# Probability Theory

Wilhelm Stannat

Technische Universität Darmstadt

Winter Term 2007/08

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.

# 3 Conditional probabilities

## 1 Elementary definitions

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

**Definition 1.1.** Let $B \in \mathcal{A}$ with $P(B) > 0$. Then

$$P[A \,|\, B] := \frac{P(A \cap B)}{P(B)}, \quad A \in \mathcal{A},$$

is said to be the *conditional probability of $A$ given $B$*. In the case $P(B) = 0$ we simply define $P[A \,|\, B] := 0$. The probability measure

$$P_B := P[\,\cdot \,|\, B]$$

on $(\Omega, \mathcal{A})$ is said to be the *conditional distribution given $B$*.

**Remark 1.2.** *(i) $P(A)$ is called the a priori probability of $A$.*

*$P[A \,|\, B]$ is called the a posteriori probability of $A$, given the information that $B$ occurred.*

*(ii) In the case of Laplace experiments*

$$P[A \,|\, B] = \frac{|A \cap B|}{|B|} = \text{fraction of all outcomes in } A \text{ that are contained in } B.$$

*(iii) If $A$ and $B$ are disjoint (hence $A \cap B = \emptyset$), then $P[A \,|\, B] = 0$.*

*(iv) If $A$ and $B$ are independent, then*

$$P[A \,|\, B] = \frac{P(A) \cdot P(B)}{P(B)} = P(A).$$

**Example 1.3.** (i) Suppose that a family has two children. Consider the following two events: $B :=$ "at least one boy" and $A :=$ "two boys". Then $P[A \,|\, B] = \frac{1}{3}$, because

$$\Omega = \{(J, J), (M, J), (J, M), (M, M)\},$$
$$P = \text{uniform distribution},$$

and thus

$$P[A \,|\, B] = \frac{|A \cap B|}{|B|} = \frac{1}{3}.$$

(ii) Let $X_1, X_2$ be independent r.v. with Poisson distribution with parameters $\lambda_1, \lambda_2$. Then

$$P[X_1 = k \,|\, X_1 + X_2 = n] = \begin{cases} 0 & \text{if } k > n \\ ? & \text{if } 0 \leqslant k \leqslant n. \end{cases}$$

According to Example 4.7 $X_1 + X_2$ has Poisson distribution with parameter $\lambda := \lambda_1 + \lambda_2$. Consequently,

$$P[X_1 = k \,|\, X_1 + X_2 = n] = \frac{P[X_1 = k,\ X_2 = n-k]}{P[X_1 + X_2 = n]}$$

$$= \frac{e^{-\lambda_1} \frac{\lambda_1^k}{k!} \cdot e^{-\lambda_2} \frac{\lambda_2^{n-k}}{(n-k)!}}{e^{-\lambda} \frac{\lambda^n}{n!}} = \binom{n}{k} \cdot \left(\frac{\lambda_1}{\lambda}\right)^k \left(\frac{\lambda_2}{\lambda}\right)^{n-k},$$

i.e., $P[\,\cdot\, \mid X_1 + X_2 = n]$ is the binomial distribution with parameters $n$ and $p = \frac{\lambda_1}{\lambda_1 + \lambda_2}$.

(iii) Consider $n$ independent $0$-$1$-experiments $X_1, \dots, X_n$ with success probability $p \in ]0,1[$. Let

$$S_n := X_1 + \ldots + X_n$$

and

$$\begin{array}{rrcl} X_i : & \Omega := \{0,1\}^n & \to & \{0,1\}, \\ & (x_1, \dots, x_n) & \mapsto & x_i. \end{array}$$

For given $(x_1, \dots, x_n) \in \{0,1\}^n$ and fixed $k \in \{0, \dots, n\}$

$$P[X_1 = x_1, \dots, X_n = x_n \,|\, S_n = k]$$

$$= \begin{cases} 0 & \text{if } \sum_i x_i \neq k \\ \frac{p^k (1-p)^{n-k}}{\binom{n}{k} p^k (1-p)^{n-k}} = \binom{n}{k}^{-1} & \text{otherwise} \end{cases}$$

It follows that the conditional distribution $P[\,\cdot \mid S_n = k]$ is the uniform distribution on

$$\Omega_k := \left\{ (x_1, \dots, x_n) \,\Big|\, \sum_{i=1}^n x_i = k \right\}.$$

**Proposition 1.4.** *(***Formula for total probability***) Let $B_1, \dots, B_n$ be disjoint, $B_i \in \mathcal{A}$ $\forall\, 1 \leq i \leq n$. Then for all $A \subset \bigcup_{i=1}^n B_i$, $A \in \mathcal{A}$:*

$$P(A) = \sum_{i=1}^n P[A \,|\, B_i] \cdot P(B_i).$$

*Proof.* Clearly, $A = \dot\cup_{i \leqslant n}(A \cap B_i)$. Consequently,

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i) = \sum_{i=1}^{n} P(A \cap B_i) = \sum_{i=1}^{n} P(A \,|\, B_i) P(B_i)\,.$$

□

**Example 1.5.** (Simpson's paradox)

Consider applications of male ($M$) and female ($W$) students at a university in the United States

| | Applications | accepted | |
|---|---|---|---|
| $M$ | 2084 | 1036 | $P[A \,\vert\, M] \approx 0.49$ |
| $W$ | 1067 | 349 | $P[A \,\vert\, W] \approx 0.33$ |

Is this an example for discrimination of female students? A closer look to the biggest four faculties $B_1, \dots, B_4$:

| | male | | | female | | |
|---|---|---|---|---|---|---|
| | Appl. | acc. | $P_M[A \,\vert\, B_i]$ | Appl. | acc. | $P_W[A \,\vert\, B_i]$ |
| $B_1$ | 826 | 551 | 0.67 | 108 | 89 | 0.82 |
| $B_2$ | 560 | 353 | 0.63 | 25 | 17 | 0.68 |
| $B_3$ | 325 | 110 | 0.34 | 593 | 219 | 0.37 |
| $B_4$ | 373 | 22 | 0.06 | 341 | 24 | 0.07 |
| | 2084 | 1036 | | 1067 | 349 | |

It follows that for all four faculties the probability of being accepted was higher for female students than it was for male students:

$$P_M[A \,|\, B_i] < P_W[A \,|\, B_i].$$

Nevertheless, the preference turns into its opposite if looking at the total probability of admission:

$$\begin{aligned} P_W(A) := P[A \,|\, W] = \sum_{i=1}^{4} P_W[A \,|\, B_i] \cdot P_W(B_i) \\ < P_M(A) := P[A \,|\, M] = \sum_{i=1}^{4} P_M[A \,|\, B_i] \cdot P_M(B_i). \end{aligned}$$

For an explanation consider the distributions of applications:

$$P_M(B_1) = \frac{|B_1 \cap M|}{|M|} = \frac{826}{2084} \approx \frac{4}{10}, \quad P_W(B_1) = \frac{|B_1 \cap W|}{|W|} = \frac{108}{1067} \approx \frac{1}{10},$$

etc. and observe that male students mainly applied at faculties with a high probability of admission, whereas female students mainly applied at faculties with a low probability of admission.

**Proposition 1.6** (Bayes' theorem)**.** *Let $B_1, \dots, B_n \in \mathcal{A}$ be disjoint with $P(B_i) > 0$ for $i = 1, \dots, n$. Let $A \in \mathcal{A}$, $A \subset \bigcup_{i=1}^n B_i$ with $P(A) > 0$. Then:*

$$P[B_i \,|\, A] = \frac{P[A \,|\, B_i] \cdot P(B_i)}{\displaystyle\sum_{j=1}^n P[A \,|\, B_j] \cdot P(B_j)}.$$

*Proof.*

$$P[B_i \,|\, A] = \frac{P(A \cap B_i)}{P(A)} \overset{1.4}{=} \frac{P[A \mid B_i] \cdot P(B_i)}{\displaystyle\sum_{j=1}^n P[A \mid B_j] \cdot P(B_j)}. \qquad \square$$

**Example 1.7** (A posteriori probabilities in medical tests)**.** Suppose that one out of $145$ persons of the same age have the disease $K$, i.e. the a priori probability of having $K$ is $P[\mathsf{K}] = \frac{1}{145}$.

Suppose now that a medical test for $K$ is given which detects $K$ in 96 % of all cases, i.e.

$$P[\texttt{positive} \,|\, K] = 0.96\,.$$

However, the test also is positive in 6% af the cases, where the person does not have $K$, i.e.

$$P[\texttt{positive} \,|\, K^c] = 0.06\,.$$

Suppose now that the test is positive. What is the a posteriori probability of actually having $K$?

So we are interested in the conditional probability $P[K \,|\, \texttt{positive}]$:

$$\begin{aligned} P[K \,|\, \texttt{positive}] &\overset{1.6}{=} \frac{P[\texttt{positive} \,|\, K] \cdot P[K]}{P[\texttt{positive} \,|\, K] \cdot P[K] + P[\texttt{positive} \,|\, K^c] \cdot P[K^c]} \\ &= \frac{0.96 \cdot \frac{1}{145}}{0.96 \cdot \frac{1}{145} + 0.06 \cdot \frac{144}{145}} = \frac{1}{1 + \frac{6}{96} \cdot 144} = \frac{1}{10}. \end{aligned}$$

Note: in only one out of ten cases, a person with a positive result actually has $K$.

Another conditional probability of interest in this context is the probability of not having $K$, once the test is negative, i.e., $P[K^c \,|\, \texttt{negative}]$:

$$\begin{aligned} P[K^c \,|\, \texttt{negative}] &= \frac{P[\texttt{negative} \,|\, K^c] \cdot P[K^c]}{P[\texttt{negative} \,|\, K] \cdot P[K] + P[\texttt{negative} \,|\, K^c] \cdot P[K^c]} \\ &= \frac{0.94 \cdot \frac{144}{145}}{0.04 \cdot \frac{1}{145} + 0.94 \cdot \frac{144}{145}} = \frac{94 \cdot 144}{4 + 94 \cdot 144} \approx 0.9997. \end{aligned}$$

*Note:* The two conditional probabilities interchange, if the a priori probability of not having $K$ is low (e.g. $\frac{1}{145}$). If the risc of having $K$ is high and one wants to test whether or not having $K$, the a posteriori probability of not having $K$, given that the test was negative, is only $0.1$.

**Example 1.8** (computing total probabilities with conditional probabilities)**.** Let $S$ be a finite set, $\Omega := S^{n+1}$, $n \in \mathbb{N}$, and $P$ be a probability measure on $\Omega$. Let $X_i : \Omega \to S$, $i = 0, \ldots, n$, be the canonical projections $X_i(\omega) := x_i$ for $\omega = (x_0, \ldots, x_n)$.

If we interpret $0, 1, \ldots, n$ as time points, then $(X_i)_{0 \leqslant i \leqslant n}$ may be seen as a *stochastic process* and $\big(X_0(\omega), \ldots, X_n(\omega)\big)$ is said to be a *sample path* (or a *trajectory*) of the process.

For all $\omega \in \Omega$ we either have $P(\{\omega\}) = 0$ or

$$\begin{aligned}
P(\{\omega\}) &= P[X_0 = x_0, \ldots, X_n = x_n] \\
&= P[X_0 = x_0, \ldots, X_{n-1} = x_{n-1}] \\
&\qquad \cdot P[X_n = x_n \,|\, X_0 = x_0, \ldots, X_{n-1} = x_{n-1}] \\
&\;\vdots \\
&= P[X_0 = x_0] \\
&\qquad \cdot P[X_1 = x_1 \,|\, X_0 = x_0] \\
&\qquad \cdot P[X_2 = x_2 \,|\, X_0 = x_0, X_1 = x_1] \\
&\qquad \cdots \\
&\qquad \cdot P[X_n = x_n \,|\, X_0 = x_0, \ldots, X_{n-1} = x_{n-1}].
\end{aligned}$$

Note: $P(\{\omega\}) \neq 0$ implies $P[X_0 = x_0, \ldots, X_k = x_k] \neq 0$ for all $k \in \{0, \ldots, n\}$.

Conclusion: A probability measure $P$ on $\Omega$ is uniquely determined by the following:

***Initial distribution:*** $\mu := P \circ X_0^{-1}$

***Transition probabilities:*** the conditional distributions

$$P[X_k = x_k \,|\, X_0 = x_0, \ldots, X_{k-1} = x_{k-1}]$$

for any $k \in \{1, \ldots, n\}$ and $(x_0, \ldots, x_k) \in S^{(k+1)}$.

*Existence* of $P$ for given initial distribution and given transition probabilities is shown in Section 3.3.

**Example 1.9.** A stochastic process is called a *Markov chain*, if $P[X_k = x_k \,|\, X_0 = x_0, \ldots, X_{k-1} = x_{k-1}] = P[X_k = x_k \,|\, X_{k-1} = x_{k-1}]$, i.e., if the transition probabilities for $X_k$ only depend on $X_{k-1}$.

If we denote by $X_{k-1}$ the "present", by $X_k$ the "future" and by "$X_0, \ldots, X_{k-2}$" the past, then we can state the Markov property as: given the "present", the "future" of the Markov chain is independent of the "past".

# 2 Transition probabilities and Fubini's theorem

Let $(S_1, \mathcal{S}_1)$ and $(S_2, \mathcal{S}_2)$ be measurable spaces.

**Definition 2.1.** A mapping

$$\begin{array}{rccl} K: & S_1 \times \mathcal{S}_2 & \rightarrow & [0,1] \\ & (x_1, A_2) & \mapsto & K(x_1, A_2) \end{array}$$

is said to be a *transition probabilities* (from $(S_1, \mathcal{S}_1)$ to $(S_2, \mathcal{S}_2)$), if

(i) $\forall x_1 \in S_1$: $K(x_1, \cdot\,)$ is a probability measure on $(S_2, \mathcal{S}_2)$.

(ii) $\forall A_2 \in \mathcal{S}_2$: $K(\,\cdot\,, A_2)$ is $\mathcal{S}_1$-measurable.

**Example 2.2.** (i) For given probability measure $\mu$ on $(S_2, \mathcal{S}_2)$ define

$$K(x_1, \cdot\,) := \mu \qquad \forall\, x_1 \in S_1 \ \text{ no coupling!}$$

(ii) Let $T : S_1 \to S_2$ be a $\mathcal{S}_1/\mathcal{S}_2$-measurable mapping, and

$$K(x_1, \cdot\,) := \delta_{T(x_1)} \qquad \forall x_1 \in S_1\,.$$

(iii) **Stochastic matrices** Let $S_1, S_2$ be countable and $\mathcal{S}_i = \mathcal{P}(S_i)$, $i = 1, 2$. In this case, any transition probability from $(S_1, \mathcal{S}_1)$ to $(S_2, \mathcal{S}_2)$ is given by

$$K(x_1, x_2) := K\big(x_1, \{x_2\}\big), \quad x_1 \in S_1,\ x_2 \in S_2,$$

where $K : S_1 \times \mathcal{S}_2 \to [0,1]$ is a mapping, such that for all $x_1 \in S_1$ $\sum_{x_2 \in S_2} K(x_1, x_2) = 1$. Consequently, $K$ can be identified with a stochastic matrix, or a transition matrix, i.e. a matrix with nonnegative entries and row sums equal to one.

**Example 2.3.** (i) **Transition probabilities of the random walk on $\mathbb{Z}^d$**

$S_1 = S_2 = S := \mathbb{Z}^d$ with $\mathcal{S} := \mathcal{P}(\mathbb{Z}^d)$

$$K(x, \cdot\,) := \frac{1}{2d} \sum_{y \in N(x)} \delta_y\,, x \in \mathbb{Z}^d\,,$$

with

$$N(x) := \big\{ y \in \mathbb{Z}^d \ \big|\ \|x - y\| = 1 \big\}$$

denotes the set of nearest neighbours of $x$.

(ii) **Ehrenfest model** Consider a box containing $N$ balls. The box is divided into two parts ("left" and "right"). A ball is selected randomly and put into the other half.

**"microscopic level"** the state space is $S := \{0,1\}^N$ with $x = (x_1, \dots, x_N) \in S$ defined by

$$x_i := \begin{cases} 1 & \text{if the } i^{th} \text{ ball is contained in the "left" half} \\ 0 & \text{if the } i^{th} \text{ ball is contained in the "right" half} \end{cases}$$

the transition probability is given by

$$K(x, \cdot\,) := \frac{1}{N} \sum_{i=1}^{N} \delta_{(x_1, \dots, x_{i-1}, 1 - x_i, x_{i+1}, \dots, x_N)}.$$

**"macroscopic level"** the state space is $S := \{0, \ldots, N\}$, where $j \in S$ denotes the number of balls contained in the left half. The transition probabilities are given by

$$K(j, \cdot\,) := \frac{N-j}{N} \cdot \delta_{j+1} + \frac{j}{N} \cdot \delta_{j-1}.$$

(iii) **Transition probabilities of the Ornstein-Uhlenbeck process** $S = S_1 = S_2 = \mathbb{R}$, $K(x, \cdot\,) := N(\alpha x, \sigma^2)$ with $\alpha \in \mathbb{R}$, $\sigma^2 > 0$.

We now turn to Fubini's theorem. To this end, let $\mu_1$ be a probability measure on $(S_1, \mathcal{S}_1)$ and $K(\,\cdot\,, \cdot\,)$ be a transition probability from $(S_1, \mathcal{S}_1)$ to $(S_2, \mathcal{S}_2)$.
Our aim is to construct a probability measure $P$ $(:= \mu_1 \otimes K)$ on the product space $(\Omega, \mathcal{A})$, where

$$\Omega := S_1 \times S_2$$

$$\mathcal{A} := \mathcal{S}_1 \otimes \mathcal{S}_2 := \sigma(X_1, X_2) \overset{!}{=} \sigma\big(\{A_1 \times A_2 \,|\, A_1 \in \mathcal{S}_1,\ A_2 \in \mathcal{S}_2\}\big)\,,$$

and

$$\begin{array}{rrcll} X_i : & \Omega = S_1 \times S_2 & \to & S_i, & \quad i = 1, 2, \\ & (x_1, x_2) & \mapsto & x_i, & \end{array}$$

satisfying

$$P(A_1 \times A_2) = \int_{A_1} K(x_1, A_2)\, \mu_1(\mathrm{d}x_1)$$

for all $A_1 \in \mathcal{S}_1$ and $A_2 \in \mathcal{S}_2$.

**Proposition 2.4** (Fubini). *Let $\mu_1$ be a probability measure on $(S_1, \mathcal{S}_1)$, $K$ a transition probability from $(S_1, \mathcal{S}_1)$ to $(S_2, \mathcal{S}_2)$, and*

$$\Omega := S_1 \times S_2, \tag{3.1}$$

$$\mathcal{A} := \sigma\big(\{A_1 \times A_2 \,|\, A_i \in \mathcal{S}_i\}\big) =: \mathcal{S}_1 \otimes \mathcal{S}_2\,. \tag{3.2}$$

*Then there exists a probability measure $P$ ($=: \mu_1 \otimes K$) on $(\Omega, \mathcal{A})$, such that for all $\mathcal{A}$-measurable functions $f \geqslant 0$*

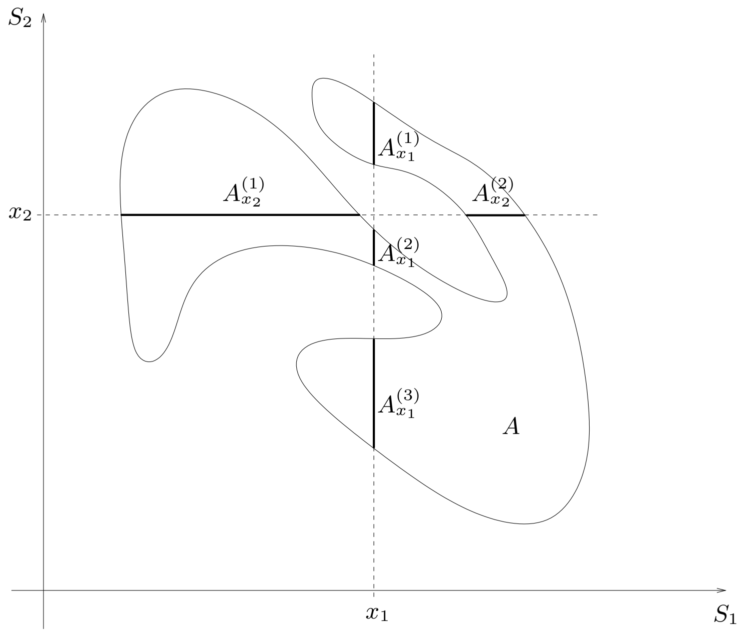
$$\int_\Omega f \,\mathrm{d}P = \int \left( \int f(x_1, x_2)\, K(x_1, \mathrm{d}x_2) \right) \mu_1(\mathrm{d}x_1), \tag{3.3}$$

*in particular, for all $A \in \mathcal{A}$*

$$P(A) = \int K(x_1, A_{x_1})\, \mu_1(\mathrm{d}x_1)\,. \tag{3.4}$$

*Here*

$$A_{x_1} = \{x_2 \in S_2 \,|\, (x_1, x_2) \in A\}$$




Note

$$A_{x_1} = A^{(1)}_{x_1} \cup A^{(2)}_{x_1} \cup A^{(3)}_{x_1} \quad \text{and} \quad A_{x_2} = A^{(1)}_{x_2} \cup A^{(2)}_{x_2}.$$

*is called the* section of $A$ by $x_1$*. In particular, for* $A_1 \in \mathcal{S}_1$*,* $A_2 \in \mathcal{S}_2$*:*

$$P(A_1 \times A_2) = \int_{A_1} K(x_1, A_2)\, \mu_1(\mathrm{d}x_1). \tag{3.5}$$

$P$ *is uniquely determined by* (3.5).

*Proof.* Uniqueness: Clearly, the collection of cylindrical sets $A_1 \times A_2$ with $A_i \in \mathcal{S}_i$ is stable under intersections and generates $\mathcal{A}$, so that the uniqueness now follows from Proposition 1.11.5.

Existence: For given $x_1 \in S_1$ let

$$\varphi_{x_1}(x_2) := (x_1, x_2)\,.$$

$\varphi_{x_1} : S_2 \to \Omega$ is measurable, because for $A_1 \in \mathcal{S}_1$, $A_2 \in \mathcal{S}_2$

$$\varphi_{x_1}^{-1}(A_1 \times A_2) = \begin{cases} \emptyset & \text{if } x_1 \notin A_1 \\ A_2 & \text{if } x_1 \in A_1 \end{cases}.$$

It follows that for any $f : \Omega \to \mathbb{R}$ $\mathcal{A}$-measurable and any $x_1 \in S_1$, the mapping

$$f_{x_1} := f \circ \varphi_{x_1} : S_2 \to \mathbb{R}\,, x_2 \mapsto f(x_1, x_2)$$

is $\mathcal{S}_2/\mathcal{B}(\mathbb{R})$-measurable.

Suppose now that $f \geqslant 0$ or bounded. Then

$$x_1 \mapsto \int f(x_1, x_2)\; K(x_1, \mathrm{d}x_2) \left(= \int f_{x_1}(x_2)\; K(x_1, \mathrm{d}x_2)\right) \tag{3.6}$$

is well-defined.

We will show in the following that this function is $\mathcal{S}_1$-measurable. We will prove the assertion for $f = 1_A$, $A \in \mathcal{A}$ first. For general $f$ the measurability then follows by measure-theoretic induction.

Note that for $f = 1_A$ we have that

$$\int \underbrace{1_A(x_1, x_2)}_{=1_{A_{x_1}}(x_2)}\; K(x_1, \mathrm{d}x_2) = K(x_1, A_{x_1}).$$

Hence, in the following we consider

$$\mathcal{D} := \big\{A \in \mathcal{A} \mid x_1 \mapsto K(x_1, A_{x_1})\ \mathcal{S}_1\text{-measurable}\big\}.$$

$\mathcal{D}$ is a Dynkin system (!) and contains all cylindrical sets $A = A_1 \times A_2$ with $A_i \in \mathcal{S}_i$, because

$$K\big(x_1, (A_1 \times A_2)_{x_1}\big) = 1_{A_1}(x_1) \cdot K(x_1, A_2)\,.$$

Since measurable cylindrical sets are stable under intersections, we conclude that $\mathcal{D} = \mathcal{A}$.

It follows that for all nonnegative or bounded $\mathcal{A}$-measurable functions $f : \Omega \to \mathbb{R}$, the integral

$$\int \left(\int f(x_1, x_2)\; K(x_1, \mathrm{d}x_2)\right) \mu(\mathrm{d}x_1)$$

is well-defined.

For all $A \in \mathcal{A}$ we can now define

$$P(A) := \int \left(\int \underbrace{1_A(x_1, x_2)}_{=1_{A_{x_1}}(x_2)}\; K(x_1, \mathrm{d}x_2)\right) \mu(\mathrm{d}x_1) = \int K(x_1, A_{x_1})\; \mu(\mathrm{d}x_1).$$

$P$ is a probability measure on $(\Omega, \mathcal{A})$, because

$$P(\Omega) = \int K(x_1, S_2)\; \mu(\mathrm{d}x_1) = \int 1\; \mu(\mathrm{d}x_1) = 1.$$

For the proof of the $\sigma$-additivity, let $A_1, A_2, \dots$ be pairwise disjoint subsets in $\mathcal{A}$. It follows that for all $x_1 \in S_1$ the subsets $(A_1)_{x_1}, (A_2)_{x_1}, \dots$ are pairwise disjoint too, hence

$$\begin{aligned} P\Big(\dot{\bigcup_{n\in\mathbb{N}}} A_n\Big) &= \int K\bigg(x_1, \Big(\dot{\bigcup_{n\in\mathbb{N}}} A_n\Big)_{x_1}\bigg)\, \mu(\mathrm{d}x_1) \\ &= \int \sum_{n=1}^{\infty} K\big(x_1, (A_n)_{x_1}\big)\, \mu(\mathrm{d}x_1) \\ &= \sum_{n=1}^{\infty} \int K\big(x_1, (A_n)_{x_1}\big)\, \mu(\mathrm{d}x_1) = \sum_{n=1}^{\infty} P(A_n). \end{aligned}$$

In the second equality we used that $K(x_1, \cdot)$ is a probability measure for all $x_1$ and in the third equality we used monotone integration.

Finally, (3.3) follows from measure-theoretic induction. □

## 2.1 Examples and Applications

**Remark 2.5.** *The classical Fubini theorem is a particular case of Proposition 2.4:* $K(x_1, \cdot\,) = \mu_2$. *In this case, the measure* $\mu_1 \otimes K$*, constructed in Fubini's theorem, is called the* product measure *of* $\mu_1$ *and* $\mu_2$ *and is denoted by* $\mu_1 \otimes \mu_2$. *Moreover, in this case*

$$\int f\, \mathrm{d}P = \int \bigg( \int f(x_1, x_2)\, \mu_2(\mathrm{d}x_2) \bigg)\, \mu_1(\mathrm{d}x_1).$$

**Remark 2.6** (Marginal distributions)**.** *Let* $X_i : \Omega \to S_i$, $i = 1, 2$, *be the natural projections* $X_i\big((x_1, x_2)\big) := x_i$. *The distributions of* $X_i$ *under the measure* $\mu_1 \otimes K$ *are called the* marginal distributions *and they are given by*

$$\begin{aligned} (P \circ X_1^{-1})(A_1) &= P[X_1 \in A_1] = P(A_1 \times S_2) \\ &= \int_{A_1} \underbrace{K(x_1, S_2)}_{=1}\, \mu_1(\mathrm{d}x_1) = \mu_1(A_1) \end{aligned}$$

*and*

$$\begin{aligned} (P \circ X_2^{-1})(A_2) &= P[X_2 \in A_2] = P(S_1 \times A_2) \\ &= \int K(x_1, A_2)\, \mu_1(\mathrm{d}x_1) =: (\mu_1 K)(A_2). \end{aligned}$$

*So, the marginal distributions are*

$$P \circ X_1^{-1} = \mu_1 \qquad P \circ X_2^{-1} = \mu_1 K\,.$$

**Definition 2.7.** Let $S_1 = S_2 = S$ and $\mathcal{S}_1 = \mathcal{S}_2 = \mathcal{S}$. A probability measure $\mu$ on $(S, \mathcal{S})$ is said to be an *equilibrium distribution for* $K$ (or *invariant distribution under* $K$) if $\mu = \mu K$.

**Example 2.8.** (i) **Ehrenfest model** (macroscopic) Let $S = \{0, 1, \dots, N\}$ and

$$K(y, \cdot\,) = \frac{y}{N} \cdot \delta_{y-1} + \frac{N-y}{N} \cdot \delta_{y+1}.$$

In this case, the binomial distribution $\mu$ with parameter $N, \frac{1}{2}$ is an equilibirum distribution, because

$$\begin{aligned}
(\mu K)(\{x\}) &= \sum_{y \in S} \mu(\{y\}) \cdot K(y, x) \\
&= \mu(\{x+1\}) \cdot \frac{x+1}{N} + \mu(\{x-1\}) \cdot \frac{N-(x-1)}{N} \\
&= 2^{-N} \binom{N}{x+1} \cdot \frac{x+1}{N} + 2^{-N} \binom{N}{x-1} \cdot \frac{N-(x-1)}{N} \\
&= 2^{-N} \left[ \binom{N-1}{x} + \binom{N-1}{x-1} \right] = 2^{-N} \cdot \binom{N}{x} = \mu(\{x\}).
\end{aligned}$$

(ii) **Ornstein-Uhlenbeck process** Let $S = \mathbb{R}$ and $K(x, \cdot\,) = N(\alpha x, \sigma^2)$ with $|\alpha| < 1$. Then

$$\mu = N\left(0, \frac{\sigma^2}{1-\alpha^2}\right)$$

is an equilibrium distribution. (Exercise.)

We now turn to the converse problem: Given a probability measure $P$ on the product space $(\Omega, \mathcal{A})$. Can we "disintegrate" $P$, i.e., can we find a probability measure $\mu_1$ on $(S_1, \mathcal{S}_1)$ and a transition probability from $S_1$ to $S_2$ such that

$$P = \mu_1 \otimes K\,?$$

*Answer:* In most cases - yes, e.g. if $S_1$ and $S_2$ are Polish spaces (i.e., a topological space having a countable basis, whose topology is induced by some complete metric), using conditional expectations (see below).

**Example 2.9.** In the particular case, when $S_1$ is countable (and $\mathcal{S}_1 = \mathcal{P}(S_1)$), we can disintegrate $P$ explicitly as follows: Necessarily, $\mu_1$ has to be the distribution of the projection $X_1$ onto the first coordinate. To define the kernel $K$, let $\nu$ be any probability measure on $(S_2, \mathcal{S}_2)$ and define

$$K(x_1, A_2) := \begin{cases} P[X_2 \in A_2 \,|\, X_1 = x_1] & \text{if } \underbrace{\mu_1(\{x_1\})}_{=P[X_1 = x_1]} > 0 \\ \\ \nu(A_2) & \text{if } \mu_1(\{x_1\}) = 0. \end{cases}$$

Then

$$P(A_1 \times A_2) = P[X_1 \in A_1,\ X_2 \in A_2] = \sum_{x_1 \in A_1} P[X_1 = x_1,\ X_2 \in A_2]$$
$$= \sum_{\substack{x_1 \in A_1, \\ \mu_1(\{x_1\})>0}} P[X_1 = x_1] \cdot P[X_2 \in A_2 \,|\, X_1 = x_1]$$
$$= \sum_{x_1 \in A_1} \mu_1(\{x_1\}) \cdot K(x_1, A_2) = \int_{A_1} K(x_1, A_2)\ \mu_1(\mathrm{d}x_1)$$
$$= (\mu_1 \otimes K)(A_1 \times A_2),$$

hence $P = \mu_1 \otimes K$.

In the next proposition we are interested in an explicit formula for the disintegration in the case of absolute continuous probability measures.

Note: If $P$ is a probability measure on $(\Omega, \mathcal{A})$ and $\varphi : \Omega \to \mathbb{R}_+$ is $\mathcal{A}$-measurable with $\int \varphi \, \mathrm{d}P = 1$. Then

$$(\varphi P)(A) := \int_A \varphi \, \mathrm{d}P$$

defines another probability measure on $(\Omega, \mathcal{A})$.

For a given transition probability $K$ from $S_1$ to $S_2$ and a function $\varphi : \Omega \to \mathbb{R}_+$, $\mathcal{A}$-measurable, let

$$K\varphi(x) := \int K(x, \mathrm{d}y)\ \varphi(x, y)\,, x \in S_1\,.$$

**Proposition 2.10.** *Let $P = \mu \otimes K$ and $\tilde{P} := \varphi P$. Then $\tilde{P} = \tilde{\mu} \otimes \tilde{K}$ with*

$$\tilde{\mu} = (K\varphi)\mu \quad \textit{und} \quad \tilde{K}(x, \mathrm{d}y) := \frac{\varphi(x, y)}{K\varphi(x)} \cdot K(x, \mathrm{d}y)$$

*for all $x \in S_1$ with $K\varphi(x) > 0$ (and $\tilde{K}(x, \cdot\,) = \nu$, $\nu$ any probability measure on $(S_2, \mathcal{S}_2)$ if $x \in S_2$ is such that $K\varphi(x) = 0$).*

*Proof.* (i) Let $\tilde{\mu}$ be the distribution of $X_1$ under $\tilde{P}$. Then for all $A \in \mathcal{S}_1$

$$\tilde{\mu}(A) = \tilde{P}(A \times S_2) = \int_{A \times S_2} \varphi(x, y)\ \mathrm{d}P$$
$$= \int 1_A(x) \Big( \int \varphi(x, y)\ K(x, \mathrm{d}y) \Big)\ \mu(\mathrm{d}x) = \int_A (K\varphi)(x)\ \mu(\mathrm{d}x),$$

hence $\tilde{\mu} = (K\varphi)\mu$. In particular, $\tilde{\mu}$-a.s. $K\varphi > 0$, because

$$\tilde{\mu}(K\varphi = 0) = \int_{\{K\varphi = 0\}} (K\varphi)(x)\,\mu(dx) = 0\,.$$

(ii) Let $\tilde{K}$ be as above. Clearly, $\tilde{K}$ is a transition probability, because

$$\int \varphi(x,y)\ K(x,\mathrm{d}y) = K\varphi(x)\,,\ \text{so that } \tilde{K}(x,S_2) = 1 \qquad \forall x \in S_1\,.$$

For all $A \in \mathcal{S}_1$ and $B \in \mathcal{S}_2$ we then have

$$\begin{aligned}
\tilde{P}(A\times B) &= \int \Big(\underbrace{\int 1_{A\times B}(x,y)\cdot\varphi(x,y)\ K(x,\mathrm{d}y)}_{\leqslant K\varphi(x)}\Big)\ \mu(\mathrm{d}x) \\
&= \int_{\{K\varphi>0\}} \Big(\int 1_{A\times B}(x,y)\cdot\varphi(x,y)\ K(x,\mathrm{d}y)\Big)\ \mu(\mathrm{d}x) \\
&= \int_A K\varphi(x)\ \tilde{K}(x,B)\ \mu(\mathrm{d}x) = \int_A \tilde{K}(x,B)\ \tilde{\mu}(\mathrm{d}x) \\
&= (\tilde{\mu}\otimes\tilde{K})(A\times B).
\end{aligned}$$

□

# 3 The canonical model for the evolution of a stochastic system in discrete time

Consider the following situation: suppose we are given

- measurable spaces $(S_i, \mathcal{S}_1)$, $i = 0, 1, 2, \ldots$ and we define

$$\begin{aligned}
S^n &:= S_0 \times S_1 \times \cdots \times S_n \\
\mathcal{S}^n &:= \mathcal{S}_0 \otimes \mathcal{S}_1 \otimes \cdots \otimes \mathcal{S}_n = \sigma\big(\{A_0 \times \cdots \times A_n \mid A_i \in \mathcal{S}_i\}\big).
\end{aligned}$$

- – an initial distribution $\mu_0$ on $(S_0, \mathcal{S}_0)$
  – transition probabilities

  $$K_n\big((x_0, \ldots, x_{n-1}), \mathrm{d}x_n\big)$$

  from $(S^{n-1}, \mathcal{S}^{n-1})$ to $(S_n, \mathcal{S}_n)$, $n = 1, 2, \ldots$.

Using Fubini's theorem, we can then define probability measures $P^n$ on $S^n$, $n = 0, 1, 2, \ldots$ as follows:

$$\begin{aligned}
P^0 &:= \mu_0 && \text{on } S_0, \\
P^n &:= P^{n-1} \otimes K_n && \text{on } S^n = S^{n-1} \times S_n
\end{aligned}$$

Note that Fubini's theorem (see Proposition 2.4) implies that for any $\mathcal{S}^n$-measurable function $f : S^n \to \mathbb{R}_+$:

$$\begin{aligned}&\int f \, \mathrm{d}P^n \\ &\quad = \int P^{n-1}\big(\mathrm{d}(x_0, \dots, x_{n-1})\big) \int K_n\big((x_0, \dots, x_{n-1}), \mathrm{d}x_n\big) \, f(x_0, \dots, x_{n-1}, x_n) \\ &\quad = \cdots \\ &\quad = \int \mu_0(\mathrm{d}x_0) \int K_1(x_0, \mathrm{d}x_1) \cdots \int K_n\big((x_0, \dots, x_{n-1}), \mathrm{d}x_n\big) \, f(x_0, \dots, x_n).\end{aligned}$$

## 3.1 The canonical model

Let $\Omega := S_0 \times S_1 \times \dots$ be the set of all paths (or trajectories) $\omega = (x_0, x_1, \dots)$ with $x_i \in S_i$, and

$$\begin{aligned}&X_n(\omega) := x_n \qquad \text{(projection onto } n^{th}\text{-coordinate)}, \\ &\mathcal{A}_n := \sigma(X_0, \dots, X_n) \quad (\subset \mathcal{A}), \\ &\mathcal{A} := \sigma(X_0, X_1, \dots) = \sigma\Big(\bigcup_{n=1}^{\infty} \mathcal{A}_n\Big).\end{aligned}$$

Our main goal in this section is to construct a probability measure $P$ on $(\Omega, \mathcal{A})$ satisfying

$$\int f(X_0, \dots, X_n) \, dP = \int f \, dP^n \qquad \forall \, n = 1, 2, \dots$$

In other words, the "finite dimensional distributions" of $P$, i.e., the joint distributions of $(X_0, \dots, X_n)$ under $P$, are given by $P^n$.

**Proposition 3.1** (Ionescu-Tulcea)**.** *There exists a unique probability measure $P$ on $(\Omega, \mathcal{A})$ such that for all $n \geqslant 0$ und all $\mathcal{S}^n$-measurable functions $f : S^n \to \mathbb{R}_+$:*

$$\int_{S^n} f \, \mathrm{d}P_{(X_0, \dots, X_n)} = \int_{\Omega} f(X_0, \dots, X_n) \, \mathrm{d}P = \int_{S^n} f \, \mathrm{d}P^n. \tag{3.7}$$

*In other words: there exists a unique $P$ such that $P^n = P \circ (X_0, \dots, X_n)^{-1}$.*

*Proof. Uniqueness:* Obvious, because the collection of finite cylindrical subsets

$$\mathcal{E} := \Big\{ \bigcap_{i=0}^{n} \{X_i \in A_i\} \;\Big|\; n \geqslant 0, \; A_i \in \mathcal{S}_i \Big\}$$

is closed under intersections and generates $\mathcal{A}$.
*Existence:* Let $A \in \mathcal{A}_n$, hence

$$A = (X_0, \dots, X_n)^{-1}(A^n) \quad \text{for some } A^n \in \mathcal{S}^n, \; 1_A = 1_{A^n}(X_0, \dots, X_n)$$

In order to have (3.7) we thus have to define

$$P(A) := P^n(A^n)\,. \tag{3.8}$$

We have to check that $P$ is well-defined. To this end note that $A \in \mathcal{A}_n \subset \mathcal{A}_{n+1}$ implies

$$A = A^n \times S_{n+1} \times S_{n+2} \times \cdots \quad = A^{n+1} \times S_{n+2} \times \cdots\,,$$

hence $A^{n+1} = A^n \times S_{n+1}$. Consequently,

$$\begin{aligned} P^{n+1}(A^{n+1}) &= P^{n+1}(A^n \times S_{n+1}) \\ &= \int_{A^n} \underbrace{K_{n+1}\big((x_0,\ldots,x_n), S_{n+1}\big)}_{=1} \,\mathrm{d}P^n = P^n(A^n)\,. \end{aligned}$$

It follows that $P$ is well-defined by (3.8) on $\mathcal{B} = \bigcup_{n+1}^{\infty} \mathcal{A}$. $\mathcal{B}$ is an algebra (i.e., a collection of subsets of $\Omega$ containing $\Omega$, that is closed under complements and finite (!) unions), and $P$ is finitely additive on $\mathcal{B}$, since $P$ is ($\sigma$-) additive on $\mathcal{A}_n$ for every $n$. To extend $P$ to a $\sigma$-additive probability measure on $\mathcal{A} = \sigma(\mathcal{B})$ with the help of Caratheodory's extension theorem, it suffices now to show that $P$ is $\emptyset$-continuous, i.e., the following condition is satisfied:

$$B_n \in \mathcal{B}\,, B_n \searrow \emptyset \quad \Rightarrow \quad P(B_n) \xrightarrow{n\to\infty} 0\,.$$

(For Caratheodory's extension theorem see text books on measure theory, or Satz 1.41 in Klenke.)

W.l.o.g. $B_0 = \Omega$ and $B_n \in \mathcal{A}_n$ (if $B_n \in \mathcal{A}_m$, just repeat $B_{n-1}$ $m$-times!). Then

$$B_n = A^n \times S_{n+1} \times S_{n+2} \times \ldots$$

with

$$A^{n+1} \subset A^n \times S_{n+1}$$

and we have to show that

$$\big(P(B_n) =\big) \quad P^n(A^n) \xrightarrow{n\to\infty} 0$$

(i.e., $\inf_n P^n(A^n) = 0$).

Suppose on the contrary that

$$\inf_{n\in\mathbb{N}} P^n(A^n) > 0\,.$$

We have to show that this implies

$$\bigcap_{n=0}^{\infty} B_n \neq \emptyset\,.$$

Note that

$$P^n(A^n) = \int \mu_0(\mathrm{d}x_0)\; f_{0,n}(x_0)$$

with

$$f_{0,n}(x_0) := \int K_1(x_0, \mathrm{d}x_1) \cdots \int K_n\big((x_0, \dots, x_{n-1}), \mathrm{d}x_n\big)\ 1_{A^n}(x_0, \dots, x_n).$$

It is easy to see that the sequence $(f_{0,n})_{n\in\mathbb{N}}$ is decreasing, because

$$\begin{aligned}\int K_{n+1}\big((x_0, \dots, x_n), \mathrm{d}x_{n+1}\big)\ 1_{A^{n+1}}(x_0, \dots, x_{n+1})&\\ \leqslant \int K_{n+1}\big((x_0, \dots, x_n), \mathrm{d}x_{n+1}\big)\ 1_{A^n\times S_{n+1}}(x_0, \dots, x_{n+1})&\\ = 1_{A^n}(x_0, \dots, x_n)\,,&\end{aligned}$$

hence

$$\begin{aligned}f_{0,n+1}(x_0) &= \int K_1(x_0, \mathrm{d}x_1) \cdots \int K_{n+1}\big((x_0, \dots, x_n), \mathrm{d}x_{n+1}\big)\ 1_{A^{n+1}}(x_0, \dots, x_{n+1})\\ &\le \int K_1(x_0, \mathrm{d}x_1) \cdots \int K_n\big((x_0, \dots, x_{n-1}), \mathrm{d}x_n\big)\ 1_{A^n}(x_0, \dots, x_n) = f_{0,n}(x_0)\,.\end{aligned}$$

In particular,

$$\int \inf_{n\in\mathbb{N}} f_{0,n}\ \mathrm{d}\mu_0 = \inf_{n\in\mathbb{N}} \int f_{0,n}\ \mathrm{d}\mu_0 = \inf_{n\in\mathbb{N}} P^n(A^n) > 0\,.$$

Therefore we can find some $\bar{x}_0 \in S_0$ with

$$\inf_{n\in\mathbb{N}} f_{0,n}(\bar{x}_0) > 0\,.$$

On the other hand we can write

$$f_{0,n}(\bar{x}_0) = \int K_1(\bar{x}_0, \mathrm{d}x_1)\ f_{1,n}(x_1)$$

with

$$\begin{aligned}f_{1,n}(x_1) := \int K_2\big((\bar{x}_0, x_1), \mathrm{d}x_2\big)&\\ \cdots \int K_n\big((\bar{x}_0, x_1, \dots, x_{n-1}), \mathrm{d}x_n\big)\ \mathbb{I}_{A^n}(\bar{x}_0, x_1, \dots, x_n)\,.&\end{aligned}$$

Using the same argument as above (with $\mu_1 = K_1(\bar{x}_0, \cdot)$) we can find some $\bar{x}_1 \in S_1$ with

$$\inf_{n\in\mathbb{N}} f_{1,n}(\bar{x}_1) > 0\,.$$

Iterating this procedure, we find for any $i = 0, 1, \dots$ some $\bar{x}_i \in S_i$ such that for all

$m \geqslant 0$

$$\inf_{n\in\mathbb{N}} \int K_m\big((\bar{x}_0,\dots,\bar{x}_{m-1}),\mathrm{d}x_m\big)$$
$$\cdots \int K_n\big((\bar{x}_0,\dots,\bar{x}_{m-1},x_m,\dots,x_{n-1}),\mathrm{d}x_n\big)$$
$$1_{A^n}(\bar{x}_0,\dots,\bar{x}_{m-1},x_m,\dots,x_n)$$
$$> 0.$$

In particular, if $m = n$

$$0 < \int K_m\big((\bar{x}_0,\dots,\bar{x}_{m-1}),\mathrm{d}x_m\big)\; 1_{A^m}(\bar{x}_0,\dots,\bar{x}_{m-1},x_m)$$
$$\leqslant 1_{A^{m-1}}(\bar{x}_0,\dots\bar{x}_{m-1})\,,$$

so that

$$(\bar{x}_0,\dots,\bar{x}_{m-1}) \in A^{m-1} \quad \text{and} \quad \bar{\omega} := (\bar{x}_0,\bar{x}_1,\dots) \in \underbrace{B_{m-1}}_{=A^{m-1}\times S_m\times S_{m+1}\times\cdots}$$

for all $m \geq 1$, i.e.,

$$\bar{\omega} \in \bigcap_{m=0}^{\infty} B_m\,.$$

Hence the assertion is proven. □

**Definition 3.2.** Suppose that $(S_i,\mathcal{S}_i) = (S,\mathcal{S})$ for all $i = 0,1,2,\dots$. Then $(X_n)_{n\geqslant 0}$ on $(\Omega,\mathcal{A},P)$ (with $P$ as in the previous proposition is said to be a *stochastic process* (in discrete time) with *state space* $(S,\mathcal{S})$, *initial distribution* $\mu_0$ and *transition probabilities* $\big(K_n(\,\cdot\,,\,\cdot\,)\big)_{n\in\mathbb{N}}$.

## 3.2 Examples

**1) Infinite product measures**

Let

$$K_n\big((x_0,\dots,x_{n-1}),\,\cdot\,\big) = \mu_n\,,$$

independent of $(x_0,\dots,x_{n-1})$: Then

$$P =: \bigotimes_{n=0}^{\infty} \mu_n$$

is said to be the *product measure associated with* $\mu_0,\mu_1,\dots$.

For all $n \geqslant 0$ and $A_0, \ldots, A_n \in \mathcal{S}$ we have that

$$\begin{aligned} P[X_0 \in A_0, \ldots, X_n \in A_n] &\overset{\text{I.-T.}}{=} P^n(A_0 \times \cdots \times A_n) \\ &= \int \mu_0(\mathrm{d}x_0) \int \mu_1(\mathrm{d}x_1) \cdots \int \mu_n(\mathrm{d}x_n)\, \mathbb{I}_{A_0 \times \cdots \times A_n}(x_0, \ldots, x_n) \\ &= \mu_0(A_0) \cdot \mu_1(A_1) \cdots \mu_n(A_n). \end{aligned}$$

In particular, $P_{X_n} = \mu_n$ for all $n$, and the natural projections $X_0, X_1, \ldots$ are independent. We thus have the following:

**Proposition 3.3.** *Let $(\mu_n)$ be a sequence of probability measures on a measurable space $(S, \mathcal{S})$. Then there exists a probability space $(\Omega, \mathcal{A}, P)$ and a sequence $(X_n)$ of independent r.v. with $P_{X_n} = \mu_n$ for all $n$.*

We have thus proven in particular the existence of a probability space modelling infinitely many independent $0-1$-experiments!

**2) Markov chains**

$$K_n\big((x_0, \ldots, x_{n-1}), \cdot\,\big) = \tilde{K}_n(x_{n-1}, \cdot\,)$$

time-homogeneous, if $\tilde{K}_n = K$ for all $n$.

For given initial distribution $\mu$ and transition probabilities $K$ there exists a unique probability measure $P$ on $(\Omega, \mathcal{A})$, which is said to be the canonical model for the time evolution of a *Markov chain*.

**Example 3.4.** Let $S = \mathbb{R}$, $\beta > 0$, $x_0 \in \mathbb{R} \setminus \{0\}$, $\mu_0 = \delta_{x_0}$ and $K(x, \cdot\,) = N(0, \beta x^2)$ $(K(0, \cdot\,) = \delta_0)$
For which $\beta$ does the sequence $(X_n)$ converge and what is its limit?
For $n \geqslant 1$

$$\begin{aligned} \mathbb{E}[X_n^2] &\overset{\text{I.-T.}}{=} \int x_n^2\, P^n\big(\mathrm{d}(x_0, \ldots, x_n)\big) \\ &= \int \Bigg( \underbrace{\int x_n^2\, K(x_{n-1}, \mathrm{d}x_n)}_{\substack{=\beta x_{n-1}^2, \\ K(x_{n-1}, \mathrm{d}x_n) = N(0, \beta x_{n-1}^2)}} \Bigg)\, P^{n-1}(\mathrm{d}x_0, \ldots, \mathrm{d}x_{n-1}) \\ &= \beta \cdot \mathbb{E}[X_{n-1}^2] = \cdots = \beta^n x_0^2\,. \end{aligned}$$

If $\beta < 1$ it follows that

$$\mathbb{E}\Big[\sum_{n=1}^{\infty} X_n^2\Big] = \sum_{n=1}^{\infty} \mathbb{E}[X_n^2] = \sum_{n=1}^{\infty} \beta^n x_0^2 < \infty,$$

hence $\sum\limits_{n=1}^{\infty} X_n^2 < \infty$ $P$-a.s., and therefore

$$\lim_{n \to \infty} X_n = 0 \quad P\text{-a.s.}$$

A similar calculation as above for the first absolute moment yields

$$\mathbb{E}\big[|X_n|\big] = \cdots = \sqrt{\frac{2}{\pi} \cdot \beta} \cdot \mathbb{E}\big[|X_{n-1}|\big] = \cdots = \left(\frac{2}{\pi} \cdot \beta\right)^{\frac{n}{2}} \cdot \underbrace{\mathbb{E}\big(|X_0|\big)}_{=|x_0|},$$

because

$$\int |X_n| \; K(x_{n-1}, \mathrm{d}x_n) = \sqrt{\frac{2}{\pi}} \cdot \sigma = \sqrt{\frac{2}{\pi} \cdot \beta \, |x_{n-1}|}\,.$$

Consequently,

$$\mathbb{E}\Big[\sum_{n=1}^{\infty} |X_n|\Big] = \sum_{n=1}^{\infty} \left(\frac{2}{\pi} \cdot \beta\right)^{\frac{n}{2}} \cdot |x_0|\,,$$

so that also for $\beta < \frac{\pi}{2}$:

$$\lim_{n\to\infty} X_n = 0 \quad P\text{-a.s.}$$

In fact, if we define

$$\begin{aligned}\beta_0 &:= \exp\Big(-\frac{4}{\sqrt{2\pi}} \int_0^{\infty} \log x \cdot e^{-\frac{x^2}{2}} \; \mathrm{d}x\Big) \\ &= 2e^{C} \approx 3.56,\end{aligned}$$

where

$$C := \lim_{n\to\infty} \Big(\sum_{k=1}^{n} \frac{1}{k} - \log n\Big) \approx 0.577$$

denotes the Euler-Mascheroni constant, it follows that

$$\forall\, \beta < \beta_0 : X_n \xrightarrow{n\to\infty} 0 \quad P\text{-a.s. with exp. rate}$$

$$\forall\, \beta > \beta_0 : |X_n| \xrightarrow{n\to\infty} \infty \quad P\text{-a.s. with exp. rate.}$$

*Proof.* It is easy to see that for all $n$: $X_n \neq 0$ $P$-a.s. For $n \in \mathbb{N}$ we can then define

$$Y_n := \begin{cases} \frac{X_n}{X_{n-1}} & \text{on } \{X_{n-1} \neq 0\} \\ 0 & \text{on } \{X_{n-1} = 0\}. \end{cases}$$

Then $Y_1, Y_2, \ldots$ are independent r.v. with distribution $N(0, \beta)$, because for all measurable functions $f : \mathbb{R}^n \to \mathbb{R}_+$

$$\begin{aligned}\int f(Y_1, \ldots, Y_n) \, \mathrm{d}P &\overset{\text{I.-T.}}{=} \int f\Big(\frac{x_1}{x_0}, \ldots, \frac{x_n}{x_{n-1}}\Big) \cdot \Big(\frac{1}{2\pi\beta}\Big)^{\frac{n}{2}} \cdot \Big(\frac{1}{x_0^2 \cdots x_{n-1}^2}\Big)^{\frac{1}{2}} \\ &\qquad \cdot \exp\Big(-\frac{x_1^2}{2\beta x_0^2} - \cdots - \frac{x_n^2}{2\beta x_{n-1}^2}\Big) \, \mathrm{d}x_1 \ldots \mathrm{d}x_n \\ &= \int f(y_1, \ldots, y_n) \cdot \Big(\frac{1}{2\pi\beta}\Big)^{\frac{n}{2}} \cdot \exp\Big(-\frac{y_1^2 + \cdots + y_n^2}{2\beta}\Big) \, \mathrm{d}y_1 \ldots \mathrm{d}y_n\,.\end{aligned}$$

Note that

$$|X_n| = |x_0| \cdot |Y_1| \cdots |Y_n|$$

and thus

$$\frac{1}{n} \cdot \log|X_n| = \frac{1}{n} \cdot \log|x_0| + \frac{1}{n} \sum_{i=1}^{n} \log|Y_i| \, .$$

Note that $(\log|Y_i|)_{i \in \mathbb{N}}$ are independent and identically distributed with

$$\mathbb{E}\big[\log|Y_i|\big] = 2 \cdot \frac{1}{\sqrt{2\pi\beta}} \int_0^\infty \log x \cdot e^{-\frac{x^2}{2\beta}} \, \mathrm{d}x \, .$$

Kolmogorov's law of large numbers now implies that

$$\lim_{n\to\infty} \frac{1}{n} \cdot \log|X_n| = \frac{2}{\sqrt{2\pi\beta}} \int_0^\infty \log x \cdot e^{-\frac{x^2}{2\beta}} \, \mathrm{d}x \quad P\text{-a.s.}$$

Consequently,

$$|X_n| \xrightarrow{n\to\infty} 0 \quad \text{with exp. rate, if} \int \cdots < 0,$$

$$|X_n| \xrightarrow{n\to\infty} \infty \quad \text{with exp. rate, if} \int \cdots > 0.$$

Note that

$$\begin{aligned} &\frac{2}{\sqrt{2\pi\beta}} \int_0^\infty \log x \cdot e^{-\frac{x^2}{2\beta}} \, \mathrm{d}x \overset{y=\frac{x}{\sqrt{\beta}}}{=} \frac{2}{\sqrt{2\pi}} \int_0^\infty \log(\sqrt{\beta} y) \cdot e^{-\frac{y^2}{2}} \, \mathrm{d}y \\ &= \frac{1}{2} \cdot \log \beta + \frac{2}{\sqrt{2\pi}} \int_0^\infty \log y \cdot e^{-\frac{y^2}{2}} \, \mathrm{d}y \\ &< 0 \quad \Leftrightarrow \quad \beta < \beta_0. \end{aligned}$$

□

It remains to check that

$$-\frac{4}{\sqrt{2\pi}} \int_0^\infty \log x \cdot e^{-\frac{x^2}{2}} \, \mathrm{d}x = \log 2 + C$$

where $C$ is the Euler-Mascheroni constant (Exercise!)

**Example 3.5.** Consider independent 0-1-experiments with success probability $p \in [0,1]$ but suppose that $p$ ist unknown. In the canonical model:

$$S_i := \{0,1\}, \ i \in \mathbb{N}; \quad \Omega := \{0,1\}^{\mathbb{N}},$$

$$X_i : \Omega \to \{0,1\}, \ i \in \mathbb{N}, \quad \text{projections},$$

$$\mu_i := p\varepsilon_1 + (1-p)\varepsilon_0, \ i \in \mathbb{N}; \quad P_p := \bigotimes_{i=1}^{\infty} \mu_i$$

$\mathcal{A}_n$ and $\mathcal{A}$ are defined as above.

Since $p$ is unknown, we choose an a priori distribution $\mu$ on $\big([0,1], \mathcal{B}([0,1])\big)$ (as a distribution for the unknown parameter $p$).

*Claim:* $K(p, \cdot\,) := P_p(\,\cdot\,)$ is a transition probability from $\big([0,1], \mathcal{B}([0,1])\big)$ to $(\Omega, \mathcal{A})$.

*Proof.* We only need to show that for given $A \in \mathcal{A}$ the mapping $p \mapsto P_p(A)$ is measurable on $[0,1]$. To this end define

$$\mathcal{D} := \big\{ A \in \mathcal{A} \;\big|\; p \mapsto P_p(A) \text{ is } \mathcal{B}([0,1])\text{-measurable} \big\}$$

Then $\mathcal{D}$ is a Dynkin system and contains all finite cylindrical sets

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

because

$$P_p(X_1 = x_1, \ldots, X_n = x_n) = p^{\sum_{i=1}^n x_i}(1-p)^{n - \sum_{i=1}^n x_i}$$

is measurable (even continuous) in $p$!

The claim now follows from the fact, that the finite cylindrical sets are closed under intersections and generate $\mathcal{A}$. □

Let $\bar{P} := \mu \otimes K$ on $\bar{\Omega} := [0,1] \times \Omega$ with $\mathcal{B}([0,1]) \otimes \mathcal{A}$. Using Remark 2.6 it follows that $\bar{P}$ has marginal distributions $\mu$ and

$$P(\,\cdot\,) := \int P_p(\,\cdot\,)\; \mu(\mathrm{d}p) \tag{3.9}$$

on $(\Omega, \mathcal{A})$. The integral can be seen as mixture of $P_p$ according to the a priori distribution $\mu$.

Note: The $X_i$ are no longer independent under $P$!

We now calculate the initial distribution $P_{X_1}$ and the transition probabilities in the particular case where $\mu$ is the Lebesgue measure (i.e., the uniform distribution on the unknown parameter $p$):

$$\begin{aligned} P \circ X_1^{-1} &= \int \big(p\varepsilon_1 + (1-p)\varepsilon_0\big)(\,\cdot\,)\; \mu(\mathrm{d}p) \\ &= \int p\; \mu(\mathrm{d}p) \cdot \varepsilon_1 + \int (1-p)\; \mu(\mathrm{d}p) \cdot \varepsilon_0 = \frac{1}{2} \cdot \varepsilon_1 + \frac{1}{2} \cdot \varepsilon_0. \end{aligned}$$

For given $n \in \mathbb{N}$ and $x_1, \dots, x_n \in \{0,1\}$ with $k := \sum_{i=1}^n x_i$ it follows that

$$\begin{aligned} P[X_{n+1} = 1 \,|\, X_1 = x_1, \dots, X_n = x_n] \\ &= \frac{P[X_{n+1} = 1, X_n = x_n, \dots, X_1 = x_1]}{P[X_n = x_n, \dots, X_1 = x_1]} \\ &\overset{(3.9)}{=} \frac{\displaystyle\int p^{k+1}(1-p)^{n-k}\,\mu(\mathrm{d}p)}{\displaystyle\int p^k(1-p)^{n-k}\,\mu(\mathrm{d}p)} \\ &= \frac{\Gamma(k+2)\Gamma(n-k+1)}{\Gamma(n+3)} \frac{\Gamma(n+2)}{\Gamma(k+1)\Gamma(n-k+1)} = \frac{k+1}{n+2} \\ &= \underbrace{\left(1 - \frac{n}{n+2}\right)\cdot\frac{1}{2} + \frac{n}{n+2}\cdot\frac{k}{n}}_{\text{convex combination}}. \end{aligned}$$

**Proposition 3.6.** *Let $P$ be a probability measure on $(\Omega, \mathcal{A})$ ("canonical model"), and*

$$\mu_n := P \circ X_n^{-1}, \quad n \in \mathbb{N}_0.$$

*Then:*

$$X_n,\ n \in \mathbb{N},\ \textit{independent} \quad \Leftrightarrow \quad P = \bigotimes_{n=0}^{\infty} \mu_n.$$

*Proof.* Let $\tilde{P} := \bigotimes_{n=0}^{\infty} \mu_n$. Then

$$P = \tilde{P}$$

if and only if for all $n \in \mathbb{N}_0$ and all $A_0 \in \mathcal{S}_0, \dots, A_n \in \mathcal{S}_n$

$$\begin{aligned} P[X_0 \in A_0, \dots, X_n \in A_n] &= \tilde{P}[X_0 \in A_0, \dots, X_n \in A_n] \\ &= \prod_{i=0}^n \mu_i(A_i) = \prod_{i=0}^n P[X_i \in A_i], \end{aligned}$$

which is the case if and only if $X_n$, $n \in \mathbb{N}_0$, are independent. □

**Definition 3.7.** Let $S_i := S$, $i \in \mathbb{N}_0$, $(\Omega, \mathcal{A})$ be the canonical model and $P$ be a probability measure on $(\Omega, \mathcal{A})$. In particular, $(X_n)_{n \geqslant 0}$ is a stochastich process in the sense of Definition 3.2. Let $J \subset \mathbb{N}_0$, $|J| < \infty$. Then the distribution of $(X_j)_{j \in J}$ under $P$

$$\mu_J := P \circ (X_i)_{i \in J}^{-1}$$

is said to be the *finite dimensional distribution* (w.r.t. $J$) on $(S^J, \mathcal{S}^J)$.

**Remark 3.8.** *$P$ is uniquely determined by its finite-dimensional distributions resp. by*

$$\mu_{\{0,\dots,n\}}, \quad n \in \mathbb{N}.$$

# 4 Stationarity

Let $(S, \mathcal{S})$ be a measurable space, $\Omega = S^{\mathbb{N}_0}$ and $(\Omega, \mathcal{A})$ be the associated canonical model. Let $P$ be a probability measure on $(\Omega, \mathcal{A})$.

**Definition 4.1.** The mapping $T : \Omega \to \Omega$, defined by

$$\omega = (x_0, x_1, \dots) \mapsto T\omega := (x_1, x_2, \dots)$$

is called the *shift-operator* on $\Omega$.

**Remark 4.2.** *For all* $n \in \mathbb{N}_0$, $A_0, \dots, A_n \in \mathcal{S}$

$$T^{-1}\big(\{X_0 \in A_0, \dots, X_n \in A_n\}\big) = \{X_1 \in A_0, \dots, X_{n+1} \in A_n\}.$$

*In particular:* $T$ *is* $\mathcal{A}/\mathcal{A}$*-measurable*

**Definition 4.3.** The measure $P$ is said to be *stationary* (or *shift-invariant*) if

$$P \circ T^{-1} = P\,.$$

**Proposition 4.4.** *The measure* $P$ *is stationary if and only if for all* $k, n \in \mathbb{N}_0$*:*

$$\mu_{\{0,\dots,n\}} = \mu_{\{k,\dots,k+n\}}.$$

*Proof.*

$$\begin{aligned} &P \circ T^{-1} = P \\ &\quad \Leftrightarrow P \circ T^{-k} = P \quad \forall k \in \mathbb{N}_0 \\ &\quad \overset{3.8}{\Leftrightarrow} (P \circ T^{-k}) \circ (X_0, \dots, X_n)^{-1} = P \circ (X_0, \dots, X_n)^{-1} \quad \forall\, k, n \in \mathbb{N}_0 \\ &\quad \Leftrightarrow \mu_{\{k,\dots,n+k\}} = \mu_{\{0,\dots,n\}}. \end{aligned}$$

□

**Remark 4.5.** *(i) The last proposition implies in the particular case*

$$P = \bigotimes_{i=1}^{\infty} \mu_n \quad \textit{with} \quad \mu_n := P \circ X_n^{-1}$$

*that*

$$P \textit{ stationary} \quad \Leftrightarrow \quad \mu_n = \mu_0 \;\; \forall\, n \in \mathbb{N}.$$

*(ii) If* $P = \bigotimes \mu_n$ *as in (i), hence* $X_0, X_1, X_2, \dots$ *independent, Kolmogorov's zero-one law implies that* $P = 0 - 1$ *on the tail-field*

$$\mathcal{A}^* := \bigcap_{n \geqslant 0} \sigma(X_n, X_{n+1}, \dots)$$

**Proposition 4.6.** *Let* $P = \bigotimes_{n=0}^{\infty} \mu_n$, $\mu_n := P \circ X_n^{-1}$, $n \in \mathbb{N}_0$. *Then* $P$ *is* ergodic, *i.e.*

$$P = 0 - 1 \text{ on } \mathcal{I} := \{A \in \mathcal{A} \,|\, T^{-1}(A) = A\}\,.$$

$\mathcal{I}$ *is called the* $\sigma$*-algebra of shift-invariant sets.*

*Proof.* Using part (ii) of the previous remark, it suffices to show that $\mathcal{I} \subset \mathcal{A}^*$. But

$$\begin{aligned} A \in \mathcal{I} \Rightarrow A &= T^{-n}(A) \in \sigma(X_n, X_{n+1}, \dots) \quad \forall\, n \in \mathbb{N} \\ &\Rightarrow A \in \mathcal{A}^*. \end{aligned}$$

□