7 The Radon-Nikodym-Theorem

Given: a measure space $(\Omega, \mathfrak{A}, \mu)$. Put $\overline{\mathfrak{Z}}_+ = \overline{\mathfrak{Z}}_+(\Omega, \mathfrak{A})$.

Definition 1. For f (quasi-)μ-integrable and $A \in \mathfrak{A}$, the *integral of f over A* is

$$\int_A f \, d\mu = \int 1_A \cdot f \, d\mu.$$

(Note: $|1_A \cdot f| \leq |f|$.)

Theorem 1. Let $f \in \overline{\mathfrak{Z}}_+$ and put

$$\nu(A) = \int_A f \, d\mu, \qquad A \in \mathfrak{A}.$$

Then ν is a measure on $\mathfrak{A}$.

Proof. Clearly $\nu(\emptyset) = 0$ and $\nu \geq 0$. For $A_1, A_2, \ldots \in \mathfrak{A}$ pairwise disjoint

$$\begin{aligned}\nu\Big(\bigcup_{i=1}^{\infty} A_i\Big) &= \int \sum_{i=1}^{\infty} 1_{A_i} \cdot f \, d\mu = \int \lim_{n\to\infty} \Big(\sum_{i=1}^{n} 1_{A_i} \cdot f\Big) \, d\mu \\ &= \lim_{n\to\infty} \int \sum_{i=1}^{n} 1_{A_i} \cdot f \, d\mu = \sum_{i=1}^{\infty} \int 1_{A_i} \cdot f \, d\mu \\ &= \sum_{i=1}^{\infty} \nu(A_i)\end{aligned}$$

follows from Theorem 5.1. □

Definition 2. The mapping ν in Theorem 1 is called *measure with μ-density f*. Notation: $\nu = f \cdot \mu$. If $\int f \, d\mu = 1$ then f is called *probability density*.

Example 1. The introductory examples of probability spaces were defined by means of probability densities.

(i) Let $(\Omega, \mathfrak{A}, \mu) = (\mathbb{R}^k, \mathfrak{B}_k, \lambda_k)$. For

$$f(x) = (2\pi)^{-k/2} \cdot \exp\Big(-\tfrac{1}{2} \textstyle\sum_{i=1}^{k} x_i^2\Big)$$

we get the *k-dimensional standard normal distribution ν*.

For $B \in \mathfrak{B}_k$ such that $0 < \lambda_k(B) < \infty$ and

$$f = \frac{1}{\lambda_k(B)} \cdot 1_B$$

we get the *uniform distribution on B*.

(ii) Suppose that Ω is countable, $\mathfrak{A} = \mathfrak{P}(\Omega)$, and μ is the counting measure on $\mathfrak{A}$. Take $f : \Omega \to \mathbb{R}_+ \cup \{\infty\}$ and use Theorem 5.1 to obtain

$$\forall\, A \in \mathfrak{A} : \nu(A) = \int_A f \, d\mu = \sum_{\omega \in A} f(\omega). \tag{1}$$

Conversely, for any measure ν on $\mathfrak{A}$ put $f(\omega) = \nu(\{\omega\})$. Then we have (1).

Theorem 2. Let $\nu = f \cdot \mu$ with $f \in \overline{\mathfrak{Z}}_+$ and $g \in \overline{\mathfrak{Z}}(\Omega, \mathfrak{A})$. Then

$$g \text{ (quasi)-}\nu\text{-integrable} \quad \Leftrightarrow \quad g \cdot f \text{ (quasi)-}\mu\text{-integrable},$$

in which case

$$\int g \, d\nu = \int g \cdot f \, d\mu$$

Proof. First, assume that $g = 1_A$ with $A \in \mathfrak{A}$. Then the statements hold by definition. For $g \in \mathfrak{S}_+(\Omega, \mathfrak{A})$ we now use linearity of the integral. For $g \in \overline{\mathfrak{Z}}_+$ we take a sequence $(g_n)_{n \in \mathbb{N}}$ in $\mathfrak{S}_+(\Omega, \mathfrak{A})$ such that $g_n \uparrow g$. Then $g_n \cdot f \in \overline{\mathfrak{Z}}_+$ and $g_n \cdot f \uparrow g \cdot f$. Hence, by Theorem 5.1 and the previous part of the proof

$$\int g \, d\nu = \lim_{n \to \infty} \int g_n \, d\nu = \lim_{n \to \infty} \int g_n \cdot f \, d\mu = \int g \cdot f \, d\mu.$$

Finally, for $g \in \overline{\mathfrak{Z}}(\Omega, \mathfrak{A})$ we already know that

$$\int g^{\pm} d\nu = \int g^{\pm} \cdot f \, d\mu = \int (g \cdot f)^{\pm} \, d\mu.$$

Use linearity of the integral. □

Remark 1.

$$f, g \in \overline{\mathfrak{Z}}_+ \wedge f = g \ \mu\text{-a.e.} \quad \Rightarrow \quad f \cdot \mu = g \cdot \mu.$$

Theorem 3 (Uniqueness of densities)**.** Let $f, g \in \overline{\mathfrak{Z}}_+$ such that $f \cdot \mu = g \cdot \mu$. Then

(i) f μ-integrable $\Rightarrow f = g$ μ-a.e.,

(ii) μ σ-finite $\Rightarrow f = g$ μ-a.e.

Proof. Ad (i): It suffices to verify that

$$f, g \ \mu\text{-integrable} \wedge \left(\forall\, A \in \mathfrak{A} : \int_A f \, d\mu \le \int_A g \, d\mu \right) \quad \Rightarrow \quad f \le g \ \mu\text{-a.e.}$$

To this end, take $A = \{f > g\}$. By assumption,

$$-\infty < \int_A f \, d\mu \le \int_A g \, d\mu < \infty$$

and therefore $\int_A (f - g) \, d\mu \le 0$. However,

$$1_A \cdot (f - g) \ge 0,$$

hence $\int_A (f-g)\,d\mu \geq 0$. Thus

$$\int 1_A \cdot (f-g)\,d\mu = 0.$$

Theorem 5.3 implies $1_A \cdot (f-g) = 0$ μ-a.e., and by definition of A we get $\mu(A) = 0$.
Ad (ii): see Elstrodt (1996, p. 141). □

Remark 2. Let $(\Omega, \mathfrak{A}, \mu) = (\mathbb{R}^k, \mathfrak{B}_k, \lambda_k)$ and $x \in \mathbb{R}^k$. There is no density $f \in \overline{\mathfrak{Z}}_+$ w.r.t. λ_k such that $\varepsilon_x = f \cdot \lambda_k$. This follows from $\varepsilon_x(\{x\}) = 1$ and

$$(f \cdot \lambda_k)(\{x\}) = \int_{\{x\}} f\,d\lambda_k = 0.$$

Definition 3. A measure ν on $\mathfrak{A}$ is *absolutely continuous w.r.t.* μ if

$$\forall\, A \in \mathfrak{A} : \mu(A) = 0 \Rightarrow \nu(A) = 0.$$

Notation: $\nu \ll \mu$.

Remark 3.

(i) $\nu = f \cdot \mu \Rightarrow \nu \ll \mu$.

(ii) In Remark 2 neither $\varepsilon_x \ll \lambda_1$ nor $\lambda_1 \ll \varepsilon_x$.

(iii) Let μ denote the counting measure on $\mathfrak{A}$. Then $\nu \ll \mu$ for every measure ν on $\mathfrak{A}$.

(iv) Let μ denote the counting measure on $\mathfrak{B}_1$. Then there is no density $f \in \overline{\mathfrak{Z}}_+$ such that $\lambda_1 = f \cdot \mu$.

Lemma 1. Let $f_n \xrightarrow{\mathfrak{L}^p} f$ and $A \in \mathfrak{A}$. If $p = 1$ or $\mu(A) < \infty$ then

$$\int_A f_n\,d\mu \to \int_A f\,d\mu.$$

Proof. Übung 6.2. See also L^p and its dual space in Elstrodt (1996, §VII.3), e.g. □

Theorem 4 (Radon, Nikodym)**.** For every σ-finite measure μ and every measure ν on $\mathfrak{A}$ we have

$$\nu \ll \mu \quad \Rightarrow \quad \exists\, f \in \overline{\mathfrak{Z}}_+ : \nu = f \cdot \mu.$$

Proof. See Elstrodt (1996, §VII.2).
Here we consider the particular case

$$\forall\, A \in \mathfrak{A} : \nu(A) \leq \mu(A) \wedge \mu(\Omega) < \infty.$$

A class $\mathfrak{U} = \{A_1, \dots, A_n\}$ is called a (finite measurable) partition of Ω if $A_1, \dots, A_n \in \mathfrak{A}$ are pairwise disjoint and $\bigcup_{i=1}^n A_i = \Omega$. The set of all partitions is partially ordered by

$$\mathfrak{U} \sqsubset \mathfrak{V} \quad \text{if} \quad \forall\, A \in \mathfrak{U}\ \exists\, B \in \mathfrak{V} : A \subset B.$$

The infimum of two partitions is given by

$$\mathfrak{U} \wedge \mathfrak{V} = \{A \cap B : A \in \mathfrak{U}, B \in \mathfrak{V}\}.$$

For any partition $\mathfrak{U}$ we define

$$f_{\mathfrak{U}} = \sum_{A \in \mathfrak{U}} \alpha_A \cdot 1_A$$

with

$$\alpha_A = \begin{cases} \nu(A)/\mu(A) & \text{if } \mu(A) > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Clearly $f_{\mathfrak{U}} \in \mathfrak{S}_+(\Omega, \sigma(\mathfrak{U})) \subset \mathfrak{S}_+(\Omega, \mathfrak{A})$, $\sigma(\mathfrak{U}) = \mathfrak{U}^+ \cup \{\emptyset\}$, and

$$\forall\, A \in \sigma(\mathfrak{U}) : \nu(A) = \int_A f_{\mathfrak{U}}\, d\mu.$$

(Thus we have $\nu|_{\sigma(\mathfrak{U})} = f_{\mathfrak{U}} \cdot \mu|_{\sigma(\mathfrak{U})}$.) Let $\mathfrak{U} \sqsubset \mathfrak{V}$ and $A \in \mathfrak{V}$. Then

$$\nu(A) = \int_A f_{\mathfrak{V}}\, d\mu = \int_A f_{\mathfrak{U}}\, d\mu,$$

since $A \in \sigma(\mathfrak{U})$. Hence

$$\int_A f_{\mathfrak{V}}^2\, d\mu = \int_A f_{\mathfrak{V}} \cdot f_{\mathfrak{U}}\, d\mu,$$

since $f_{\mathfrak{V}}|_A$ is constant, and therefore

$$0 \le \int (f_{\mathfrak{U}} - f_{\mathfrak{V}})^2\, d\mu = \int f_{\mathfrak{U}}^2\, d\mu - \int f_{\mathfrak{V}}^2\, d\mu. \tag{2}$$

Put

$$\beta = \sup\left\{ \int f_{\mathfrak{U}}^2\, d\mu : \mathfrak{U} \text{ partition} \right\},$$

and note that $0 \le \beta \le \mu(\Omega) < \infty$, since $f_{\mathfrak{U}} \le 1$. Consider a sequence of functions $f_n = f_{\mathfrak{U}_n}$ such that

$$\lim_{n\to\infty} \int f_n^2\, d\mu = \beta.$$

Due to (2) we may assume that $\mathfrak{U}_{n+1} \sqsubset \mathfrak{U}_n$. Then, by (2), $(f_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathfrak{L}^2$, so that there exists $f \in \mathfrak{L}^2$ with

$$\lim_{n\to\infty} \|f_n - f\|_2 = 0 \quad \wedge \quad 0 \le f \le 1 \ \mu\text{-a.e.},$$

see Theorem 6.3.

We claim that $\nu = f \cdot \mu$. Let $A \in \mathfrak{A}$. Put

$$\widetilde{\mathfrak{U}}_n = \mathfrak{U}_n \wedge \{A, A^c\}$$

and

$$\widetilde{f}_n = f_{\widetilde{\mathfrak{U}}_n}.$$

Then

$$\nu(A) = \int_A \widetilde{f}_n\, d\mu = \int_A f_n\, d\mu + \int_A (\widetilde{f}_n - f_n)\, d\mu,$$

and (2) yields $\lim_{n\to\infty} \|\widetilde{f}_n - f_n\|_2 = 0$. It remains to apply Lemma 1. □

8 Kernels and Product Measures

Given: measurable spaces $(\Omega_1, \mathfrak{A}_1)$ and $(\Omega_2, \mathfrak{A}_2)$.
Motivation: two-stage experiment. Output $\omega_1 \in \Omega_1$ of the first stage determines probabilistic model for the second stage.

Example 1. Choose one out of n coins and throw it once. Parameters $a_1, \dots, a_n \geq 0$ such that $\sum_{i=1}^n a_i = 1$ and $b_1, \dots, b_n \in [0,1]$.
Let

$$\Omega_1 = \{1, \dots, n\}, \qquad \mathfrak{A}_1 = \mathfrak{P}(\Omega_1)$$

and define

$$\mu(\{i\}) = a_i, \qquad i \in \Omega_1,$$

to be the probability of choosing the i-th coin. Moreover, let

$$\Omega_2 = \{\mathrm{H}, \mathrm{T}\}, \qquad \mathfrak{A}_2 = \mathfrak{P}(\Omega_2)$$

and define

$$K(i, \{\mathrm{H}\}) = b_i$$

to be the probability for obtaining H when throwing the i-th coin. Thus, for $A_2 \in \mathfrak{A}_2$,

$$K(i, A_2) = b_i \cdot \varepsilon_{\mathrm{H}}(A_2) + (1 - b_i) \cdot \varepsilon_{\mathrm{T}}(A_2).$$

Definition 1. $K : \Omega_1 \times \mathfrak{A}_2 \to \overline{\mathbb{R}}$ is a *kernel (from* $(\Omega_1, \mathfrak{A}_1)$ *to* $(\Omega_2, \mathfrak{A}_2)$*)*, if

(i) $K(\omega_1, \cdot)$ is a measure on $\mathfrak{A}_2$ for every $\omega_1 \in \Omega_1$,

(ii) $K(\cdot, A_2)$ is $\mathfrak{A}_1$-$\overline{\mathfrak{B}}$-measurable for every $A_2 \in \mathfrak{A}_2$.

K is a *Markov (transition) kernel*, if, additionally, $K(\omega_1, \Omega_2) = 1$ for every $\omega_1 \in \Omega_1$.
K is a *σ-finite* kernel if, additionally,

$$\exists\, A_{2,1}, A_{2,2}, \ldots \in \mathfrak{A}_2 \text{ pairwise disjoint}:$$
$$\Omega_2 = \bigcup_{i=1}^{\infty} A_{2,i} \quad \wedge \quad \forall\, i \in \mathbb{N} : \sup_{\omega_1 \in \Omega_1} K(\omega_1, A_{2,i}) < \infty.$$

Example 2. Extremal cases, non-disjoint.

(i) Model for the second stage not influenced by output of the first stage, i.e., for a (probability) measure ν on $\mathfrak{A}_2$

$$\forall\, \omega_1 \in \Omega_1 : \quad K(\omega_1, \cdot) = \nu.$$

In Example 1 this means $b_1 = \cdots = b_n$.

(ii) Output of the first stage determines the output of the second stage, i.e., for a $\mathfrak{A}_1$-$\mathfrak{A}_2$-measurable mapping $f : \Omega_1 \to \Omega_2$

$$\forall\, \omega_1 \in \Omega_1 : \quad K(\omega_1, \cdot) = \varepsilon_{f(\omega_1)}.$$

In Example 1 this means $b_1, \dots, b_n \in \{0, 1\}$.

Notation: $\int f\, d\mu = \int_\Omega f(\omega)\, \mu(d\omega)$.
Given: a (probability) measure μ on $\mathfrak{A}_1$ and a (Markov) kernel K from $(\Omega_1, \mathfrak{A}_1)$ to $(\Omega_2, \mathfrak{A}_2)$. Question: stochastic model $(\Omega, \mathfrak{A}, P)$ for a compound experiment? Reasonable, and assumed in the sequel,

$$\Omega = \Omega_1 \times \Omega_2, \qquad \mathfrak{A} = \mathfrak{A}_1 \otimes \mathfrak{A}_2.$$

Question: How to define P?

Example 3. In Example 1, a reasonable requirement for P is

$$P(\{i\} \times \Omega_2) = a_i, \qquad P(\{i\} \times \{\mathrm{H}\}) = a_i \cdot b_i$$

for every $i \in \Omega_1$. Consequently, for $A_2 \subset \Omega_2$

$$P(\{i\} \times A_2) = K(i, A_2) \cdot a_i$$

and for $A \subset \Omega$

$$\begin{aligned} P(A) &= \sum_{i=1}^n P(\{(\omega_1, \omega_2) \in A : \omega_1 = i\}) = \sum_{i=1}^n P(\{i\} \times \{\omega_2 \in \Omega_2 : (i, \omega_2) \in A\}) \\ &= \sum_{i=1}^n K(i, \{\omega_2 \in \Omega_2 : (i, \omega_2) \in A\}) \cdot a_i \\ &= \int_{\Omega_1} K(i, \{\omega_2 \in \Omega_2 : (i, \omega_2) \in A\})\, \mu(di). \end{aligned}$$

May we generally use the right-hand side integral for the definition of P?

Lemma 1. Let $f \in \overline{\mathfrak{Z}}(\Omega, \mathfrak{A})$. Then, for $\omega_1 \in \Omega_1$, the ω_1-*section*

$$f(\omega_1, \cdot) : \Omega_2 \to \overline{\mathbb{R}}$$

of f is $\mathfrak{A}_2$-$\overline{\mathfrak{B}}$-measurable, and for $\omega_2 \in \Omega_2$ the ω_2-*section*

$$f(\cdot, \omega_2) : \Omega_1 \to \overline{\mathbb{R}}$$

of f is $\mathfrak{A}_1$-$\overline{\mathfrak{B}}$-measurable.

Proof. In the case of an ω_1-section. Fix $\omega_1 \in \Omega_1$. Then $\Omega_2 \to \Omega_1 \times \Omega_2 : \ \omega_2 \mapsto (\omega_1, \omega_2)$ is $\mathfrak{A}_2$-$\mathfrak{A}$-measurable due to Corollary 3.1.(i). Apply Theorem 2.1. □

Remark 1. In particular, for $A \in \mathfrak{A}$ and $f = 1_A$

$$f(\omega_1, \cdot) = 1_A(\omega_1, \cdot) = 1_{A(\omega_1)}$$

where[1]

$$A(\omega_1) = \{\omega_2 \in \Omega_2 : (\omega_1, \omega_2) \in A\}$$

[1] poor notation

is the ω_1-*section* of A. By Lemma 1

$$\forall\, \omega_1 \in \Omega_1 : \quad A(\omega_1) \in \mathfrak{A}_2.$$

Analogously for the ω_2-*section*

$$A(\omega_2) = \{\omega_1 \in \Omega_1 : (\omega_1, \omega_2) \in A\}$$

of A.

Given:

- a σ-finite kernel K from $(\Omega_1, \mathfrak{A}_1)$ to $(\Omega_2, \mathfrak{A}_2)$,
- a σ-finite measure μ on $\mathfrak{A}_1$.

Lemma 2. Let $f \in \overline{\mathfrak{Z}}_+(\Omega, \mathfrak{A})$. Then

$$g: \ \Omega_1 \to \mathbb{R}_+ \cup \{\infty\}, \quad \omega_1 \mapsto \int_{\Omega_2} f(\omega_1, \omega_2)\, K(\omega_1, d\omega_2)$$

is $\mathfrak{A}_1$-$\mathfrak{B}([0,\infty])$-measurable.

Proof. First we additionally assume

$$\forall\, \omega_1 \in \Omega_1 : K(\omega_1, \Omega_2) < \infty. \tag{1}$$

Let $\mathfrak{F}$ denote the set of all functions $f \in \overline{\mathfrak{Z}}_+(\Omega, \mathfrak{A})$ with the measurability property as claimed. We show that

$$\forall\, A_1 \in \mathfrak{A}_1,\ A_2 \in \mathfrak{A}_2 : \quad 1_{A_1 \times A_2} \in \mathfrak{F}. \tag{2}$$

Indeed,

$$\int_{\Omega_2} 1_{A_1 \times A_2}(\omega_1, \omega_2) K(\omega_1, d\omega_2) = 1_{A_1}(\omega_1) K(\omega_1, A_2).$$

Furthermore, we show that

$$\forall\, A \in \mathfrak{A} : \quad 1_A \in \mathfrak{F}. \tag{3}$$

To this end let

$$\mathfrak{D} = \{A \in \mathfrak{A} : 1_A \in \mathfrak{F}\}$$

and

$$\mathfrak{E} = \{A_1 \times A_2 : A_1 \in \mathfrak{A}_1 \wedge A_2 \in \mathfrak{A}_2\}.$$

Then $\mathfrak{E} \subset \mathfrak{D}$ by (2), $\mathfrak{E}$ is closed w.r.t. intersections, and $\sigma(\mathfrak{E}) = \mathfrak{A}$. From (1) it easily follows that $\mathfrak{D}$ is a Dynkin class. Hence Theorem 1.2 yields

$$\mathfrak{A} = \sigma(\mathfrak{E}) = \delta(\mathfrak{E}) \subset \mathfrak{D} \subset \mathfrak{A},$$

which implies (3). From Lemma 5.2 and Theorem 2.6 we get

$$f_1, f_2 \in \mathfrak{F} \wedge \alpha \in \mathbb{R}_+ \quad \Rightarrow \quad \alpha f_1 + f_2 \in \mathfrak{F}. \tag{4}$$

Finally, Theorem 5.1 and Theorem 2.5.(iii) imply that

$$f_n \in \mathfrak{F} \wedge f_n \uparrow f \quad \Rightarrow \quad f \in \mathfrak{F}. \tag{5}$$

Use Theorem 2.7 together with (3)–(5) to conclude that $\mathfrak{F} = \overline{\mathfrak{Z}}_+$.
In the general case we take $A_{2,1}, A_{2,2}, \ldots \in \mathfrak{A}_2$ pairwise disjoint such that

$$\bigcup_{i=1}^{\infty} A_{2,i} = \Omega_2 \quad \wedge \quad \forall\, i \in \mathbb{N} : \sup_{\omega_1 \in \Omega_1} K(\omega_1, A_{2,i}) < \infty,$$

and we define

$$K_i(\omega_1, \cdot) = K(\omega_1, \cdot \cap A_{2,i}) = 1_{A_{2,i}} \cdot K(\omega_1, \cdot).$$

Then, using Theorems 5.1 and 7.2,

$$\begin{aligned} \int_{\Omega_2} f(\omega_1, \omega_2)\, K(\omega_1, d\omega_2) &= \sum_{i=1}^{\infty} \int_{\Omega_2} 1_{A_{2,i}}(\omega_2)\, f(\omega_1, \omega_2)\, K(\omega_1, d\omega_2) \\ &= \sum_{i=1}^{\infty} \int_{\Omega_2} f(\omega_1, \omega_2)\, K_i(\omega_1, d\omega_2). \end{aligned}$$

Since $K_i(\omega_1, \Omega_2) < \infty$ for every $\omega_1 \in \Omega_1$, we conclude that $\int_{\Omega_2} f(\cdot, \omega_2)\, K_i(\cdot, d\omega_2)$ is $\mathfrak{A}_1$-$\mathfrak{B}([0, \infty])$-measurable. Apply Theorems 2.5 and 2.6. □

Theorem 1.

$$\begin{aligned} &\exists_1 \text{ measure } \mu \times K \text{ on } \mathfrak{A} \ \ \forall\, A_1 \in \mathfrak{A}_1 \ \ \forall\, A_2 \in \mathfrak{A}_2 : \\ &\qquad \mu \times K(A_1 \times A_2) = \textstyle\int_{A_1} K(\omega_1, A_2)\, \mu(d\omega_1). \end{aligned} \tag{6}$$

Moreover, $\mu \times K$ is σ-finite, and

$$\forall\, A \in \mathfrak{A} : \quad \mu \times K(A) = \int_{\Omega_1} K(\omega_1, A(\omega_1))\, \mu(d\omega_1). \tag{7}$$

If μ is a probability measure and K is a Markov kernel then $\mu \times K$ is a probability measure, too.

Proof. 'Existence': For $A \in \mathfrak{A}$ and $\omega_1 \in \Omega_1$

$$K(\omega_1, A(\omega_1)) = \int_{\Omega_2} 1_{A(\omega_1)}(\omega_2)\, K(\omega_1, d\omega_2) = \int_{\Omega_2} 1_A(\omega_1, \omega_2)\, K(\omega_1, d\omega_2).$$

According to Lemma 8.2 $\mu \times K$ is well-defined via (7). Using Theorem 5.1, it is easy to verify that $\mu \times K$ is a measure on $\mathfrak{A}$.
For $A_1 \in \mathfrak{A}_1$ and $A_2 \in \mathfrak{A}_2$

$$K(\omega_1, (A_1 \times A_2)(\omega_1)) = \begin{cases} K(\omega_1, A_2) & \text{if } \omega_1 \in A_1 \\ 0 & \text{otherwise.} \end{cases}$$

Hence $\mu \times K$ satisfies (6).

By assumption we have $A_{1,1}, A_{1,2}, \ldots \in \mathfrak{A}_1$ pairwise disjoint such that

$$\bigcup_{i=1}^{\infty} A_{1,i} = \Omega_1 \quad \wedge \quad \forall\, i \in \mathbb{N} : \mu(A_{1,i}) < \infty$$

and $A_{2,1}, A_{2,2}, \ldots \in \mathfrak{A}_2$ pairwise disjoint such that

$$\bigcup_{j=1}^{\infty} A_{2,j} = \Omega_2 \quad \wedge \quad \forall\, j \in \mathbb{N} : \sup_{\omega_1 \in \Omega_1} K(\omega_1, A_{2,j}) < \infty.$$

Consider the sets $A_{1,i} \times A_{2,j}$ with $i, j \in \mathbb{N}$ and note that

$$\begin{aligned}(\mu \times K)(A_{1,i} \times A_{2,j}) &= \int_{A_{1,i}} K(\omega_1, A_{2,j})\, \mu(d\omega_1) \\ &\leq \sup_{\omega_1 \in \Omega_1} K(\omega_1, A_{2,j})\, \mu(A_{1,i}) < \infty,\end{aligned}$$

to conclude that $\mu \times K$ ist σ-finite.
'Uniqueness': Apply Theorem 4.4 with $\mathfrak{A}_0 = \{A_1 \times A_2 : A_i \in \mathfrak{A}_i\}$. □

Example 4. In Example 3 we have $P = \mu \times K$.

Remark 2. Particular case of Theorem 1 with

$$\mu = \mu_1, \qquad \forall\, \omega_1 \in \Omega_1 : \ K(\omega_1, \cdot) = \mu_2$$

for σ-finite measures μ_i on $(\Omega_i, \mathfrak{A}_i)$:

$$\begin{aligned}&\underset{1}{\exists}\,\text{measure } \mu_1 \times \mu_2 \text{ on } \mathfrak{A} \ \ \forall\, A_1 \in \mathfrak{A}_1 \ \ \forall\, A_2 \in \mathfrak{A}_2 : \\ &\quad \mu_1 \times \mu_2(A_1 \times A_2) = \mu_1(A_1) \cdot \mu_2(A_2).\end{aligned} \tag{8}$$

Moreover, $\mu_1 \times \mu_2$ is σ-finite and satisfies

$$\forall\, A \in \mathfrak{A} : \quad \mu_1 \times \mu_2(A) = \int_{\Omega_1} \mu_2(A(\omega_1))\, \mu(d\omega_1). \tag{9}$$

We add that σ-finiteness is used for the definition (9) and the uniqueness in (8). In general, we only have existence of a measure $\mu_1 \times \mu_2$ with (8). See Elstrodt (1996, §V.1).

Definition 2. $\mu = \mu_1 \times \mu_2$ is called the *product measure* corresponding to μ_1 and μ_2, and $(\Omega, \mathfrak{A}, \mu)$ is called the *product measure space* corresponding to $(\Omega_1, \mathfrak{A}_1, \mu_1)$ and $(\Omega_2, \mathfrak{A}_2, \mu_2)$.

Example 5.

(i) In Example 3 with $b = b_1 = \cdots = b_n$ and $\nu = b \cdot \varepsilon_{\mathrm{H}} + (1 - b) \cdot \varepsilon_{\mathrm{T}}$ we have $P = \mu \times \nu$.

(ii) For countable spaces Ω_i and σ-algebras $\mathfrak{A}_i = \mathfrak{P}(\Omega_i)$ we get

$$\mu_1 \times \mu_2(A) = \sum_{\omega_1 \in \Omega_1} \mu_2(A(\omega_1)) \cdot \mu_1(\{\omega_1\}), \qquad A \subset \Omega.$$

In particular, for uniform distributions μ_i on finite spaces, $\mu_1 \times \mu_2$ is the uniform distribution on Ω. Cf. Example 3.1 in the case $n = 2$.

(iii) The multi-dimensional Lebesgue measure is a product measure. Namely, for $k, \ell \in \mathbb{N}$ and $A_1 \in \mathfrak{I}_k$, $A_2 \in \mathfrak{I}_\ell$ we have

$$\lambda_{k+\ell}(A_1 \times A_2) = \lambda_k(A_1) \cdot \lambda_\ell(A_2) = \lambda_k \times \lambda_\ell(A_1 \times A_2),$$

see Example 4.1.(i). Corollary 4.1 yields

$$\lambda_{k+\ell} = \lambda_k \times \lambda_\ell.$$

From (9) we get

$$\lambda_{k+\ell}(A) = \int_{\mathbb{R}^k} \lambda_\ell(A(\omega_1))\, \lambda_k(d\omega_1), \qquad A \in \mathfrak{B}_{k+\ell},$$

cf. *Cavalieri's Principle.*

Theorem 2 (Fubini's Theorem)**.**

(i) For $f \in \overline{\mathfrak{Z}}_+(\Omega, \mathfrak{A})$

$$\int_\Omega f \, d(\mu \times K) = \int_{\Omega_1} \int_{\Omega_2} f(\omega_1, \omega_2)\, K(\omega_1, d\omega_2)\, \mu(d\omega_1).$$

(ii) For f $(\mu \times K)$-integrable and

$$A_1 = \{\omega_1 \in \Omega_1 : f(\omega_1, \cdot) \;\; K(\omega_1, \cdot)\text{-integrable}\}$$

we have

(a) $A_1 \in \mathfrak{A}_1$ and $\mu(A_1^c) = 0$,

(b) $A_1 \to \mathbb{R} : \omega_1 \mapsto \int_{\Omega_2} f(\omega_1, \cdot)\, dK(\omega_1, \cdot)$ is integrable w.r.t. $\mu|_{A_1 \cap \mathfrak{A}_1}$,

(c)
$$\int_\Omega f \, d(\mu \times K) = \int_{A_1} \int_{\Omega_2} f(\omega_1, \omega_2)\, K(\omega_1, d\omega_2)\, \mu|_{A_1 \cap \mathfrak{A}_1}(d\omega_1).$$

Proof. Ad (i): algebraic induction. Ad (ii): consider f^+ and f^- and use (i). □

Remark 3. For brevity, we write

$$\int_{\Omega_1} \int_{\Omega_2} f(\omega_1, \omega_2)\, K(\omega_1, d\omega_2)\, \mu(d\omega_1) = \int_{A_1} \int_{\Omega_2} f(\omega_1, \omega_2)\, K(\omega_1, d\omega_2)\, \mu|_{A_1 \cap \mathfrak{A}_1}(d\omega_1),$$

if f is $(\mu \times K)$-integrable. For $f \in \overline{\mathfrak{Z}}(\Omega, \mathfrak{A})$

$$f \text{ is } (\mu \times K)\text{-integrable} \qquad \Leftrightarrow \qquad \int_{\Omega_1} \int_{\Omega_2} |f|(\omega_1, \omega_2)\, K(\omega_1, d\omega_2)\, \mu(d\omega_1) < \infty.$$

Corollary 1 (Fubini's Theorem)**.** For σ-finite measures μ_i on $\mathfrak{A}_i$ and a $(\mu_1 \times \mu_2)$-integrable function f

$$\int_\Omega f\, d(\mu_1 \times \mu_2) = \int_{\Omega_1} \int_{\Omega_2} f(\omega_1, \omega_2)\, \mu_2(d\omega_2)\, \mu_1(d\omega_1)$$
$$= \int_{\Omega_2} \int_{\Omega_1} f(\omega_1, \omega_2)\, \mu_1(d\omega_1)\, \mu_2(d\omega_2).$$

Proof. Theorem 2 yields the first equality. For the second equality, put $\widetilde{f}(\omega_2, \omega_1) = f(\omega_1, \omega_2)$ and note that $\int_\Omega f\, d(\mu_1 \times \mu_2) = \int_\Omega \widetilde{f}\, d(\mu_2 \times \mu_1)$. □

Corollary 2. For every measurable space $(\Omega, \mathfrak{A})$, every σ-finite measure μ on $\mathfrak{A}$, and every $f \in \overline{\mathfrak{Z}}_+(\Omega, \mathfrak{A})$

$$\int_\Omega f\, d\mu = \int_{]0,\infty[} \mu(\{f > x\})\, \lambda_1(dx).$$

Proof. Übung 7.2. □

Now we construct a stochastic model for a series of experiments, where the outputs of the first $i-1$ stages determine the model for the ith stage.

Given: measurable spaces $(\Omega_i, \mathfrak{A}_i)$ for $i \in I$, where $I = \{1, \dots, n\}$ or $I = \mathbb{N}$. Put

$$\left(\Omega'_i, \mathfrak{A}'_i\right) = \left(\bigtimes_{j=1}^{i} \Omega_j, \bigotimes_{j=1}^{i} \mathfrak{A}_j \right),$$

and note that

$$\bigtimes_{j=1}^{i} \Omega_j = \Omega'_{i-1} \times \Omega_i \qquad \wedge \qquad \bigotimes_{j=1}^{i} \mathfrak{A}_j = \mathfrak{A}'_{i-1} \otimes \mathfrak{A}_i$$

for $i \in I \setminus \{1\}$. Furthermore, let

$$\Omega = \bigtimes_{i \in I} \Omega_i, \qquad \mathfrak{A} = \bigotimes_{i \in I} \mathfrak{A}_i. \tag{10}$$

Given:

- σ-finite kernels K_i from $\left(\Omega'_{i-1}, \mathfrak{A}'_{i-1}\right)$ to $(\Omega_i, \mathfrak{A}_i)$ for $i \in I \setminus \{1\}$,
- a σ-finite measure μ on $\mathfrak{A}_1$.

Theorem 3. For $I = \{1, \dots, n\}$

$$\underset{1}{\exists} \text{ measure } \nu \text{ on } \mathfrak{A} \ \ \forall A_1 \in \mathfrak{A}_1 \dots \forall A_n \in \mathfrak{A}_n :$$
$$\nu(A_1 \times \cdots \times A_n)$$
$$= \int_{A_1} \cdots \int_{A_{n-1}} K_n((\omega_1, \dots, \omega_{n-1}), A_n)\, K_{n-1}((\omega_1, \dots, \omega_{n-2}), d\omega_{n-1}) \cdots \mu(d\omega_1).$$

Moreover, ν is σ-finite and for f ν-integrable (the short version)

$$\int_\Omega f\, d\nu = \int_{\Omega_1} \cdots \int_{\Omega_n} f(\omega_1, \dots, \omega_n) K_n((\omega_1, \dots, \omega_{n-1}), d\omega_n) \cdots \mu(d\omega_1). \tag{11}$$

Notation: $\nu = \mu \times K_2 \times \cdots \times K_n$.

Proof. Induction, using Theorems 1 and 2. □

Remark 4. Particular case of Theorem 3 with

$$\mu = \mu_1, \qquad \forall\, i \in I \setminus \{1\}\ \forall\, \omega'_{i-1} \in \Omega'_{i-1} :\ K_i(\omega'_{i-1}, \cdot) = \mu_i \tag{12}$$

for σ-finite measures μ_i on $\mathfrak{A}_i$:

$$\underset{1}{\exists}\,\text{measure } \mu_1 \times \cdots \times \mu_n \text{ on } \mathfrak{A}\ \ \forall\, A_1 \in \mathfrak{A}_1 \ldots \forall\, A_n \in \mathfrak{A}_n :$$
$$\mu_1 \times \cdots \times \mu_n(A_1 \times \cdots \times A_n) = \mu_1(A_1) \cdot \cdots \cdot \mu_n(A_n).$$

Moreover, $\mu_1 \times \cdots \times \mu_n$ is σ-finite and for every $\mu_1 \times \cdots \times \mu_n$-integrable function f

$$\int_\Omega f\, d(\mu_1 \times \cdots \times \mu_n) = \int_{\Omega_1} \ldots \int_{\Omega_n} f(\omega_1, \ldots, \omega_n)\ \mu_n(d\omega_n) \cdots \mu_1(d\omega_1).$$

Definition 3. $\mu = \mu_1 \times \cdots \times \mu_n$ is called the *product measure* corresponding to μ_i for $i = 1, \ldots, n$, and $(\Omega, \mathfrak{A}, \mu)$ is called the *product measure space* corresponding to $(\Omega_i, \mathfrak{A}_i, \mu_i)$ for $i = 1, \ldots, n$.

Example 6.

(i) For uniform distributions μ_i on finite spaces Ω_i, $\mu_1 \times \cdots \times \mu_n$ is the uniform distribution on Ω. Cf. Example 3.1 in the case $n \in \mathbb{N}$.

(ii)
$$\lambda_n = \lambda_1 \times \cdots \times \lambda_1.$$

Theorem 4 (Ionescu-Tulcea)**.** Assume that μ is a probability measure and that K_i are Markov kernels for $i \in \mathbb{N} \setminus \{1\}$. Then, for $I = \mathbb{N}$,

$$\underset{1}{\exists}\,\text{probability measure } P \text{ on } \mathfrak{A}\ \ \forall\, n \in \mathbb{N}\ \ \forall\, A_1 \in \mathfrak{A}_1 \ldots \forall\, A_n \in \mathfrak{A}_n :$$
$$P\Big(A_1 \times \cdots \times A_n \times \bigtimes_{i=n+1}^{\infty} \Omega_i\Big) = (\mu \times K_2 \times \cdots \times K_n)(A_1 \times \cdots \times A_n). \tag{13}$$

Proof. ‘Existence’: Consider σ-algebras

$$\bigotimes_{i=1}^{n} \mathfrak{A}_i, \qquad \widetilde{\mathfrak{A}}_n = \sigma\left(\pi^{\mathbb{N}}_{\{1,\ldots,n\}}\right)$$

on $\bigtimes_{i=1}^{n} \Omega_i$ and $\bigtimes_{i=1}^{\infty} \Omega_i$, respectively. Define a probability measure $\widetilde{P}_n$ on $\widetilde{\mathfrak{A}}_n$ by

$$\widetilde{P}_n\Big(A \times \bigtimes_{i=n+1}^{\infty} \Omega_i\Big) = (\mu \times K_2 \times \cdots \times K_n)(A), \qquad A \in \bigotimes_{i=1}^{n} \mathfrak{A}_i.$$

Then (11) yields the following *consistency property*

$$\widetilde{P}_{n+1}\Big(A \times \Omega_{n+1} \times \bigtimes_{i=n+2}^{\infty} \Omega_i\Big) = \widetilde{P}_n\Big(A \times \bigtimes_{i=n+1}^{\infty} \Omega_i\Big), \qquad A \in \bigotimes_{i=1}^{n} \mathfrak{A}_i.$$

Thus

$$\widetilde{P}(\widetilde{A}) = \widetilde{P}_n(\widetilde{A}), \qquad \widetilde{A} \in \widetilde{\mathfrak{A}}_n,$$

yields a well-defined mapping on the algebra

$$\widetilde{\mathfrak{A}} = \bigcup_{n\in\mathbb{N}} \widetilde{\mathfrak{A}}_n$$

of cylinder sets. Obviously, $\widetilde{P}$ is a content and (13) holds for $P = \widetilde{P}$.
Claim: $\widetilde{P}$ is σ-continuous at $\emptyset$. See Gänssler, Stute (1977, p. 49–50) for the proof.
Then, by Theorem 4.1, $\widetilde{P}$ is σ-additive and it remains to apply Theorem 4.3.
'Uniqueness': By (13), P is uniquely determined on the class of measurable rectangles. Apply Theorem 4.4. □

Example 7. The queueing model, see Übung 7.3. Here $K_i((\omega_1, \dots, \omega_{i-1}), \cdot)$ only depends on ω_{i-1}. Outlook: Markov processes.

Given: a non-empty set I and probability spaces $(\Omega_i, \mathfrak{A}_i, \mu_i)$ for $i \in I$. Recall the definition (10).

Theorem 5.

$$\underset{1}{\exists}\text{probability measure } P \text{ on } \mathfrak{A} \ \ \forall\, S \in \mathfrak{P}_0(I) \ \ \forall\, A_i \in \mathfrak{A}_i,\ i \in S:$$

$$P\Big(\bigtimes_{i\in S} A_i \times \bigtimes_{i\in I\setminus S} \Omega_i\Big) = \prod_{i\in S} \mu_i(A_i). \tag{14}$$

Notation: $P = \bigtimes_{i\in I} \mu_i$.

Proof. See Remark 4 in the case of a finite set I.
If $|I| = |\mathbb{N}|$, assume $I = \mathbb{N}$ without loss of generality. The particular case of Theorem 4 with (12) for probability measures μ_i on $\mathfrak{A}_i$ shows

$$\underset{1}{\exists}\text{probability measure } P \text{ on } \mathfrak{A} \ \ \forall\, n \in \mathbb{N} \ \ \forall\, A_1 \in \mathfrak{A}_1 \dots \forall\, A_n \in \mathfrak{A}_n:$$

$$P\Big(A_1 \times \dots \times A_n \times \bigtimes_{i=n+1}^{\infty} \Omega_i\Big) = \mu_1(A_1) \cdot \dots \cdot \mu_n(A_n).$$

If I is uncountable, we use Theorem 3.2. For $S \subset I$ non-empty and countable and for $B \in \bigotimes_{i\in S} \mathfrak{A}_i$ we put

$$P\big(\big(\pi^I_S\big)^{-1} B\big) = \bigtimes_{i\in S} \mu_i(B).$$

Hereby we get a well-defined mapping $P : \mathfrak{A} \to \mathbb{R}$, which clearly is a probability measure and satisfies (14). Use Theorem 4.4 to obtain the uniqueness result. □

Definition 4. $P = \bigtimes_{i\in I} \mu_i$ is called the *product measure* corresponding to μ_i for $i \in I$, and $(\Omega, \mathfrak{A}, P)$ is called the *product measure space* corresponding to $(\Omega_i, \mathfrak{A}_i, \mu_i)$ for $i \in I$.

Remark 5. Theorem 5 answers the question that is posed in Example 3.1 in full generality. Moreover, it is the basis for a positive answer to the question from the introductory Example I.2, see Theorem **??**.**??**.**??**.

9 Image Measures

Given: a measure space $(\Omega, \mathfrak{A}, \mu)$, a measurable space $(\Omega', \mathfrak{A}')$, and an $\mathfrak{A}$-$\mathfrak{A}'$-measurable mapping $f : \Omega \to \Omega'$.

Lemma 1.

$$\begin{aligned} f(\mu) : \mathfrak{A}' &\to \mathbb{R}_+ \cup \{\infty\} \\ A' &\mapsto \mu(f^{-1}(A')) = \mu(\{f \in A'\}) \end{aligned}$$

defines a measure on $\mathfrak{A}'$.

Proof. $f(\mu)$ is well-defined, since $f^{-1}(A') \in \mathfrak{A}$ for any $A' \in \mathfrak{A}'$. The respective properties of $f(\mu)$ are easy to verify. □

Definition 1. $f(\mu)$ is called the *image measure* of μ under f.

Example 1. Let

$$(\Omega, \mathfrak{A}, \mu) = (\mathbb{R}^k, \mathfrak{B}_k, \lambda_k), \qquad (\Omega', \mathfrak{A}') = (\mathbb{R}^k, \mathfrak{B}_k).$$

(i) Fix $a \in \mathbb{R}^k$. For $f(\omega) = \omega + a$ we get

$$f(\lambda_k)(A') = \lambda_k(A' - a) = \lambda_k(A'),$$

see Analysis IV ('or' verify this identity for measurable rectangles and apply Theorem 4.4). Thus

$$f(\lambda_k) = \lambda_k.$$

(ii) Fix $r \in \mathbb{R} \setminus \{0\}$. For $f(\omega) = r \cdot \omega$ we get

$$f(\lambda_k)(A') = \lambda_k(1/r \cdot A') = \frac{1}{|r|^k} \cdot \lambda_k(A'),$$

see Analysis IV ('or' verify this identity for measurable rectangles and apply Theorem 4.4). Thus

$$f(\lambda_k) = \frac{1}{|r|^k} \cdot \lambda_k.$$

Theorem 1 (Transformation 'Theorem')**.**

(i) for $g \in \overline{\mathfrak{Z}}_+(\Omega', \mathfrak{A}')$

$$\int_{\Omega'} g \, df(\mu) = \int_{\Omega} g \circ f \, d\mu \tag{1}$$

(ii) for $g \in \overline{\mathfrak{Z}}(\Omega', \mathfrak{A}')$

$$g \text{ is } f(\mu)\text{-integrable} \quad \Leftrightarrow \quad g \circ f \text{ is } \mu\text{-integrable},$$

in which case (1) holds.

Proof. Algebraic induction. □

Example 2. Consider open sets $U, V \subset \mathbb{R}^k$ and a $\mathfrak{C}^1$-diffeomorphism $f : U \to V$. Let

$$(\Omega, \mathfrak{A}, \mu) = (U, U \cap \mathfrak{B}_k, \lambda_k|_{U\cap\mathfrak{B}_k}), \qquad (\Omega', \mathfrak{A}') = (V, V \cap \mathfrak{B}_k).$$

Put

$$\nu = \lambda_k|_{V\cap\mathfrak{B}_k}.$$

Then

$$f(\mu)(A') = \int_{f^{-1}(A')} d\mu = \int_{A'} |\det Df^{-1}|\, d\nu,$$

see Analysis IV for the case of an open set $A' \subset V$. Thus

$$f(\mu) = |\det Df^{-1}| \cdot \nu,$$

and therefore

$$\int_U g \circ f\, d\mu = \int_V g\, df(\mu) = \int_V g \cdot |\det Df^{-1}|\, d\nu.$$

Put $g = h \circ f^{-1}$ and $\varphi = f^{-1}$ to obtain

$$\int_U h\, d\mu = \int_V h \circ \varphi \cdot |\det D\varphi|\, d\nu.$$

Literature

H. Bauer, *Probability Theory*, de Gruyter, Berlin, 1996.

P. Billingsley, *Probability and Measure*, Wiley, New York, first edition 1979, third edition 1995.

Y. S. Chow, H. Teicher, *Probability Theory*, Springer, New York, first editon 1978, third edition 1997.

R. M. Dudley, *Real Analysis and Probability*, Cambridge University Press, Cambridge, 2002.

J. Elstrodt, *Maß- und Integrationstheorie*, Springer, Berlin, first edition 1996, fifth edition, 2007.

K. Floret, *Maß- und Integrationstheorie*, Teubner, Stuttgart, 1981.

P. Gänssler, W. Stute, *Wahrscheinlichkeitstheorie*, Springer, Berlin, 1977.

E. Hewitt, K. Stromberg, *Real and Abstract Analysis*, Springer, Berlin, 1965.

A. Irle, *Finanzmathematik*, Teubner, Stuttgart, 1998.

A. Klenke, *Wahrscheinlichkeitstheorie*, Springer, Berlin, first edition 2006, second edition 2008.

K. R. Parthasarathy, *Probability Measures on Metric Spaces*, Academic Press, New York, 1967.

A. N. Širjaev, *Wahrscheinlichkeit*, Deutscher Verlag der Wissenschaften, Berlin, 1988.
A. N. Shiryayev, *Probability*, Springer, New York, 1984.

J. Yeh, *Martingales and Stochastic Analysis*, World Scientific, Singapore, 1995.

Index