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)*

$$\mathfrak{L}(X) := P_X(A') := P(\{X^{-1}(A')\}), \qquad A' \in \mathfrak{A}'$$

Notation: Q Prob. measure on $(\Omega', \mathfrak{A}')$, then $X \sim Q$ iff $P_X = Q$.

(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) 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 $Y(\omega) = 1 - \omega$.

(ii) 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 $X \sim Q$: Choose $(\Omega, \mathfrak{A}, P) = (\Omega', \mathfrak{A}', Q)$ and $X = \mathrm{id}_\Omega$.

(iii) 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 Statistics.
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 Statistics.
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 $X = (X_1, \ldots, X_k)$ be a random vector. Then

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

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. '$\Leftarrow$': $(P_1)_{X^1}$ and $(P_2)_{X^2}(A)$ coincide by assumption on the $\cap$–stable class

$$\mathfrak{E} = \Big\{\prod_{i=1}^{k}]-\infty, x_i] : x_1, \ldots, x_k \in \mathbb{R}\Big\}\ ;$$

hence by Theorem II.4.4 they coincide on $\sigma(\mathfrak{E}) = \mathfrak{B}_k$ (see Remark II.1.6). □

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 4.1 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 4.1.b). □