular, there exists a sequence $\delta_k \in \,]0,\infty[\, \setminus D$, $\delta_k \downarrow 0$ such that the set

$$F_k := \big\{ d(\,\cdot\,, F) \leqslant \delta_k \big\}$$

satisfies $\mu(\bar{F}_k \setminus \mathring{F}_k) = 0$. $F_k \searrow F$ now implies that

$$\limsup_{n\to\infty} \mu_n(F) \leqslant \limsup_{n\to\infty} \mu_n(F_k) \overset{\text{(v)}}{=} \mu(F_k) \xrightarrow{k\to\infty} \mu(F).$$

(iii)⇒(i): Let $f \in C_b(S)$. It suffices to prove that

$$\limsup_{n\to\infty} \int f \, d\mu_n \leqslant \int f \, d\mu,$$

(since then

$$-\liminf \int f \, d\mu_n \leqslant \int (-f) \, \mathrm{d}\mu,$$

hence $\liminf \int f \, d\mu_n \geqslant \int f \, \mathrm{d}\mu$)

W.l.o.g. $0 \leqslant f \leqslant 1$

Fix $k \in \mathbb{N}$ and let $F_j := \left\{ f \geqslant \frac{j}{k} \right\}$, $j \in \mathbb{N}$ (F_j closed!)

Then

$$\frac{1}{k} \sum_{i=1}^{k} 1_{F_i} \leqslant f \leqslant \frac{1}{k} + \frac{1}{k} \sum_{i=1}^{k} 1_{F_i}$$

Hence for all probability measures ν on $(S, \mathcal{S})$:

$$\frac{1}{k} \sum_{i=1}^{k} \nu(F_i) \underset{\dagger}{\leqslant} \int f \, d\nu \underset{\ddagger}{\leqslant} \frac{1}{k} + \frac{1}{k} \sum_{i=1}^{k} \nu(F_i).$$

and

$$\begin{aligned} \limsup_{n\to\infty} \int f \, \mathrm{d}\mu_n - \frac{1}{k} &\overset{(\ddagger)}{\leqslant} \frac{1}{k} \cdot \limsup_{n\to\infty} \sum_{i=1}^{k} \mu_n(F_i) \\ &\leqslant \frac{1}{k} \sum_{i=1}^{k} \limsup_{n\to\infty} \mu_n(F_i) \overset{\text{(iii)}}{\leqslant} \frac{1}{k} \sum_{i=1}^{k} \mu(F_i) \overset{(\dagger)}{\leqslant} \int f \, d\mu \end{aligned}$$

□

Corollary 10.4. *Let X, X_n, $n \in \mathbb{N}$, be measurable mappings from $(\Omega, \mathcal{A}, P)$ to $(S, \mathcal{S})$ with distributions μ, μ_n, $n \in \mathbb{N}$. Then:*

$$X_n \xrightarrow{n\to\infty} X \quad \text{in probability} \quad \Rightarrow \quad \mu_n \xrightarrow{n\to\infty} \mu \quad \text{weakly}$$

Here, $\lim_{n\to\infty} X_n = X$ in probability, if $\lim_{n\to\infty} P(d(X, X_n) > \delta) = 0$ for all $\delta > 0$.

Proof. Let $f \in C_b(S)$ be uniformly continuous and $\varepsilon > 0$. Then there exists a $\delta = \delta(\varepsilon) > 0$ such that:
$x, y \in S$ with $d(x,y) \leqslant \delta$ implies $|f(x) - f(y)| < \varepsilon$
Hence

$$\begin{aligned}
&\left| \int f \,\mathrm{d}\mu - \int f \,\mathrm{d}\mu_n \right| = \Big| \mathbb{E}\big[f(X)\big] - \mathbb{E}\big[f(X_n)\big] \Big| \\
&\leqslant \int_{\{d(X,X_n)\leqslant\delta\}} \big|f(X) - f(X_n)\big| \,\mathrm{d}P + \int_{\{d(X,X_n)>\delta\}} \big|f(X) - f(X_n)\big| \,\mathrm{d}P \\
&\leqslant \varepsilon + 2\|f\|_\infty \cdot \underbrace{P\big[d(X_n, X) > \delta\big]}_{\xrightarrow{n\to\infty} 0}.
\end{aligned}$$

□

Corollary 10.5. *Let $S = \mathbb{R}^1$ and let μ, μ_n, $n \in \mathbb{N}$, be probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with distributions functions F, F_n. Then the following statements are equivalent:*

(i) $\mu_n \xrightarrow{n\to\infty} \mu$ *vaguely, i.e.* $\lim_{n\to\infty} \int f \, d\mu_n = \int f \, d\mu$ *for all* $f \in \mathbb{C}_0(\mathbb{R}^1)$ *(= the space of continuous functions with compact support)*

(ii) $\mu_n \xrightarrow{n\to\infty} \mu$ *weakly*

(iii) $F_n(x) \xrightarrow{n\to\infty} F(x)$ *for all x where F is continuous.*

(iv) $\mu_n\big(]a,b]\big) \xrightarrow{n\to\infty} \mu\big(]a,b]\big)$ *for all $]a,b]$ with $\mu(\{a\}) = \mu(\{b\}) = 0$.*

Proof. **(i)⇒(ii):** Exercise.

(ii)⇒(iii): Let x be such that F is continuous in x. Then $\mu(\{x\}) = 0$, which implies by the Portmanteau theorem:

$$F_n(x) = \mu_n\big(]-\infty, x]\big) \xrightarrow{n\to\infty} \mu\big(]-\infty, x]\big) = F(x).$$

(iii)⇒(iv): Let $]a,b]$ be such that $\mu(\{a\}) = \mu(\{b\}) = 0$ then F is continuous in a and b and thus

$$\begin{aligned}
\mu\big(]a,b]\big) = F(b) - F(a) &\overset{\text{(iii)}}{=} \lim_{n\to\infty} F_n(b) - \lim_{n\to\infty} F_n(a) \\
&= \lim_{n\to\infty} \mu_n\big(]a,b]\big).
\end{aligned}$$

(iv)⇒(i): Let $D := \{x \in \mathbb{R} \mid \mu(\{x\}) = 0\}$. Then $\mathbb{R} \setminus D$ is countable, hence $D \subset \mathbb{R}$ dense. Let $f \in C_0(\mathbb{R})$, then f is uniformly continuous, hence for $\varepsilon > 0$ we find $c_0 < \cdots < c_m \in D$ such that

$$\Big\| \underbrace{f - \sum_{k=1}^{m} f(c_{k-1}) \cdot \mathbb{I}_{]c_{k-1},c_k]}}_{=:g} \Big\|_\infty \leqslant \sup_k \sup_{x \in [c_{k-1},c_k]} \big| f(x) - f(c_{k-1}) \big| < \varepsilon.$$

Then

$$\begin{aligned} &\left| \int f \,\mathrm{d}\mu - \int f \,\mathrm{d}\mu_n \right| \\ &\leqslant \underbrace{\int |f-g| \,\mathrm{d}\mu}_{<\varepsilon} + \left| \int g \,\mathrm{d}\mu - \int g \,\mathrm{d}\mu_n \right| + \underbrace{\int |f-g| \,\mathrm{d}\mu_n}_{<\varepsilon} \\ &\leqslant 2\varepsilon + \sum_{k=1}^{m} f(c_{k-1}) \cdot \Big| \mu\big(]c_{k-1}, c_k]\big) - \mu_n\big(]c_{k-1}, c_k]\big) \Big| \xrightarrow[n\to\infty]{\text{(iv)}} 2\varepsilon. \end{aligned}$$

□

11 Dynkin-systems and Uniqueness of probability measures

Let $\Omega \neq \emptyset$.

Definition 11.1. A collection of subsets $\mathcal{D} \subset \mathcal{P}(\Omega)$ is called a *Dynkin-system*, if:

(i) $\Omega \in \mathcal{D}$.

(ii) $A \in \mathcal{D} \quad \Rightarrow \quad A^c \in \mathcal{D}$.

(iii) $A_i \in \mathcal{D}$, $i \in \mathbb{N}$, pairwise disjoint, then

$$\dot{\bigcup_{i \in \mathbb{N}}} A_i \in \mathcal{D}.$$

Example 11.2. (i) Every σ-Algebra $\mathcal{A} \subset \mathcal{P}(\Omega)$ is a Dynkin-system

(ii) Let P_1, P_2 be probability measures on $(\Omega, \mathcal{A})$. Then

$$\mathcal{D} := \{A \in \mathcal{A} \mid P_1(A) = P_2(A)\}$$

is a Dynkin-system

Remark 11.3. *(i) Let $\mathcal{D}$ be a Dynkin-system. Then*

$$A, B \in \mathcal{D}, A \subset B \qquad \Rightarrow \qquad B \setminus A = (B^c \,\dot\cup\, A)^c \in \mathcal{D}$$

(ii) *Every Dynkin-system which is closed under finite unions (short notation: $\cap$-stable), is a σ-algebra, because:*

(a) $$A, B \in \mathcal{D} \quad \Rightarrow \quad A \cup B = A \,\dot{\cup}\, \underbrace{\big(B \setminus \underbrace{(A \cap B)}_{\in \mathcal{D}}\big)}_{\overset{\textit{(i)}}{\in} \mathcal{D}} \in \mathcal{D}.$$

(b) $$A_i \in \mathcal{D},\ i \in \mathbb{N} \quad \Rightarrow \quad \bigcup_{i \in \mathbb{N}} A_i = \dot{\bigcup_{i \in \mathbb{N}}} \underbrace{\Big[A_i \cap \Big(\underbrace{\bigcup_{n=1}^{i-1} A_n}_{\overset{\textit{(a)}}{\in} \mathcal{D}}\Big)^c\Big]}_{\substack{\in \mathcal{D} \text{ by ass.,} \\ \text{pairwise disjoint}}} \in \mathcal{D}.$$

Proposition 11.4. *Let $\mathcal{B} \subset \mathcal{P}(\Omega)$ be a $\cap$-stable collection of subsets. Then*

$$\sigma(\mathcal{B}) = \mathcal{D}(\mathcal{B}),$$

where

$$\mathcal{D}(\mathcal{B}) := \bigcap_{\substack{\mathcal{D} \text{ Dynkin-system} \\ \mathcal{B} \subset \mathcal{D}}} \mathcal{D}$$

is called the Dynkin-system generated by $\mathcal{B}$.

Proof. See text books on measure theory. □

Proposition 11.5 (Uniqueness of probability measures)**.** *Let P_1, P_2 be probability measures on $(\Omega, \mathcal{A})$, and $\mathcal{B} \subset \mathcal{A}$ be a $\cap$-stable collection of subsets. Then:*

$$P_1(A) = P_2(A) \ \text{ for all } \ A \in \mathcal{B} \quad \Rightarrow \quad P_1 = P_2 \ \text{ on } \sigma(\mathcal{B}).$$

Proof. The collection of subsets

$$\mathcal{D} := \big\{A \in \mathcal{A} \ \big|\ P_1(A) = P_2(A)\big\}$$

is a Dynkin-system containing $\mathcal{B}$. Consequently,

$$\sigma(\mathcal{B}) \overset{11.4}{=} \mathcal{D}(\mathcal{B}) \subset \mathcal{D}.$$ □

Example 11.6. (i) For $p \in\,]0, 1[$ the probability measure P_p on $(\Omega := \{0,1\}^{\mathbb{N}}, \mathcal{A})$ is uniquely determined by

$$P_p[X_1 = x_1, \dots, X_n = x_n] = p^k (1-p)^{n-k}, \quad \text{with } k := \sum_{i=1}^n x_i$$

for all $x_1, \dots, x_n \in \{0,1\}$, $n \in \mathbb{N}$, because the collection of cylindrical sets

$$\{X_1 = x_1, \dots, X_n = x_n\}, \quad n \in \mathbb{N}_0,\ x_1, \dots, x_n \in \{0,1\}$$

is $\cap$-stable, generating $\mathcal{A}$ (cf. Example 1.7).

(Existence of P_p for $p = \frac{1}{2}$ see Example 3.6. Existence for $p \in\,]0,1[\, \setminus \{\frac{1}{2}\}$ later.)

(ii) A probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is uniquely determined through its distribution function F $(:= \mu(]-\infty, \cdot\,]))$, because

$$\mu(]a, b]) = F(b) - F(a),$$

and the collection of intervals $]a, b]$, $a, b \in \mathbb{R}$, is $\cap$-stable, generating $\mathcal{B}(\mathbb{R})$.

2 Independence

1 Independent events

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

Definition 1.1. A collection of events $A_i \in \mathcal{A}$, $i \in I$, are said to be *independent* (w.r.t. P), if for any finite subset $J \subset I$

$$P\Big(\bigcap_{j\in J} A_j\Big) = \prod_{j\in J} P(A_j).$$

A family of collection of subsets $\mathcal{B}_i \subset \mathcal{A}$, $i \in I$, is said to be *independent*, if for all finite subsets $J \subset I$ and for all subsets $A_j \in \mathcal{B}_j$, $j \in J$

$$P\Big(\bigcap_{j\in J} A_j\Big) = \prod_{j\in J} P(A_j).$$

Proposition 1.2. *Let $\mathcal{B}_i$, $i \in I$, be independent and closed under intersections. Then:*

(i) $\sigma(\mathcal{B}_i)$, $i \in I$, are independent.

(ii) Let J_k, $k \in K$, be a partition of the index set I. Then the σ-algebras

$$\sigma\Big(\bigcup_{i\in J_k} \mathcal{B}_i\Big), \quad k \in K,$$

are independent.

Proof. (i) Let $J \subset I$, J finite, be of the form $J = \{j_1, \dots, j_n\}$. Let $A_{j_1} \in \sigma(\mathcal{B}_{j_1}), \dots, A_{j_n} \in \sigma(\mathcal{B}_{j_n})$.

We have to show that

$$P(A_{j_1} \cap \cdots \cap A_{j_n}) = P(A_{j_1}) \cdots P(A_{j_n}). \tag{2.1}$$

To this end suppose first that $A_{j_2} \in \mathcal{B}_{j_2}, \dots, A_{j_n} \in \mathcal{B}_{j_n}$, and define

$$\mathcal{D}_{j_1} := \big\{A \in \sigma(\mathcal{B}_{j_1}) \bigm| P(A \cap A_{j_2} \cap \cdots \cap A_{j_n}) = P(A) \cdot P(A_{j_2}) \cdots P(A_{j_n})\big\}.$$

Then $\mathcal{D}_{j_1}$ is a Dynkin system (!) containing $\mathcal{B}_{j_1}$. Proposition 1.11.4 now implies

$$\sigma(\mathcal{B}_{j_1}) = \mathcal{D}(\mathcal{B}_{j_1}) \subset \mathcal{D}_{j_1}\ ,$$

hence $\sigma(\mathcal{B}_{j_1}) = \mathcal{D}_{j_1}$. Iterating the above argument for $\mathcal{D}_{j_2}$, $\mathcal{D}_{j_3}$, implies (2.1).

(ii) For $k \in K$ define

$$\mathcal{C}_k := \Big\{ \bigcap_{j \in J} A_j \ \Big| \ J \subset J_k,\ J \text{ finite},\ A_j \in \mathcal{B}_j \Big\}.$$

Then $\mathcal{C}_k$ is closed under intersections and the collection of subsets $\mathcal{C}_k$, $k \in K$, are still independent, because: given $k_1, \dots, k_n \in K$ and finite subsets $J^1 \subset J_{k_1}, \dots, J^n \subset J_{k_n}$, then

$$P\Big(\underbrace{\Big(\bigcap_{i \in J^1} A_i \Big)}_{\in \mathcal{C}_{k_1}} \cap \dots \cap \underbrace{\Big(\bigcap_{i \in J^n} A_i \Big)}_{\in \mathcal{C}_{k_n}} \Big) \overset{\mathcal{B}_i, i \in I}{\underset{\text{ind.}}{=}} \prod_{j=1}^{n} P\Big(\bigcap_{i \in J^j} A_i \Big).$$

(i) now implies that

$$\sigma(\mathcal{C}_k) = \sigma\Big(\bigcup_{i \in J_k} \mathcal{B}_i \Big), \quad k \in K,$$

are independent too. □

Example 1.3. Let $A_i \in \mathcal{A}$, $i \in I$, be independent. Then A_i, A_i^c, $i \in I$, are independent too.

Remark 1.4. *Pairwise independence does not imply independence in general.*
Beispiel: *Consider two tosses with a fair coin, i.e.*

$$\Omega := \big\{ (i,k) \ \big| \ i,k \in \{0,1\} \big\}, \quad P := \textit{uniform distribution}.$$

Consider the events

$$A := \textit{"1. toss 1"} = \big\{ (1,0),\ (1,1) \big\}$$
$$B := \textit{"2. toss 1"} = \big\{ (0,1),\ (1,1) \big\}$$
$$C := \textit{"1. and 2. toss equal"} = \big\{ (0,0),\ (1,1) \big\}.$$

Then $P(A) = P(B) = P(C) = \frac{1}{2}$ and A, B, C are pairwise independent

$$P(A \cap B) = P(B \cap C) = P(C \cap A) = \frac{1}{4}.$$

But on the other hand

$$P(A \cap B \cap C) = 14 \neq P(A) \cdot P(B) \cdot P(C).$$

Example 1.5. Independent 0-1-experiments with success probability $p \in [0,1]$. Let $\Omega := \{0,1\}^{\mathbb{N}}$, $X_i(\omega) := x_i$ and $\omega := (x_i)_{i \in \mathbb{N}}$. Let P_p be a probability measure on $\mathcal{A} := \sigma\big(\{X_i = 1\},\ i = 1, 2, \dots \big)$, with

(i) $P_p[X_i = 1] = p$ (hence $P_p[X_i = 0] = P_p\big(\{X_i = 1\}^c \big) = 1 - p$).

(ii) $\{X_i = 1\}$, $i \in \mathbb{N}$, are *independent* w.r.t. P_p.

Existence of such a probability measure later! Then for any $x_1, \dots, x_n \in \{0,1\}$:

$$P_p[X_{i_1} = x_1, \dots, X_{i_n} = x_n] \overset{\text{(ii) and 1.3}}{=} \prod_{j=1}^{n} P_p[X_{i_j} = x_j] \overset{\text{(i)}}{=} p^k (1-p)^{n-k},$$

where $k := \sum_{i=1}^n x_i$ gilt. Hence P_p is uniquely determined by (i) and (ii).

Proposition 1.6 (Kolmogorov's Zero-One Law)**.** *Let $\mathcal{B}_n$, $n \in \mathbb{N}$, be independent σ-algebras, and*

$$\mathcal{B}_\infty := \bigcap_{n=1}^{\infty} \sigma\Big(\bigcup_{m=n}^{\infty} \mathcal{B}_m\Big)$$

be the tail-field *(resp.* σ-algebra of terminal events*). Then*

$$P(A) \in \{0,1\} \qquad \forall\, A \in \mathcal{B}_\infty$$

i.e., P is deterministic on $\mathcal{B}_\infty$.

Illustration: Independent 0-1-experiments
Let $\mathcal{B}_i = \sigma\big(\{X_i = 1\}\big)$. Then

$$\mathcal{B}_\infty = \bigcap_{n \in \mathbb{N}} \sigma\Big(\bigcup_{m \geqslant n} \mathcal{B}_m\Big)$$

is the σ-algebra containing the events of the remote future, e.g.

$$\limsup_{i \to \infty} \{X_i = 1\} = \{\text{``infinitely many `1'''}\}$$

$$\Big\{\omega \in \{0,1\}^{\mathbb{N}} \;\Big|\; \lim_{n \to \infty} \underbrace{\frac{1}{n} \sum_{i=1}^{n} X_i(\omega)}_{=: \frac{S_n(\omega)}{n}} \quad \text{exists}\Big\}$$

Proof of the Zero-One Law. Proposition 1.2 implies that for all n

$$\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_{n-1}, \sigma\Big(\bigcup_{m=n}^{\infty} \mathcal{B}_m\Big)$$

are independent. Since $\mathcal{B}_\infty \subset \sigma\Big(\bigcup_{m \geqslant n} \mathcal{B}_m\Big)$, this implies that for all n

$$\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_{n-1}, \mathcal{B}_\infty$$

are independent. By definition this implies that

$$\mathcal{B}_\infty\,, \mathcal{B}_n\,, n \in \mathbb{N} \text{ are independent}$$

and now Proposition 1.2(ii) implies that

$$\sigma\Big(\bigcup_{n\in\mathbb{N}} \mathcal{B}_n\Big) \quad \text{und} \quad \mathcal{B}_\infty$$

are idependent. Since $\mathcal{B}_\infty \subset \sigma\Big(\bigcup_{n\geqslant 1} \mathcal{B}_n\Big)$ we finally obtain that $\mathcal{B}_\infty$ and $\mathcal{B}_\infty$ are independent. The conclusion now follows from the next lemma. □

Lemma 1.7. *Let $\mathcal{B} \subset \mathcal{A}$ be a σ-algebra such that $\mathcal{B}$ is independent from $\mathcal{B}$. Then*

$$P(A) \in \{0,1\} \qquad \forall\, A \in \mathcal{B}\,.$$

Proof. For all $A \in \mathcal{B}$

$$P(A) = P(A \cap A) = P(A) \cdot P(A) = P(A)^2.$$

Hence $P(A) = 0$ or $P(A) = 1$. □

For any sequence A_n, $n \in \mathbb{N}$, of independent events in $\mathcal{A}$, Kolmogorov's Zero-One Law implies in particular for

$$A_\infty := \bigcap_{n\in\mathbb{N}} \bigcup_{m\geqslant n} A_m \quad \big(=: \limsup_{n\to\infty} A_n\big)$$

that $P(A_\infty) = 0 - 1$.
Proof: The σ-algebras $\mathcal{B}_n := \sigma\{A_n\} = \{\emptyset, \Omega, A, A^c\}$, $n \in \mathbb{N}$, are independent by Proposition 1.2 and $A_\infty \in \mathcal{B}_\infty$.

Lemma 1.8 (Borel-Cantelli)**.** *(i) Let $A_i \in \mathcal{A}$, $i \in \mathbb{N}$. Then*

$$\sum_{i=1}^{\infty} P(A_i) < \infty \quad \Rightarrow \quad P\big(\limsup_{i\to\infty} A_i\big) = 0.$$

(ii) Assume that $A_i \in \mathcal{A}$, $i \in \mathbb{N}$, are independent. Then

$$\sum_{i=1}^{\infty} P(A_i) = \infty \quad \Rightarrow \quad P\big(\limsup_{i\to\infty} A_i\big) = 1.$$

Proof. (i) See Lemma 1.1.11.

(ii) It suffices to show that

$$P\Big(\bigcup_{m=n}^{\infty} A_m\Big) = 1 \quad \text{resp.} \quad P\Big(\bigcap_{m=n}^{\infty} A_m^c\Big) = 0 \qquad \forall\, n\,.$$

The last equality follows from the fact that

$$P\Big(\bigcap_{m=n}^{\infty} A_m^c\Big) = \lim_{k\to\infty} \underbrace{P\Big(\bigcap_{m=n}^{n+k} A_m^c\Big)}_{=\prod_{m=n}^{n+k} P(A_m^c) \quad \text{ind.}}$$

$$= \prod_{m=n}^{n+k} (1 - P(A_m)) \le \exp\left(\sum_{m=n}^{n+k} P(A_m)\right) = 0$$

where we used the inequality $1 - \alpha \leqslant e^{-\alpha}$ for all $\alpha \in \mathbb{R}$.

□

Example 1.9. Independent 0-1-**experiments with success probability** $p \in]0,1[$**.** Let $(x_1, \dots, x_N) \in \{0,1\}^N$ ("binary text of length N").

P_p["text occurs"]   ?

To calculate this probability we partition the infinite sequence $\omega = (y_n) \in \{0,1\}^{\mathbb{N}}$ into blocks of length N

$$(\underbrace{y_1, y_2, \dots}_{\substack{\text{1. block} \\ \text{length } = N}} \; \underbrace{\dots\dots}_{\substack{\text{2. block} \\ \text{length } = N}} \; \dots) \in \Omega := \{0,1\}^{\mathbb{N}}.$$

and consider the events A_i = "text occurs in the i^{th} block". Clearly, A_i, $i \in \mathbb{N}$, are independent events (!) by Proposition 1.2(ii) with equal probability

$$P_p(A_i) = p^K (1-p)^{N-K} =: \alpha > 0.$$

where $K := \sum_{i=1}^N x_i$ is the total sum of ones. In particular, $\sum_{i=1}^\infty P_p(A_i) = \sum_{i=1}^\infty \alpha = \infty$, and now Borel-Cantelli implies $P_p(A_\infty) = 1$, where

$$A_\infty = \limsup_{i\to\infty} A_i := \text{"text occurs infinitely many times"}\,.$$

Moreover: since the indicator functions $1_{A_1}, 1_{A_2}, \dots$ are uncorrelated (since they are independent r.v. (see below)), the strong law of large numbers implies that

$$\frac{1}{n}\sum_{i=1}^n 1_{A_i} \xrightarrow{P_p\text{-a.s.}} \mathbb{E}[1_{A_i}] = \alpha\,,$$

i.e. the relative frequency of the given text in the infinite sequence is strictly positive.

2 Independent random variables

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

Definition 2.1. A family X_i, $i \in I$, of r.v. on $(\Omega, \mathcal{A}, P)$ is said to be *independent*, if the σ-algebras

$$\sigma(X_i) := X_i^{-1}\big(\mathcal{B}(\bar{\mathbb{R}})\big) \quad \Big(= \big\{\{X_i \in A\} \mid A \in \mathcal{B}(\bar{\mathbb{R}})\big\}\Big), \quad i \in I,$$

are independent, i.e. for all finite subsets $J \subset I$ and any Borel subsets $A_j \in \mathcal{B}(\bar{\mathbb{R}})$

$$P\Big(\bigcap_{j \in J}\{X_j \in A_j\}\Big) = \prod_{j \in J} P[X_j \in A_j].$$

Remark 2.2. *Let X_i, $i \in I$, be independent and $h_i : \bar{\mathbb{R}} \to \bar{\mathbb{R}}$, $i \in I$, $\mathcal{B}(\bar{\mathbb{R}})/\mathcal{B}(\bar{\mathbb{R}})$-measurable. Then $Y_i := h_i(X_i)$, $i \in I$, are again independent, because $\sigma(Y_i) \subset \sigma(X_i)$ for all $i \in I$.*

Proposition 2.3. *Let $X_1, \ldots, X_n$ be independent r.v., ≥ 0. Then*

$$\mathbb{E}[X_1 \cdots X_n] = \mathbb{E}[X_1] \cdots \mathbb{E}[X_n].$$

Proof. W.l.o.g. $n = 2$. (Proof of the general case by induction, using the fact that $X_1 \cdot \ldots \cdot X_{n-1}$ and X_n are independent , since $X_1 \cdot \ldots \cdot X_{n-1}$ is measurable w.r.t $\sigma\big(\sigma(X_1) \cup \cdots \cup \sigma(X_{n-1})\big)$ and $\sigma\big(\sigma(X_1) \cup \cdots \cup \sigma(X_{n-1})\big)$ and $\sigma(X_n)$ are independent by Proposition 1.2.)

It therefore suffices to consider two independent r.v. $X, Y, \geq 0$, and we have to show that

$$\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]. \tag{2.2}$$

W.l.o.g. X, Y simple
(for general X and Y there exist increasing sequences of simple r.v. X_n (resp. Y_n), which are $\sigma(X)$-measurable (resp. $\sigma(Y)$-measurable), converging pointwise to X (resp. Y).

Then $\mathbb{E}[X_n Y_n] = \mathbb{E}[X_n] \cdot \mathbb{E}[Y_n]$ for all n implies (2.2) using monotone integration.)

But for X, Y simple, hence

$$X = \sum_{i=1}^{m} \alpha_i 1_{A_i} \quad \text{and} \quad Y = \sum_{j=1}^{n} \beta_j 1_{B_j},$$

with $\alpha_i, \beta_j \geqslant 0$ and $A_i \in \sigma(X)$ resp. $B_j \in \sigma(Y)$ it follows that

$$\mathbb{E}[XY] = \sum_{i,j} \alpha_i \beta_j \cdot P(A_i \cap B_j) = \sum_{i,j} \alpha_i \beta_j \cdot P(A_i) \cdot P(B_j) = \mathbb{E}[X] \cdot \mathbb{E}[Y]. \qquad \square$$

Corollary 2.4. *X, Y independent, $X, Y \in \mathcal{L}^1$*

$$\Rightarrow \qquad XY \in \mathcal{L}^1 \qquad \textit{and} \qquad \mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y].$$

Proof. Let $\varepsilon_1, \varepsilon_2 \in \{+, -\}$. Then X^{ε_1} and Y^{ε_2} are independent by Remark 2.2 and nonnegative. Proposition 2.3 implies

$$\mathbb{E}[X^{\varepsilon_1} \cdot Y^{\varepsilon_2}] = \mathbb{E}[X^{\varepsilon_1}] \cdot \mathbb{E}[Y^{\varepsilon_2}].$$

In particular $X^{\varepsilon_1} \cdot Y^{\varepsilon_2}$ in $\mathcal{L}^1$, because $\mathbb{E}[X^{\varepsilon_1}] \cdot \mathbb{E}[Y^{\varepsilon_2}] < \infty$. Hence

$$X \cdot Y = X^+ \cdot Y^+ + X^- \cdot Y^- - (X^+ \cdot Y^- + X^- \cdot Y^+) \quad \in \mathcal{L}^1,$$

and $\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$. □

Remark 2.5. *(i) In general the converse to the above corollary does not hold: For example let X be $N(0,1)$-distributed and $Y = X^2$. Then X and Y are not independent, but*

$$\mathbb{E}[XY] = \mathbb{E}[X^3] = \mathbb{E}[X] \cdot \mathbb{E}[Y] = 0\,.$$

(ii)

$$X, Y \in \mathcal{L}^2 \qquad \text{independent} \Rightarrow X, Y \qquad \text{uncorelated}$$

because

$$\operatorname{cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X] \cdot \mathbb{E}[Y] = 0\,.$$

Corollary 2.6 (to the strong law of large numbers). *Let $X_1, X_2, \dots \in \mathcal{L}^2$ be independent with $\sup_{i \in \mathbb{N}} \operatorname{var}(X_i) < \infty$. Then*

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \big(X_i(\omega) - \mathbb{E}[X_i]\big) = 0 \qquad P\text{-a.s.}$$

If $\mathbb{E}[X_i] \equiv m$ then $\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_i(\omega) = m \quad P$-a.s.

3 Kolmogorov's law of large numbers

Proposition 3.1 (Kolmogorov, 1930). *Let $X_1, X_2, \dots \in \mathcal{L}^1$ be independent, identically distributed, $m = \mathbb{E}[X_i]$. Then*

$$\underbrace{\frac{1}{n} \sum_{i=1}^{n} X_i(\omega)}_{\substack{\text{empirical} \\ \text{mean}}} \xrightarrow{n \to \infty} m \quad P\text{-a.s.}$$

Proposition 3.1 follows from the following more general result:

Proposition 3.2 (Etemadi, 1981). *Let $X_1, X_2, \dots \in \mathcal{L}^1$ be pairwise independent, identically distributed, $m = \mathbb{E}[X_i]$. Then*

$$\frac{1}{n} \sum_{i=1}^{n} X_i(\omega) \xrightarrow{n \to \infty} m \quad P\text{-a.s.}$$

Proof. W.l.o.g. $X_i \geqslant 0$
(otherwise consider $X_1^+, X_2^+, \dots$ (pairwise independent, identically distributed)
and $X_1^-, X_2^-, \dots$ (pairwise independent, identically distributed))

1. Replace X_i by $\tilde{X}_i := 1_{\{X_i < i\}} X_i$.

   Clearly,

$$\tilde{X}_i = h_i(X_i) \quad \text{with} \quad h_i(x) := \begin{cases} x & \text{if } x < i \\ 0 & \text{if } x \geqslant i \end{cases}$$

   Then $\tilde{X}_1, \tilde{X}_2, \dots$ are pairwise independent by Remark 2.2. For the proof it is now sufficient to show that for $\tilde{S}_n := \sum_{i=1}^n \tilde{X}_i$ we have that

$$\frac{\tilde{S}_n}{n} \xrightarrow{n\to\infty} m \quad P\text{-a.s.}$$

   Indeed,

$$\begin{aligned} \sum_{n=1}^\infty P[X_n \neq \tilde{X}_n] &= \sum_{n=1}^\infty P[X_n \geqslant n] = \sum_{n=1}^\infty P[X_1 \geqslant n] \\ &= \sum_{n=1}^\infty \sum_{k=n}^\infty P\big[X_1 \in [k, k+1[\big] = \sum_{k=1}^\infty k \cdot P\big[X_1 \in [k, k+1[\big] \\ &= \sum_{k=1}^\infty \mathbb{E}\big[\underbrace{k \cdot 1_{\{X_1 \in [k,k+1[\}}}_{\leqslant X_1 \cdot 1_{\{X_1 \in [k,k+1[\}}}\big] \leqslant \mathbb{E}[X_1] < \infty \end{aligned}$$

   implies by the Borel-Cantelli lemma

$$P[X_n \neq \tilde{X}_n \text{ infinitely often}] = 0\,.$$

2. Reduce the proof to convergence along the subsequence $k_n = \lfloor \alpha^n \rfloor$ (= largest natural number $\leq \alpha^n$), $\alpha > 1$.

   We will show in Step 3. that

$$\frac{\tilde{S}_{k_n} - \mathbb{E}[\tilde{S}_{k_n}]}{k_n} \xrightarrow{n\to\infty} 0 \qquad \text{P-a.s.} \tag{2.3}$$

   This will imply the assertion of the Proposition, because

$$\mathbb{E}[\tilde{X}_i] = \mathbb{E}\big[1_{\{X_i < i\}} \cdot X_i\big] = \mathbb{E}\big[1_{\{X_1 < i\}} \cdot X_1\big] \overset{i\to\infty}{\nearrow} \mathbb{E}[X_1] (= m)$$

   hence

$$\frac{1}{k_n} \cdot \mathbb{E}[\tilde{S}_{k_n}] = \frac{1}{k_n} \sum_{i=1}^{k_n} \mathbb{E}[\tilde{X}_i] \xrightarrow{n\to\infty} m,$$

and thus

$$\frac{1}{k_n}\cdot \tilde S_{k_n} \xrightarrow{n\to\infty} m \quad P\text{-a.s.}$$

If $l \in \mathbb{N}\cap [k_n, k_{n+1}[$, then

$$\underbrace{\frac{k_n}{k_{n+1}}}_{\xrightarrow{n\to\infty}\frac{1}{\alpha}} \cdot \underbrace{\frac{\tilde S_{k_n}}{k_n}}_{\substack{\xrightarrow{n\to\infty} m\\ P\text{-a.s.}}} \leqslant \frac{\tilde S_l}{l} \leqslant \underbrace{\frac{\tilde S_{k_{n+1}}}{k_{n+1}}}_{\substack{\xrightarrow{n\to\infty} m\\ P\text{-a.s.}}} \cdot \underbrace{\frac{k_{n+1}}{k_n}}_{\xrightarrow{n\to\infty}\alpha} .$$

Hence there exists a P-null set $N_\alpha \in \mathcal{A}$, such that for all $\omega \notin N_\alpha$

$$\frac{1}{\alpha}\cdot m \leqslant \liminf_{l\to\infty} \frac{\tilde S_l(\omega)}{l} \leqslant \limsup_{l\to\infty}\frac{\tilde S_l(\omega)}{l} \leqslant \alpha\cdot m.$$

Finally choose a subsequence $\alpha_n \searrow 1$. Then for all $\omega \notin N := \bigcup_{n\geqslant 1} N_{\alpha_n}$

$$\lim_{l\to\infty}\frac{\tilde S_l(\omega)}{l} = m.$$

3. Due to Lemma 1.7.7 it suffices for the proof of (2.3) to show that

$$\forall \varepsilon > 0: \qquad \sum_{n=1}^\infty P\left[\left|\frac{\tilde S_{k_n} - \mathbb{E}[\tilde S_{k_n}]}{k_n}\right| \geqslant \varepsilon\right] < \infty$$

(fast convergence in probability towards 0)

Pairwise independence of $\tilde X_i$ implies $\tilde X_i$ pairwise uncorrelated, hence

$$P\left[\left|\frac{\tilde S_{k_n} - \mathbb{E}[\tilde S_{k_n}]}{k_n}\right| \geqslant \varepsilon\right] \leqslant \frac{1}{k_n^2\varepsilon^2}\cdot \operatorname{var}(\tilde S_{k_n}) = \frac{1}{k_n^2\varepsilon^2}\sum_{i=1}^{k_n}\operatorname{var}(\tilde X_i)$$
$$\leqslant \frac{1}{k_n^2\varepsilon^2}\sum_{i=1}^{k_n}\mathbb{E}\big[(\tilde X_i)^2\big].$$

It therefore suffices to show that

$$s := \sum_{n=1}^\infty\left(\frac{1}{k_n^2}\sum_{i=1}^{k_n}\mathbb{E}\big[(\tilde X_i)^2\big]\right) = \sum_{\substack{(i,n)\in\mathbb{N}^2,\\ i\leqslant k_n}} \frac{1}{k_n^2}\cdot\mathbb{E}\big[(\tilde X_i)^2\big] < \infty\,.$$

To this end note that

$$s = \sum_{i=1}^\infty\left(\sum_{n\,:\,k_n\geqslant i}\frac{1}{k_n^2}\right)\cdot \mathbb{E}\big[(\tilde X_i)^2\big].$$

We will show in the following that there exists a constant c such that

$$\sum_{n\,:\,k_n\geqslant i}\frac{1}{k_n^2}\leqslant\frac{c}{i^2}. \tag{2.4}$$

This will then imply that

$$\begin{aligned}
s &\overset{(2.4)}{\leqslant} c\sum_{i=1}^{\infty}\frac{1}{i^2}\cdot\mathbb{E}\big[(\tilde{X}_i)^2\big]=c\sum_{i=1}^{\infty}\frac{1}{i^2}\cdot\mathbb{E}\big[1_{\{X_1<i\}}\cdot X_1^2\big]\\
&\leqslant c\sum_{i=1}^{\infty}\bigg(\frac{1}{i^2}\sum_{l=1}^{i}l^2\cdot P\big[X_1\in[l-1,l[\big]\bigg)\\
&= c\sum_{l=1}^{\infty}\Bigg(l^2\cdot\underbrace{\bigg(\sum_{i=l}^{\infty}\frac{1}{i^2}\bigg)}_{\leqslant 2l^{-1}}\cdot P\big[X_1\in[l-1,l[\big]\Bigg)\\
&\leqslant 2c\sum_{l=1}^{\infty}l\cdot P\big[X_1\in[l-1,l[\big]=2c\sum_{l=1}^{\infty}\mathbb{E}\big[\underbrace{l\cdot 1_{\{X_1\in[l-1,l[\}}}_{\leqslant(X_1+1)\cdot 1_{\{X_1\in[l-1,l[\}}}\big]\\
&\leqslant 2c\cdot\big(\mathbb{E}[X_1]+1\big)<\infty,
\end{aligned}$$

where we used the fact that

$$\sum_{i=l}^{\infty}\frac{1}{i^2}\leqslant\frac{1}{l^2}+\sum_{i=l+1}^{\infty}\frac{1}{(i-1)i}=\frac{1}{l^2}+\sum_{i=l+1}^{\infty}\bigg(\frac{1}{i-1}-\frac{1}{i}\bigg)=\frac{1}{l^2}+\frac{1}{l}\leqslant\frac{2}{l}\,.$$

It remains to show (2.4). To this end note that

$$\begin{aligned}
&\lfloor\alpha^n\rfloor=k_n\leqslant\alpha^n<k_n+1\\
\Rightarrow\quad & k_n>\alpha^n-1\overset{\alpha>1}{\geqslant}\alpha^n-\alpha^{n-1}=\underbrace{\bigg(\frac{\alpha-1}{\alpha}\bigg)}_{=:c_\alpha}\alpha^n.
\end{aligned}$$

Let n_i be the smallest natural number satisfying $k_{n_i}=\lfloor\alpha^{n_i}\rfloor\geqslant i$, hence $\alpha^{n_i}\geqslant i$, then

$$\sum_{n\,:\,k_n\geqslant i}\frac{1}{k_n^2}\leqslant c_\alpha^{-2}\sum_{n\geqslant n_i}\frac{1}{\alpha^{2n}}=c_\alpha^{-2}\cdot\frac{1}{1-\alpha^{-2}}\cdot\alpha^{-2n_i}\leqslant\frac{c_\alpha^{-2}}{1-\alpha^{-2}}\cdot\frac{1}{i^2}. \qquad\square$$

Corollary 3.3. *Let $X_1, X_2, \dots$ be pairwise independent, identically distributed (iid) with $X_i\geqslant 0$. Then*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}X_i(\omega)=\mathbb{E}[X_1]\quad\big(\in[0,\infty]\big)\qquad P\text{-a.s.}$$

Proof. W.l.o.g. $\mathbb{E}[X_1] = \infty$. Then $\frac{1}{n}\sum_{i=1}^{n}\big(X_i(\omega) \wedge N\big) \xrightarrow{n\to\infty} \mathbb{E}[X_1 \wedge N]$, P-a.s. for all N, hence

$$\frac{1}{n}\sum_{i=1}^{n} X_i(\omega) \geqslant \frac{1}{n}\sum_{i=1}^{n}\big(X_i(\omega) \wedge N\big) \xrightarrow{n\to\infty} \mathbb{E}[X_1 \wedge N] \overset{N\to\infty}{\nearrow} \mathbb{E}[X_1] \quad P\text{-a.s.} \qquad \square$$

Example 3.4. Growth in random media Let $Y_1, Y_2, \dots$ be i.i.d., $Y_i > 0$, with $m := \mathbb{E}[Y_i]$ (existence of such a sequence later!)

Define $X_0 = 1$ and inductively $X_n := X_{n-1} \cdot Y_n$

Clearly, $X_n = Y_1 \cdots Y_n$ and $\mathbb{E}[X_n] = \mathbb{E}[Y_1] \cdots \mathbb{E}[Y_n] = m^n$, hence

$$\mathbb{E}[X_n] \to \begin{cases} +\infty & \text{if } m > 1 \quad \text{exponential growth (supercritical)} \\ 1 & \text{if } m = 1 \quad \text{critical} \\ 0 & \text{if } m < 1 \quad \text{exponential decay (subcritical)} \end{cases}$$

What will be the long-time behaviour of $X_n(\omega)$?

Surprisingly, in the supercritical case $m > 1$, one may observe that $\lim_{n\to\infty} X_n = 0$ with positive probability.

Explanation: Suppose that $\log Y_i \in \mathcal{L}^1$. Then

$$\frac{1}{n}\log X_n = \frac{1}{n}\sum_{i=1}^{n}\log Y_i \xrightarrow{n\to\infty} \mathbb{E}[\log Y_1] =: \alpha \quad P\text{-a.s.}$$

and

$\alpha < 0$**:** $\exists\, \varepsilon > 0$ with $\alpha + \varepsilon < 0$, so that $X_n(\omega) \leqslant e^{n(\alpha+\varepsilon)}$ $\forall\, n \geqslant n_0(\omega)$, hence P-a.s. exponential decay

$\alpha > 0$**:** $\exists\, \varepsilon > 0$ with $\alpha - \varepsilon > 0$, so that $X_n(\omega) \geqslant e^{n(\alpha-\varepsilon)}$ $\forall\, n \geqslant n_0(\omega)$, hence P-a.s. exponential growth

Note that Jensen's inequality

$$\alpha = \mathbb{E}[\log Y_1] \quad \leqslant \quad \log \underbrace{\mathbb{E}[Y_1]}_{=m},$$

and in general the inequality is strict, i.e. $\alpha < \log m$, so that it might happen that $\alpha < 0$ although $m > 1$ (!)

Illustration As a particular example let

$$Y_i := \begin{cases} \frac{1}{2}(1+c) & \text{with prob.} \frac{1}{2} \\ \frac{1}{2} & \text{with prob.} \frac{1}{2} \end{cases}$$

, so that $\mathbb{E}[Y_i] = \frac{1}{4}(1+c) + \frac{1}{4} = \frac{1}{2} + \frac{1}{4}c$ (supercritical if $c > 2$)

On the other hand

$$\mathbb{E}[\log Y_1] = \frac{1}{2} \cdot \left[\log\left(\frac{1}{2}(1+c)\right) + \log\frac{1}{2}\right] = \frac{1}{2} \cdot \log\frac{1+c}{4} \overset{c<3}{<} 0.$$

Hence $X_n \xrightarrow{n\to\infty} 0$ P-a.s. with exponential rate for $c < 3$, whereas at the same time for $c > 2$ $\mathbb{E}[X_n] = m^n \nearrow \infty$ with exponential rate.

Back to Kolmogorov's law of large numbers:
Let $X_1, X_2, \ldots \in \mathcal{L}^1$ i.i.d. with $m := \mathbb{E}[X_i]$. Then

$$\frac{1}{n}\sum_{i=1}^{n} X_i(\omega) \xrightarrow{n\to\infty} \mathbb{E}[X_1] \quad P\text{-a.s.}$$

Define the "random measure"

$$\begin{aligned}\varrho_n(\omega, A) &:= \frac{1}{n}\sum_{i=1}^{n} 1_A\big(X_i(\omega)\big) \\ &= \text{"relative frequency of the event } X_i \in A\text{"}\end{aligned}$$

Then

$$\varrho_n(\omega, \,\cdot\,) = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i(\omega)}$$

is a probability measure on $\big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$ for fixed ω and it is called the *empirical distribution of the first n observations*

Proposition 3.5. *For P-almost every $\omega \in \Omega$:*

$$\varrho_n(\omega, \,\cdot\,) \xrightarrow{n\to\infty} \mu := P \circ X_1^{-1} \quad \textit{weakly.}$$

Proof. Clearly, Kolmogorov's law of large numbers implies that for any $x \in \mathbb{R}$

$$\begin{aligned}F_n(\omega, x) := \varrho_n\big(\omega,]-\infty, x]\big) &= \frac{1}{n}\sum_{i=1}^{n} 1_{]-\infty,x]}\big(X_i(\omega)\big) \\ &\to \mathbb{E}[1_{]-\infty,x]}(X_1)] = P[X_1 \le x] = \mu\big(]-\infty, x]\big) =: F(x)\end{aligned}$$

P-a.s., hence for every $\omega \notin N(x)$ for some P-null set $N(x)$.
Then

$$N := \bigcup_{r\in\mathbb{Q}} N(r).$$

is a P-null set too, and for all $x \in \mathbb{R}$ and all $s, r \in \mathbb{Q}$ with $s < x < r$ and $\omega \notin N$:

$$\begin{aligned}F(s) := \lim_{n\to\infty} F_n(\omega, s) &\leqslant \liminf_{n\to\infty} F_n(\omega, x) \\ &\leqslant \limsup_{n\to\infty} F_n(\omega, x) \leqslant \lim_{n\to\infty} F_n(\omega, r) = F(r).\end{aligned}$$

Hence, if F is continuous at x, then for $\omega \notin N$

$$\lim_{n\to\infty} F_n(\omega, x) = F(x).$$

Now the assertion follows from the Portmanteau theorem. □