Stochastik I

Sommersemester 2004

Prof.Dr. L. Baringhaus

Überarbeitet von Alexander Seifert.

Anregungen und Fehler schickt bitte an alexseifert@gmx.net.

Inhaltsverzeichnis

Kapitel 1

Maß und Wahrscheinlichkeit

Definition 1.1

Wir definieren eine nichtleere Menge Ω als die Menge der möglichen Ereignisse eines Zufallsexperiments und $\mathcal{A} \subset \mathfrak{P}(\Omega)$ sei ein System von Teilmengen von Ω mit folgenden Eigenschaften:
(i) $\Omega \in \mathcal{A}$,
(ii) aus $A \in \mathcal{A}$ folgt stets $A^c \in \mathcal{A}$,
(iii) für jede Folge $A_1, A_2, ... \in \mathcal{A}$ folgt $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.
Dann heißt $\mathcal{A}$ die σ**-Algebra** der interessierenden Ereignisse.
Des Weiteren definieren wir eine Abbildung $P : \mathcal{A} \to [0,1]$ mit folgenden Eigenschaften:
(i) $P(\Omega) = 1$ (**Normiertheit** von P),
(ii) Für jede Folge $(A_n)_{n=1}^{\infty}$ von paarweise disjunkten Mengen $A_n \in \mathcal{A}$ gilt:

$$P\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} P(A_n) \quad (\sigma\textbf{-Additivität} \text{ von } P).$$

Die Abbildung P heißt **Wahrscheinlichkeitsmaß**. Das Tripel $(\Omega, \mathcal{A}, P)$ heißt **Wahrscheinlichkeitsraum**.
Für $A \in \mathcal{A}$ ist $P(A)$ die Wahrscheinlichkeit für das Eintreten des **Ereignisses** A.

Beispiel 1.2

Würfelwurf: Sei $\Omega = \{1,...,6\}$, $A = \{2,4,6\}$. Möglicher Ansatz für P:

$$P(A) = \frac{|A|}{|\Omega|},$$

mit $|\cdot|$ definiert durch die Anzahl der Elemente einer Menge.

Eigenschaften einer σ-Algebra

Das unmögliche Ereignis $\varnothing = \Omega^c \in \mathcal{A}$ liegt in jeder σ-Algebra. Somit gilt:
(i) Aus $A, B \in \mathcal{A}$ folgt $A \cup B = A \cup B \cup \varnothing \cup \varnothing \cup ... \in \mathcal{A}$,
(ii) Aus $A, B \in \mathcal{A}$ folgt $A \cap B = (A^c \cup B^c)^c \in \mathcal{A}$,
(iii) Für jede Folge $A_1, A_2, ... \in \mathcal{A}$ folgt $\bigcap_{n=1}^{\infty} A_n = \Big(\bigcup_{n=1}^{\infty} A_n^c\Big)^c \in \mathcal{A}$.

Logische Beschreibung von Ereignissen

Für $A \in \mathcal{A}$ bezeichnen wir A^c mit „nicht A" oder „A tritt nicht ein". Für $A, B \in \mathcal{A}$ bezeichnen wir:
(i) $A \cup B$ mit „A oder B tritt ein",
(ii) $A \cap B$ mit „A und B treten ein",
(iii) $A \subset B$ mit „das Ereignis A hat das Ereignis B zur Folge",
(iv) $A \triangle B := (A \cup B) \cap (A \cap B)^c = (A \cap B^c) \cup (B \cap A^c)$ mit „entweder A oder B tritt ein".
Allgemein: Ist $\omega \in \Omega$ ein Ereignis, so bedeutet $\omega \in A$, dass A eintritt.

Beispiel 1.3

Sei $\Omega \neq \varnothing$ eine endliche Menge, $\mathcal{A} = \mathfrak{P}(\Omega)$, $A \in \mathcal{A}$,

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\text{Anzahl der für das Eintreten von } A \text{ günstigen Ergebnisse}}{\text{Anzahl der möglichen Ergebnisse}}.$$

Dann heißt $(\Omega, \mathcal{A}, P)$ ein **diskreter Laplacescher Wahrscheinlichkeitsraum**.

Grundregeln der Kombinatorik

Für $M = \{1,...,n\}$, $n \in \mathbb{N}$, definieren wir:
(i) $\mathrm{P}_n^r = \{(x_1,...,x_r) \in \mathbb{R}^r \mid x_1,...,x_r \in M\} = M^r = \underbrace{M \times ... \times M}_{r\text{-mal}}$,
(ii) $\mathrm{P}_n^{(r)} = \{(x_1,...,x_r) \in \mathrm{P}_n^r \mid x_1,...,x_r \text{ paarweise verschieden}\}$, $r \leq n$,
(iii) $\mathrm{K}_n^{(r)} = \{(x_1,...,x_r) \in \mathrm{P}_n^r \mid x_1 < ... < x_r\}$, $r \leq n$,
(iv) $\mathrm{K}_n^r = \{(x_1,...,x_r) \in \mathrm{P}_n^r \mid x_1 \leq ... \leq x_r\}$.

Satz 1.4

(i) $|P_n^r| = n^r$,

(ii) $|P_n^{(r)}| = n(n-1)\cdot\ldots\cdot(n-r+1)$, insbesondere: $|P_n^{(n)}| = n!$,

(iii) $|K_n^{(r)}| = \binom{n}{r}$,

(iv) $|K_n^r| = \binom{n+r-1}{r}$.

Beweis

(i) klar!

(ii) klar!

(iii) Aus $|P_n^{(r)}| = |K_n^{(r)}| r!$ folgt

$$|K_n^{(r)}| = \frac{n(n-1)\cdot\ldots\cdot(n-r+1)}{r!} = \binom{n}{r}.$$

(iv) Es gibt eine Bijektion $K_n^r \ni (x_1,\ldots,x_r) \leftrightarrow (x_1, x_2+1, x_3+2, \ldots, x_r+r-1) \in K_{n+r-1}^{(r)}$. Somit erhalten wir:

$$|K_n^r| = |K_{n+r-1}^{(r)}| = \binom{n+r-1}{r}.$$

□

Definition 1.5

Jedes $(x_1,\ldots,x_r) \in K_n^r$ ist eindeutig bestimmt durch seinen **Besetzungszahlvektor**

$$\left(\sum\nolimits_{j=1}^r \delta_{1,x_j}, \ldots, \sum\nolimits_{j=1}^r \delta_{n,x_j}\right) \in \{(k_1,\ldots,k_n) \in \mathbb{N}_0^n \mid k_1 + \ldots + k_n = r\},$$

wobei δ_{i,x_j} das Kronecker-Symbol „Kronecker delta" ist und ist definiert durch

$$\delta_{i,x_j} := \begin{cases} 1, & x_j = i, \\ 0, & \text{sonst.} \end{cases}$$

Beispiel 1.6

Sei $n = 6$, $r = 7$. Dann haben wir zu $(2,2,2,3,3,5,5) \in K_6^7$ den Besetzungszahlvektor $(0,3,2,0,2,0)$.

Folgerung 1.7

$$|\{(k_1,\ldots,k_n) \in \mathbb{N}_0^n \mid k_1 + \ldots + k_n = r\}| = \binom{n+r-1}{r}.$$

Bemerkung 1.8

(i) Sei $M = \{1,\ldots,n\}$, $r \le n$. Dann gilt:

$$|\mathfrak{P}_r(M)| = |\{A \subset M \mid |A| = r\}| = |K_n^{(r)}| = \binom{n}{r} \quad (\text{Stimmt auch für } r = 0)$$

und damit

$$|\mathfrak{P}(M)| = \left|\bigcup_{r=0}^n \mathfrak{P}_r(M)\right| = \sum_{r=0}^n |\mathfrak{P}_r(M)| = \sum_{r=0}^n \binom{n}{r} = \sum_{r=0}^n \binom{n}{r} \cdot 1^r \cdot 1^{n-r} = (1+1)^n = 2^n.$$

(ii) Aus Folgerung 1.7 folgt:

$$|\{(l_1,\ldots,l_n) \in \mathbb{N}^n \mid l_1 + \ldots + l_n = r\}| = |\{(k_1,\ldots,k_n) \in \mathbb{N}_0^n \mid k_1 + \ldots + k_n = r - n\}| = \binom{r-1}{r-n} = \binom{r-1}{n-1}$$

für $n \le r$.

Folgerung 1.9

Es sollen r nicht unterscheidbare Kugeln auf n unterscheidbare Urnen verteilt werden:

(i) Es gibt $\binom{n+r-1}{r}$ verschiedene Aufteilungen.

(ii) Es gibt $\binom{r-1}{n-1}$ verschiedene Aufteilungen, so dass keine Urne leer bleibt. Es gibt n^r Möglichkeiten, r unterscheidbare Kugeln auf n unterscheidbare Urnen zu verteilen.

Es gibt n^r Möglichkeiten, r unterscheidbare Kugeln auf n unterscheidbare Urnen zu verteilen.

Satz 1.10

Es sei $k_1,\ldots,k_n \in \mathbb{N}_0$, $k_1 + \ldots + k_n = r$, $r,n \in \mathbb{N}$. Dann gilt:

$$\left|\{(x_1,\dots,x_r) \in \mathrm{P}_n^r \mid |\{j \mid x_j = i\}| = k_i, i = 1,\dots,n\}\right| = \frac{r!}{k_1! \cdot \ldots \cdot k_n!}.$$

Beweis

$$\begin{aligned}\left|\{(x_1,\dots,x_r) \in \mathrm{P}_n^r \mid |\{j \mid x_j = i\}| = k_i, i = 1,\dots,n\}\right| &= \binom{r}{k_1}\binom{r-k_1}{k_2}\binom{r-k_1-k_2}{k_3}\cdot\ldots\cdot\binom{r-k_1-\ldots-k_{n-1}}{k_n} \\ &= \frac{r!}{k_1!\,\cancel{(r-k_1)!}}\,\frac{\cancel{(r-k_1)!}}{k_2!\,\cancel{(r-k_1-k_2)!}}\cdot\ldots\cdot\frac{\cancel{(r-k_1-\ldots-k_{n-1})!}}{k_n!\,\underbrace{(r-k_1-\ldots-k_n)!}_{=0!}} \\ &= \frac{r!}{k_1!\cdot\ldots\cdot k_n!}.\end{aligned}$$

Genau k_i der Komponenten von $(x_1,\dots,x_r)$ sind gleich i. □

Beispiel 1.11

Sei $A = \{1,\dots,r\}$ und $B = \{1,\dots,n\}$. Dann gilt:

(i) $|\{f \mid f : A \to B\}| = n^r$, $r,n \in \mathbb{N}$,

(ii) $|\{f \mid f : A \to B \text{ ist injektiv}\}| = n(n-1)\cdot\ldots\cdot(n-r+1)$, $r \le n$,

(iii) Sei $\Omega = \{f \mid f : A \to B \text{ ist surjektiv}\}$. Dann ist $|\Omega| = a(r,n)$ für $r,n \in \mathbb{N}$ mit:

$$\begin{aligned}a(r,n) &= \left|\bigcup_{\substack{(k_1,\dots,k_n)\in\mathbb{N}^n \\ k_1+\ldots+k_n=r}} \{f \in \Omega \mid |f^{-1}(\{i\})| = k_i, i = 1,\dots,n\}\right| \\ &= \sum_{\substack{(k_1,\dots,k_n)\in\mathbb{N}^n \\ k_1+\ldots+k_n=r}} |\{f \in \Omega \mid |f^{-1}(\{i\})| = k_i, i = 1,\dots,n\}| \\ &= \sum_{\substack{(k_1,\dots,k_n)\in\mathbb{N}^n \\ k_1+\ldots+k_n=r}} \frac{r!}{k_1!\cdot\ldots\cdot k_n!}.\end{aligned}$$

Durch einen Potenzreihenansatz erhalten wir

$$\begin{aligned}\sum_{r=n}^{\infty} \underline{\frac{a(r,n)}{r!}} x^r &= \sum_{r=n}^{\infty}\left(\sum\nolimits_{\substack{(k_1,\dots,k_n)\in\mathbb{N}^n \\ k_1+\ldots+k_n=r}} \frac{1}{k_1!\cdot\ldots\cdot k_n!}\right)x^r \\ &= \left(\sum\nolimits_{l=1}^{\infty}\frac{1}{l!}x^l\right)^n = (e^x-1)^n = \sum_{k=0}^{n}\binom{n}{k}(-1)^k e^{x(n-k)} \\ &= \sum_{k=0}^{n}\binom{n}{k}(-1)^k \sum\nolimits_{r=0}^{\infty}\frac{(n-k)^r}{r!}x^r \\ &= \sum_{r=0}^{\infty}\underline{\frac{1}{r!}\left(\sum\nolimits_{k=0}^{n}\binom{n}{k}(-1)^k(n-k)^r\right)}x^r.\end{aligned}$$

Da für $r < n$ gilt

$$a(r,n) = \sum\nolimits_{k=0}^{n}\binom{n}{k}(-1)^k(n-k)^r = 0,$$

liefert der Koeffizientenvergleich:

$$a(r,n) = \sum\nolimits_{k=0}^{n}\binom{n}{k}(-1)^k(n-k)^r \quad \text{für } r,n \in \mathbb{N}.$$

Außerdem gilt:

$$a(n,n) = n! = \sum\nolimits_{k=0}^{n}\binom{n}{k}(-1)^k(n-k)^n.$$

Folgerung 1.12

(i) Die Anzahl der Möglichkeiten, r unterscheidbare Kugeln auf n unterscheidbare Urnen zu verteilen, so dass keine Urne leer bleibt, ist

$$a(r,n) = \sum_{k=0}^{n}\binom{n}{k}(-1)^k(n-k)^r.$$

(ii) Die Anzahl der Möglichkeiten, r unterscheidbare Kugeln auf n unterscheidbare Urnen zu verteilen, so dass genau m Urnen leer bleiben, ist

$$\binom{n}{m}a(r,n-m) = \binom{n}{m}\sum_{k=0}^{n-m}\binom{n-m}{k}(-1)^k(n-m-k)^r.$$

(iii) Sei A, B wie im vorigen Beispiel. Betrachten wir $\Omega := \{ f \mid f : A \to B \}$, $\mathcal{A} = \mathfrak{P}(\Omega)$, P das diskrete Laplacesche Wahrscheinlichkeitsmaß auf Ω. Identifiziere jedes $f \in \Omega$ mit einer Aufteilung r unterscheidbarer Kugeln auf n unterscheidbare Urnen. Sei

$$A_{m,n} = \{\text{genau } m \text{ Urnen bleiben leer}\}.$$

Dann gilt für die Wahrscheinlichkeit:

$$P(A_{m,n}) = \frac{|A_{m,n}|}{|\Omega|} = \frac{\binom{n}{m} a(r, n-m)}{n^r} = \binom{n}{m} \sum_{k=0}^{n-m} (-1)^k \left(1 - \frac{m+k}{n}\right)^r.$$

Beispiel 1.13

Sei $\Omega_n = \{\pi \mid \pi : \{1,\ldots,n\} \to \{1,\ldots,n\} \text{ bijektiv}\}$. Dann heißt $i \in \{1,\ldots,n\}$ ein **Fixpunkt** von $\pi \in \Omega_n$, wenn $\pi(i) = i$. Wir definieren:

$$a_n(k) := \left|\{\pi \in \Omega_n \mid \pi \text{ hat genau } k \text{ Fixpunkte}\}\right| \text{ für } 0 \leq k \leq n,\ a_0 :\equiv 1.$$

Es gilt

$$n! = |\Omega_n| = \left|\bigcup\nolimits_{k=0}^{n} \{\pi \in \Omega_n \mid \pi \text{ hat genau } k \text{ Fixpunkte}\}\right| = \sum_{k=0}^{n} a_n(k) = \sum_{k=0}^{n} \binom{n}{k} a_{n-k}(0)$$

und damit erhalten wir

$$1 = \sum_{k=0}^{n} \frac{1}{k!} \frac{a_{n-k}(0)}{(n-k)!}.$$

Nun gilt für $|x| < 1$

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n \cdot 1 = \sum_{n=0}^{\infty} \left(\sum\nolimits_{k=0}^{n} \frac{1}{k!} \frac{a_{n-k}(0)}{(n-k)!}\right) x^n \overset{l:=n-k}{=} \left(\sum\nolimits_{k=0}^{\infty} \frac{1}{k!} x^k\right)\left(\sum\nolimits_{l=0}^{\infty} \frac{a_l(0)}{l!} x^l\right) = e^x \sum_{l=0}^{\infty} \frac{a_l(0)}{l!} x^l.$$

Es folgt:

$$\sum_{n=0}^{\infty} \underline{\frac{a_n(0)}{n!}} x^n = \frac{1}{1-x} e^{-x} = \left(\sum\nolimits_{k=0}^{\infty} 1 \cdot x^k\right)\left(\sum\nolimits_{l=0}^{\infty} \frac{(-1)^l}{l!} x^l\right) \overset{l=n-k}{=} \sum_{n=0}^{\infty} \underline{\left(\sum\nolimits_{k=0}^{n} \frac{(-1)^k}{k!}\right)} x^n.$$

Koeffizientenvergleich liefert

$$a_n(0) = n! \sum_{k=0}^{n} \frac{(-1)^k}{k!}.$$

Also folgt aus $a_n(k) = \binom{n}{k} a_{n-k}(0)$

$$a_n(k) = \binom{n}{k} (n-k)! \sum_{j=0}^{n-k} \frac{(-1)^j}{j!} = \frac{n!}{k!} \sum_{j=0}^{n-k} \frac{(-1)^j}{j!}.$$

Sei $\mathcal{A} = \mathfrak{P}(\Omega_n)$, P das diskrete Wahrscheinlichkeitsmaß, $A_{n,k} = \{\pi \in \Omega_n \mid \pi \text{ hat genau } k \text{ Fixpunkte}\}$. Dann erhalten wir für die Wahrscheinlichkeit:

$$P(A_{n,k}) = \frac{|A_{n,k}|}{|\Omega_n|} = \frac{a_n(k)}{n!} = \frac{1}{k!} \sum_{j=0}^{n-k} \frac{(-1)^j}{j!}.$$

Spezielle Fälle sind:

$$P(A_{n,0}) = \sum_{j=0}^{n} \frac{(-1)^j}{j!} \approx e^{-1},$$

$$P(A_{n,0}^c) = \frac{|\Omega_n \setminus A_{n,0}|}{|\Omega_n|} = 1 - P(A_{n,0}) = 1 - \sum_{j=0}^{n} \frac{(-1)^j}{j!} = \sum_{j=1}^{n} \frac{(-1)^{j-1}}{j!} \approx 1 - e^{-1} \approx 0{,}63.$$

$P(A_{n,0}^c)$ ist die Wahrscheinlichkeit für mindestens einen Fixpunkt.

Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum. Dann gilt:

$$P(\varnothing) = P(\varnothing \cup \varnothing \cup \ldots) = \sum_{k=1}^{\infty} P(\varnothing) \ \Rightarrow\ P(\varnothing) = 0.$$

Sei $A, B \in \mathcal{A}$ mit $A \cap B = \varnothing$. Dann folgt:

$$\begin{aligned} P(A \cup B) &= P(A \cup B \cup \varnothing \cup \varnothing \cup \ldots) \\ &= P(A) + P(B) + \underbrace{P(\varnothing) + P(\varnothing) + \ldots}_{=0} \\ &= P(A) + P(B). \end{aligned}$$

Sei $A, B \in \mathcal{A}$ mit $A \subset B$. Dann ist $B = A \cup (B \cap A^c)$, wobei A und $B \cap A^c$ disjunkt sind. Es gilt:

$$P(B) = P(A \cup (B \cap A^c)) = P(A) + P(B \cap A^c)$$

Also erhalten wir

$$P(B \cap A^c) = P(B) - P(A) \text{ (\textbf{Subtraktivität} des Wahrscheinlichkeitsmaßes).}$$

Ferner gilt wegen $P(B \cap A^c) \geq 0$:

$$P(A) \leq P(B) \text{ (\textbf{Isotonie} des Wahrscheinlichkeitsmaßes).}$$

Definition 1.14

Sind $A, B \in \mathcal{A}$ mit $\mathcal{A}$ σ-Algebra und $A \cap B = \varnothing$, so sei $A + B := A \cup B$.
Sind $A_1, A_2, ... \in \mathcal{A}$ paarweise disjunkt, so sei $\sum_{n=1}^{\infty} A_n := \bigcup_{n=1}^{\infty} A_n$.
Aus der **σ-Additivität** des Wahrscheinlichkeitsmaßes folgt:

$$P\left(\sum_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n) \text{ für } A_1, A_2, ... \in \mathcal{A} \text{ paarweise disjunkt.}$$

Für paarweise disjunkte $A_1, ..., A_n \in \mathcal{A}$ gilt:

$$P(A_1 + ... + A_n) = P(A_1) + ... + P(A_n) \text{ (\textbf{endliche } \boldsymbol{\sigma}\textbf{-Additivität}).}$$

Seien $A_1, A_2, ... \in \mathcal{A}$ (nicht notwendig disjunkt). Dann gilt:

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} P(A_n) \text{ (\textbf{Sub-}\boldsymbol{\sigma}\textbf{-Additivität}).}$$

Denn es gilt

$$\bigcup_{n=1}^{\infty} A_n = A_1 + \sum_{n=2}^{\infty} (A_n \cap \underbrace{(A_1^c \cap ... \cap A_{n-1}^c)}_{=(A_1 \cup ... \cup A_{n-1})^c})$$

und damit

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = P(A_1) + \sum_{n=2}^{\infty} \underbrace{P(\overbrace{A_n \cap (A_1^c \cap ... \cap A_{n-1}^c)}^{\subset A_n})}_{\leq P(A_n)}.$$

Ebenso gilt $P(A_1 \cup ... \cup A_n) \leq P(A_1) + ... + P(A_n)$ (**Subadditivität von** P).

Definition 1.15

Sei $A_1, A_2, ... \in \mathcal{A}$ mit $A_1 \subset A_2 \subset ...$. Dann schreiben wir $A_n \uparrow A$, wenn $A = \bigcup_{n=1}^{\infty} A_n$ (**Isotonie**).
Sei $A_1, A_2, ... \in \mathcal{A}$ mit $A_1 \supset A_2 \supset ...$. Dann schreiben wir $A_n \downarrow A$, wenn $A = \bigcap_{n=1}^{\infty} A_n$ (**Antitonie**).

Satz 1.16

(i) Gilt für eine Folge von Ereignissen A_n, $n \in \mathbb{N}$, mit $A_n \uparrow A$, so ist $P(A) = \lim_{n\to\infty} P(A_n)$,
(ii) Gilt für eine Folge von Ereignissen A_n, $n \in \mathbb{N}$, mit $A_n \downarrow A$, so ist $P(A) = \lim_{n\to\infty} P(A_n)$.

Beweis

(i) Es gilt

$$A = \bigcup_{n=1}^{\infty} A_n = A_1 + \sum_{n=2}^{\infty} (A_n \cap A_{n-1}^c).$$

Also folgt:

$$\begin{aligned} P(A) &= P(A_1) + \sum_{n=2}^{\infty} P(A_n \cap A_{n-1}^c) = P(A_1) + \sum_{n=2}^{\infty} (P(A_n) - P(A_{n-1})) \\ &= P(A_1) + \lim_{m\to\infty} \underbrace{\sum_{n=2}^{m} (P(A_n) - P(A_{n-1}))}_{=P(A_m) - P(A_1) \text{ (Teleskopsumme)}} = P(A_1) + \lim_{m\to\infty} (P(A_m) - P(A_1)) \\ &= \lim_{m\to\infty} P(A_m). \end{aligned}$$

(ii) $A_n \downarrow A$ ist äquivalent zu $A_n^c \uparrow A^c$. Also folgt aus (i):

$$\begin{aligned} P(\underbrace{A^c}_{=\Omega \cap A^c}) &= P(\Omega) - P(A) = 1 - P(A) \text{ (Subtraktivität von } P\text{)} \\ &= \lim_{n\to\infty} P(A_n^c) = \lim_{n\to\infty} (1 - P(A_n)) = 1 - \lim_{n\to\infty} P(A_n). \end{aligned}$$

Also folgt:

$$P(A) = \lim_{n\to\infty} P(A_n).$$

□

Grundlagen der Maßtheorie

Definition 1.17

Sei Ω eine nichtleere Mange. Ein Mengensystem $\mathcal{D} \subset \mathfrak{P}(\Omega)$ heißt **Dynkin-System auf** Ω, wenn
(i) $\Omega \in \mathcal{D}$,
(ii) aus $A \in \mathcal{D}$ folgt $A^c \in \mathcal{D}$,
(iii) für paarweise disjunkte Mengen $A_n \in \mathcal{D}$, $n \in \mathbb{N}$, folgt $\sum_{n=1}^{\infty} A_n \in \mathcal{D}$.
Ein Mengensystem $\mathcal{E} \subset \mathfrak{P}(\Omega)$ heißt $\cap$**-stabil**, wenn aus $A, B \in \mathcal{E}$ stets $A \cap B \in \mathcal{E}$ folgt.

Lemma 1.18

Ein $\cap$-stabiles Dynkin-System ist eine σ-Algebra.
Beweis
Für $A_n \in \mathcal{D}$, $n \in \mathbb{N}$, ist wegen der $\cap$-Stabilität
$$\bigcup_{n=1}^{\infty} A_n = A_1 + \sum_{n=2}^{\infty} \underbrace{(A_n \cap (A_{n-1}^c \cap \dots \cap A_1^c))}_{\in \mathcal{D}} \in \mathcal{D}. \qquad \square$$

Definition 1.19

Für ein Mengensystem $\mathcal{E} \subset \mathfrak{P}(\Omega)$ ist
$$\sigma(\mathcal{E}) = \bigcap_{\substack{\mathcal{A} \text{ ist } \sigma\text{-Algebra auf } \Omega \\ \mathcal{E} \subset \mathcal{A}}} \mathcal{A}$$
die kleinste vom Mengensystem $\mathcal{E}$ erzeugte σ-Algebra auf Ω,
$$\mathfrak{D}(\mathcal{E}) = \bigcap_{\substack{\mathcal{D} \text{ ist Dynkin-System auf } \Omega \\ \mathcal{E} \subset \mathcal{D}}} \mathcal{D}$$
das kleinste vom Mengensystem $\mathcal{E}$ erzeugte Dynkin-System auf Ω.

Lemma 1.20

Ist $\mathcal{E} \subset \mathfrak{P}(\Omega)$ $\cap$-stabil, so ist $\sigma(\mathcal{E}) = \mathfrak{D}(\mathcal{E})$.
Beweis
Da jede σ-Algebra ein Dynkin-System ist, ist $\mathfrak{D}(\mathcal{E}) \subset \sigma(\mathcal{E})$. Zum Nachweis von $\sigma(\mathcal{E}) \subset \mathfrak{D}(\mathcal{E})$ genügt es zu zeigen, dass $\mathfrak{D}(\mathcal{E})$ $\cap$-stabil ist. Dazu betrachten wir für $E \in \mathfrak{D}(\mathcal{E})$ das Mengensystem
$$\mathcal{D}_E := \{ B \in \mathfrak{D}(\mathcal{E}) \mid B \cap E \in \mathfrak{D}(\mathcal{E}) \}.$$
$\mathcal{D}_E$ ist ein Dynkin-System auf Ω. Die Eigenschaft $B^c \in \mathcal{D}_E$, falls $B \in \mathcal{D}_E$, folgt dabei beispielsweise aus
$$(B^c \cap E)^c = E^c \cup B = E^c + B \cap E \in \mathfrak{D}(\mathcal{E}),$$
da dann nämlich auch $B^c \cap E \in \mathfrak{D}(\mathcal{E})$ ist. Ist speziell $E \in \mathcal{E}$, so ist $\mathcal{E} \subset \mathcal{D}_E$ wegen der $\cap$-Stabilität von $\mathcal{E}$, also auch $\mathfrak{D}(\mathcal{E}) \subset \mathcal{D}_E$. Für jedes $E \in \mathcal{E}$ gilt also $E \cap B \in \mathfrak{D}(\mathcal{E})$ für jedes $B \in \mathfrak{D}(\mathcal{E})$. Für beliebiges $E \in \mathfrak{D}(\mathcal{E})$ folgt jetzt auch $\mathcal{E} \subset \mathcal{D}_E$, also auch $\mathfrak{D}(\mathcal{E}) \subset \mathcal{D}_E$ und damit die $\cap$-Stabilität von $\mathfrak{D}(\mathcal{E})$. $\square$

Definition 1.21

Ein Mengensystem $\mathcal{R} \subset \mathfrak{P}(\Omega)$ heißt **Ring auf** Ω, wenn
(i) $\varnothing \in \mathcal{R}$,
(ii) aus $A, B \in \mathcal{R}$ folgt $A \cup B \in \mathcal{R}$,
(iii) aus $A, B \in \mathcal{R}$ folgt $A \cap B^c \in \mathcal{R}$.
Sind A, B Elemente eines Rings $\mathcal{R}$, so ist auch $A \cap B = A \cap (A \cap B^c)^c \in \mathcal{R}$.

Definition 1.22

Ein Mengensystem $\mathcal{S} \subset \mathfrak{P}(\Omega)$ heißt **Semi-Ring auf** Ω, wenn
(i) $\varnothing \in \mathcal{S}$,
(ii) aus $A, B \in \mathcal{S}$ folgt $A \cap B \in \mathcal{S}$,
(iii) aus $A, B \in \mathcal{S}$ die Existenz von endlich vielen paarweise disjunkten Mengen $E_1, \dots, E_n \in \mathcal{S}$ mit

$$A \cap B^c = \sum_{j=1}^{n} E_j$$

folgt.

Beispiel 1.23

(i) Es ist

$$\mathcal{E} = \{(a,b] \mid -\infty < a \le b < \infty\}$$

ein Semi-Ring auf $\mathbb{R}$. Es gilt

$$\begin{aligned}\mathcal{B} &:= \sigma(\mathcal{E}) = \sigma(\mathfrak{R}(\mathcal{E})) \\ &= \sigma(\{U \mid U \subset \mathbb{R} \text{ offen}\}) = \sigma(\{A \mid A \subset \mathbb{R} \text{ abgeschlossen}\}) = \sigma(\{K \mid K \subset \mathbb{R} \text{ kompakt}\}).\end{aligned}$$

$\mathcal{B}$ heißt die **Borelsche σ-Algebra auf** $\mathbb{R}$.

(ii) Es ist

$$\mathcal{E} = \{(a_1,b_1] \times \ldots \times (a_d,b_d] \mid -\infty < a_i \le b_i < \infty, i = 1,\ldots,d\}$$

ein Semi-Ring auf $\mathbb{R}^d$. Es gilt

$$\begin{aligned}\mathcal{B}^d &:= \sigma(\mathcal{E}) = \sigma(\mathfrak{R}(\mathcal{E})) \\ &= \sigma\left(\{U \mid U \subset \mathbb{R}^d \text{ offen}\}\right) = \sigma\left(\{A \mid A \subset \mathbb{R}^d \text{ abgeschlossen}\}\right) = \sigma\left(\{K \mid K \subset \mathbb{R}^d \text{ kompakt}\}\right).\end{aligned}$$

$\mathcal{B}^d$ heißt die **Borelsche σ-Algebra auf** $\mathbb{R}^d$.

Definition 1.24

Für ein Mengensystem $\mathcal{E} \subset \mathfrak{P}(\Omega)$ ist

$$\mathfrak{R}(\mathcal{E}) = \bigcap_{\substack{\mathcal{R} \text{ ist Ring auf } \Omega \\ \mathcal{E} \subset \mathcal{R}}} \mathcal{R}$$

der vom Mengensystem $\mathcal{E}$ erzeugte Ring auf Ω.

Lemma 1.25

Ist $\mathcal{E} \subset \mathfrak{P}(\Omega)$ ein Semi-Ring, so ist

$$\mathfrak{R}(\mathcal{E}) = \left\{\sum_{j=1}^{n} E_j \mid E_1,\ldots,E_n \in \mathcal{E} \text{ paarweise disjunkt, } n \in \mathbb{N}\right\}.$$

Beweis

Übungsaufgabe! □

Definition 1.26

Eine Abbildung $\mu : \mathcal{E} \to [0,\infty]$ auf einem Mengensystem $\mathcal{E} \subset \mathfrak{P}(\Omega)$ mit $\varnothing \in \mathcal{E}$ heißt **Mengenfunktion**.

Eine **Mengenfunktion** μ heißt **σ-additiv**, wenn $\mu(\varnothing) = 0$ ist und für jede Folge von paarweise disjunkten Mengen $E_n \in \mathcal{E}$, $n \in \mathbb{N}$, mit $\sum_{n=1}^{\infty} E_n \in \mathcal{E}$ gilt

$$\mu\left(\sum_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \mu(E_n).$$

Eine Mengenfunktion μ heißt **endlich additiv**, wenn $\mu(\varnothing) = 0$ und für je endlich viele paarweise disjunkte Mengen $E_1,\ldots,E_n \in \mathcal{E}$, $n \in \mathbb{N}$, mit $\sum_{j=1}^{n} E_j \in \mathcal{E}$ gilt

$$\mu\left(\sum_{j=1}^{n} E_j\right) = \sum_{j=1}^{n} \mu(E_j).$$

Ist $\mathcal{E}$ eine σ-Algebra, so heißt μ ein **Maß**. Mit ∞ soll in der üblichen Weise gerechnet werden. Nicht generell üblich ist hier die zweckmäßige Fortsetzung $0 \cdot \infty = \infty \cdot 0 = 0$.

Eine Mengenfunktion μ heißt **σ-endlich**, wenn es eine Folge von Mengen $E_n \in \mathcal{E}$, $n \in \mathbb{N}$, gibt mit $\mu(E_n) < \infty$ für jedes $n \in \mathbb{N}$ und $\Omega = \bigcup_{n \in \mathbb{N}} E_n$.

Für eine endlich additive bzw. σ-additive Mengenfunktion μ auf einem Ring $\mathcal{R}$ gelten die folgenden, wie für Wahrscheinlichkeitsmaße leicht zu beweisenden Eigenschaften:

(i) Aus $A,B \in \mathcal{R}$, $A \subset B$ folgt $\mu(A) \le \mu(B)$ (Isotonie),

(ii) Aus $A,B \in \mathcal{R}$, $A \subset B$, $\mu(B) < \infty$ folgt $\mu(B \cap A^c) = \mu(B) - \mu(A)$ (Subtraktivität),

(iii) Aus $A_1,\ldots,A_n \in \mathcal{R}$ folgt $\mu\left(\bigcup_{j=1}^{n} A_j\right) \le \sum_{j=1}^{n} \mu(A_j)$ (Subadditivität),

(iv) Aus μ σ-additiv, $A_n \in \mathcal{R}$, $n \in \mathbb{N}$, $\bigcup_{n=1}^{\infty} A_n \in \mathcal{R}$ folgt $\mu\left(\bigcup_{n=1}^{\infty} A_n\right) \le \sum_{n=1}^{\infty} \mu(A_n)$ (Sub-σ-Additivität),

(v) Aus μ σ-additiv, $A_n \in \mathcal{R}$, $n \in \mathbb{N}$, $A_n \uparrow A \in \mathcal{R}$ folgt $\lim_{n\to\infty} \mu(A_n) = \mu(A)$ (Stetigkeit von unten),

(vi) Aus μ σ-additiv, $A_n \in \mathcal{R}$, $n \in \mathbb{N}$, $\mu(A_1) < \infty$, $A_n \downarrow A \in \mathcal{R}$ folgt $\lim_{n\to\infty} \mu(A_n) = \mu(A)$ (Stetigkeit von oben).

Für eine endlich additive Mengenfunktion μ auf einem Ring $\mathcal{R}$ gilt darüber hinaus:

(vii) Aus $A_n \in \mathcal{R}$, $n \in \mathbb{N}$, paarweise disjunkt, $\sum_{n=1}^{\infty} A_n \in \mathcal{R}$ folgt $\mu\left(\sum_{n=1}^{\infty} A_n\right) \geq \sum_{n=1}^{\infty} \mu(A_n)$.
Dies folgt aus

$$\mu\left(\sum\nolimits_{n=1}^{\infty} A_n\right) \geq \mu\left(\sum\nolimits_{n=1}^{m} A_n\right) = \sum_{n=1}^{m} \mu(A_n) \text{ für jedes } m \in \mathbb{N}.$$

Lemma 1.27

Ist $\mathcal{S} \subset \mathfrak{P}(\Omega)$ ein Semi-Ring und $\alpha : \mathcal{S} \to [0,\infty]$ σ-additiv (endlich additiv), so lässt sich α auf genau eine Weise, nämlich durch die Festsetzung

$$\mu(E) = \sum_{j=1}^{n} \alpha(E_j) \text{ mit } E = \sum_{j=1}^{n} E_j, \; E_1,\ldots,E_n \in \mathcal{S} \text{ paarweise disjunkt,}$$

zu einer σ-additiven (endlich additiven) Mengenfunktion auf

$$\mathfrak{R}(\mathcal{S}) = \left\{\sum\nolimits_{j=1}^{n} E_j \mid E_1,\ldots,E_n \in \mathcal{S} \text{ paarweise disjunkt, } n \in \mathbb{N}\right\}$$

fortsetzen.

Beweis

Die Mengenfunktion μ ist sinnvoll definiert. Aus $E = \sum_{i=1}^{m} D_i = \sum_{j=1}^{n} E_j$ mit paarweise disjunkten $D_1,\ldots,D_m \in \mathcal{S}$ und paarweise disjunkten $E_1,\ldots,E_n \in \mathcal{S}$ folgt nämlich

$$\sum_{i=1}^{m} \alpha(D_i) = \sum_{i=1}^{m}\sum_{j=1}^{n} \alpha\underbrace{(D_i \cap E_j)}_{\in \mathcal{S}} = \sum_{j=1}^{n}\sum_{i=1}^{m} \alpha(E_j \cap D_i) = \sum_{j=1}^{n} \alpha(E_j).$$

Für paarweise disjunkte $A_n = \sum_{j=1}^{m_n} E_{nj}$ mit für jedes $n \in \mathbb{N}$ paarweise disjunkten $E_{n1},\ldots,E_{nm_n} \in \mathcal{S}$, $m_n \in \mathbb{N}$, und $A = \sum_{n=1}^{\infty} A_n = \sum_{i=1}^{k} D_i$ mit paarweise disjunkten $D_1,\ldots,D_k \in \mathcal{S}$ ist

$$D_i = \sum_{n=1}^{\infty}\sum_{j=1}^{m_n} \underbrace{(D_i \cap E_{nj})}_{\in \mathcal{S}}, \; i = 1,\ldots,k,$$

und

$$E_{nj} = \sum_{i=1}^{k} (E_{nj} \cap D_i), \; j = 1,\ldots,m_n, \; n \in \mathbb{N},$$

und daher

$$\begin{aligned}\sum_{n=1}^{\infty} \mu(A_n) &= \sum_{n=1}^{\infty}\sum_{j=1}^{m_n} \alpha(E_{nj}) = \sum_{n=1}^{\infty}\sum_{j=1}^{m_n}\sum_{i=1}^{k} \alpha(E_{nj} \cap D_i) = \sum_{i=1}^{k}\sum_{n=1}^{\infty}\sum_{j=1}^{m_n} \alpha(D_i \cap E_{nj}) \\ &= \sum_{i=1}^{k} \alpha(D_i) = \mu(A).\end{aligned}$$

Im Falle von endlich additiven Mengenfunktionen verläuft die Argumentation analog. □

Satz 1.28 (Eindeutigkeitssatz für Maße)

Es sei $\mathcal{E} \subset \mathfrak{P}(\Omega)$ ein $\cap$-stabiles Mengensystem auf der nichtleeren Menge Ω. Es seien μ_1, μ_2 Maße auf $\sigma(\mathcal{E})$ mit der Eigenschaft

$$\mu_1(E) = \mu_2(E) \text{ für jedes } E \in \mathcal{E}.$$

Es gebe eine Folge von Mengen $E_n \in \mathcal{E}$, $n \in \mathbb{N}$, mit $\Omega = \bigcup_{n\in\mathbb{N}} E_n$ und

$$\mu_1(E_n) = \mu_2(E_n) < \infty \text{ für jedes } n \in \mathbb{N}.$$

Dann ist $\mu_1 = \mu_2$.

Beweis

Für $E \in \mathcal{E}$ mit $\mu_1(E), \mu_2(E) < \infty$ sei

$$\mathcal{D}_E = \{A \in \sigma(\mathcal{E}) \mid \mu_1(A \cap E) = \mu_2(A \cap E)\}.$$

$\mathcal{D}_E$ ist ein Dynkin-System, welches $\mathcal{E}$ enthält. Also ist $\sigma(\mathcal{E}) \subset \mathcal{D}_E$, d.h. es ist $\mu_1(A \cap E) = \mu_2(A \cap E)$ für jedes $A \in \sigma(\mathcal{E})$. Angewendet auf die Mengen E_n, $n \in \mathbb{N}$, impliziert dies für jedes $A \in \sigma(\mathcal{E})$

$$\begin{aligned}\mu_1(A) &= \mu_1\Big(A \cap \bigcup\nolimits_{n=1}^{\infty} E_n\Big) = \mu_1\Big(A \cap \Big(E_1 + \sum\nolimits_{n=2}^{\infty}(E_n \cap E_{n-1}^c \cap \ldots \cap E_1^c)\Big)\Big) \\ &= \mu_1\Big(A \cap E_1 + \sum\nolimits_{n=2}^{\infty}(A \cap E_n \cap E_{n-1}^c \cap \ldots \cap E_1^c)\Big) \\ &= \mu_1(A \cap E_1) + \sum_{n=2}^{\infty} \mu_1(A \cap E_n \cap E_{n-1}^c \cap \ldots \cap E_1^c) \\ &= \mu_2(A \cap E_1) + \sum_{n=2}^{\infty} \mu_2(A \cap E_n \cap E_{n-1}^c \cap \ldots \cap E_1^c) \\ &= \mu_2\Big(A \cap E_1 + \sum\nolimits_{n=2}^{\infty}(A \cap E_n \cap E_{n-1}^c \cap \ldots \cap E_1^c)\Big) \\ &= \mu_2\Big(A \cap \bigcup\nolimits_{n=1}^{\infty} E_n\Big) = \mu_2(A).\end{aligned}$$

□

Satz 1.29 (Maßerweiterungssatz, C. Caratheodory)
Es sei $\mathcal{R}$ ein Ring über der Menge Ω und $\mu : \mathcal{R} \to [0,\infty]$ eine σ-additive Mengenfunktion. Dann gibt es ein Maß $\tilde{\mu}$ auf $\sigma(\mathcal{R})$ mit der Eigenschaft, dass $\tilde{\mu}|_{\mathcal{R}} = \mu$ ist. $\tilde{\mu}$ ist eindeutig bestimmt, wenn μ σ-endlich ist.

Beweis
Wir führen den Beweis in sechs Schritten.

(i) Es sei $\overline{\mu} : \mathfrak{P}(\Omega) \to \overline{\mathbb{R}}$ definiert durch

$$\overline{\mu}(T) = \inf\Big\{\sum\nolimits_{n=1}^{\infty} \mu(A_n) \mid T \subset \bigcup\nolimits_{n=1}^{\infty} A_n, A_n \in \mathcal{R}, n \in \mathbb{N}\Big\},\ T \subset \Omega,$$

wobei $\inf \varnothing := \infty$ ist. Dann hat $\overline{\mu}$ die Eigenschaften

(1) $0 \le \overline{\mu}(T) \le \infty$ für alle $T \subset \Omega$,
(2) $\overline{\mu}(\varnothing) = 0$,
(3) $\overline{\mu}(T_1) \le \overline{\mu}(T_2)$ für $T_1 \subset T_2 \subset \Omega$,
(4) $\overline{\mu}\Big(\bigcup\nolimits_{n=1}^{\infty} T_n\Big) \le \sum\nolimits_{n=1}^{\infty} \overline{\mu}(T_n)$ für alle Folgen $(T_n)_{n\in\mathbb{N}}$ von Teilmengen von Ω.

(Diese Eigenschaften besagen, dass $\overline{\mu}$ ein **äußeres Maß** ist.) Die Eigenschaften (1)-(3) sind klar. Die Eigenschaft (4) ergibt sich so:
Wegen der Eigenschaft (3) nehmen wir o.B.d.A. $\overline{\mu}(T_n) < \infty$ für alle $n \in \mathbb{N}$ an. Dann gibt es zu einem beliebigen $\varepsilon > 0$ für jedes $n \in \mathbb{N}$ eine Folge $(A_{nm})_{m=1}^{\infty}$ von Mengen $A_{nm} \in \mathcal{R}$ mit der Eigenschaft, dass

$$T_n \subset \bigcup\nolimits_{m=1}^{\infty} A_{nm} \ \text{ und } \ \overline{\mu}(T_n) \ge \sum_{m=1}^{\infty} \mu(A_{nm}) - \varepsilon 2^{-n}$$

ist. Dies impliziert

$$\bigcup_{n=1}^{\infty} T_n \subset \bigcup_{n=1}^{\infty}\bigcup_{m=1}^{\infty} A_{nm} \ \text{ und } \ \overline{\mu}\Big(\bigcup\nolimits_{n=1}^{\infty} T_n\Big) \le \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} \mu(A_{nm}) \le \sum_{n=1}^{\infty} \overline{\mu}(T_n) + \varepsilon.$$

Da $\varepsilon > 0$ beliebig war, folgt $\overline{\mu}\Big(\bigcup\nolimits_{n=1}^{\infty} T_n\Big) \le \sum\limits_{n=1}^{\infty} \overline{\mu}(T_n)$.

(ii) Es ist $\overline{\mu}|_{\mathcal{R}} = \mu$. Denn:
Sei $A \in \mathcal{R}$. Offenbar ist $\overline{\mu}(A) \le \mu(A)$. Andererseits folgt aus $A \subset \bigcup\nolimits_{n=1}^{\infty} A_n$ mit Mengen $A_n \in \mathcal{R}$ bekanntlich $\mu(A) \le \sum\nolimits_{n=1}^{\infty} \mu(A_n)$ und damit $\mu(A) \le \overline{\mu}(A)$.

(iii) Es ist

$$\mathcal{M} := \{A \in \mathfrak{P}(\Omega) \mid \text{Für alle } T \subset \Omega \text{ gilt } \overline{\mu}(T) = \overline{\mu}(T \cap A) + \overline{\mu}(T \cap A^c)\}$$

eine σ-Algebra über Ω und $\overline{\mu}|_{\mathcal{M}}$ ein Maß.
Offenbar ist $\Omega \in \mathcal{M}$ und mit $A \in \mathcal{M}$ auch $A^c \in \mathcal{M}$. Mit $A, B \in \mathcal{M}$ ist auch $A \cap B \in \mathcal{M}$, denn für $T \subset \Omega$ ist

$$\begin{aligned}\overline{\mu}(T) &= \overline{\mu}(T \cap B) + \overline{\mu}(T \cap B^c), && \text{da } B \in \mathcal{M}, \\ &= \overline{\mu}(T \cap B \cap A) + \overline{\mu}(T \cap B \cap A^c) + \overline{\mu}(T \cap B^c), && \text{da } A \in \mathcal{M}, \\ &= \overline{\mu}(T \cap (B \cap A)) + \overline{\mu}(T \cap (B \cap A)^c), && \text{da } B \in \mathcal{M}.\end{aligned}$$

Also ist $\mathcal{M}$ eine σ-Algebra. Sind $A, B \in \mathcal{M}$ disjunkt, so gilt für $T \subset \Omega$, dass

$$\overline{\mu}(T \cap (A \cup B)) = \overline{\mu}(T \cap A \cap (A \cup B)) + \overline{\mu}(T \cap A^c \cap (A \cup B)) = \overline{\mu}(T \cap A) + \overline{\mu}(T \cap B)$$

ist. Mittels vollständiger Induktion ergibt sich für disjunkte Mengen $A_1, \ldots, A_m \in \mathcal{M}$:

$$\overline{\mu}\Big(T \cap \bigcup\nolimits_{n=1}^{m} A_n\Big) = \sum_{n=1}^{m} \overline{\mu}(T \cap A_n) \ \text{ für alle } T \subset \Omega$$

und damit insbesondere, dass $\overline{\mu}|_{\mathcal{M}}$ endlich additiv ist. Ist $(A_n)_{n\in\mathbb{N}}$ eine Folge von paarweise disjunkten Mengen in $\mathcal{M}$ und $A=\bigcup_{n=1}^{\infty}A_n$, so ist $\overline{\mu}(A)\geq\sum_{n=1}^{\infty}\overline{\mu}(A_n)$ und damit wegen der Eigenschaft (i)(4) auch $\overline{\mu}(A)=\sum_{n=1}^{\infty}\overline{\mu}(A_n)$. Folglich ist $\overline{\mu}|_{\mathcal{M}}$ σ-additiv. Es bleibt nur noch zu zeigen, dass auch $A\in\mathcal{M}$ zutrifft. Es ist $B_m=\bigcup_{n=1}^{m}A_n\in\mathcal{M}$ und $A^c\subset B_m^c$ für jedes $m\in\mathbb{N}$, so dass für jedes $T\subset\Omega$

$$\overline{\mu}(T)=\overline{\mu}(T\cap B_m)+\overline{\mu}(T\cap B_m^c)\geq\overline{\mu}(T\cap A^c)+\sum_{n=1}^{m}\overline{\mu}(T\cap A_n)\ \text{ für } m\in\mathbb{N}$$

und damit

$$\overline{\mu}(T)\geq\sum_{n=1}^{\infty}\overline{\mu}(T\cap A_n)+\overline{\mu}(T\cap A^c)\geq\overline{\mu}(T\cap A)+\overline{\mu}(T\cap A^c)$$

gilt. Da wegen der Eigenschaft (i)(4)

$$\overline{\mu}(T)\leq\overline{\mu}(T\cap A)+\overline{\mu}(T\cap A^c) \qquad \circledast$$

zutrifft, ist $\overline{\mu}(T)=\overline{\mu}(T\cap A)+\overline{\mu}(T\cap A^c)$, also $A\in\mathcal{M}$.

(iv) Es ist $\mathcal{R}\subset\mathcal{M}$. Denn:
Seien $A\in\mathcal{R}$ und $T\subset\Omega$. Wegen der Gültigkeit der Ungleichung $\circledast$ ist nur die Ungleichung

$$\overline{\mu}(T)\geq\overline{\mu}(T\cap A)+\overline{\mu}(T\cap A^c)$$

zu zeigen. Ist $(A_n)_{n\in\mathbb{N}}$ eine Folge von Mengen in $\mathcal{R}$ mit $T\subset\bigcup_{n=1}^{\infty}A_n$, so gilt

$$\sum_{n=1}^{\infty}\mu(A_n)=\sum_{n=1}^{\infty}\mu(A_n\cap A)+\sum_{n=1}^{\infty}\mu(A_n\cap A^c)$$

und daher wegen $T\cap A^c\subset\bigcup_{n=1}^{\infty}(A_n\cap A^c)$ und $T\cap A\subset\bigcup_{n=1}^{\infty}(A_n\cap A)$ die Ungleichung

$$\sum_{n=1}^{\infty}\mu(A_n)\geq\overline{\mu}(T\cap A)+\overline{\mu}(T\cap A^c).$$

Da die Folge $(A_n)_{n\in\mathbb{N}}$ mit $A_n\in\mathcal{R}$ und $T\subset\bigcup_{n=1}^{\infty}A_n$ beliebig wählbar ist, folgt die gesuchte Aussage

$$\overline{\mu}(T)\geq\overline{\mu}(T\cap A)+\overline{\mu}(T\cap A^c).$$

(v) Es ist $\mathcal{R}\subset\sigma(\mathcal{R})\subset\mathcal{M}$ und $\tilde{\mu}:=\overline{\mu}|_{\sigma(\mathcal{R})}$ eine Fortsetzung von μ zu einem Maß auf $\sigma(\mathcal{R})$.

(vi) Die Eindeutigkeit der Fortsetzung bei σ-endlichem μ folgt aus dem Eindeutigkeitssatz für Maße. □

Satz 1.30

Es sei $F:\mathbb{R}\to\mathbb{R}$ monoton wachsend und rechtsseitig stetig. Dann gibt es genau ein Maß $\mu_F:\mathcal{B}\to[0,\infty]$ mit der Eigenschaft, dass

$$\mu_F((a,b])=F(b)-F(a)\ \text{ für alle } -\infty<a\leq b<\infty$$

ist.

Beweis

Auf dem Semi-Ring

$$\mathcal{S}=\{(a,b]\mid-\infty<a\leq b<\infty\}$$

wird durch $\mu_F((a,b])=F(b)-F(a)$, $-\infty<a\leq b<\infty$ offenbar eine additive Mengenfunktion definiert. Diese Mengenfunktion lässt sich zu einer endlich additiven – auch wieder mit μ_F bezeichneten – Mengenfunktion auf dem von $\mathcal{S}$ erzeugten Ring fortsetzen. Es genügt zu zeigen, dass μ_F auf $\mathcal{S}$ σ-additiv ist. Dazu sei $(a,b]=\sum_{j=1}^{\infty}(a_j,b_j]$ mit $-\infty<a<b<\infty$, $-\infty<a_j\leq b_j<\infty$, $j\in\mathbb{N}$. Wegen der rechtsseitigen Stetigkeit von F gibt es zu jedem $\varepsilon>0$ ein $0<\delta<b-a$ und ein $\delta_j>0$, $j\in\mathbb{N}$, mit $0\leq F(a+\delta)-F(a)\leq\frac{\varepsilon}{2}$ und $0\leq F(b_j+\delta_j)-F(b_j)\leq\varepsilon 2^{-(j+1)}$, $j\in\mathbb{N}$. Also gilt

$$\mu_F((a,b])\leq\mu_F((a+\delta,b])+\frac{\varepsilon}{2}$$

und

$$\mu_F\big((a_j,b_j+\delta_j]\big)\leq\mu_F\big((a_j,b_j]\big)+\varepsilon 2^{-(j+1)},\ j\in\mathbb{N}.$$

Wegen

$$[a+\delta,b]\subset\bigcup_{j=1}^{\infty}(a_j,b_j+\delta_j)$$

und der Kompaktheit von $[a+\delta,b]$ existieren endlich viele Intervalle $(a_{j_k},b_{j_k}+\delta_{j_k})$, $k=1,\dots,n$, mit

$$[a+\delta,b]\subset\bigcup_{k=1}^{n}(a_{j_k},b_{j_k}+\delta_{j_k}).$$

Da μ_F isoton und subadditiv ist, folgt

$$\mu((a+\delta,b]) \le \mu_F\left(\bigcup\nolimits_{k=1}^{n}(a_{j_k},b_{j_k}+\delta_{j_k})\right) \le \sum_{k=1}^{n}\mu_F\left((a_{j_k},b_{j_k}+\delta_{j_k}]\right) \le \sum_{k=1}^{n}\mu_F\left((a_{j_k},b_{j_k}]\right)+\frac{\varepsilon}{2}.$$

Damit haben wir

$$\mu_F((a,b]) \le \mu_F((a+\delta,b])+\frac{\varepsilon}{2} \le \sum_{j=1}^{\infty}\mu_F\left((a_j,b_j]\right)+\varepsilon \text{ für jedes } \varepsilon>0,$$

also $\mu_F((a,b]) \le \sum_{j=1}^{\infty}\mu_F\left((a_j,b_j]\right)$. Die Ungleichung $\mu_F((a,b]) \ge \sum_{j=1}^{\infty}\mu_F\left((a_j,b_j]\right)$ trifft zu, da μ_F endlich additiv ist. □

Beispiel 1.31

Sei $F(x)=x$. Das zu diesem F gehörige Maß μ_F ist das **Lebesgue-Borelsche Maß** auf der Borelschen σ-Algebra $\mathcal{B}$ von $\mathbb{R}$. Also: $\mu_F((a,b])=b-a$. Wir schreiben $\boldsymbol{\lambda}$ für das Lebesgue-Borelsche Maß auf $\mathcal{B}$.

Sei $F:\mathbb{R}\to\mathbb{R}$ mit folgenden Eigenschaften:

(P1) F ist monoton wachsend und rechtsseitig stetig,

(P2) $\lim\limits_{x\to\infty}F(x)=1$ und $\lim\limits_{x\to-\infty}F(x)=0$.

Dann gibt es nach Satz 1.30 genau ein Wahrscheinlichkeitsmaß P_F mit

$$P_F((a,b])=F(b)-F(a) \text{ für alle } -\infty<a\le b<\infty.$$

Für $a\to-\infty$ folgt:

$$P_F((-\infty,b])=F(b) \text{ für alle } b\in\mathbb{R}.$$

$b\to\infty$ liefert

$$P_F(\underbrace{(-\infty,\infty)}_{=\mathbb{R}})=\lim_{b\to\infty}P_F((-\infty,b])=\lim_{b\to\infty}F(b)=1.$$

Beispiel 1.32

Sei

$$F(x)=\int_{-\infty}^{x}f(t)\,dt \text{ mit } \int_{-\infty}^{\infty}f(t)\,dt=1.$$

Ist umgekehrt P ein Wahrscheinlichkeitsmaß auf $\mathcal{B}$, so hat die durch $F(x)=P((-\infty,x])$, $x\in\mathbb{R}$, definierte Funktion $F:\mathbb{R}\to\mathbb{R}$ die Eigenschaften (P1) und (P2).

Beispiel 1.33

Sei $f:\mathbb{R}\to\mathbb{R}$, $f\ge 0$ und (uneigentlich) Riemann-integrierbar mit $\int_{-\infty}^{\infty}f(t)\,dt=1$. Dann hat die Funktion $F:\mathbb{R}\to\mathbb{R}$, $F(x)=\int_{-\infty}^{x}f(t)\,dt$, $x\in\mathbb{R}$, genau die oben genannten Eigenschaften (i) und (ii).

Das zu diesem F gehörige Wahrscheinlichkeitsmaß auf $\mathcal{B}$ heißt **Wahrscheinlichkeitsmaß mit der Wahrscheinlichkeitsdichte** oder **Dichte** f. Spezialfälle sind:

(i) $f(t)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{t^2}{2}\right)$, $t\in\mathbb{R}$. Das zu diesem f gehörige Wahrscheinlichkeitsmaß auf $\mathcal{B}$ heißt **Standardnormalverteilung**.

Es sei $\Phi(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{t^2}{2}\right)dt$, $x\in\mathbb{R}$. Es ist $\Phi(x)=1-\Phi(-x)$ für $x\in\mathbb{R}$.

(ii) Sei für $-\infty<a<b<\infty$

$$f(t)=\begin{cases}\frac{1}{b-a}, & t\in[a,b],\\ 0, & \text{sonst.}\end{cases}$$

Dann gilt:

$$F(x)=\int_{-\infty}^{x}f(t)\,dt=\begin{cases}\frac{x-a}{b-a}, & a\le x\le b,\\ 0, & x<a,\\ 1, & x>b.\end{cases}$$

Das zu diesem F gehörige Wahrscheinlichkeitsmaß P auf $\mathcal{B}$ heißt **Rechteckverteilung** oder **Gleichverteilung** auf $[a,b]$. Für $a\le\alpha\le\beta\le b$ gilt:

$$P((\alpha,\beta])=P((-\infty,\beta])-P((-\infty,\alpha])=\frac{\beta-a}{b-a}-\frac{\alpha-a}{b-a}=\frac{\beta-\alpha}{b-a}.$$

Im Fall $a = 0$, $b = 1$ ist $P((\alpha,\beta]) = \beta - \alpha$.

(iii) Sei

$$f(t) = \begin{cases} 0, & t < 0, \\ \lambda \exp(-\lambda t), & t \geq 0, \end{cases}$$

wobei $\lambda > 0$ ein gegebener Parameter ist. Dann ist

$$F(x) = \int_{-\infty}^{x} \lambda \exp(-\lambda t)\, dt = \begin{cases} 0, & x < 0, \\ 1 - \exp(-\lambda x), & x \geq 0. \end{cases}$$

Das zu F gehörige Wahrscheinlichkeitsmaß P auf $\mathcal{B}$ heißt **Exponentialverteilung** mit dem Parameter $\lambda > 0$:

$$P((t,\infty)) = 1 - P((-\infty,t]) = 1 - F(t) = \exp(-\lambda t).$$

Beispiel 1.34 (Weitere Beispiele für Maße und Wahrscheinlichkeitsräume)

(i) Sei $\Omega \neq \varnothing$, $\mathcal{A}$ σ-Algebra auf Ω, $a \in \Omega$. Dann ist $\delta_a : \mathcal{A} \to [0,1]$ definiert durch

$$\delta_a(A) = \begin{cases} 1, & a \in A, \\ 0, & a \notin A, \end{cases} \quad \text{für } A \in \mathcal{A}$$

ein Wahrscheinlichkeitsmaß auf $\mathcal{A}$. δ_a heißt **Einpunktmaß** oder **Dirac-Maß** im Punkt $a \in \Omega$.

(ii) Sei μ_n, $n \in \mathbb{N}$, eine Folge von Maßen auf einer σ-Algebra $\mathcal{A}$ auf Ω. Sei α_n, $n \in \mathbb{N}$, eine Folge von nichtnegativen Zahlen auf $\overline{\mathbb{R}}$. Dann ist $\mu = \sum_{n\in\mathbb{N}} \alpha_n \mu_n$ ein Maß auf $\mathcal{A}$. Denn es gilt:

$$\mu(A) = \sum_{n\in\mathbb{N}} \underbrace{\alpha_n \mu_n(A)}_{\geq 0} \geq 0 \text{ für jedes } A \in \mathcal{A} \text{ und } \mu(\varnothing) = \sum_{n\in\mathbb{N}} \alpha_n \underbrace{\mu_n(\varnothing)}_{=0} = 0.$$

(iii) Sei A_j, $j \in \mathbb{N}$, eine Folge von paarweise disjunkten Mengen mit $A_j \in \mathcal{A}$. Dann gilt:

$$\begin{aligned} \mu\Big(\sum_{j\in\mathbb{N}} A_j\Big) &= \sum_{n\in\mathbb{N}} \alpha_n \mu_n\Big(\sum_{j\in\mathbb{N}} A_j\Big) = \sum_{n\in\mathbb{N}} \alpha_n \sum_{j\in\mathbb{N}} \mu_n(A_j) \\ &= \sum_{n\in\mathbb{N}} \sum_{j\in\mathbb{N}} \underbrace{\alpha_n \mu_n(A_j)}_{\geq 0} \\ &= \sum_{j\in\mathbb{N}} \sum_{n\in\mathbb{N}} \alpha_n \mu_n(A_j) \text{ (für nichtnegative Zahlen kann man die Summen vertauschen)} \\ &= \sum_{j\in\mathbb{N}} \mu(A_j). \end{aligned}$$

Sei speziell $\mu_n = \delta_{a_n}$, $n \in \mathbb{N}$, mit $a_n \in \Omega$. Dann ist $\mu = \sum_{n\in\mathbb{N}} \alpha_n \delta_{a_n}$ ein Maß. Sei $\sum_{n\in\mathbb{N}} \alpha_n = 1$. Dann ist $\mu = \sum_{n\in\mathbb{N}} \alpha_n \delta_{a_n}$ ein Wahrscheinlichkeitsmaß. Sei $\{\omega\} \in \mathcal{A}$ für jedes $\omega \in \Omega$ und sei $A = \{a_n \in \Omega \mid n \in \mathbb{N}\}$ mit paarweise verschiedenen a_n. Dann gilt:

$$\mu(A) = \sum_{n\in\mathbb{N}} \alpha_n \underbrace{\delta_{a_n}(A)}_{=1} = 1, \; \mu(A^c) = 1 - \mu(A) = 0.$$

Damit gilt für jedes $B \in \mathcal{A}$: $0 \leq \mu(A^c \cap B) \leq \mu(A^c) = 0$. Also ist $\mu(A^c \cap B) = 0$. Es folgt:

$$\mu(B) = \mu(A \cap B) + \underbrace{\mu(A^c \cap B)}_{=0} = \mu(A \cap B).$$

Ferner gilt:

$$\mu(\{\omega\}) = \mu(A \cap \{\omega\}) = \begin{cases} 0, & \omega \notin A, \\ \underbrace{\sum_{n\in\mathbb{N}} \alpha_n \delta_{a_n}(\{\omega\})}_{=\alpha_{n_0}}, & \omega \in A, \text{ etwa } \omega = a_{n_0}. \end{cases}$$

Also ist $\mu(\{\omega\}) = \alpha_{n_0}$, falls $\omega = a_{n_0}$ ist. Es gilt:

$$\mu(B) = \mu(A \cap B) = \mu\Big(\sum_{\substack{n\in\mathbb{N}, \\ a_n \in A, a_n \in B}} \{a_n\}\Big) = \sum_{\substack{n\in\mathbb{N}, \\ a_n \in A, a_n \in B}} \underbrace{\mu(\{a_n\})}_{=\alpha_n} = \sum_{\substack{n\in\mathbb{N}, \\ a_n \in A, a_n \in B}} \alpha_n.$$

Ein solches Maß μ heißt **diskretes Wahrscheinlichkeitsmaß**.

Ist also $\mathcal{A}$ eine σ-Algebra auf Ω mit $\{\omega\} \in \mathcal{A}$ für jedes $\omega \in \Omega$, so heißt ein Wahrscheinlichkeitsmaß μ auf $\mathcal{A}$ diskret, wenn es eine abzählbare Menge $A \subset \Omega$ gibt, so dass $\mu = \sum_{a\in A} \mu(\{a\}) \delta_a$ ist.
Ist etwa $\Omega = \mathbb{R}$, $\mathcal{A} = \mathcal{B}$ (insbesondere ist $\{x\} \in \mathcal{B}$ für $x \in \mathbb{R}$) und $A = \{a_1, a_2, \ldots\}$ mit paarweise verschiedenen $a_n \in \mathbb{N}$, so gilt für das Wahrscheinlichkeitsmaß einer Folge p_n, $n \in \mathbb{N}$, $p_n \geq 0$, $\sum_{n\in\mathbb{N}} p_n = 1$:

$$P = \sum_{n\in\mathbb{N}} p_n \delta_{a_n}, \; F(x) = P((-\infty,x]) = \sum_{n\in\mathbb{N}} p_n \underbrace{\delta_{a_n}((-\infty,x])}_{=\begin{cases} 1, & a_n \leq x, \\ 0, & \text{sonst} \end{cases}}.$$

Sei $\mathcal{B}^d$ die Borelsche σ-Algebra auf $\mathbb{R}^d$ mit $\mu_F((a,b]) = F(b) - F(a)$ und sei μ ein endliches Maß auf $\mathcal{B}^d$, d.h. $\mu(\mathbb{R}^d) < \infty$. Sei $F : \mathbb{R}^d \to \mathbb{R}$ definiert durch $F(x_1,...,x_d) = \mu((-\infty, x_1] \times ... \times (-\infty, x_d])$ für $(x_1,...,x_d) \in \mathbb{R}^d$ mit folgenden Eigenschaften:

(B1) $\lim\limits_{\varepsilon_i \downarrow 0, i=1,...,d} F(x_1 + \varepsilon_1,..., x_d + \varepsilon_d) = F(x_1,...,x_d)$ für jedes $(x_1,...,x_d) \in \mathbb{R}^d$ (F stetig von oben),

(B2) Für alle $-\infty < a_i < b_i < \infty$, $i = 1,...,d$, mit $S = (a_1,b_1] \times ... \times (a_d,b_d]$ gilt

$$\Delta_S F := \sum_{(\varepsilon_1,...,\varepsilon_d) \in \{0,1\}^d} (-1)^{\varepsilon_1 + ... + \varepsilon_d} F(\varepsilon_1 a_1 + (1-\varepsilon_1) b_1, ..., \varepsilon_d a_d + (1 - \varepsilon_d) b_d) \geq 0 .$$

Analog zum eindimensionalen Fall zeigt man:

Satz 1.35

Ist $F : \mathbb{R}^d \to \mathbb{R}$ mit den Eigenschaften (B1) und (B2), so existiert genau ein Maß μ_F auf $\mathcal{B}^d$ mit der Eigenschaft, dass $\mu_F(S) = \Delta_S F$ für jedes $S = (a_1,b_1] \times ... \times (a_d,b_d]$ mit $-\infty < a_i < b_i < \infty$, $i = 1,...,d$, gilt.

Das zu $F : \mathbb{R}^d \to \mathbb{R}$ mit $F(x_1,...,x_d) = x_1 \cdot ... \cdot x_d$, $(x_1,...,x_d) \in \mathbb{R}^d$, gehörige Maß μ_F auf $\mathcal{B}^d$ heißt das **Lebesgue-Borelsche Maß** auf $\mathcal{B}^d$. Es wird mit $\boldsymbol{\lambda}^d$ bezeichnet. Es gilt:

$$\boldsymbol{\lambda}^d((a_1,b_1] \times ... \times (a_d,b_d]) = \prod_{i=1}^{d} (b_i - a_i) \text{ (Volumen des Quaders } (a_1,b_1] \times ... \times (a_d,b_d]).$$

Ist $F : \mathbb{R}^d \to \mathbb{R}$ mit den Eigenschaften (B1) und (B2) und gilt

(B3) $\lim\limits_{x_{i_0} \to -\infty} F(x_1,...,x_{i_0},...,x_d) = 0$ für jedes $i_0 \in \{1,...,d\}$, $x_1,...,x_d \in \mathbb{R}$,

(B4) $\lim\limits_{\substack{x_1 \to \infty \\ \vdots \\ x_d \to \infty}} F(x_1,...,x_d) = 1$,

so existiert genau ein Wahrscheinlichkeitsmaß P_F auf $\mathcal{B}^d$ mit $F(x_1,...,x_d) = P_F((-\infty,x_1] \times ... \times (-\infty,x_d])$ für alle $(x_1,...,x_d) \in \mathbb{R}^d$. Umgekehrt hat für jedes Wahrscheinlichkeitsmaß P auf $\mathcal{B}^d$ die durch $F(x_1,...,x_d) = P((-\infty,x_1] \times ... \times (-\infty,x_d])$, $(x_1,...,x_d) \in \mathbb{R}^d$, definierte Funktion $F : \mathbb{R}^d \to \mathbb{R}$ die Eigenschaften (B1)-(B4).

Beispiel 1.36

Sei $f : \mathbb{R}^d \to \mathbb{R}$, $f \geq 0$, integrierbar mit

$$\int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} f(t_1,...,t_d)\, dt_1 ... dt_d = 1 .$$

Dann hat $F : \mathbb{R}^d \to \mathbb{R}$ mit

$$F(x_1,...,x_d) = \int_{-\infty}^{x_d} ... \int_{-\infty}^{x_1} f(t_1,...,t_d)\, dt_1 ... dt_d \;,\; (x_1,...,x_d) \in \mathbb{R}^d ,$$

die Eigenschaften (B1)-(B4). Das zu F gehörige Wahrscheinlichkeitsmaß heißt **Wahrscheinlichkeitsmaß mit der Dichte** f:

$$f(t_1,...,t_d) = \prod_{k=1}^{d} f(t_k) .$$

Kapitel 2

Bedingte Wahrscheinlichkeiten

Sei $\Omega \neq \varnothing$ endlich, $(\Omega, \mathcal{A}, P)$ ein diskreter Laplacescher Wahrscheinlichkeitsraum mit $\mathcal{A} = \mathfrak{P}(\Omega)$. Sei $A, B \subset \Omega$ mit $P(B) = \frac{|B|}{|\Omega|} > 0$. Dann ist
$$\frac{|A \cap B|}{|B|} = \frac{P(A \cap B)}{P(B)}$$
die bedingte Wahrscheinlichkeit für das Eintreten von A unter der Bedingung, dass B eintritt.

Definition 2.1

Sei $(\Omega, \mathcal{A}, P)$ ein diskreter Laplacescher Wahrscheinlichkeitsraum, $A, B \subset \Omega$, $P(B) > 0$. Dann heißt
$$P(A \mid B) := \frac{P(A \cap B)}{P(B)}$$
die **bedingte Wahrscheinlichkeit** für das Eintreten von A unter der Bedingung, dass B eintritt.

Satz 2.2

Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und I eine abzählbare Indexmenge. Sei weiter $A, A_i \in \mathcal{A}$, $i \in I$. Die A_i seien paarweise disjunkt und es sei $P(A_i) > 0$ für alle $i \in I$. Ferner sei $A \subset \sum_{i \in I} A_i$. Dann gilt:

(i) $P(A) = \sum_{i \in I} P(A \mid A_i) P(A_i)$ („**Formel von der totalen Wahrscheinlichkeit**"),

(ii) Ist $P(A) > 0$, $j \in I$, so gilt:
$$P(A_j \mid A) = \frac{P(A \mid A_j) P(A_j)}{\sum_{i \in I} P(A \mid A_i) P(A_i)} \quad \text{(„\textbf{Formel von Bayes}")}.$$

Beweis

(i) Aus $A = \sum_{i \in I} (A_i \cap A)$ folgt
$$P(A) = \sum_{i \in I} P(A_i \cap A) = \sum_{i \in I} P(A \mid A_i) P(A_i).$$

(ii) Nach (i) gilt:
$$P(A_j \mid A) = \frac{P(A_j \cap A)}{P(A)} = \frac{P(A \mid A_j) P(A_j)}{\sum_{i \in I} P(A \mid A_i) P(A_i)}.$$
□

Beispiel 2.3

Spieler und Losverkäufer vereinbaren folgendes Spiel: Lostrommel enthält a Gewinne und b Nieten mit $a + b \geq 3$. Spieler darf zwischen zwei Strategien wählen:

(i) Spieler zieht Los. Das Los ist ein Gewinn oder eine Niete. Das Spiel ist beendet.

(ii) Spieler zieht ein Los und wirft es unbesehen weg. Daraufhin entfernt der Losverkäufer eine Niete aus der Trommel. Spieler zieht erneut ein Los. Das Los ist ein Gewinn oder eine Niete. Das Spiel ist beendet.

Sei A das Ereignis, dass der Spieler beim ersten Zug ein Gewinnlos zieht und B sei das Ereignis, dass der Spieler beim zweiten Zug eine Niete zieht.

(i) $P(A) = \dfrac{a}{a+b}$,

(ii) Da $B = B \cap A + B \cap A^c$, erhalten wir:
$$\begin{aligned} P(B) &= P(B \mid A) P(A) + P(B \mid A^c) P(A^c) = \frac{a-1}{a+b-2} \frac{a}{a+b} + \frac{a}{a+b-2} \frac{b}{a+b} \\ &= \frac{a}{a+b} \frac{a+b-1}{a+b-2} > P(A). \end{aligned}$$

Satz 2.4

Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, seien $A_0, \ldots, A_n \in \mathcal{A}$, $n \geq 1$, $P(A_0 \cap \ldots \cap A_{n-1}) > 0$. Dann gilt:
$$P(A_0 \cap \ldots \cap A_n) = P(A_0) P(A_1 \mid A_0) \cdot \ldots \cdot P(A_n \mid A_0 \cap \ldots \cap A_{n-1}) \quad \text{(„\textbf{Multiplikationsformel}")}.$$

Beweis

Es gilt:

$$P(A_0)\cdot\ldots\cdot P(A_n \mid A_0\cap\ldots\cap A_{n-1}) = \cancel{P(A_0)}\,\frac{\cancel{P(A_1\cap A_0)}}{\cancel{P(A_0)}}\,\frac{\cancel{P(A_2\cap A_0\cap A_1)}}{\cancel{P(A_0\cap A_1)}}\cdot\ldots\cdot\frac{P(A_0\cap\ldots\cap A_n)}{\cancel{P(A_0\cap\ldots\cap A_{n-1})}}$$
$$= P(A_0\cap\ldots\cap A_n).$$

□

Beispiel 2.5

Eine Urne enthält r rote und s schwarze Kugeln ($a := r+s$). Es wird n-mal je eine Kugel gezogen. Diese Kugel wird zusammen mit c Kugeln derselben Farbe in die Urne zurückgelegt. Gesucht wird die Wahrscheinlichkeit, in diesen n Ziehungen genau k rote Kugeln zu ziehen. Seien:

$$0 \triangleq \text{schwarze Kugel}\,,\ 1 \triangleq \text{rote Kugel}\,,$$

$$A_j^{(0)} \triangleq \text{im } j\text{-ten Zug wird eine schwarze Kugel gezogen}\,,\ A_j^{(1)} \triangleq \text{im } j\text{-ten Zug wird eine rote Kugel gezogen}\,.$$

Dann gilt wegen der Unabhängigkeit der einzelnen Ziehungen für $\omega_i \in \{0,1\}$

$$P(A_1^{(\omega_1)}\cap\ldots\cap A_n^{(\omega_n)}) = \underbrace{P(A_1^{(\omega_1)})}_{=\begin{cases} s/a, & \omega_1=0,\\ r/a, & \omega_1=1\end{cases}}\, P(A_2^{(\omega_2)}\mid A_1^{(\omega_1)})\cdot\ldots\cdot P(A_n^{(\omega_n)}\mid A_1^{(\omega_1)}\cap\ldots\cap A_{n-1}^{(\omega_{n-1})})\,,$$

wobei

$$P\big(A_j^{(0)}\mid A_1^{(\omega_1)}\cap\ldots\cap A_{j-1}^{(\omega_{j-1})}\big) = \frac{s+((j-1)-(\omega_1+\ldots+\omega_{j-1}))c}{a+(j-1)c}$$

und

$$P\big(A_j^{(1)}\mid A_1^{(\omega_1)}\cap\ldots\cap A_{j-1}^{(\omega_{j-1})}\big) = \frac{r+(\omega_1+\ldots+\omega_{j-1})c}{a+(j-1)c}$$

für $2\le j\le n$ ist. Nach $(j-1)$ Ziehungen haben wir dann $a+(j-1)c$ Kugeln und $r+(\omega_1+\ldots+\omega_{j-1})c$ rote Kugeln in der Urne beim Eintreten des Ereignisses $A_1^{(\omega_1)}\cap\ldots\cap A_{j-1}^{(\omega_{j-1})}$. Folglich haben wir dann

$$a+(j-1)c-(r+(\omega_1+\ldots+\omega_{j-1})c) = s+((j-1)-(\omega_1+\ldots+\omega_{j-1}))c$$

schwarze Kugeln in der Urne. Also gilt für $\omega_1+\ldots+\omega_{j-1}=k$, $0\le k\le j-1$:

$$P\big(A_j^{(0)}\mid A_1^{(\omega_1)}\cap\ldots\cap A_{j-1}^{(\omega_{j-1})}\big) = \frac{s+((j-1)-k)c}{a+(j-1)c}\,,$$
$$P\big(A_j^{(1)}\mid A_1^{(\omega_1)}\cap\ldots\cap A_{j-1}^{(\omega_{j-1})}\big) = \frac{r+kc}{a+(j-1)c}\,.$$

Daraus folgt für $\omega_1+\ldots+\omega_n=k$, $0\le k\le n$:

$$P(A_1^{(\omega_1)}\cap\ldots\cap A_n^{(\omega_n)}) = \frac{\prod_{j=1}^{k}(r+(j-1)c)\prod_{j=1}^{n-k}(s+(j-1)c)}{\prod_{j=1}^{n}(a+(j-1)c)} =: p_k\,.$$

Sei A das Ereignis, genau k rote Kugeln in n Ziehungen zu erhalten, dann gilt:

$$P(A) = P\Big(\sum\nolimits_{\substack{(\omega_1,\ldots,\omega_n)\in\{0,1\}^n,\\ \omega_1+\ldots+\omega_n=k}} A_1^{(\omega_1)}\cap\ldots\cap A_n^{(\omega_n)}\Big)$$
$$= \sum_{\substack{(\omega_1,\ldots,\omega_n)\in\{0,1\}^n,\\ \omega_1+\ldots+\omega_n=k}} p_k = \binom{n}{k}p_k \ \text{(Polyarchisches Urnenmodell).}$$

Kapitel 3

Messbare Funktionen und Bildmaße, Zufallsvariablen und ihre Verteilungen

Messbare Funktionen und Bildmaße

Definition 3.1

Sei Ω eine nichtleere Menge, $\mathcal{A}$ eine σ-Algebra über Ω. Das Paar $(\Omega,\mathcal{A})$ heißt **Messraum**.
Seien $(\Omega,\mathcal{A})$ und $(R,\mathcal{S})$ Messräume. Eine Abbildung $f:\Omega\to R$ heißt $(\mathcal{A},\mathcal{S})$**-messbar**, wenn $f^{-1}(B)\in\mathcal{A}$ für jedes $B\in\mathcal{S}$ zutrifft. Wir schreiben in diesem Fall $f:(\Omega,\mathcal{A})\to(R,\mathcal{S})$.
Ist $\mathcal{S}=\sigma(\mathcal{E})$ für ein Mengensystem $\mathcal{E}\subset\mathfrak{P}(R)$, so ist eine Abbildung $f:\Omega\to R$ genau dann $(\mathcal{A},\mathcal{S})$-messbar, wenn $f^{-1}(E)\in\mathcal{A}$ für jedes $E\in\mathcal{E}$ zutrifft. Es ist nämlich $\{B\in\mathcal{S}\mid f^{-1}(B)\in\mathcal{A}\}$ eine σ-Algebra auf R, die das Mengensystem $\mathcal{E}$ enthält.
Hieraus folgt, dass jede stetige Funktion $f:\mathbb{R}^d\to\mathbb{R}^p$ $(\mathcal{B}^d,\mathcal{B}^p)$-messbar ist.
Im Folgenden betrachten wir Funktionen mit Werten in $\overline{\mathbb{R}}$. Mit

$$\overline{\mathcal{B}}=\mathcal{B}\cup\{B\cup\{\infty\}\mid B\in\mathcal{B}\}\cup\{B\cup\{-\infty\}\mid B\in\mathcal{B}\}\cup\{B\cup\{\pm\infty\}\mid B\in\mathcal{B}\}$$

liegt eine σ-Algebra auf $\overline{\mathbb{R}}$ mit $\overline{\mathcal{B}}_{\mathbb{R}}:=\{B\cap\mathbb{R}\mid B\in\overline{\mathcal{B}}\}=\mathcal{B}$ vor. Ferner ist

$$\overline{\mathcal{B}}=\sigma(\{(a,\infty]\mid a\in\mathbb{R}\})=\sigma(\{[a,\infty]\mid a\in\mathbb{R}\})=\sigma(\{[-\infty,b)\mid b\in\mathbb{R}\})=\sigma(\{[-\infty,b]\mid b\in\mathbb{R}\}),$$

woraus unmittelbar folgt, dass eine Funktion $f:\Omega\to\overline{\mathbb{R}}$ genau dann $(\mathcal{A},\overline{\mathcal{B}})$-messbar ist, wenn eine der folgenden äquivalenten Bedingungen erfüllt ist:

$$\begin{aligned}
\{f>a\}&:=\{\omega\in\Omega\mid f(\omega)>a\}=f^{-1}((a,\infty])\in\mathcal{A}\ \text{für jedes}\ a\in\mathbb{R},\\
\{f\geq a\}&:=\{\omega\in\Omega\mid f(\omega)\geq a\}=f^{-1}([a,\infty])\in\mathcal{A}\ \text{für jedes}\ a\in\mathbb{R},\\
\{f<b\}&:=\{\omega\in\Omega\mid f(\omega)<b\}=f^{-1}([-\infty,b))\in\mathcal{A}\ \text{für jedes}\ b\in\mathbb{R},\\
\{f\leq b\}&:=\{\omega\in\Omega\mid f(\omega)\leq b\}=f^{-1}([-\infty,b])\in\mathcal{A}\ \text{für jedes}\ b\in\mathbb{R}.
\end{aligned}$$

Schreibweisen wie $\{f>a\}$ oder $\{f\in B\}(=\{\omega\in\Omega\mid f(\omega)\in B\})$ für das Urbild einer Menge $B\subset R$ unter einer Abbildung $f:\Omega\to R$ sind in der Maß- und Wahrscheinlichkeitstheorie üblich und werden in der Folge häufig benutzt. Funktionen $f:(\Omega,\mathcal{A})\to(\overline{\mathbb{R}},\overline{\mathcal{B}})$ nennen wir ab jetzt kurz messbar.

Lemma 3.2

Seien f_n, $n\in\mathbb{N}$, messbare Funktionen auf Ω. Dann sind auch die Funktionen $\sup_{n\in\mathbb{N}}f_n$ und $\inf_{n\in\mathbb{N}}f_n$ messbar.
Beweis
Es ist

$$\{\sup_{n\in\mathbb{N}}f_n\leq a\}=\bigcap_{n\in\mathbb{N}}\{f_n\leq a\}\ \text{für jedes}\ a\in\mathbb{R}$$

und

$$\{\inf_{n\in\mathbb{N}}f_n\geq a\}=\bigcap_{n\in\mathbb{N}}\{f_n\geq a\}\ \text{für jedes}\ a\in\mathbb{R}.$$

Als Folgerung erhalten wir, dass mit messbaren Funktionen f_n, $n\in\mathbb{N}$, auch $\liminf_{n\to\infty}f_n=\sup_{n\in\mathbb{N}}\inf_{k\geq n}f_k$ und $\limsup_{n\to\infty}f_n=\inf_{n\in\mathbb{N}}\sup_{k\geq n}f_k$ messbar sind.
Ist f messbar und $\alpha\in\overline{\mathbb{R}}$, so ist auch αf messbar. Ferner ist $f=f^+-f^-$ mit den nicht negativen messbaren Funktionen $f^+:=\max\{f,0\}$ und $f^-:=\max\{-f,0\}$. □

Beispiel 3.3

(i) Für eine Menge $A\subset\Omega$ heißt die Funktion $I_A:\Omega\to\mathbb{R}$ mit

$$I_A(\omega)=\begin{cases}1, & \omega\in A,\\ 0, & \omega\notin A\end{cases}$$

der **Indikator von** A oder auch die **Indikatorfunktion von** A . Anstelle von I_A schreiben wir auch $I(A)$, im Falle $A = \{f \in B\}$ für eine Funktion $f : \Omega \to R$ und eine Menge $B \subset R$ auch $I(f \in B)$. Eine Indikatorfunktion I_A ist genau dann messbar, wenn $A \in \mathcal{A}$ ist.

(ii) Eine reelle Funktion $f : \Omega \to \mathbb{R}$, die nur endlich viele Werte, etwa die paarweise verschiedenen Werte $\alpha_1,\dots,\alpha_n$ annimmt, hat die Darstellung $f = \sum_{i=1}^n \alpha_i I(A_i)$ mit den paarweise disjunkten Mengen $A_i = \{f = \alpha_i\}$, $i = 1,\dots,n$, deren Vereinigung Ω ist. Eine solche Funktion ist genau dann messbar, wenn $A_i \in \mathcal{A}$ für jedes $i = 1,\dots,n$ gilt. Reelle nichtnegative messbare Funktionen, die nur endlich viele Werte annehmen, nennen wir **primitive** Funktionen. Jede Funktion $f = \sum_{i=1}^m \beta_i I(B_i)$ mit beliebigen nichtnegativen reellen Zahlen $\beta_i \in \mathbb{R}$ und Mengen $B_i \in \mathcal{A}$, $i = 1,\dots,m$, ist eine primitive Funktion. Eine primitive Funktion hat stets eine **Normaldarstellung** der Form $f = \sum_{i=1}^n \alpha_i I(A_i)$ mit nichtnegativen reellen Zahlen α_i und paarweise disjunkten Mengen $A_i \in \mathcal{A}$, $i = 1,\dots,n$, deren Vereinigung Ω ist. Wir bezeichnen die primitiven Funktionen mit $\mathcal{P}$. Wir nennen eine Folge von nichtnegativen Funktionen auf $f_n : \Omega \to \overline{\mathbb{R}}$, $n \in \mathbb{N}$, **isoton konvergent** gegen die nichtnegative Funktion $f : \Omega \to \overline{\mathbb{R}}$, wenn $f_n \le f_{n+1}$ für jedes $n \in \mathbb{N}$ und $f = \sup_{n\in\mathbb{N}} f_n$ ist. Wir drücken dies durch die Schreibweise $f_n \uparrow f$ aus.

Satz 3.4

Sei $f \ge 0$ messbar. Dann gibt es eine Folge von Funktionen $f_n \in \mathcal{P}$, $n \in \mathbb{N}$, mit $f_n \uparrow f$.

Beweis

Für $n \in \mathbb{N}$ sind

$$A_{in} = \begin{cases} \{i2^{-n} \le f < (i+1)2^{-n}\}, & i = 0,\dots,n2^n - 1, \\ \{f \ge n\}, & i = n2^n, \end{cases}$$

paarweise disjunkte Mengen aus $\mathcal{A}$. Die Folge der Funktionen $f_n = \sum_{i=0}^{n2^n} i2^{-n} I_{A_{in}}$, $n \in \mathbb{N}$, leisten das Verlangte. □

Korollar 3.5

Sind f,g messbar, so sind auch $f+g$, $f-g$, $f \cdot g$ und f/g (sofern definiert in dem Sinne, dass $f(\omega) \pm g(\omega)$ niemals von der Form $\infty - \infty$ oder $-\infty + \infty$ und $f(\omega)/g(\omega)$ niemals von der Form $\pm\infty/\pm\infty$ oder $a/0$ ist) messbar.

Beweis

Übungsaufgabe. □

Definition 3.6

Ist μ ein Maß auf der σ -Algebra $\mathcal{A}$ auf Ω, $(R,\mathcal{S})$ ein Messraum und $f : (\Omega,\mathcal{A}) \to (R,\mathcal{S})$, so wird durch $\mu^f(B) := \mu(f^{-1}(B))$, $B \in \mathcal{S}$, ein Maß auf $\mathcal{S}$ definiert. μ^f heißt **Bildmaß** von μ unter f .

Sei $(\Omega,\mathcal{A})$ ein Messraum, d.h. $\Omega \ne \varnothing$, $\mathcal{A}$ σ -Algebra über Ω. Sei weiter $(R,\mathcal{S})$ ein Messraum. Ist μ ein Maß auf $\mathcal{A}$ und ist $f : \Omega \to R$ messbar, so gilt für paarweise disjunkte $B_1, B_2, \dots \in \mathcal{S}$ mit $\sum_{n=1}^\infty B_n \in \mathcal{S}$:

$$\mu^f\left(\sum_{n=1}^\infty B_n\right) = \mu\left(f^{-1}\left(\sum_{n=1}^\infty B_n\right)\right) = \mu\left(\sum_{n=1}^\infty f^{-1}(B_n)\right) = \sum_{n=1}^\infty \mu(f^{-1}(B_n)) = \sum_{n=1}^\infty \mu^f(B_n).$$

Definition 3.7

Sei $(\Omega,\mathcal{A},P)$ ein Wahrscheinlichkeitsraum, $(R,\mathcal{S})$ ein Messraum. Dann heißt $X : (\Omega,\mathcal{A}) \to (R,\mathcal{S})$ eine **Zufallsvariable**. Das Bildmaß P^X ist ein Wahrscheinlichkeitsmaß auf $(R,\mathcal{S})$ und heißt die **Verteilung** von X .

Es sei $\mathcal{E} \subset \mathfrak{P}(R)$, $\mathcal{S} = \sigma(\mathcal{E})$ (die von f auf R erzeugte σ -Algebra). Sei $f : \Omega \to R$, $f^{-1}(E) \in \mathcal{A}$ für jedes $E \in \mathcal{E}$. Dann ist $f : (\Omega,\mathcal{A}) \to (R,\mathcal{S})$.
Sei $f : \mathbb{R}^d \to \mathbb{R}^p$ stetig. Dann ist $f^{-1}(U) \subset \mathbb{R}^d$ offen für jedes offene $U \subset \mathbb{R}^p$. Also ist $f : (\mathbb{R}^d,\mathcal{B}^d) \to (\mathbb{R}^p,\mathcal{B}^p)$.

Beispiel 3.8

(i) Sei $\mathcal{B}$ die Borelsche σ -Algebra auf $\mathbb{R}$. Dann ist $I_A : (\Omega,\mathcal{A}) \to (\mathbb{R},\mathcal{B})$ genau dann, wenn $A \in \mathcal{A}$ ist.

(ii) Sei $f : \Omega \to \overline{\mathbb{R}}$, f nehme genau die endlich vielen Werte $a_1,\dots,a_m$ an. Dann gilt:

$$f = \sum_{j=1}^m a_i I_{\{f=a_i\}}.$$

Es gilt zudem $\Omega = \sum_{i=1}^{m} \{f = a_i\} = \sum_{i=1}^{m} f^{-1}(\{a_i\})$ und es ist $f : (\Omega, \mathcal{A}) \to (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ genau dann, wenn $\{f = a_i\} \in \mathcal{A}$ für alle $i = 1,\dots,m$.

Beispiel 3.9 (Zufallsvariablen und ihre Verteilungen)

Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, $(R, \mathcal{S})$ ein Messraum, $X : (\Omega, \mathcal{A}) \to (R, \mathcal{S})$ $(X^{-1}(B) \in \mathcal{A}$ für alle $B \in \mathcal{S})$. Es sei $\{x\} \in \mathcal{S}$ für jedes $x \in R$ und es gebe eine abzählbare Menge $A \subset R$ mit

$$P(X \in A) = P(X^{-1}(A)) = P(\{X \in A\}) = 1 .$$

Wir schreiben auch $P^X(A)$. Dann gilt $P^X(A^c) = 1 - P^X(A) = 0$. Damit erhalten wir

$$P^X = \sum_{a \in A} P(X = a)\delta_a ,$$

wobei $P(X = a) = P(\{X = a\}) = P(X^{-1}(\{a\})) = P^X(\{a\})$ ist. Für jede Menge $B \in \mathcal{S}$ gilt dann:

$$\begin{aligned} P^X(B) &= P^X(A \cap B) + \underbrace{P^X(A^c \cap B)}_{=0} = P^X(A \cap B) = P^X\Big(\sum\nolimits_{a \in A \cap B} \{a\}\Big) = \sum_{a \in A \cap B} P^X(\{a\}) \\ &= \sum_{a \in A \cap B} P(X = a) = \sum_{a \in A} P(X = a)\delta_a(B) = \Big(\sum\nolimits_{a \in A} P(X = a)\delta_a\Big)(B). \end{aligned}$$

In diesem Fall ist die Verteilung P^X eindeutig festgelegt durch die **Zähldichte** $x \mapsto P(X = x)$, $x \in \mathbb{R}$.
Die Zufallsvariable X heißt in diesem Fall **diskret verteilt**. P^X ist die **diskrete Verteilung** von X.

Eine Urne enthält r rote und s schwarze Kugeln, $a := r + s$. Es wird n-mal ohne Zurücklegen je eine Kugel gezogen, $n \le a$. Wie groß ist die Wahrscheinlichkeit, dass in den n Ziehungen genau k rote Kugeln gezogen werden?
Wir nummerieren die roten Kugeln mit den Zahlen $1,\dots,r$ und die schwarzen mit $r+1,\dots,a$.
Sei $\Omega = \{(\omega_1,\dots,\omega_n) \in \{1,\dots,a\}^n \mid \omega_i \text{ paarweise verschieden}\}$ und $\mathcal{A} = \mathfrak{P}(\Omega)$.
Wir nehmen an, P ist die **diskrete Laplace-Verteilung** auf Ω:

$$P(A) = \frac{|A|}{|\Omega|} \quad \text{für } A \subset \Omega$$

und $X : \Omega \to \mathbb{R}$ mit

$$X(\omega_1,\dots,\omega_n) = \big|\{j \in \{1,\dots,n\} \mid \omega_j \in \{1,\dots,r\}\}\big| \quad \text{für } (\omega_1,\dots,\omega_n) \in \Omega .$$

Wir suchen $P(X = k)$ (Zähldichte hier: $k \mapsto P(X = k)$, $k \in \{0,\dots,n\}$). Es gilt:

$$\begin{aligned} |\{X = k\}| &= \big|\{(\omega_1,\dots,\omega_n) \in \Omega \mid X(\omega_1,\dots,\omega_n) = k\}\big| \\ &= \big|\{(\omega_1,\dots,\omega_n) \in \Omega \mid \big|\{j \in \{1,\dots,n\} \mid \omega_j \in \{1,\dots,r\}\}\big| = k\}\big| \\ &= \binom{n}{k} r(r-1)\cdot\ldots\cdot(r-k+1)s(s-1)\cdot\ldots\cdot(s-(n-k)+1). \end{aligned}$$

Es folgt für die Wahrscheinlichkeit, in den n Ziehungen genau k rote Kugeln zu erhalten:

$$\begin{aligned} P(X = k) = \frac{|\{X = k\}|}{|\Omega|} &= \binom{n}{k} \frac{r(r-1)\cdot\ldots\cdot(r-k+1)s(s-1)\cdot\ldots\cdot(s-(n-k)+1)}{a(a-1)\cdot\ldots\cdot(a-n+1)} \\ &= \cancel{\binom{n}{k}} \frac{\binom{r}{k}\binom{s}{n-k}}{\binom{a}{n}} \cancel{\frac{k!(n-k)!}{n!}} = \frac{\binom{r}{k}\binom{a-r}{n-k}}{\binom{a}{n}} \end{aligned}$$

mit $k \in \mathbb{N}_0$, $\max\{0, \underbrace{n + r - a}_{=n-s}\} \le k \le \min\{r, n\}$.

Definition 3.10

Die Verteilung einer reellen Zufallsvariable X mit der Zähldichte

$$P(X = k) = \frac{\binom{r}{k}\binom{a-r}{n-k}}{\binom{a}{n}} \quad \text{mit } k \in \mathbb{N}_0,\ \max\{0, n + r - a\} \le k \le \min\{r, n\},$$

heißt **hypergeometrische Verteilung** mit den Parametern $a, r, n \in \mathbb{N}$, $r, n \le a$.
Schreibweise: $X \sim \mathcal{H}(a, r, n)$.

Eine Anwendung dieser Verteilung ist das Zahlenlotto „6 aus 49“ mit $a = 49$, $r = 6$, $n = 6$.
Ist X die Anzahl der richtig getippten Zahlen auf einem Tippschein, so gilt:

k	0	1	2	3	4	5	6
$P(X = k)$	0,43596	0,41302	0,13238	$0,1765 \cdot 10^{-1}$	$0,9686 \cdot 10^{-3}$	$0,1845 \cdot 10^{-4}$	$0,7151 \cdot 10^{-7}$

Eine Urne enthalte r rote und s schwarze Kugeln, $a := r + s$. Seien die Kugeln nummeriert wie oben. Es wird n-mal mit Zurücklegen je eine Kugel aus der Urne gezogen. Wie groß ist die Wahrscheinlichkeit, genau k rote Kugeln zu ziehen?
Offensichtlich gilt:

$$k \in \{0,\dots,n\}\,,\ \Omega = \{1,\dots,a\}^n\,.$$

Wir nehmen an, P ist das diskrete Laplacesche Wahrscheinlichkeitsmaß auf Ω, $X : \Omega \to \mathbb{R}$ die Anzahl der gezogenen roten Kugeln mit

$$X(\omega_1,\dots,\omega_n) = \left|\{j \in \{1,\dots,n\} \mid \omega_j \in \{1,\dots,r\}\}\right|,\ (\omega_1,\dots,\omega_n) \in \Omega\,.$$

Dann gilt:

$$P(X = k) = \frac{|\{X = k\}|}{|\Omega|} = \binom{n}{k}\frac{r^k(a-r)^{n-k}}{a^n} = \binom{n}{k}\left(\frac{r}{a}\right)^k\left(1-\frac{r}{a}\right)^{n-k}.$$

Definition 3.11

Die Verteilung einer reellen Zufallsvariable X mit der Zähldichte

$$P(X = k) = \binom{n}{k}\left(\frac{r}{a}\right)^k\left(1-\frac{r}{a}\right)^{n-k} \text{ mit } k \in \{0,\dots,n\}\,,$$

heißt **Binomialverteilung** mit den Parametern $n \in \mathbb{N}$ und

$$p = \frac{r}{a} \in [0,1] \text{ mit } r,a \in \mathbb{N}_0\,,\ 0 \le r \le a \ne 0\,.$$

Schreibweise: $X \sim \mathcal{B}(n,p)$.

Kapitel 4

Maßintegrale, Erwartungswerte von Zufallsvariablen

Das μ-Integral

Sei $(\Omega, \mathcal{A})$ ein Messraum, μ ein Maß auf $\mathcal{A}$. Für eine Funktion $f \in \mathcal{P}$, der Menge der reellen nichtnegativen primitiven Funktionen in Normaldarstellung $f = \sum_{j=1}^{n} \alpha_j I_{A_j}$ mit reellen nichtnegativen Zahlen $\alpha_1, \ldots, \alpha_n$ und paarweise disjunkten $A_1, \ldots, A_n \in \mathcal{A}$ mit $\sum_{j=1}^{n} A_j = \Omega$, definieren wir das μ**-Integral** von f durch

$$\int f\, d\mu = \sum_{j=1}^{n} \alpha_j \mu(A_j)\,.$$

Die Definition ist sinnvoll, denn für eine andere Normaldarstellung von f, etwa $f = \sum_{i=1}^{m} \beta_i I_{B_i}$ mit reellen, nichtnegativen Zahlen $\beta_1, \ldots, \beta_m$ und paarweise disjunkten Mengen $B_1, \ldots, B_m \in \mathcal{A}$ mit $\sum_{i=1}^{m} B_i = \Omega$ ist $\alpha_j = \beta_i$, falls $A_j \cap B_i \neq \varnothing$ ist, also

$$\begin{aligned}\sum_{j=1}^{n} \alpha_j \mu(A_j) &= \sum_{j=1}^{n} \alpha_j \mu\Big(\sum\nolimits_{i=1}^{m} (A_j \cap B_i)\Big) = \sum_{j=1}^{n}\sum_{i=1}^{m} \alpha_j \mu(A_j \cap B_i) \\ &= \sum_{i=1}^{m}\sum_{j=1}^{n} \beta_i \mu(B_i \cap A_j) = \sum_{i=1}^{m} \beta_i \mu\Big(\sum\nolimits_{j=1}^{n} (B_i \cap A_j)\Big) = \sum_{i=1}^{m} \beta_i \mu(B_i).\end{aligned}$$

Einige elementare Eigenschaften des μ-Integrals sind unmittelbar klar:

$$\int I_A\, d\mu = \mu(A) \text{ für jedes } A \in \mathcal{A}\,, \tag{4.1}$$

$$\int \alpha f\, d\mu = \alpha \int f\, d\mu \text{ für } f \in \mathcal{P} \text{ und } \alpha \in \mathbb{R}^+\,, \tag{4.2}$$

$$\int (f+g)\, d\mu = \int f\, d\mu + \int g\, d\mu \text{ für } f, g \in \mathcal{P}\,, \tag{4.3}$$

$$\int f\, d\mu \leq \int g\, d\mu \text{ für } f, g \in \mathcal{P}\,,\ f \leq g\,. \tag{4.4}$$

Aus den Eigenschaften (4.1)-(4.3) ergibt sich, dass im Falle $f = \sum_{j=1}^{n} \alpha_j I_{A_j} \in \mathcal{P}$ mit reellen nichtnegativen Zahlen $\alpha_1, \ldots, \alpha_n$ und beliebigen $A_1, \ldots, A_n \in \mathcal{A}$ gilt $\int f\, d\mu = \sum_{j=1}^{n} \alpha_j \mu(A_j)$.

Lemma 4.1

Seien $f, f_n \in \mathcal{P}$, $n \in \mathbb{N}$, mit $f_n \leq f_{n+1}$ für jedes $n \in \mathbb{N}$ und $f \leq \sup_{n\in\mathbb{N}} f_n$. Dann gilt:

$$\int f\, d\mu \leq \sup_{n\in\mathbb{N}} \int f_n\, d\mu\,.$$

Beweis

Für $\alpha \in (0,1)$ und $B_n = \{f_n \geq \alpha f\}$ ist $B_n \uparrow \Omega$ und $f_n \geq \alpha f I_{B_n}$ für jedes $n \in \mathbb{N}$, so dass für

$$f = \sum_{j=1}^{n} \alpha_j I_{A_j}$$

folgt:

$$\int f_n\, d\mu \geq \alpha \sum_{j=1}^{n} \alpha_j \mu(B_n \cap A_j)\,.$$

Der Grenzübergang $n \to \infty$ liefert

$$\sup_{n\in\mathbb{N}} \int f_n\, d\mu \geq \alpha \sum_{j=1}^{n} \alpha_j \mu(A_j) = \alpha \int f\, d\mu\,,$$

so dass mit dem anschließenden Grenzübergang $\alpha \uparrow 1$ die Behauptung folgt. □

Lemma 4.2

Seien $f_n, g_n \in \mathcal{P}$, $n \in \mathbb{N}$, mit $f_n \uparrow \sup_{n\in\mathbb{N}} f_n$, $g_n \uparrow \sup_{n\in\mathbb{N}} g_n$ und $\sup_{n\in\mathbb{N}} f_n = \sup_{n\in\mathbb{N}} g_n$. Dann gilt:

$$\sup_{n\in\mathbb{N}} \int f_n\, d\mu = \sup_{n\in\mathbb{N}} \int g_n\, d\mu\,.$$

Beweis
Aus $f_n \leq \sup_{n\in\mathbb{N}} g_n$ folgt $\int f_n \, d\mu \leq \sup_{n\in\mathbb{N}} \int g_n \, d\mu$ mit Lemma 4.1. Also ist $\sup_{n\in\mathbb{N}} \int f_n \, d\mu \leq \sup_{n\in\mathbb{N}} \int g_n \, d\mu$. Genauso ergibt sich $\sup_{n\in\mathbb{N}} \int g_n \, d\mu \leq \sup_{n\in\mathbb{N}} \int f_n \, d\mu$. □

Es sei
$$\mathcal{M} := \mathcal{M}(\Omega,\mathcal{A}) := \{ f \mid f : (\Omega,\mathcal{A}) \to (\overline{\mathbb{R}},\overline{\mathcal{B}}) \}$$
die Menge der messbaren Funktionen auf Ω ,
$$\mathcal{M}_+ := \mathcal{M}_+(\Omega,\mathcal{A}) := \{ f \mid f : (\Omega,\mathcal{A}) \to (\overline{\mathbb{R}},\overline{\mathcal{B}}), f \geq 0 \}$$
die Menge der nichtnegativen messbaren Funktionen auf Ω . Aufgrund von Lemma 4.2 ist die folgende Definition sinnvoll:

Definition 4.3

(i) Es sei $f \in \mathcal{M}_+$ und $f_n \in \mathcal{P}$ für jedes $n \in \mathbb{N}$ mit $f_n \uparrow f$. Dann heißt
$$\int f \, d\mu = \sup_{n\in\mathbb{N}} \int f_n \, d\mu$$
das μ **-Integral** von f .

(ii) Es sei $f \in \mathcal{M}$ mit der Eigenschaft, dass $\int f^+ \, d\mu < \infty$ oder $\int f^- \, d\mu < \infty$ ist. Dann heißt f eine **Funktion mit existierendem** μ **-Integral**
$$\int f \, d\mu = \int f^+ \, d\mu - \int f^- \, d\mu \, .$$

(iii) Es sei $f \in \mathcal{M}$ mit der Eigenschaft, dass $\int f^+ \, d\mu < \infty$ und $\int f^- \, d\mu < \infty$ ist. Dann heißt f μ **-integrierbar** und die reelle Zahl
$$\int f \, d\mu = \int f^+ \, d\mu - \int f^- \, d\mu$$
ist das μ **-Integral** von f .

Einige elementare Eigenschaften des μ -Integrals sind klar:
$$\int \alpha f \, d\mu = \alpha \int f \, d\mu \quad \text{für } f \in \mathcal{M}_+ \text{ und } \alpha \in \overline{\mathbb{R}} \, , \ \alpha \geq 0 \, , \tag{4.5}$$
$$\int (f+g) \, d\mu = \int f \, d\mu + \int g \, d\mu \quad \text{für } f,g \in \mathcal{M}_+ \, , \tag{4.6}$$
$$\int f \, d\mu \leq \int g \, d\mu \quad \text{für } f,g \in \mathcal{M}_+ \, , \ f \leq g \, . \tag{4.7}$$
Aus den Eigenschaften (4.6) und (4.7) folgt wegen $|f| = f^+ + f^-$ für eine Funktion $f \in \mathcal{M}$, dass f genau dann μ -integrierbar ist, wenn $|f|$ μ -integrierbar ist. Allgemeiner gilt für jedes $f \in \mathcal{M}$ mit existierendem μ -Integral:
$$\left| \int f \, d\mu \right| \leq \int |f| \, d\mu \, .$$
Im Falle $f,g \in \mathcal{M}$ mit $|f| \leq |g|$ folgt aus der μ -Integrierbarkeit von g die μ -Integrierbarkeit von f . Ferner folgt, dass für reellwertige μ -integrierbare Funktionen $f,g \in \mathcal{M}$ und reelle Zahlen $\alpha,\beta \in \mathbb{R}$ auch die reellwertige Funktion $\alpha f + \beta g \in \mathcal{M}$ μ -integrierbar ist, wobei
$$\int (\alpha f + \beta g) \, d\mu = \alpha \int f \, d\mu + \beta \int g \, d\mu \tag{4.8}$$
ist. Darüber hinaus folgt für reellwertige μ -integrierbare Funktionen $f,g \in \mathcal{M}$ mit $f \leq g$:
$$\int f \, d\mu \leq \int g \, d\mu \, . \tag{4.9}$$
Die Menge
$$\mathcal{L} := \mathcal{L}(\Omega,\mathcal{A},\mu) := \{ f \in \mathcal{M}(\Omega,\mathcal{A}) \mid f \text{ reellwertig und } \mu\text{-integrierbar} \}$$
ist daher ein reeller Vektorraum. Das μ -Integral ist ein positives lineares Funktional auf $\mathcal{L}(\Omega,\mathcal{A},\mu)$.

Satz 4.4 (Satz von der monotonen Konvergenz)

Seien $f, f_n \in \mathcal{M}_+$, $n \in \mathbb{N}$, mit $f_n \uparrow f$. Dann gilt:
$$\sup_{n\in\mathbb{N}} \int f_n \, d\mu = \int f \, d\mu \, .$$

Beweis
Für jedes $n \in \mathbb{N}$ sei die Folge der reellen nichtnegativen primitiven Funktionen u_{nk} , $k \in \mathbb{N}$, isoton konvergent gegen f_n . Für jedes $k \in \mathbb{N}$ ist $v_k := \max_{1\leq n\leq k} u_{nk} \in \mathcal{P}$ und $v_k \leq v_{k+1}$. Aus
$$0 \leq u_{nk} \leq v_k \leq f_k \quad \text{für } n \leq k$$

folgt

$$0 \le f_n \le \sup_{k\in\mathbb{N}} v_k \le f \quad \text{für jedes } n \in \mathbb{N}$$

und damit $v_k \uparrow f$. Hieraus resultiert wegen

$$\int f\,d\mu = \sup_{k\in\mathbb{N}} \int v_k\,d\mu \le \sup_{k\in\mathbb{N}} \int f_k\,d\mu \le \int f\,d\mu$$

die Behauptung. □

Lemma 4.5 (Lemma von Fatou)
Seien $f_n \in \mathcal{M}_+$, $n \in \mathbb{N}$. Dann ist

$$\int \liminf_{n\to\infty} f_n\,d\mu \le \liminf_{n\to\infty} \int f_n\,d\mu\,.$$

Beweis
Mit $g_n := \inf_{k\ge n} f_k$ für $n \in \mathbb{N}$ ist $g_n \uparrow f = \liminf_{n\to\infty} f_n$, also

$$\int f\,d\mu = \sup_{n\in\mathbb{N}} \int g_n\,d\mu = \liminf_{n\to\infty} \int g_n\,d\mu \le \liminf_{n\to\infty} \int f_n\,d\mu\,.$$ □

Satz 4.6 (Satz von der majorisierten Konvergenz)
Seien $g, f_n \in \mathcal{M}$ reellwertig, $n \in \mathbb{N}$. Sei $|f_n| \le |g|$ für jedes $n \in \mathbb{N}$. Sei $f := \lim_{n\to\infty} f_n$ und g μ-integrierbar. Dann sind f, f_n, $n \in \mathbb{N}$, μ-integrierbar und

$$\lim_{n\to\infty} \int |f_n - f|\,d\mu = 0\,.$$

Beweis
Die μ-Integrierbarkeit von f und f_n ist klar. Es ist $0 \le 2g - |f_n - f|$ für jedes $n \in \mathbb{N}$, so dass mit dem Lemma von Fatou

$$\begin{aligned} 0 \le \int 2g\,d\mu &= \int \liminf_{n\to\infty} (2g - |f_n - f|)\,d\mu \\ &\le \liminf_{n\to\infty} \int (2g - |f_n - f|)\,d\mu \\ &= \int 2g\,d\mu - \limsup_{n\to\infty} \int |f_n - f|\,d\mu, \end{aligned}$$

also $\limsup_{n\to\infty} \int |f_n - f|\,d\mu = 0$ folgt. □

Bemerkung 4.7

(i) Sei $(\Omega, \mathcal{A})$ ein Messraum, μ ein Maß auf $\mathcal{A}$, $A \in \mathcal{A}$. Dann ist $\mathcal{A}_A := \{A \cap B \mid B \in \mathcal{A}\}$ eine σ-Algebra auf A. $\mathcal{A}_A$ heißt **Spur-σ-Algebra** von $\mathcal{A}$ auf A.
Sei $0 < \mu(A) < \infty$. Durch $\mu_A : \mathcal{A}_A \to [0,\infty)$ mit $\mu_A(B) = \mu(B)$, $B \in \mathcal{A}_A$, wird ein Maß auf $\mathcal{A}_A$ definiert. Dann ist

$$\frac{1}{\mu(A)}\mu_A$$

ein Wahrscheinlichkeitsmaß auf $\mathcal{A}_A$. Ist $\nu : \mathcal{A}_A \to [0,\infty]$ ein weiteres Maß, so wird durch $\tilde{\mu} : \mathcal{A} \to [0,\infty]$ mit $\tilde{\mu}(B) = \nu(A \cap B)$, $B \in \mathcal{A}$, ein Maß auf $\mathcal{A}$ definiert mit der Eigenschaft

$$\tilde{\mu}|_{\mathcal{A}_A} = \nu\,.$$

Sei speziell $(\Omega, \mathcal{A}) = (\mathbb{R}^d, \mathcal{B}^d)$, $\mu = \boldsymbol{\lambda}^d$ und sei $A \in \mathcal{B}^d$ mit $0 < \boldsymbol{\lambda}^d(A) < \infty$. Das Wahrscheinlichkeitsmaß

$$\frac{1}{\boldsymbol{\lambda}^d(A)}\boldsymbol{\lambda}^d_A$$

heißt **Gleichverteilung** auf A. Ist $A = (0,1)$ für $d = 1$, so ist $\boldsymbol{\lambda}_{(0,1)} := \boldsymbol{\lambda}^1_{(0,1)}$ ein Wahrscheinlichkeitsmaß auf $\mathcal{B}_{(0,1)} := \{(0,1) \cap B \mid B \in \mathcal{B}\}$.
Beachte: Es ist $\boldsymbol{\lambda}_{(0,1)}(\{x\}) = 0$ für jedes $x \in \mathbb{R}$, denn:

$$\{x\} = \bigcap_{n=1}^{\infty} \left(x - \tfrac{1}{n}, x\right].$$

Also folgt

$$\boldsymbol{\lambda}_{(0,1)}(\{x\}) = \lim_{n\to\infty} \underbrace{\boldsymbol{\lambda}_{(0,1)}\left(\left(x - \tfrac{1}{n}, x\right]\right)}_{=1/n} = 0\,.$$

$\boldsymbol{\lambda}_{(0,1)}$ ist die Gleichverteilung auf $(0,1)$. Sei $(\Omega,\mathcal{A},P)$ ein Wahrscheinlichkeitsraum, $X:(\Omega,\mathcal{A})\to(\mathbb{R}^d,\mathcal{B}^d)$. Dann heißt X auf A **gleichverteilt**, wenn gilt

$$P_A^X = \frac{1}{\boldsymbol{\lambda}^d(A)}\boldsymbol{\lambda}_A^d .$$

Die Zufallsvariable $X:(\Omega,\mathcal{A})\to(A,\mathcal{B}_A)$ heißt ebenfalls auf A gleichverteilt, wenn gilt:

$$P^X = \frac{1}{\boldsymbol{\lambda}^d(A)}\boldsymbol{\lambda}_A^d .$$

Speziell ist $X:(\Omega,\mathcal{A})\to(\mathbb{R},\mathcal{B})$ auf $(0,1)$ gleichverteilt, wenn $P_{(0,1)}^X = \boldsymbol{\lambda}_{(0,1)}$ ist.

(ii) Sei $(\Omega,\mathcal{A},P)$ ein Wahrscheinlichkeitsraum, $X:(\Omega,\mathcal{A})\to(\mathbb{R},\mathcal{B})$. Dann ist die Verteilung von X eindeutig festgelegt durch ihre **Verteilungsfunktion**

$$F(x) = P^X((-\infty,x]) = P\big(X^{-1}((-\infty,x])\big) = P(X\in(-\infty,x]) = P(X\le x),\ x\in\mathbb{R}.$$

Ist $X:(\Omega,\mathcal{A})\to(\mathbb{R}^d,\mathcal{B}^d)$, $X=(X_1,\dots,X_d)$, so ist die Verteilung von X eindeutig festgelegt durch ihre Verteilungsfunktion

$$\begin{aligned}F(x_1,\dots,x_d) &= P^X((-\infty,x_1]\times\dots\times(-\infty,x_d]) = P(\{X_1\le x_1\}\cap\dots\cap\{X_d\le x_d\})\\ &= P(X_1\le x_1,\dots,X_d\le x_d)\end{aligned}$$

für $(x_1,\dots,x_d)\in\mathbb{R}^d$.

Definition 4.8

Sei $(\Omega,\mathcal{A},P)$ ein Wahrscheinlichkeitsraum, $X:(\Omega,\mathcal{A})\to(\overline{\mathbb{R}},\overline{\mathcal{B}})$. Ist $\int X^+\,dP<\infty$ oder $\int X^-\,dP<\infty$, so heißt

$$E(X) = \int X\,dP = \int X^+\,dP - \int X^-\,dP$$

der **Erwartungswert** von X. X ist integrierbar, wenn $\int |X|\,dP<\infty$ gilt.

Beispiel 4.9

Sei $\mathcal{A}$ σ-Algebra. Dann gilt:

(i) Sei $A\in\mathcal{A}$, $X=I_A$. Dann ist $E(X)=P(A)$.

(ii) Seien $A_1,\dots,A_n\in\mathcal{A}$, $X=\sum_{i=1}^n I_{A_i}$. Dann gilt:

$$E(X) = E\Big(\sum_{i=1}^n I_{A_i}\Big) = \sum_{i=1}^n E(I_{A_i}) = \sum_{i=1}^n P(A_i).$$

Beispiel 4.10

Eine Urne enthält r rote und s schwarze Kugeln, $a:=r+s$. Wir nummerieren die roten Kugeln mit $1,\dots,r$ und die schwarzen mit $r+1,\dots,a$. Es wird n-mal

(1) mit Zurücklegen,

(2) ohne Zurücklegen

je eine Kugel gezogen. Die Zufallsvariable X beschreibe die Anzahl der gezogenen roten Kugeln in n Ziehungen. Gesucht wird der Erwartungswert von X.

Sei B_i das Ereignis, dass bei der i-ten Ziehung eine rote Kugel gezogen wird. Dann ist

$$X = \sum_{i=1}^n I_{B_i}$$

und es gilt:

(1) Offensichtlich ist $\Omega_1=\{1,\dots,a\}^n$. Sei $\mathcal{A}=\mathfrak{P}(\Omega_1)$. Für die B_i gilt:

$$B_i = \{(\omega_1,\dots,\omega_n)\in\Omega_1 \mid \omega_i\in\{1,\dots,r\}\},\ i=1,\dots,n.$$

P ist hier das diskrete Laplacesche Wahrscheinlichkeitsmaß und es gilt:

$$E(X) = E\Big(\sum_{i=1}^n I_{B_i}\Big) = \sum_{i=1}^n P(B_i) = \sum_{i=1}^n \frac{ra^{n-1}}{a^n} = \sum_{i=1}^n \frac{r}{a} = n\underbrace{\Big(\frac{r}{a}\Big)}_{=:p} = np.$$

Bekanntlich ist $X\sim\mathcal{B}(n,p)$. Also gilt für den Erwartungswert binomialverteilter Zufallsvariablen:

$$E(X)=np.$$

(2) Es ist $\Omega_2=\{(\omega_1,\dots,\omega_n)\in\Omega_1 \mid \omega_i \text{ paarweise verschieden}, i=1,\dots,n\}$ und $n\le a$. Sei $\mathcal{A}=\mathfrak{P}(\Omega_2)$. P ist hier wieder das diskrete Laplacesche Wahrscheinlichkeitsmaß und für die B_i gilt:

$$B_i = \{(\omega_1,\dots,\omega_n)\in\Omega_2 \mid \omega_i\in\{1,\dots,r\}\},\ i=1,\dots,n.$$

Also folgt:

$$E(X) = \sum_{i=1}^{n} P(B_i) = \sum_{i=1}^{n} \frac{r(a-1)\cdot\ldots\cdot(a-n+1)}{a(a-1)\cdot\ldots\cdot(a-n+1)} = \sum_{i=1}^{n} \frac{r}{a} = n\frac{r}{a}.$$

Bekanntlich ist $X \sim \mathcal{H}(a,r,n)$. Also gilt für den Erwartungswert hypergeometrisch verteilter Zufallsvariablen:

$$E(X) = n\frac{r}{a}.$$

Eigenschaften von μ-Integralen und Beispiele zur Berechnung von μ-Integralen

Definition 4.11

Es sei $(\Omega, \mathcal{A})$ ein Messraum, μ ein Maß auf $\mathcal{A}$. Mengen $N \in \mathcal{A}$ mit $\mu(N) = 0$ heißen μ**-Nullmengen**.
Ist E eine Eigenschaft derart, dass für jedes $\omega \in \Omega$ definiert ist, ob ω diese Eigenschaft hat oder nicht, so sagt man: **„E gilt μ-fast überall (μ-f.ü.)“**, wenn es eine μ-Nullmenge $N \in \mathcal{A}$ gibt, so dass E für jedes $\omega \in N^c$ gilt.

Satz 4.12

Es sei $f \in \mathcal{M}_+$. Genau dann ist $f = 0$ μ-f.ü., wenn $\int f\,d\mu = 0$ ist.

Beweis

Ist $f = 0$ μ-f.ü., so ist $\mu(f \neq 0) = 0$. Aus $f \leq \sup_{n\in\mathbb{N}} nI(f \neq 0)$ folgt

$$0 \leq \int f\,d\mu \leq \sup_{n\in\mathbb{N}} n \int I(f \neq 0)\,d\mu = 0.$$

Wegen

$$\int f\,d\mu = \int fI(f > 0)\,d\mu = \sup_{n\in\mathbb{N}} \int fI\Big(f > \frac{1}{n}\Big)d\mu \geq \sup_{n\in\mathbb{N}} \frac{1}{n}\mu\Big(f > \frac{1}{n}\Big)$$

folgt umgekehrt aus $0 = \int f\,d\mu$ auch $\mu(f > 0) = 0$. □

Korollar 4.13

Ist $f : (\Omega, \mathcal{A}) \to (\overline{R}, \overline{\mathcal{A}})$ und ist $N \in \mathcal{A}$ eine μ-Nullmenge, so gilt

$$\int fI_N\,d\mu = 0.$$

Beweis

Für eine Funktion $f : (\Omega, \mathcal{A}) \to (\overline{R}, \overline{\mathcal{A}})$ mit existierendem μ-Integral und Mengen $A \in \mathcal{A}$ schreiben wir

$$\int_A f\,d\mu \text{ für } \int fI_A\,d\mu.$$ □

Satz 4.14

Sei $f \in \mathcal{M}_+$. Durch

$$\nu(A) = \int_A f\,d\mu, \; A \in \mathcal{A},$$

wird ein Maß ν auf $(\Omega, \mathcal{A})$ definiert. Dann ist jede μ-Nullmenge ist eine ν-Nullmenge.

Beweis

Natürlich ist $\nu(\varnothing) = 0$. Die σ-Additivität von ν folgt aus dem Satz (4.4) von der monotonen Konvergenz. Ist $A \in \mathcal{A}$ und $\mu(A) = 0$, so ist nach Korollar 4.13 auch $\nu(A) = 0$. □

Definition 4.15

Ein Maß ν mit der in Satz 4.14 gegebenen Darstellung heißt **Maß mit der μ-Dichte** f. Wir drücken diesen Sachverhalt durch die Schreibweise $d\nu = f\,d\mu$ aus. Diese Schreibweise liegt vor allem auf Grund des folgenden Satzes nahe.

Satz 4.16

Sei $f \in \mathcal{M}_+$, $d\nu = f\,d\mu$. Für jede Funktion $g \in \mathcal{M}_+$ ist

$$\int g\,d\nu = \int gf\,d\mu. \tag{4.1}$$

Für eine Funktion $g \in \mathcal{M}$ existiert das ν-Integral von g genau dann, wenn das μ-Integral von gf existiert. In diesem Fall gilt die Identität (4.1).

Beweis

Wir verwenden das **Prinzip der algebraischen Induktion**: der Beweis der behaupteten Aussage erfolgt sukzessive für eine Indikatorfunktion g, eine nichtnegative primitive Funktion g, eine nichtnegative $(\mathcal{A},\overline{\mathcal{B}})$-messbare Funktion g und schließlich über die Zerlegung $g = g^+ - g^-$ für eine beliebige $(\mathcal{A},\overline{\mathcal{B}})$-messbare Funktion g. Für eine Indikatorfunktion g ist die Aussage unmittelbare Konsequenz aus der Definition von ν. Wegen der Linearitätseigenschaften der μ- und ν-Integrale ist sie damit auch für primitive Funktionen klar. Für $u_n \in \mathcal{P}$ mit $u_n \uparrow g \in \mathcal{M}_+$ folgt aus dem Satz (4.4) von der monotonen Konvergenz

$$\int g\, d\nu = \sup_{k\in\mathbb{N}} \int u_k f\, d\mu = \int gf\, d\mu\,.$$

Für $g \in \mathcal{M}$ liefert dieses Resultat bei Anwendung auf g^+ und g^- die behauptete Aussage. □

Satz 4.17 (Transformationssatz für Integrale)

Sei $T : (\Omega,\mathcal{A}) \to (R,\mathcal{S})$, μ ein Maß auf $(\Omega,\mathcal{A})$. Für jede Funktion $g \in \mathcal{M}_+(R,\mathcal{S})$ gilt

$$\int g \circ T\, d\mu = \int g\, d\mu^T\,. \tag{4.2}$$

Für eine Funktion $g : (R,\mathcal{S}) \to (\overline{\mathbb{R}},\overline{\mathcal{B}})$ existiert das μ-Integral von $g \circ T$ genau dann, wenn das μ^T-Integral von g existiert. In diesem Fall gilt die Identität (4.2).

Beweis

Ergibt sich sehr einfach wieder durch Anwendung des Prinzips der algebraischen Induktion. □

Beispiel 4.18

(i) Es sei $\mu = \delta_a$ das Einpunktmaß in $a \in \Omega$ und $f : (\Omega,\mathcal{A}) \to (\overline{\mathbb{R}},\overline{\mathcal{B}})$. Mit Hilfe des Prinzips der algebraischen Induktion macht man sich sofort klar, dass

$$\int f\, d\delta_a = f(a)$$

ist.

(ii) Es sei $(\Omega,\mathcal{A}) = (\mathbb{N},\mathfrak{P}(\mathbb{N}))$ und $\tau = \sum_{n=1}^\infty \delta_n$ das **Zählmaß** auf $(\Omega,\mathcal{A})$. Für jede nichtnegative Funktion $f : \mathbb{N} \to \overline{\mathbb{R}}$ folgt aus dem Satz (4.4) von der monotonen Konvergenz:

$$\int f\, d\tau = \int \sum\nolimits_{n=1}^\infty fI_{\{n\}}\, d\tau = \sum_{n=1}^\infty \int fI_{\{n\}}\, d\tau = \sum_{n=1}^\infty f(n)\tau(\{n\}) = \sum_{n=1}^\infty f(n)\,.$$

Damit ist klar, dass eine Funktion $f : \mathbb{N} \to \overline{\mathbb{R}}$ genau dann τ-integrierbar ist, wenn $\sum_{n=1}^\infty |f(n)| < \infty$ ist. Das τ-Integral von f ist in diesem Fall $\int f\, d\tau = \sum_{n=1}^\infty f(n)$.

(iii) Es sei $(\mu_n)_{n\in\mathbb{N}}$ eine Folge von Maßen μ_n auf $(\Omega,\mathcal{A})$, $(\alpha_n)_{n\in\mathbb{N}}$ eine Folge von Zahlen $\alpha_n \in [0,\infty]$ und $\mu = \sum_{n\in\mathbb{N}} \alpha_n\mu_n$. Für eine nichtnegative Funktion $f : (\Omega,\mathcal{A}) \to (\overline{\mathbb{R}},\overline{\mathcal{B}})$ ist

$$\int f\, d\mu = \sum_{n\in\mathbb{N}} \alpha_n \int f\, d\mu_n \tag{4.3}$$

Dies zeigt man wieder mit Hilfe des Prinzips der algebraischen Induktion: Für eine Indikatorfunktion f ist die Aussage unmittelbare Konsequenz aus der Definition von μ. Wegen der Linearitätseigenschaften der μ- und μ_n-Integrale ist sie damit auch für nichtnegative primitive Funktionen klar. Ist $(u_k)_{k=1}^\infty$ eine monoton wachsende Folge von nichtnegativen primitiven Funktion u_k mit $\sup_{k\in\mathbb{N}} u_k = f$, so haben wir

$$\int f\, d\mu = \sup_{k\in\mathbb{N}} \int u_k\, d\mu = \sup_{k\in\mathbb{N}} \sum_{n=1}^\infty \alpha_n \int u_k\, d\mu_n = \sup_{k\in\mathbb{N}} \int h_k\, d\tau\,, \tag{4.4}$$

wobei τ das Maß aus (ii) auf $(\mathbb{N},\mathfrak{P}(\mathbb{N}))$ und $h_k : \mathbb{N} \to [0,\infty]$ die durch $h_k(n) = \alpha_n \int u_k\, d\mu_n$, $n \in \mathbb{N}$, definierte Funktion ist. Es folgt aus der Definition und der Isotonie-Eigenschaft des μ_n-Integrals, dass $(h_k)_{k\in\mathbb{N}}$ eine monoton wachsende Folge von Funktionen mit $\sup_{k\in\mathbb{N}} h_k = h$ ist. Die Anwendung des Satzes (4.4) von der monotonen Konvergenz führt auf

$$\sup_{k\in\mathbb{N}} \int h_k\, d\tau = \int h\, d\tau = \sum_{n\in\mathbb{N}} h(n) = \sum_{n\in\mathbb{N}} \alpha_n \int f\, d\mu_n$$

und daher wegen (4.4) auf die behauptete Aussage (4.3).

Ist speziell $(\mu_n)_{n\in\mathbb{N}}$ eine Folge von Einpunktmaßen $\mu_n = \delta_{a_n}$, $a_n \in \Omega$, so erhalten wir

$$\int f\, d\mu = \sum_{n\in\mathbb{N}} \alpha_n f(a_n)\,.$$

In diesem Fall ist eine Funktion $f : (\Omega,\mathcal{A}) \to (\overline{\mathbb{R}},\overline{\mathcal{B}})$ offenbar genau dann μ-integrierbar, wenn $\sum_{n=1}^\infty \alpha_n |f(a_n)| < \infty$ ist. Es gilt dann

$$\int f\,d\mu = \sum_{n=1}^{\infty} \alpha_n f(a_n).$$

Sei $\boldsymbol{\lambda}^d$ das Lebesgue-Borelsche Maß auf der Borelschen σ-Algebra $\mathcal{B}^d$ von $\mathbb{R}^d$. Für jedes $a \in \mathbb{R}$ und alle $-\infty < a_j < b_j < \infty$ für $j = 2,\dots,d$ ist

$$\begin{aligned}\boldsymbol{\lambda}^d(\{a\} \times (a_2,b_2] \times \dots \times (a_d,b_d]) &= \lim_{n\to\infty} \boldsymbol{\lambda}^d\left(\bigcap\nolimits_{n=1}^{\infty}\left(a - \frac{1}{n}, a\right] \times (a_2,b_2] \times \dots \times (a_d,b_d]\right) \\ &= \lim_{n\to\infty} \frac{1}{n} \prod_{j=2}^{d} (b_j - a_j) = 0.\end{aligned}$$

Für $a_j \to -\infty$, $b_j \to \infty$, $j = 2,\dots,d$, folgt hieraus $\boldsymbol{\lambda}^d(\{a\} \times \mathbb{R}^{d-1}) = 0$. Mit analoger Argumentation ergibt sich allgemeiner $\boldsymbol{\lambda}^d(\mathbb{R}^{j-1} \times \{a\} \times \mathbb{R}^{d-j}) = 0$ für $j = 2,\dots,d$. Achsenparallele Hyperebenen haben somit das $\boldsymbol{\lambda}^d$-Maß 0. Im Fall $d = 1$ schreiben wir $\boldsymbol{\lambda}$ für $\boldsymbol{\lambda}^1$. Es ist $\boldsymbol{\lambda}(\{x\}) = 0$. Daher ist für jedes Intervall I, gleichgültig ob offen, halboffen oder abgeschlossen, $\boldsymbol{\lambda}(I)$ gleich der Länge von I. Für eine Menge $A \in \mathcal{B}^d$ heißt $\boldsymbol{\lambda}_A^d = \boldsymbol{\lambda}^d_{\mathcal{B}_A^d}$ das **Lebesgue-Borelsche Maß auf der Spur-σ-Algebra**

$$\mathcal{B}_A^d = \{A \cap B \mid B \in \mathcal{B}^d\}.$$

Satz 4.19

(i) Sei $-\infty < a < b < \infty$, $f : ([a,b], \mathcal{B}_{[a,b]}) \to (\mathbb{R}, \mathcal{B})$ Riemann-integrierbar. Dann ist f $\boldsymbol{\lambda}_{[a,b]}$-integrierbar und

$$\int_a^b f(x)\,dx = \int f\,d\boldsymbol{\lambda}_{[a,b]}.$$

(ii) Sei $f : (\mathbb{R}, \mathcal{B}) \to (\mathbb{R}, \mathcal{B})$ nichtnegativ und uneigentlich Riemann-integrierbar. Dann ist f $\boldsymbol{\lambda}$-integrierbar und

$$\int_{-\infty}^{\infty} f(x)\,dx = \int f\,d\boldsymbol{\lambda}.$$

Beweis

(i) Für jede Zerlegung $a = a_0 < a_1 < \dots < a_{m-1} < a_m = b$ von $[a,b]$ gilt

$$\begin{aligned}\sum_{n=1}^{m} \inf\{f(x) \mid a_{n-1} \le x \le a_n\}(a_n - a_{n-1}) &\le \sum_{n=1}^{m} \int_{[a_{n-1},a_n]} f\,d\boldsymbol{\lambda}_{[a,b]} \\ &\le \sum_{n=1}^{m} \sup\{f(x) \mid a_{n-1} \le x \le a_n\}(a_n - a_{n-1})\end{aligned}$$

und darüber hinaus

$$\sum_{n=1}^{m} \int_{[a_{n-1},a_n]} f\,d\boldsymbol{\lambda}_{[a,b]} = \int f\,d\boldsymbol{\lambda}_{[a,b]},$$

so dass die Behauptung aus der Definition des Riemann-Integrals folgt.

(ii) Dies folgt aus der Definition des uneigentlichen Riemann-Integrals, dem Satz (4.4) von der monotonen Konvergenz und der Tatsache, dass für alle $-\infty < a < b < \infty$ gilt:

$$\int f I_{[a,b]}\,d\boldsymbol{\lambda} = \int f|_{[a,b]}\,d\boldsymbol{\lambda}_{[a,b]}.$$ □

Anstelle von

$$\int f\,d\boldsymbol{\lambda}_{[a,b]} \text{ bzw. } \int f\,d\boldsymbol{\lambda}$$

schreiben wir auch

$$\int_a^b f(x)\,dx \text{ bzw. } \int_{-\infty}^{\infty} f(x)\,dx.$$

Treten solche Integrale in der Folge auf, wollen wir sie als $\boldsymbol{\lambda}_{[a,b]}$- bzw. $\boldsymbol{\lambda}$-Integrale auffassen.

Ist $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, $X : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B})$ eine Zufallsvariable, $f : (\mathbb{R}, \mathcal{B}) \to (\mathbb{R}, \mathcal{B})$ eine $\boldsymbol{\lambda}$-Dichte (oder spezieller sogar eine **Riemannsche Wahrscheinlichkeitsdichte**) von X (genauer: von P^X), so gilt für jede nichtnegative Funktion $g : (\mathbb{R}, \mathcal{B}) \to (\overline{\mathbb{R}}, \overline{\mathcal{B}})$

$$E(g \circ X) = \int g \circ X\,dP = \int g\,dP^X = \int gf\,d\boldsymbol{\lambda}. \tag{4.5}$$

Für eine beliebige Funktion $g : (\mathbb{R}, \mathcal{B}) \to (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ sind die P-Integrierbarkeit von $g \circ X$, die P^X-Integrierbarkeit von g und die $\boldsymbol{\lambda}$-Integrierbarkeit von gf äquivalent, und es gilt (4.5) für die entsprechenden endlichen Erwartungswerte und Integrale.

Definition 4.20

Ist $(R,\mathcal{S})$ ein Messraum mit der Eigenschaft, dass $\{x\} \in \mathcal{S}$ für jedes $x \in \mathbb{R}$ ist, so heißt eine Zufallsvariable $X:(\Omega,\mathcal{A}) \to (R,\mathcal{S})$ **diskret verteilt**, wenn es eine abzählbare Menge $S \subset R$ mit $P^X(S) = P(X \in S) = 1$ gibt. In diesem Fall ist $P(X = a) = 0$ für jedes $a \notin S$ und

$$P^X = \sum_{a \in S} P(X = a)\delta_a = \sum_{a \in R} P(X = a)\delta_a .$$

Für jede nichtnegative Funktion $g:(R,\mathcal{S}) \to (\overline{\mathbb{R}},\overline{\mathcal{B}})$ ist der Erwartungswert der Zufallsvariable $g \circ X$ darstellbar in der Form

$$E(g \circ X) = \int g\,dP^X = \sum_{a \in S} g(a)P(X = a) = \sum_{a \in \mathbb{R}} g(a)P(X = a) . \tag{4.6}$$

Für eine beliebige Funktion $g:(R,\mathcal{S}) \to (\overline{\mathbb{R}},\overline{\mathcal{B}})$ ist die Integrierbarkeit von $g \circ X$ äquivalent mit

$$\sum_{a \in S} |g(a)|P(X = a) < \infty$$

und es gilt die Darstellung (4.6).

Sei $(\Omega,\mathcal{A},P)$ ein Wahrscheinlichkeitsraum, $X:(\Omega,\mathcal{A}) \to (\mathbb{R},\mathcal{B})$. X habe eine $\boldsymbol{\lambda}$-Dichte f, d.h. $dP^X = f\,d\boldsymbol{\lambda}$, also

$$P(X \in B) = P^X(B) = \int fI_B\,d\boldsymbol{\lambda} = \int_B f\,d\boldsymbol{\lambda} \text{ für } B \in \mathcal{B} .$$

Ist f uneigentlich Riemann-integrierbar, so gilt

$$P(X \in [a,b]) = P^X([a,b]) = \int_{[a,b]} f\,d\boldsymbol{\lambda} = \int f\,d\boldsymbol{\lambda}_{[a,b]} = \int_a^b f(x)dx \text{ für } a,b \in [-\infty,\infty] ,$$

als Riemann-Integral aufgefasst. Es gilt:

$$E(X) = \int xf(x)\,d\boldsymbol{\lambda}(x) .$$

Sind die Funktionen $x \mapsto \max\{xf(x),0\}$ und $x \mapsto -\min(xf(x),0)$, $x \in \mathbb{R}$, uneigentlich Riemann-integrierbar, so gilt

$$E(X) = \int_{-\infty}^{\infty} \max\{xf(x),0\}\,dx + \int_{-\infty}^{\infty} \min\{xf(x),0\}\,dx$$

und damit

$$E(X) = \int_{-\infty}^{\infty} xf(x)\,dx .$$

Gilt $dP^X = f\,d\boldsymbol{\lambda}$, so heißt die Zufallsvariable X **absolut-stetig verteilt** mit der $\boldsymbol{\lambda}$-Dichte f.

Beispiel 4.21

(i) Sei X eine reelle Zufallsvariable. X (genauer: P^X) habe die $\boldsymbol{\lambda}$-Dichte

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

Dann heißt X **gleichverteilt** mit den Parametern $a \in \mathbb{R}$ und $\sigma^2 > 0$.
Schreibweise: $X \sim \mathcal{N}(a,\sigma^2)$.
Für den Erwartungswert gilt:

$$\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} x\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-a)^2}{2\sigma^2}\right)dx \\
&= \int_{-\infty}^{\infty} \underbrace{(x-a)}_{=:y}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-a)^2}{2\sigma^2}\right)dx + a\underbrace{\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-a)^2}{2\sigma^2}\right)dx}_{=1} \\
&= \underbrace{\int_{-\infty}^{\infty} y\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{y^2}{2\sigma^2}\right)dy}_{=0} + a = a.
\end{aligned}$$

(ii) X (genauer: P^X) habe die Dichte

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b, \\ 0, & \text{sonst.} \end{cases}$$

Dann gilt:

$$P(X = x) = \int_{-\infty}^{x} f(t)\,dt = \begin{cases} 0, & x \leq a, \\ \frac{x-a}{b-a}, & a < x < b, \\ 1, & x \geq b. \end{cases}$$

Dann heißt X **rechteckverteilt** auf $[a,b]$.
Schreibweise: $X \sim \mathcal{R}(a,b)$.
Für den Erwartungswert gilt:

$$E(X) = \int_{-\infty}^{\infty} xf(x)\,dx = \int_{a}^{b} x\frac{1}{b-a}\,dx = \frac{a+b}{2}.$$

(iii) X habe die Dichte

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

Dann gilt:

$$\begin{aligned} P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt &= \begin{cases} 0, & x < 0, \\ \int_0^x \lambda\exp(-\lambda t)\,dt, & x \geq 0 \end{cases} \\ &= \begin{cases} 0, & x < 0, \\ 1 - \exp(-\lambda x), & x \geq 0. \end{cases} \end{aligned}$$

Dann heißt X **exponentialverteilt** mit dem Parameter $\lambda > 0$.
Schreibweise: $X \sim \mathrm{Exp}(\lambda)$.
Für den Erwartungswert gilt:

$$E(X) = \int_{-\infty}^{\infty} xf(x)\,dx = \int_{0}^{\infty} x\lambda\exp(-\lambda x)\,dx = \frac{1}{\lambda}.$$

(iv) X habe die Dichte

$$f(x) = \frac{\sigma}{\pi}\frac{1}{\sigma^2 + (x-\mu)^2},\ x \in \mathbb{R}.$$

Dann gilt:

$$P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt = \frac{1}{2} + \frac{1}{\pi}\arctan\left(\frac{x-\mu}{\sigma}\right) \text{ für } x \in \mathbb{R}.$$

Dann heißt X **Cauchy-verteilt** mit den Parametern $\mu \in \mathbb{R}$ und $\sigma^2 > 0$.
Schreibweise: $X \sim \mathcal{C}(\mu,\sigma)$.
Aus

$$\int_0^{\infty} xf(x)\,dx = \int_0^{\infty} x\frac{\sigma}{\pi}\frac{1}{\sigma^2 + (x-\mu)^2}\,dx = \infty$$

und

$$\int_{-\infty}^{0} xf(x)\,dx = ... = 0$$

folgt, dass der Erwartungswert von X nicht existiert.

Sei $(\Omega,\mathcal{A},P)$ ein Wahrscheinlichkeitsraum, $(R,\mathcal{S})$ ein Messraum, $\{x\} \in \mathcal{S}$ für alle $x \in R$. Sei des Weiteren $X : (\Omega,\mathcal{A}) \to (R,\mathcal{S})$. X habe eine diskrete Verteilung, d.h. es gibt eine abzählbare Teilmenge $A \subset R$ mit $P(X \in A) = 1$. Dann gilt:

$$P^X = \sum_{a\in A} P(X = a)\delta_a.$$

Sei $g : (R,\mathcal{S}) \to (\overline{\mathbb{R}},\overline{\mathcal{B}})$. Dann ist $g \circ X : (\Omega,\mathcal{A}) \to (\overline{\mathbb{R}},\overline{\mathcal{B}})$ und es gibt zwei Fälle für den Erwartungswert:

(i) $g \geq 0$:

$$E(g \circ X) = \int g \circ X\,dP = \int g\,dP^X = \sum_{a\in A} g(a)P(X = a).$$

(ii) $g = g^+ - g^-$. Ist $E(g^+ \circ X) < \infty$ oder $E(g^- \circ X) < \infty$, so gilt

$$E(g \circ X) = \sum_{a\in A} g^+(a)P(X = a) - \sum_{a\in A} g^-(a)P(X = a).$$

Es ist $E(|g \circ X|) < \infty$ genau dann, wenn

$$\sum_{a\in A} |g(a)|P(X = a) < \infty.$$

Es gilt dann

$$E(g \circ X) = \sum_{a \in A} g(a) P(X = a).$$

Sei speziell $(R, \mathcal{S}) = (\mathbb{R}, \mathcal{B})$, $g(x) = x$. Dann ist $E(|X|) < \infty$ genau dann, wenn
$$\sum_{a \in A} |a| P(X = a) < \infty.$$

Es ist dann
$$E(X) = \sum_{a \in A} a P(X = a) = \sum_{x \in \mathbb{R}} x P(X = x).$$

Es gilt
$$E(X) = \sum_{x \in \mathbb{R}} x P(X = x) \quad (\text{„}X \text{ diskret“}),$$
$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx = \int x f(x)\, d\boldsymbol{\lambda}(x) \quad (\text{„}X \text{ absolut stetig verteilt mit } \boldsymbol{\lambda}\text{-Dichte } f\text{“}).$$

Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, $X : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B})$ eine reelle Zufallsvariable mit $E(|X|^s) < \infty$ für ein $s \geq 2$. Dann gilt für $0 < r < s$, dass $E(|X|^r) < \infty$ wegen $|X|^r \leq |X|^s + 1$ ist.

Definition 4.22

Sei $E(X^2) < \infty$. Dann ist auch $E(|X|) < \infty$ und $\operatorname{Var}(X) := E([X - E(X)]^2)$ heißt die **Varianz** von X.

Es gilt für $a \in \mathbb{R}$:
$$\begin{aligned} E([X - a]^2) &= E([(X - E(X)) - (a - E(X))]^2) \\ &= \operatorname{Var}(X) + [a - E(X)]^2 - 2 \underbrace{E([X - E(X)][a - E(X)])}_{=0} \\ &= \operatorname{Var}(X) + [a - E(X)]^2. \end{aligned}$$

Somit erhalten wir:
$$\min_{a \in \mathbb{R}} E([X - a]^2) = \operatorname{Var}(X).$$

Beispiel 4.23

(a) Sei $X \sim \mathcal{N}(a, \sigma^2)$. Dann folgt:
$$\begin{aligned} \operatorname{Var}(X) &= \int (x - E(X))^2\, dP = \int (x - a)^2\, dP^X(x) = \int_{-\infty}^{\infty} (x - a)^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(x - a)^2}{\sigma^2}\right) dx \\ &= \int_{-\infty}^{\infty} y^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{y^2}{\sigma^2}\right) dy = \sigma^2. \end{aligned}$$

(b) Sei $X \sim \mathcal{R}(a, b)$. Dann gilt:
$$\operatorname{Var}(X) = \int_a^b \frac{1}{b - a}\left(x - \frac{a + b}{2}\right)^2 dx = \frac{(b - a)^2}{12}.$$

(c) Sei $X \sim \operatorname{Exp}(\lambda)$. Dann ist
$$\operatorname{Var}(X) = \int_0^{\infty} \left(x - \frac{1}{\lambda}\right)^2 \lambda \exp(-\lambda x)\, dx = \frac{1}{\lambda^2}.$$

Definition 4.24

Seien X, Y reelle Zufallsvariablen, $E(X^2) < \infty$, $E(Y^2) < \infty$. Dann heißt
$$\operatorname{Cov}(X, Y) := E([X - E(X)][Y - E(Y)]) = E(XY) - E(X)E(Y)$$
die **Kovarianz** von X und Y.

Beispiel 4.25 (diskreter Fall)

Sei $X \sim \mathcal{H}(a, r, n)$. Dann folgt für $s := \max\{n + r - a, 0\}$ und $t := \min\{n, r\}$:
$$\operatorname{Var}(X) = E\left(\left[X - n\frac{r}{a}\right]^2\right) = \sum_{\substack{k \in \mathbb{N}_0, \\ s \leq k \leq t}} \left(k - n\frac{r}{a}\right)^2 \frac{\binom{r}{k}\binom{a-r}{n-k}}{\binom{a}{n}}.$$

Alternativ gilt für die Ereignisse B_i, dass beim n-fachen Ziehen je einer Kugel ohne Zurücklegen aus einer Urne mit r roten und s schwarzen Kugeln ($a := r + s$), wobei in der i-ten Ziehung eine rote Kugel gezogen wird:
$$X = \sum_{i=1}^{n} I_{B_i}.$$

Dazu seien X,Y reelle Zufallsvariablen mit $E(X^2) < \infty$ und $E(Y^2) < \infty$. Dann gilt für $(|X| - |Y|)^2 \geq 0$

$$|XY| \leq \frac{1}{2}(|X|^2 + |Y|^2),$$

also $E(|XY|) < \infty$ und damit $E(|X+Y|^2) < \infty$.
Dann gilt:

$$\begin{aligned}\operatorname{Var}(X+Y) &= E([X+Y-E(X+Y)]^2) = E(([X-E(X)]+[Y-E(Y)])^2) \\ &= \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\underbrace{E([X-E(X)][Y-E(Y)])}_{=\operatorname{Cov}(X,Y)}.\end{aligned}$$

Also gilt:

$$\operatorname{Var}(X+Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\operatorname{Cov}(X,Y).$$

Mittels vollständiger Induktion zeigt man: Sind $X_1,\ldots,X_n$ Zufallsvariablen mit $E(X_i^2) < \infty$, $i = 1,\ldots,n$, so gilt

$$\operatorname{Var}(X_1 + \ldots + X_n) = \operatorname{Var}(X_1) + \ldots + \operatorname{Var}(X_n) + 2\sum_{1 \leq i < j \leq n} \operatorname{Cov}(X_i, X_j).$$

Dann erhalten wir

$$\operatorname{Var}(X) = \operatorname{Var}\left(\sum\nolimits_{i=1}^n I_{B_i}\right) = \operatorname{Var}(I_{B_1}) + \ldots + \operatorname{Var}(I_{B_n}) + 2\sum_{1 \leq i < j \leq n} \operatorname{Cov}(I_{B_i}, I_{B_j}).$$

Allgemein gilt für Indikatorfunktionen:

$$\begin{aligned}\operatorname{Var}(I_A) &= E([I_A - P(A)]^2) = E(I_A - 2I_A P(A) + P(A)^2) = P(A) - P(A)^2 \\ &= P(A)(1 - P(A)).\end{aligned}$$

Also gilt:

$$\operatorname{Var}(I_{B_i}) = P(B_i)(1 - P(B_i)) = \frac{r}{a}\left(1 - \frac{r}{a}\right) \text{ für } i = 1,\ldots,n$$

und somit

$$\begin{aligned}\operatorname{Cov}(I_{B_i}, I_{B_j}) &= \operatorname{Cov}(I_{B_1}, I_{B_2}) \\ &= E(I_{B_1} I_{B_2}) - E(I_{B_1}) E(I_{B_2}) \\ &= P(B_1 \cap B_2) - P(B_1)P(B_2) = \frac{r(r-1)}{a(a-1)} - \left(\frac{r}{a}\right)^2\end{aligned}$$

für alle $1 \leq i < j \leq n$. Daraus folgt

$$\operatorname{Var}(X) = n\frac{r}{a}\left(1 - \frac{r}{a}\right) + n(n-1)\left(\frac{r(r-1)}{a(a-1)} - \left(\frac{r}{a}\right)^2\right) = n\frac{r}{a}\left(1 - \frac{r}{a}\right)\left(\frac{a-n}{a-1}\right).$$

Ungleichungen der mathematischen Stochastik

Satz 4.26

Sei X eine reelle Zufallsvariable mit $E(X^2) < \infty$. Dann gilt für alle $\varepsilon > 0$:

$$P\left(|X - E(X)| \underset{(-)}{\geq} \varepsilon\right) \leq \frac{1}{\varepsilon^2}\operatorname{Var}(X) \quad (\textbf{Chebyshevsche Ungleichung})$$

Beweis
Es gilt:

$$\begin{aligned}\operatorname{Var}(X) &= E([X - E(X)]^2) \geq E(\underbrace{[X - E(X)]^2 I\left(|X - E(X)| \underset{(-)}{\geq} \varepsilon\right)}_{\geq \varepsilon^2 I(|X - E(X)| \underset{(-)}{\geq} \varepsilon)}) \geq E\left(\varepsilon^2 I\left(|X - E(X)| \underset{(-)}{\geq} \varepsilon\right)\right) \\ &= \varepsilon^2 P\left(|X - E(X)| \underset{(-)}{\geq} \varepsilon\right).\end{aligned}$$

□

Satz 4.27

Sei X eine reelle Zufallsvariable, $f : [0,\infty) \to [0,\infty)$ monoton wachsend. Dann gilt für alle $\varepsilon > 0$:

$$P(|X| > \varepsilon) \leq \frac{1}{f(\varepsilon)} E(f \cdot |X|) \quad (\textbf{Markovsche Ungleichung}).$$

Beweis
Es gilt:

$$E(f \cdot |X|) \geq E(f \cdot |X| I(|X| > \varepsilon)) \geq f(\varepsilon) E(I(|X| > \varepsilon)) = f(\varepsilon) P(|X| > \varepsilon).$$ □

Für den Spezialfall $f(x) = x^k$ für $x \geq 0$, $k \in \mathbb{N}$, gilt:

$$P(|X| \geq \varepsilon) \leq \frac{1}{\varepsilon^k} E(|X|^k).$$

Es sei $E(|X|^k) < \infty$. Dann ist auch $E(X) < \infty$ und es gilt

$$P(|X - E(X)| \geq \varepsilon) \leq \frac{1}{\varepsilon^k} E(|X - E(X)|^k).$$

Satz 4.28

Es seien X,Y reelle Zufallsvariablen mit $E(X^2) < \infty$ und $E(Y^2) < \infty$. Dann gilt:

$$E(|XY|) \leq \sqrt{E(X^2)E(Y^2)} \quad \textbf{(Cauchy-Schwarzsche Ungleichung)}.$$

Beweis

1. Fall: $E(Y^2) = 0$: Aus $P(|Y| = 0) = 1$ folgt $P(|XY| = 0) = 1$ und damit $E(|XY|) = 0$.

2. Fall: $E(Y^2) > 0$: Für alle $t \in \mathbb{R}$ gilt

$$0 \leq E([|X| - t|Y|]^2) = E(|X|^2) - 2tE(|XY|) + t^2 E(|Y|^2) =: f(t),$$

also folgt

$$0 \overset{!}{=} f'(t) = 2tE(|Y|^2) - 2E(|XY|).$$

Somit erhalten wir

$$t_0 = \frac{E(|XY|)}{E(|Y|^2)}$$

und damit

$$0 \leq f(t_0) = E(|X|^2) - 2\frac{E(|XY|)^2}{E(|Y|^2)} + \frac{E(|XY|)^2}{E(|Y|^2)} = E(|X|^2) - \frac{E(|XY|)^2}{E(|Y|^2)}.$$

Also gilt

$$E(|XY|)^2 \leq E(|X|^2)E(|Y|^2).$$

□

Korollar 4.29

Für die Kovarianz gilt:

$$|\mathrm{Cov}(X,Y)| \leq \sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}.$$

Beweis

Es gilt:

$$\begin{aligned}|\mathrm{Cov}(X,Y)| &= |E([X - E(X)][Y - E(Y)])| \leq E(|[X - E(X)][Y - E(Y)]|) \\ &\leq \sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}.\end{aligned}$$

□

Definition 4.30

Seien X,Y reelle Zufallsvariablen mit $E(X^2) < \infty$ und $E(Y^2) < \infty$. Dann heißt

$$\rho(X,Y) := \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

der **Korrelationskoeffizient** von X und Y . Dabei ist $\rho(X,Y) := 0$, falls $\mathrm{Var}(X) = 0$ oder $\mathrm{Var}(Y) = 0$.

Es gilt:

$$-1 \leq \rho(X,Y) \leq 1.$$

Interpretation des Korrelationskoeffizienten

Seien X,Y reelle Zufallsvariablen mit $E(X^2) = 1$, $E(Y^2) = 1$, $E(X) = 0$ und $E(Y) = 0$. Sei

$$f(a,b) := E([X - (aY + b)]^2).$$

Gesucht wird $\min_{a,b\in\mathbb{R}} f(a,b)$.

Es gilt:

$$\begin{aligned}f(a,b) &= E(X^2) + E([aY + b]^2) - 2E(X[aY + b]) \\ &= E(X^2) + a^2E(Y^2) + b^2 - 2aE(XY)\end{aligned}.$$

Daraus folgt für $f(a,b) \to \min_{a,b\in\mathbb{R}} f(a,b)$, dass $b = 0$. Also ist

$$f(a,b) = 1 + a^2 - 2aE(XY)$$

minimal für $a = E(XY)$.

Es folgt:

$$\min_{a,b\in\mathbb{R}} E\big([X-(aY+b)]^2\big) = 1 + E(XY)^2 - 2E(XY)^2 = 1 - E(XY)^2 = 1 - \rho(X,Y)^2 .$$

Seien X,Y beliebige reelle Zufallsvariablen mit $E(X^2) < \infty$, $E(Y^2) < \infty$, $\operatorname{Var}(X) > 0$ und $\operatorname{Var}(Y) > 0$. Aus der Standardisierung

$$X^* = \frac{X - E(X)}{\sqrt{\operatorname{Var}(X)}},\; Y^* = \frac{Y - E(Y)}{\sqrt{\operatorname{Var}(Y)}}$$

folgt $E(X^*) = E(Y^*) = 0$ und somit

$$\operatorname{Var}(X^*) = E\big((X^*)^2\big) = 1 = \operatorname{Var}(Y^*) = E\big((Y^*)^2\big)$$

und

$$\begin{aligned}
\min_{a,b\in\mathbb{R}} E\big([X-(aY+b)]^2\big) &= E\Big(\big[\sqrt{\operatorname{Var}(X)}X^* + E(X) - \big(aY^*\sqrt{\operatorname{Var}(Y)} + E(Y) + b\big)\big]^2\Big) \\
&= \min_{a,b\in\mathbb{R}} \operatorname{Var}(X) E\left(\left[X^* - \left(\frac{aY^*\sqrt{\operatorname{Var}(Y)} + E(Y) - E(X)}{\sqrt{\operatorname{Var}(X)}}\right)\right]^2\right) \\
&= \min_{a,b\in\mathbb{R}} \operatorname{Var}(X) E\big([X^* - aY^*]^2\big) \\
&= \operatorname{Var}(X)\big(1 - \rho(X^*,Y^*)^2\big) \\
&= \operatorname{Var}(X)\big(1 - \rho(X,Y)^2\big).
\end{aligned}$$

Der Korrelationskoeffizient ist also ein Maß für den (affin-)linearen Zusammenhang zwischen X und Y.

Kapitel 5

Stochastische Unabhängigkeit von Zufallsvariablen und Ereignissen, Produkträume

Definition 5.1

Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, $(R_i, \mathcal{S}_i)$, $i \in I$, seien Messräume, $X_i : (\Omega, \mathcal{A}) \to (R_i, \mathcal{S}_i)$, $i \in I$, Zufallsvariablen. Dann heißen die Zufallsvariablen X_i, $i \in I$, heißen **(stochastisch) unabhängig**, wenn für jede endliche Teilmenge $K \subset I$ gilt:

$$P\Big(\bigcap\nolimits_{i \in K} \{X_i \in B_i\}\Big) = \prod_{i \in K} P(\{X_i \in B_i\}) \text{ für jede Auswahl } B_j \in \mathcal{S}_j,\ j \in I.$$

Konvention:

$$\bigcap\nolimits_{i \in \varnothing} \{X_i \in B_j\} := \Omega,\ \prod\nolimits_{i \in \varnothing} P(\{X_i \in B_j\}) := 1.$$

Definition 5.2

Seien $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und A_i, $i \in I$, eine Familie von Ereignissen, $A_i \in \mathcal{A}$, $i \in I$. Die A_i, $i \in I$, heißen **stochastisch unabhängig**, wenn die Zufallsvariablen I_{A_i}, $i \in I$, (stochastisch) unabhängig sind.

Beispiel 5.3

Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, $A_1, \ldots, A_n \in \mathcal{A}$ seien unabhängig. Sei $p = P(A_i)$, $i = 1, \ldots, n$, und

$$X = \sum\nolimits_{i=1}^{n} I_{A_i}.$$

Dann gilt für $J = \{1, \ldots, n\}$:

$$\begin{aligned}
P(X = k) &= P\Big(\bigcup\nolimits_{\substack{H \subset J \\ |H| = k}} \Big(\bigcap\nolimits_{l \in H} A_l \cap \bigcap\nolimits_{j \in J \setminus H} A_j^c\Big)\Big) = \sum_{\substack{H \subset J \\ |H| = k}} P\Big(\bigcap\nolimits_{l \in H} A_l \cap \bigcap\nolimits_{j \in J \setminus H} A_j^c\Big) \\
&= \sum_{\substack{H \subset J \\ |H| = k}} P\Big(\sum\nolimits_{l \in H} \{I_{A_l} = 1\} \cap \bigcap\nolimits_{j \in J \setminus H} \{I_{A_j} = 0\}\Big) \\
&= \sum_{\substack{H \subset J \\ |H| = k}} \prod_{l \in H} P(I_{A_l} = 1) \prod_{j \in J \setminus H} P(I_{A_j} = 0) \\
&= \sum_{\substack{H \subset J \\ |H| = k}} p^k (1 - p)^{n-k} = \binom{n}{k} p^k (1 - p)^{n-k}
\end{aligned}$$

für $k \in \{0, \ldots, n\}$.

Definition 5.4

Eine reelle Zufallsvariable X, deren Verteilung festgelegt ist durch

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \text{ für } k \in \{0, \ldots, n\},$$

heißt **binomialverteilt** mit den Parametern $n \in \mathbb{N}$ und $p \in [0,1]$ und heißt **Binomialverteilung**.
Schreibweise: $X \sim \mathcal{B}(n, p)$.

Bemerkung 5.5

Seien $(\Omega, \mathcal{A}, P)$, $(R_i, \mathcal{S}_i)$, $i = 1, \ldots, n$, Messräume. Die Zufallsvariablen $X_i : (\Omega, \mathcal{A}) \to (R_i, \mathcal{S}_i)$, $i = 1, \ldots, n$, sind genau dann unabhängig, wenn gilt

$$P(X_1 \in B_1 \cap \ldots \cap X_n \in B_n) = P(X_1 \in B_1) \cdot \ldots \cdot P(X_n \in B_n)$$

für jede Auswahl von Mengen $B_i \in \mathcal{S}_i$, $i = 1, \ldots, n$.

Produktmaße

Für zwei Messräume $(\Omega_i, \mathcal{A}_i)$, $i = 1,2$, sei $\Omega_1 \times \Omega_2$ das kartesische Produkt von Ω_1 und Ω_2 und

$$\mathcal{E} = \{ A_1 \times A_2 \mid A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2 \}$$

das System der **Rechteck-Mengen** auf $\Omega_1 \times \Omega_2$. $\mathcal{E}$ ist ein Semi-Ring auf $\Omega := \Omega_1 \times \Omega_2$. Die von diesem Semi-Ring erzeugte σ-Algebra

$$\mathcal{A}_1 \otimes \mathcal{A}_2 := \sigma(\mathcal{E})$$

heißt **Produkt-σ-Algebra** von $\mathcal{A}_1$ und $\mathcal{A}_2$ auf $\Omega_1 \times \Omega_2$. Für $\omega_1 \in \Omega_1$ und $A \subset \Omega$ heißt

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

der ω_1-Schnitt von A und analog für $\omega_2 \in \Omega_2$

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

der ω_2-Schnitt von A. Für jedes $\omega_1 \in \Omega_1$ ist $\{ A \in \mathcal{A}_1 \otimes \mathcal{A}_2 \mid A_{\omega_1} \in \mathcal{A}_2 \}$ eine σ-Algebra, die den Semi-Ring $\mathcal{E}$ enthält. Also ist

$$A_{\omega_1} \in \mathcal{A}_2 \text{ für jedes } A \in \mathcal{A}_1 \times \mathcal{A}_2 \text{ und jedes } \omega_1 \in \Omega_1. \tag{5.1}$$

Analog ist auch $A_{\omega_2} \in \mathcal{A}_1$ für jedes $A \in \mathcal{A}_1 \times \mathcal{A}_2$ und jedes $\omega_2 \in \Omega_2$. Für einen Messraum $(R, \mathcal{S})$ und eine Funktion $f : (\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2) \to (R, \mathcal{S})$ ist die für jedes $\omega_1 \in \Omega_1$ durch

$$f_{\omega_1}(\omega_2) = f(\omega_1, \omega_2), \; \omega_2 \in \Omega_2,$$

definierte Funktion $f_{\omega_1} : \Omega_2 \to R$ $(\mathcal{A}_2, \mathcal{S})$-messbar. Dies folgt aus der für jedes $B \subset R$ gültigen Identität

$$f_{\omega_1}^{-1}(B) = \left(f^{-1}(B) \right)_{\omega_1},$$

der $(\mathcal{A}_1 \otimes \mathcal{A}_2, \mathcal{S})$-Messbarkeit von f und der Eigenschaft (5.1) für ω_1-Schnittmengen. Analog ist die für jedes $\omega_2 \in \Omega_2$ durch

$$f_{\omega_2}(\omega_1) = f(\omega_1, \omega_2), \; \omega_1 \in \Omega_1,$$

definierte Funktion $f_{\omega_2} : \Omega_1 \to R$ $(\mathcal{A}_1, \mathcal{S})$-messbar. Ist μ_1 ein Maß auf $\mathcal{A}_1$ und μ_2 ein Maß auf $\mathcal{A}_2$, so wird durch

$$\tilde{\mu}(A \times B) = \mu_1(A)\mu_2(B), \; A \in \mathcal{A}_1, \; B \in \mathcal{A}_2,$$

eine σ-additive Mengenfunktion $\tilde{\mu}$ auf dem Semi-Ring $\mathcal{E}$ definiert. Es ist nämlich $\tilde{\mu}(\varnothing) = 0$ und im Falle

$$A \times B = \sum_{n=1}^{\infty} A_n \times B_n$$

mit $A \times B \in \mathcal{E}$ und paarweise disjunkten $A_n \times B_n \in \mathcal{E}$, $n \in \mathbb{N}$,

$$I_A(\omega_1) I_B(\omega_2) = \sum_{n=1}^{\infty} I_{A_n}(\omega_1) I_{B_n}(\omega_2), \; (\omega_1, \omega_2) \in \Omega_1 \times \Omega_2. \tag{5.2}$$

Die Integration der bei festem $\omega_2 \in \Omega_2$ $(\mathcal{A}_1, \overline{\mathcal{B}})$-messbaren nichtnegativen $\overline{\mathbb{R}}$-wertigen Funktionen auf der rechten und der linken Seite in (5.2) bezüglich μ_1 führt unter Anwendung des Satzes (4.4) von der monotonen Konvergenz auf die Identität

$$\mu_1(A) I_B(\omega_2) = \sum_{n=1}^{\infty} \mu_1(A_n) I_{B_n}(\omega_2). \tag{5.3}$$

Die Integration der $(\mathcal{A}_2, \overline{\mathcal{B}})$-messbaren nichtnegativen $\overline{\mathbb{R}}$-wertigen Funktionen auf der rechten und der linken Seite in (5.3) bezüglich μ_2 liefert

$$\tilde{\mu}(A \times B) = \sum_{n=1}^{\infty} \tilde{\mu}(A_n \times B_n).$$

Aus dem Maßerweiterungssatz (1.29) folgt die Existenz eines Maßes $\mu_1 \otimes \mu_2$ auf $\mathcal{A}_1 \otimes \mathcal{A}_2$ mit der Eigenschaft $\mu_1 \otimes \mu_2(A \times B) = \mu_1(A)\mu_2(B)$ für jedes $A \in \mathcal{A}_1$ und jedes $B \in \mathcal{A}_2$. Im Falle σ-endlicher Maße μ_1 und μ_2 ist $\mu_1 \otimes \mu_2$ eindeutig bestimmt und heißt **Produktmaß** von μ_1 und μ_2. Der nachfolgende Satz zeigt, wie unter dieser Voraussetzung $\mu_1 \otimes \mu_2(C)$ für beliebige Mengen $C \in \mathcal{A}_1 \otimes \mathcal{A}_2$ berechnet werden kann. Zur Verdeutlichung vereinbaren wir für existierende μ-Integrale $\overline{\mathbb{R}}$-wertiger messbarer Funktionen f noch die alternative Schreibweise

$$\int f \, d\mu = \int f(\omega) \, d\mu(\omega).$$

Satz 5.6
Es seien μ_1 und μ_2 σ-endliche Maße auf $\mathcal{A}_1$ und $\mathcal{A}_2$. Für das durch
$$\mu_1 \otimes \mu_2(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2) \text{ für } A_1 \in \mathcal{A}_1,\ A_2 \in \mathcal{A}_2,$$
eindeutig festgelegte Produktmaß $\mu_1 \otimes \mu_2$ auf $\mathcal{A}_1 \otimes \mathcal{A}_2$ gilt
$$\mu_1 \otimes \mu_2(C) = \int \mu_2(C_{\omega_1})\,d\mu_1(\omega_1) = \int \mu_1(C_{\omega_2})\,d\mu_2(\omega_2) \text{ für jedes } C \in \mathcal{A}_1 \otimes \mathcal{A}_2. \tag{5.4}$$
Beweis
Für $E \in \mathcal{A}_2$ mit $\mu_2(E) < \infty$ sei
$$\mathcal{D}_E = \left\{C \in \mathcal{A}_1 \otimes \mathcal{A}_2 \mid \omega_1 \to \mu_2(E \cap C_{\omega_1}),\ \omega_1 \in \Omega_1, \text{ ist } (\mathcal{A}_1, \overline{\mathcal{B}})\text{-messbar}\right\}.$$
Es ist $\mathcal{D}_E$ ein Dynkin-System auf $\Omega_1 \times \Omega_2$, welches den Semi-Ring $\mathcal{E}$ enthält. Also ist $\mathcal{D}_E = \mathcal{A}_1 \otimes \mathcal{A}_2$. Wegen der σ-Endlichkeit von μ_2 gibt es eine Folge von Mengen $E_n \in \mathcal{A}_2$ mit $E_n \uparrow \Omega_2$ und $\mu_2(E_n) < \infty$ für jedes $n \in \mathbb{N}$. Aus
$$\mu_2(E_n \cap C_{\omega_1}) \uparrow \mu_2(C_{\omega_1}) \text{ für jedes } \omega_1 \in \Omega_1$$
folgt die $(\mathcal{A}_1, \overline{\mathcal{B}})$-Messbarkeit von
$$\omega_1 \to \mu_2(C_{\omega_1}),\ \omega_1 \in \Omega_1.$$
Das linke Integral in (5.4) ist daher sinnvoll. Durch
$$\mu(C) = \int \mu_2(C_{\omega_1})\,d\mu_1(\omega_1),\ C \in \mathcal{A}_1 \otimes \mathcal{A}_2,$$
wird ein Maß μ auf $\mathcal{A}_1 \otimes \mathcal{A}_2$ mit der Eigenschaft
$$\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2),\ A_1 \in \mathcal{A}_1,\ A_2 \in \mathcal{A}_2,$$
definiert. Wegen $\mu_1 \otimes \mu_2|_{\mathcal{E}} = \mu|_{\mathcal{E}}$ ist $\mu = \mu_1 \otimes \mu_2$ und damit die erste Identität in (5.4) bewiesen. Der Beweis der zweiten Identität erfolgt analog. □

Satz 5.7 (Satz von Fubini)
Sei $f : (\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2) \to (\overline{\mathbb{R}}, \overline{\mathcal{B}})$.

(i) Ist $f \geq 0$, so ist
$$\omega_1 \to \int f_{\omega_1}\,d\mu_2,\ \omega_1 \in \Omega_1,$$
$(\mathcal{A}_1, \overline{\mathcal{B}})$-messbar bzw.
$$\omega_2 \to \int f_{\omega_2}\,d\mu_1,\ \omega_2 \in \Omega_2,$$
$(\mathcal{A}_2, \overline{\mathcal{B}})$-messbar und es gilt
$$\int f\,d\mu_1 \otimes \mu_2 = \int\Big(\int f_{\omega_1}\,d\mu_2\Big)d\mu_1(\omega_1) = \int\Big(\int f_{\omega_2}\,d\mu_1\Big)d\mu_2(\omega_2). \tag{5.5}$$

(ii) Ist f $\mu_1 \otimes \mu_2$-integrierbar, so ist f_{ω_1} für μ_1-f.a. ω_1 μ_2-integrierbar und f_{ω_2} für μ_2-f.a. ω_2 μ_1-integrierbar. Die μ_1-f.ü. bzw. μ_2-f.ü. definierte Funktion
$$\omega_1 \to \int f_{\omega_1}\,d\mu_2 \text{ bzw. } \omega_2 \to \int f_{\omega_2}\,d\mu_1$$
ist μ_1- bzw. μ_2-integrierbar und es gilt die Identität (5.5).

Beweis
Den Beweis führt man leicht mittels algebraischer Induktion. Man hat für den Nachweis der Aussagen von Teil (ii) lediglich noch die f.ü.-Eigenschaft der angegebenen Funktionen nachzuweisen. Diese ergibt sich unter Beachtung von $|f|_{\omega_i} = |f_{\omega_i}|$, $(f^+)_{\omega_i} = (f_{\omega_i})^+$ und $(f^-)_{\omega_i} = (f_{\omega_i})^-$ für $\omega_i \in \Omega_i$, $i = 1,2$, und
$$\int\Big(\int |f_{\omega_2}|\,d\mu_1\Big)d\mu_2(\omega_2) = \int\Big(\int |f_{\omega_1}|\,d\mu_2\Big)d\mu_1(\omega_1) = \int |f|\,d\mu_1 \otimes \mu_2 < \infty$$
aus der folgenden Hilfsüberlegung. □

Lemma 5.8
Sei μ ein Maß auf der σ-Algebra $\mathcal{A}$ auf Ω, $f : (\Omega, \mathcal{A}) \to (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ μ-integrierbar. Dann ist f μ-f.ü. endlich.
Beweis
Es ist
$$\infty > a := \int_{\{|f|=\infty\}} |f|\,d\mu \geq n\mu(|f| = \infty) \text{ für jedes } n \in \mathbb{N},$$
also $\mu(|f| = \infty) = 0$.
Sind μ_i σ-endliche Maße auf den σ-Algebren $\mathcal{A}_i$ auf Ω_i, $i = 1,\ldots,n$, so lassen sich die Produkte $\Omega_1 \times \ldots \times \Omega_n$, $\mathcal{A}_1 \otimes \ldots \otimes \mathcal{A}_n$, $\mu_1 \otimes \ldots \otimes \mu_n$ induktiv definieren, wobei sich herausstellt, dass die Produktbildung assoziativ ist. Insbesondere ist

$$(\mathbb{R}^n, \mathcal{B}^n) = (\mathbb{R} \times ... \times \mathbb{R}, \mathcal{B} \otimes ... \otimes \mathcal{B}).$$

Das n-fache Produktmaß $\boldsymbol{\lambda}^n = \boldsymbol{\lambda} \otimes ... \otimes \boldsymbol{\lambda}$ des Lebesgue-Borelschen Maßes $\boldsymbol{\lambda}$ ist das n-dimensionale Lebesgue-Borelsche Maß auf $\mathcal{B}^n$. Sind $(\Omega_i, \mathcal{A}_i, P_i)$, $i = 1,...,n$, Wahrscheinlichkeitsräume, so heißt

$$(\Omega_1 \times ... \times \Omega_n, \mathcal{A}_1 \otimes ... \otimes \mathcal{A}_n, P_1 \otimes ... \otimes P_n)$$

Produktraum der $(\Omega_i, \mathcal{A}_i, P_i)$, $i = 1,...,n$. Ist für $i = 1,...,n$ $\pi_i : \Omega_1 \times ... \times \Omega_n \to \Omega_i$ die Projektion auf Ω_i, also

$$\pi_i(\omega_1,...,\omega_n) = \omega_i \text{ für } (\omega_1,...,\omega_n) \in \Omega_1 \times ... \times \Omega_n,$$

so ist

$$(P_1 \otimes ... \otimes P_n)^{\pi_i} = P_i \text{ für jedes } i = 1,...,n. \qquad \square$$

Produkträume

Für nur endlich viele $(\Omega_i, \mathcal{A}_i, \mu_i)$, $i = 1,...,n$, lässt sich der Produktraum

$$(\Omega_1 \times ... \times \Omega_n, \mathcal{A}_1 \otimes ... \otimes \mathcal{A}_n, \mu_1 \otimes ... \otimes \mu_n)$$

induktiv definieren. Sei im Folgenden $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, $(R_i, \mathcal{S}_i)$, $i = 1,...,n$, Messräume, $X_i : (\Omega, \mathcal{A}) \to (R_i, \mathcal{S}_i)$. Sei $\pi_i : R_1 \times ... \times R_n \to R_i$ mit $\pi_i(x_1,...,x_n) = x_i$ für $(x_1,...,x_n) \in R_1 \times ... \times R_n$, $i = 1,...,n$. Dann ist $X : (\Omega, \mathcal{A}) \to (R_1 \times ... \times R_n, \mathcal{S}_1 \otimes ... \otimes \mathcal{S}_n)$, wobei $X = (X_1,...,X_n)^{\intercal}$. Denn es gilt:

$$X^{-1}(B_1 \times ... \times B_n) = \bigcap_{i=1}^{n} \{X_i \in B_i\} \in \mathcal{A} \text{ für } B_i \in \mathcal{S}_i,\ i = 1,...,n.$$

Ist umgekehrt $X = (X_1,...,X_n)^{\intercal}$, so gilt, dass $X_i : \Omega \to R_i$ $(\mathcal{A}, \mathcal{S}_i)$-messbar ist für jedes $i = 1,...,n$. Denn:

$$X_i = \pi_i \circ X \text{ und } \pi_i^{-1}(B_i) = R_1 \times ... \times B_i \times ... \times R_n \in \mathcal{S}_1 \otimes ... \otimes \mathcal{S}_n \text{ für } B_i \in \mathcal{S}_i,\ i = 1,...,n.$$

Ist $(R_i, \mathcal{S}_i) = (\mathbb{R}, \mathcal{B})$, dann ist jede Zufallsvariable $X : (\Omega, \mathcal{A}) \to (\mathbb{R}^n, \mathcal{B}^n)$ mit $X = (X_1,...,X_n)$ ein n-dimensionaler Zufallsvektor – in der Regel ein Spaltenvektor. Es ist

$$(\mathbb{R} \times ... \times \mathbb{R}, \mathcal{B} \otimes ... \otimes \mathcal{B}, \boldsymbol{\lambda} \otimes ... \otimes \boldsymbol{\lambda}) = (\mathbb{R}^n, \mathcal{B}^n, \boldsymbol{\lambda}^n).$$

Der n-dimensionale Zufallsvektor X habe die $\boldsymbol{\lambda}^n$-Dichte f. D.h.

$$P(X \in B) = \int_B f\, d\boldsymbol{\lambda}^n = \int f I_B\, d\boldsymbol{\lambda}^n = \int_B f(x_1,...,x_n)\, d\boldsymbol{\lambda} \otimes ... \otimes \boldsymbol{\lambda}(x_1,...,x_n),\ B \in \mathcal{B}^n.$$

Sei $T : \mathbb{R}^n \to \mathbb{R}^n$ bijektiv und stetig differenzierbar mit Nichtverschwindender Funktionaldeterminante

$$\Delta : \mathbb{R}^n \to \mathbb{R}.$$

Dann hat $Y = T \circ X$ die $\boldsymbol{\lambda}^n$-Dichte

$$\frac{1}{|\Delta \circ T^{-1}(y)|} f \circ T^{-1}(y) \text{ für } y \in \mathbb{R}^n.$$

Ein Transformationssatz für Dichten

Sei X ein n-dimensionaler Zufallsvektor mit der $\boldsymbol{\lambda}^n$-Dichte f, welche außerhalb der Vereinigung von abzählbar vielen paarweise disjunkten offenen Mengen $M_i \subset \mathbb{R}^n$, $i \in I$, verschwinde. Die Abbildung

$$T : (\mathbb{R}^n, \mathcal{B}^n) \to (\mathbb{R}^n, \mathcal{B}^n)$$

besitze folgende Eigenschaften:

(i) die Restriktion T_i von T auf M_i ist stetig differenzierbar,

(ii) die Funktionaldeterminante Δ_i von T_i ist auf M_i von 0 verschieden,

(iii) T_i ist injektiv.

Es sei $S_i : T_i(M_i) \to M_i$ die Inverse von T_i. Dann ist $T_i(M_i)$ offen, S_i $(\mathcal{B}^n_{T_i(M_i)}, \mathcal{B}^n_{M_i})$-messbar, und

$$\sum_{i \in I} \frac{f \circ S_i}{|\Delta_i \circ S_i|} I(T_i(M_i))$$

ist $\boldsymbol{\lambda}^n$-Dichte des n-dimensionalen Zufallsvektors $Y = T \circ X$.

Folgerung 5.9

(i) Hat die reelle Zufallsvariable X die $\boldsymbol{\lambda}$-Dichte $f(x)$, $x \in \mathbb{R}$, so hat die Zufallsvariable $Y = aX + b$, $(a \neq 0,\ b \in \mathbb{R})$ die $\boldsymbol{\lambda}$-Dichte $\frac{1}{|a|} f\left(\frac{x-b}{a}\right)$, $x \in \mathbb{R}$.

(ii) Hat der Zufallsvektor (X,Y) die $\boldsymbol{\lambda}^2$-Dichte $f(x,y)$, so hat die Zufallsvariable

(1) $X \pm Y$ die $\boldsymbol{\lambda}$-Dichte $h(z) = \int_{-\infty}^{\infty} f(z \mp y, y)\,dy$, $z \in \mathbb{R}$,

(2) XY die $\boldsymbol{\lambda}$-Dichte $h(z) = \int_0^{\infty} (f(\frac{z}{y}, y) + f(-\frac{z}{y}, -y))\frac{1}{y}\,dy$, $z \in \mathbb{R}$,

(3) $\frac{X}{Y}$ die $\boldsymbol{\lambda}$-Dichte $h(z) = \int_0^{\infty} (f(zy, y) + f(-zy, -y))\,y\,dy$, $z \in \mathbb{R}$.

Die Aussage (ii)(2) oben beweist man z.B. so:
Sei $T : \mathbb{R} \times (\mathbb{R} \setminus \{0\}) \to \mathbb{R}^2$, $T(x,y) = (xy, y)$ für $(x,y) \in \mathbb{R} \times (\mathbb{R} \setminus \{0\})$. Sei weiter $T_i = T|_{M_i}$, $i = 1,2$, mit $M_1 = \mathbb{R} \times (0,\infty)$, $M_2 = \mathbb{R} \times (-\infty, 0)$. Es ist $|\Delta_i(x,y)| = |y|$ für $(x,y) \in M_i$, $i = 1,2$, $S_i(u,v) = (\frac{u}{v}, v)$ für $(u,v) \in M_i$, $i = 1,2$, also

$$\frac{1}{|v|}\Big(f\Big(\frac{u}{v}, v\Big) I(\mathbb{R} \times (0,\infty))(u,v) + f\Big(\frac{u}{v}, v\Big) I(\mathbb{R} \times (-\infty, 0))(u,v)\Big),\ (u,v) \in \mathbb{R}^2,$$

die $\boldsymbol{\lambda}^2$-Dichte von (XY, Y). Damit ist

$$\int_{-\infty}^{\infty} \Big(f\Big(\frac{u}{v}, v\Big) I(\mathbb{R} \times (0,\infty))(u,v) + f\Big(\frac{u}{v}, v\Big) I(\mathbb{R} \times (-\infty,0))(u,v)\Big)\frac{1}{|v|}\,dv = \int_0^{\infty} \Big(f\Big(\frac{u}{v}, v\Big) + f\Big(-\frac{u}{v}, -v\Big)\Big)\frac{1}{v}\,dv$$

für $u \in \mathbb{R}$ die $\boldsymbol{\lambda}$-Dichte von XY.

Sind die Zufallsvariablen X und Y in (ii) darüber hinaus unabhängig mit den $\boldsymbol{\lambda}$-Dichten f und g, so hat

(1) $X \pm Y$ die $\boldsymbol{\lambda}$-Dichte $h(z) = \int_{-\infty}^{\infty} f(z \mp y) g(y)\,dy$, $z \in \mathbb{R}$,

(2) XY die $\boldsymbol{\lambda}$-Dichte $h(z) = \int_0^{\infty} (f(\frac{z}{y}) g(y) + f(-\frac{z}{y}) g(-y))\frac{1}{y}\,dy$, $z \in \mathbb{R}$,

(3) $\frac{X}{Y}$ die $\boldsymbol{\lambda}$-Dichte $h(z) = \int_0^{\infty} (f(zy) g(y) + f(-zy) g(-y))\,y\,dy$, $z \in \mathbb{R}$.

Sei $(\Omega_i, \mathcal{A}_i, P_i)$ ein Modell für ein Zufallsexperiment i, $i = 1,\dots,n$. Dann ist

$$(\Omega_1 \times \dots \times \Omega_n, \mathcal{A}_1 \otimes \dots \otimes \mathcal{A}_n, P_1 \otimes \dots \otimes P_n)$$

ein mathematisches Modell für das Gesamtexperiment, das aus der „unabhängigen“ Ausführung der Zufallsexperimente i, $i = 1,\dots,n$, besteht. Wichtiger Spezialfall $(\Omega_i, \mathcal{A}_i, P_i) = (\Omega_0, \mathcal{A}_0, P_0)$, $i = 1,\dots,n$. Dann ist das Gesamtexperiment die n-fache „unabhängige“ Wiederholung des Einzelexperiments.

Beispiel 5.10

Sei

$$(\Omega, \mathcal{A}, P) = (\Omega_0^n, \mathcal{A}_0 \otimes \dots \otimes \mathcal{A}_0, P_0 \otimes \dots \otimes P_0).$$

Im Einzelexperiment $(\Omega_0, \mathcal{A}_0, P_0)$ seien $A_1,\dots,A_s \subset \Omega_0$ Ereignisse mit $A_i \cap A_j = \varnothing$ für $i \neq j$, $\Omega_0 = \sum_{i=1}^{s} A_i$, $X = (X_1,\dots,X_s)$. Sei weiter

$$X_i = \sum_{j=1}^{n} I_{B_i^{(j)}} \text{ mit } B_i^{(j)} = \underbrace{\Omega_0 \times \dots \times \Omega_0}_{j-1\text{-mal}} \times A_i \times \underbrace{\Omega_0 \times \dots \times \Omega_0}_{n-j\text{-mal}}$$

(beschreibt im Gesamtexperiment das Ereignis, dass bei der j-ten Wiederholung A_i eintritt). X_i ist die Anzahl der Wiederholungen, bei denen A_i eintritt. Wir suchen

$$P(X = (k_1,\dots,k_s)) = P(X_1 = k_1,\dots,X_s = k_s),\ (k_1,\dots,k_s) \in \mathbb{N}_0^s,\ k_1 + \dots + k_s = n.$$

Es gilt:

$$\sum_{\substack{(l_1,\dots,l_n) \in \{1,\dots,s\}^n \\ |\{j | l_j = i\}| = k_i, i=1,\dots,s}} \underbrace{P(A_{l_1} \times \dots \times A_{l_n})}_{=P_0(A_{l_1}) \cdot \dots \cdot P_0(A_{l_n})} = \sum_{i=1}^{\frac{n!}{k_1! \cdot \dots \cdot k_s!}} p_1^{k_1} \cdot \dots \cdot p_s^{k_s} = \frac{n!}{k_1! \cdot \dots \cdot k_s!} p_1^{k_1} \cdot \dots \cdot p_s^{k_s}.$$

Definition 5.11

Ein s-dimensionaler Zufallsvektor X heißt **multinomialverteilt** mit den Parametern $n \in \mathbb{N}$ und $p_1,\dots,p_s \in (0,1)$ mit $p_1 + \dots + p_s = 1$, wenn gilt:

$$P(X_1 = k_1,\dots,X_s = k_s) = \frac{n!}{k_1! \cdot \dots \cdot k_s!} p_1^{k_1} \cdot \dots \cdot p_s^{k_s} \text{ mit } (k_1,\dots,k_s) \in \mathbb{N}_0^s,\ k_1 + \dots + k_s = n.$$

Schreibweise: $X \sim \mathcal{M}(n; p_1,\dots,p_s)$.

Sei $(\Omega,\mathcal{A},P)$ ein Wahrscheinlichkeitsraum, $(R_i,\mathcal{S}_i)$, $i=1,\dots,n$, Messräume, $X_i:(\Omega,\mathcal{A})\to(R_i,\mathcal{S}_i)$, $i=1,\dots,n$, Zufallsvariablen. Sei weiter $X=(X_1,\dots,X_n)$ mit $X:(\Omega,\mathcal{A})\to(R_1\times\dots\times R_n,\mathcal{S}_1\otimes\dots\otimes\mathcal{S}_n)$ und

$$P^X = P^{X_1}\otimes\dots\otimes P^{X_n}.$$

Es gilt $P^X(B_1\times\dots\times B_n)=P^{X_1}(B_1)\cdot\ldots\cdot P^{X_n}(B_n)$ genau dann, wenn

$$P(X_1\in B_1,\dots,X_n\in B_n)=P(X_1\in B_1)\cdot\ldots\cdot P(X_n\in B_n) \text{ für jede Wahl } B_i\in\mathcal{S}_i,\ i=1,\dots,n.$$

Also sind $X_1,\dots,X_n$ unabhängig.

Definition 5.12

Die Verteilung von X heißt die **gemeinsame** Verteilung von $X_1,\dots,X_n$. Für $1\le i_1,\dots,i_k\le n$, $1\le k\le n$, heißt die Verteilung von $(X_{i_1},\dots,X_{i_k})$ die **Randverteilung** von $(X_1,\dots,X_n)$. Die Verteilung der X_i, $i=1,\dots,n$, heißt die **1-dim-Randverteilung** von $(X_1,\dots,X_n)$.

Die $X_1,\dots,X_n$ sind genau dann unabhängig, wenn die gemeinsame Verteilung von $X_1,\dots,X_n$ das Produkt ihrer 1-dim-Randverteilung ist.

Für jedes $i=1,\dots,n$ sei $\mathcal{E}_i$ ein $\cap$-stabiles Erzeugendensystem von $\mathcal{S}_i$. Es gilt also $\mathcal{S}_i=\sigma(\mathcal{E}_i)$ und $E\cap F\in\mathcal{E}_i$ für jedes $E,F\in\mathcal{E}_i$. Ferner gebe es für jedes $i=1,\dots,n$ eine Folge von Mengen $E_{i_k}\in\mathcal{E}_i$, $k\in\mathbb{N}$, mit $\bigcup_{k\in\mathbb{N}}E_{i_k}=R_i$. Dann sind $X_1,\dots,X_n$ genau dann unabhängig, wenn für alle $E_i\in\mathcal{E}_i$, $i=1,\dots,n$, gilt:

$$P(X_1\in E_1,\dots,X_n\in E_n)=P(X_1\in E_1)\cdot\ldots\cdot P(X_n\in E_n).$$

Betrachten wir den Spezialfall $(R_i,\mathcal{S}_i)=(\mathbb{R},\mathcal{B})$ für $i=1,\dots,n$, $X=(X_1,\dots,X_n)$ n-dimensionaler Zufallsvektor, $\mathcal{S}_i=\mathcal{B}=\sigma(\mathcal{E})$ mit $\mathcal{E}=\{(-\infty,a]\mid a\in\mathbb{R}\}$. Dann sind $X_1,\dots,X_n$ genau dann stochastisch unabhängig, wenn

$$P(X_1\le x_1,\dots,X_n\le x_n)=P(X_1\le x_1)\cdot\ldots\cdot P(X_n\le x_n)=F(x_1,\dots,x_n)=F_1(x_1)\cdot\ldots\cdot F_n(x_n)$$

für alle $(x_1,\dots,x_n)\in\mathbb{R}^n$ mit F als Verteilungsfunktion von X_i, $i=1,\dots,n$.

Satz 5.13

Sei μ_i ein σ-endliches Maß auf dem Messraum $(R_i,\mathcal{S}_i)$, $i=1,\dots,n$, und sei $f_i:(R_i,\mathcal{S}_i)\to(\mathbb{R},\mathcal{B})$ mit $f\ge 0$ und $\int f_i\,d\mu_i=1$ für $i=1,\dots,n$. Dann ist $f:R_1\times\dots\times R_n\to\mathbb{R}$ definiert durch

$$f(x_1,\dots,x_n)=\prod_{i=1}^n f_i(x_i) \text{ für } (x_1,\dots,x_n)\in R_1\times\dots\times R_n$$

und ist $(\mathcal{S}_1\otimes\dots\otimes\mathcal{S}_n)$-messbar. Ferner gilt:

(i) Hat X_i die μ_i-Dichte f_i, d.h. gilt

$$P(X_i\in B_i)=\int_{B_i}f_i\,d\mu_i \text{ für alle } B_i\in\mathcal{S}_i,\ i=1,\dots,n,$$

so hat $X=(X_1,\dots,X_n)$ im Fall der Unabhängigkeit der $X_1,\dots,X_n$ die $\mu_1\otimes\dots\otimes\mu_n$-Dichte f, d.h. es gilt

$$P(X\in B)=\int_B f\,d\mu_1\otimes\dots\otimes\mu_n \text{ für alle } B\in\mathcal{S}_1\otimes\dots\otimes\mathcal{S}_n).$$

(ii) Hat $X=(X_1,\dots,X_n)$ die $\mu_1\otimes\dots\otimes\mu_n$-Dichte f, so sind $X_1,\dots,X_n$ unabhängig.

Beweis

Die Messbarkeit von f ist klar. Unter den Voraussetzungen in (i) oder (ii) gilt:

$$\begin{aligned}
P(X_1\in B_1,\dots,X_n\in B_n) &= P(X_1\in B_1)\cdot\ldots\cdot P(X_n\in B_n)\\
&=\int_{B_1}f_1\,d\mu_1\cdot\ldots\cdot\int_{B_n}f_n\,d\mu_n\\
&=\prod_{i=1}^n\int f_iI_{B_i}\,d\mu_i\\
&=\int\prod_{i=1}^n f_i(x_i)I_{B_i}(x_i)\,d\mu_1\otimes\dots\otimes\mu_n(x_1,\dots,x_n)\\
&=\int_{B_1\times\dots\times B_n}f\,d\mu_1\otimes\dots\otimes d\mu_n
\end{aligned}$$

für jede Wahl $B_i\in\mathcal{S}_i$, $i=1,\dots,n$. Ferner gilt

$$\int f\,d\mu_1\otimes\dots\otimes\mu_n=\int\prod_{i=1}^n f_i\,d\mu_1\otimes\dots\otimes\mu_n=\prod_{i=1}^n\int f_i\,d\mu_i=1.$$

Also sind die $X_1,\dots,X_n$ genau dann unabhängig, wenn die $\mu_1\otimes\dots\otimes\mu_n$-Dichte der gemeinsamen Verteilung das Produkt der Dichten ihrer 1-dim-Randverteilung ist.

Seien $(R_i, \mathcal{S}_i) = (\mathbb{R}, \mathcal{B})$, $i = 1,\dots,n$, $X = (X_1,\dots,X_n)$ ein n-dimensionaler Zufallsvektor und sei $X_i \sim \mathcal{N}(0,1)$, $i = 1,\dots,n$. Die $X_1,\dots,X_n$ seien unabhängig. Dann hat X die $\boldsymbol{\lambda}^n = \boldsymbol{\lambda} \otimes \dots \otimes \boldsymbol{\lambda}$-Dichte:

$$f(x_1,\dots,x_n) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} x_i^2\right) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2}\sum\nolimits_{i=1}^n x_i^2\right) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2}|x|^2\right)$$

mit $|x| = \sqrt{\sum\nolimits_{i=1}^n x_i^2}$ als euklidische Norm von $x = (x_1,\dots,x_n) \in \mathbb{R}^n$.
Sei $A \in \mathbb{R}^{n\times n}$ eine reguläre $n\times n$-Matrix, $b \in \mathbb{R}^n$, $Y = AX + b$. Die Anwendung der Transformationsformel liefert, dass Y die $\boldsymbol{\lambda}^n$-Dichte g hat mit

$$\begin{aligned} g(y) &= \frac{1}{|\det A|} f\left(A^{-1}(y-b)\right) \\ &= \frac{1}{|\det A|}\left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2}(y-b)^\intercal (A^{-1})^\intercal A^{-1}(y-b)\right) \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{\sqrt{\det\Sigma}} \exp\left(-\frac{1}{2}(y-b)^\intercal \Sigma^{-1}(y-b)\right) \end{aligned}$$

mit $\Sigma := AA^\intercal$, $|X|^2 = X^\intercal X$, $y \in \mathbb{R}^n$.

Definition 5.14 ($\boldsymbol{\lambda}^n$-Dichte von Y)

Ein n-dimensionaler Zufallsvektor Y mit der $\boldsymbol{\lambda}^n$-Dichte

$$\left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{\sqrt{\det\Sigma}} \exp\left(-\frac{1}{2}(y-b)^\intercal \Sigma^{-1}(y-b)\right),\ y \in \mathbb{R}^n,$$

heißt n**-dimensional normalverteilt** mit den Parametern $b \in \mathbb{R}^n$ und $\Sigma \in \mathbb{R}^{n\times n}$. Σ ist symmetrisch und positiv definit.
Schreibweise: $Y \sim \mathcal{N}_n(b,\Sigma)$.

Definition 5.15

Sei $X = (X_1,\dots,X_n)$ ein n-dimensionaler Zufallsvektor.

(i) Gilt $E(|X_i|) < \infty$, $i = 1,\dots,n$, so heißt

$$E(X) = \begin{pmatrix} E(X_1) \\ \vdots \\ E(X_n) \end{pmatrix}$$

der **Erwartungswertvektor** von X.

(ii) Gilt $E\left(X_i^2\right) < \infty$, $i = 1,\dots,n$, so heißt

$$\mathrm{Cov}(X) := \begin{pmatrix} \mathrm{Cov}(X_1,X_1) & \cdots & \mathrm{Cov}(X_1,X_n) \\ \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_n,X_1) & \cdots & \mathrm{Cov}(X_n,X_n) \end{pmatrix}$$

die **Kovarianzmatrix** von X.

Eigenschaften von Erwartungswertvektoren und Kovarianzmatrizen

(1) Sei $X = (X_1,\dots,X_n)$, $B \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$. Dann gilt:

$$E(BX+b) = BE(X) + b,$$

$$\mathrm{Cov}(BX+b) = \mathrm{Cov}(BX) = B\,\mathrm{Cov}(X)B^\intercal.$$

(2) Für $B = (z_1,\dots,z_n) \in \mathbb{R}^{1\times n}$ ergibt sich

$$0 \le \mathrm{Var}(zX) = z\,\mathrm{Cov}(X)z^\intercal.$$

Also ist die Kovarianz von X positiv semidefinit.

Folgerung 5.16

Sei $X = (X_1,\dots,X_n) \sim \mathcal{N}_n(0, I_n)$ mit I_n als $n\times n$-Einheitsmatrix. Dann sind $X_1,\dots,X_n$ unabhängig und es gilt $E(X) = 0$ und $\mathrm{Cov}(X) = I_n$.
Aus $Y = AX + b$ folgt $Y \sim \mathcal{N}_n(b,\Sigma)$ mit $b = E(Y)$ und $\Sigma = \mathrm{Cov}(Y)$.

Satz 5.17

Seien X, Y unabhängige reelle Zufallsvariablen mit $E(|X|) < \infty$ und $E(|Y|) < \infty$. Dann gilt:

$$E(XY) = E(X)E(Y).$$

Die Gleichung gilt auch ohne Voraussetzung $E(|X|) < \infty$ und $E(|Y|) < \infty$, falls $X,Y \geq 0$.

Beweis

Seien $X,Y \geq 0$. Dann gilt:

$$\begin{aligned} E(XY) &= \int xy\, dP^{(X,Y)}(x,y) = \int xy\, dP^X \otimes P^Y(x,y) \\ &= \int x\, dP^X(x) \int y\, dP^Y(y) = E(X)E(Y). \end{aligned} \tag{5.6}$$

Es seien $E(|X|) < \infty$ und $E(|Y|) < \infty$. Dann wende man (5.6) auf $|X|$ und $|Y|$ an. □

Folgerung 5.18

Seien X,Y reelle unabhängige Zufallsvariablen mit $E(X^2) < \infty$ und $E(Y^2) < \infty$. Dann gilt:

$$\operatorname{Cov}(X,Y) = E(XY) - E(X)E(Y) = 0.$$

Damit folgt: $\operatorname{Var}(X+Y) = \operatorname{Var}(X) + \operatorname{Var}(Y)$.

Beispiel 5.19

Sei $X \sim \mathcal{B}(n,p)$. Dann gilt:

$$\operatorname{Var}(X) = E([X - E(X)]^2) = E([X - np]^2) = \sum_{k=0}^{n} (k - np)^2 \binom{n}{k} p^k (1-p)^{n-k}.$$

Sei dazu $X = \sum_{i=1}^n I_{A_i}$ mit unabhängigen Ereignissen $A_1,\dots,A_n$ und $p = P(A_i)$. Dann erhalten wir:

$$\operatorname{Var}(X) = \sum_{\nu=1}^{n} \operatorname{Var}(I_{A_\nu}) = \sum_{i=1}^{n} \big(E(I_{A_i}) - (E(I_{A_i}))^2 \big) = \sum_{i=1}^{n} p(1-p) = np(1-p).$$

Sei $(\Omega,\mathcal{A},P)$ ein Wahrscheinlichkeitsraum, $(R_i,\mathcal{S}_i)$, $i = 1,\dots,n$, Messräume. Seien $X_i : (\Omega,\mathcal{A}) \to (R_i,\mathcal{S}_i)$, $i = 1,\dots,n$, $(R,\mathcal{S}) = (R_1 \times \dots \times R_n, \mathcal{S}_1 \otimes \dots \otimes \mathcal{S}_n)$, $X = (X_1,\dots,X_n)$, $X : (\Omega,\mathcal{A}) \to (R,\mathcal{S})$. Seien $\{x\} \in \mathcal{S}_i$ für jedes $x \in R_i$ und $i = 1,\dots,n$. Die X_i seien in $\mathcal{S}_i$ diskret verteilt, d.h. zu jedem $i = 1,\dots,n$ existiert eine abzählbare Teilmenge $A_i \subset R_i$ mit der Eigenschaft $P(X_i \in A_i) = 1$.

Sei $\mu_i = \sum_{a \in A_i} \delta_a$. Dann ist μ_i σ-endliches Maß auf $(R_i,\mathcal{S}_i)$. Dann hat P^{X_i} die μ_i-Dichte

$$f_i(x) = \begin{cases} P(X_i = x), & x \in A_i, \\ 0, & \text{sonst.} \end{cases}$$

Denn für $B \in \mathcal{S}_i$ gilt:

$$P^{X_i}(B) = P(X_i \in B \cap A_i) + P(X_i \in B \cap A_i^c) = \sum_{x \in B \cap A_i} P(X_i = x) = \int_B f_i(x)\, d\mu_i.$$

Damit sind die $X_1,\dots,X_n$ genau dann unabhängig, wenn die $\mu_1 \otimes \dots \otimes \mu_n$-Dichte von X, also die Funktion $f(x_1,\dots,x_n) = P(X_1 = x_1,\dots,X_n = x_n)$, von der Form

$$f(x_1,\dots,x_n) = \prod_{i=1}^{n} f_i(x_i) \text{ für jedes } (x_1,\dots,x_n) \in R_1 \times \dots \times R_n$$

ist. Also sind die $X_1,\dots,X_n$ genau dann unabhängig, wenn für alle $(x_1,\dots,x_n) \in R_1 \times \dots \times R_n$ gilt:

$$P(X_1 = x_1,\dots,X_n = x_n) = P(X_1 = x_1) \cdot \ldots \cdot P(X_n = x_n).$$

Bemerkung 5.20

(i) Sind $(\Omega_i,\mathcal{A}_i,P_i)$, $i = 1,\dots,n$, Modelle für n Einzelexperimente. So ist

$$(\Omega_1 \times \dots \times \Omega_n, \mathcal{A}_1 \otimes \dots \otimes \mathcal{A}_n, P_1 \otimes \dots \otimes P_n)$$

ein Modell für das Gesamtexperiment, das aus der unabhängigen Durchführung der Einzelexperimente besteht.

(ii) Seien $(R_i,\mathcal{S}_i,P_i)$, $i = 1,\dots,n$, Wahrscheinlichkeitsräume. Dann gibt es einen Wahrscheinlichkeitsraum $(\Omega,\mathcal{A},P)$ und unabhängige Zufallsvariablen $X_i : (\Omega,\mathcal{A}) \to (R_i,\mathcal{S}_i)$ mit $P^{X_i} = P_i$, $i = 1,\dots,n$.
Der Raum $(\Omega,\mathcal{A},P) = (R_1 \times \dots \times R_n, \mathcal{S}_1 \otimes \dots \otimes \mathcal{S}_n, P_1 \otimes \dots \otimes P_n)$ und die Projektionen

$$X_i : R_1 \times \dots \times R_n \to R_i,\ X_i(x_1,\dots,x_n) = x_i \text{ für } (x_1,\dots,x_n) \in R_1 \times \dots \times R_n,\ i = 1,\dots,n,$$

leisten das Verlangte.

(iii) Die Aussage in (ii) lässt sich verallgemeinern auf den Fall beliebig vieler Wahrscheinlichkeitsräume $(R_i,\mathcal{S}_i,P_i)$, $i \in I$.

(iv) Wegen $P^{X_1} \otimes ... \otimes P^{X_n} = (P^{X_1} \otimes ... \otimes P^{X_m}) \otimes (P^{X_{m+1}} \otimes ... \otimes P^{X_n})$, $1 \le m < n$, folgt aus der Unabhängigkeit von $X_1,...,X_n$ die Unabhängigkeit von $(X_1,...,X_m)$ und $(X_{m+1},...,X_n)$. Also sind damit $g(X_1,...,X_m)$ und $h(X_{m+1},...,X_n)$ unabhängig mit

$$g : (R_1 \times ... \times R_m, \mathcal{S}_1 \otimes ... \otimes \mathcal{S}_m) \to (S_1, \mathcal{Z}_1)$$

und

$$h : (R_{m+1} \times ... \times R_n, \mathcal{S}_{m+1} \otimes ... \otimes \mathcal{S}_n) \to (S_2, \mathcal{Z}_2).$$

Allgemein gilt: Sind die X_i, $i \in I$, unabhängige Zufallsvariablen und $I = \sum_{i \in K} I_i$ mit paarweise disjunkten $I_i \subset I$, $i \in K$. Dann sind $(X_j)_{j \in I_i}$, $i \in K$, unabhängig.

Unendliche Produkte von Wahrscheinlichkeitsräumen

Zum Beweis der Existenz von unendlich dimensionalen Produktmaßen benötigen wir die folgende Hilfsaussage.

Lemma 5.21

Es sei $\mu : \mathcal{R} \to [0,\infty)$ eine endlich additive Mengenfunktion auf dem Ring $\mathcal{R}$ auf Ω. Ist μ $\varnothing$-stetig in dem Sinne, dass für jede Folge $A_n \in \mathcal{A}$, $n \in \mathbb{N}$, $A_n \downarrow \varnothing$ gilt $\lim_{n\to\infty} \mu(A_n) = 0$, so ist μ σ-additiv.

Beweis

Für paarweise disjunkte Mengen $A_j \in \mathcal{R}$, $j \in \mathbb{N}$, mit $A = \sum_{j=1}^{\infty} A_j \in \mathcal{R}$ ist $A \cap \left(\sum_{j=1}^{n} A_j\right)^c \downarrow \varnothing$ für $n \to \infty$, also

$$0 = \lim_{n\to\infty} \mu\left(A \cap \left(\sum_{j=1}^{n} A_j\right)^c\right) = \mu(A) - \lim_{n\to\infty} \mu\left(\sum_{j=1}^{n} A_j\right) = \mu(A) - \lim_{n\to\infty} \sum_{j=1}^{n} \mu(A_j)$$
$$= \mu(A) - \sum_{j=1}^{\infty} \mu(A_j).$$

□

Es seien $(\Omega_i, \mathcal{A}_i, P_i)$, $i \in \mathbb{N}$, Wahrscheinlichkeitsräume. Es sei $\Omega = \times_{i=1}^{\infty} \Omega_i$ das kartesische Produkt der Mengen Ω_i. Die Elemente von Ω schreiben wir in der Form $(\omega_1, \omega_2, ...)$ mit $\omega_i \in \Omega_i$. Für jedes $i \in \mathbb{N}$ sei $\pi_i : \Omega \to \Omega_i$ die natürliche Projektion von Ω auf Ω_i. Für jedes $n \in \mathbb{N}$ sei $\pi_{1-n} : \Omega \to \times_{i=1}^{n} \Omega_i$ die natürliche Projektion von Ω auf $\times_{i=1}^{n} \Omega_i$. Offenbar ist

$$\mathcal{A}_0 = \bigcup_{n=1}^{\infty} \pi_{1-n}^{-1}\left(\bigotimes_{i=1}^{n} \mathcal{A}_i\right)$$

eine Algebra auf Ω. Das ist ein Ring auf Ω mit der Eigenschaft, dass auch $\Omega \in \mathcal{A}_0$ ist. Die von dieser Algebra erzeugte σ-Algebra,

$$\bigotimes_{i=1}^{\infty} \mathcal{A}_i = \sigma(\mathcal{A}_0) = \sigma\left(\bigcup_{i=1}^{\infty} \pi_i^{-1}(\mathcal{A}_i)\right)$$

heißt **Produkt-σ-Algebra** auf Ω.

Sind $m, n \in \mathbb{N}$, $m < n$, und gilt für zwei Mengen $B \in \bigotimes_{i=1}^{m} \mathcal{A}_i$ und $C \in \bigotimes_{i=1}^{n} \mathcal{A}_i$ die Identität

$$\pi_{1-m}^{-1}(B) = \pi_{1-n}^{-1}(C),$$

so ist $C = B \times \Omega_{m+1} \times ... \times \Omega_n$ und daher $\bigotimes_{i=1}^{m} P_i(B) = \bigotimes_{i=1}^{n} P_i(C)$. Durch $P(A) = \bigotimes_{i=1}^{m} P_i(B)$ für $A \in \mathcal{A}_0$ mit $A = \pi_{1-m}^{-1}(B)$, $B \in \bigotimes_{i=1}^{m} \mathcal{A}_i$, ist somit eine Mengenfunktion P auf $\mathcal{A}_0$ wohldefiniert. Aufgrund dieser Definition ist aber auch klar, dass P endlich additiv auf $\mathcal{A}_0$ ist. Wir zeigen jetzt, dass P $\varnothing$-stetig, also auch σ-additiv auf $\mathcal{A}_0$ ist. Sei dazu $A_n \downarrow \varnothing$, $A_n = \pi_{1-n}^{-1}(B_n)$ mit $B_n \in \bigotimes_{i=1}^{n} \mathcal{A}_i$, $n \in \mathbb{N}$. Den Darstellungen

$$A_n = B_n \times \Omega_{n+1} \times ..., \; A_{n+1} = B_{n+1} \times \Omega_{n+2} \times ...$$

und der Beziehung $A_{n+1} \subset A_n$ entnehmen wir, dass

$$B_{n+1} \subset B_n \times \Omega_{n+1} \tag{5.7}$$

ist. Angenommen, es ist

$$\alpha = \inf_{n\in\mathbb{N}} P(A_n) = \inf_{n\in\mathbb{N}} \bigotimes_{i=1}^{n} P_i(B_n) > 0. \tag{5.8}$$

Mit

$$Q_n^1 = \left\{\omega_1 \in \Omega_1 \mid \bigotimes_{i=2}^{n} P_i(B_{n,\omega_1}) \ge \frac{\alpha}{2}\right\}, \; n \ge 2,$$

ist

$$\alpha \le \bigotimes_{i=1}^{n} P_i(B_n) \le \int_{Q_n^1} \bigotimes_{i=2}^{n} P_i(B_{n,\omega_1})\, dP_1(\omega_1) + \frac{\alpha}{2} \le P_1(Q_n^1) + \frac{\alpha}{2},$$

also

$$P_1(Q_n^1) \ge \frac{\alpha}{2} \text{ für jedes } n \ge 2.$$

Aus der Inklusion (5.7) folgt

$$B_{n+1,\omega_1} \subset B_{n,\omega_1} \times \Omega_{n+1},$$

was

$$\bigotimes_{i=2}^{n+1} P_i(B_{n+1,\omega_1}) \le \bigotimes_{i=2}^{n} P_i(B_{n,\omega_1}),$$

also $Q_n^1 \downarrow \bigcap_{n=2}^{\infty} Q_n^1$ und damit

$$P_1\Big(\bigcap_{n=2}^{\infty} Q_n^1\Big) \ge \frac{\alpha}{2}$$

zur Folge hat. Also gibt es ein $\omega_1 \in \bigcap_{n=2}^{\infty} Q_n^1$ mit

$$\inf_{n\in\mathbb{N}, n\ge 2} \bigotimes_{i=2}^{n} P_i(B_{n,\omega_1}) \ge \frac{\alpha}{2}. \tag{5.9}$$

Insbesondere ist $\omega_1 \in B_1$. Definieren wir

$$Q_n^2 = \Big\{\omega_2 \in \Omega_2 \mid \bigotimes_{i=3}^{n} P_i\big((B_{n,\omega_1})_{\omega_2}\big) \ge \frac{\alpha}{4}\Big\},\ n \ge 3,$$

so erhalten wir aus (5.9) mit denselben Überlegungen wie zuvor aus (5.8), dass ein $\omega_2 \in \bigcap_{n=3}^{\infty} Q_n^2$ mit

$$\inf_{n\in\mathbb{N}, n\ge 3} \bigotimes_{i=3}^{n} P_i\big((B_{n,\omega_1})_{\omega_2}\big) \ge \frac{\alpha}{4},$$

also insbesondere $(\omega_1,\omega_2) \in B_2$ existiert. Mittels vollständiger Induktion erhalten wir auf diese Weise eine Folge von Elementen $\omega_i \in \Omega_i$, $i \in \mathbb{N}$, mit $(\omega_1,\dots,\omega_n) \in B_n$ für jedes $n \in \mathbb{N}$. Also ist $\bigcap_{n=1}^{\infty} A_n \ne \varnothing$, im Widerspruch zur Voraussetzung.
Mit dem Maßerweiterungssatz folgt die Existenz und Eindeutigkeit des Produktmaßes auf $\bigotimes_{i=1}^{\infty} \mathcal{A}_i$.

Satz 5.22

Es gibt genau ein Wahrscheinlichkeitsmaß P auf $\big(\times_{i=1}^{\infty}\Omega_i, \bigotimes_{i=1}^{\infty}\mathcal{A}_i\big)$ mit der Eigenschaft, dass für jedes $n \in \mathbb{N}$ gilt

$$P^{\pi_{1-n}} = \bigoplus_{i=1}^{n} P_i.$$

Wir schreiben $P = \bigotimes_{i=1}^{\infty} P_i$ und nennen P das **Produktmaß** der P_i, $i \in \mathbb{N}$.

Korollar 5.23

Ist für jedes $n \in \mathbb{N}$ ein Wahrscheinlichkeitsmaß Q_n auf einem Messraum $(R_n, \mathcal{S}_n)$ gegeben, so existieren ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ und unabhängige Zufallsvariablen $X_n : (\Omega, \mathcal{A}) \to (R_n, \mathcal{S}_n)$, $n \in \mathbb{N}$, mit der Eigenschaft, dass $P^{X_n} = Q_n$ für jedes $n \in \mathbb{N}$ ist.

Beweis

$\Omega = \times_{n=1}^{\infty} R_n$, $\mathcal{A} = \bigotimes_{n=1}^{\infty} \mathcal{S}_n$, $P = \bigotimes_{n=1}^{\infty} P_n$ und $X_n = \pi_n$, die natürliche Projektion von $\times_{n=1}^{\infty} R_n$ auf R_n, $n \in \mathbb{N}$, leisten das Verlangte. □

Ist $I \ne \varnothing$ eine beliebige nicht abzählbare Indexmenge und ist $(\Omega_i, \mathcal{A}_i, P_i)$ ein Wahrscheinlichkeitsraum für jedes $i \in I$, so erhalten wir mit

$$\Omega = \times_{i\in I} \Omega_i = \Big\{\omega \mid \omega : I \to \bigcup_{i\in I} \Omega_i \text{ mit } \omega(i) \in \Omega_i \text{ für alle } i \in I\Big\}$$

und

$$\mathcal{A} = \sigma\Big(\bigcup_{i\in I} \pi_i^{-1}(\mathcal{A}_i)\Big)$$

mit $\pi_i : \Omega \to \Omega_i$, $\pi_i(\omega) = \omega(i)$, $\omega \in \Omega$, als natürliche Projektion von Ω auf Ω_i, $i \in I$, einen neuen Messraum $(\Omega, \mathcal{A})$. Wir bezeichnen für eine beliebige nichtleere Teilmenge $K \subset I$ mit π_K die natürliche Projektion von $\times_{i\in I}\Omega_i$ auf $\times_{i\in K}\Omega_i$ und für beliebige nichtleere Teilmengen $K \subset J \subset I$ mit π_K^J die natürliche Projektion von $\times_{i\in J}\Omega_i$ auf $\times_{i\in K}\Omega_i$. Aus

$$\mathcal{A} = \bigcup_{\substack{\varnothing \ne J \subset I,\\ J \text{ abzählbar}}} \pi_J^{-1}\big(\bigotimes_{i\in J} \mathcal{A}_i\big)$$

(Übungsaufgabe!) folgt für jedes $A \in \mathcal{A}$ die Existenz einer nichtleeren abzählbaren Menge $J \subset I$ und die einer Menge $A_J \in \bigotimes_{i\in J}\mathcal{A}_i$ mit $A = \pi_J^{-1}(A_J)$. Nach Satz 5.22 haben wir ein Produktmaß $\bigotimes_{i\in J}P_i$ auf dem Messraum $(\times_{i\in J}\Omega_i, \bigotimes_{i\in J}\mathcal{A}_i)$. Mit der durch $P(A) = \bigotimes_{i\in J}P_i(A_J)$ Wohldefinierten Mengenfunktion auf $\mathcal{A}$ (Übungsaufgabe!) liegt ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{A})$ vor. Da für abzählbares $J \subset I$ das Maß $\bigotimes_{i\in J}P_i$ eindeutig festgelegt ist durch seine Werte auf der Algebra

$$\bigcup_{\substack{\varnothing \neq H \subset J,\\ H \text{ endlich}}} (\pi_H^J)^{-1}(\bigoplus_{i\in H}\mathcal{A}_i),$$

ist insbesondere auch P festgelegt durch seine Werte auf der Algebra

$$\bigcup_{\substack{\varnothing \neq H \subset J,\\ H \text{ endlich}}} \pi_H^{-1}(\bigotimes_{i\in H}\mathcal{A}_i).$$

Damit haben wir die folgenden Verallgemeinerungen von Satz 5.22 und Korollar 5.23.

Satz 5.24

Es gibt genau ein Wahrscheinlichkeitsmaß P auf $(\times_{i\in I}\Omega_i, \bigotimes_{i\in I}P_i)$ mit der Eigenschaft, dass für jede nichtleere endliche Menge $H \subset I$ gilt

$$P^{\pi_H} = \bigotimes_{i\in H}P_i.$$

Korollar 5.25

Ist für jedes $i \in I$ ein Maß Q_i auf einem Messraum $(R_i, \mathcal{S}_i)$ gegeben, so existieren ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ und unabhängige Zufallsvariablen $X_i : (\Omega, \mathcal{A}) \to (R_i, \mathcal{S}_i)$, $i \in I$, mit der Eigenschaft, dass $P^{X_i} = Q_i$ für jedes $i \in \mathbb{N}$ ist.

Kapitel 6

Gesetze der Großen Zahlen

Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum.

Lemma 6.1 (Borel-Cantelli)
Es seien $A_n \in \mathcal{A}$, $n \in \mathbb{N}$, Ereignisse:
(i) Aus $\sum_{i=1}^{\infty} P(A_i) < \infty$ folgt $P(\limsup_{n\to\infty} A_n) = 0$, wobei
$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$$
bedeutet: „Unendlich viele der A_n, $n \in \mathbb{N}$, treten ein".
(ii) Sind die A_n, $n \in \mathbb{N}$, unabhängig und ist $\sum_{i=1}^{\infty} P(A_i) = \infty$, so ist $P(\limsup_{n\to\infty} A_n) = 1$.

Beweis
(i) Es gilt:
$$P\Big(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k\Big) = \lim_{n\to\infty} P\Big(\bigcup_{k=n}^{\infty} A_k\Big) \le \lim_{n\to\infty} \sum_{k=n}^{\infty} P(A_k) = 0\,.$$
(ii) Es gilt:
$$\begin{aligned} P(\limsup_{n\to\infty} A_n) &= 1 - P\Big(\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k^c\Big) = 1 - \lim_{n\to\infty} P\Big(\bigcap_{k=n}^{\infty} A_k^c\Big) \\ &= 1 - \lim_{n\to\infty}\lim_{m\to\infty} \prod_{k=n}^{m} P(A_k^c) = 1 - \lim_{n\to\infty}\lim_{m\to\infty} \prod_{k=n}^{m} (1 - P(A_k)) \\ &\ge 1 - \liminf_{n\to\infty}\liminf_{m\to\infty} \prod_{k=n}^{m} \exp(-P(A_k)) \\ &= 1 - \liminf_{n\to\infty}\liminf_{m\to\infty} \exp\Big(-\sum_{k=n}^{m} P(A_k)\Big) = 1, \end{aligned}$$
da $1 - x \le e^{-x}$ für $x \ge 0$. Man beachte:
$$\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k^c = \liminf_{n\to\infty} A_n^c$$
bedeutet: „Fast alle A_n^c, $n \in \mathbb{N}$, treten ein". □

Definition 6.2
(i) Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, X, X_n, $n \in \mathbb{N}$, seien d-dimensionale Zufallsvektoren. Dann **konvergiert** die Folge $(X_n)_{n\in\mathbb{N}}$ P**-f.s.** gegen X, wenn gilt:
$$P(\lim_{n\to\infty} X_n = X) = 1\,.$$
Schreibweise: $X_n \to X\,[P]$ oder $X_n \to X$ P-f.s. (P-f.ü.).
(ii) Die Folge $(X_n)_{n\in\mathbb{N}}$ **konvergiert stochastisch** gegen X, wenn für alle $\varepsilon > 0$ gilt:
$$\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0\,.$$
Schreibweise: $X_n \xrightarrow{P} X$.

Bemerkung 6.3
(i) Es gilt:
$$\begin{aligned} \{\omega \in \Omega \mid \lim_{n\to\infty} X_n(\omega) = X(\omega)\} &= \bigcap_{k\in\mathbb{N}}\bigcup_{n\in\mathbb{N}}\bigcap_{m=n}^{\infty} \Big\{|X_m - X| < \frac{1}{k}\Big\} \\ &= \bigcap_{\varepsilon>0}\bigcup_{n\in\mathbb{N}}\bigcap_{m=n}^{\infty} \{|X_m - X| < \varepsilon\} \\ &= \bigcap_{\varepsilon>0}\bigcup_{n\in\mathbb{N}}\bigcap_{m=n}^{\infty} \{\omega \in \Omega \mid |X_m(\omega) - X(\omega)| < \varepsilon\} \in \mathcal{A}. \end{aligned}$$
(ii) Sei $M := \{\lim_{n\to\infty} X_n = X\}$. Dann gilt:
$$0 = P(M^c) = P\Big(\bigcup_{\varepsilon>0}\bigcap_{n\in\mathbb{N}}\bigcup_{m=n}^{\infty} \{|X_m - X| \ge \varepsilon\}\Big).$$
Für alle $\varepsilon > 0$ folgt:

$$0 = P\Big(\bigcap_{n\in\mathbb{N}}\bigcup_{m=n}^{\infty}\{|X_m - X| \geq \varepsilon\}\Big) = \lim_{n\to\infty} P\Big(\bigcup_{m=n}^{\infty}\{|X_m - X| \geq \varepsilon\}\Big)$$
$$\geq \limsup_{n\to\infty} P(\{|X_n - X| \geq \varepsilon\}).$$

Also gilt $X_n \to X[P]$ und somit folgt $X_n \xrightarrow{P} X$.

Satz 6.4 (Ungleichung von Kolmogorov)
Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, $X_i : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B})$ seien unabhängig mit $E(|X_i|^2) < \infty$, $E(X_i) = 0$, $i = 1,...,n$. Sei weiter $S_k = \sum_{i=1}^{k} X_i$ und $S_0 = 0$. Dann gilt für alle $\varepsilon > 0$:

$$P(\max_{1\leq k\leq n} |S_k| \geq \varepsilon) \leq \frac{1}{\varepsilon^2}\sum_{i=1}^{n}\operatorname{Var}(X_i).$$

Beweis
Sei $A = \{\max_{1\leq k\leq n} |S_k| \geq \varepsilon\}$, $A_k = \{|S_k| \geq \varepsilon \mid |S_i| < \varepsilon \text{ für } i = 1,...,k-1\}$. Dann gilt $A = \sum_{k=1}^{n} A_k$ und

$$\sum_{i=1}^{n}\operatorname{Var}(X_i) = \operatorname{Var}(S_n) = E(S_n^2) \geq E(S_n^2 I_A) = \sum_{k=1}^{n} E(S_n^2 I_{A_k}) = \sum_{k=1}^{n} E\big([S_k + (S_n - S_k)]^2 I_{A_k}\big)$$
$$\geq \sum_{k=1}^{n}\big(E(S_k^2 I_{A_k}) + 2E((S_n - S_k)S_k I_{A_k})\big) = \sum_{k=1}^{n} E\big(\underbrace{S_k^2 I_{A_k}}_{\geq \varepsilon^2 I_{A_k}}\big) \geq \varepsilon^2 \sum_{k=1}^{n} P(A_k) = \varepsilon^2 P(A)$$
$$= \varepsilon^2 P(\max_{1\leq k\leq n} |S_k| \geq \varepsilon),$$

wobei wegen der Unabhängigkeit von $S_n - S_k$ und $S_k I_{A_k}$ gilt: $E((S_n - S_k)S_k I_{A_k}) = E(S_n - S_k)E(S_k I_{A_k})$. □

Satz 6.5
Seien $X_i : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B})$ unabhängig, $E(X_i) = 0$, $i \in \mathbb{N}$. Es gelte $\sum_{i=1}^{\infty}\operatorname{Var}(X_i) < \infty$. Dann konvergiert die Folge $\left(\sum_{i=1}^{n} X_i\right)_{n\in\mathbb{N}}$ P-f.s. gegen eine reelle Zufallsvariable.
Beweis
Für $r \in \mathbb{N}$, $n \in \mathbb{N}$ ist mit $S_m := \sum_{i=1}^{m} X_i$, $m \in \mathbb{N}$,

$$P(\max_{1\leq k\leq r} |S_{n+k} - S_n| > \varepsilon) \leq \frac{1}{\varepsilon^2}\sum_{i=n+1}^{n+r}\operatorname{Var}(X_i) \text{ (Kolmogorovsche Ungleichung).}$$

$r \uparrow \infty$ liefert

$$P(\sup_{k\in\mathbb{N}} |S_{n+k} - S_n| > \varepsilon) \leq \frac{1}{\varepsilon^2}\sum_{i=n+1}^{\infty}\operatorname{Var}(X_i),$$

und daraus folgt

$$P(\inf_{n\in\mathbb{N}}\sup_{k\in\mathbb{N}} |S_{n+k} - S_n| > \varepsilon) = 0 \text{ für jedes } \varepsilon > 0.$$

Also ist die Folge $(S_n)_{n\in\mathbb{N}}$ mit Wahrscheinlichkeit 1 eine Cauchy-Folge und daraus folgt die Behauptung. □

Lemma 6.6 (Kronecker)
Sei a_n, $n \in \mathbb{N}$, eine Folge von reellen Zahlen mit der Eigenschaft, dass $\sum_{k=1}^{\infty}\frac{1}{k}a_k$ konvergiert, d.h. die Folge b_n, $n \in \mathbb{N}$, mit $b_n = \sum_{k=1}^{n}\frac{1}{k}a_k$ konvergiert gegen eine reelle Zahl. Dann gilt:

$$\frac{1}{n}\sum_{k=1}^{n} a_k \to 0.$$

Beweis
Seien $b_0 := 0$, $a_n = b_n - b_{n-1}$, $n \in \mathbb{N}$. Dann gilt:

$$\frac{1}{n}\sum_{k=1}^{n} a_k = \frac{1}{n}\sum_{k=1}^{n} k(b_k - b_{k-1}) = \frac{1}{n}\sum_{k=1}^{n}(kb_k - (k-1)b_{k-1}) - \frac{1}{n}\sum_{k=1}^{n-1} b_k = b_n - \frac{1}{n}\sum_{k=1}^{n-1} b_k \to 0,$$

da $b_n \to b \in \mathbb{R}$ und auch

$$\frac{1}{n}\sum_{k=1}^{n-1} b_k \to b.$$ □

Satz 6.7 (Kolmogorovsches Kriterium)
Seien X_n, $n \in \mathbb{N}$, unabhängige reelle Zufallsvariablen mit $E(X_n) = 0$ für alle $n \in \mathbb{N}$. Des Weiteren sei

$$\sum_{k=1}^{\infty}\frac{1}{k^2}\operatorname{Var}(X_k) < \infty.$$

Dann gilt:

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to 0 \quad P\text{-f.s.}$$

Beweis

Sei $Y_n = \frac{1}{n} X_n$, $n \in \mathbb{N}$. Dann sind die Y_n unabhängig, $E(Y_n) = 0$ und

$$\sum_{n=1}^{\infty} \operatorname{Var}(Y_n) = \sum_{n=1}^{\infty} \tfrac{1}{n^2} \operatorname{Var}(X_n) < \infty .$$

Hieraus folgt, dass $\sum_{n=1}^{\infty} Y_n$ P-f.s. konvergiert und somit gilt nach dem Lemma von Kronecker:

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to 0 \quad P\text{-f.s.}$$

□

Satz 6.8 (Kolmogorov – Das starke Gesetz der großen Zahlen, 1933)

Seien X_n, $n \in \mathbb{N}$, unabhängige reelle Zufallsvariablen, je mit derselben identischen Verteilung und sei $E(|X_1|) < \infty$, $\mu = E(X_1)$. Dann gilt:

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mu \quad P\text{-f.s.}$$

Beweis

O.B.d.A. sei $\mu = 0$ (betrachte sonst $X_i' = X_i - \mu$).

Betrachte $Y_n = X_n I(|X_i| \le n)$. Die Y_n, $n \in \mathbb{N}$, sind unabhängig und es gilt:

$$\sum_{n=1}^{\infty} P(X_n \neq Y_n) = \sum_{n=1}^{\infty} P(|X_1| > n) \le E(|X_1|) < \infty .$$

Aus dem Lemma von Borel-Cantelli (6.1) folgt

$$P(\limsup_{n\to\infty} \{X_n \neq Y_n\}) = 0 .$$

Es genügt zu zeigen, dass $\frac{1}{n}\sum_{i=1}^{n} Y_i \to 0$ P-f.s. ist. Es gilt:

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{1}{n^2} \operatorname{Var}(Y_n) &\le \sum_{n=1}^{\infty} \frac{1}{n^2} E(Y_n^2) = \sum_{n=1}^{\infty} \frac{1}{n^2} E\big(|X_1|^2 I(|X_1| \le n)\big) \\
&= \sum_{n=1}^{\infty} \frac{1}{n^2} E\Big(|X_1|^2 \textstyle\sum_{k=1}^{n} I(k-1 < |X_1| \le k)\Big) \\
&= \sum_{n=1}^{\infty} \frac{1}{n^2} \sum_{k=1}^{n} E(\underbrace{|X_1|}_{<k} |X_1| I(k-1 < |X_1| \le k)) \\
&< \sum_{k=1}^{\infty} \Big(kE(|X_1| I(k-1 < |X_1| \le k)) \textstyle\sum_{n=k}^{\infty} \dfrac{1}{n^2} \Big) \\
&\le 2E(|X_1|) < \infty.
\end{aligned}$$

Aus dem Kolmogorovschen Kriterium (6.7) folgt:

$$\frac{1}{n}\sum_{i=1}^{n} (Y_i - E(Y_i)) \to 0 \quad P\text{-f.s.}$$

□

Noch zu zeigen bleibt:

$$\frac{1}{n}\sum_{i=1}^{n} E(Y_i) \to 0 .$$

Sei $(a_n)_{n\in\mathbb{N}}$ eine Folge reeller Zahlen mit $a_n \to a \in \mathbb{R}$. Dann gilt:

$$\frac{1}{n}\sum_{i=1}^{n} a_i \to a .$$

Dazu genügt es zu zeigen, dass $E(Y_i) \to 0$. Dies folgt aus dem Satz (4.6) von der majorisierten Konvergenz:

$$E(Y_i) = E(X_1 I(|X_1| \le k)) = E(X_1 I(|X_1| \le k)) \to E(X_1) = 0 .$$

□

Bemerkung 6.9

(i) Sind X_n, $n \in \mathbb{N}$, unabhängige und identisch verteilte d-dimensionale Zufallsvektoren, je mit dem Erwartungswertvektor $\mu \in \mathbb{R}^d$, so gilt

$$\frac{1}{n}\sum_{k=1}^{n} X_k \to \mu \quad P\text{-f.s.}$$

(ii) Aus dem starken Gesetz der großen Zahlen (6.8) folgt das **schwache Gesetz der großen Zahlen**:

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{P} \mu \,.$$

Wird zudem $E(X_n^2) < \infty$ vorausgesetzt, so folgt das schwache Gesetz der großen Zahlen aus der Chebyshevschen Ungleichung: Für alle $\varepsilon > 0$ gilt

$$P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\right| > \varepsilon\right) \leq \frac{1}{\varepsilon^2}\operatorname{Var}\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n\varepsilon^2}\operatorname{Var}(X_1) \to 0 \,.$$

(iii) Seien $X_n \sim \mathcal{B}(1,p)$, $n \in \mathbb{N}$, unabhängig. Dann gilt:

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to E(X_1) = p \quad P\text{-f.s.}$$

Anwendung: A_n, $n \in \mathbb{N}$, eine Folge von unabhängigen Ereignissen, $p = P(A_n)$. Dann gilt:

$$\frac{1}{n}\sum_{i=1}^{n} I_{A_i} \to p \quad P\text{-f.s.}$$

(iv) Monte-Carlo-Interpretation: Seien U_n, $n \in \mathbb{N}$, unabhängige und identisch verteilte reelle Zufallsvariablen $U_\nu \sim \mathcal{P}(0,1)$. Sei $f : ([0,1], \mathcal{B}_{[0,1]}) \to (\mathbb{R}, \mathcal{B})$ mit

$$\int_0^1 |f(x)|\, d\boldsymbol{\lambda}_{[0,1]}(x) < \infty \,.$$

Dann sind $X_n = f \cdot U_n$ unabhängige reelle Zufallsvariablen mit

$$E(X_n) = \int_0^1 f(x)\, d\boldsymbol{\lambda}(x) \,.$$

Es gilt

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to \int_0^1 f(x)\, d\boldsymbol{\lambda}(x) \quad P\text{-f.s.}$$

Man verschaffe sich Beobachtungen $u_1,\ldots,u_n$ aus $U_1,\ldots,U_n$ und bilde $X_i = f(u_i)$, $i = 1,\ldots,n$. Verwende

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}\sum_{i=1}^{n} f(u_i)$$

dann als Approximation für

$$\int_0^1 f(x)\, d\boldsymbol{\lambda}(x) \,.$$

Kapitel 7

Nichtnegative ganzzahlige Zufallsvariablen und erzeugende Funktionen

Definition 7.1
Sei X eine $\mathbb{N}_0$-wertige Zufallsvariable. Dann heißt

$$f_X(t) = \sum_{k=0}^{\infty} P(X = k) t^k \quad \text{mit } |t| \leq 1,$$

erzeugende Funktion von X. Es ist

$$f_X'(t) = \sum_{k=1}^{\infty} kP(X = k) t^{k-1} \quad \text{für } t \in (-1,1).$$

Damit folgt aus dem Satz von der monotonen Konvergenz (4.4)

$$\begin{aligned} f_X'(1-) &= \lim_{\varepsilon \downarrow 0} f_X(1-\varepsilon) = \lim_{\varepsilon \downarrow 0} \sum_{k=1}^{\infty} kP(X = k)(1-\varepsilon)^{k-1} \\ &= \sum_{k=1}^{\infty} kP(X = k)\big(\lim_{\varepsilon \downarrow 0}(1-\varepsilon)\big)^k = \sum_{k=1}^{\infty} kP(X = k). \end{aligned}$$

Wir betrachten nun

$$\begin{aligned} g_X(t) &= \sum_{n=0}^{\infty} P(X > n) t^n = \sum_{n=0}^{\infty} (1 - P(X \leq n)) t^n = \frac{1}{1-t} - \sum_{n=0}^{\infty} P(X \leq n) t^n \\ &= \frac{1}{1-t} - \sum_{n=0}^{\infty} \Big(\sum\nolimits_{k=0}^{n} P(X = k)\Big) t^n = \frac{1}{1-t} - \Big(\sum\nolimits_{k=0}^{\infty} t^k\Big)\Big(\sum\nolimits_{l=0}^{\infty} P(X = l) t^l\Big) \\ &= \frac{1}{1-t} - \frac{1}{1-t} f_X(t) = \frac{1 - f_X(t)}{1-t} = \frac{f_X(t) - 1}{t-1} \end{aligned}$$

für $t \in (-1,1)$. Es gilt

$$\begin{aligned} g_X(1-) &= \lim\nolimits_{\varepsilon \downarrow 0} g_X(1-\varepsilon) = \sum_{n=0}^{\infty} P(X > n) = \sum_{n=0}^{\infty} \sum_{k=n+1}^{\infty} P(X = k) = \sum_{k=1}^{\infty} P(X = k) \sum_{n=0}^{k-1} 1 \\ &= \sum_{k=1}^{\infty} kP(X = k) = E(X). \end{aligned}$$

Dies ist äquivalent zur Existenz von $f_X'(1)$. Es gilt dann

$$E(X) = g_X(1-) = f_X'(1-) = f_X'(1) < \infty.$$

Weiter gilt z.B.

$$f_X''(1-) = \lim_{\varepsilon \downarrow 0} f_X''(1-\varepsilon) = \lim_{\varepsilon \downarrow 0} \sum_{k=2}^{\infty} k(k-1) P(X = k)(1-\varepsilon)^{k-2} = \sum_{k=2}^{\infty} k(k-1) P(X = k).$$

Es ist

$$E(X(X-1)) = \sum_{k=2}^{\infty} k(k-1) P(X = k).$$

Ist $E(X^2) < \infty$, so gilt für die Varianz

$$\operatorname{Var}(X) = f''(X)(1-) + f_X'(1-) - \big(f_X'(1-)\big)^2.$$

Satz 7.2 (Eindeutigkeitssatz für erzeugende Funktionen)
Seien X, Y $\mathbb{N}_0$-wertige Zufallsvariablen mit erzeugenden Funktionen f_X und f_Y. Dann gilt

$$f_X(t) = f_Y(t) \quad \text{für } |t| \leq 1,$$

was äquivalent zu $P^X = P^Y$ ist.
Beweis
Identitätssatz für Potenzreihen. □

Satz 7.3

Seien X,Y $\mathbb{N}_0$-wertige Zufallsvariablen mit erzeugenden Funktionen f_X und f_Y. Seien X,Y unabhängig. Dann gilt:

$$f_{X+Y}(t) = f_X(t) f_Y(t) \text{ für } |t| \le 1 .$$

Beweis

Es gilt:

$$f_{X+Y}(t) = E(t^{X+Y}) = E(t^X t^Y) = E(t^X)E(t^Y) = f_X(t) f_Y(t) \text{ mit } |t| \le 1 . \qquad \square$$

Beispiel 7.4

Seien X_n, $n \in \mathbb{N}$, unabhängige Zufallsvariablen mit je derselben Verteilung $\mathcal{B}(1,p)$, $p \in (0,1)$. Sei

$$T_1 = \inf\{n \in \mathbb{N} \mid X_n = 1\} .$$

Sei $T_1 := \infty$, falls $X_n = 0$ für jedes $n \in \mathbb{N}$. Dann gilt:

$$P(T_1 = k) = P(X_1 = 0,\dots,X_{k-1} = 0, X_k = 1) = P(X_1 = 0)^{k-1} P(X_1 = 1) = (1-p)^{k-1} p \text{ für } k \in \mathbb{N} .$$

Weiterhin gilt:

$$\sum_{k=1}^{\infty} P(T_1 = k) = \sum_{k=1}^{\infty} (1-p)^{k-1} p = p \sum_{k=0}^{\infty} (1-p)^k = \frac{1}{1-(1-p)} p = 1 .$$

Es folgt:

$$P(T_1 = \infty) = 0 .$$

Definition 7.5

Eine Zufallsvariable W heißt **geometrisch verteilt** mit den Parametern $p \in (0,1)$ auf $\mathbb{N}$, wenn gilt

$$P(W = k) = (1-p)^{k-1} p \text{ für } k \in \mathbb{N} .$$

Für die erzeugende Funktion von W gilt:

$$f_W(t) = E(t^W) = \sum_{k=1}^{\infty} P(W = k) t^k = \sum_{k=1}^{\infty} (1-p)^{k-1} p t^k = tp \sum_{k=1}^{\infty} ((1-p)t)^{k-1} = tp \sum_{k=0}^{\infty} ((1-p)t)^k$$
$$= \frac{pt}{1-(1-p)t} \text{ für } |t| < \frac{1}{1-p}.$$

Außerdem gilt

$$f_W'(1) = E(W) = \frac{1}{p}.$$

Sei $r \in \mathbb{N}$. Definiere T_j induktiv wie folgt, wobei $T_0 := 0$ ist:

$$T_{j+1} := \inf\{n \in \mathbb{N} \mid n > T_j, X_n = 1\} \text{ für } j \in \mathbb{N}$$

und D_j induktiv wie folgt:

$$D_j := T_j - T_{j-1} ,\ j \in \mathbb{N} .$$

Dann gilt:

$$\begin{aligned} P(D_1 = j_1,\dots,D_r = j_r) &= P(T_1 = j_1, T_2 = j_1 + j_2,\dots,T_r = j_1 + \dots + j_r) \\ &= P(X_1 = 0,\dots,X_{j_1-1} = 0, X_{j_1} = 1, X_{j_1+1} = 0,\dots,X_{j_1+j_2-1} = 0, X_{j_1+j_2} = 1,\dots,X_{j_1+\dots+j_r} = 1) \\ &= (1-p)^{j_1+\dots+j_r-r} p^r \end{aligned}$$

für $j_1,\dots,j_r \in \mathbb{N}$. Es folgt

$$P(D_k = j_k) = (1-p)^{j_k-1} p \left(\sum_{j=1}^{\infty} (1-p)^{j-1} p\right)^{r-1} = (1-p)^{j_k-1} p \text{ für } j_k \in \mathbb{N} ,\ k \in \{1,\dots,r\} .$$

Also gilt

$$P(D_1 = j_1,\dots,D_r = j_r) = P(D_1 = j_1) \cdot \ldots \cdot P(D_r = j_r) \text{ für alle } r \in \mathbb{N} \text{ mit } j_1,\dots,j_r \in \mathbb{N} .$$

Es folgt, dass D_j, $j \in \mathbb{N}$, stochastisch unabhängig und identisch verteilt sind (und zwar je geometrisch mit dem Parameter p). Man kann $T_r = D_1 + \dots + D_r$ als „Wartezeit“ bis zum r-ten Erfolg interpretieren. Es gilt:

$$P(T_r = k) = \sum_{\substack{(j_1,\dots,j_r) \in \mathbb{N}^r \\ j_1+\dots+j_r = k}} P(D_1 = j_1,\dots,D_r = j_r) = \sum_{\substack{(j_1,\dots,j_r) \in \mathbb{N}^r \\ j_1+\dots+j_r = k}} (1-p)^{k-r} p^r = \binom{k-1}{k-r} (1-p)^{k-r} p^r$$

mit $k \in \{r, r+1,\dots\}$.

Definition 7.6
Eine Zufallsvariable V mit der Verteilung
$$P(V=k)=\binom{k-1}{k-r}(1-p)^{k-r}\,p^r \text{ für } k\in\{r,r+1,\ldots\},$$
heißt **negativ binomialverteilt** mit den Parametern $r\in\mathbb{N}$ und $p\in(0,1)$ auf der Menge $\{r,r+1,\ldots\}$.
Schreibweise: $V\sim \mathrm{Nb}(r,p)$ auf $\{r,r+1,\ldots\}$.

Sei $C_j=D_j-1$, $j\in\mathbb{N}$, die Anzahl der Misserfolge zwischen dem $j-1$-ten und dem j-ten Erfolg, sei weiter $S_j=T_j-j=C_1+\ldots+C_j$ die Anzahl aller Misserfolge bis zum j-ten Erfolg. Dann gilt:
$$P(C_j=k)=P(D_j=k+1)=(1-p)^k\,p \text{ für } k\in\mathbb{N}_0$$
und
$$P(S_r=k)=P(T_r=k+r)=\binom{k+r-1}{k}(1-p)^k\,p^r \text{ für } k\in\mathbb{N}_0.$$

Definition 7.7
Eine Zufallsvariable U mit der Verteilung
$$P(U=k)=\binom{k+r-1}{k}(1-p)^k\,p^r \text{ für } k\in\mathbb{N}_0$$
heißt **negativ binomialverteilt** mit den Parametern $r\in\mathbb{N}$ und $p\in(0,1)$ auf $\mathbb{N}_0$.
Schreibweise: $U\sim \mathcal{N}b(r,p)$ auf $\mathbb{N}_0$.

Sei $U\sim\mathcal{N}b(r,p)$ auf $\mathbb{N}_0$, $V\sim\mathrm{Nb}(r,p)$ auf $\{r,r+1,\ldots\}$. Dann gilt
$$f_U(t)=(f_{U_1}(1))^r=\left(\frac{pt}{1-(1-p)t}\right)^r \text{ für } |t|<\frac{1}{1-p},$$
mit $U_1\sim\mathcal{N}b(1,p)$ und $E(U)=r/p$. Es ist
$$f_V(t)=(f_{V_1}(t))^r=E(t^{V_1})=\frac{1}{t}\frac{pt}{1-(1-p)t}=\frac{p}{1-(1-p)t} \text{ für } |t|<\frac{1}{1-p}$$
die erzeugende Funktion von $V_1\sim\mathcal{N}b(1,p)$. Dann gilt:
$$E(V)=rE(V_1)=r\left(\frac{1}{p}-1\right)=r\frac{1-p}{p} \text{ und } \mathrm{Var}(U)=\mathrm{Var}(V)=r\frac{1-p}{p^2}.$$

Satz 7.8 (Poissonscher Grenzwertsatz)
Sei $X_n\sim\mathcal{B}(n,p_n)$ für $n\in\mathbb{N}$, $p_n\in(0,1)$. Es gelte $\lim_{n\to\infty}np_n=\lambda\in(0,\infty)$. Dann gilt für $n\to\infty$:
$$P(X_n=k)\to e^{-\lambda}\frac{\lambda^k}{k!} \text{ für jedes } k\in\mathbb{N}_0.$$

Beweis
Es gilt:
$$P(X_n=k)=\binom{n}{k}p_n^k(1-p_n)^{n-k}=\frac{1}{k!}\underbrace{n(n-1)\cdot\ldots\cdot(n-k+1)\,p_n^k}_{\to\lambda^k}\underbrace{\left(1-\frac{np_n}{n}\right)^{n-k}}_{\to e^{-\lambda}} \text{ für } k\in\mathbb{N}_0.$$
Außerdem gilt:
$$\sum_{k=0}^{\infty}e^{-\lambda}\frac{\lambda^k}{k!}=e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}=e^{-\lambda}e^{\lambda}=1.$$
□

Definition 7.9
Eine Zufallsvariable X mit der Verteilung
$$P(X=k)=e^{-\lambda}\frac{\lambda^k}{k!} \text{ für } k\in\mathbb{N}_0$$
heißt **Poisson-verteilt** mit dem Parameter $\lambda>0$.
Also: Bei $X_n\sim\mathcal{B}(n,p_n)$ mit $\lim_{n\to\infty}np_n=\lambda\in(0,\infty)$ ist X Poisson-verteilt mit dem Parameter $\lambda>0$.
Schreibweise: $X\sim\mathcal{P}(\lambda)$.

Es gilt:
$$\lim_{n\to\infty}P(X_n=k)=P(X=k) \text{ für alle } k\in\mathbb{N}_0.$$

Approximation für $P(X = k)$ bei $X \sim \mathcal{B}(n, p)$ liefert:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \approx e^{-np} \frac{(np)^k}{k!} \text{ für } k \in \{0,\dots,n\}.$$

Für die erzeugende Funktion von $X_n \sim \mathcal{B}(n, p_n)$ gilt:

$$f_{X_n}(t) = E(t^{V_n}) = \sum_{k=0}^{n} \binom{n}{k} p_n^k (1-p_n)^{n-k} t^k, \; t \in \mathbb{R}.$$

Für die erzeugende Funktion von $X \sim \mathcal{P}(\lambda)$ gilt:

$$f_X(t) = E(t^X) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} t^k = e^{-\lambda} e^{\lambda t} = e^{\lambda(t-1)}, \; t \in \mathbb{R}.$$

Außerdem gilt:

$$f_X'(1) = \lambda = E(X) \text{ und } f_X''(1) + f_X'(1) - f_X'(1)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda = \operatorname{Var} X.$$

Für $np_n \to \lambda \in (0,\infty)$ gilt

$$f_{X_n}(t) = (1 - p_n + p_n t)^n = \left(1 + \frac{np_n(t-1)}{n}\right)^n \to e^{\lambda(t-1)} = f_X(t), \; t \in \mathbb{R}.$$

Satz 7.10 (Stetigkeitssatz)

Seien $a_{k,n} \geq 0$ für $n \in \mathbb{N}$, $k \in \mathbb{N}_0$ und es gelte

$$\sum_{k=0}^{\infty} a_{k,n} = 1 \text{ für jedes } n \in \mathbb{N}.$$

Sei

$$A_n(t) = \sum_{k=0}^{\infty} a_{k,n} t^k \text{ für } |t| < 1, \; n \in \mathbb{N}.$$

Dann gilt:

(i) Aus der Existenz von $\lim_{n\to\infty} a_{k,n} = a_k$ für jedes $k \in \mathbb{N}_0$ folgt die Existenz von
$$\lim_{n\to\infty} A_n(t) = A(t) \text{ für } |t| < 1.$$

(ii) Aus der Existenz von $\lim_{n\to\infty} A_n(t) = A(t)$ für jedes $t \in (0,1)$ folgt die Existenz von
$$\lim_{n\to\infty} a_{k,n} = a_k \text{ für jedes } k \in \mathbb{N}_0.$$

Beweis

(i) Sei $t \in (-1,1)$. Dann gilt:

$$|A_n(t) - A(t)| = \left|\sum_{k=0}^{\infty} (a_{k,n} - a_k) t^k\right| \leq \sum_{k=0}^{l} |a_{k,n} - a_k| |t|^k + \sum_{k=l+1}^{\infty} |t|^k = \sum_{k=0}^{l} |a_{k,n} - a_k| |t|^k + \frac{|t|^{l+1}}{1-|t|}.$$

Somit gilt:

$$\limsup_{n\to\infty} |A_n(t) - A(t)| \leq 0 + \frac{|t|^{l+1}}{1-|t|} \text{ für alle } l \in \mathbb{N}.$$

$l \to \infty$ liefert $\lim_{n\to\infty} |A_n(t) - A(t)| = 0$.

(ii) Vollständige Induktion nach $k \in \mathbb{N}_0$ liefert:

$k = 0$: Sei $a_0 = \lim_{t\downarrow 0} A(t)$. Dann gilt:

$$a_{0,n} \leq A_n(t) \leq a_{0,n} + \frac{t}{1-t} \text{ für } t \in (0,1) \text{ und alle } n \in \mathbb{N}.$$

Daraus folgt

$$A_n(t) - \frac{t}{1-t} \leq a_{0,n} \leq A_n(t)$$

und somit

$$A(t) - \frac{t}{1-t} \leq \liminf_{n\to\infty} a_{0,n} \leq \limsup_{n\to\infty} a_{0,n} \leq A(t).$$

$t \downarrow 0$ liefert

$$a_0 \leq \liminf_{n\to\infty} a_{0,n} \leq \limsup_{n\to\infty} a_{0,n} \leq a_0.$$

Also folgt $a_0 = \lim_{n\to\infty} a_{0,n}$.

$k \to k+1$: Es sei

$$B_n(t) = \frac{1}{t^{k+1}}\left(A_n(t) - \sum\nolimits_{j=0}^{k} a_{j,n} t^j\right) = \sum_{j=k+1}^{\infty} a_{j,n} t^{j-(k+1)}.$$

Sei $B(t) = \lim_{n\to\infty} B_n(t)$ für $t \in (0,1)$, $a_{k+1} := \lim_{t\downarrow 0} B(t)$. Dann gilt:

$$a_{k+1,n} \le B_n(t) \le a_{k+1,n} + \frac{t}{1-t}$$

und somit

$$B_n(t) - \frac{t}{1-t} \le a_{k+1,n} \le B_n(t),$$

also

$$B(t) - \frac{t}{1-t} \le \liminf_{n\to\infty} a_{k+1,n} \le \limsup_{n\to\infty} a_{k+1,n} \le B(t).$$

Für $t \downarrow 0$ gilt:

$$a_{k+1} \le \liminf_{n\to\infty} a_{k+1,n} \le \limsup_{n\to\infty} a_{k+1,n} \le a_{k+1}$$

und damit

$$a_{k+1} = \lim_{n\to\infty} a_{k+1,n}.$$

□

Korollar 7.11
Seien X und X_n, $n \in \mathbb{N}$, $\mathbb{N}_0$-wertige Zufallsvariablen mit den erzeugenden Funktionen f_X und f_{X_n}, $n \in \mathbb{N}$. Genau dann gilt

$$\lim_{n\to\infty} P(X_n = k) = P(X = k) \text{ für jedes } k \in \mathbb{N}_0,$$

wenn

$$\lim_{n\to\infty} f_{X_n}(t) = f_X(t) \text{ für jedes } t \in (-1,1).$$

Beispiel 7.12
Sei $X_r \sim \mathcal{N}b(r, p_r)$ auf $\mathbb{N}_0$, $r \in \mathbb{N}$, $p_r \in (0,1)$, und gelte $r(1-p_r) \to \lambda \in (0,\infty)$ für $r \to \infty$. Sei zudem $X \sim \mathcal{P}(\lambda)$. Dann gilt:

$$f_{X_r}(t) = \left(\frac{p_r}{1-(1-p_r)t}\right)^r = \frac{\left(1 - \dfrac{r(1-p_r)}{r}\right)^r}{\left(1 - \dfrac{r(1-p_r)t}{r}\right)^r} \to \frac{e^{-\lambda}}{e^{-\lambda t}} = e^{\lambda(t-1)} = f_X(t) \text{ für } |t| < 1.$$

Verallgemeinerung auf den Fall $\mathbb{N}_0^m$-wertiger Zufallsvektoren

Definition 7.13
Sei $X = (X_1, \dots, X_m)$ ein $\mathbb{N}_0^m$-wertiger Zufallsvektor. Dann heißt

$$f_X(t_1,\dots,t_m) = E\left(t_1^{X_1} \cdot \ldots \cdot t_m^{X_m}\right) = \sum_{(k_1,\dots,k_m)\in\mathbb{N}_0^m} P(X_1 = k_1, \dots, X_m = k_m)\, t_1^{k_1} \cdot \ldots \cdot t_m^{k_m}$$

für $|t_i| \le 1$, $i = 1,\dots,m$, **erzeugende Funktion** von X.

Satz 7.14 (Eindeutigkeitssatz)
Seien $X = (X_1,\dots,X_m)$ und $Y = (Y_1,\dots,Y_m)$ $\mathbb{N}_0^m$-wertig mit den erzeugenden Funktionen f_X und f_Y. Dann gilt:

$$f_X(t_1,\dots,t_m) = f_Y(t_1,\dots,t_m) \text{ für alle } (t_1,\dots,t_m) \in [-1,1]^m,$$

was äquivalent zu $P^X = P^Y$ ist.
Beweis
Eindeutigkeitssatz für Potenzreihen. □

Sei $X = (X_1,\dots,X_m)$ eine $\mathbb{N}_0^m$-wertiger Zufallsvektor mit erzeugender Funktion f_X.

(i) Dann gilt für die erzeugenden Funktionen von X_i:

$$f_{X_i}(t) = f_X(1,\dots,1,t,1,\dots,1) \text{ mit } t \text{ als } i\text{-te Komponente und } |t| \le 1.$$

(ii) Die Unabhängigkeit von $X_1,\dots,X_m$ ist äquivalent zu

$$f_X(t_1,\dots,t_m) = \prod_{i=1}^m f_{X_i}(t_i) \text{ für alle } |t_i| \le 1,\ i = 1,\dots,m.$$

(iii) Die erzeugende Funktion von $S = X_1 + \ldots + X_m$ ist

$$f_S(t) = E(t^S) = f_X(t,\ldots,t) \text{ für } |t| \le 1 .$$

(iv) Für die erzeugende Funktion von $X_i + X_j$ für $1 \le i \le j \le m$ gilt

$$f_{X_i+X_j}(1,\ldots,1,t,1,\ldots,1,t,1,\ldots,1) \text{ mit } t \text{ als } i\text{-te und } j\text{-te Komponente und } |t| \le 1 .$$

Beispiel 7.15

Sei $X = (X_1,\ldots,X_s) \sim \mathcal{M}(n;p_1,\ldots,p_s)$ mit $n \in \mathbb{N}$, $p_i \in (0,1)$, $p_1 + \ldots + p_s = 1$, $s \in \mathbb{N}$. Dann gilt

$$P(X_1 = k_1,\ldots,X_s = k_s) = \frac{n!}{k_1! \cdot \ldots \cdot k_s!} p_1^{k_1} \cdot \ldots \cdot p_s^{k_s} \text{ für } (k_1,\ldots,k_s) \in \mathbb{N}_0^s .$$

Für die erzeugende Funktion folgt nach dem polynomischen Lehrsatz:

$$f_X(t_1,\ldots,t_s) = E\left(t_1^{X_1} \cdot \ldots \cdot t_s^{X_s}\right) = \sum_{\substack{(k_1,\ldots,k_s)\in\mathbb{N}_0^s \\ k_1+\ldots+k_s=n}} \frac{n!}{k_1! \cdot \ldots \cdot k_s!} p_1^{k_1} \cdot \ldots \cdot p_s^{k_s} t_1^{k_1} \cdot \ldots \cdot t_s^{k_s} = (p_1 t_1 + \ldots + p_s t_s)^n .$$

Denn es gilt

$$(x_1 + \ldots + x_s)^n = \sum_{\substack{(k_1,\ldots,k_s)\in\mathbb{N}_0^s \\ k_1+\ldots+k_s=n}} \frac{n!}{k_1! \cdot \ldots \cdot k_s!} x_1^{k_1} \cdot \ldots \cdot x_s^{k_s} .$$

Beispiel 7.16

Betrachten wir das Einzelexperiment, bei dem die sich paarweise ausschließenden Ereignisse $A_0,\ldots,A_s$ auftreten können, wobei genau eines dieser Ereignisse auftritt. Dieses Einzelexperiment wird unbeschränkt oft wiederholt. Es sei X_j die Anzahl der Versuchswiederholungen, bei denen A_j auftritt bis zum r-ten Auftreten von A_0, $j = 1,\ldots,s$. Wir suchen die Verteilung von $(X_1,\ldots,X_s)$. Es gilt:

$$P(X_1 = k_1,\ldots,X_s = k_s) = \binom{k+r-1}{k} \frac{k!}{k_1! \cdot \ldots \cdot k_s!} p_1^{k_1} \cdot \ldots \cdot p_s^{k_s} p_0^r$$

für $k,k_1,\ldots,k_s \in \mathbb{N}_0$, $k = k_1 + \ldots + k_s$, $p_0,\ldots,p_s \in (0,1)$, $r \in \mathbb{N}$.

Definition 7.17

Ein $\mathbb{N}_0^s$-wertiger Zufallsvektor $X = (X_1,\ldots,X_s)$ mit der Verteilung

$$P(X_1 = k_1,\ldots,X_s = k_s) = \binom{k+r-1}{k} \frac{k!}{k_1! \cdot \ldots \cdot k_s!} p_1^{k_1} \cdot \ldots \cdot p_s^{k_s} p_0^r$$

heißt **mehrdimensional negativ binomialverteilt** mit den Parametern $k,k_1,\ldots,k_s \in \mathbb{N}_0$ mit $k_1 + \ldots + k_s = k$, $r \in \mathbb{N}$ und $p_0,\ldots,p_s \in (0,1)$.
Schreibweise: $X \sim \mathrm{MNb}(r;p_0;p_1,\ldots,p_s)$.

Die zugehörige erzeugende Funktion lautet:

$$\begin{aligned} f_X(t_1,\ldots,t_s) &= \sum_{(k_1+\ldots+k_s)\in\mathbb{N}_0^s} \binom{k_1+\ldots+k_s+r-1}{k_1+\ldots+k_s} \frac{(k_1+\ldots+k_s)!}{k_1! \cdot \ldots \cdot k_s!} p_1^{k_1} \cdot \ldots \cdot p_s^{k_s} p_0^r t_1^{k_1} \cdot \ldots \cdot t_s^{k_s} \\ &= \sum_{k=0}^{\infty} \left(\sum_{\substack{(k_1,\ldots,k_s)\in\mathbb{N}_0^s \\ k_1+\ldots+k_s=k}} \frac{k!}{k_1! \cdot \ldots \cdot k_s!} (p_1 t_1)^{k_1} \cdot \ldots \cdot (p_s t_s)^{k_s} \right) \binom{k+r-1}{k} p_0^r \\ &= \sum_{k=0}^{\infty} \binom{k+r-1}{k} p_0^r (p_1 t_1 + \ldots + p_s t_s)^k = \left(\frac{p_0}{1-(p_1 t_1 + \ldots + p_s t_s)} \right)^r \end{aligned}$$

für $|t_i| \le 1$, $i = 1,\ldots,s$.

Kapitel 8

Zentrale Grenzwertsätze

Sei $X_\lambda \sim \mathcal{P}(\lambda)$. Dann ist $E(X_\lambda) = \lambda = \operatorname{Var}(X_\lambda)$ und es gilt für großes λ

$$P\Big(a < \frac{X_\lambda - \lambda}{\sqrt{\lambda}} \le b\Big) = \sum_{\lambda + a\sqrt{\lambda} < k \le \lambda + b\sqrt{\lambda}} P(X_\lambda = k) = \sum_{\lambda + a\sqrt{\lambda} < k \le \lambda + b\sqrt{\lambda}} e^{-\lambda} \frac{\lambda^k}{k!}.$$

Sei $a > 0$, $C_\lambda(a) = \{k \in \mathbb{N}_0 \mid |k - \lambda| \le a\sqrt{\lambda}\}$, $Z_\lambda(a) = \{z > 0 \mid z - k = \lambda \text{ für } k \in C_\lambda(a)\}$. Sei $k_0 = [\lambda]$ definiert durch die größte ganze Zahl, die kleiner oder gleich λ ist. Es sei $g_\lambda(k) = \log P(X_\lambda = k)$ für $k \in \mathbb{N}_0$. Dann gilt für $k = z + \lambda$, $z \in Z_\lambda(a)$:

$$\frac{P(X_\lambda = k+1)}{P(X_\lambda = k)} = \frac{e^{-\lambda}\frac{\lambda^{k+1}}{(k+1)!}}{e^{-\lambda}\frac{\lambda^k}{k!}} = \frac{\lambda}{k+1} = \frac{\lambda}{1 + \lambda + z} = \frac{1}{1 + \frac{z}{\lambda} + \frac{1}{\lambda}}.$$

Es folgt:

$$g_\lambda(k+1) - g_\lambda(k) = -\log\Big(1 + \frac{z}{\lambda} + \frac{1}{\lambda}\Big) = -\frac{z}{\lambda} + \frac{r_1^\lambda(z)}{\lambda} = -\frac{k - k_0}{\lambda} + \frac{r_2^\lambda(k)}{\lambda}$$

mit $\sup_{z \in Z_\lambda(a)} |r_1^\lambda(z)| = 0$ bzw. 1 und $\sup_{k \in C_\lambda(a)} |r_2^\lambda(k)| = 0$ bzw. 1, für $\lambda \to \infty$.
Sei $k > k_0$ mit $k \in C_\lambda(a)$. Dann gilt:

$$\begin{aligned} g_\lambda(k) - g_\lambda(k_0) &= \sum_{j=k_0}^{k-1} (g_\lambda(j+1) - g_\lambda(j)) = -\sum_{j=k_0}^{k-1} \frac{j - k_0}{\lambda} + \frac{r_3^\lambda(k)}{\sqrt{\lambda}} = -\frac{1}{\lambda} \sum_{j=1}^{k-k_0-1} j + \frac{r_3^\lambda(k)}{\sqrt{\lambda}} \\ &= -\frac{1}{2\lambda}(k - k_0)^2 + \frac{r_4^\lambda(k)}{\sqrt{\lambda}} = -\frac{1}{2\lambda}(k - \lambda)^2 + \frac{r_5^\lambda(k)}{\sqrt{\lambda}} \end{aligned}$$

mit $\sup_{k \in C_\lambda(a)} |r_3^\lambda(k)| = 0$ bzw. 1, $\sup_{k \in C_\lambda(a)} |r_4^\lambda(k)| = 0$ bzw. 1 und $\sup_{k \in C_\lambda(a)} |r_5^\lambda(k)| = 0$ bzw. 1, für $\lambda \to \infty$. Es folgt:

$$P(X_\lambda = k) = e^{g_\lambda(k_0)} e^{-\frac{1}{2\lambda}(k-\lambda)^2} \Big(1 + \frac{s^\lambda(k)}{\sqrt{\lambda}}\Big) = c(\lambda) e^{-\frac{1}{2\lambda}(k-\lambda)^2} \Big(1 + \frac{s^\lambda(k)}{\sqrt{\lambda}}\Big)$$

mit $\sup_{k \in C_\lambda(a)} |s^\lambda(k)| = 0$ bzw. 1 und $c(\lambda) = e^{g_\lambda(k_0)} = P(X_\lambda = k_0)$. Weiterhin gilt mit der Chebyshevschen Ungleichung:

$$1 \ge P(X_\lambda \in C_\lambda(a)) = P\Big(\Big|\frac{X_\lambda - \lambda}{\sqrt{\lambda}}\Big| \le a\Big) = 1 - P\Big(\Big|\frac{X_\lambda - \lambda}{\sqrt{\lambda}}\Big| > a\Big) \ge 1 - \frac{1}{a^2} \quad \text{für } a > 0.$$

Zudem gilt:

$$P(X_\lambda \in C_\lambda(a)) = \sum_{\lambda - a\sqrt{\lambda} \le k \le \lambda + a\sqrt{\lambda}} P(X_\lambda = k) = \sqrt{\lambda} c(\lambda) \underbrace{\sum_{\lambda - a\sqrt{\lambda} \le k \le \lambda + a\sqrt{\lambda}} \frac{1}{\sqrt{\lambda}} e^{-\frac{1}{2\lambda}(k-\lambda)^2}}_{\to I(a)} \Big(1 + \frac{s^\lambda(k)}{\sqrt{\lambda}}\Big),$$

wobei

$$I(a) := \int_{-a}^{a} e^{-\frac{x^2}{2}}\, dx.$$

Es folgt:

$$\frac{1}{I(a)} \ge \limsup_{\lambda \to \infty} \sqrt{\lambda} c(\lambda) \ge \liminf_{\lambda \to \infty} \sqrt{\lambda} c(\lambda) \ge \Big(1 - \frac{1}{a^2}\Big) \frac{1}{I(a)} \quad \text{für alle } a > 0.$$

Für $a \to \infty$ gilt

$$\lim_{a \to \infty} I(a) = \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\, dx = \sqrt{2\pi}.$$

Damit gilt:

$$\lim_{\lambda \to \infty} \sqrt{\lambda} c(\lambda) = \frac{1}{\sqrt{2\pi}}.$$

Satz 8.1 (Zentraler Grenzwertsatz für die Poisson-Verteilung, lokale Form)
Es gilt

$$\sup_{k \in C_\lambda(k)} \left| \frac{P(X_\lambda = k)}{\frac{1}{\sqrt{2\pi\lambda}} \exp\Big(-\frac{1}{2\lambda}(k - \lambda)^2\Big)} - 1 \right| \to 0.$$

Schreibweise:

$$P(X_\lambda = k) \sim \frac{1}{\sqrt{2\pi\lambda}} \exp\Big(-\frac{1}{2\lambda}(k-\lambda)^2\Big) \text{ für } \lambda \to \infty \text{ gleichmäßig in } k_0 \in \mathbb{N}_0$$

mit $|k-\lambda| \le c\sqrt{\lambda}$ für $c > 0$.
Spezialfall: $k = \lambda \in \mathbb{N}$. Hierfür gilt:

$$P(X_\lambda = \lambda) = e^{-k}\frac{k^k}{k!} \sim \frac{1}{\sqrt{2\pi k}} \text{ für } k \to \infty .$$

Es folgt:

$$k! \sim \sqrt{2\pi k}\Big(\frac{k}{e}\Big)^k \text{ für } k \to \infty \text{ (\textbf{Stirlingsche Formel}).}$$

Satz 8.2 (Zentraler Grenzwertsatz für die Poisson-Verteilung, kumulative Form)
Für $-\infty < a < b < \infty$ gilt

$$\lim_{\lambda\to\infty} P\Big(a \underset{(-)}{<} \frac{X_\lambda - \lambda}{\sqrt{\lambda}} \underset{(-)}{<} b\Big) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{x^2}{2}}\,dx .$$

Beweis
Es gilt:

$$P\Big(a \underset{(-)}{<} \frac{X_\lambda - \lambda}{\sqrt{\lambda}} \underset{(-)}{<} b\Big) = \Bigg[\sum_{\lambda + a\sqrt{\lambda} \underset{(-)}{<} k \underset{(-)}{<} \lambda + b\sqrt{\lambda}} \frac{1}{\sqrt{2\pi\lambda}} e^{-\frac{1}{2\lambda}(k-\lambda)^2}\Bigg](1 + 0 \text{ bzw. } 1) \to \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{x^2}{2}}\,dx$$

für $\lambda \to \infty$. □

Definition 8.3
Wir nennen

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^x e^{-\frac{t^2}{2}}\,dt \text{ für } x \in \mathbb{R}$$

die **Verteilungsfunktion** der $(0,1)$-Verteilung.

Es gilt dann:

$$\lim_{\lambda\to\infty} P\Big(a \underset{(-)}{<} \frac{X_\lambda - \lambda}{\sqrt{\lambda}} \underset{(-)}{<} b\Big) = \Phi(b) - \Phi(a) .$$

Korollar 8.4
Es gilt:

$$\lim_{\lambda\to\infty} P\Big(\frac{X_\lambda - \lambda}{\sqrt{\lambda}} \le x\Big) = \Phi(x) \text{ für alle } x \in \mathbb{R} .$$

Beweis
Es gilt:

$$\lim_{\lambda\to\infty} P\Big(\frac{X_\lambda - \lambda}{\sqrt{\lambda}} \le x\Big) \ge \lim_{\lambda\to\infty} P\Big(a < \frac{X_\lambda - \lambda}{\sqrt{\lambda}} \le x\Big) = \Phi(x) - \Phi(a) \text{ für alle } a \in \mathbb{R} .$$

Für $a \to -\infty$ gilt $\Phi(a) \to 0$. Es folgt:

$$\begin{aligned}\liminf_{\lambda\to\infty} P\Big(\frac{X_\lambda - \lambda}{\sqrt{\lambda}} \le x\Big) &= \limsup_{\lambda\to\infty}\Big(1 - P\Big(\frac{X_\lambda - \lambda}{\sqrt{\lambda}} > x\Big)\Big) \ge \limsup_{\lambda\to\infty}\Big(1 - P\Big(x < \frac{X_\lambda - \lambda}{\sqrt{\lambda}} \le b\Big)\Big)\\ &= 1 - (\Phi(b) - \Phi(x))\end{aligned}$$

für alle $b > x$. Für $b \to \infty$ gilt $\Phi(b) \to 1$. Also folgt:

$$\limsup_{\lambda\to\infty} P\Big(\frac{X_\lambda - \lambda}{\sqrt{\lambda}} \le x\Big) \ge \Phi(x) .$$

□

Beispiel 8.5
Sei $X \sim \mathcal{P}(\lambda)$. Dann gilt:

$$P(a \le X \le b) = P\Big(\frac{a-\lambda}{\sqrt{\lambda}} \le \frac{X-\lambda}{\sqrt{\lambda}} \le \frac{b-\lambda}{\sqrt{\lambda}}\Big) \approx \Phi\Big(\frac{b-\lambda}{\sqrt{\lambda}}\Big) - \Phi\Big(\frac{a-\lambda}{\sqrt{\lambda}}\Big).$$

Sei $X \sim \mathcal{N}b(r,p)$ auf $\mathbb{N}_0$, $Y \sim \mathrm{Nb}(r,p)$ auf $\{r, r+1, \ldots\}$. Dann gilt für die erzeugenden Funktionen:

$$f_X(t) = \Big(\frac{p}{1-(1-p)t}\Big)^r,\ f_Y(t) = \Big(\frac{pt}{1-(1-p)t}\Big)^r \text{ für } |t| < \frac{1}{1-p} .$$

Tabelle der Verteilungsfunktion $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt$ **,** $x \geq 0$ **, der** $\mathcal{N}(0{,}1)$ **-Verteilung**

x	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990

Ablesebeispiele:

$\Phi(1{,}96) = 0{,}9750$

$\Phi(-0{,}75) = 1 - \Phi(0{,}75) = 1 - 0{,}7734 = 0{,}2266$

Tabelle der Werte u_α

$$\alpha = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u_\alpha} e^{-\frac{t^2}{2}}\, dt\,,\ \alpha \in (0{,}1)$$

α	0,9	0,95	0,975	0,99	0,995
u_α	1,2816	1,6449	1,9600	2,3264	2,5758

Es ist $u_{1-\alpha} = -u_\alpha$.

Verteilungen von nichtnegativen ganzzahligen Zufallsvariablen

Binomialverteilung: $\mathcal{B}(n,p),\ n \in \mathbb{N},\ p \in (0,1)$

Zähldichte: $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ für $k \in \{0,\dots,n\}$
Erwartungswert: $E(X) = np$
Varianz: $\mathrm{Var}(X) = np(1-p)$
Erzeugende Funktion: $f_X(t) = (1-p+pt)^n$ für $t \in \mathbb{R}$

Poisson-Verteilung: $\mathcal{P}(\lambda),\ \lambda > 0$

Zähldichte: $P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$ für $k \in \mathbb{N}_0$
Erwartungswert: $E(X) = \lambda$
Varianz: $\mathrm{Var}(X) = \lambda$
Erzeugende Funktion: $f_X(t) = e^{\lambda(t-1)}$ für $t \in \mathbb{R}$

Hypergeometrische Verteilung: $\mathcal{H}(a,r,n),\ a,r,n \in \mathbb{N},\ n \le a,\ r \le a$

Zähldichte: $P(X = k) = \frac{\binom{r}{k}\binom{a-r}{n-k}}{\binom{a}{n}}$ für $\max\{0, n+r-a\} \le k \le \min\{r,n\}$
Erwartungswert: $E(X) = n\frac{r}{a}$
Varianz: $\mathrm{Var}(X) = n\frac{r}{a}\left(1-\frac{r}{a}\right)\frac{a-n}{a-1}$

Negative Binomialverteilung: $\mathcal{N}b(r,p)$ auf $\mathbb{N}_0,\ r \in \mathbb{N},\ p \in (0,1)$

Zähldichte: $P(X = k) = \binom{k+r-1}{k} p^r (1-p)^k$ für $k \in \mathbb{N}_0$
Erwartungswert: $E(X) = r\frac{1-p}{p}$
Varianz: $\mathrm{Var}(X) = r\frac{1-p}{p^2}$
Erzeugende Funktion: $f_X(t) = \left(\frac{p}{1-(1-p)t}\right)^r$ für $|t| < \frac{1}{1-p}$

Negative Binomialverteilung: $\mathrm{Nb}(r,p)$ auf $\{r, r+1, \dots\},\ r \in \mathbb{N},\ p \in (0,1)$

Zähldichte: $P(X = k) = \binom{k-1}{k-r} p^r (1-p)^{k-r}$ für $k \in \{r, r+1, \dots\}$
Erwartungswert: $E(X) = r\frac{1}{p}$
Varianz: $\mathrm{Var}(X) = r\frac{1-p}{p^2}$
Erzeugende Funktion: $f_X(t) = \left(\frac{pt}{1-(1-p)t}\right)^r$ für $|t| < \frac{1}{1-p}$

Verteilungen von reellen Zufallsvariablen mit Dichten

Normalverteilung: $\mathcal{N}(\mu, \sigma^2),\ \mu \in \mathbb{R},\ \sigma^2 > 0$

Dichte: $\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ für $x \in \mathbb{R}$
Erwartungswert: $E(X) = \mu$
Varianz: $\mathrm{Var}(X) = \sigma^2$

Exponentialverteilung: $\operatorname{Exp}(\lambda),\ \lambda > 0$

$$\text{Dichte: } \lambda e^{-\lambda x} \text{ für } x \geq 0$$
$$\text{Erwartungswert: } E(X) = \frac{1}{\lambda}$$
$$\text{Varianz: } \operatorname{Var}(X) = \frac{1}{\lambda^2}$$

Rechteckverteilung: $\mathcal{R}(a,b),\ -\infty < a < b < \infty$

$$\text{Dichte: } \frac{1}{b-a} \text{ für } x \in [a,b]$$
$$\text{Erwartungswert: } E(X) = \frac{a+b}{2}$$
$$\text{Varianz: } \operatorname{Var}(X) = \frac{1}{12}(b-a)^2$$

Gammaverteilung: $\mathcal{G}(\alpha,\lambda),\ \alpha > 0,\ \lambda > 0$

$$\text{Dichte: } \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \text{ für } x > 0$$
$$\text{Erwartungswert: } E(X) = \frac{\alpha}{\lambda}$$
$$\text{Varianz: } \operatorname{Var}(X) = \frac{\alpha}{\lambda^2}$$

Cauchy-Verteilung: $\mathcal{C}(\mu,\sigma),\ \mu \in \mathbb{R},\ \sigma > 0$

$$\text{Dichte: } \frac{\sigma}{\pi} \frac{1}{\left(\sigma^2 + (x-\mu)^2\right)} \text{ für } x \in \mathbb{R}$$

Verteilungen von Zufallsvektoren

Normalverteilung: $\mathcal{N}_d(\mu,\Sigma),\ \mu \in \mathbb{R}^d,\ \Sigma \in \mathbb{R}^{d\times d}$ symmetrisch und positiv definit

$$\text{Dichte: } \left(\frac{1}{\sqrt{2\pi}}\right)^d \frac{1}{\sqrt{\det \Sigma}} e^{-\frac{1}{2}(x-\mu)^\intercal \Sigma^{-1}(x-\mu)} \text{ für } x = (x_1,\dots,x_d) \in \mathbb{R}^d$$
$$\text{Erwartungswertvektor: } E(X) = \mu$$
$$\text{Kovarianzmatrix: } \operatorname{Cov}(X) = \Sigma$$

Multinomialverteilung: $\mathcal{M}(n,p_1,\dots,p_s),\ n \in \mathbb{N},\ p_1,\dots,p_s \in (0,1),\ p_1 + \dots + p_s = 1$

$$\text{Zähldichte: } P(X_1 = k_1,\dots,X_s = k_s) = \frac{n!}{k_1! \cdot \ldots \cdot k_s!} p_1^{k_1} \cdot \ldots \cdot p_s^{k_s},\ (k_1,\dots,k_s) \in \mathbb{N}_0^s,\ k_1 + \dots + k_s = n$$
$$\text{Erwartungswertvektor: } E(X) = np \text{ für } p = (p_1,\dots,p_s)$$
$$\text{Kovarianzmatrix: } \operatorname{Cov}(X) = n[\operatorname{diag}(p_1,\dots,p_s) - pp^\intercal]$$
$$\text{Erzeugende Funktion: } f_X(t_1,\dots,t_s) = (p_1t_1 + \dots + p_st_s)^n \text{ für } t_i \in \mathbb{R} \text{ mit } i = 1,\dots,s$$

Mehrdimensionale negative Binomialverteilung: $\operatorname{MNb}(r;p_0;p_1,\dots p_s),\ r \in \mathbb{N},\ p_0,\dots,p_s \in (0,1),\ p_1 + \dots + p_s = 1$

$$\text{Zähldichte: } P(X_1 = k_1,\dots,X_s = k_s) = \binom{k+r-1}{k} \frac{k!}{k_1! \cdot \ldots \cdot k_s!} p_1^{k_1} \cdot \ldots \cdot p_s^{k_s} p_0^r$$
$$\text{für } (k_1,\dots,k_s) \in \mathbb{N}_0^s \text{ mit } k_1 + \dots + k_s = k$$
$$\text{Erwartungswertvektor: } E(X) = \frac{r}{p_0} p,\ p = (p_1,\dots,p_s)$$
$$\text{Kovarianzmatrix: } \operatorname{Cov}(X) = r\left[\operatorname{diag}\left(\frac{p_1}{p_0},\dots,\frac{p_s}{p_0}\right) + \frac{1}{p_0^2} pp^\intercal\right]$$
$$\text{Erzeugende Funktion: } f_X(t_1,\dots,t_s) = \left(\frac{p_0}{1-(p_1t_1 + \dots + p_st_s)}\right)^r \text{ für } |t_i| < \frac{1}{1-p_0} \text{ mit } i = 1,\dots,s$$

Satz 8.6 (Zentraler Grenzwertsatz für die Binomialverteilung)
Sei $X_n \sim \mathcal{B}(n, p_n)$ für $n \in \mathbb{N}$. Es gelte $\operatorname{Var}(X_n) = np_n(1-p_n) \to \infty$, $E(X_n) = np_n$. Dann folgt:

$$P(X_n = k) = \binom{n}{k} p_n^k (1-p_n)^{n-k} = \frac{e^{-np_n}\frac{(np_n)^k}{k!}\, e^{-n(1-p_n)}\frac{(n(1-p_n))^{n-k}}{(n-k)!}}{e^{-n}\frac{n^n}{n!}} \sim \frac{\frac{1}{\sqrt{2\pi np_n}} e^{-\frac{1}{2}\frac{(k-np_n)^2}{np_n}} \frac{1}{\sqrt{2\pi n(1-p_n)}} e^{-\frac{1}{2}\frac{((n-k)-n(1-p_n))^2}{n(1-p_n)}}}{\frac{1}{\sqrt{2\pi n}}}$$

für $n \to \infty$. Also erhalten wir

$$P(X_n = k) \sim \frac{1}{\sqrt{2\pi np_n(1-p_n)}} e^{-\frac{1}{2}\frac{(k-np_n)^2}{np_n(1-p_n)}}$$

für $n \to \infty$ gleichmäßig in $k \in \mathbb{N}_0$ mit $|k - np_n| \le c\sqrt{np_n(1-p_n)}$, $c > 0$.

Satz 8.7 (Zentraler Grenzwertsatz von Laplace)
Für $k \to \infty$ gilt

$$P(X_n = k) \sim \frac{1}{\sqrt{2\pi np_n(1-p_n)}} e^{-\frac{1}{2}\frac{(k-np_n)^2}{np_n(1-p_n)}}$$

gleichmäßig in $k \in \mathbb{N}_0$ mit $|k - np_n| \le c\sqrt{np_n(1-p_n)}$, $c > 0$.

Satz 8.8 (Zentraler Grenzwertsatz von Moivre-Laplace, kumulative Form)
Für alle $-\infty < a < b < \infty$ gilt:

$$\lim_{n\to\infty} P\left(a \underset{(-)}{\le} \frac{X_n - np_n}{\sqrt{np_n(1-p_n)}} \underset{(-)}{\le} b\right) = \Phi(b) - \Phi(a) = \frac{1}{2\pi}\int_a^b e^{-\frac{t^2}{2}}\, dt.$$

Korollar 8.9
Es gilt

$$\lim_{n\to\infty} P\left(\frac{X_n - np_n}{\sqrt{np_n(1-p_n)}} \le x\right) = \Phi(x) \text{ für alle } x \in \mathbb{R}.$$

Sei $X \sim \mathcal{B}(n, p)$. Dann gilt:

$$P\big(a \underset{(-)}{\le} X \underset{(-)}{\le} b\big) = P\left(\frac{a-np}{\sqrt{np(1-p)}} \underset{(-)}{\le} \frac{X-np}{\sqrt{np(1-p)}} \underset{(-)}{\le} \frac{b-np}{\sqrt{np(1-p)}}\right) \approx \Phi\left(\frac{b-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a-np}{\sqrt{np(1-p)}}\right).$$

Satz 8.10 (Zentraler Grenzwertsatz für die Multinomialverteilung)
Sei $X = (X_1,\dots,X_s) \sim \mathcal{M}(1; p_1,\dots,p_s)$, $X_s = 1 - (X_1 + \dots + X_{s-1})$, $p_i \in (0,1)$, $p_1 + \dots + p_s = 1$, $i = 1,\dots,s$ und sei $Y = (X_1,\dots,X_{s-1})$ mit $E(Y) = (p_1,\dots,p_{s-1}) = p$, $C := \operatorname{Cov}(X,Y) = \operatorname{diag}(p_1,\dots,p_s) - pp^\intercal$. Dann gilt:

$$C^{-1} = \operatorname{diag}\left(\frac{1}{p_1},\dots,\frac{1}{p_{s-1}}\right) + \frac{1}{p_s} 1_{(s-1)\times(s-1)},$$

wobei $1_{(s-1)\times(s-1)}$ eine Matrix, bestehend nur aus 1-Einträgen ist.
Nun seien $X^{(n)} = (X_1^{(n)},\dots,X_s^{(n)}) \sim \mathcal{M}(n; p_1,\dots,p_s)$, $X_s^{(n)} = n - (X_1^{(n)} + \dots + X_{s-1}^{(n)})$, $Y^{(n)} = (X_1^{(n)},\dots,X_{s-1}^{(n)})$.
Dann gilt

$$\begin{aligned} P(X^{(n)} = (k_1,\dots,k_s)) &= \frac{n!}{k_1! \cdot \ldots \cdot k_s!} p_1^{k_1} \cdot \ldots \cdot p_s^{k_s} \\ &= \frac{e^{-np_1}\frac{(np_1)^{k_1}}{k_1!} \cdot \ldots \cdot e^{-np_s}\frac{(np_s)^{k_s}}{k_s!}}{e^{-n}\frac{n^n}{n!}} \sim \frac{\frac{1}{\sqrt{2\pi np_1}} e^{-\frac{1}{2}\frac{(k_1-np_1)^2}{np_1}} \cdot \ldots \cdot \frac{1}{\sqrt{2\pi np_s}} e^{-\frac{1}{2}\frac{(k_s-np_s)^2}{np_s}}}{\frac{1}{\sqrt{2\pi n}}} \end{aligned}$$

gleichmäßig in $(k_1,\dots,k_s) \in \mathbb{N}_0^s$ mit $k_1 + \dots + k_s = n$ und $|k_i - np_i| \le c\sqrt{n}$ für $i = 1,\dots,s$ und $c > 0$.

Satz 8.11 (Zentraler Grenzwertsatz für die Multinomialverteilung, lokale Form)
Es gilt für $n \to \infty$:

$$P(X^{(n)} = (k_1,\dots,k_s)) \sim \left(\frac{1}{\sqrt{2\pi n}}\right)^{s-1} \frac{1}{\sqrt{p_1 \cdot \ldots \cdot p_s}} e^{-\frac{1}{2}\sum_{i=1}^s \frac{(k_i-np_i)^2}{np_i}}$$

gleichmäßig in $(k_1,\dots,k_s) \in \mathbb{N}_0^s$ mit $k_1 + \dots + k_s = n$, $|k_i - np_i| \le c\sqrt{n}$ für $i = 1,\dots,s$ und $c > 0$.

Satz 8.12 (Zentraler Grenzwertsatz für die Multinomialverteilung, kumulative Form)
Sei $Q = (a_1, b_1] \times \dots \times (a_{s-1}, b_{s-1}]$, $-\infty < a_i < b_i < \infty$, $i = 1,\dots,s-1$. Dann gilt:

$$\begin{aligned}P\left(\frac{1}{\sqrt{n}}(Y^{(n)}-np)\in Q\right) &= \sum_{w\in Q_n} P(Y^{(n)}=np+\sqrt{n}w)) \\ &= \left[\sum_{w\in Q_n}\left(\frac{1}{\sqrt{2\pi n}}\right)^{s-1}\frac{1}{\sqrt{p_n\cdot\ldots\cdot p_s}}e^{-\frac{1}{2}\sum_{i=1}^{s}\frac{w_i^2}{p_i}}\right](1+0 \text{ bzw. } 1) \\ &= \left[\sum_{w\in Q_n}\left(\frac{1}{\sqrt{2\pi n}}\right)^{s-1}\frac{1}{\sqrt{\det C}}e^{-\frac{1}{2}w^\top C^{-1}w}\right](1+0 \text{ bzw. } 1)\end{aligned}$$

mit

$$Q_n=\left\{\left(\frac{k_1-np_1}{\sqrt{n}},\ldots,\frac{k_{s-1}-np_{s-1}}{\sqrt{n}}\right)\in Q \mid (k_1,\ldots,k_{s-1})\in\mathbb{N}_0^{s-1}, k_1+\ldots+k_{s-1}\le n\right\}$$

und $k_s=n-(k_1+\ldots+k_{s-1})$, $w_s=-(w_1+\ldots+w_{s-1})$. Also erhalten wir:

$$\lim_{n\to\infty}P\left(\frac{1}{\sqrt{n}}(Y^{(n)}-np)\in Q\right)=\left(\frac{1}{\sqrt{2\pi}}\right)^{s-1}\int_Q\frac{1}{\sqrt{\det C}}e^{-\frac{1}{2}w^\top Cw}\,dw\,.$$

Sei nun $E=\{w\in\mathbb{R}^{s-1}\mid w^\top C^{-1}w\le z\}$, $z>0$, ein Ellipsoid. Analog gilt:

$$\begin{aligned}P\left(\frac{1}{\sqrt{n}}(Y^{(n)}-np)\in E\right) &= P\left((Y^{(n)}-np)^\top C^{-1}(Y^{(n)}-np)\le nz\right)=\left(\frac{1}{\sqrt{2\pi}}\right)^{s-1}\int_E\frac{1}{\sqrt{\det C}}e^{-\frac{1}{2}w^\top C^{-1}w}\,dw \\ &= \int_{\sum_{i=1}^{s-1}z_i^2\le z}\left(\frac{1}{\sqrt{2\pi}}\right)^{s-1}e^{-\sum_{i=1}^{s-1}z_i^2}\,dz_1\ldots dz_{s-1}.\end{aligned}$$

Also gilt:

Satz 8.13

Es gilt für $n\to\infty$:

$$P\left(\sum_{i=1}^{s}\frac{(X_i^{(n)}-np_i)^2}{np_i}\le z\right)\to\left(\frac{1}{\sqrt{2\pi}}\right)^{s-1}\int_{\sum_{i=1}^{s-1}z_i^2\le z}e^{-\frac{1}{2}\sum_{i=1}^{s-1}z_i^2}\,dz_1\ldots dz_{s-1}$$

für jedes $z>0$.

Definition 8.14

Die von L. Euler eingeführte **Gammafunktion** ist für komplexe Zahlen definiert. Wir betrachten sie nur für reelle Zahlen $x>0$:

$$\Gamma(x)=\int_0^\infty t^{x-1}e^{-t}\,dt\,.$$

Es gilt $\Gamma(x+1)=x\Gamma(x)$ für jedes $x>0$ und $\Gamma(1)=1$. Also ist $\Gamma(n)=(n-1)!$ für $n\in\mathbb{N}$.

Satz 8.15

Sei $X^{(n)}=(X_1^{(n)},\ldots,X_s^{(n)})\sim\mathcal{M}(n;p_1,\ldots,p_s)$, $(p_1,\ldots,p_s)\in(0,1)^s$, $p_1+\ldots+p_s=1$. Dann gilt:

$$P\left(\sum_{i=1}^{s}\frac{(X_i^{(n)}-np_i)^2}{np_i}\le z\right)\to\left(\frac{1}{\sqrt{2\pi}}\right)^{s-1}\int_{\sum_{i=1}^{s-1}w_i^2}e^{-\frac{1}{2}\sum_{i=1}^{s-1}w_i^2}\,dw_1\ldots dw_{s-1}=\frac{1}{2^{\frac{s-1}{2}}\Gamma\left(\frac{s-1}{2}\right)}\int_0^z w^{\frac{s-1}{2}-1}e^{-\frac{w}{2}}\,dw$$

für $z>0$. Es gilt also

$$\lim_{n\to\infty}P\left(\sum_{i=1}^{s}\frac{(x_i^{(n)}-np_i)^2}{np_i}\le z\right)=\frac{1}{2^{\frac{s-1}{2}}\Gamma\left(\frac{s-1}{2}\right)}\int_0^z w^{\frac{s-1}{2}-1}e^{-\frac{w}{2}}\,dw$$

für jedes $z>0$.

Bemerkung 8.16

Sind $X_1,\ldots,X_n$ unabhängige und identisch verteilte reelle Zufallsvariablen mit $\mu=E(X_i)\in\mathbb{R}$ und $0<\sigma^2=\mathrm{Var}(X_1)<\infty$, so gilt

$$\lim_{n\to\infty}P\left(\left(\frac{1}{n}\sum_{i=1}^{n}X_i-\mu\right)\sqrt{n}\le z\right)=\Phi(z)\ \text{ für jedes } z\in\mathbb{R}\,.$$

Satz 8.17 (Zentraler Grenzwertsatz von Lindberg-Levy)

Seien $X_1,\ldots,X_n$ unabhängige Zufallsvariablen mit derselben Verteilung $\mathcal{B}(1,p)$. Dann gilt $\sum_{i=1}^{n}X_i\sim\mathcal{B}(n,p)$ mit $\mu=p$, $\sigma^2=p(1-p)$, und es gilt für $n\to\infty$

$$P\left(\frac{\sum_{i=1}^{n}X_i-np}{\sqrt{np(1-p)}}\le z\right)\to\Phi(z)\ \text{ für alle } z\in\mathbb{R}\,.$$

Seien $X_1,...,X_n$ unabhängig, je mit derselben Verteilung $\mathcal{P}(\lambda)$, $\lambda > 0$. Dann gilt $\mu = \lambda = \sigma^2$ und

$$P\left(\frac{\sum_{i=1}^n X_i - n\lambda}{\sqrt{n\lambda}} \leq z\right) \to \Phi(z) \text{ für alle } z \in \mathbb{R} \text{ und } n \to \infty .$$

Außerdem gilt:

$$\sum_{i=1}^n X_i \sim \mathcal{P}(n\lambda).$$

Allgemeiner gilt: Sind X,Y unabhängige Zufallsvariablen mit $X \sim \mathcal{P}(\lambda_1)$, $Y \sim \mathcal{P}(\lambda_2)$. Dann folgt:

$$X + Y \sim \mathcal{P}(\lambda_1 + \lambda_2).$$

Denn: Für die erzeugenden Funktionen von $X+Y$ gilt:

$$f_{X+Y}(z) = f_X(t) f_Y(t) = e^{\lambda_1(t-1)} e^{\lambda_2(t-1)} = e^{(\lambda_1+\lambda_2)(t-1)} \text{ für } t \in \mathbb{R}.$$

Sind X,Y unabhängig mit $X \sim \mathcal{B}(m,p)$, $Y \sim \mathcal{B}(n,p)$, so ist $X+Y \sim \mathcal{B}(m+n,p)$.
Sind X,Y unabhängig mit $X \sim \mathcal{N}b(r,p)$ auf $\mathbb{N}_0$, $Y \sim \mathcal{N}b(s,p)$ auf $\mathbb{N}_0$, so gilt dann:

$$X + Y \sim \mathcal{N}b(r+s,p) \text{ auf } \mathbb{N}_0.$$

Es sei $z > 0$, $s \in \mathbb{N}$, $s \geq 2$, $E = \{x \in \mathbb{R}^{s-1} \mid x^\intercal C^{-1} x \leq z\}$ mit $C \in \mathbb{R}^{(s-1)\times(s-1)}$ eine symmetrische positiv definite Matrix und sei

$$I = \left(\frac{1}{\sqrt{2\pi}}\right)^{s-1} \frac{1}{\sqrt{\det C}} \int_E e^{-\frac{1}{2}x^\intercal C^{-1} x}\, dx .$$

Die Substitution $x = C^{1/2} y$, wobei $C^{1/2} \in \mathbb{R}^{(s-1)\times(s-1)}$ symmetrisch und positiv definit mit $C = C^{1/2} C^{1/2}$ ist, führt auf die Darstellung

$$I = \left(\frac{1}{\sqrt{2\pi}}\right)^{s-1} I_0 ,$$

wobei I_0 das Integral

$$I_0 = \int_{y^\intercal y \leq z} e^{-\frac{1}{2}y^\intercal y}\, dy = \int_{\sum_{i=1}^{s-1} y_i^2 \leq z} e^{-\frac{1}{2}\sum_{i=1}^{s-1} y_i^2}\, dy_1 \ldots dy_{s-1}$$

ist. Im Fall $s > 2$ lässt sich dieses Integral durch Transformation auf Kugelkoordinaten

$$r > 0,\ 0 < \varphi_i \leq \pi,\ i = 1,...,s-3,\ 0 < \varphi_{s-2} \leq 2\pi,$$

$$\begin{aligned}
y_1 &= r \sin\varphi_1 \cdot \ldots \cdot \sin\varphi_{s-4} \sin\varphi_{s-3} \sin\varphi_{s-2} \\
y_2 &= r \sin\varphi_1 \cdot \ldots \cdot \sin\varphi_{s-4} \sin\varphi_{s-3} \cos\varphi_{s-2} \\
y_3 &= r \sin\varphi_1 \cdot \ldots \cdot \sin\varphi_{s-4} \cos\varphi_{s-3} \\
&\vdots \\
y_{s-1} &= r \cos\varphi_1,
\end{aligned}$$

wobei der Betrag der Funktionaldeterminante dieser Transformation

$$r^{s-2} \prod_{i=1}^{s-3} (\sin\varphi_i)^{s-2-i}$$

ist, in der Form

$$I_0 = 2\pi \int_0^{\sqrt{z}} r^{s-2} e^{-\frac{r^2}{2}}\, dr \prod_{k=1}^{s-3} \int_0^\pi (\sin\varphi)^k\, d\varphi$$

bringen. Es ist

$$\int_0^\pi (\sin\varphi)^k\, d\varphi = 2\int_0^{\pi/2} (\sin\varphi)^k\, d\varphi .$$

Die Substitution $\varphi = \arcsin\sqrt{x}$ liefert

$$\int_0^\pi (\sin\varphi)^k\, d\varphi = \int_0^1 x^{\frac{k+1}{2}-1} (1-x)^{\frac{1}{2}-1}\, dx = \frac{\Gamma\left(\frac{k+1}{2}\right)\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{k+2}{2}\right)} = \sqrt{\pi}\, \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k+2}{2}\right)}.$$

Hieraus folgt

$$\prod_{k=1}^{s-3} \int_0^\pi (\sin\varphi)^k\, d\varphi = \frac{\pi^{\frac{s-3}{2}}}{\Gamma\left(\frac{s-1}{2}\right)}.$$

Wegen

$$\int_0^{\sqrt{z}} r^{s-2} e^{-\frac{r^2}{2}}\, dr = \frac{1}{2} \int_0^z w^{\frac{s-1}{2}-1} e^{-\frac{w}{2}}\, dw$$

erhalten wir insgesamt

$$I = \frac{1}{2^{\frac{s-1}{2}} \Gamma\left(\frac{s-1}{2}\right)} \int_0^z w^{\frac{s-1}{2}-1} e^{-\frac{w}{2}}\, dw\,.$$

Diese Darstellung für I gilt auch im Fall $s = 2$.