Chapter III

Basic Concepts of Probability Theory

Context for probability theoretical concepts: a probability space $(\Omega, \mathfrak{A}, P)$.
Terminology: $A \in \mathfrak{A}$ *event*, $P(A)$ *probability of the event* $A \in \mathfrak{A}$.

1 Random Variables and Distributions

Given: a probability space $(\Omega, \mathfrak{A}, P)$ and a measurable space $(\Omega', \mathfrak{A}')$.

Definition 1. $X : \Omega \to \Omega'$ *random element* if X is $\mathfrak{A}$-$\mathfrak{A}'$-measurable. Particular cases:

(i) X *(real) random variable* if $(\Omega', \mathfrak{A}') = (\mathbb{R}, \mathfrak{B})$,

(ii) X *numerical random variable* if $(\Omega', \mathfrak{A}') = (\overline{\mathbb{R}}, \overline{\mathfrak{B}})$,

(iii) X *k-dimensional (real) random vector* if $(\Omega', \mathfrak{A}') = (\mathbb{R}^k, \mathfrak{B}_k)$,

(iv) X *k-dimensional numerical random vector* if $(\Omega', \mathfrak{A}') = (\overline{\mathbb{R}}^k, \overline{\mathfrak{B}}_k)$.

Definition 2.

(i) *Distribution (probability law)* of a random element $X : \Omega \to \Omega'$ *(with respect to P)*

$$P_X = X(P).$$

Notation: $X \sim Q$ if $P_X = Q$.

(ii) Given: probability spaces $(\Omega_1, \mathfrak{A}_1, P_1)$, $(\Omega_2, \mathfrak{A}_2, P_2)$ and random elements

$$X_1 : \Omega_1 \to \Omega', \qquad X_2 : \Omega_2 \to \Omega'.$$

X_1 and X_2 are *identically distributed* if

$$(P_1)_{X_1} = (P_2)_{X_2}.$$

Remark 1.

(i) $P_X(A') = P(\{X \in A'\})$ for every $A' \in \mathfrak{A}'$.

(ii) For random elements $X, Y : \Omega \to \Omega'$

$$X = Y \ P\text{-a.s.} \quad \Rightarrow \quad P_X = P_Y,$$

but the converse is not true in general. For instance, let P be the uniform distribution on $\Omega = \{0, 1\}$ and define $X(\omega) = \omega$ and $Y(\omega) = 1 - \omega$.

(iii) For every probability measure Q on $(\Omega', \mathfrak{A}')$ there exists a probability space $(\Omega, \mathfrak{A}, P)$ and a random element $X : \Omega \to \Omega'$ such that $X \sim Q$. Take $(\Omega, \mathfrak{A}, P) = (\Omega', \mathfrak{A}', Q)$ and $X = \mathrm{id}_\Omega$.

(iv) A major part of probability theory deals with properties of random elements that can be formulated in terms of their distributions.

Example 1.

(i) *Discrete distributions*, specified by a countable set $\emptyset \neq D \subset \Omega'$ and a mapping $p : D \to \mathbb{R}$ such that

$$\forall\, r \in D : p(r) \geq 0 \qquad \wedge \qquad \sum_{r \in D} p(r) = 1,$$

namely,

$$P_X = \sum_{r \in D} p(r) \cdot \varepsilon_r.$$

Thus, if $\{r\} \in \mathfrak{A}'$ for every $r \in D$,

$$P(\{X = r\}) = p(r).$$

If $|D| < \infty$ then $p(r) = \frac{1}{|D|}$ yields the *uniform distribution on* D.
For $(\Omega', \mathfrak{A}') = (\mathbb{R}, \mathfrak{B})$

$$B(n, p) = \sum_{k=0}^{n} \binom{n}{k} \cdot p^k (1-p)^{n-k} \cdot \varepsilon_k$$

is the *binomial distribution* with parameters $n \in \mathbb{N}$ and $p \in [0, 1]$. In particular, for $n = 1$ we get the *Bernoulli distribution*

$$B(1, p) = (1-p) \cdot \varepsilon_0 + p \cdot \varepsilon_1.$$

Further examples include the *geometric distribution* with parameter $p \in]0, 1]$,

$$G(p) = \sum_{k=1}^{\infty} p \cdot (1-p)^{k-1} \cdot \varepsilon_k,$$

and the *Poisson distribution* with parameter $\lambda > 0$,

$$\pi(\lambda) = \sum_{k=0}^{\infty} \exp(-\lambda) \cdot \frac{\lambda^k}{k!} \cdot \varepsilon_k.$$

(ii) *Distributions on* $(\mathbb{R}^k, \mathfrak{B}_k)$ *that are absolutely continuous w.r.t.* λ_k, namely, due to the Radon-Nikodym-Theorem

$$P_X = f \cdot \lambda_k,$$

where

$$f \in \overline{\mathfrak{Z}}_+(\mathbb{R}^k, \mathfrak{B}_k) \qquad \wedge \qquad \int f \, d\lambda_k = 1.$$

Thus

$$P(\{X \in A'\}) = \int_{A'} f \, d\lambda_k$$

for every $A' \in \mathfrak{B}_k$.

We present some examples in the case $k = 1$. The *normal distribution*

$$N(\mu, \sigma^2) = f \cdot \lambda_1 \, ,$$

with parameters $\mu \in \mathbb{R}$ and σ^2, where $\sigma > 0$, is obtained by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2} \right), \qquad x \in \mathbb{R}.$$

The *exponential distribution* with parameter $\lambda > 0$ is obtained by

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ \lambda \cdot \exp(-\lambda \, x) & \text{if } x \geq 0. \end{cases}$$

The *uniform distribution* on $D \in \mathfrak{B}$ with $\lambda_1(D) \in]0, \infty[$ is obtained by

$$f = \frac{1}{\lambda_1(D)} \cdot 1_D.$$

(iii) Distributions on product spaces can be constructed by means of the results from Section II.8.

Remark 2. Define $\infty^r = \infty$ for $r > 0$. For $1 \leq p < q < \infty$ and $X \in \overline{\mathfrak{Z}}(\Omega, \mathfrak{A})$

$$\int |X|^p \, dP \leq \left(\int |X|^q \, dP \right)^{p/q},$$

due to Hölder's inequality.

Notation:

$$\mathfrak{L} = \mathfrak{L}(\Omega, \mathfrak{A}, P) = \left\{ X \in \mathfrak{Z}(\Omega, \mathfrak{A}) : \int |X| \, dP < \infty \right\}$$

is the class of P-integrable random variables, and analogously

$$\overline{\mathfrak{L}} = \overline{\mathfrak{L}}(\Omega, \mathfrak{A}, P) = \left\{ X \in \overline{\mathfrak{Z}}(\Omega, \mathfrak{A}) : \int |X| \, dP < \infty \right\}$$

is the class of P-integrable numerical random variables. We consider P_X as a distribution on $(\mathbb{R}, \mathfrak{B})$ if $P(\{X \in \mathbb{R}\}) = 1$ for a numerical random variable X, and we consider $\mathfrak{L}$ as a subspace of $\overline{\mathfrak{L}}$.

Definition 3. For $X \in \overline{\mathfrak{L}}$

$$\mathrm{E}(X) = \int X \, dP$$

is the *expectation* of X. For $X \in \overline{\mathfrak{Z}}(\Omega, \mathfrak{A})$ such that $X^2 \in \overline{\mathfrak{L}}$

$$\mathrm{Var}(X) = \int (X - \mathrm{E}(X))^2 \, dP$$

and $\sqrt{\mathrm{Var}(X)}$ are the *variance* and the *standard deviation* of X, respectively.

Remark 3. Theorem II.9.1 implies

$$\int_\Omega |X|^p \, dP < \infty \qquad \Leftrightarrow \qquad \int_{\mathbb{R}} |x|^p \, P_X(dx) < \infty$$

for $X \in \overline{\mathfrak{Z}}(\Omega, \mathfrak{A})$, in which case, for $p = 1$

$$\mathrm{E}(X) = \int_{\mathbb{R}} x \, P_X(dx),$$

and for $p = 2$

$$\mathrm{Var}(X) = \int_{\mathbb{R}} (x - \mathrm{E}(X))^2 \, P_X(dx).$$

Thus $\mathrm{E}(X)$ and $\mathrm{Var}(X)$ depend only on P_X.

Example 2.

$$\begin{array}{lll} X \sim B(n,p) & \mathrm{E}(X) = n \cdot p & \mathrm{Var}(X) = n \cdot p \cdot (1-p) \\ X \sim G(p) & \mathrm{E}(X) = \dfrac{1}{p} & \mathrm{Var}(X) = \dfrac{1-p}{p^2} \\ X \sim \pi(\lambda) & \mathrm{E}(X) = \lambda & \mathrm{Var}(X) = \lambda, \end{array}$$

see Introduction to Stochastics.

X is *Cauchy distributed* with parameter $\alpha > 0$ if $X \sim f \cdot \lambda_1$ where

$$f(x) = \frac{\alpha}{\pi(\alpha^2 + x^2)}, \qquad x \in \mathbb{R}.$$

Since $\int_0^t \frac{x}{1+x^2} dx = \frac{1}{2} \log(1 + t^2)$ neither $\mathrm{E}(X^+) < \infty$ nor $\mathrm{E}(X^-) < \infty$, and therefore $X \notin \mathfrak{L}$.

If $X \sim N(\mu, \sigma^2)$ then

$$\mathrm{E}(X) = \mu \qquad \wedge \qquad \mathrm{Var}(X) = \sigma^2,$$

see Introduction to Stochastics.

If X is exponentially distributed with parameter $\lambda > 0$ then

$$\mathrm{E}(X) = \frac{1}{\lambda} \qquad \wedge \qquad \mathrm{Var}(X) = \frac{1}{\lambda^2}.$$

Definition 4. Let $X = (X_1, \dots, X_k)$ be a random vector. Then

$$F_X : \mathbb{R}^k \to [0,1]$$
$$(x_1, \dots, x_k) \mapsto P_X\Big(\bigtimes_{i=1}^{k} \,]-\infty\,,\ x_i] \Big) = P\Big(\bigcap_{i=1}^{k} \{X_i \leq x_i\} \Big)$$

is called the *distribution function* of X.

Theorem 1. Given: probability spaces $(\Omega_1, \mathfrak{A}_1, P_1)$, $(\Omega_2, \mathfrak{A}_2, P_2)$ and random vectors

$$X^1 : \Omega_1 \to \mathbb{R}^k, \qquad X^2 : \Omega_2 \to \mathbb{R}^k.$$

Then

$$(P_1)_{X^1} = (P_2)_{X^2} \qquad \Leftrightarrow \qquad F_{X^1} = F_{X^2}.$$

Proof. '$\Rightarrow$' holds trivially. '$\Leftarrow$': By Remark II.1.6, $\mathfrak{B}_k = \sigma(\mathfrak{E})$ for

$$\mathfrak{E} = \Big\{ \bigtimes_{i=1}^{k} \,]-\infty, x_i] : x_1, \dots, x_k \in \mathbb{R} \Big\}.$$

Use Theorem II.4.4. □

For notational convenience, we consider the case $k = 1$ in the sequel.

Theorem 2.

(i) F_X is non-decreasing,

(ii) F_X is right-continuous,

(iii) $\lim_{x \to -\infty} F_X(x) = 0$ and $\lim_{x \to \infty} F_X(x) = 1$,

(iv) F_X is continuous at x iff $P(\{X = x\}) = 0$.

Proof. Übung 3.4.a. □

Theorem 3. For every function F that satisfies (i)–(iii) from Theorem 2,

$$\exists_1 Q \text{ probability measure on } \mathfrak{B} : \forall\, x \in \mathbb{R} : Q(]-\infty, x]) = F(x).$$

Proof. Analogously to the construction of the Lebesgue measure, see Übung 3.4.b. □

2 Convergence in Probability

Motivated by the Examples II.5.2 and II.6.1 we introduce a notion of convergence that is weaker than convergence in mean and convergence almost surely.

In the sequel, X, X_n, etc. random variables on a common probability space $(\Omega, \mathfrak{A}, P)$.

Lemma 1.

$$X_n \xrightarrow{P\text{-a.s.}} X \quad \Leftrightarrow \quad \forall\, \varepsilon > 0 : \lim_{n\to\infty} P\left(\left\{\sup_{m\geq n} |X_m - X| > \varepsilon\right\}\right) = 0.$$

Proof. Put

$$C_{k,n} = \bigcap_{m\geq n} \{|X_m - X| \leq 1/k\}, \qquad B_k = \bigcup_{n\in\mathbb{N}} C_{k,n}, \qquad A = \bigcap_{k\in\mathbb{N}} B_k.$$

Hence

$$A = \{\lim_{n\to\infty} X_n = X\}.$$

Clearly $B_k \downarrow A$ and $C_{k,n} \uparrow B_k$. Thus, using the σ-continuity of P,

$$\begin{aligned} & X_n \xrightarrow{P\text{-a.s.}} X \\ \Leftrightarrow \quad & \forall\, k \in \mathbb{N} : \ P(B_k) = 1 \\ \Leftrightarrow \quad & \forall\, k \in \mathbb{N} : \ \lim_{n\to\infty} P(C_{k,n}) = 1 \\ \Leftrightarrow \quad & \forall\, k \in \mathbb{N} : \ \lim_{n\to\infty} P\left(\left\{\sup_{m\geq n} |X_m - X| > 1/k\right\}\right) = 0. \end{aligned}$$

□

Definition 1. $(X_n)_n$ *converges* to X *in probability* if

$$\forall\, \varepsilon > 0 : \lim_{n\to\infty} P(\{|X_n - X| > \varepsilon\}) = 0.$$

Notation: $X_n \xrightarrow{P} X$.

Remark 1. By Lemma 1,

$$X_n \xrightarrow{P\text{-a.s.}} X \quad \Rightarrow \quad X_n \xrightarrow{P} X.$$

Example II.6.1 shows that '⇐' does not hold in general. The Law of Large Numbers deals with convergence almost surely or convergence in probability, see the introductory Example I.1 and Sections **??**.**??** and **??**.**??**.

Theorem 1 (Chebyshev-Markov Inequality)**.** Let $(\Omega, \mathfrak{A}, \mu)$ be a measure space and $f \in \overline{\mathfrak{Z}}(\Omega, \mathfrak{A})$. For every $\varepsilon > 0$ and every $1 \leq p < \infty$

$$\mu(\{|f| \geq \varepsilon\}) \leq \frac{1}{\varepsilon^p} \cdot \int |f|^p \, d\mu.$$

Proof. We have

$$\int_{\{|f|\geq\varepsilon\}} \varepsilon^p \, d\mu \leq \int_\Omega |f|^p \, d\mu.$$

□

Corollary 1. If $\mathrm{E}(X^2) < \infty$, then

$$P(\{|X - \mathrm{E}(X)| \geq \varepsilon\}) \leq \frac{1}{\varepsilon^2} \cdot \mathrm{Var}(X).$$

Theorem 2.

$$d(X,Y) = \int \min(1, |X - Y|) \, dP$$

defines a semi-metric on $\mathfrak{Z}(\Omega, \mathfrak{A})$, and

$$X_n \xrightarrow{P} X \quad \Leftrightarrow \quad \lim_{n\to\infty} d(X_n, X) = 0.$$

Proof. '⇒' For $\varepsilon > 0$

$$\begin{aligned} &\int \min(1,\, |X_n - X|) \, dP \\ &\quad = \int_{\{|X_n - X| > \varepsilon\}} \min(1,\, |X_n - X|) \, dP + \int_{\{|X_n - X| \leq \varepsilon\}} \min(1,\, |X_n - X|) \, dP \\ &\quad \leq P(\{|X_n - X| > \varepsilon\}) + \min(1, \varepsilon). \end{aligned}$$

'⇐': Let $0 < \varepsilon < 1$. Use Theorem 1 to obtain

$$\begin{aligned} P(\{|X_n - X| > \varepsilon\}) &= P(\{\min(1, |X_n - X|) > \varepsilon\}) \\ &\leq \frac{1}{\varepsilon} \cdot \int \min(1,\, |X_n - X|) \, dP = \frac{1}{\varepsilon} \cdot d(X_n, X). \end{aligned}$$

□

Remark 2. By Theorem 2,

$$X_n \xrightarrow{\mathfrak{L}^p} X \quad \Rightarrow \quad X_n \xrightarrow{P} X.$$

Example II.5.2 shows that '⇐' does not hold in general.

Corollary 2.

$$X_n \xrightarrow{P} X \quad \Rightarrow \quad \exists \,\text{subsequence } (X_{n_k})_{k\in\mathbb{N}} : \; X_{n_k} \xrightarrow{P\text{-a.s.}} X.$$

Proof. Due to Theorems II.6.3 and 2 there exists a subsequence $(X_{n_k})_{k\in\mathbb{N}}$ such that

$$\min(1, |X_{n_k} - X|) \xrightarrow{P\text{-a.s.}} 0.$$

□

Remark 3. In any semi-metric space (M, d) a sequence $(a_n)_{n\in\mathbb{N}}$ converges to a iff

$$\forall\,\text{subsequence } (a_{n_k})_{k\in\mathbb{N}}\ \exists\,\text{subsequence } (a_{n_{k_\ell}})_{\ell\in\mathbb{N}} :\ \lim_{\ell\to\infty} d(a_{n_{k_\ell}}, a) = 0.$$

Corollary 3. $X_n \xrightarrow{P} X$ iff

$$\forall\,\text{subsequence } (X_{n_k})_{k\in\mathbb{N}}\ \exists\,\text{subsequence } (X_{n_{k_\ell}})_{\ell\in\mathbb{N}} :\ X_{n_{k_\ell}} \xrightarrow{P\text{-a.s.}} X.$$

Proof. '⇒': Corollary 2. '⇐': Remarks 1 and 3 together with Theorem 2. □

Remark 4. We conclude that, in general, there is no semi-metric on $\mathfrak{Z}(\Omega, \mathfrak{A})$ that defines a.s.-convergence. However, if Ω is countable, then

$$X_n \xrightarrow{P\text{-a.s.}} X \quad\Leftrightarrow\quad X_n \xrightarrow{P} X.$$

Proof: Übung 8.2.

Lemma 2. Let $\longrightarrow$ denote convergence almost everywhere or convergence in probability. If $X_n^{(i)} \longrightarrow X^{(i)}$ for $i = 1, \dots, k$ and $f : \mathbb{R}^k \to \mathbb{R}$ is continuous, then

$$f \circ (X_n^{(1)}, \dots, X_n^{(k)}) \longrightarrow f \circ (X^{(1)}, \dots, X^{(k)}).$$

Proof. Trivial for convergence almost everywhere, and by Corollary 3 the conclusion holds for convergence in probability, too. □

Corollary 4. Let $X_n \xrightarrow{P} X$. Then

$$X_n \xrightarrow{P} Y \quad\Leftrightarrow\quad X = Y\ P\text{-a.s.}$$

Proof. Corollary 3 and Lemma II.6.1. □

3 Convergence in Distribution

Given: a metric space (M, ρ). Put

$$C^b(M) = \{f : M \to \mathbb{R} : f \text{ bounded, continuous}\},$$

and consider the Borel-σ-algebra $\mathfrak{B}(M)$ in M. Moreover, let $\mathfrak{M}(M)$ denote the set of all probability measures on $\mathfrak{B}(M)$.

Definition 1.

(i) A sequence $(Q_n)_{n\in\mathbb{N}}$ in $\mathfrak{M}(M)$ *converges weakly* to $Q \in \mathfrak{M}(M)$ if

$$\forall\, f \in C^b(M) :\ \lim_{n\to\infty} \int f\, dQ_n = \int f\, dQ.$$

Notation: $Q_n \xrightarrow{\text{w}} Q$.

(ii) A sequence $(X_n)_{n\in\mathbb{N}}$ of random elements with values in M *converges in distribution* to a random element X with values in M if $Q_n \xrightarrow{\text{w}} Q$ for the distributions Q_n of X_n and Q of X, respectively.
Notation: $X_n \xrightarrow{\text{d}} X$.

Remark 1. For convergence in distribution the random elements need not be defined on a common probability space.

In the sequel: Q_n, $Q \in \mathfrak{M}(M)$ for $n \in \mathbb{N}$.

Example 1.

(i) For $x_n,\ x \in M$

$$\varepsilon_{x_n} \xrightarrow{\text{w}} \varepsilon_x \quad \Leftrightarrow \quad \lim_{n\to\infty} \rho(x_n, x) = 0.$$

For the proof of '$\Leftarrow$', note that

$$\int f\, d\varepsilon_{x_n} = f(x_n), \qquad \int f\, d\varepsilon_x = f(x).$$

For the proof of '$\Rightarrow$', suppose that $\limsup_{n\to\infty} \rho(x_n, x) > 0$. Take

$$f(y) = \min(\rho(y, x), 1), \qquad y \in M,$$

and observe that $f \in C^b(M)$ and

$$\limsup_{n\to\infty} \int f\, d\varepsilon_{x_n} = \limsup_{n\to\infty} \min(\rho(x_n, x), 1) > 0$$

while $\int f\, d\varepsilon_x = 0$.

(ii) For the euclidean distance ρ on $M = \mathbb{R}^k$

$$(M, \mathfrak{B}(M)) = (\mathbb{R}^k, \mathfrak{B}_k).$$

Now, in particular, $k = 1$ and

$$Q_n = N(\mu_n, \sigma_n^2)$$

where $\sigma_n > 0$. For $f \in C^b(\mathbb{R})$

$$\int f\, dQ_n = 1/\sqrt{2\pi} \cdot \int_{\mathbb{R}} f(\sigma_n \cdot x + \mu_n) \cdot \exp(-1/2 \cdot x^2)\, \lambda_1(dx).$$

Put $N(\mu, 0) = \varepsilon_\mu$. Then

$$\lim_{n\to\infty} \mu_n = \mu \ \wedge \ \lim_{n\to\infty} \sigma_n = \sigma \quad \Rightarrow \quad Q_n \xrightarrow{\text{w}} N(\mu, \sigma^2).$$

Otherwise $(Q_n)_{n\in\mathbb{N}}$ does not converge weakly. Übung 8.4.

(iii) For $M = C([0, T])$ let $\rho(x, y) = \sup_{t\in[0,T]} |x(t) - y(t)|$. Cf. the introductory Example I.3.

Remark 2. Note that $Q_n \xrightarrow{\text{w}} Q$ does not imply

$$\forall\, A \in \mathfrak{B}(M) : \lim_{n\to\infty} Q_n(A) = Q(A).$$

For instance, assume $\lim_{n\to\infty} \rho(x_n, x) = 0$ with $x_n \neq x$ for every $n \in \mathbb{N}$. Then

$$\varepsilon_{x_n}(\{x\}) = 0, \qquad \varepsilon_x(\{x\}) = 1.$$

Theorem 1 (Portmanteau Theorem)**.** The following properties are equivalent:

(i) $Q_n \xrightarrow{\text{w}} Q$,

(ii) $\forall\, f \in C^b(M)$ uniformly continuous : $\lim_{n\to\infty} \int f\, dQ_n = \int f\, dQ$,

(iii) $\forall\, A \subset M$ closed : $\limsup_{n\to\infty} Q_n(A) \leq Q(A)$,

(iv) $\forall\, A \subset M$ open : $\liminf_{n\to\infty} Q_n(A) \geq Q(A)$,

(v) $\forall\, A \in \mathfrak{B}(M)$: $Q(\partial A) = 0 \;\Rightarrow\; \lim_{n\to\infty} Q_n(A) = Q(A)$.

Proof. See Gänssler, Stute (1977, Satz 8.4.9). □

In the sequel, we study the particular case $(M, \mathfrak{B}(M)) = (\mathbb{R}, \mathfrak{B})$, i.e., convergence in distribution for random variables. The Central Limit Theorem deals with this notion of convergence, see the introductory Example I.1 and Section **??**.**??**.

Notation: for any $Q \in \mathfrak{M}(\mathbb{R})$

$$F_Q(x) = Q(]-\infty, x]), \qquad x \in \mathbb{R},$$

and for any function $F : \mathbb{R} \to \mathbb{R}$

$$\text{Cont}(F) = \{x \in \mathbb{R} : F \text{ continuous at } x\}.$$

Theorem 2.

$$Q_n \xrightarrow{\text{w}} Q \quad \Leftrightarrow \quad \forall\, x \in \text{Cont}(F_Q) : \lim_{n\to\infty} F_{Q_n}(x) = F_Q(x).$$

Moreover, if $Q_n \xrightarrow{\text{w}} Q$ and $\text{Cont}(F_Q) = \mathbb{R}$ then

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |F_{Q_n}(x) - F_Q(x)| = 0.$$

Proof. '⇒': If $x \in \text{Cont}(F_Q)$ and $A =]-\infty, x]$ then $Q(\partial A) = Q(\{x\}) = 0$, see Theorem 1.2. Hence Theorem 1 implies

$$\lim_{n\to\infty} F_{Q_n}(x) = \lim_{n\to\infty} Q_n(A) = Q(A) = F_Q(x).$$

'⇐': Consider a non-empty open set $A \subset \mathbb{R}$. Take pairwise disjoint open intervals $A_1, A_2, \dots$ such that $A = \bigcup_{i=1}^\infty A_i$. Fatou's Lemma implies

$$\liminf_{n\to\infty} Q_n(A) = \liminf_{n\to\infty} \sum_{i=1}^\infty Q_n(A_i) \geq \sum_{i=1}^\infty \liminf_{n\to\infty} Q_n(A_i).$$

Note that $\mathbb{R} \setminus \text{Cont}(F_Q)$ is countable. Fix $\varepsilon > 0$, and take

$$A_i' =]a_i', b_i'] \subset A_i$$

for $i \in \mathbb{N}$ such that

$$a_i', b_i' \in \text{Cont}(F_Q) \;\wedge\; Q(A_i) \leq Q(A_i') + \varepsilon \cdot 2^{-i}.$$

Then

$$\liminf_{n\to\infty} Q_n(A_i) \geq \liminf_{n\to\infty} Q_n(A_i') = Q(A_i') \geq Q(A_i) - \varepsilon \cdot 2^{-i}.$$

We conclude that

$$\liminf_{n\to\infty} Q_n(A) \geq Q(A) - \varepsilon,$$

and therefore $Q_n \xrightarrow{\text{w}} Q$ by Theorem 1.

Uniform convergence, Übung 9.1. □

Corollary 1.

$$Q_n \xrightarrow{\text{w}} Q \ \wedge \ Q_n \xrightarrow{\text{w}} \widetilde{Q} \quad \Rightarrow \quad Q = \widetilde{Q}.$$

Proof. By Theorem 2 $F_Q(x) = F_{\widetilde{Q}}(x)$ if $x \in D = \mathrm{Cont}(F_Q) \cap \mathrm{Cont}(F_{\widetilde{Q}})$. Since D is dense in $\mathbb{R}$ and F_Q as well as $F_{\widetilde{Q}}$ are right-continuous, we get $F_Q = F_{\widetilde{Q}}$. Apply Theorem 1.3. □

Given: random variables X_n, X on $(\Omega, \mathfrak{A}, P)$ for $n \in \mathbb{N}$.

Theorem 3.

$$X_n \xrightarrow{P} X \quad \Rightarrow \quad X_n \xrightarrow{\text{d}} X$$

and

$$X_n \xrightarrow{\text{d}} X \ \wedge \ X \text{ constant a.s.} \quad \Rightarrow \quad X_n \xrightarrow{P} X.$$

Proof. Assume $X_n \xrightarrow{P} X$. For $\varepsilon > 0$ and $x \in \mathbb{R}$

$$\begin{aligned}
&P(\{X \leq x - \varepsilon\}) - P(\{|X - X_n| > \varepsilon\}) \\
&\qquad \leq P(\{X \leq x - \varepsilon\} \cap \{|X - X_n| \leq \varepsilon\}) \\
&\qquad \leq P(\{X_n \leq x\}) \\
&\qquad = P(\{X_n \leq x\} \cap \{X \leq x + \varepsilon\}) + P(\{X_n \leq x\} \cap \{X > x + \varepsilon\}) \\
&\qquad \leq P(\{X \leq x + \varepsilon\}) + P(\{|X - X_n| > \varepsilon\}).
\end{aligned}$$

Thus

$$F_X(x - \varepsilon) \leq \liminf_{n\to\infty} F_{X_n}(x) \leq \limsup_{n\to\infty} F_{X_n}(x) \leq F_X(x + \varepsilon).$$

For $x \in \mathrm{Cont}(F_X)$ we get $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$. Apply Theorem 2.

Now, assume that $X_n \xrightarrow{\text{d}} X$ and $P_X = \varepsilon_x$. Let $\varepsilon > 0$ and take $f \in C^b(\mathbb{R})$ such that $f \geq 0$, $f(x) = 0$, and $f(y) = 1$ if $|x - y| \geq \varepsilon$. Then

$$P(\{|X - X_n| > \varepsilon\}) = P(\{|x - X_n| > \varepsilon\}) = \int 1_{\mathbb{R}\setminus[x-\varepsilon, x+\varepsilon]} \, dP_{X_n} \leq \int f \, dP_{X_n}$$

and

$$\lim_{n\to\infty} \int f \, dP_{X_n} = \int f \, dP_X = 0.$$

□

Example 2. Consider the uniform distribution P on $\Omega = \{0, 1\}$. Put

$$X_n(\omega) = \omega, \qquad X(\omega) = 1 - \omega.$$

Then $P_{X_n} = P_X$ and therefore

$$X_n \xrightarrow{\text{d}} X.$$

However, $\{|X_n - X| < 1/2\} = \emptyset$ and therefore

$$X_n \xrightarrow{P} X \text{ does not hold.}$$

Theorem 4 (Skorohod)**.** There exists a probability space $(\Omega, \mathfrak{A}, P)$ with the following property. If

$$Q_n \xrightarrow{\text{w}} Q,$$

then there exist X_n, $X \in \mathfrak{Z}(\Omega, \mathfrak{A})$ for $n \in \mathbb{N}$ such that

$$\forall\, n \in \mathbb{N}: \; Q_n = P_{X_n} \quad \wedge \quad Q = P_X \quad \wedge \quad X_n \xrightarrow{P\text{-a.s.}} X.$$

Proof. Take $\Omega = \,]0, 1[$, $\mathfrak{A} = \mathfrak{B}(\Omega)$, and consider the uniform distribution P on Ω. Define

$$X_Q(\omega) = \inf\{z \in \mathbb{R} : \omega \leq F_Q(z)\}, \qquad \omega \in \,]0, 1[\,,$$

for any $Q \in \mathfrak{M}(\mathbb{R})$. Since X_Q is non-decreasing, we have $X_Q \in \mathfrak{Z}(\Omega, \mathfrak{A})$. It turns out that

$$P_{X_Q} = Q, \tag{1}$$

see Übung 9.2. Moreover, if $Q_n \xrightarrow{\text{w}} Q$ then $X_{Q_n} \xrightarrow{P\text{-a.s.}} X_Q$, see Gänssler, Stute (1977, p. 67–68). □

Remark 3. By (1) we have a general method to transform uniformly distributed 'random numbers' from $]0, 1[$ into 'random numbers' with distribution Q.

Remark 4.

(i) Put

$$C^{(r)} = \{f : \mathbb{R} \to \mathbb{R} : f, f^{(1)}, \dots, f^{(r)} \text{ bounded, uniformly continuous}\}.$$

Then

$$Q_n \xrightarrow{\text{w}} Q \quad \Leftrightarrow \quad \exists\, r \in \mathbb{N}_0 \;\forall\, f \in C^{(r)} : \; \lim_{n \to \infty} \int f \, dQ_n = \int f \, dQ,$$

see Gänssler, Stute (1977, p. 66).

(ii) The *Lévy distance*

$$d(Q, R) = \inf\{h \in \,]0, \infty[\, : \forall\, x \in \mathbb{R} : F_Q(x - h) - h \leq F_R(x) \leq F_Q(x + h) + h\}$$

defines a metric on $\mathfrak{M}(\mathbb{R})$, and

$$Q_n \xrightarrow{\text{w}} Q \quad \Leftrightarrow \quad \lim_{n \to \infty} d(Q_n, Q) = 0,$$

see Chow, Teicher (1978, Thm. 8.1.3).

(iii) Suppose that (M, ρ) is a complete separable metric space. Then there exists a metric d on $\mathfrak{M}(M)$ such that $(\mathfrak{M}(M), d)$ is complete and separable as well, and

$$Q_n \xrightarrow{\text{w}} Q \quad \Leftrightarrow \quad \lim_{n\to\infty} d(Q_n, Q) = 0,$$

see Parthasarathy (1967, Sec. II.6).

Finally, we present a compactness criterion, which is very useful for construction of probability measures on $\mathfrak{B}(M)$.

Lemma 1. Let $x_{n,\ell} \in \mathbb{R}$ for $n, \ell \in \mathbb{N}$ with

$$\forall\, \ell \in \mathbb{N}: \ \sup_{n\in\mathbb{N}} |x_{n,\ell}| < \infty.$$

Then there exists an increasing sequence $(n_i)_{i\in\mathbb{N}}$ in $\mathbb{N}$ such that

$$\forall\, \ell \in \mathbb{N}: \ (x_{n_i,\ell})_{i\in\mathbb{N}} \text{ converges.}$$

Proof. See Billingsley (1979, Thm. 25.13). □

Definition 2.

(i) $\mathfrak{P} \subset \mathfrak{M}(M)$ *tight* if

$$\forall\, \varepsilon > 0\ \exists\, K \subset M \text{ compact } \forall\, P \in \mathfrak{P}: \quad P(K) \geq 1 - \varepsilon.$$

(ii) $\mathfrak{P} \subset \mathfrak{M}(M)$ *relatively compact* if every sequence in $\mathfrak{P}$ contains a subsequence that converges weakly.

Theorem 5 (Prohorov)**.** Assume that M is a complete separable metric space and $\mathfrak{P} \subset \mathfrak{M}(M)$. Then

$$\mathfrak{P} \text{ relatively compact} \quad \Leftrightarrow \quad \mathfrak{P} \text{ tight.}$$

Proof. See Parthasarathy (1967, Thm. II.6.7). Here: $M = \mathbb{R}$.

'$\Rightarrow$': Suppose that $\mathfrak{P}$ is not tight. Then, for some $\varepsilon > 0$, there exists a sequence $(P_n)_{n\in\mathbb{N}}$ in $\mathfrak{P}$ such that

$$P_n([-n, n]) < 1 - \varepsilon.$$

For a suitable subsequence, $P_{n_k} \xrightarrow{\text{w}} P \in \mathfrak{M}(\mathbb{R})$. Take $m > 0$ such that

$$P(]-m, m[) > 1 - \varepsilon.$$

Theorem 1 implies

$$P(]-m, m[) \leq \liminf_{k\to\infty} P_{n_k}(]-m, m[) \leq \liminf_{k\to\infty} P_{n_k}([-n_k, n_k]) < 1 - \varepsilon,$$

which is a contradiction.

'$\Leftarrow$': Consider any sequence $(P_n)_{n\in\mathbb{N}}$ in $\mathfrak{P}$ and the corresponding sequence $(F_n)_{n\in\mathbb{N}}$ of distribution functions. Use Lemma 1 to obtain a subsequence $(F_{n_i})_{i\in\mathbb{N}}$ and a non-decreasing function $G : \mathbb{Q} \to [0,1]$ with

$$\forall\, q \in \mathbb{Q} : \lim_{i\to\infty} F_{n_i}(q) = G(q).$$

Put

$$F(x) = \inf\{G(q) : q \in \mathbb{Q} \wedge x < q\}, \qquad x \in \mathbb{R}.$$

Claim (*Helly's Theorem*):

(i) F is non-decreasing and right-continuous,

(ii) $\forall\, x \in \mathrm{Cont}(F) : \lim_{i\to\infty} F_{n_i}(x) = F(x)$.

Proof: Ad (i): Obviously F is non-decreasing. For $x \in \mathbb{R}$ and $\varepsilon > 0$ take $\delta_2 > 0$ such that

$$\forall\, q \in \mathbb{Q} \cap\,]x, x+\delta_2[\, : \quad G(q) \le F(x) + \varepsilon.$$

Thus, for $z \in\,]x, x+\delta_2[$,

$$F(x) \le F(z) \le F(x) + \varepsilon.$$

Ad (ii): If $x \in \mathrm{Cont}(F)$ and $\varepsilon > 0$ take $\delta_1 > 0$ such that

$$F(x) - \varepsilon \le F(x - \delta_1).$$

Thus, for $q_1, q_2 \in \mathbb{Q}$ with

$$x - \delta_1 < q_1 < x < q_2 < x + \delta_2,$$

we get

$$\begin{aligned} F(x) - \varepsilon \le F(x-\delta_1) \le G(q_1) &\le \liminf_{i\to\infty} F_{n_i}(x) \le \limsup_{i\to\infty} F_{n_i}(x) \\ &\le G(q_2) \le F(x) + \varepsilon. \end{aligned}$$

Claim:

$$\lim_{x\to-\infty} F(x) = 0 \ \wedge \ \lim_{x\to\infty} F(x) = 1.$$

Proof: For $\varepsilon > 0$ take $m \in \mathbb{Q}$ such that

$$\forall\, n \in \mathbb{N} : P_n(]-m, m]) \ge 1 - \varepsilon.$$

Thus

$$G(m) - G(-m) = \lim_{i\to\infty} \big(F_{n_i}(m) - F_{n_i}(-m)\big) = \lim_{i\to\infty} P_{n_i}(]-m, m]) \ge 1 - \varepsilon.$$

Since $F(m) \ge G(m)$ and $F(-m-1) \le G(-m)$, we obtain

$$F(m) - F(-m-1) \ge 1 - \varepsilon.$$

It remains to apply Theorems 1.3 and 2. $\square$