KAPITEL 8

Zufallsvariablen: Die allgemeine Definition

8.1. Zufallsvariablen

Bis zu diesem Zeitpunkt haben wir ausschließlich Zufallsvariablen mit endlich oder abzählbar vielen Werten (also diskrete Zufallsvariablen) betrachtet. Jetzt werden wir allgemeine Zufallsvariablen einführen.

Definition 8.1.1. Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Eine Funktion $X : \Omega \to \mathbb{R}$ heißt *messbar*, wenn für alle $a \in \mathbb{R}$ gilt:

$$\{X \leq a\} \in \mathcal{F}.$$

Hierbei ist $\{X \leq a\}$ die Menge aller Punkte im Wahrscheinlichkeitsraum, wo die Funktion X einen Wert $\leq a$ annimmt:

$$\{X \leq a\} = \{\omega \in \Omega : X(\omega) \leq a\} \subset \Omega.$$

Eine messbare Funktion nennen wir auch eine *Zufallsvariable*.

Für eine Zufallsvariable X ist also die Wahrscheinlichkeit $\mathbb{P}[X \leq a]$ wohldefiniert. Der nächste Satz besagt, dass auch die Wahrscheinlichkeit $\mathbb{P}[X \in B]$ wohldefiniert ist, wobei $B \subset \mathbb{R}$ eine beliebige Borel-Menge ist.

Satz 8.1.2. *Sei $X : \Omega \to \mathbb{R}$ eine Zufallsvariable. Dann gilt für jede Borel-Menge $B \subset \mathbb{R}$:*

$$\{X \in B\} \in \mathcal{F}.$$

Hierbei ist

$$\{X \in B\} = X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}.$$

Bemerkung 8.1.3. Aus diesem Satz folgt, dass für eine Zufallsvariable X gilt:

(1) Das Ereignis $\{X = a\}$ ist messbar, für alle $a \in \mathbb{R}$.
(2) Das Ereignis $\{X \in A\}$ ist messbar, für jede höchstens abzählbare Menge $A \subset \mathbb{R}$.
(3) Somit ist auch das Ereignis $\{X \notin A\} = \{X \in A\}^c$ ebenfalls messbar, für jede höchstens abzählbare Menge $A \subset \mathbb{R}$.
(4) Insbesondere ist das Ereignis $\{X \in \mathbb{Q}\}$ messbar.
(5) Ereignisse $\{a < X < b\}$, $\{a \leq X \leq b\}$, $\{a < X \leq b\}$, $\{a \leq X < b\}$ sind messbar.

Für den Beweis von Satz 8.1.2 benötigen wir eine Hilfsaussage.

Proposition 8.1.4. *Seien $(\Omega, \mathcal{F})$ ein Messraum und E eine Menge. Außerdem seien $X : \Omega \to E$ eine Abbildung und $\mathcal{E} \subset 2^E$ eine Mengenfamilie mit der Eigenschaft, dass $X^{-1}(B) \in \mathcal{F}$ für jede Menge $B \in \mathcal{E}$. Dann gilt auch $X^{-1}(B) \in \mathcal{F}$ für jede Menge $B \in \sigma(\mathcal{E})$.*

Bemerkung 8.1.5. Mit anderen Worten: Um zu zeigen, dass die Urbilder aller Mengen aus einer σ-Algebra messbar sind, reicht es zu zeigen, dass die Urbilder aller Mengen aus einem Erzeuger dieser σ-Algebra messbar sind.

Beweis von Proposition 8.1.4. Wir wollen zeigen, dass das Urbild jeder Menge aus $\sigma(\mathcal{E})$ ein Element von $\mathcal{F}$ ist. Deshalb betrachten wir die Familie

$$\mathcal{A} = \{B \subset E : X^{-1}(B) \in \mathcal{F}\} \subset 2^E.$$

Wir werden im Weiteren zeigen, dass die Familie $\mathcal{A}$ eine σ-Algebra ist. Außerdem gilt $\mathcal{E} \subset \mathcal{A}$ laut Voraussetzung. Die Familie $\mathcal{A}$ ist also eine σ-Algebra, die $\mathcal{E}$ enthält. Die von $\mathcal{E}$ erzeugte σ-Algebra $\sigma(\mathcal{E})$ ist (laut Definition der erzeugten σ-Algebra) die *kleinste* σ-Algebra, die $\mathcal{E}$ enthält. Somit muss $\sigma(\mathcal{E}) \subset \mathcal{A}$ gelten. Für jede Menge $B \in \sigma(\mathcal{E})$ gilt dann $B \in \mathcal{A}$. Laut Definition von $\mathcal{A}$ bedeutet das, dass $X^{-1}(B) \in \mathcal{F}$ für jede Menge $B \in \sigma(\mathcal{E})$. Das beweist die Behauptung der Proposition.

Wir werden nun zeigen, dass für die Familie $\mathcal{A}$ alle drei Bedingingen aus der Definition einer σ-Algebra gelten.

Bedingung 1. Es gilt $E \in \mathcal{A}$, denn $X^{-1}(E) = \Omega$ und $\Omega \in \mathcal{F}$.

Bedingung 2. Wir zeigen, dass $\mathcal{A}$ komplementstabil ist. Sei also $A \in \mathcal{A}$. Wir zeigen, dass $A^c \in \mathcal{A}$. Es gilt

$$X^{-1}(A^c) = \{\omega \in \Omega : X(\omega) \in A^c\} = \{\omega \in \Omega : X(\omega) \notin A\} = \{\omega \in \Omega : X(\omega) \in A\}^c = (X^{-1}(A))^c.$$

Aus $A \in \mathcal{A}$ folgt, dass $X^{-1}(A) \in \mathcal{F}$. Außerdem ist die Familie $\mathcal{F}$ eine σ-Algebra und somit komplementstabil. Es folgt, dass $X^{-1}(A^c) = (X^{-1}(A))^c \in \mathcal{F}$. Das bedeutet aber, dass $A^c \in \mathcal{A}$.

Bedingung 3. Schließlich zeigen wir, dass die Familie $\mathcal{A}$ σ-vereinigungsstabil ist. Seien also $A_1, A_2, \ldots \in \mathcal{A}$. Wir zeigen, dass $\cup_{n\in\mathbb{N}} A_n \in \mathcal{A}$. Es gilt

$$X^{-1}(\cup_{n\in\mathbb{N}} A_n) = \{\omega \in \Omega : X(\omega) \in \cup_{n\in\mathbb{N}} A_n\} = \cup_{n\in\mathbb{N}}\{\omega \in \Omega : X(\omega) \in A_n\} = \cup_{n\in\mathbb{N}} X^{-1}(A_n).$$

Aus $A_n \in \mathcal{A}$ folgt, dass $X^{-1}(A_n) \in \mathcal{F}$ für alle $n \in \mathbb{N}$. Außerdem ist die Familie $\mathcal{F}$ eine σ-Algebra und somit σ-vereinigungsstabil. Es folgt, dass $X^{-1}(\cup_{n\in\mathbb{N}} A_n) = \cup_{n\in\mathbb{N}} X^{-1}(A_n) \in \mathcal{F}$. Das bedeutet, dass $\cup_{n\in\mathbb{N}} A_n \in \mathcal{A}$.

Somit haben wir gezeigt, dass die Mengenfamilie $\mathcal{A}$ eine σ-Algebra ist. □

Beweis von Satz 8.1.2. Betrachte die Mengenfamilie

$$\mathcal{E} = \{(-\infty, a], a \in \mathbb{R}\} \subset 2^{\mathbb{R}}.$$

Da $X : \Omega \to \mathbb{R}$ messbar ist, gilt $X^{-1}(B) = \{X \leq a\} \in \mathcal{F}$ für jedes $B = (-\infty, a] \in \mathcal{E}$. Proposition 8.1.4 besagt, dass $X^{-1}(B) \in \mathcal{F}$ für alle $B \in \sigma(\mathcal{E})$. Dabei ist aber $\sigma(\mathcal{E})$ nichts anderes als die Borel-σ-Algebra. □

Beispiel 8.1.6. Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $A \in \mathcal{F}$ ein messbares Ereignis. Wir zeigen, dass die Indikatorfunktion von A

$$X(\omega) = \mathbb{1}_A(\omega) = \begin{cases} 1, & \omega \in A, \\ 0, & \omega \in A^c \end{cases}$$

eine Zufallsvariable ist.

Lösung. Sei $a \in \mathbb{R}$ beliebig. Wir betrachten das Ereignis

$$\{X \le a\} = \begin{cases} \Omega, & a \ge 1, \\ \emptyset, & a < 0, \\ A^c, & a \in [0, 1). \end{cases}$$

Es gilt $\Omega, \emptyset, A^c \in \mathcal{F}$, denn $A \in \mathcal{F}$ und $\mathcal{F}$ ist eine σ-Algebra. Somit ist X messbar.

8.2. Zufallsvektoren

Definition 8.2.1. Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Seien $X_1, \ldots, X_d : \Omega \to \mathbb{R}$ Funktionen, wobei $d \in \mathbb{N}$. Betrachte nun die Funktion $X : \Omega \to \mathbb{R}^d$ mit

$$X(\omega) = (X_1(\omega), \ldots, X_d(\omega)) \in \mathbb{R}^d.$$

Die Funktion X heißt ein d-dimensionaler *Zufallsvektor* (oder *messbar*), wenn $X_1, \ldots, X_d$ messbar sind.

Satz 8.2.2. *Sei $X : \Omega \to \mathbb{R}^d$ eine Funktion. Folgende Bedingungen sind äquivalent:*

(1) *X ist ein Zufallsvektor.*
(2) *Für jede Borel-Menge $B \subset \mathbb{R}^d$ gilt $\{X \in B\} \in \mathcal{F}$.*

Beweis von (1) $\Rightarrow$ (2). Seien $X_1, \ldots, X_d$ messbar. Sei $A = (-\infty, a_1] \times \ldots \times (-\infty, a_d]$ ein “Oktant”. Dann gilt:

$$X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\} = \{\omega \in \Omega : X_1(\omega) \le a_1, \ldots, X_d(\omega) \le a_d\} = \cap_{k=1}^d \{X_k \le a_k\}.$$

Wegen der Messbarkeit von X_k gilt $\{X_k \le a_k\} \in \mathcal{F}$ für alle $k = 1, \ldots, d$. Da $\mathcal{F}$ eine σ-Algebra ist, folgt, dass $X^{-1}(A) \in \mathcal{F}$. Wir haben gezeigt, dass das Urbild jedes Oktanten messbar ist. Die Familie der Oktanten erzeugt die Borel-σ-Algebra $\mathcal{B}^d$. Mit Proposition 8.1.4 folgt daraus, dass das Urbild jeder Borel-Menge messbar ist. □

Beweis von (2) $\Rightarrow$ (1). Wir nehmen an, dass für jede Borel-Menge $B \subset \mathbb{R}^d$ gilt, dass $\{X \in B\} \in \mathcal{F}$. Sei $k \in \{1, \ldots, d\}$ fest. Sei $B = \{(x_1, \ldots, x_d) \in \mathbb{R}^d : x_k \le a\}$. Diese Menge ist Borel, da abgeschlossen. Es folgt, dass $X^{-1}(B) = \{X_k \le a\} \in \mathcal{F}$. Somit ist die Funktion X_k messbar. Das gilt für jedes $k \in \{1, \ldots, d\}$. Somit ist X messbar. □

Die Familie der Borel-Mengen in $\mathbb{R}^d$ wird mit $\mathcal{B}^d$ bezeichnet.

Definition 8.2.3. Eine Funktion $f : \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$ heißt *Borel-messbar* (oder *Borel-Funktion*), wenn gilt:

$$f^{-1}(A) \in \mathcal{B}^{d_1} \text{ für alle } A \in \mathcal{B}^{d_2}.$$

Bemerkung 8.2.4. Eine Funktion ist also Borel-messbar, wenn das Urbild jeder Borel-Menge wieder eine Borel-Menge ist. Zum Vergleich: Eine Funktion ist stetig, wenn das Urbild jeder offenen Menge offen ist.

Proposition 8.2.5. *Jede stetige Funktion $f : \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$ ist Borel-messbar.*

Beweis. Die Funktion f sei stetig. Es folgt, dass für jede offene Menge $A \subset \mathbb{R}^{d_2}$ das Urbild $f^{-1}(A)$ offen ist. Das Urbild jeder offenen Menge ist also eine Borel-Menge. Die Familie der offenen Mengen erzeugt die Borel-σ-Algebra. Mit Proposition 8.1.4 folgt, dass auch das Urbild jeder Borel-Menge eine Borel-Menge ist. Somit ist f Borel-messbar. $\square$

Satz 8.2.6. *Sei $X : \Omega \to \mathbb{R}^{d_1}$ ein Zufallsvektor und $f : \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$ eine Borel-Funktion. Dann ist auch die Verknüpfung*

$$f \circ X : \Omega \to \mathbb{R}^{d_2}$$

ein Zufallsvektor.

Beweis. Sei $A \in \mathcal{B}^{d_2}$. Dann gilt $f^{-1}(A) \in \mathcal{B}^{d_1}$, denn f ist eine Borel-Funktion. Es gilt

$$(f \circ X)^{-1}(A) = X^{-1}\left(f^{-1}(A)\right) \in \mathcal{F},$$

da X messbar ist. Nach Satz 8.2.2 ist $f \circ X$ ein Zufallsvektor. $\square$

Korollar 8.2.7. *Sind X, Y Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$, so sind auch $X + Y$, $X \cdot Y$ und $a \cdot X$, wobei $a \in \mathbb{R}$, Zufallsvariablen.*

Beweis. Die Funktionen $(x, y) \mapsto x + y$, $(x, y) \mapsto xy$, $x \mapsto ax$ sind Borel-Funktionen, da sie stetig sind. Die Behauptung folgt aus Satz 8.2.6. $\square$

8.3. Verteilungsfunktion einer Zufallsvariable

Die Familie der Borel-Teilmengen von $\mathbb{R}$ wird mit $\mathcal{B}$ bezeichnet.

Definition 8.3.1. Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Sei $X : \Omega \to \mathbb{R}$ eine Zufallsvariable.

(1) Die *Verteilung* von X ist die Funktion

$$P_X : \mathcal{B} \to [0,1] \text{ mit } P_X(A) = \mathbb{P}[X \in A], \;\; A \in \mathcal{B}.$$

P_X ist ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B})$.

(2) Die *Verteilungsfunktion* von X ist die Funktion

$$F_X : \mathbb{R} \to [0,1] \text{ mit } F_X(t) = \mathbb{P}[X \le t], \;\; t \in \mathbb{R}.$$

Beispiel 8.3.2. Im Einheitskreis werde ein Punkt (X, Y) zufällig und gleichverteilt gewählt. Es sei R der Abstand von (X, Y) zum Mittelpunkt des Kreises. Bestimme die Verteilungsfunktion von R.

Lösung. Als Grundmenge wählen wir den Einheitskreis $\Omega = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$. Sei $\mathcal{F} = \mathcal{B}^2_\Omega$ die σ-Algebra der Borel-Teilmengen von Ω. Als Wahrscheinlichkeitsmaß wählen wir

$$\mathbb{P}[A] = \frac{\lambda(A)}{\pi}, \;\; A \in \mathcal{F},$$

wobei λ das Lebesgue-Maß ist. Das entspricht der Annahme, dass der Punkt gleichverteilt ist. Der Abstand zum Ursprung ist dann die Zufallsvariable $R : \Omega \to \mathbb{R}$ mit

$$R(x,y) = \sqrt{x^2+y^2}, \;\; (x,y) \in \Omega.$$

Beachte, dass R stetig und somit messbar ist. Um die Verteilungsfunktion von R zu bestimmen, schauen wir uns das Ereignis $\{R \leq t\}$ an:

$$\{R \leq t\} = \{(x,y) \in \Omega : \sqrt{x^2+y^2} \leq t\} = \begin{cases} \Omega, & t \geq 1, \\ \emptyset, & t < 0, \\ \text{Kreis vom Radius } t, & t \in [0,1]. \end{cases}$$

Es folgt, dass

$$F_R(t) = \mathbb{P}[R \leq t] = \begin{cases} 1, & t \geq 1, \\ 0, & t < 0, \\ \frac{\pi t^2}{\pi}, & t \in [0,1] \end{cases} = \begin{cases} 1, & t \geq 1, \\ 0, & t < 0, \\ t^2, & t \in [0,1]. \end{cases}$$

Dies ist die Verteilungsfunktion von R.

Beispiel 8.3.3. Ein Zufallsgenerator erzeugt zwei unabhängige und in $[0,1]$ gleichverteilte Zufallszahlen X, Y. Bestimme die Verteilungsfunktion von $Z := X + Y$.

Lösung. Als Grundmenge wählen wir $\Omega = [0,1]^2$. Sei $\mathcal{F} = \mathcal{B}^2_\Omega$ die Einschränkung der Borel-σ-Algebra $\mathcal{B}^2$ auf Ω. Die Bedingung der Gleichverteilung und Unabhängigkeit von X und Y wird so interpretiert: die Wahrscheinlichkeit eines Ereignisses $A \in \mathcal{F}$ ist

$$\mathbb{P}[A] = \lambda(A).$$

Wir können die Zufallsvariablen $X, Y, Z : \Omega \to \mathbb{R}$ definieren:

$$X(x,y) = x, \;\; Y(x,y) = y, \;\; Z(x,y) = x + y, \;\; (x,y) \in [0,1]^2.$$

Das Ereignis, das uns hier interessiert, ist

$$\{Z \leq t\} = \{(x,y) \in [0,1]^2 : x + y \leq t\}.$$

Es gilt

$$\{Z \leq t\} = \begin{cases} \emptyset, & t < 0, \\ \text{gleichschenkliges rechtwinkliges Dreieck mit Kathetenlänge } t, & t \in [0,1], \\ \text{das Komplement eines solchen Dreiecks mit Kathetenlänge } 2-t, & t \in [1,2], \\ \Omega, & t \geq 2. \end{cases}$$

Somit erhalten wir

$$F_Z(t) = \mathbb{P}[Z \le t] = \begin{cases} 0, & t < 0, \\ \frac{t^2}{2}, & t \in [0,1], \\ 1 - \frac{(2-t)^2}{2}, & t \in [1,2], \\ 1, & t \ge 2. \end{cases}$$

Dies ist die Verteilungsfunktion von Z. Sie ist stetig.

Beispiel 8.3.4. Betrachte eine konstante Zufallsvariable, also $X = c$. Wie sieht dann die Verteilungsfunktion F_X aus?

Lösung. Es gilt

$$F_X(t) = \mathbb{P}[c \le t] = \begin{cases} 0, & t < c, \\ 1, & t \ge c. \end{cases}$$

Dieses Beispiel zeigt, dass eine Verteilungsfunktion Unstetigkeitsstellen haben kann.

Satz 8.3.5. *Sei X eine Zufallsvariable mit Verteilungsfunktion F_X. Dann hat F_X die folgenden drei Eigenschaften:*

(1) *Grenzwerte:* $\lim_{t\to-\infty} F_X(t) = 0$ *und* $\lim_{t\to+\infty} F_X(t) = 1$.
(2) *Monotonie: Für alle $t_1 \le t_2$ gilt $F_X(t_1) \le F_X(t_2)$.*
(3) *Rechtsstetigkeit: Für alle $t_0 \in \mathbb{R}$ gilt* $\lim_{t\downarrow t_0} F_X(t) = F_X(t_0)$.

Bemerkung 8.3.6. Der linksseitige Grenzwert $\lim_{t\uparrow t_0} F_X(t)$ existiert ebenfalls, da die Funktion F_X monoton ist. Allerdings muss der linksseitige Grenzwert nicht mit $F_X(t_0)$ übereinstimmen, siehe Beispiel 8.3.4 mit $t_0 = c$.

Beweis von (2). Seien $t_1 \le t_2$ beliebig. Betrachte die Ereignisse

$$\{X \le t_1\} = \{\omega \in \Omega : X(\omega) \le t_1\} \subset \{\omega \in \Omega : X(\omega) \le t_2\} = \{X \le t_2\}$$

Deshalb gilt für die Wahrscheinlichkeiten dieser Ereignisse $\mathbb{P}[X \le t_1] \le \mathbb{P}[X \le t_2]$, was gleichbedeutend ist mit $F_X(t_1) \le F_X(t_2)$. $\square$

Beweis von (1). Wir betrachten den Fall $t \to -\infty$. Führe die Ereignisse $A_n = \{X \le -n\}$ mit $n \in \mathbb{N}$ ein. Dann gilt $A_1 \supset A_2 \supset \ldots$ und $\cap_{n\in\mathbb{N}} A_n = \emptyset$. Aufgrund der Stetigkeit der Wahrscheinlichkeit folgt daraus, dass

$$\lim_{n\to\infty} F_X(-n) = \lim_{n\to\infty} \mathbb{P}[A_n] = \mathbb{P}[\emptyset] = 0.$$

Dabei darf allerdings n nur natürliche Werte annehmen. Für ein beliebiges $t < 0$ kann man immer ein $n \in \mathbb{N}$ mit $-n \le t < -n+1$ finden. Wegen der Monotonie von F und des Sandwichprinzips gilt dann

$$0 = \lim_{n\to\infty} F_X(-n) \le \lim_{t\to-\infty} F_X(t) \le \lim_{n\to\infty} F_X(-n+1) = 0.$$

Somit ist $\lim_{t\to-\infty} F_X(t) = 0$, wie behauptet. $\square$

Beweis von (3). Sei $t_0 \in \mathbb{R}$ beliebig. Wir definieren die Ereignisse $A_n = \{X \le t_0 + 1/n\}$ mit $n \in \mathbb{N}$. Es gilt dann $A_1 \supset A_2 \supset \dots$ und $\cap_{n\in\mathbb{N}} A_n = \{X \le t_0\}$. Aufgrund der Stetigkeit der Wahrscheinlichkeit gilt für den Grenzwert:

$$\lim_{n\to\infty} F_X\left(t_0 + \frac{1}{n}\right) = \lim_{n\to\infty} \mathbb{P}[A_n] = \mathbb{P}[X \le t_0] = F_X(t_0).$$

Das gilt wieder nur für $n \in \mathbb{N}$. Für ein beliebiges $t > t_0$ können wir $n \in \mathbb{N}$ mit $t_0 + \frac{1}{n} \le t < t_0 + \frac{1}{n-1}$ finden. Aufgrund der Monotonie von F_X und des Sandwichprinzips erhalten wir

$$F_X(t_0) = \lim_{n\to\infty} F_X\left(t_0 + \frac{1}{n}\right) \le \lim_{t\downarrow t_0} F_X(t) \le \lim_{n\to\infty} F_X\left(t_0 + \frac{1}{n-1}\right) = F_X(t_0).$$

Somit ist $\lim_{t\downarrow t_0} F_X(t) = F_X(t_0)$, wie behauptet. $\square$

Der nächste Satz besagt, dass die Verteilung einer Zufallsvariable durch ihre Verteilungsfunktion eindeutig festgelegt wird.

Satz 8.3.7. *Seien X_1 und X_2 Zufallsvariablen mit*

$$F_{X_1}(t) = F_{X_2}(t) \text{ für alle } t \in \mathbb{R}.$$

Dann gilt für alle Borel-Mengen $B \subset \mathbb{R}$:

$$\mathbb{P}[X_1 \in B] = \mathbb{P}[X_2 \in B].$$

Für den Beweis benötigen wir einen Satz aus der Maßtheorie.

Satz 8.3.8 (Eindeutigkeit der Maß-Fortsetzung). *Sei Ω eine Menge und $\mathcal{E} \subset 2^\Omega$ eine schnittstabile Mengenfamilie. Die Schnittstabilität bedeutet: für $A, B \in \mathcal{E}$ gilt auch $A \cap B \in \mathcal{E}$. Es sei $\mathcal{A} = \sigma(\mathcal{E})$ die von $\mathcal{E}$ erzeugte σ-Algebra. Seien $\mathbb{P}_1$ und $\mathbb{P}_2$ zwei Wahrscheinlichkeitsmaße auf $(\Omega, \mathcal{A})$ mit der Eigenschaft, dass*

$$\mathbb{P}_1[A] = \mathbb{P}_2[A] \text{ für alle } A \in \mathcal{E}.$$

Dann gilt sogar

$$\mathbb{P}_1[A] = \mathbb{P}_2[A] \text{ für alle } A \in \mathcal{A}.$$

Bemerkung 8.3.9. Mit anderen Worten: stimmen zwei Wahrscheinlichkeitsmaße auf dem Erzeuger einer σ-Algebra überein, so stimmen sie auch auf der ganzen σ-Algebra überein, wenn der Erzeuger schnittstabil ist.

Beweis von Satz 8.3.7. Die Mengenfamilie $\mathcal{E} = \{(-\infty, t], t \in \mathbb{R}\}$ ist schnittstabil, denn

$$(-\infty, t] \cap (-\infty, s] = (-\infty, \min(t, s)] \in \mathcal{E}.$$

Für die Verteilungen von X_1 und X_2 gilt nun:

$$P_{X_1}((-\infty, t]) = \mathbb{P}[X_1 \le t] = F_{X_1}(t) = F_{X_2}(t) = \mathbb{P}[X_2 \le t] = P_{X_2}((-\infty, t]).$$

Die Wahrscheinlichkeitsmaße P_{X_1} und P_{X_2} stimmen also auf $\mathcal{E}$ überein. Nach der Eindeutigkeit der Maß-Fortsetzung stimmen sie auch auf der Borel-σ-Algebra $\mathcal{B} = \sigma(\mathcal{E})$ überein.

Somit gilt $P_{X_1}(B) = P_{X_2}(B)$ für alle $B \in \mathcal{B}$. Mit anderen Worten, $\mathbb{P}[X_1 \in B] = \mathbb{P}[X_2 \in B]$ für alle $B \in \mathcal{B}$. □

Beispiel 8.3.10. Sei X eine Zufallsvariable mit Verteilungsfunktion $F_X(t) = \mathbb{P}[X \leq t]$. Bestimme $\mathbb{P}[X < t]$.

Lösung. Betrachte Ereignisse $A_n = \{X \leq t - \frac{1}{n}\}$ mit $n \in \mathbb{N}$. Es gilt $A_1 \subset A_2 \subset \ldots$ und $\cup_{n\in\mathbb{N}} A_n = \{X < t\}$. Mit der Stetigkeit der Wahrscheinlichkeit folgt, dass

$$\mathbb{P}[X < t] = \lim_{n\to\infty} \mathbb{P}[A_n] = \lim_{n\to\infty} \mathbb{P}\left[X \leq t - \frac{1}{n}\right] = \lim_{s\uparrow t} F_X(s).$$

Beachte: dieser Grenzwert muss im Allgemeinen nicht mit $F(t)$ übereinstimmen.

Beispiel 8.3.11. Sei X eine Zufallsvariable mit Verteilungsfunktion $F_X(t) = \mathbb{P}[X \leq t]$. Bestimme $\mathbb{P}[X = t]$.

Lösung. Es gilt

$$\mathbb{P}[X = t] = \mathbb{P}[X \leq t] - \mathbb{P}[X < t] = F_X(t) - \lim_{s\uparrow t} F_X(s).$$

Die Wahrscheinlichkeit, dass $X = t$ ist, ist also gleich dem Sprung, den die Verteilungsfunktion F_X an der Stelle t macht. Die Funktion F_X ist stetig an der Stelle t genau dann wenn $\mathbb{P}[X = t] = 0$.

Definition 8.3.12. Sei X eine Zufallsvariable. Ein Wert $t \in \mathbb{R}$ mit $\mathbb{P}[X = t] > 0$ heißt ein *Atom* von X. Atome sind also Unstetigkeitsstellen der Verteilungsfunktion F_X.

Proposition 8.3.13. *Jede Zufallsvariable hat höchstens abzählbar viele Atome. Mit anderen Worten, jede Verteilungsfunktion hat höchstens abzählbar viele Unstetigkeitsstellen (Sprünge).*

Beweis. Es gilt: F_X ist monoton, $\lim_{t\to-\infty} F_X(t) = 0$ und $\lim_{t\to+\infty} F_X(t) = 1$. Für jedes $n \in \mathbb{N}$ kann es also höchstens n Sprünge geben, die eine Höhe von $\geq 1/n$ haben. Sonst wäre die Summe der Sprunghöhen > 1, was ein Widerspruch ist. Die Menge der Sprünge ist eine abzählbare Vereinigung endlicher Mengen und somit selbst abzählbar. □

Beispiel 8.3.14. Sei X eine Zufallsvariable mit Verteilungsfunktion $F_X(t) = \mathbb{P}[X \leq t]$. Für $a < b$ bestimme $\mathbb{P}[a \leq X \leq b]$, $\mathbb{P}[a \leq X < b]$, $\mathbb{P}[a < X \leq b]$, $\mathbb{P}[a < X < b]$.

Lösung. Es gilt

$$\begin{aligned}
\mathbb{P}[a \leq X \leq b] &= \mathbb{P}[X \leq b] - \mathbb{P}[X < a] = F_X(b) - \lim_{s\uparrow a} F_X(s),\\
\mathbb{P}[a \leq X < b] &= \mathbb{P}[X < b] - \mathbb{P}[X < a] = \lim_{s\uparrow b} F_X(s) - \lim_{s\uparrow a} F_X(s),\\
\mathbb{P}[a < X \leq b] &= \mathbb{P}[X \leq b] - \mathbb{P}[X \leq a] = F_X(b) - F_X(a),\\
\mathbb{P}[a < X < b] &= \mathbb{P}[X < b] - \mathbb{P}[X \leq a] = \lim_{s\uparrow b} F_X(s) - F_X(a).
\end{aligned}$$

8.4. Definition und Eigenschaften des Erwartungswerts

Wir haben den Erwartungswert nur für Zufallsvariablen auf einem diskreten Wahrscheinlichkeitsraum definiert. Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein diskreter Wahrscheinlichkeitsraum und $X : \Omega \to \mathbb{R}$ eine Zufallsvariable. Eine der Definitionen des Erwartungswerts war:

$$\mathbb{E}X = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}[\{\omega\}].$$

Nun definieren wir den Erwartungswert für beliebige Zufallsvariablen. Sei dazu $(\Omega, \mathcal{F}, \mathbb{P})$ ein beliebiger Wahrscheinlichkeitsraum und $X : \Omega \to \mathbb{R}$ eine beliebige Zufallsvariable. Da der Wahrscheinlichkeitsraum Ω im allgemeinen nicht abzählbar ist, können wir die Summe $\sum_{\omega \in \Omega}$ nicht bilden. Wir werden die Summe durch das Integral (das sogenannte Lebesgue-Integral) ersetzen:

$$\mathbb{E}X = \int_\Omega X(\omega)\, d\mathbb{P}(\omega). \tag{8.4.1}$$

Nun definieren wir den Erwartungswert (bzw. das Lebesgue-Integral). Das soll in mehreren Schritten geschehen.

Sei $X : \Omega \to \mathbb{R}$ eine Zufallsvariable (d.h. eine messbare Funktion) auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$.

Schritt 1. (Elementarfunktionen).

Definition 8.4.1. Seien $A_1, \ldots, A_n \in \mathcal{F}$ paarweise disjunkte Mengen mit $\Omega = A_1 \cup \ldots \cup A_n$. Außerdem seien $y_1, \ldots, y_n \in \mathbb{R}$. Eine Elementarfunktion ist eine Funktion der Form

$$X = \sum_{k=1}^n y_k \mathbb{1}_{A_k}.$$

Ist nun $X : \Omega \to \mathbb{R}$ eine Elementarfunktion, so definieren wir den Erwartungswert von X wie folgt:

$$\mathbb{E}X \overset{def}{=} \sum_{k=1}^n y_k \mathbb{P}[A_k].$$

Schritt 2. (Nicht-negative, messbare Funktionen)

Sei $X : \Omega \to [0, \infty)$ eine messbare, nichtnegative Funktion. Nun definieren wir

$$\mathbb{E}X \overset{def}{=} \sup\{\mathbb{E}Y | Y : \Omega \to [0,\infty), \text{Elementarfunktion mit } 0 \le Y(\omega) \le X(\omega)\ \forall \omega \in \Omega\}.$$

Die Idee ist also, dass wir X von unten mit Elementarfunktionen approximieren. Der so definierte Erwartungswert $\mathbb{E}X$ nimmt nichtnegative Werte oder den Wert $+\infty$ an.

Schritt 3. (Beliebige, messbare Funktionen)

Sei $X : \Omega \to \mathbb{R}$ eine beliebige messbare Funktion. Wir werden X als eine Differenz von zwei nichtnegativen Funktionen X^+ und X^- darstellen. Definiere

$$X^+(\omega) \overset{def}{=} \begin{cases} X(\omega), & \text{falls } X(\omega) \ge 0, \\ 0, & \text{falls } X(\omega) < 0, \end{cases} \qquad X^-(\omega) \overset{def}{=} \begin{cases} 0, & \text{falls } X(\omega) \ge 0, \\ |X(\omega)|, & \text{falls } X(\omega) < 0. \end{cases}$$

Dann sind X^+ und X^- messbar (Übung) und es gilt

$$X^+, X^- \geq 0, \qquad X = X^+ - X^-, \quad |X| = X^+ + X^-.$$

Für die nichtnegativen Zufallsvariablen X^+ und X^- haben wir den Erwartungswert bereits in Schritt 2 definiert. Mit dieser Definition können wir nun folgende Fälle betrachten:

Fall 1: Gilt $\mathbb{E}X^+ < \infty$ und $\mathbb{E}X^- < \infty$, so heißt die Zufallsvariable X *integrierbar*. Bezeichnung: $X \in L^1$. Für eine integrierbare Zufallsvariable X definieren wir

$$\mathbb{E}X \stackrel{def}{=} \mathbb{E}X^+ - \mathbb{E}X^-.$$

In allen anderen Fällen heißt die Zufallsvariable X nicht integrierbar.

Fall 2: Gilt $\mathbb{E}X^+ = +\infty$ und $\mathbb{E}X^- < +\infty$, so definieren wir

$$\mathbb{E}X \stackrel{def}{=} +\infty.$$

Fall 3: Gilt $\mathbb{E}X^+ < +\infty$ und $\mathbb{E}X^- = +\infty$, so definieren wir

$$\mathbb{E}X \stackrel{def}{=} -\infty.$$

Fall 4: Gilt: $\mathbb{E}X^+ = \mathbb{E}X^- = +\infty$, so kann man den Erwartungswert nicht definieren.

Wir werden einige Eigenschaften des Erwartungswerts (bzw. des Lebesgue-Integrals) ohne Beweis auflisten.

Satz 8.4.2. *Für eine beliebige Zufallsvariable X gilt $\mathbb{E}|X| = \mathbb{E}X^+ + \mathbb{E}X^-$. Insbesondere ist X genau dann integrierbar, wenn $\mathbb{E}|X| < \infty$. Für eine integrierbare Zufallsvariable X gilt die Ungleichung*

$$|\mathbb{E}X| \leq \mathbb{E}|X|.$$

Satz 8.4.3. *Der Erwartungswert ist linear:*

(1) *Sind X und Y integrierbare Zufallsvariablen, so ist auch $X + Y$ integrierbar und es gilt*

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y].$$

(2) *Ist X integrierbar und ist $a \in \mathbb{R}$, so ist auch aX integrierbar und es gilt*

$$\mathbb{E}[aX] = a\mathbb{E}[X].$$

In dem Fall, wenn die Zufallsvariable höchstens abzählbar viele Werte annimmt, stimmt die neue Definition des Erwartungswerts mit der alten Definition überein.

Satz 8.4.4. *Sei $X : \Omega \to \mathbb{R}$ eine Zufallsvariable, die höchstens abzählbar viele Werte $y_1, y_2, \ldots$ annimmt. Dann ist X integrierbar genau dann, wenn*

$$\mathbb{E}|X| = \sum_n |y_n| \cdot \mathbb{P}[X = y_n] < \infty.$$

Ist X integrierbar, so gilt

$$\mathbb{E}X = \sum_n y_n \cdot \mathbb{P}[X = y_n].$$

Der Erwartungswert ist monoton:

SATZ 8.4.5. *Sei $Y : \Omega \to \mathbb{R}$ eine Zufallsvariable mit $\mathbb{E}Y < \infty$ und X eine weitere Zufallsvariable mit $X(\omega) \leq Y(\omega)$ für alle $\omega \in \Omega$. Dann gilt $\mathbb{E}X \leq \mathbb{E}Y$.*

Der Erwartungswert einer Zufallsvariable ändert sich nicht, wenn man die Werte der Zufallsvariable auf einer Nullmenge verändert. Dies wird im nächsten Satz beschrieben.

DEFINITION 8.4.6. Zwei Zufallsvariablen $X, Y : \Omega \to \mathbb{R}$ heißen *fast überall gleich*, wenn

$$\mathbb{P}[\{\omega \in \Omega : X(\omega) \neq Y(\omega)\}] = 0.$$

SATZ 8.4.7. *Sind X und Y fast überall gleich und eine der Zufallsvariablen integrierbar, so ist auch die andere Zufallsvariable integrierbar und es gilt $\mathbb{E}X = \mathbb{E}Y$.*

8.5. Diskrete und absolut stetige Verteilungen

DEFINITION 8.5.1. Eine Zufallsvariable X heißt *diskret*, wenn X nur endlich oder abzählbar viele Werte annimmt. Die *Zähldichte* von X ist die Funktion

$$p_X(y) = \mathbb{P}[X = y].$$

BEMERKUNG 8.5.2. Für die Verteilungsfunktion einer diskreten Zufallsvariable gilt

$$F_X(t) = \sum_{y \leq t} p_X(y) \text{ für alle } t \in \mathbb{R}.$$

Die Verteilungsfunktion einer diskreten Zufallsvariable ist eine "Sprungfunktion".

DEFINITION 8.5.3. Eine Zufallsvariable X mit Verteilungsfunktion F_X heißt *absolut stetig*, wenn es eine Borel-Funktion $f_X : \mathbb{R} \to [0, \infty)$ gibt, so dass

$$F_X(t) = \int_{-\infty}^{t} f_X(y) dy \text{ für alle } t \in \mathbb{R}.$$

Die Funktion f_X heißt die *Dichte von X*.

BEMERKUNG 8.5.4. Für die Dichte gelten die folgenden zwei Eigenschaften:

(1) $f_X(y) \geq 0$ für alle $y \in \mathbb{R}$.
(2) $\int_{\mathbb{R}} f_X(y) dy = 1$.

SATZ 8.5.5. *Sei X eine absolut stetige Zufallsvariable. Dann gilt für jede Borel-Menge $B \subset \mathbb{R}$*

$$\mathbb{P}[X \in B] = \int_B f_X(y) dy.$$

BEMERKUNG 8.5.6. Zum Vergleich: Im diskreten Fall gilt

$$\mathbb{P}[X \in B] = \sum_{y \in B} p_X(y).$$

BEWEIS. Definiere zwei Wahrscheinlichkeitsmaße auf $(\mathbb{R}, \mathcal{B})$:

$$\mathbb{P}_1(B) = \mathbb{P}[X \in B], \qquad \mathbb{P}_2(B) = \int_B f_X(y) dy, \qquad B \in \mathcal{B}.$$

Diese Wahrscheinlichkeitsmaße stimmen auf allen Mengen der Form $B = (-\infty, t]$ überein, denn

$$\mathbb{P}_2(B) = \int_{-\infty}^{t} f_X(y) dy = F_X(t) = \mathbb{P}[X \leq t] = \mathbb{P}_1(B).$$

Die Mengen der Form $(-\infty, t]$ bilden einen schnittstabilen Erzeuger der Borel-σ-Algebra. Durch die Eindeutigkeit der Maßfortsetzung folgt, dass

$$\mathbb{P}_1(B) = \mathbb{P}_2(B) \text{ für alle } B \in \mathcal{B}.$$

Somit gilt $\mathbb{P}[X \in B] = \int_B f_X(y)dy$ für alle $B \in \mathcal{B}$. □

Die Verteilungsfunktion ist das Integral der Dichte. Umgekehrt, ist die Dichte die Ableitung der Verteilungsfunktion. Das gilt allerdings nicht an allen, sondern an fast allen Stellen.

SATZ 8.5.7. *Sei X eine absolut stetige Zufallsvariable mit Dichte f_X und Verteilungsfunktion F_X. Dann gibt es eine Borel-Menge $N \subset \mathbb{R}$ mit Lebesgue-Maß 0, so dass die Funktion F_X differenzierbar an allen Stellen $t \in \mathbb{R}\backslash N$ ist und*

$$F_X'(t) = f_X(t) \textit{ für alle } t \in \mathbb{R}\backslash N.$$

OHNE BEWEIS.

Die wichtigsten Eigenschaften der diskreten und der absolut stetigen Verteilungen werden in der folgenden Tabelle zusammengefasst. Einige dieser Formeln werden wir später beweisen.

Diskrete Zufallsvariablen	**Absolut stetige Zufallsvariablen**
Zähldichte p_X	Dichte f_X
Verteilungsfunktion: $F_X(t) = \sum_{y \leq t} p_X(y)$	Verteilungsfunktion: $F_X(t) = \int_{-\infty}^t f_X(y)dy$
$p_X(y) \in [0,1]$ für alle $y \in \mathbb{R}$	$f_X(y) \geq 0$ für alle $y \in \mathbb{R}$
$\sum_{y \in \mathbb{R}} p_X(y) = 1$	$\int_{-\infty}^{\infty} f_X(y)dy = 1$
$p_X(y) = \mathbb{P}[X = y]$, $y \in \mathbb{R}$	$f_x(y) = F_X'(y) = \lim_{\varepsilon \downarrow 0} \frac{\mathbb{P}[y \leq X \leq y+\varepsilon]}{\varepsilon}$ f. ü.
$\mathbb{P}[X \in B] = \sum_{y \in B} p_X(y)$, $B \subset \mathbb{R}$ Borel	$\mathbb{P}[X \in B] = \int_B f_X(y)dy$, $B \subset \mathbb{R}$ Borel
X integrierbar, wenn $\sum_{y \in \mathbb{R}} \|y\| \cdot p_X(y) < \infty$	X integrierbar, wenn $\int_{\mathbb{R}} \|y\| \cdot f_X(y)dy < \infty$
Erwartungswert: $\mathbb{E}X = \sum_{y \in \mathbb{R}} y \cdot p_X(y)$	Erwartungswert: $\mathbb{E}X = \int_{\mathbb{R}} y \cdot f_X(y)dy$

8.6. Beispiele von absolut stetigen Verteilungenen

Gleichverteilung auf einem Intervall

DEFINITION 8.6.1. Eine Zufallsvariable X heißt *gleichverteilt* auf einem Intervall $[a,b]$, wobei $a, b \in \mathbb{R}$ und $a < b$, wenn X absolut stetig ist und die Dichte von X durch die folgende Formel gegeben ist:

$$f_X(y) = \begin{cases} \frac{1}{b-a}, & y \in [a,b], \\ 0, & y \notin [a,b]. \end{cases}$$

BEZEICHNUNG. $X \sim \mathrm{U}[a,b]$. Dabei steht “U” für “uniform”.

BEMERKUNG 8.6.2. Die Dichte ist also konstant auf dem Intervall $[a,b]$. Die Werte außerhalb von $[a,b]$ werden nicht angenommen, da die Dichte außerhalb von $[a,b]$ verschwindet. Die Konstante $\frac{1}{b-a}$ wurde so gewählt, dass die Bedingung $\int_{\mathbb{R}} f_X(y)dy = 1$ erfüllt ist.

Bemerkung 8.6.3. Die Verteilungsfunktion einer gleichverteilten Zufallsvariable $X \sim \mathrm{U}[a,b]$ ist

$$F_X(t) = \int_{-\infty}^t f_X(y)dy = \begin{cases} 0, & t \le a, \\ \frac{t-a}{b-a}, & t \in [a,b], \\ 1, & t \ge b. \end{cases}$$

Die Verteilungsfunktion ist also linear auf dem Intervall $[a,b]$ und konstant sonst. Die verteilungsfunktion ist stetig. Leitet man die Verteilungsfunktion ab, so erhält man die Dichte. Es gibt allerdings zwei Ausnahmestellen 0 und 1, an denen die Verteilungsfunktion nicht differenzierbar ist.

Satz 8.6.4. *Sei X eine Zufallsvariable, die auf $[a,b]$ gleichverteilt ist. Dann gilt*

$$\mathbb{E}X = \frac{a+b}{2}.$$

Beweis.

$$\mathbb{E}X = \int_{\mathbb{R}} y f_X(y)dy = \int_a^b \frac{y}{b-a}dy = \frac{1}{b-a}\int_a^b ydy = \frac{1}{b-a}\left(\frac{b^2}{2} - \frac{a^2}{2}\right) = \frac{a+b}{2}.$$

□

Exponentialverteilung

Definition 8.6.5. Eine Zufallsvariable X heißt *exponentialverteilt* mit Parameter $\lambda > 0$, wenn X absolut stetig ist und die Dichte von X durch die folgende Formel gegeben ist:

$$f_X(y) = \begin{cases} \lambda e^{-\lambda y}, & y > 0, \\ 0, & y < 0. \end{cases}$$

Bezeichnung. $X \sim \mathrm{Exp}(\lambda)$.

Bemerkung 8.6.6. Das Integral der Dichte ist gleich 1, denn $\int_0^\infty \lambda e^{-\lambda y} dy = 1$. Eine Exponentialverteilte Zufallsvariaböe kann nur positive Werte annehmen: $\mathbb{P}[X > 0] = 1$.

Bemerkung 8.6.7. Die Verteilungsfunktion einer Zufallsvariable $X \sim \mathrm{Exp}(\lambda)$ ist gegeben durch

$$F_X(t) = \int_{-\infty}^t f_X(y)dy = \begin{cases} 0, & t \le 0 \\ 1 - e^{-\lambda t}, & t \ge 0. \end{cases}$$

Beweis. Für $t < 0$ gilt $F_X(t) = \int_{-\infty}^t f_X(t) = \int_{-\infty}^t 0 = 0$, denn die Dichte f_X verschwindet auf der negativen Halbachse. Für $t \ge 0$ gilt

$$F_X(t) = \int_{-\infty}^t f_X(y)dy = \int_0^t \lambda e^{-\lambda y} dy = -e^{-\lambda y}\Big|_0^t = 1 - e^{-\lambda t}.$$

Bemerkung 8.6.8. Alternativ kann man die Exponentialverteilung durch die folgende Formel definieren:

$$\mathbb{P}[X > t] = e^{-\lambda t}, \qquad t \ge 0.$$

Satz 8.6.9. *Für eine Exp(λ)-verteilte Zufallsvariable X gilt*

$$\mathbb{E}X = \frac{1}{\lambda}.$$

Beweis.

$$\mathbb{E}X = \int_0^\infty y f_X(y) dy = \int_0^\infty \lambda y e^{-\lambda y} dy = \int_0^\infty y \left(-e^{-\lambda y}\right)' dy.$$

Nun benutzen wir die partielle Integration:

$$\mathbb{E}X = \int_0^\infty y \left(-e^{-\lambda y}\right)' dy = -ye^{-\lambda y}\Big|_0^\infty + \int_0^\infty e^{-\lambda y} dy = \frac{1}{\lambda}.$$

□

Satz 8.6.10 (Vergessenseigenschaft der Exponentialverteilung). *Sei X eine Zufallsvariable. Die folgenden zwei Eigenschaften sind äquivalent:*

(1) *X ist exponentialverteilt mit einem Parameter $\lambda > 0$.*
(2) *Für alle $s, t > 0$ gilt:*

$$\mathbb{P}[X > t + s | x > t] = \mathbb{P}[X > s].$$

Beweis von (1) ⇒ (2). Sei $X \sim \text{Exp}(\lambda)$. Dann gilt für alle $t, s > 0$:

$$\mathbb{P}[X > t + s | X > t] = \frac{\mathbb{P}[X > t + s | X > t]}{\mathbb{P}[X > t]} = \frac{\mathbb{P}[X > t + s]}{\mathbb{P}[X > t]} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} = \mathbb{P}[X > s].$$

□

Für den Beweis der Rückrichtung benötigen wir ein Lemma.

Lemma 8.6.11 (Cauchy-Funktionalgleichung). *Sei $g : [0, \infty) \to \mathbb{R}$ eine monotone Funktion mit $g(t + s) = g(t) + g(s)$ für alle $t, s \geq 0$. Dann ist g linear, d.h. es gibt ein $\lambda \in \mathbb{R}$ mit $g(t) = \lambda t$ für alle $t \geq 0$.*

Beweis. Sei $a \in \mathbb{R}$. Dann gilt

$$g(2a) = g(a) + g(a) = 2g(a).$$

Analog ergibt sich

$$g(3a) = g(2a) + g(a) = 2g(a) + g(a) = 3g(a).$$

Induktiv erhalten wir für jedes $n \in \mathbb{N}$

$$g(na) = ng(a).$$

Nun sei $a = 1/n$. Es ergibt sich $g(1) = ng(\frac{1}{n})$ und somit

$$g\left(\frac{1}{n}\right) = \frac{g(1)}{n} \stackrel{def}{=} \frac{\lambda}{n},$$

wobei wir $\lambda := g(1)$ gesetzt haben. Sei $m \in \mathbb{N}$, dann gilt

$$g\left(\frac{m}{n}\right) = g\left(m \cdot \frac{1}{n}\right) = m \cdot g\left(\frac{1}{n}\right) = \frac{m}{n} \cdot \lambda.$$

Wir haben somit gezeigt, dass $g(r) = \lambda r$ für alle $r \in \mathbb{Q}$, $r \geq 0$ gilt. Nun müssen wir zeigen, dass das auch für irrationale Werte gilt. Sei $t \in \mathbb{R}\backslash\mathbb{Q}$, $t > 0$ beliebig. Es existieren Folgen $\{r_i\} \subset \mathbb{Q}$ und $\{s_i\} \subset \mathbb{Q}$ mit den folgenden Eigenschaften:

(1) r_i ist monoton fallend mit $r_i \downarrow t$ für $i \to \infty$.
(2) s_i ist monoton steigend mit $s_i \uparrow t$ für $i \to \infty$.

Die Funktion g ist nach Voraussetzung monoton. Sei g zum Beispiel monoton fallend. (Für g monoton steigend ist der Beweis analog). Dann gilt

(1) $g(r_i)$ ist monoton steigend mit $g(t) \geq g(r_i)$ für alle $i \in \mathbb{N}$.
(2) $g(s_i)$ ist monoton fallend mit $g(t) \leq g(s_i)$ für alle $i \in \mathbb{N}$.

Da die Zahlen r_i und s_i rational sind, gilt $g(r_i) = \lambda r_i$ und $g(s_i) = \lambda s_i$. Somit erhalten wir $\lambda r_i \leq t \leq \lambda s_i$. Nun lassen wir $i \to \infty$ und benutzen den Sandwich-Satz: $g(t) = \lambda t$. $\square$

Beweis von (2) $\Rightarrow$ (1) in Satz 8.6.10. Sei F_X die Verteilungsfunktion von X. Betrachte die Funktion

$$g(t) \stackrel{def}{=} \log(1 - F_X(t)), \quad t \geq 0.$$

Diese Funktion ist monoton fallend, da die Verteilungsfunktion F_X monoton steigend ist. Es gilt

$$\mathbb{P}[X > t] = 1 - F_X(t) = e^{g(t)}, \quad t \geq 0.$$

Mit der Vergessenseigenschaft erhalten wir

$$\frac{e^{g(t+s)}}{e^{g(t)}} = \mathbb{P}[X > t + s | X > t] = \mathbb{P}[X > s] = e^{g(s)}.$$

Somit erfüllt g die Cauchy-Funktionalgleichung $g(t+s) = g(t)+g(s)$ und ist monoton fallend. Es gibt nach Lemma 8.6.11 ein λ mit $g(t) = -\lambda t$ für alle $t \geq 0$. Dabei muss λ positiv sein, da g monoton fallend ist. Es folgt, dass

$$F_X(t) = 1 - e^{-\lambda t}, \quad t \geq 0$$

Damit ist gezeigt, dass $X \sim \text{Exp}(\lambda)$. $\square$

Beispiel 8.6.12 (Radioaktiver Zerfall). Die Lebensdauer eines Atoms bis zum radioaktiven Zerfall sei eine Zufallsvariable $X \sim \text{Exp}(\lambda)$. Die Halbwertszeit m lässt sich bestimmen durch

$$\mathbb{P}[X > m] = \frac{1}{2}.$$

Löst man diese Gleichung auf, so erhält man

$$e^{-\lambda m} = \frac{1}{2} \Rightarrow \lambda m = \log(2) \Rightarrow m = \frac{\log(2)}{\lambda}.$$

Warum benutzt man Exponentialverteilung als die Verteilung der Lebensdauer eines Atoms? Weil man davon ausgeht, dass für die Lebensdauer die Vergessenseigenschaft gilt. Und daraus folgt, dass die Lebensdauer exponentialverteilt sein muss.

Bemerkung 8.6.13. Sei X eine beliebige Zufallsvariable mit Dichte f_X. Dann ist die Zerfallrate (oder die Ausfallrate) definiert durch

$$r_X(t) = \frac{f_X(t)}{1 - F_X(t)}.$$

Bei einem Atom geht man davon aus, dass die Zerfallrate konstant (d.h. unabhängig von t) ist. Die Exponentialverteilung erfüllt diese Eigenschaft: Für $X \sim \mathrm{Exp}(\lambda)$ gilt

$$r_X(t) = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda.$$

Normalverteilung (oder Gauß-Verteilung)

DEFINITION 8.6.14. Eine Zufallsvariable X heißt *standardnormalverteilt*, wenn X absolut stetig ist und die Dichte durch die folgende Formel gegeben ist:

$$f_X(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}, \quad y \in \mathbb{R}.$$

BEZEICHNUNG. $X \sim \mathrm{N}(0,1)$.

Die Funktion f_X ist trivialerweise nicht negativ. Aber warum ist das Integral dieser Funktion gleich 1? Das beantwortet der folgende Satz.

SATZ 8.6.15. *Es gilt* $I := \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy = \sqrt{2\pi}$.

BEWEIS. Die Stammfunktion von $e^{-y^2/2}$ lässt sich nicht mit Hilfe der 4 Grungrechenarten und Potenzen als eine endliche Kombination von $y, e^y, \log y, \sin y, \cos y$, usw. darstellen. Man muss also einen Trick anwenden. Um das Integral I zu berechnen, betrachten wir I^2:

$$I^2 = \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \cdot \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy = \int_{\mathbb{R}^2} e^{-\left(\frac{x^2}{2} + \frac{y^2}{2}\right)} dx dy.$$

Und nun gehen wir zu Polarkoordinaten über:

$$I^2 = \int_0^{2\pi} \int_0^{\infty} e^{-\frac{r^2}{2}} r dr d\varphi = 2\pi \int_0^{\infty} r e^{-\frac{r^2}{2}} dr = 2\pi \left(-e^{-\frac{r^2}{2}} \Big|_0^{\infty} \right) = 2\pi.$$

Da $I \geq 0$, können wir die Wurzel ziehen: $I = \sqrt{2\pi}$. □

BEMERKUNG 8.6.16. Die Verteilungsfunktion der Standardnormalverteilung ist

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{y^2}{2}} dy.$$

Ihre Werte können numerisch berechnet werden.

BEMERKUNG 8.6.17. Aus $X \sim N(0,1)$ folgt für den Erwartungswert $\mathbb{E}X = 0$.

BEWEIS. Erstens, ist X integrierbar, denn $\int_{\mathbb{R}} |y| e^{-\frac{y^2}{2}} dy < \infty$. Für den Erwartungswert erhalten wir dann

$$\mathbb{E}X = \int_{\mathbb{R}} y e^{-\frac{y^2}{2}} dy = 0,$$

denn es wird eine ungerade Funktion integriert. □

Definition 8.6.18. Eine Zufallsvariable X heißt *normalverteilt* mit Parameter $\mu \in \mathbb{R}$ und $\sigma^2 > 0$, wenn X absolut stetig ist und die Dichte durch die folgende Formel gegeben ist:

$$f_X(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}}, \quad y \in \mathbb{R}.$$

Bezeichnung. $X \sim \mathrm{N}(\mu, \sigma^2)$.

Bemerkung 8.6.19. Ist $Y \sim \mathrm{N}(0,1)$ standardnormalverteilt, so ist $\mu + \sigma Y$ normalverteilt mir Parametern μ und σ^2 (Übung). Daraus folgt, dass der Erwartungswert einer $\mathrm{N}(\mu, \sigma^2)$-verteilten Zufallsvariable gleich μ ist. Später werden wir zeigen, dass der zweite Parameter σ^2 mit der Varianz der Zufallsvariable übereinstimmt.

Cauchy-Verteilung

Definition 8.6.20. Eine Zufallsvariable X heißt *Cauchy-verteilt*, wenn X absolut stetig ist und die Dichte durch die folgende Formel gegeben ist:

$$f_X(y) = \frac{1}{\pi} \frac{1}{1+y^2}, \quad y \in \mathbb{R}.$$

Bezeichnung. $X \sim$ Cauchy.

Bemerkung 8.6.21. Die Verteilungsfunktion der Cauchy-Verteilung ist gegeben durch

$$F_X(t) = \int_{-\infty}^t f_X(y)dy = \frac{1}{\pi} \int_{-\infty}^t \frac{1}{1+y^2} dy = \frac{1}{\pi} \arctan y \Big|_{-\infty}^t = \frac{1}{2} + \frac{1}{\pi} \arctan t,$$

da $\arctan(-\infty) = -\frac{\pi}{2}$. Insbesondere gilt

$$\int_{-\infty}^{\infty} f_X(y)dy = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{1}{1+y^2} dy = \frac{1}{2} + \frac{1}{\pi} \arctan(+\infty) = 1,$$

so dass f_X tatsächlich eine Dichte ist.

Beispiel 8.6.22. Sei $\varphi \sim \mathrm{U}[-\pi/2, \pi/2]$ gleichverteilt auf dem Intervall $[-\pi/2, \pi/2]$ und sei $X = \tan(\varphi)$. Nun behaupten wir, dass $X \sim$ Cauchy.

Beweis. Wir berechnen die Verteilungsfunktion von X:

$$\mathbb{P}[X \le t] = \mathbb{P}[\varphi \le \arctan(t)] = \frac{\arctan t - \left(-\frac{\pi}{2}\right)}{\pi} = \frac{1}{2} + \frac{1}{\pi} \arctan t.$$

Bemerkung 8.6.23. Sei $X \sim$ Cauchy. Die Dichte von X ist eine gerade Funktion und man könnte meinen, dass der Erwartungswert von X aus diesem Grund gleich 0 ist. Dies ist jedoch nicht der Fall. Wir behaupten, dass X nicht integrierbar ist. Wir berechnen $\mathbb{E}X^+$:

$$\mathbb{E}X^+ = \int_0^\infty y f_X(y)dy = \frac{2}{\pi} \int_0^\infty \frac{y}{1+y^2} dy = \frac{1}{\pi} \log(1+y^2)\Big|_0^\infty = +\infty.$$

Analog ist $\mathbb{E}X^- = +\infty$. Wir haben eine Unbestimmtheit $\mathbb{E}X = (+\infty) - (+\infty)$ und die Zufallsvariable X ist somit nicht integrierbar. Dabei ist der Erwartungswert von X weder $+\infty$ noch $-\infty$, sondern er ist überhaupt nicht definiert.

8.7. Singuläre Verteilungen

Die Verteilung einer diskreten Zufallsvariable besteht ausschließlich aus Atomen. Sei nun X eine Zufallsvariable, die keine Atome hat, d.h. es gelte $\mathbb{P}[X = y] = 0$ für alle $y \in \mathbb{R}$. Eine solche Zufallsvariable ist nicht diskret. Ist sie dann absolut stetig? Es stellt sich heraus, dass die Antwort im Allgemeinen "nein" ist. Es gibt nämlich eine dritte Klasse von Verteilungen, die sogenannten singulären Verteilungen. Außerdem kann man Mischungen aus Verteilungen von diesen drei Typen (diskret, absolut stetig, singulär) bilden. Nach dem Zerlegungssatz von Lebesgue kann jede Verteilung als eine solche Mischung dargestellt werden.

Definition 8.7.1. Eine Zufallsvariable X heißt *singulär*, wenn die folgenden zwei Bedingungen gelten:

(1) X hat keine Atome, d.h. es gilt $\mathbb{P}[X = y] = 0$ für alle $y \in \mathbb{R}$.
(2) Es gibt eine Borel-Menge $N \subset \mathbb{R}$ mit Lebesgue-Maß $\lambda(N) = 0$, so dass
$$\mathbb{P}[X \in N] = 1.$$

Eine singuläre Zufallsvariable nimmt also ausschließlich Werte in einer Nullmenge N an.

Beispiel 8.7.2 (Cantor-Menge). Jede Zahl $x \in [0, 1]$ kann man in einer Darstellung zur Basis 3 (d.h. im Ternärsystem) schreiben:

$$x = [0.\varepsilon_1\varepsilon_2 \ldots]_3 = \sum_{k=1}^{\infty} \frac{\varepsilon_k}{3^k}.$$

Dabei sind $\varepsilon_k \in \{0, 1, 2\}$ die "Ziffern" von x. Die Cantor-Menge C ist die Menge aller Punkte $x \in [0, 1]$, dessen Ternärdarstellung ausschließlich aus den Ziffern 0 und 2 besteht (so dass die Ziffer 1 also nicht vorkommt):

$$C = \left\{ x \in [0, 1] : x = \sum_{k=1}^{\infty} \frac{\varepsilon_k}{3^k}, \, \varepsilon_k \in \{0, 2\} \right\}.$$

Bei manchen Zahlen (nämlich, bei den Zahlen der Form $\frac{k}{3^m}$) ist die Ternärdarstellung nicht eindeutig, z.B.

$$1 = [1.000\ldots]_3 = [0.222\ldots]_3 \text{ und } \frac{1}{3} = [0.1000\ldots]_3 = [0.0222\ldots]_3.$$

Solche Zahlen werden wir dann in die Cantor-Menge aufnehmen, wenn mindestens eine Darstellung nur aus den Ziffern 0 und 2 besteht. Man kann zeigen, dass die Cantor-Menge abgeschlossen ist. Wir werden nun zeigen, dass das Lebesgue-Maß von C gleich 0 ist. Betrachte die Menge C_n, die aus allen Zahlen in $[0, 1]$ besteht, die keine einzige 1 unter den ersten n Ziffern in ihrer Ternärdarstellung haben. Für die ersten n Ziffern gibt es also 2^n Möglichkeiten. Die Menge aller Zahlen in $[0, 1]$, bei denen die ersten n Ziffern festgelegt sind, ist ein Intervall der Länge $1/3^n$. Somit ist das Lebesgue-Maß von C_n gleich

$$\lambda(C_n) = \left(\frac{2}{3}\right)^n.$$

Es ist aber klar, dass $C \subset C_n$, und zwar für jedes n. Somit gilt

$$\lambda(C) \le \left(\frac{2}{3}\right)^n \text{ für alle } n \in \mathbb{N}.$$

Das kann aber nur dann gelten, wenn $\lambda(C) = 0$. Das Lebesgue-Maß der Cantor-Menge ist also 0. Dabei ist die Cantor-Menge überabzählbar!

Beispiel 8.7.3 (Cantor-Verteilung). Nun konstruiren wir eine singuläre Verteilung. Es handelt sich dabei um eine Art Gleichverteilung auf der Cantor-Menge. Seien dazu $\varepsilon_1, \varepsilon_2, \ldots$ unabhängige Zufallsvariablen mit

$$\mathbb{P}[\varepsilon_k = 0] = \mathbb{P}[\varepsilon_k = 2] = 1/2.$$

Diese interpretieren wir als Ziffern in einer Ternärdarstellung einer zufälligen Zahl. D.h., wir betrachten die Zufallsvariable

$$X = [0.\varepsilon_1\varepsilon_2\ldots]_3 = \sum_{k=1}^{\infty} \frac{\varepsilon_k}{3^k}.$$

Nach Definition gilt $\mathbb{P}[X \in C] = 1$. Dabei ist die Cantor-Menge C eine Nullmenge. Wir zeigen noch, dass X keine Atome hat. Sei $y = [0.\eta_1\eta_2\ldots]_3 \in [0,1]$. Dann gilt für jedes $n \in \mathbb{N}$

$$\mathbb{P}[X = y] \leq \mathbb{P}[\varepsilon_1 = \eta_1, \ldots, \varepsilon_n = \eta_n] \leq \frac{1}{3^n}.$$

Daraus folgt, dass $\mathbb{P}[X = y] = 0$. Somit ist X singulär.

Bemerkung 8.7.4. Die Verteilungsfunktion der Cantor-Verteilung hat viele interessante Eigenschaften. Sie ist stetig, da die Cantor-Verteilung keine Atome hat. Außerdem ist sie konstant auf jedem Intervall, das komplett außerhalb der Cantor-Menge liegt. Somit ist die Ableitung der Verteilungsfunktion gleich 0 außerhalb der Cantor-Menge, also fast überall. Dennoch ist die Verteilungsfunktion nicht konstant, denn sie ist gleich 0 für $t = 0$ und gleich 1 für $t = 1$. Für diese Verteilungsfunktion gilt der Satz von Newton–Leibniz nicht, denn

$$1 = F_X(1) - F_X(0) \neq \int_0^1 F_X'(t)dt = 0.$$

Dabei gilt die letzte Gleichheit, weil F_X' fast überall gleich 0 ist.

Nun zeigen wir, dass die drei Klassen von Verteilungen (diskret, absolut stetig und singulär) sich nicht überschneiden. Diskrete und absolut stetige Verteilungen überschneiden sich nicht, denn die einen haben eine unstetige und die anderen eine stetige Verteilungsfunktion. Sei nun X eine singuläre Zufallsvariable. Da X keine Atome hat, kann X nicht diskret sein. Es bleibt zu zeigen, dass X nicht absolut stetig sein kann. Das wird im folgenden Satz gemacht.

Satz 8.7.5. *Sei X eine Zufallsvariable und $N \subset \mathbb{R}$ eine Borel-Menge mit $\lambda(N) = 0$ und $\mathbb{P}[X \in N] = 1$. Dann ist X nicht absolut stetig.*

Beweis. Durch Widerspruch. Sei X absolut stetig und f_X die Dichte von X. Dann gilt

$$0 = \mathbb{P}[X \notin N] = \int_{\mathbb{R}\backslash N} f_X(y)dy = \int_{\mathbb{R}} f_X(y) \cdot \mathbb{1}_{\mathbb{R}\backslash N}(y)dy.$$

Die Dichte f_X ist nichtnegativ. Das Lebesgue-Integral einer nichtnegativen Funktion kann nur dann 0 sein, wenn die Funktion fast überall 0 ist. Das heißt, es muss eine Nullmenge M geben mit

$$f_X(y)\mathbb{1}_{\mathbb{R}\backslash N}(y) = 0 \text{ für alle } y \notin M.$$

Dann gilt aber $f_X(y) = 0$ für alle $y \notin N \cup M$. Dabei ist $N \cup M$ eine Nullmenge, also ist $f_X(y) = 0$ fast überall. Daraus folgt aber, dass $\int_{\mathbb{R}} f_X(y)dy = 0$. Somit ist f_X keine Dichte, denn für eine Dichte müsste dieses Integral 1 sein. Widerspruch. □

8.8. Zerlegungssatz von Lebesgue

Der Zerlegungssatz von Lebesgue besagt, dass man jede Verteilung als eine Mischung aus einer diskreten, einer absolut stetigen und einer singulären Verteilung darstellen kann.

Satz 8.8.1 (Lebesgue). *Sei F eine Verteilungsfunktion. Dann existieren drei Verteilungsfunktionen F_1, F_2, F_3 und drei Zahlen $p_1, p_2, p_3 \leq 0$ mit $p_1 + p_2 + p_3 = 1$, sodass*

$$F(t) = p_1 F_1(t) + p_2 F_2(t) + p_3 F_3(t) \text{ für alle } t \in \mathbb{R}$$

und dabei F_1 diskret ist, F_2 absolut stetig und F_3 singulär.

Ohne Beweis.

Beispiel 8.8.2 (Gemischte Verteilung). Man betrachte eine Bahnschranke, die für 10 Minuten offen steht, dann für 10 Minuten geschlossen, dann wieder für 10 Minuten offen, usw. Ein Fußgänger kommt zu einem zufälligen Zeitpunkt an dieser Bahnschranke an. Es sei X die Zeit, die er warten muss, bis die Bahnschranke offen ist. Wie sieht dann die Verteilungsfunktion F_X aus?

Lösung. Eigentlich kann der Fußgänger zu einem beliebigen Zeitpunkt ankommen, also könnte man versuchen, die ganze Gerade $\mathbb{R}$ mit dem Lebesgue-Maß als Wahrscheinlichkeitsraum zu nehmen. Allerdings ist das Lebesgue-Maß kein Wahrscheinlichkeitsmaß: $\lambda(\mathbb{R}) = +\infty$. Deshalb werden wir benutzen, dass die Bahnschranke periodisch funktioniert, und nur eine Periode als Grundmenge betrachten. Die Grundmenge sei also $\Omega = (0, 20)$, versehen mit der σ-Algebra der Borel-Teilmengen. Dabei sei die Bahnschranke im Zeitintervall $(0, 10)$ offen und im Zeitintervall $(10, 20)$ geschlossen. Die Ankunft des Fußgängers kann man sich nun so vorstellen: es wird im Intervall $(0, 20)$ ein zufälliger, gleichverteilter Punkt ausgewählt. Die Wahrscheinlichkeit einer Borel-Menge $A \subset (0, 20)$ ist also

$$\mathbb{P}[A] = \frac{\lambda(A)}{20},$$

wobei λ das Lebesgue-Maß sei. Kommt nun der Fußgänger zu einem Zeitpunkt $\omega \in (0, 20)$ an, so sieht seine Wartezeit wie folgt aus:

$$X(\omega) = \begin{cases} 0, & \omega \in (0, 10), \text{ denn da ist die Bahnschranke offen,} \\ 20 - \omega, & \omega \in (10, 20), \text{ denn die Bahnschranke öffnet zum Zeitpunkt 20.} \end{cases}$$

Den Wert an der Stelle 10 (der entweder als 0 oder als 10 festgelegt werden kann) kann man ignorieren, den ein einzelner Punkt ist lediglich eine Nullmenge. Es gilt also

$$\mathbb{P}[X = 0] = \frac{10}{20} = \frac{1}{2}.$$

Nun bestimmen wir die Verteilungsfunktion von X:

$$F_X(t) = \mathbb{P}[X \leq t] = \begin{cases} 0, & t < 0, \\ \frac{1}{2}, & t = 0, \\ \frac{10+t}{20}, & t \in [0,10], \\ 1, & t \geq 10. \end{cases}$$

Die Verteilungsfunktion ist nicht absolut stetig und nicht singulär (denn es gibt ein Atom an der Stelle 0). Sie ist auch nicht diskret, denn sie ist keine reine Sprungfunktion. Hier haben wir es mit einer gemischten Verteilung zu tun. Die Verteilung von X ist eine Mischung aus einer diskreten Verteilung (die nur den Wert 0 annimmt, ein Atom mit der Wahrscheinlichkeit 1 an der Stelle 0) und einer absolut stetigen Verteilung (Gleichverteilung auf $(0, 10)$). Bezeichnet man mit F_1 und F_2 die entsprechenden Verteilungsfunktionen, so hat man die Darstellung

$$F_X = \frac{1}{2}F_1 + \frac{1}{2}F_2.$$

8.9. Verteilungsfunktion eines Zufallsvektors

Die Familie der Borel-Teilmengen von $\mathbb{R}^d$ wird mit $\mathcal{B}^d$ bezeichnet.

Definition 8.9.1. Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X = (X_1, \ldots, X_d) : \Omega \to \mathbb{R}^d$ ein Zufallsvektor.

(1) Die *Verteilung* von X ist die Funktion

$$P_X : \mathcal{B}^d \to [0,1] \text{ mit } P_X(A) = \mathbb{P}[X \in A], \quad A \in \mathcal{B}^d.$$

P_X ist ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}^d, \mathcal{B}^d)$.

(2) Die *Verteilungsfunktion* von X ist die Funktion

$$F_X : \mathbb{R}^d \to [0,1] \text{ mit } F_X(t) = F_X(t_1, \ldots, t_d) = \mathbb{P}[X_1 \leq t_1, \ldots, X_d \leq t_d],$$

wobei $t = (t_1, \ldots, t_d) \in \mathbb{R}^d$.

Satz 8.9.2. *Sei $X : \Omega \to \mathbb{R}^d$ ein Zufallsvektor. Dann hat die Verteilungsfunktion von X die folgenden drei Eigenschaften.*

(1) *Grenzwerte:*

(a) *Die Verteilungsfunktion konvergiert gegen 0, wenn* **mindestens eine** *der Koordinaten gegen $-\infty$ konvergiert. Das heißt, für jedes $i = 1, \ldots, d$ gilt*

$$\lim_{t_i \to -\infty} F_X(t_1, \ldots, t_d) = 0.$$

(b) *Die Verteilungsfunktion konvergiert gegen 1, wenn* **alle** *Koordinaten* **gleichzeitig** *gegen $+\infty$ konvergieren. Das heißt,*

$$\lim_{t_1 \to +\infty, \ldots, t_d \to +\infty} F_X(t_1, \ldots, t_d) = 1.$$

(2) *Monotonie: Für alle $t_1 \leq t_1', \ldots, t_d \leq t_d'$ gilt*

$$F_X(t_1, \ldots, t_d) \leq F_X(t_1', \ldots, t_d').$$

(3) *Rechtsstetigkeit: Für alle $(t_1, \ldots, t_d) \in \mathbb{R}^d$ gilt*

$$F_X(t_1, \ldots, t_d) = \lim_{s_1 \downarrow t_1, \ldots, s_d \downarrow t_d} F_X(s_1, \ldots, s_d).$$

Beweis. Der Beweis geht analog zum eindimensionalen Fall. □

8.10. Diskrete und absolut stetige Zufallsvektoren

Wie im Fall von Zufallsvariablen lassen sich diskrete, absolut stetige und singuläre Vektoren definieren.

Definition 8.10.1. Ein Zufallsvektor $X : \Omega \to \mathbb{R}^d$ heißt *diskret*, wenn X höchstens abzählbar viele Werte annimmt. Die *Zähldichte* von X ist die Funktion

$$p_X(y) = \mathbb{P}[X = y], \quad y \in \mathbb{R}^d.$$

Definition 8.10.2. Ein Zufallsvektor $X : \Omega \to \mathbb{R}^d$ heißt *absolut stetig*, wenn eine Borel-Funktion $f_X : \mathbb{R}^d \to [0, \infty)$ existiert mit

$$F_X(t_1, \ldots, t_d) = \int_{-\infty}^{t_1} \cdots \int_{-\infty}^{t_d} f_X(y_1, \ldots, y_d) dy_1 \ldots dy_d$$

für alle $t_1, \ldots, t_d \in \mathbb{R}$. Die Funktion f_X heißt die *Dichte* von X.

Bemerkung 8.10.3. Genauso wie im eindimensionalen Fall zeigt man, dass

$$\mathbb{P}[X \in B] = \int_B f_X(y_1, \ldots, y_d) dy_1 \ldots dy_d \text{ für alle } B \in \mathcal{B}^d.$$

Beispiel 8.10.4. Ein Zufallsvektor X heißt *gleichverteilt* auf einer Borel-Menge $A \subset \mathbb{R}^d$ mit $A \in \mathcal{B}^d$, $0 < \lambda(A) < \infty$, wenn X absolut stetig ist mit Dichte

$$f_X(y) = \frac{\mathbb{1}_A(y)}{\lambda(A)} = \begin{cases} \frac{1}{\lambda(A)}, & y \in A, \\ 0, & y \notin A. \end{cases}$$

8.11. Randverteilungen eines Zufallsvektors und Unabhängigkeit

Sei $X = (X_1, \ldots, X_d)$ ein Zufallsvektor. Seine Komponenten $X_1, \ldots, X_d$ sind dann Zufallsvariablen. Die Verteilungen von $X_1, \ldots, X_d$ bezeichnet man als *Randverteilungen* von X.

Beispiel 8.11.1. Sei $X = (X_1, \ldots, X_d)$ ein d-dimensionaler diskreter Zufallsvektor mit Zähldichte

$$p_X(t) = p_{X_1,\ldots,X_d}(t_1, \ldots, t_d) = \mathbb{P}[X_1 = t_1, \ldots, X_d = t_d], \quad t = (t_1, \ldots, t_d) \in \mathbb{R}^d.$$

Dann kan man die Zähldichten der einzelnen Komponenten $X_1, \ldots, X_d$ wie folgt berechnen:

$$p_{X_1}(t) = \mathbb{P}[X_1 = t_1] = \sum_{t_2, t_3, \ldots, t_d \in \mathbb{R}} p_{X_1,\ldots,X_d}(t_1, \ldots, t_d),$$
$$p_{X_2}(t) = \mathbb{P}[X_2 = t_2] = \sum_{t_1, t_3, \ldots, t_d \in \mathbb{R}} p_{X_1,\ldots,X_d}(t_1, \ldots, t_d),$$

und so weiter.

BEISPIEL 8.11.2. Wir würfeln 2-mal mit einem fairen Würfel. Dabei seien Z_1 und Z_2 die Augenzahlen. Weiter definiere

$$X_1 = \max\{Z_1, Z_2\}, \quad X_2 = \min\{Z_1, Z_2\}.$$

Dann ist $X = (X_1, X_2)$ ein diskreter Zufallsvektor. Wir bestimmen die Randverteilung (also in diesem Fall die Zähldichte) von X_1.

LÖSUNG. Die Zähldichte von (X_1, X_2) stellt sich wie folgt dar:

$$p_X(i,i) = \mathbb{P}[X = (i,i)] = \mathbb{P}[\{(i,i)\}] = \frac{1}{36}, \quad i = 1, \ldots, 6,$$

$$p_X(i,j) = \mathbb{P}[X = (i,j)] = \mathbb{P}[\{(i,j),(j,i)\}] = \frac{2}{36} = \frac{1}{18}, \quad 1 \le j < i \le 6.$$

Somit kann man die Zähldichte von X_1 bestimmen:

$$p_{X_1}(2) = p_X(2,1) + p_X(2,2) = \frac{1}{36} + \frac{1}{18} = \frac{1}{12},$$

$$p_{X_1}(3) = p_X(3,1) + p_X(3,2) + p_X(3,3) = \frac{1}{36} + \frac{1}{36} + \frac{1}{18} = \frac{1}{9},$$

und so weiter.

Nun zeigen wir, wie man die Randverteilungen eines absolut stetigen Zufallsvektors berechnet.

SATZ 8.11.3. *Sei $X = (X_1, \ldots, X_d)$ ein absolut stetiger Zufallsvektor mit Dichte f_X. Dann gilt: Für alle $i = 1, \ldots, d$ ist X_i absolut stetig mit Dichte*

$$f_{X_i}(t_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_X(t_1, \ldots, t_d) dt_1 \ldots dt_{i-1} dt_{i+1} \ldots dt_d \tag{8.11.1}$$

für fast alle $t_i \in \mathbb{R}$.

BEWEIS. Wir beweisen den Satz für $i = 1$. Der allgemeine Fall ist analog. Betrachte die Menge $B = \{(X_1, \ldots, X_d) : X_1 \le s\}$. Die Verteilungsfunktion von X_1 ist

$$\begin{aligned} F_{X_1}(s) &= \mathbb{P}[X_1 \le s] \\ &= \int_B f_X(t_1, \ldots, t_d) dy_1 \ldots dy_d \\ &= \int_{-\infty}^{s} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_X(y_1, \ldots, y_d) dy_1 \ldots dy_d \\ &= \int_{-\infty}^{s} g(y_1) dy_1, \end{aligned}$$

wobei g die Funktion auf der rechten Seite von (8.11.1) ist:

$$g(y_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_X(y_1, \ldots, y_d) dy_2 \ldots dy_d.$$

Also ist g die Dichte von X_1. $\square$

Beispiel 8.11.4. Die Zufallsvariable X sei gleichverteilt auf dem Einheitskreis $D = \{(y_1, y_2) : y_1^2 + y_2^2 \leq 1\}$. Die Dichte von X ist

$$f_{X_1,X_2}(y_1, y_2) = \begin{cases} \frac{1}{\pi}, & (y_1, y_2) \in D, \\ 0, & (y_1, y_2) \notin D. \end{cases}$$

Wie berechnet man die Randdichten f_{X_1}, f_{X_2}?

Lösung. Die Randdichte von X_1 ist

$$f_{X_1}(t_1) = \int_{-\infty}^{\infty} f_{X_1,X_2}(t_1, t_2) dt_2 = \begin{cases} 0, & |t_1| > 1, \\ \int_{-\sqrt{1-t_1^2}}^{+\sqrt{1-t_1^2}} \frac{1}{\pi} dt_2 = \frac{2}{\pi}\sqrt{1 - t_1^2}, & \text{sonst.} \end{cases}$$

Die Randdichte von X_2 bestimmt man genauso.

8.12. Unabhängigkeit

Für diskrete Zufallsvariablen haben wir die Unabhängigkeit wie folgt definiert: Zufallsvariablen $X_1, \ldots, X_d$ heißen unabhängig, wenn für alle $y_1, \ldots, y_d \in \mathbb{R}$ gilt

$$\mathbb{P}[X_1 = y_1, \ldots, X_d = y_d] = \mathbb{P}[X_1 = y_1] \cdot \ldots \mathbb{P}[X_d = y_d].$$

Diese Definition macht im Allgemeinen Fall allerdings keinen Sinn. Sind z.B. $X_1, \ldots, X_d$ absolut stetig, so sind beide Seiten der Gleichung gleich 0. Somit wären beliebige absolut stetige Zufallsvariablen unabhängig, was dem intuitiven Verständnis der Unabhängigkeit nicht entspricht. Für allgemeine Zufallsvariablen benutzt man deshalb eine andere Definition.

Definition 8.12.1. Seien $X_1, \ldots, X_d$ Zufallsvariablen. Diese Zufallsvariablen heißen *unabhängig*, wenn für alle Borel-Mengen $B_1, \ldots, B_d \subset \mathbb{R}$ gilt:

$$\mathbb{P}[X_1 \in B_1, \ldots, X_d \in B_d] = \mathbb{P}[X_1 \in B_1] \cdot \ldots \cdot \mathbb{P}[X_d \in B_d].$$

Eine unendliche Famielie $X_1, X_2, \ldots$ von Zufallsvariablen heißt unabhängig, wenn für jedes $d \in \mathbb{N}$ die Zufallsvariablen $X_1, \ldots, X_d$ unabhängig sind.

Bemerkung 8.12.2. Wenn die Zufallsvariablen $X_1, \ldots, X_d$ unabhängig sind, so folgt daraus, dass

$$\begin{aligned} F_{X_1,\ldots X_d}(t_1, \ldots, t_d) &= \mathbb{P}[X_1 \leq t_1, \ldots, X_d \leq t_d] \\ &= \mathbb{P}[X_1 \leq t_1] \cdot \ldots \cdot \mathbb{P}[X_d \leq t_d] \\ &= F_{X_1}(t_1) \cdot \ldots \cdot F_{X_d}(t_d). \end{aligned}$$

Die gemeinsame Verteilungsfunktion ist also das Produkt der Verteilungsfunktionen der einzelnen Komponenten.

Eine ähnliche Produktformel gilt auch für Dichten, wie der nächste Satz zeigt.

Satz 8.12.3. *Seien $X_1, \ldots, X_d$ unabhängige absolut stetige Zufallsvariablen mit Dichten $f_{X_1}, \ldots, f_{X_d}$. Dann ist der Zufallsvektor $X = (X_1, \ldots, X_d)$ absolut stetig mit*

$$f_{X_1,\ldots,X_d}(y_1, \ldots, y_d) = f_{X_1}(y_1) \cdot \ldots \cdot f_{X_d}(y_d) \text{ für fast alle } (y_1, \ldots, y_d) \in \mathbb{R}^d.$$

Das heißt: die gemeinsame Dichte ist das Produkt der Randdichten.

Beweis. Wir können die Verteilungsfunktion des Vektors $(X_1, \dots, X_d)$ wie folgt darstellen:

$$\begin{aligned} F_{X_1,\dots,X_d}(t_1, \dots, t_d) &= F_{X_1}(t_1) \cdot \ldots \cdot F_{X_d}(t_d) \\ &= \int_{-\infty}^{t_1} f_{X_1}(y_1)dy_1 \cdot \ldots \cdot \int_{-\infty}^{t_d} f_{X_d}(y_d)dy_d \\ &= \int_{-\infty}^{t_1} \dots \int_{-\infty}^{t_d} f_{X_1}(y_1) \cdot \ldots \cdot f_{X_d}(y_d)dy_1 \dots dy_d. \end{aligned}$$

Es folgt aus der Definition der absoluten Stetigkeit, dass $f_{X_1}(y_1) \cdot \ldots \cdot f_{X_d}(y_d)$ die Dichte von $(X_1, \dots, X_d)$ ist. □

Beispiel 8.12.4. Seien die Zufallsvariablen X_1, X_2 unabhängig und gleichverteilt auf dem Intervall $[0, 1]$. Dann können wir die Dichte des Vektors (X_1, X_2) wie folgt bestimmen:

$$f_{X_1,X_2}(y_1, y_2) = f_{X_1}(y_1) \cdot f_{X_2}(y_2) = \mathbb{1}_{y_1\in[0,1]} \cdot \mathbb{1}_{y_2\in[0,1]} = \mathbb{1}_{(y_1,y_2)\in[0,1]^2}.$$

Somit ist der Vektor (X_1, X_2) gleichverteilt auf dem Quadrat $[0, 1]^2$.

8.13. Transformationsformel für die Dichte

Sei X eine absolut stetige Zufallsvariable und $\varphi : \mathbb{R} \to \mathbb{R}$ eine Funktion. Wie bestimmt man die Dichte von $\varphi(X)$?

Satz 8.13.1. *Seien*

(1) *X eine absolut stetige Zufallsvariable mit Dichte f_X.*
(2) *$I \subset \mathbb{R}$ ein offenes Intervall mit $\mathbb{P}[X \in I] = 1$, wobei I auch unendlich sein darf.*
(3) *$\varphi : I \to \mathbb{R}$ eine stetig differenzierbare Funktion mit $\varphi'(x) \neq 0$ für alle $x \in I$.*

Dann ist $\varphi(X)$ eine absolut stetige Zufallsvariable mit Dichte

$$f_{\varphi(X)}(y) = f_X(\varphi^{-1}(y)) \cdot \left|(\varphi^{-1})'(y)\right| \text{ für fast alle } y \in \varphi(I).$$

Dabei ist φ^{-1} die inverse Funktion von φ.

Beweis. Wegen $\varphi'(x) \neq 0$ für alle $x \in I$ ist φ entweder monoton steigend oder monoton fallend. Sei o.B.d.A. φ monoton steigend. Die Verteilungsfunktion von $\varphi(X)$, an der Stelle $t \in \varphi(I)$, ist

$$F_{\varphi(X)}(t) = \mathbb{P}[\varphi(X) \leq t] = \mathbb{P}[X \leq \varphi^{-1}(t)] = F_X\left(\varphi^{-1}(t)\right).$$

Die Dichte von $\varphi(X)$ erhalten wir, indem wir die Verteilungsfunktion ableiten:

$$f_{\varphi(X)}(t) = \frac{d}{dt}F_{\varphi(X)}(t) = \frac{d}{dt}F_X\left(\varphi^{-1}(t)\right) = F_X'\left(\varphi^{-1}(t)\right)(\varphi^{-1})'(t) = f_X\left(\varphi^{-1}(t)\right)(\varphi^{-1})'(t).$$

□

Bemerkung 8.13.2. Für monoton fallendes φ geht dies analog, jedoch mit $|(\varphi^{-1})'(t)|$ anstelle von $(\varphi^{-1})'(t)$.

Beispiel 8.13.3. Sei X eine absolut stetige Zufallsvariable mit Dichte f_X und $a \neq 0$ und $b \in \mathbb{R}$ zwei Konstanten. Bestimme die Dichte von $aX + b$.

LÖSUNG. Die Funktion $\varphi(X) = aX + b$ ist stetig differenzierbar auf $I = \mathbb{R}$. Die Ableitung ist $\varphi'(X) = a \neq 0$. Die Umkehrfunktion und ihre Ableitung sind

$$\varphi^{-1}(x) = \frac{x-b}{a}, \qquad (\varphi^{-1})'(x) = \frac{1}{a}.$$

Nun setzt man das in die Transformationformel ein und erhält für die Dichte von $aX + b$:

$$f_{aX+b}(t) = \frac{1}{|a|} \cdot f_X\left(\frac{t-b}{a}\right).$$

BEISPIEL 8.13.4. Sei $X \sim \mathrm{N}(0,1)$ standardnormalverteilt und $\mu \in \mathbb{R}, \sigma > 0$ zwei Konstanten. Die Dichte von $Y := \mu + \sigma X$ ist, mit der Formel aus dem obigen Beispiel,

$$f_Y(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}}.$$

Also ist $Y \sim \mathrm{N}(\mu, \sigma^2)$. Umgekehrt folgt aus $Y \sim \mathrm{N}(\mu, \sigma^2)$, dass die Zufallsvariable

$$X = \frac{Y - \mu}{\sigma}$$

standardnormalverteilt ist. Mit einer linearen Transformation kann man also eine beliebige Normalverteilung auf die Standardnormalverteilung zurückführen und umgekehrt.

8.14. Faltungsformeln

Seien X_1, X_2 unabhängige Zufallsvariablen. Wie bestimmt man dann die Verteilung der Zufallsvariablen

$$X_1 + X_2, X_1 - X_2, \quad X_1 \cdot X_2, \quad \frac{X_1}{X_2}?$$

Zuerst bestimmen wir die Verteilung von $X_1 + X_2$.

SATZ 8.14.1 (Faltungsformel für diskrete Zufallsvariablen). *Seien X_1, X_2 unabhängige Zufallsvariablen. Es seien beide Zufallsvariablen diskret. Dann gilt für die Zähldichte von $X_1 + X_2$ die Formel*

(8.14.1) $$p_{X_1+X_2}(z) = \sum_{y \in \mathrm{Im}(X_1)} p_{X_1}(y) \cdot p_{X_2}(z-y) \text{ für alle } z \in \mathbb{R}.$$

DEFINITION 8.14.2. Die Zähldichte $p_{X_1+X_2} = p_{X_1} * p_{X_2}$ heißt die *Faltung* der Zähldichten p_{X_1} und p_{X_2}.

BEMERKUNG 8.14.3. Zur Erinnerung: $p_{X_1}(y) = \mathbb{P}[X_1 = y]$ ist die Zähldichte von X_1 und $\mathrm{Im}(X_1) = \{y \in \mathbb{R} : \mathbb{P}[X_1 = y] \neq 0\}$ ist die Menge der Werte, die X_1 annehmen kann.

Beweis von Satz 8.14.1. Sei $z \in \mathbb{R}$. Es gilt dann

$$\begin{aligned} p_{X_1+X_2}(z) &= \mathbb{P}[X_1 + X_2 = z] \\ &= \sum_{y \in \mathrm{Im}(X_1)} \mathbb{P}[X_1 = y, X_2 = z - y] \\ &= \sum_{y \in \mathrm{Im}(X_1)} \mathbb{P}[X_1 = y] \cdot \mathbb{P}[X_2 = z - y] \\ &= \sum_{y \in \mathrm{Im}(X_1)} p_{X_1}(y) \cdot p_{X_2}(z - y). \end{aligned}$$

□

Satz 8.14.4. *Es seien $X_1 \sim Poi(\lambda_1)$ und $X_2 \sim Poi(\lambda_2)$ unabhängige, Poisson-verteilte Zufallsvariablen. Dabei sind $\lambda_1, \lambda_2 > 0$ Parameter. Dann ist auch die Summe $X_1 + X_2$ Poisson-verteilt:*

$$X_1 + X_2 \sim Poi(\lambda_1 + \lambda_2).$$

Das heißt: $Poi(\lambda_1) * Poi(\lambda_2) = Poi(\lambda_1 + \lambda_2)$.

Beweis. Für alle $n \in \mathbb{N}_0$ gilt

$$\begin{aligned} \mathbb{P}[X_1 + X_2 = n] &= \sum_{k=0}^{n} \mathbb{P}[X_1 = k] \cdot \mathbb{P}[X_2 = n - k] \\ &= \sum_{k=0}^{n} e^{-(\lambda_1+\lambda_2)} \frac{\lambda_1^k \cdot \lambda_2^{n-k}}{k!(n-k)!} \frac{n!}{n!} \\ &= e^{-(\lambda_1+\lambda_2)} \frac{1}{n!} \sum_{k=0}^{n} \binom{n}{k} \lambda_1^k \lambda_2^{n-k} \\ &= e^{-(\lambda_1+\lambda_2)} \frac{(\lambda_1 + \lambda_2)^n}{n!}. \end{aligned}$$

Somit ist $X_1 + X_2 \sim \mathrm{Poi}(\lambda_1 + \lambda_2)$. □

Übung 8.14.5. Seien $X_1 \sim \mathrm{Bin}(n_1, p)$ und $X_2 \sim \mathrm{Bin}(n_2, p)$ unabhängig. Dann gilt

$$X_1 + X_2 \sim \mathrm{Bin}(n_1 + n_2, p).$$

Das heißt: $\mathrm{Bin}(n_1, p) * \mathrm{Bin}(n_2, p) = \mathrm{Bin}(n_1 + n_2, p)$.

Übung 8.14.6. Seien $X_1 \sim \mathrm{NB}(r_1, p)$ und $X_2 \sim \mathrm{NB}(r_2, p)$ unabhängig. Dann gilt

$$X_1 + X_2 \sim \mathrm{NB}(r_1 + r_2, p).$$

Das heißt: $\mathrm{NB}(r_1, p) * \mathrm{NB}(r_2, p) = \mathrm{NB}(r_1 + r_2, p)$.

Als Spezialfall dieser Aussage ergibt sich

Übung 8.14.7. Seien $X_1, \ldots, X_n \sim \mathrm{Geo}(p)$ unabhängig. Dann gilt

$$X_1 + \ldots + X_n \sim \mathrm{NB}(n, p).$$

Das heißt: $\mathrm{Geo}(p) * \ldots * \mathrm{Geo}(p) = \mathrm{NB}(n, p)$.

Satz 8.14.8 (Faltungsformel für absolut stetige Zufallsvariablen). *Seien X_1, X_2 unabhängige und absolut stetige Zufallsvariablen mit Dichten f_{X_1} und f_{X_2}. Dann ist auch die Summe $X_1 + X_2$ absolut stetig mit Dichte*

$$f_{X_1+X_2}(z) = \int_{\mathbb{R}} f_{X_1}(y) f_{X_2}(z-y) dy \text{ für fast alle } z \in \mathbb{R}. \tag{8.14.2}$$

Definition 8.14.9. Die Dichte $f_{X_1+X_2} = f_{X_1} * f_{X_2}$ heißt die *Faltung* der Dichten f_{X_1} und f_{X_2}.

Bemerkung 8.14.10. Eine intuitive Erklärung dieser Formel: Damit $X_1 + X_2 = z$ ist, muss X_1 irgendeinen Wert y annehmen (diesem Ereignis entspricht die Dichte $f_{X_1}(y)$) und dann muss $X_2 = z - y$ gelten (diesem Ereignis entspricht die Dichte $f_{X_2}(z-y)$). Da X_1 und X_2 unabhängig sind, multipliziert man beide Dichten. Da außerdem y beliebige Werte annehmen kann, integriert man zum Schluss über y.

Beweis von Satz 8.14.8. Die rechte Seite der Faltungsformel bezeichnen wir mit $g(z)$, d.h.

$$g(z) = \int_{\mathbb{R}} f_{X_1}(y) f_{X_2}(z-y) dy.$$

Für alle $t \in \mathbb{R}$ gilt mit dem Satz von Fubini (Maßtheorie)

$$\begin{aligned}
\int_{-\infty}^{t} g(z) dz &= \int_{-\infty}^{t} \int_{\mathbb{R}} f_{X_1}(y) f_{X_2}(z-y) dy dz \\
&= \int_{\mathbb{R}} \int_{-\infty}^{t} f_{X_1}(y) f_{X_2}(z-y) dz dy \\
&= \int_{\mathbb{R}} \int_{-\infty}^{t-y} f_{X_1}(y) f_{X_2}(u) du dy,
\end{aligned}$$

wobei wir die neue Variable $u = z - y$ eingeführt haben. Definiere die Menge

$$B \stackrel{def}{=} \{(u, y) : u + y \leq t\}.$$

Dann kann man das Integral so darstellen:

$$\int_{-\infty}^{t} g(z) dz = \int_{B} f_{X_1}(y) f_{X_2}(u) du dy = \int_{B} f_{X_1, X_2}(y, u) d(u, y) = \mathbb{P}[(X_1, X_2) \in B].$$

Dabei gilt die Formel $f_{X_1}(y) f_{X_2}(u) = f_{X_1, X_2}(y, u)$ wegen der Unabhängigkeit von X_1 und X_2. Somit haben wir gezeigt, dass

$$\int_{-\infty}^{t} g(z) dz = \mathbb{P}[X_1 + X_2 \leq t] = F_{X_1+X_2}(t).$$

Daraus folgt, dass g die Dichte von $X_1 + X_2$ ist. $\square$

Satz 8.14.11. *Seien $X_1 \sim N(\mu_1, \sigma_1^2)$ und $X_2 \sim N(\mu_2, \sigma_2^2)$ normalverteilt und unabhängig. Dann ist auch die Summe $X_1 + X_2$ normalverteilt:*

$$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

Beweis. Schritt 1. Seien zuerst $\mu_1 = \mu_2 = 0$ und somit $X_1 \sim \mathrm{N}(0, \sigma_1^2)$ und $X_2 \sim \mathrm{N}(0, \sigma_2^2)$. Mit der Faltungsformel ergibt sich

$$\begin{aligned}
f_{X_1+X_2}(z) &= \int_{-\infty}^{\infty} f_{X_1}(y) f_{X_2}(z-y) dy \\
&= \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{y^2}{\sigma_1^2} + \frac{(z-y)^2}{\sigma_2^2}\right)} dy \\
&= \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{z^2}{2(\sigma_1^2+\sigma_2^2)}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{y^2}{\sigma_1^2} + \frac{(z-y)^2}{\sigma_2^2} - \frac{z^2}{2(\sigma_1^2+\sigma_2^2)}\right)} dy \\
&= \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{z^2}{2(\sigma_1^2+\sigma_2^2)}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{y(\sigma_1^2+\sigma_2^2) - z\sigma^2}{\sigma_1\sigma_2\sqrt{\sigma_1^2+\sigma_2^2}}\right)^2} dy.
\end{aligned}$$

Nun führen wir eine neue Variable $w = \frac{y(\sigma_1^2+\sigma_2^2)-z\sigma^2}{\sigma_1\sigma_2\sqrt{\sigma_1^2+\sigma_2^2}}$ ein:

$$\begin{aligned}
f_{X_1+X_2}(z) &= \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{z^2}{2(\sigma_1^2+\sigma_2^2)}} \frac{\sigma_1\sigma_2}{\sqrt{\sigma_1^2+\sigma_2^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}w^2} dw \\
&= \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{z^2}{2(\sigma_1^2+\sigma_2^2)}} \frac{\sigma_1\sigma_2}{\sqrt{\sigma_1^2+\sigma_2^2}} \cdot \sqrt{2\pi} \\
&= \frac{1}{\sqrt{2\pi}\sqrt{\sigma_1^2+\sigma_2^2}} e^{-\frac{z^2}{2(\sigma_1^2+\sigma_2^2)}}.
\end{aligned}$$

Also gilt $X_1 + X_2 \sim \mathrm{N}(0, \sigma_1^2 + \sigma_2^2)$.

Schritt 2. Seien nun $X_1 \sim \mathrm{N}(\mu_1, \sigma_1^2)$ und $X_2 \sim \mathrm{N}(\mu_2, \sigma_2^2)$ normalverteilt mit beliebigen Parametern. Betrachte die zentrierten Zufallsvariablen $X_1' := X_1 - \mu_1 \sim \mathrm{N}(0, \sigma_1^2)$ und $X_2' := X_2 - \mu_2 \sim \mathrm{N}(0, \sigma_2^2)$. Für die Zufallsvariablen X_1' und X_2' haben wir in Schritt 1 bereits gezeigt, dass $X_1' + X_2' \sim \mathrm{N}(0, \sigma_1^2 + \sigma_2^2)$. Somit gilt

$$X_1 + X_2 = (\mu_1 + \mu_2) + (X_1' + X_2') \sim \mathrm{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

Später werden wir diesen Satz auf einem viel schönerem Weg mithilfe von charakteristischen Funktionen beweisen. □

Beispiel 8.14.12. Die Dauer eines Telefongesprächs sei eine exponentialverteilte Zufallsvariable mit Parameter 1. Betrachte zwei unabhängig stattfindende Gespräche. Wie ist dann die Gesamtdauer der beiden Gespräche verteilt?

Lösung. Wir betrachten also zwei unabhängige Zufallsvariablen $X_1 \sim \text{Exp}(1)$ und $X_2 \sim \text{Exp}(1)$ und bestimmen die Verteilung von $X_1 + X_2$. Die Dichten von X_1 und X_2 sind

$$f_{X_1}(y) = f_{X_2}(y) = \begin{cases} e^{-y}, & y > 0, \\ 0, & y \le 0. \end{cases}$$

Die Zufallsvariable $X_1 + X_2$ nimmt nur positive Werte an. Sei $z > 0$. Mit der Faltungsformel ergibt sich

$$f_{X_1+X_2}(z) = \int_{\mathbb{R}} f_{X_1}(y) f_{X_2}(z-y) dy = \int_0^z e^{-y} e^{-(z-y)} dy = e^{-z} \int_0^z 1 dy = z e^{-z}.$$

Es gilt somit

$$f_{X_1+X_2}(z) = \begin{cases} z e^{-z} & z > 0, \\ 0, & z \le 0. \end{cases}$$

Übung 8.14.13. Seien $X_1, \ldots, X_n \sim \text{Exp}(1)$ unabhängig. Dann ist die Dichte von $X_1 + \ldots + X_n$ gegeben durch

$$f_{X_1+\ldots+X_n}(z) = \begin{cases} \frac{1}{n!} z^n e^{-z}, & z > 0, \\ 0, & z \le 0. \end{cases}$$

Diese Verteilung nennt man *Erlang–Verteilung*.

Seien X_1, X_2 eine unabhängige Zufallsvariablen. Wir bestimmen die Verteilung des Produkts $X_1 X_2$. Sind X_1 und X_2 diskret, so kann man die Zähldichte von $X_1 X_2$ wie folgt berechnen

$$p_{X_1 X_2}(z) = \sum_{y \in \text{Im}(X_1) \backslash \{0\}} p_{X_1}(y) p_{X_2}\left(\frac{z}{y}\right), \quad z \neq 0.$$

Sind X_1 und X_2 absolut stetig, so gilt für die Dichte von $X_1 X_2$ die Formel

$$f_{X_1 X_2}(z) = \int_{-\infty}^{\infty} \frac{1}{|y|} f_{X_1}(y) f_{X_2}\left(\frac{z}{y}\right) dy.$$

Wir werden diese Formel nicht beweisen, sondern nur erklären, warum in ihr der Faktor $\frac{1}{|y|}$ auftaucht. Sei $\varepsilon > 0$ sehr klein. Dann gilt

$$f_{X_1 X_2}(z) \approx \frac{1}{\varepsilon} \cdot \mathbb{P}[z \le X_1 X_2 \le z + \varepsilon].$$

Was muss nun geschehen, damit das Ereignis $z \le X_1 X_2 \le z + \varepsilon$ eintritt? Zuerst muss X_1 irgendeinen Wert y annehmen. Diesem Ereignis entspricht die Dichte $f_{X_1}(y)$. Ist nun $y > 0$, so muss für X_2 gelten: $z/y \le X_2 \le z/y + \varepsilon/y$. Dieses Ereignis hat eine Wahrscheinlichkeit von $\approx f_{X_2}(z/y) \cdot \varepsilon/y$. Analog kommt man im Fall $y < 0$ auf das Ereignis $z/y + \varepsilon/y \le X_2 \le z/y$ mit einer Wahrscheinlichkeit von $\approx -f_{X_2}(z/y) \cdot \varepsilon/y$. Da nun y beliebig sein kann, integriert man über y und erhält

$$\mathbb{P}[z \le X_1 X_2 \le z + \varepsilon] \approx \varepsilon \int_{-\infty}^{\infty} \frac{1}{|y|} f_{X_1}(y) f_{X_2}\left(\frac{z}{y}\right) dy.$$

Daraus ergibt sich die Formel für die Dichte von $X_1 X_2$.

Analog kann man zeigen: Sind X_1 und X_2 unabhängig und beide absolut stetig, so gilt für die Dichte von X_1/X_2 die Formel

$$f_{X_1/X_2}(z) = \int_{-\infty}^{\infty} |y| f_{X_1}(zy) f_{X_2}(y) dy.$$

8.15. Transformationsformel für den Erwartungswert

Sei X eine absolut stetige Zufallsvariable mit Dichte f_X. Sei $\varphi : \mathbb{R} \to \mathbb{R}$ eine Borel-Funktion. Wie bestimmt man dann den Erwartungswert von $\varphi(X)$?

Satz 8.15.1 (Transformationsformel für den Erwartungswert). *Sei X eine absolut stetige Zufallsvariable und $\varphi : \mathbb{R} \to \mathbb{R}$ eine Borel-Funktion. Dann gilt*

$$\mathbb{E}\varphi(X) = \int_{-\infty}^{\infty} \varphi(y) f_X(y) dy, \tag{8.15.1}$$

falls $\int_{-\infty}^{\infty} |\varphi(y)| f_X(y) dy < \infty$.

Bemerkung 8.15.2. Intuitiv kann man die Formel so verstehen: Die Zufallsvariable X nimmt einen Wert y mit "Dichte" $f_X(y)$ an. Ist $X = y$, so gilt $\varphi(X) = \varphi(y)$. Der entsprechende Beitrag zum Erwartungswert ist also $\varphi(y) f_X(y)$. Da nun y beliebig sein kann, integrieren wir über y.

Beweis von Satz 8.15.1. Schritt 1. Sei zuerst $\varphi(y) = \mathbb{1}_A(y)$ eine Indikatorfunktion, wobei $A \subset \mathbb{R}$ eine Borel-Menge ist. Dann nimmt $\varphi(X)$ nur die Werte 0 und 1 an und es gilt

$$\mathbb{E}\varphi(X) = \mathbb{E}\mathbb{1}_A(X) = 1 \cdot \mathbb{P}[X \in A] + 0 \cdot \mathbb{P}[X \notin A] = \mathbb{P}[X \in A].$$

Auf der anderen Seite, gilt

$$\int_{-\infty}^{\infty} \varphi(y) f_X(y) dy = \int_{-\infty}^{\infty} \mathbb{1}_A(y) f_X(y) dy = \int_A f_X(y) dy = \mathbb{P}[X \in A].$$

Es folgt, dass Formel (8.15.1) erfüllt ist.

Schritt 2. Sei $\varphi(y) = \sum_{i=1}^n \alpha_i \mathbb{1}_{A_i}(y)$ eine Elementarfunktion. Dabei seien $A_1, \ldots, A_n \subset \mathbb{R}$ disjunkte Borel-Mengen und $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$. Dann gilt

$$\mathbb{E}\varphi(X) = \mathbb{E}\left[\sum_{i=1}^n \alpha_i \mathbb{1}_{A_i}(X)\right] = \sum_{i=1}^n \alpha_i \mathbb{E}[\mathbb{1}_{A_i}(X)] = \sum_{i=1}^n \alpha_i \mathbb{P}[X \in A_i],$$

wobei wir im letzten Schritt das Ergebnis von Schritt 1 benutzt haben. Auf der anderen Seite gilt

$$\int_{-\infty}^{\infty} \varphi(y) f_X(y) dy = \int_{-\infty}^{\infty} \sum_{i=1}^n \alpha_i \mathbb{1}_{A_i}(y) f_X(y) dy = \sum_{i=1}^n \alpha_i \int_{-\infty}^{\infty} \mathbb{1}_{A_i}(y) f_X(y) dy = \sum_{i=1}^n \alpha_i \mathbb{P}[X \in A_i],$$

wobei wir im letzten Schritt wieder das Ergebnis von Schritt 1 benutzt haben. Also gilt die Formel (8.15.1).

SCHRITT 3. Sei nun $\varphi \geq 0$ eine beliebige, nichtnegative Borel-Funktion. Wir approximieren φ mit Elementarfunktionen. Dazu definieren wir $\varphi_n : \mathbb{R} \to \mathbb{R}$ wie folgt

$$\varphi_n(y) = \sum_{k=1}^{n \cdot 2^n} \frac{k-1}{2^n} \mathbb{1}_{\frac{k-1}{2^n} \leq \varphi(y) < \frac{k}{2^n}} + n \cdot \mathbb{1}_{\varphi(y) \geq n}.$$

Die Folge φ_n hat folgende Eigenschaften:

(1) Für alle $y \in \mathbb{R}$ gilt $\varphi_n(y) \uparrow \varphi(y)$ für $n \to \infty$.
(2) Für alle $\omega \in \Omega$ gilt $\varphi_n(X(\omega)) \uparrow \varphi(X(\omega))$ für $n \to \infty$.

Der Satz von der monotonen Konvergenz (Maßtheorie) ergibt dann

(1) $\int \varphi_n(y) f_X(y) dy \to \int \varphi(y) f_X(y) dy$ für $n \to \infty$.
(2) $\mathbb{E}\varphi_n(X) \to \mathbb{E}\varphi(X)$ für $n \to \infty$.

Hier gilt $\int \varphi_n(y) f_X(y) dy = \mathbb{E}\varphi_n(X)$ nach Schritt 2 (denn φ_n ist eine Elementarfunktion), also sind die Terme $\int \varphi(y) f_X(y) dy$ und $\mathbb{E}\varphi(X)$ auch gleich und die Formel (8.15.1) gilt.

SCHRITT 4. Sei $\varphi : \mathbb{R} \to \mathbb{R}$ eine beliebige Borel-Funktion mit $\int |\varphi(y)| f_X(y) dy \leq \infty$. Wir schreiben dann

$$\varphi = \varphi_+ - \varphi_-,$$

wobei φ_+ und φ_- der positive bzw. der negative Anteil von φ ist:

$$\varphi_+(y) = \begin{cases} \varphi(y), & \text{falls } \varphi(y) \geq 0, \\ 0, & \text{falls } \varphi(y) < 0; \end{cases} \qquad \varphi_-(y) = \begin{cases} 0, & \text{falls } \varphi(y) \geq 0, \\ |\varphi(y)|, & \text{falls } \varphi(y) < 0. \end{cases}$$

Es gilt $\varphi_+ \geq 0$ und $\varphi_- \geq 0$. In Schritt 3 haben wir gezeigt, dass

$$\mathbb{E}\varphi_+(X) = \int \varphi_+(y) f_X(y) dy < \infty, \qquad \mathbb{E}\varphi_-(X) = \int \varphi_-(y) f_X(y) dy < \infty.$$

Dabei sind beide Erwartungswerte endlich wegen der Annahme $\int |\varphi(y)| f_X(y) dy \leq \infty$. Wir bilden die Differenz:

$$\mathbb{E}[\varphi(X)] = \mathbb{E}[\varphi_+(X) - \varphi_-(X)] = \int (\varphi_+(y) - \varphi_-(y)) f_X(y) dy = \int \varphi(y) f_X(y) dy.$$

Somit gilt die Formel (8.15.1). □

BEISPIEL 8.15.3. Mit $\varphi(y) = y$ erhalten wir die Formel

$$\mathbb{E}X = \int_{-\infty}^{\infty} y f_X(y) dy.$$

Etwas allgemeiner, mit $\varphi(y) = y^n$ erhalten wir

$$\mathbb{E}[X^n] = \int_{-\infty}^{\infty} y^n f_X(y) dy.$$

Die Zahlen $\mathbb{E}X, \mathbb{E}[X^2], \mathbb{E}[X^3], \ldots$ nennt man auch *Momente* von X.

BEISPIEL 8.15.4. Sei $X \sim \text{Exp}(1)$. Bestimme $\mathbb{E}[X^n]$.

LÖSUNG. Die Dichte von X ist $f_X(y) = e^{-y}$ für $y > 0$. Dann gilt:

$$\mathbb{E}[X^n] = \int_0^\infty y^n f_X(y)dy = \int_0^\infty y^n e^{-y}dy = n!.$$

BEMERKUNG 8.15.5. Eine ähnliche Transformationsformel gilt auch für diskrete Zufallsvariablen: ist X diskret, so haben wir

$$\mathbb{E}\varphi(X) = \sum_{y\in \mathrm{Im}(X)} \varphi(y)p_X(y),$$

falls $\sum_{y\in \mathrm{Im}(X)} |\varphi(y)|p_X(y) < \infty$.

8.16. Multiplikativität des Erwartungswerts

Sind $X : \Omega \to \mathbb{R}$ und $Y : \Omega \to \mathbb{R}$ zwei (möglicherweise abhängige) integrierbare Zufallsvariablen, so ist die Summe $X + Y$ ebenfalls integrierbar und es gilt

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y].$$

Dies ist eine der Eigenschaften des Lebesgue-Integrals. Nun beweisen wir, dass eine ähnliche Formel auch für das Produkt gilt, wenn man zusätzlich voraussetzt, dass die Zufallsvariablen unabhängig sind.

SATZ 8.16.1. *Die Zufallsvariablen $X, Y : \Omega \to \mathbb{R}$ seien integrierbar und unabhängig. Dann ist auch XY integrierbar und es gilt*

$$\mathbb{E}[XY] = (\mathbb{E}X) \cdot (\mathbb{E}Y). \tag{8.16.1}$$

BEWEIS. SCHRITT 1. Wenn die Zufallsvariablen nur endlich viele Werte annehmen, haben wir diesen Satz bereits im Kapitel über den Erwartungswert der diskreten Zufallsvariablen bewiesen.

SCHRITT 2. Annahme: Seien $X(\omega) \geq 0$ und $Y(\omega) \geq 0$ für alle $\omega \in \Omega$. Dann existieren zwei Folgen von Zufallsvariablen $X_n, Y_n : \Omega \to \mathbb{R}$ mit den Eigenschaften:

(1) Für jedes $\omega \in \Omega$ gilt $X_n(\omega) \uparrow X(\omega)$ für $n \to \infty$.
(2) Für jedes $\omega \in \Omega$ gilt $Y_n(\omega) \uparrow Y(\omega)$ für $n \to \infty$.
(3) Für jedes $\omega \in \Omega$ gilt $X_n(\omega) \geq 0, \quad Y_n(\omega) \geq 0$.
(4) X_n und Y_n nehmen nur endlich viele Werte an.

Beispielsweise kann man X_n und Y_n so konstruieren: $X_n = f_n(X)$ und $Y_n = f_n(Y)$ mit

$$f_n(z) = \sum_{k=1}^{n\cdot 2^n} \frac{k-1}{2^n} \mathbb{1}_{\frac{k-1}{2^n} \leq z \leq \frac{k}{2^n}} + n \cdot \mathbb{1}_{z\geq n}.$$

Damit gilt dann für alle $\omega \in \Omega$:

$$X_n(\omega) \cdot Y_n(\omega) \uparrow X(\omega) \cdot Y(\omega) \text{ für } n \to \infty.$$

Da die Zufallsvariablen X und Y unabhängig sind, sind auch $X_n = f_n(X)$ und $Y_n = f_n(Y)$ unabhängig. Aus dem Satz von der monotonen Konvergenz folgt, dass

$$\begin{aligned}\mathbb{E}[XY] &= \lim_{n\to\infty} \mathbb{E}[X_n \cdot Y_n] \\ &= \lim_{n\to\infty} \mathbb{E}[X_n] \cdot \mathbb{E}[Y_n] \\ &= \lim_{n\to\infty} \mathbb{E}[X_n] \cdot \lim_{n\to\infty} \mathbb{E}[Y_n] \\ &= \mathbb{E}[X] \cdot \mathbb{E}[Y].\end{aligned}$$

Dabei haben wir benutzt, dass $\mathbb{E}[X_n \cdot Y_n] = \mathbb{E}[X_n] \cdot \mathbb{E}[Y_n]$. Dies wurde in Schritt 1 gezeigt, denn X_n und Y_n nehmen nur endlich viele Werte an. Also gilt die Formel (8.16.1).

Schritt 3. Seien nun $X, Y : \Omega \to \mathbb{R}$ unabhängig und integrierbar. Dann sind $|X|$ und $|Y|$ nichtnegativ und wir können Schritt 2 anwenden:

$$\mathbb{E}|XY| = \mathbb{E}\left[|X| \cdot |Y|\right] = \mathbb{E}|X| \cdot \mathbb{E}|Y| < \infty.$$

Also ist XY integrierbar. Nun schreiben wir $X = X^+ - X^-$ und $Y = Y^+ - Y^-$ mit

$$X^+(\omega) = \begin{cases} X(\omega), & X(\omega) \geq 0, \\ 0, & X(\omega) < 0, \end{cases} \qquad X^-(\omega) = \begin{cases} 0, & X(\omega) \geq 0, \\ |X(\omega)|, & X(\omega) < 0, \end{cases}$$

$$Y^+(\omega) = \begin{cases} Y(\omega), & Y(\omega) \geq 0, \\ 0, & Y(\omega) < 0, \end{cases} \qquad Y^-(\omega) = \begin{cases} 0, & Y(\omega) \geq 0, \\ |Y(\omega)|, & Y(\omega) < 0. \end{cases}$$

Da X und Y unabhängig sind, gilt auch

(1) X^+ und Y^+ sind unabhängig.
(2) X^+ und Y^- sind unabhängig.
(3) X^- und Y^+ sind unabhängig.
(4) X^- und Y^- sind unabhängig.

Da nun X^+, Y^+, X^-, Y^- alle nichtnegativ sind, können wir auf diese Zufallsvariablen Schritt 2 anwenden:

$$\begin{aligned}\mathbb{E}[XY] &= \mathbb{E}\left[(X^+ - X^-)(Y^+ - Y^-)\right] \\ &= \mathbb{E}\left[X^+Y^+ - X^-Y^+ - X^+Y^- + X^-Y^-\right] \\ &= \mathbb{E}\left[X^+Y^+\right] - \mathbb{E}\left[X^-Y^+\right] - \mathbb{E}\left[X^+Y^-\right] + \mathbb{E}\left[X^-Y^-\right] \\ &= \mathbb{E}\left[X^+\right] \cdot \mathbb{E}\left[Y^+\right] - \mathbb{E}\left[X^-\right] \cdot \mathbb{E}\left[Y^+\right] - \mathbb{E}\left[X^+\right] \cdot \mathbb{E}\left[Y^-\right] + \mathbb{E}\left[X^-\right] \cdot \mathbb{E}\left[Y^-\right] \\ &= \mathbb{E}\left[X^+ - X^-\right] \cdot \mathbb{E}\left[Y^+ - Y^-\right] \\ &= \mathbb{E}[X] \cdot \mathbb{E}[Y].\end{aligned}$$

Also gilt die Formel $\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$. $\square$