2 Zufallsvariablen

2.1 Induzierter Raum und Verteilung

Wir kommen zum wichtigsten Begriff der W-Theorie. In den meisten Situationen sind wir nicht an der gesamten Verteilung interessiert, sondern nur an bestimmten Daten, die vom Experiment abhängen.

Beispiele. Im Lotto interessiert die Anzahl der richtigen Zahlen, beim zweimaligen Würfeln zum Beispiel die Augensumme, und beim Münzwurf, im wievielten Wurf zum ersten Mal Kopf kommt.

Definition. Sei $(\Omega, \mathcal{E}, P)$ ein W-Raum, $X : \Omega \to W$ (Wertebereich), $\mathcal{F}$ σ-Algebra auf W. Falls $X^{-1}(F) \in \mathcal{E}$ für alle $F \in \mathcal{F}$, so heißt X *Zufallsvariable* auf Ω, genauer $(W, \mathcal{F})$-wertige Zufallsvariable.

Wenn W endlich oder abzählbar ist, so sei $\mathcal{F} = 2^W$. Wir werden fast ausschließlich $W \subseteq \mathbb{R}$ betrachten, $\mathcal{F}$ = Borelmengen. X heißt dann *reelle* Zufallsvariable.

Definition. Der durch X *induzierte* W-Raum ist $(W, \mathcal{F}, P_X)$ mit $P_X(F) = P(X^{-1}(F))$. Wir schreiben kurz $P(X \in F)$ für $P(\{\omega \in \Omega : X(\omega) \in F\})$. Zum Beispiel $P(X = x) = P(\{\omega \in \Omega : X(\omega) = x\})$ oder $P(X \leq x) = P(\{\omega \in \Omega : X(\omega) \leq x\}$, falls $W = \mathbb{R}$.

Satz 2.1. *P_X ist W-Maß auf $(W, \mathcal{F})$.*

Beweis. $P_X(W) = P(X^{-1}(W)) = P(\Omega) = 1$, und ferner $P_X(\dot{\bigcup\limits_i} F_i) = P(X^{-1}(\dot{\bigcup\limits_i} F_i)) = P(\dot{\bigcup\limits_i} X^{-1}(F_i)) = \sum\limits_i P(X^{-1}(F_i)) = \sum\limits_i P_X(F_i)$. $\square$

Beispiele. 1. Münzwurf, $\Omega = \{K, Z\}$, $P(K) = p$, $P(Z) = 1 - p$. Sei X = Nummer des ersten Kopfwurfs, $W = \mathbb{N}$. Es ist $P(X = n) = (1-p)^{n-1}p$, und wir erhalten die geometrische Verteilung für X.

2. Zweimaliges Würfeln, $\Omega = \{1, 2, \ldots, 6\}^2$ mit Gleichverteilung. Sei X= Augensumme, $W = \{1, 2, \ldots, 12\}$. Dann ist $P(X = s) = \frac{1}{36} \cdot \#\{(i,j) \in \Omega : i + j = s\}$, zum Beispiel $P(X = 10) = \frac{3}{36} = \frac{1}{12}$.

Definition. Sei $X : \Omega \to \mathbb{R}$. Die *Verteilungsfunktion* von X ist $F_X : \mathbb{R} \to [0, 1]$ mit

$$F_X(x) = P(X \leq x) = P(\{\omega \in \Omega : X(\omega) \leq x\}) \,.$$

Offenbar gilt

$$\lim_{x\to-\infty} F(x) = 0, \ \lim_{x\to\infty} F(x) = 1\,,$$
$$x < y \Rightarrow F(x) \leq F(y)\,.$$

Ferner ist $F(x)$ *rechtsstetig*, das heißt mit $h \searrow 0$ geht $F(x+h) \to F(x)$.

Beispiele. 1. Münzwurf. $\Omega = \{K, Z\}$, $X(K) = 1$, $X(Z) = 0$, $P(K) = p$, $P(Z) = 1 - p$. Es ist

$$F(x) = \begin{cases} 0 & x < 0 \\ 1-p & 0 \leq x < 1 \\ 1 & x \geq 1\,. \end{cases}$$

2. Indikatorvariable. Sei $A \in \mathcal{E}$, $I_A : \Omega \to \mathbb{R}$ mit

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

Es ist

$$F_{I_A}(x) = \begin{cases} 0 & x < 0 \\ 1-P(A) & 0 \leq x < 1 \\ 1 & x \geq 1\,. \end{cases}$$

Angenommen, X und Y sind zwei reelle Zufallsvariablen, $X, Y : \Omega \to \mathbb{R}$. Wir fassen (X, Y) als Zufallsvariable auf mit $(X, Y) : \Omega^2 \to \mathbb{R}^2$, $(X, Y)(\omega_1, \omega_2) = (X(\omega_1), Y(\omega_2))$. Allgemein seien $X_1, \ldots, X_n : \Omega^n \to \mathbb{R}$, dann ist $(X_1, \ldots, X_n) : \Omega^n \to \mathbb{R}^n$. Die *gemeinsame* Verteilung ist

$$F(x_1, \ldots, x_n) = P(X_1 \leq x_1 \wedge X_2 \leq x_2 \wedge \ldots \wedge X_n \leq x_n)\,.$$

2.2 Diskrete Zufallsvariablen

Eine (reelle) Zufallsvariable $X : \Omega \to W \subseteq \mathbb{R}$ heißt *diskret*, falls W endlich oder abzählbar ist.

Seien $X, Y : \Omega \to W \subseteq \mathbb{R}$ reelle Zufallsvariablen. Wir werden X und Y als unabhängig ansehen, falls

$$P(X \in A \wedge Y \in B) = P(X \in A)P(X \in B)$$

für alle $A, B \subseteq W$ gilt. Das folgende (leicht zu beweisende) Resultat zeigt, dass man sich bei diskreten Zufallsvariablen auf einelementige Mengen A, B beschränken kann.

Lemma 2.2. *Zwei diskrete Zufallsvariablen $X, Y : \Omega \to W \subseteq \mathbb{R}$ sind genau dann unabhängig, wenn*

$$P(X = x \wedge Y = y) = P(X = x)P(Y = y) \quad \textit{für alle } x, y \in W$$

gilt.

Mit Induktion gilt analog das allgemeine Resultat: Die Zufallsvariablen $X_1, \ldots, X_n : \Omega \to W \subseteq \mathbb{R}$ sind genau dann unabhängig, wenn

$$P(X_1 = x_1 \wedge \ldots \wedge X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i)$$

ist für alle $(x_1, \ldots, x_n) \in W^n$.

Beispiel. Betrachten wir den Münzwurf, $\Omega = \{K, Z\}$, $P(K) = p$, $P(Z) = 1 - p$, $0 < p < 1$. Es sei $X = \#K$, $Y = \#Z$ bei einmaligem Wurf. Dann ist

$$P(X = 1 \wedge Y = 1) = 0,\ P(X = 1)P(Y = 1) = p(1 - p)\,,$$

also sind, wie zu erwarten, X und Y nicht unabhängig.

Nehmen wir nun an, die Münze wird N Mal geworfen, wobei N Poisson verteilt ist, mit $\lambda > 0$. Das heißt, die W-keit N Mal zu werfen, ist $\frac{\lambda^N}{N!}e^{-\lambda}$. Dann ist

$$\begin{aligned} P(X = x \wedge Y = y) = &\ P(X = x \wedge Y = y \big| N = x + y)P(N = x + y) \\ = &\ \binom{x+y}{x} p^x (1-p)^y \frac{\lambda^{x+y}}{(x+y)!} e^{-\lambda} = \frac{(\lambda p)^x}{x!} \frac{(\lambda(1-p))^y}{y!} e^{-\lambda}\,, \end{aligned}$$

und

$$\begin{aligned} P(X = x) &= \sum_{n \geq x} P(X = x \big| N = n)P(N = n) = \sum_{n \geq x} \binom{n}{x} p^x (1-p)^{n-x} \frac{\lambda^n}{n!} e^{-\lambda} \\ &= e^{-\lambda} \frac{(\lambda p)^x}{x!} \sum_{n \geq x} \frac{(\lambda(1-p))^{n-x}}{(n-x)!} = e^{-\lambda} \frac{(\lambda p)^x}{x!} e^{\lambda(1-p)} = \frac{(\lambda p)^x}{x!} e^{-\lambda p}\,. \end{aligned}$$

Analog erhalten wir $P(Y = y) = \frac{(\lambda(1-p))^y}{y!} e^{-\lambda(1-p)}$, und wir sehen, dass X und Y unabhängig sind.

Nun führen wir die wichtigsten Maßzahlen ein.

Definition. Sei $X : \Omega \to \mathbb{R}$ diskrete Zufallsvariable. Der *Erwartungswert* $\mathrm{E}[X]$ ist

$$\mathrm{E}[X] = \sum_x xP(X = x) = \sum_{\omega \in \Omega} X(\omega)P(\omega)\,,$$

falls die Summe absolut konvergiert.

Der Erwartungswert gibt also an, was für ein Wert von X im Mittel zu erwarten ist. Es gibt also stets ω, ω' mit

$$X(\omega) \leq \mathrm{E}[X],\ \ X(\omega') \geq \mathrm{E}[X]\,.$$

Für endliche Mengen Ω oder endlichem Wertebereich existiert $\mathrm{E}[X]$ natürlich immer.

Im Fall der Gleichverteilung auf endlichem Ω, $P(\omega) = \frac{1}{|\Omega|}$ für $\omega \in \Omega$, haben wir

$$\mathrm{E}[X] = \frac{1}{|\Omega|} \sum_\omega X(\omega)\,.$$

Ist $\Omega = \{\omega_1, \ldots, \omega_n\}$, $X(\omega_i) = x_i$, so können wir

$$\mathrm{E}[X] = \frac{x_1 + \cdots + x_n}{n}$$

schreiben. $\mathrm{E}[X]$ ist somit der übliche Durchschnittswert.

Lemma 2.3. *Ist $X : \Omega \to \mathbb{R}$ Zufallsvariable, $g : \mathbb{R} \to \mathbb{R}$ beliebig, so gilt*

$$\mathrm{E}[g(X)] = \sum_x g(x)P(X = x) = \sum_\omega g(X(\omega))P(\omega)\,,$$

falls die Summe konvergiert.

Beweis. Wir setzen $Y = g \circ X : \Omega \to \mathbb{R}$, $\Omega \xrightarrow{X} \mathbb{R} \xrightarrow{g} \mathbb{R}$. Dann ist

$$E[Y] = \sum_\omega Y(\omega)P(\omega) = \sum_\omega g(X(\omega))P(\omega) = \sum_x \sum_{\omega : X(\omega) = x} g(x)P(\omega) =$$

$$= \sum_x g(x)P(X = x). \qquad \square$$

Künftig werden wir den Zusatz, "falls die Summe existiert" weglassen. Alle Sätze sind mit dieser Einschränkung zu verstehen.

Beispiel. X nehme die Werte $-2, -1, 1, 3$ mit W-keiten $\frac{1}{4}, \frac{1}{8}, \frac{1}{4}, \frac{3}{8}$ an. Es sei $Y = X^2$. Dann ist

$$\mathrm{E}[X^2] = 4 \cdot \frac{1}{4} + 1 \cdot \frac{1}{8} + 1 \cdot \frac{1}{4} + 9 \cdot \frac{3}{8} = \frac{38}{8} = \frac{19}{4}\,.$$

Definition. Sei $k \in \mathbb{N}$. Das k-te *Moment* von X ist

$$m_k = \mathrm{E}[X^k]\,.$$

Das k-te *zentrale Moment* ist

$$\sigma_k = E[(X - \mathrm{E}[X])^k]\,.$$

Besonders wichtig sind

$$\mathrm{E}[X] = m_1,\ \mathrm{Var}[X] = \sigma_2 = E[(X - \mathrm{E}[X])^2]\,.$$

$\mathrm{Var}[X]$ heißt die *Varianz*, $\sqrt{\mathrm{Var}[X]}$ die *Streuung*.

Die Varianz gibt also an, welche (quadratische) *Abweichung* vom Erwartungswert im Mittel zu erwarten ist. Ist zum Beispiel X eine konstante Funktion, $X(\omega) = c$ für alle ω, so haben wir $\mathrm{E}[X] = c$, $\mathrm{Var}[X] = 0$.

Satz 2.4 (Linearität des Erwartungswertes)**.** *Seien* $X_1, \ldots, X_n : \Omega \to \mathbb{R}$, $X = \lambda_1 X_1 + \cdots + \lambda_n X_n$. *Dann ist* $\mathrm{E}[X] = \sum\limits_{i=1}^{n} \lambda_i \mathrm{E}[X_i]$.

Beweis. Wir haben

$$\mathrm{E}[X] = \sum_{\omega} X(\omega)P(\omega) = \sum_{\omega}(\sum_{i=1}^{n} \lambda_i X_i(\omega))P(\omega) = \sum_{i=1}^{n} \lambda_i \mathrm{E}[X_i]\,. \qquad \square$$

Folgerung 2.5. *Es ist* $\mathrm{Var}[X] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2$.

Beweis. Es gilt

$$\begin{aligned} \mathrm{Var}[X] &= \mathrm{E}[(X - \mathrm{E}[X])^2] = E[X^2 - 2\mathrm{E}[X]X + (\mathrm{E}[X])^2] \\ &= \mathrm{E}[X^2] - 2\mathrm{E}[X]\mathrm{E}[X] + (\mathrm{E}[X])^2 \\ &= \mathrm{E}[X^2] - (\mathrm{E}[X])^2\,. \qquad \square \end{aligned}$$

Satz 2.6. *Es seien* $X : \Omega \to T \subseteq \mathbb{R}$, $Y : \Omega \to U \subseteq \mathbb{R}$ *unabhängige Zufallsvariablen. Dann gilt* $\mathrm{E}[XY] = \mathrm{E}[X]\mathrm{E}[Y]$.

Beweis. Wir haben

$$
\begin{aligned}
\mathrm{E}[XY] &= \sum_{\omega} X(\omega)Y(\omega)P(\omega) = \sum_{(x,y)\in T\times U} xyP(X=x \wedge Y=y) \\
&= \sum_{(x,y)} xyP(X=x)P(Y=y) = \sum_{x} xP(X=x)\sum_{y} yP(Y=y) \\
&= \mathrm{E}[X]\mathrm{E}[Y]\,. \qquad \square
\end{aligned}
$$

Satz 2.7. *Für die Varianz gilt*

a. $\mathrm{Var}[X+c] = \mathrm{Var}[X]\,,$

b. $\mathrm{Var}[cX] = c^2\mathrm{Var}[X]\,,$

c. $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$, *falls* X, Y *unabhängig sind.*

Beweis. a. $\mathrm{Var}[X+c] = \mathrm{E}[(X+c-\mathrm{E}[X+c])^2] = E[(X-\mathrm{E}[X])^2] = \mathrm{Var}[X]$.

b.
$$
\begin{aligned}
\mathrm{Var}[cX] &= \mathrm{E}[(cX-\mathrm{E}[cX])^2] = \mathrm{E}[(cX-c\mathrm{E}[X])^2] \\
&= \mathrm{E}[c^2(X-\mathrm{E}[X])^2] = c^2\mathrm{Var}[X]\,.
\end{aligned}
$$

c.
$$
\begin{aligned}
\mathrm{Var}[X+Y] &= \mathrm{E}[(X+Y)^2] - (\mathrm{E}[X+Y])^2 \\
&= \mathrm{E}[X^2] + 2\mathrm{E}[XY] + \mathrm{E}[Y^2] - (\mathrm{E}[X])^2 - 2\mathrm{E}[X]\mathrm{E}[Y] - (\mathrm{E}[Y])^2 \\
&= \mathrm{E}[X^2] - (\mathrm{E}[X])^2 + \mathrm{E}[Y^2] - (\mathrm{E}[Y])^2 \\
&= \mathrm{Var}[X] + \mathrm{Var}[Y]\,. \qquad \square
\end{aligned}
$$

Beispiele. 1. Bernoulli Verteilung. $\Omega = \{K, Z\}$, $P(K) = p$, $P(Z) = 1-p$, $X : \Omega \to \{0,1\}$, $X(K) = 1$, $X(Z) = 0$. Dann ist $P(X=1) = p$, $P(X=0) = 1-p$ und somit

$$
\mathrm{E}[X] = 1\cdot p + 0\cdot(1-p) = p,\ \mathrm{E}[X^2] = p,\ \mathrm{Var}[X] = p - p^2 = p(1-p)\,.
$$

Die größte Varianz erhalten wir also für die faire Münze $p = \frac{1}{2}$.

2. Binomialverteilung. Wir führen unabhängig n Würfe durch mit $P(X = k) = \binom{n}{k}p^k(1-p)^{n-k}$, $X =$ Anzahl Kopf, $0 \le k \le n$. Sei $X_i = \begin{cases} 1 \\ 0 \end{cases}$ die Indikatorvariable im i-ten Wurf Kopf oder Zahl. Dann ist $X = X_1 + \cdots + X_n$, also

$$
\mathrm{E}[X] = np,\ \mathrm{Var}[X] = np(1-p)\,.
$$

3. Poissonverteilung. Sei $P(X = n) = \frac{\lambda^n}{n!} e^{-\lambda}$, $n \geq 0$. Dann ist

$$\mathrm{E}[X] = \sum_{n \geq 0} n \frac{\lambda^n}{n!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{n \geq 1} \frac{\lambda^{n-1}}{(n-1)!} = \lambda\,,$$

$$\begin{aligned}\mathrm{E}[X^2] &= \sum_{n \geq 0} n^2 \frac{\lambda^n}{n!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{n \geq 1} n \frac{\lambda^{n-1}}{(n-1)!} = \lambda e^{-\lambda} \sum_{n \geq 0} (n+1) \frac{\lambda^n}{n!} \\ &= \lambda e^{-\lambda} \left(\lambda \sum_{n \geq 1} \frac{\lambda^{n-1}}{(n-1)!} + e^{\lambda} \right) = \lambda^2 + \lambda\,.\end{aligned}$$

Also ist

$$\mathrm{Var}[X] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2 = \lambda\,.$$

Definition. Seien $X, Y : \Omega \to \mathbb{R}$. Die *Kovarianz* von X und Y ist

$$\mathrm{cov}[X, Y] = \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])]$$

Ausmultiplizieren ergibt

$$\mathrm{cov}[X, Y] = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y]\,,$$

also insbesondere $\mathrm{cov}[X, X] = \mathrm{Var}[X]$.

X und Y heißen *unkorreliert*, falls $\mathrm{cov}[X, Y] = 0$ ist. Wegen Satz 2.6 sind unabhängige Variablen X, Y stets unkorreliert.

Der *Korrelationskoeffizient* ist

$$\rho(X, Y) = \frac{\mathrm{cov}[X, Y]}{\sqrt{\mathrm{Var}[X]\mathrm{Var}[Y]}}\,.$$

Lemma 2.8. *Es gilt für Konstanten a, b*

$$\mathrm{Var}[aX + bY] = a^2\mathrm{Var}[X] + b^2\mathrm{Var}[Y] + 2ab\mathrm{cov}[X, Y]\,.$$

Beweis. Ausrechnen. $\square$

Lemma 2.9. *Sei $X : \Omega \to \mathbb{R}$, dann gilt*

a. $\mathrm{E}[X^2] = 0 \Leftrightarrow P(X = 0) = 1$,

b. $\mathrm{Var}[X] = 0 \Leftrightarrow P(X = \mathrm{E}[X]) = 1$.

Beweis. a. Es gilt $\mathrm{E}[X^2] = \sum_x x^2 P(X = x)$. Ist also $\mathrm{E}[X^2] = 0$, so muß $P(X = x) = 0$ sein für alle $x \neq 0$, also $P(X \neq 0) = 0$ und somit $P(X = 0) = 1$. Die Umkehrung ist ebenso klar.
b. Mit $\mathrm{Var}[X] = \mathrm{E}[(X - \mathrm{E}[X])^2]$ brauchen wir nur a) auf $X - \mathrm{E}[X]$ anzuwenden. □

Satz 2.10 (Cauchy-Schwarz Ungleichung)**.** *Es gilt*

$$(\mathrm{E}[XY])^2 \leq \mathrm{E}[X^2]\mathrm{E}[Y^2] \,,$$

mit Gleichheit genau dann, wenn $P(Y = aX) = 1$ für ein $a \in \mathbb{R}$, oder $P(X = a'Y) = 1$ für $a' \in \mathbb{R}$.

Beweis. Angenommen $\mathrm{E}[X^2] = 0$, dann ist $P(X = 0) = 1$ nach Lemma 2.9, somit $P(X = x) = 0$ für $x \neq 0$. Daraus folgt $P(X = x \wedge Y = y) = 0$ für $x \neq 0, y$, also

$$\mathrm{E}[XY] = \sum_{(x,y)} xyP(X = x \wedge Y = y) = 0,$$

und wir haben $P(X = 0 \cdot Y) = 1$. Ist umgekehrt $P(X = 0) = 1$, so gilt $\mathrm{E}[X^2] = 0$, $\mathrm{E}[XY] = 0$, und wir erhalten Gleichheit. Wir können also $\mathrm{E}[X^2] > 0$ annehmen. Wir betrachten die Zufallsvariable $Z = Y - aX$ für $a \neq 0$. Dann gilt

$$0 \leq \mathrm{E}[Z^2] = \mathrm{E}[(Y - aX)^2] = \mathrm{E}[Y^2] - 2a\mathrm{E}[XY] + a^2\mathrm{E}[X^2] \,.$$

Als Polynom in a hat diese quadratische Gleichung höchstens eine reelle Nullstelle, also ist die Diskriminante

$$4\mathrm{E}[XY]^2 - 4\mathrm{E}[X^2]\mathrm{E}[Y^2] \leq 0 \,,$$

und das ist die Cauchy-Schwarz Ungleichung. Gleichheit gilt genau für $E[(Y - aX)^2] = 0$, also genau für $P(Y = aX) = 1$ nach Lemma 2.9. □

Folgerung 2.11. *Für den Korrelationskoeffizienten $\rho(X, Y)$ gilt $|\rho(X, Y)| \leq 1$, mit Gleichheit genau dann, wenn $P(Y = aX + b) = 1$, $a, b \in \mathbb{R}$, oder $P(X = a'Y + b') = 1$, $a', b' \in \mathbb{R}$.*

Beweis. Wir setzen $U = X - \mathrm{E}[X]$, $V = Y - E[Y]$, dann ist $\mathrm{cov}[X, Y] = \mathrm{E}[UV]$, $\mathrm{Var}[X] = \mathrm{E}[U^2]$, $\mathrm{Var}[Y] = \mathrm{E}[V^2]$. Satz 2.10 ergibt somit

$$(\mathrm{cov}[X, Y])^2 \leq \mathrm{Var}[X]\mathrm{Var}[Y]$$

also

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

Gleichheit gilt für $\mathrm{Var}[X] = 0$ oder $P(V = aU) = 1$ für $a \in \mathbb{R}$. Im ersten Fall ist nach Lemma 2.9, $P(X = \mathrm{E}[X]) = 1$, also $P(X = 0 \cdot Y + \mathrm{E}[X]) = 1$, und umgekehrt. Im zweiten Fall haben wir $P(Y - E[Y] = a(X - \mathrm{E}[X])) = 1$, also $P(Y = aX + b) = 1$ mit $b = -a\mathrm{E}[X] + E[Y]$, und umgekehrt. $\square$

Wir besprechen noch zwei weitere fundamentale Ungleichungen.

Satz 2.12. *Es gilt*

a. *Ungleichung von Markov: Sei $X \geq 0$, dann ist*

$$P(X \geq a) \leq \frac{\mathrm{E}[X]}{a} \quad (a > 0)\,.$$

b. *Ungleichung von Tschebyschev:*

$$P(|X - \mathrm{E}[X]| \geq a) \leq \frac{\mathrm{Var}[X]}{a^2} \quad (a > 0)\,.$$

Beweis. a. Wir haben

$$\begin{aligned}\mathrm{E}[X] = \sum_{\omega} X(\omega)P(\omega) &= \sum_{X(\omega)\geq a} X(\omega)P(\omega) + \underbrace{\sum_{X(\omega)<a} X(\omega)P(\omega)}_{\geq 0} \\ &\geq a \sum_{X(\omega)\geq a} P(\omega) = aP(X \geq a).\end{aligned}$$

b. Sei $Y = |X - \mathrm{E}[X]|^2$, dann ist $P(Y \geq a^2) \leq \frac{\mathrm{E}[Y]}{a^2}$ nach a), somit

$$P(|X - \mathrm{E}[X]| \geq a) = P(|X - \mathrm{E}[X]|^2 \geq a^2) \leq \frac{\mathrm{E}[Y]}{a^2} = \frac{\mathrm{Var}[X]}{a^2}\,. \qquad \square$$

Beispiel. Es seien $X_1, \ldots, X_n$ unabhängige Kopien von X (zum Beispiel n Münzwürfe), $Y = X_1 + \cdots + X_n$. Dann ist $\mathrm{E}[Y] = n\mathrm{E}[X]$, $\mathrm{Var}[Y] = n\mathrm{Var}[X]$, also

$$\mathrm{E}[\frac{X_1 + \cdots + X_n}{n}] = \mathrm{E}[X], \ \mathrm{Var}[\frac{X_1 + \cdots + X_n}{n}] = \frac{\mathrm{Var}[X]}{n} .$$

Tschebyschev ergibt

$$P(\Big|\frac{X_1 + \cdots + X_n}{n} - \mathrm{E}[X]\Big| \geq a) \leq \frac{1}{n}\frac{\mathrm{Var}[X]}{a^2}$$

und dies geht gegen 0 für $n \to \infty$.

Ebenso wie bedingte W-keiten können wir auch bedingte Erwartungswerte betrachten. Es seien $X, Y : \Omega \to \mathbb{R}$, dann sei $Y\big|X = x$ die Zufallsvariable Y unter der Annahme $X = x$. Wir setzen $\psi(x) = \mathrm{E}[Y\big|X = x]$, $\psi(X)$ ist also wieder eine Zufallsvariable.

Satz 2.13. *Es gilt* $\mathrm{E}[\psi(X)] = \mathrm{E}[Y]$*, das heißt also*

$$\mathrm{E}[Y] = \sum_x \mathrm{E}[Y\big|X = x)P(X = x) .$$

Beweis. Wir haben

$$\begin{aligned}
\mathrm{E}[\psi(X)] &= \sum_x \psi(x)P(X = x) = \sum_x \mathrm{E}[Y\big|X = x]P(X = x) \\
&= \sum_x \sum_y yP(Y = y\big|X = x)P(X = x) \\
&= \sum_x \sum_y yP(Y = y \wedge X = x) = \sum_y y \sum_x P(Y = y \wedge X = x) \\
&= \sum_y yP(Y = y) = \mathrm{E}[Y] . \qquad \square
\end{aligned}$$

Beispiel. Sei X die Augensumme bei zweimaligem Würfeln mit Gleichverteilung, $Y = \#$ gerade Ziffern. Dann ist

$$\mathrm{E}[Y\big|X = 2] = 0, \ \mathrm{E}[Y\big|X = 3] = 1, \ E[Y\big|X = 4] = \frac{2}{3}, \ldots$$

Offenbar ist $\mathrm{E}[Y] = 1$.

2.3 Wahrscheinlichkeitserzeugende Funktionen

Für diskrete Zufallsvariablen $X : \Omega \to \mathbb{N}_0$ können wir eine kompakte Form finden.

Definition. Sei $X : \Omega \to \mathbb{N}_0$. Dann heißt die formale Reihe

$$G_X(z) = \sum_{n \geq 0} P(X = n) z^n$$

die *wahrscheinlichkeitserzeugende Funktion.*

Lemma 2.14. *Falls die jeweiligen Reihen konvergieren, so gilt:*

a. $\mathrm{E}[X] = G'(1)$,

b. $\mathrm{Var}[X] = G''(1) + G'(1) - (G'(1))^2$.

Beweis. a. Wir haben

$$\mathrm{E}[X] = \sum_{n \geq 0} nP(X = n) = G'(1)\,.$$

b. $\mathrm{Var}[X] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2 = \sum\limits_{n \geq 0} n^2 P(X = n) - (\mathrm{E}[X])^2$. Nun ist

$$G''(z) = \sum_{n \geq 1} n(n-1) P(X = n) z^{n-2} = \sum_{n \geq 2} n^2 P(X = n) z^{n-2} - \sum_{n \geq 2} nP(X = n) z^{n-2}$$

also

$$\sum_{n \geq 1} n^2 P(X = n) = G''(z)\big|_{z=1} + G'(z)\big|_{z=1}$$

somit

$$\mathrm{Var}[X] = G''(1) + G'(1) - (G'(1))^2\,. \qquad \square$$

Beispiele. 1. Binomialverteilung, $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$. Hier ist $G(z) = \sum\limits_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} z^k = (pz + (1-p))^n$, also $G'(z) = np(pz + (1-p))^{n-1}$, $G''(z) = n(n-1)p^2 (pz + (1-p))^{n-2}$. Es folgt

$$\mathrm{E}[X] = np,\ \mathrm{Var}[X] = n(n-1)p^2 + np - n^2p^2 = np(1-p)\,.$$

2. Poissonverteilung, $P(X = n) = \frac{\lambda^n}{n!}e^{-\lambda}$. Hier ist $G(z) = \sum\limits_{n\geq 0} \frac{\lambda^n}{n!}e^{-\lambda}z^n = e^{-\lambda}e^{\lambda z} = e^{\lambda(z-1)}$, $G'(z) = \lambda e^{\lambda(z-1)}$, $G''(z) = \lambda^2 e^{\lambda(z-1)}$. Es folgt

$$\mathrm{E}[X] = \lambda,\ \mathrm{Var}[X] = \lambda^2 + \lambda - \lambda^2 = \lambda\,.$$

Sind $A(z) = \sum\limits_{n\geq 0} a_n z^n$, $B(z) = \sum\limits_{n\geq 0} b_n z^n$, dann können wir das Produkt bilden. Der Koeffizient von z^n in $A(z)B(z)$ ist $\sum\limits_{k=0}^{n} a_k b_{n-k}$. Wir nennen

$$A(z)B(z) = \sum_{n\geq 0}(\sum_{k=0}^{n} a_k b_{n-k})z^n$$

die *Konvolution* von $A(z)$ und $B(z)$. Wir schreiben $[z^n]A(z)$ für den Koeffizienten von z^n in $A(z)$.

Beispiel. Hypergeometrische Verteilung. $X : \Omega \to \{0, 1, \ldots, r\}$, $P(X = k) = \frac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}}$, n, m, r fest.

Sei $G_m(z) = \sum\limits_{k\geq 0} \binom{r}{k}\binom{n-r}{m-k} z^k$, und

$$H(z,t) = (1 + tz)^r (1 + t)^{n-r}\,.$$

Dann ist

$$[t^m]H(z,t) = \sum_{k=0}^{m} \binom{r}{k} z^k \binom{n-r}{m-k} = G_m(z)\,,$$

also

$$H(z,t) = \sum_{m\geq 0} G_m(z)t^m\,.$$

Es folgt

$$\begin{aligned}
G'_m(z) &= [t^m]\frac{\partial}{\partial z}H(z,t)\\
&= [t^m]rt(1 + tz)^{r-1}(1 + t)^{n-r}\\
&= r[t^{m-1}](1 + tz)^{r-1}(1 + t)^{n-r}\\
&= r\sum_{k=0}^{m-1} \binom{r-1}{k} z^k \binom{n-r}{m-1-k}\,,
\end{aligned}$$

und mit $z = 1$,

$$G'_m(1) = r \sum_{k=0}^{m-1} \binom{r-1}{k} \binom{n-r}{m-1-k} = r \binom{n-1}{m-1}.$$

Daraus erhalten wir

$$\mathrm{E}[X] = \frac{1}{\binom{n}{m}} r \binom{n-1}{m-1} = \frac{rm}{n}.$$

Beispiel. Ein berühmtes Problem betrifft Zufallswege auf $\mathbb{Z}$. Ein Wanderer startet in 0 und geht immer mit W-keit 1/2 einen Schritt nach rechts oder links. Was ist die W-keit p_0, dass er letztlich wieder nach 0 zurückkommt?

Es sei a_n die Anzahl der Wege, die im n-ten Schritt zum *ersten Mal* nach 0 zurückkehren. Weiter sei b_n die Anzahl der Wege, die im n-ten Schritt nach 0 kommen. Also zum Beispiel $a_0 = 0$, $a_1 = 0$, $a_2 = 2$, $b_0 = 1$, $b_1 = 0$, $b_2 = 2$. Sei $A(z) = \sum_{n \geq 0} a_n z^n$, $B(z) = \sum_{n \geq 0} b_n z^n$. Das Ereignis $A = \{\text{kehrt zurück}\}$ ist daher die disjunkte Summe $A = \bigcup A_n$, $A_n = \{\text{kehrt im } n\text{-ten Schritt zum ersten Mal zurück}\}$ mit $P(A_n) = \frac{a_n}{2^n}$. Wir haben also

$$p_0 = \sum_{n \geq 0} \frac{a_n}{2^n} = A(\frac{1}{2}).$$

Nun ist

$$b_n = \sum_{k=0}^{n} a_k b_{n-k}$$

durch Klassifikation, wann er zum ersten Mal zurückkehrt. Wir erhalten also

$$B(z) = A(z)B(z) + 1,$$

somit

$$A(z) = 1 - \frac{1}{B(z)},$$

und insbesondere

$$p_0 = 1 \Leftrightarrow B(\frac{1}{2}) = \infty.$$

Zur Analyse von $B(z)$ sehen wir zunächst, das der Wanderer eine *gerade* Anzahl $2n$ von Schritten tun muß, um zu 0 zurückzukehren, n in jeder Richtung. Also ist

$$b_{2n} = \binom{2n}{n}, \; b_{2n+1} = 0 \quad (n \geq 0)\,,$$

und somit

$$B(\frac{1}{2}) = \sum_{n \geq 0} \frac{\binom{2n}{n}}{2^{2n}}\,.$$

Durch Induktion verifiziert man leicht

$$\binom{2n}{n} \geq \frac{2^{2n}}{2n}\,,$$

also

$$B(\frac{1}{2}) \geq 1 + \sum_{n \geq 1} \frac{1}{2n} = 1 + \frac{1}{2} \sum_{n \geq 1} \frac{1}{n} = \infty\,,$$

da die harmonische Reihe $\sum\limits_{n \geq 1} \frac{1}{n}$ divergiert.

Ergebnis: Der Wanderer kehrt mit W-keit 1 zum Ausgangspunkt 0 zurück.

Übrigens gilt im Fall, dass n mit W-keit p nach rechts geht und mit W-keit $1-p$ nach links geht, stets $p_0 < 1$ für alle $p \neq \frac{1}{2}$. Er kehrt also nur im Fall der Gleichverteilung fast sicher zu 0 zurück.

2.4 Binomialverteilung und ihre Approximation

Wir betrachten den besonders wichtigen Fall, dass die Zufallsvariable $X : \Omega \to \mathbb{R}$ binomialverteilt ist. Es sei $p = W$-keit Erfolg, $1 - p = W$-keit Mißerfolg, $X =$ Anzahl der Erfolge bei n Versuchen. Dann ist

$$P(X = k) = b(k, n; p) = \binom{n}{k} p^k (1-p)^{n-k}\,,$$

$$\mathrm{E}[X] = np, \; \mathrm{Var}[X] = np(1-p)\,.$$

Es ergibt sich das Problem, dass die Binomialkoeffizienten $\binom{n}{k}$ wegen des raschen Wachstums von $n!$ schwer zu berechnen sind. Man wird also versuchen, gute Approximationen zu bestimmen.

Betrachten wir zunächst die hypergeometrische Verteilung

$$\frac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}}.$$

Angenommen wir haben die Situation:

$$n, n-r, r \gg m, k\,,$$

das heißt $n, n-r, r$ sind "viel größer" als m und k.

Beispiel. Wir haben 500 Schrauben, 60 davon sind defekt, $n = 500$, $r = 60$, und wir testen $m = 5$ Schrauben.

Mit Zurücklegen haben wir die Binomialverteilung, ohne Zurücklegen ist die W-keit für Erfolg $\sim \frac{r}{n}$. Die Idee ist, dass für große $n, r, n-r$ die Ziehungen von m Schrauben sich wenig unterscheiden, ob wir mit oder ohne Zurücklegen ziehen.

Aufgrund der Annahmen machen wir die folgenden Approximationen:

1. $k \ll r \Rightarrow r(r-1)\cdots(r-k+1) \sim r^k$
2. $m \ll r \Rightarrow n(n-1)\cdots(n-m+1) \sim n^m$
3. $m-k \ll n-r \Rightarrow (n-r)(n-r-1)\cdots(n-r-m+k+1) \sim (n-r)^{m-k}$.

Wir erhalten

$$\begin{aligned}
h(k, m; r, n) = \frac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}} &= \tfrac{m!}{k!(m-k)!}\,\cdot \\
&\quad \tfrac{r(r-1)\cdots(r-k+1)(n-r)(n-r-1)\cdots(n-r-m+k+1)}{n(n-1)\cdots(n-m+1)} \\
&\sim \tbinom{m}{k}\tfrac{r^k(n-r)^{m-k}}{n^m} = \tbinom{m}{k}(\tfrac{r}{n})^k(1-\tfrac{r}{n})^{m-k} \\
&= b(k, m; \tfrac{r}{n})\,.
\end{aligned}$$

Im Erwartungswert stimmen die Ausdrücke überein.

$$\mathrm{E}[X_{\mathrm{hyp}}] = \frac{rm}{n}, \quad \mathrm{E}[X_{\mathrm{bin}}] = m\frac{r}{n}\,.$$

Beispiel. $n = 60$, $r = 25$, $m = 4$. Die Tabelle zeigt, dass wir gute Approximationen erhalten:

	$k = 0$	1	2	3	4
$b(k, 4; \frac{25}{60})$	0,107	0,336	0,366	0,165	0,026
$h(k, 4; 25, 60)$	0,116	0,331	0,354	0,167	0,032

Natürlich benötigen wir auch für die hypergeometrische Verteilung Binomialkoeffizienten. In der Praxis verwendet man als Approximation die Poisson Verteilung

$$p(k; \lambda) = \frac{\lambda^k}{k!} e^{-\lambda} \,.$$

Wir machen die Voraussetzungen

1. $n \gg 0 \Rightarrow (1 + \frac{x}{n})^n \sim e^x$,
2. $p \ll 1,\ k$ klein $\Rightarrow (1 - p)^k \sim 1 \,,$
3. $k \ll n \Rightarrow n(n-1) \cdots (n-k+1) \sim n^k \,.$

Daraus erhalten wir

$$\begin{aligned} b(k, n; p) &= \tbinom{n}{k} p^k (1-p)^{n-k} = \tfrac{n(n-1)\cdots(n-k+1)}{k!} p^k (1-p)^{n-k} \\ &\sim \tfrac{n^k}{k!} p^k (1-p)^{n-k} = \tfrac{1}{k!} (np)^k \tfrac{(1-p)^n}{(1-p)^k} \\ &\sim \tfrac{1}{k!} (np)^k (1-p)^n = \tfrac{1}{k!} (np)^k (1 - \tfrac{np}{n})^k \\ &\sim \tfrac{1}{k!} (np)^k e^{-np} = p(k; np) \,, \end{aligned}$$

also

$$b(k, n; p) \sim p(k; \lambda) \text{ mit } \lambda = np \,.$$

Wiederum sehen wir Gleichheit für den Erwartungswert

$$\mathrm{E}[X_{\text{bin}}] = np, \quad \mathrm{E}[X_{\text{pois}}] = \lambda = np \,.$$

Beispiel. Jemand spielt 3 Jahre im Lotto. Mit welcher W-keit kann man mindestens 3 richtige erwarten? Wir haben $p(X \geq 3) = 0,018$, also ist die W-keit

$$1 - b(0, 3 \cdot 52; 0,018) \sim 1 - p(0; 3 \cdot 52 \cdot 0,018) = 0,93 \,.$$

2.5 Zufallsvariablen mit Dichtefunktion

Definition. Die Zufallsvariable $X : \Omega \to \mathbb{R}$ heißt *stetig verteilt* mit Dichtefunktion, falls die Verteilungsfunktion $F(x) = P(X \leq x)$ von der Form

$$F(x) = \int\limits_{-\infty}^{x} f(t)dt\,,$$

ist, wobei $f(x)$ integrierbar ist, $f(x) \geq 0$, $\int\limits_{-\infty}^{\infty} f(x)dx = 1$. Es gilt dann

$$\lim_{x\to-\infty} F(x) = 0,\ \lim_{x\to\infty} F(x) = 1,\ F'(x) = f(x)\,.$$

Für $A \in \mathcal{E}$ haben wir

$$P(X \in A) = \int\limits_{A} f(t)dt\,,$$

insbesondere

$$P(a \leq X \leq b) = \int\limits_{a}^{b} f(t)dt\,.$$

Achtung: Es ist stets $P(X = x) = 0$.

Es seien $X, Y : \Omega \to \mathbb{R}$ stetig verteilte Zufallsvariablen mit Verteilungsfunktionen $F_X(x)$, $F_Y(y)$ und Dichten $f_X(x)$, $f_Y(y)$. Die *gemeinsame Verteilung* $F(x, y)$, falls sie existiert, ist

$$F(x, y) = P(X \leq x \wedge Y \leq y) = \int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y} f(u, v)dvdu\,.$$

Definition. $X, Y : \Omega \to \mathbb{R}$ heißen *unabhängig*, falls die Ereignisse $\{X \leq x\}$, $\{Y \leq y\}$ unabhängig sind für alle x, y, also wenn

$$F(x, y) = F_X(x)F_Y(y)\,,$$

gilt. Allgemein für n Variablen.

Lemma 2.15. *Sind $X, Y : \Omega \to \mathbb{R}$ stetig verteilt mit $F(x,y)$, $f(x,y)$ und $F_X(x), f_X(x), F_Y(y), f_Y(y)$ wie zuvor, so gilt*

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy, \ f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx$$

X und Y sind unabhängig genau dann, wenn

$$f(x,y) = f_X(x)f_Y(y)\,.$$

Beweis. Wir haben

$$F_X(x) = P(X \leq x) = \lim_{y\to\infty} P(X \leq x \wedge Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(u,v)dvdu$$

$$= \int_{-\infty}^{x} \left(\int_{-\infty}^{\infty} f(u,y)dy \right) du\,,$$

und somit

$$f_X(x) = F'_X(x) = \int_{-\infty}^{\infty} f(x,y)dy\,,$$

analog für $f_Y(y)$. X und Y unabhängig bedeutet $F(x,y) = F_X(x)F_Y(y)$ für alle x, y, somit

$$\frac{\partial}{\partial x}F(x,y) = F'_X(x)F_Y(y), \ \frac{\partial^2}{\partial x \partial y}F(x,y) = F'_X(x)F'_Y(y) = f_X(x)f_Y(y)\,,$$

also

$$f(x,y) = f_X(x)f_Y(y)\,.$$

Umgekehrt folgt aus $f(x,y) = f_X(x)f_Y(y)$,

$$F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u,v)dvdu = \int_{-\infty}^{x} \int_{-\infty}^{y} f_X(u)f_Y(v)dvdu = F_X(x)F_Y(y)\,. \qquad \square$$

Allgemein gilt: $X_1, \ldots, X_n : \Omega \to \mathbb{R}$ unabhängig

$$\Leftrightarrow F(x_1, \ldots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i) \text{ für alle } (x_1, \ldots, x_n)$$

$$\Leftrightarrow f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i) \text{ für alle } (x_1, \ldots, x_n)\,.$$

Wir erklären wieder den Erwartungswert und die Varianz. Alle Sätze für diskrete Variablen gelten auch hier, die Beweise sind allerdings etwas subtiler.

Definition. Der *Erwartungswert* und die *Varianz* von X sind

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x f(x) dx, \ \mathrm{Var}[X] = \int_{-\infty}^{\infty} (x - \mathrm{E}[X])^2 f(x) dx.$$

Beispiel. Gleichverteilung auf $[a, b]$. Hier ist $f(x) = \frac{1}{b-a}$,

$$F(x) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a < x \leq b \\ 1 & x > b\,. \end{cases}$$

Wir berechnen

$$\mathrm{E}[X] = \int_a^b x \frac{1}{b-a} dx = \frac{1}{b-a} \frac{x^2}{2} \Big|_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}\,,$$

$$\mathrm{Var}[X] = \int_a^b (x - \frac{a+b}{2})^2 \frac{1}{b-a} dx = \frac{1}{b-a} \frac{(x - \frac{a+b}{2})^3}{3} \Big|_a^b = \frac{(b-a)^2}{12}\,.$$

Insbesondere erhalten wir für das Einheitsintervall $[0, 1]$, $\mathrm{E}[X] = \frac{1}{2}$, $\mathrm{Var}[X] = \frac{1}{12}$.

Der folgende Satz gibt eine einfache Methode an, falls die Dichte $f(x) = 0$ ist für $x < 0$.

Lemma 2.16. *Sei $X : \Omega \to \mathbb{R}$ mit $f(x) = 0$ für $x < 0$. Dann gilt*

$$\mathrm{E}[X] = \int_0^{\infty} (1 - F(x)) dx\,,$$

wobei $F(x)$ die Verteilungsfunktion ist.

Beweis. Wir haben

$$\int_0^\infty (1-F(x))dx = \int_0^\infty P(X>x)dx = \int_{x=0}^\infty \left(\int_{y=x}^\infty f(y)dy\right) dx$$

$$= \int_{y=0}^\infty \left(\int_{x=0}^{y} dx\right) f(y)dy = \int_{y=0}^\infty y f(y)dy = \mathrm{E}[X]\,. \qquad \square$$

Beispiel. Exponentialverteilung. $X : [0,\infty) \to [0,\infty)$ mit $f(x) = \lambda e^{-\lambda x}$, $\lambda > 0$ für $x \geq 0$, $f(x) = 0$ für $x < 0$, $F(x) = 1 - e^{-\lambda x}$. Aus dem Lemma folgt

$$\mathrm{E}[X] = \int_0^\infty e^{-\lambda x} dx = -\frac{1}{\lambda} e^{-\lambda x}\Big|_0^\infty = \frac{1}{\lambda}\,.$$

Wir wollen nun die entsprechenden Sätze für $\mathrm{E}[X]$ und $\mathrm{Var}[X]$ herleiten.

Satz 2.17. *Sei $X : \Omega \to \mathbb{R}$ und $Y = g(X)$, wobei $g : \mathbb{R} \to \mathbb{R}$ stetig ist und $g(x) \geq 0$ für alle x. Dann gilt*

$$\mathrm{E}[g(X)] = \int_{-\infty}^\infty g(x)f(x)dx\,.$$

Beweis. Da $Y \geq 0$ ist, gilt nach dem Lemma

$$\mathrm{E}[Y] = \int_0^\infty (1 - F_Y(y))dy\,,$$

wobei

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y),\ 1 - F_Y(y) = P(g(X) > y)\,.$$

Sei A das Ereignis $A = \{x : g(x) > y\}$, dann ist

$$P(g(X) > y) = \int_A f(x)dx\,,$$

und somit

$$\mathrm{E}[Y] = \int_0^\infty \left(\int_A f(x)dx\right) dy = \int_{-\infty}^\infty \left(\int_{y=0}^{g(x)} dy\right) f(x)dx = \int_{-\infty}^\infty g(x)f(x)dx\,. \qquad \square$$

Bemerkung. Der Satz gilt auch ohne die Einschränkung $g(x) \geq 0$.

Folgerung 2.18. *Wir haben für $X : \Omega \to \mathbb{R}$*

a. $\mathrm{E}[aX + b] = a\mathrm{E}[X] + b$,

b. $\mathrm{Var}[X] = \mathrm{E}[(X - \mathrm{E}[X])^2] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2$,

c. $\mathrm{Var}[aX + b] = a^2\mathrm{Var}[X]$.

Beweis. Mit $Y = aX + b$, $g(x) = ax + b$ gilt

$$\mathrm{E}[aX + b] = \int_{-\infty}^{\infty} (ax + b)f(x)dx = a\mathrm{E}[X] + b\,.$$

b. Sei $Y = (X - \mathrm{E}[X])^2$, $g(x) = (x - \mathrm{E}[X])^2$, dann ist

$$\mathrm{E}[Y] = \int_{-\infty}^{\infty} (x - \mathrm{E}[X])^2 f(x)dx = \mathrm{Var}[X]\,.$$

Ausrechnen ergibt

$$\begin{aligned}\mathrm{Var}[X] &= \int_{-\infty}^{\infty} (x^2 - 2x\mathrm{E}[X] + (\mathrm{E}[X])^2)f(x)dx \\ &= \mathrm{E}[X^2] - 2(\mathrm{E}[X])^2 + (\mathrm{E}[X])^2 = \mathrm{E}[X^2] - (\mathrm{E}[X])^2\,.\end{aligned}$$

c. $$\begin{aligned}\mathrm{Var}[aX + b] &= \int_{-\infty}^{\infty} (ax - a\mathrm{E}[X])^2 f(x)dx \\ &= a^2 \int_{-\infty}^{\infty} (x - E[X])^2 f(x)dx = a^2\mathrm{Var}[X]\,. \qquad \square\end{aligned}$$

Beispiel. Standardnormalverteilung. Sei $X : \mathbb{R} \to \mathbb{R}$ mit Dichte $f(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$. Dann ist mit der Substitution $y = -x$

$$\mathrm{E}[X] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} xe^{-\frac{x^2}{2}}dx = -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} ye^{-\frac{y^2}{2}}dy = -\mathrm{E}[X]\,,$$

also $\mathrm{E}[X] = 0$. Für die Varianz erhalten wir mit partieller Integration

$$\operatorname{Var}[X] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2 = \mathrm{E}[X^2] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x \cdot x e^{-\frac{x^2}{2}} dx$$

$$= \frac{1}{\sqrt{2\pi}} \left(-x e^{-\frac{x^2}{2}} \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \right) = 1 \,.$$

Wir haben noch nicht die allgemeine Linearität des Erwartungswertes für stetige Zufallsvariablen bewiesen. Dazu müssen wir uns überlegen, wie die Dichte der Summe $X + Y$ aussieht.

Problem I. Gegeben X mit $F_X(x)$, $f_X(x)$ und $Y = g(X)$. Was ist $F_Y(y)$, $f_Y(y)$?

Problem II. Gegeben X und Y mit $F_X(x), f_X(x)$, $F_Y(y), f_Y(y)$. Was ist $F_{X+Y}(z)$, $f_{X+Y}(z)$?

Satz 2.19. *Sei $X : \Omega \to \mathbb{R}$ mit Dichte $f(x)$, $Y = g(X)$, wobei $g(x)$ monoton steigend ist mit $g'(x) \neq 0$ für alle x. Dann ist die Dichtefunktion $h(y)$ von Y gegeben durch*

$$h(y) = \frac{f(g^{-1}(y))}{g'(g^{-1}(y))} \,, \quad g^{-1} \text{ Umkehrfunktion.}$$

Beweis. Es ist $F_X(x) = \int\limits_{-\infty}^{x} f(t)dt = P(X \leq x)$, also

$$F_Y(y) = P(Y \leq y) = P(X \leq g^{-1}(y)) = \int_{-\infty}^{g^{-1}(y)} f(t)dt \,.$$

Mit der Substitution $t = g^{-1}(u)$ haben wir $dt = \frac{1}{g'(g^{-1}(u))} du$ und erhalten

$$F_Y(y) = \int_{-\infty}^{y} \frac{f(g^{-1}(u))}{g'(g^{-1}(u))} du \,,$$

und somit

$$h(y) = F_Y(y)' = \frac{f(g^{-1}(y))}{g'(g^{-1}(y))} \,. \qquad \square$$

Bemerkung. Der Satz gilt allgemein, so lange $g'(x) \neq 0$ ist für alle x.

Beispiel. Sei X $N(0,1)$-verteilt $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$, $Y = \sigma X + \mu$, $\sigma > 0$. Sei $g(x) = \sigma x + \mu$, dann ist $g^{-1}(y) = \frac{y-\mu}{\sigma}$, $g'(x) = \sigma$. Für die Dichte von Y erhalten wir demnach

$$h(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

das heißt Y ist $N(\mu, \sigma^2)$-verteilt. Ferner ist

$$\mathrm{E}[Y] = \sigma \mathrm{E}[X] + \mu = \mu, \mathrm{Var}[X] = \sigma^2 \mathrm{Var}[X] = \sigma^2 .$$

Beispiel. Sei X $N(0,1)$-verteilt. Was ist $h(y)$ für $Y = X^2$? Wir haben $g(x) = x^2$, $g^{-1}(y) = \pm\sqrt{y}$, $g'(x) = 2x$. Indem wir die positive und negative Wurzel in Betracht ziehen, erhalten wir mit $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

$$h(y) = \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} e^{-\frac{y}{2}}, \; y \geq 0 .$$

Diese Dichtefunktion sollte bekannt vorkommen, es ist die Γ-Verteilung mit $r = \alpha = \frac{1}{2}$. Also erhalten wir das Resultat:

$$X \sim N(0,1) \Rightarrow X^2 \sim \gamma_{\frac{1}{2},\frac{1}{2}} .$$

Betrachten wir nun Problem II.

Satz 2.20. *Seien $X, Y : \Omega \to \mathbb{R}$ und $f(x,y)$ die gemeinsame Dichte. Dann hat $Z = X + Y$ die Dichte*

$$f_Z(z) = \int_{-\infty}^{\infty} f(t, z-t) dt .$$

Beweis. Sei A das Ereignis

$$A = \{(x,y) : x + y \leq z\} = \{(x,y) : Z \leq z\} .$$

Dann ist

$$P(Z \leq z) = \iint_A f(u,v) du dv = \int_{u=-\infty}^{\infty} \int_{v=-\infty}^{z-u} f(u,v) dv du .$$

Die Substitution $x = u$, $y = u + v$ mit $u = x$, $v = y - x$ ergibt die Jacobi Matrix

$$\begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} = 1$$

also $dudv = dxdy$, und wir erhalten

$$P(Z \leq z) = \int\limits_{x=-\infty}^{\infty} \int\limits_{y=-\infty}^{z} f(x, y - x) dy dx = \int\limits_{y=-\infty}^{z} \left(\int\limits_{-\infty}^{\infty} f(x, y - x) dx \right) dy \,,$$

also

$$f_Z(z) = \int\limits_{-\infty}^{\infty} f(t, z - t) dt \,. \qquad \square$$

Folgerung 2.21. *Sind X, Y unabhängig, so gilt*

$$f_{X+Y}(x) = \int\limits_{-\infty}^{\infty} f_X(t) f_Y(x - t) dt \,.$$

Wir nennen diese Operation die *Faltung* zweier Funktionen f_X, f_Y, in Zeichen $f_X * f_Y$. Somit ist für unabhängige X und Y: $f_{X+Y} = f_X * f_Y$.

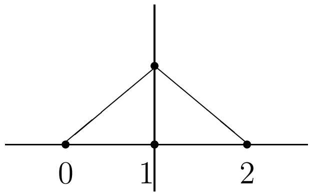
Beispiel. Es seien X und Y gleichverteilte unabhängige Variablen auf $[0, 1]$, also $f(x) = f_X(x) = 1$, $g(x) = f_Y(x) = 1$ für alle $x \in [0, 1]$ und 0 außerhalb. Für $X + Y$ erhalten wir die Dichte

$$h(x) = \int\limits_{0}^{x} f(t) g(x - t) dt \,.$$

Für $x \leq 1$ ergibt dies $h(x) = \int\limits_0^x dt = x$. Sei $x \geq 1$. Mit $t \leq 1$, $x - t \leq 1$ haben wir $x - 1 \leq t \leq 1$

$$h(x) = \int\limits_{x-1}^{1} dt = 2 - x.$$

Die Dichte von $X+Y$ ist also die Funktion



Für den Erwartungswert und die Varianz ergibt sich daraus

$$
\begin{aligned}
\mathrm{E}[X+Y] &= \int_0^2 x h(x) dx = \int\limits_0^1 x^2 dx + \int_1^2 (2x - x^2) dx \\
&= \frac{x^3}{3}\Big|_0^1 + x^2\Big|_1^2 - \frac{x^3}{3}\Big|_1^2 = \frac{1}{3} + 3 - \frac{8}{3} + \frac{1}{3} = 1, \\
\mathrm{E}[(X+Y)^2] &= \int\limits_0^2 x^2 h(x) dx = \frac{7}{6} \\
\mathrm{Var}[X+Y] &= \frac{7}{6} - 1 = \frac{1}{6}\,.
\end{aligned}
$$

Wir sehen also mit $\mathrm{E}[X] = \mathrm{E}[Y] = \frac{1}{2}$, $\mathrm{Var}[X] = \mathrm{Var}[Y] = \frac{1}{12}$, dass tatsächlich $\mathrm{E}[X+Y] = \mathrm{E}[X] + \mathrm{E}[Y]$ und $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$ gilt. Dies wollen wir allgemein zeigen.

Folgerung 2.22. *Seien $X, Y : \Omega \to \mathbb{R}$. Dann gilt* $\mathrm{E}[X+Y] = \mathrm{E}[X] + \mathrm{E}[Y]$.

Beweis. Es seien $f_X(x)$, $f_Y(y)$ die Dichten und $f(x,y)$ die gemeinsame Dichte. Wir haben

$$
\begin{aligned}
\mathrm{E}[X+Y] &= \int\limits_{z=-\infty}^{\infty} z \int\limits_{t=-\infty}^{\infty} f(t, z-t) dt dz \\
&= \int\limits_t \left(\int\limits_z z f(t, z-t) dz \right) dt \quad (z = y + t) \\
&= \int\limits_t \left(\int\limits_y (y+t) f(t, y) dy \right) dt
\end{aligned}
$$

$$= \int_t \left(\int_y y f(t,y) dy + t \int_y f(t,y) dy \right) dt$$
$$= \int_y y \left(\int_t f(t,y) dt \right) dy + \int_t t \left(\int_y f(t,y) dy \right) dt$$
$$= \int_y y f_Y(y) dy + \int_t t f_X(t) dt = \mathrm{E}[Y] + \mathrm{E}[X] \,. \qquad \square$$

Daraus folgt nach Folgerung 2.18a) auch $\mathrm{E}[aX+bY] = a\mathrm{E}[X]+b\mathrm{E}[Y]$. Allgemein verwendet man Induktion, um die Linearität $\mathrm{E}[\sum_{i=1}^{n} \lambda_i X_i] = \sum_{i=1}^{n} \lambda_i \mathrm{E}[X_i]$ zu zeigen. Die Regel $\mathrm{Var}[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} \mathrm{Var}[X_i]$ für unabhängige Variablen X_i wird ähnlich bewiesen.

Beispiel. Wir geben ohne Beweis die Resultate für die Normalverteilung und Γ-Verteilung an.

1. Ist X_1 $N(\mu_1, \sigma_1^2)$-verteilt, X_2 $N(\mu_2, \sigma_2^2)$-verteilt und X_1, X_2 unabhängig, so ist $X_1 + X_2$ $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$-verteilt.

2. Sind X_1 γ_{α,r_1}-verteilt, X_2 γ_{α,r_2}-verteilt, X_1, X_2 unabhängig, so ist $X_1 + X_2$ γ_{α,r_1+r_2}-verteilt.

3. Es seien $X_1, \ldots, X_n$ unabhängige Standard normalverteilte Zufallsvariablen. Dann ist $Y = X_1^2 + \cdots + X_n^2$ $\gamma_{\frac{1}{2},\frac{n}{2}}$-verteilt. Wir sagen, Y ist χ^2-verteilt mit n Freiheitsgraden. Die χ^2-Verteilung wird eine große Rolle in der Statistik spielen.

Beispiel. Seien $X, Y : \Omega \to [0, \infty)$ unabhängige Zufallsvariable, beide exponentiell verteilt zum selben $\lambda > 0$. Dann ist $f_X(x) = \lambda e^{-\lambda x}$, $f_Y(y) = \lambda e^{-\lambda y}$. Es gilt

$$f_{X+Y}(x) = \int_0^x \lambda e^{-\lambda t} \cdot \lambda e^{-\lambda(x-t)} dt$$
$$= \lambda^2 e^{-\lambda x} \int_0^x dt = \lambda^2 e^{-\lambda x} x \,.$$

Allgemein haben wir für unabhängige Variablen $X_1, \dots, X_n$ alle zum selben $\lambda > 0$

$$f_{X_1+\dots+X_n}(x) = \frac{\lambda^n x^{n-1}}{(n-1)!} e^{-\lambda x}\,.$$

Für $n = 1, 2$ wissen wir das. Nun verwenden wir Induktion.

$$\begin{aligned} f_{X_1+\dots+X_n}(x) = f_{(X_1+\dots+X_{n-1})+X_n}(x) &= \int\limits_0^x \frac{\lambda^{n-1} t^{n-2}}{(n-2)!} e^{-\lambda t} \cdot \lambda e^{-\lambda(x-t)} dt \\ &= \frac{\lambda^n e^{-\lambda x}}{(n-2)!} \int\limits_0^x t^{n-2} dt = \frac{\lambda^n x^{n-1}}{(n-1)!} e^{-\lambda x}\,. \end{aligned}$$

2.6 Exponentialverteilung

Wartezeiten werden im einfachsten Fall folgendermaßen modelliert: Es sei X die Wartezeit mit Werten in $[0, \infty)$ mit Dichte $f(x)$.

Als Beispiel könnten wir X als Wartezeit auf den ersten freien Parkplatz (kontinuierlich) oder als Wartezeit auf das erste Mal Kopf beim Münzwurf (diskret) interpretieren.

Definition. Wir sprechen von einer *gedächtnislosen* Wartezeit, wenn

$$P(X \geq s + t \big| X \geq t) = P(X \geq s)$$

für alle s, t gilt.

Dies bedeutet mit

$$P(X \geq s + t \big| X \geq t) = \frac{P(X \geq s + t \wedge X \geq t)}{P(X \geq t)} = \frac{P(X \geq s + t)}{P(X \geq t)}\,,$$

dass die Definition äquivalent ist zu

$$P(X \geq s + t) = P(X \geq s) P(X \geq t) \quad \text{für alle } s, t.$$

Satz 2.23. a. *Sei X exponential verteilt, $f(x) = \lambda e^{-\lambda x}$, dann gilt $P(X \geq s + t) = P(X \geq s)P(X \geq t)$.*
b. *Sei umgekehrt $f : [0, \infty) \to [0, \infty)$ stetig mit Verteilung $F(x) = P(X \leq x)$ gegeben und $P(X \geq s + t) = P(X \geq s)P(X \geq t)$, dann ist $f(x) = \lambda e^{-\lambda x}$ für ein $\lambda > 0$.*

Beweis. a. $P(X \geq x) = 1 - F(x) = e^{-\lambda x}$, also

$$P(X \geq s+t) = e^{-\lambda(s+t)} = e^{-\lambda s} e^{-\lambda t} = P(X \geq s) P(X \geq t)\,.$$

b. Sei $G(x) = \int\limits_x^\infty f(t)dt$, also

$$G(x) = P(X \geq x) = 1 - F(x)\,.$$

Wir stellen fest:

1. $G(0) = 1$,
2. $G(x) \to 0$ für $x \to \infty$,
3. $G(s+t) = G(s)G(t)$ für alle $s, t \geq 0$.

Angenommen es gibt $s \geq 0$ mit $G(s) = 0$. Dann ist wegen 3), $G(s) = G(\frac{s}{n})^n$, also $G(\frac{s}{n}) = 0$ für alle n, das heißt $G(0) = 0$ wegen der Stetigkeit von G, im Widerspruch zu 1).

Es sei $G(1) = e^\alpha$, also $\alpha = \log G(1)$.

Behauptung. $G(t) = e^{\alpha t}$ für alle t.
Zunächst haben wir $G(n) = G(1)^n = e^{\alpha n}$, dann $G(m) = G(\frac{m}{n})^n$ also $e^{\alpha m} = G(\frac{m}{n})^n$, das heißt $G(\frac{m}{n}) = e^{\alpha \frac{m}{n}}$ für alle $\frac{m}{n} \in \mathbb{Q}$. Die Funktionen $G(t)$ und $e^{\alpha t}$ stimmen also auf $\mathbb{Q}$ überein, und somit wegen der Stetigkeit auf $[0, \infty)$. Außerdem gilt $\alpha < 0$ wegen $G(x) = e^{\alpha x} \to 0$ für $x \to \infty$ nach 2).

Nun setzen wir $\lambda = -\alpha$ und erhalten

$$F(x) = 1 - G(x) = 1 - e^{-\lambda x}\,,$$

also $f(x) = F'(x) = \lambda e^{-\lambda x}\,.$ $\square$

Beispiel. Angenommen die Wartezeit im Sprechzimmer ist im Durchschnitt 20 Minuten. Was ist die W-keit, nach höchstens 10 Minuten dranzukommen? Wir setzen $\lambda = \frac{1}{20}$ ($\mathrm{E}[X] = \frac{1}{\lambda}$) und berechnen

$$P(X \leq 10) = 1 - e^{-\frac{10}{20}} = 1 - e^{-\frac{1}{2}} = 0{,}39.$$

Nach wievielen Minuten komme ich mit mindestens 90% W-keit an die Reihe? Sei T die Anzahl der Minuten, dann haben wir

$$\frac{1}{20}\int_0^T e^{-\frac{x}{20}}dx = 1 - e^{-\frac{T}{20}} \geq 0,9 \;\Rightarrow\; e^{-\frac{T}{20}} \leq 0,1$$

$$\Rightarrow -\frac{T}{20} \leq \log 0,1 \;\Rightarrow\; T \geq -20 \cdot \log 0,1 = 46,05.$$

Nun nehmen wir an, dass die Variablen $X_1, \ldots, X_n$ unabhängig sind, alle exponential verteilt zum selben $\lambda > 0$.

Satz 2.24. a. *Die Variable* $X_{\min} = \min(X_1, \ldots, X_n)$ *hat Dichtefunktion* $m_n(x) = n\lambda e^{-n\lambda x}$*, das heißt* X *ist exponential verteilt zu* $n\lambda$*.*

b. *Die Variable* $X_{\max} = \max(X_1, \ldots, X_n)$ *hat Dichtefunktion* $M_n(x) = n\lambda e^{-\lambda x} \cdot (1 - e^{-\lambda x})^{n-1}$.

Beweis. a. Gesucht ist $m_n(x)$ mit

$$P(X_{\min} \geq x) = \int_x^\infty m_n(t)dt\,.$$

Nun ist

$$\begin{aligned} P(X_{\min} \geq x) &= P(X_1 \geq x \wedge \ldots \wedge X_n \geq x) \\ &= \prod_{i=1}^n P(X_i \geq x) = \prod_{i=1}^n (1 - F(x)) \\ &= e^{-\lambda n x}\,, \end{aligned}$$

und somit

$$F_{X_{\min}}(x) = 1 - P(X_{\min} \geq x) = 1 - e^{-\lambda n x}\,,$$

und daher $m_n(x) = \lambda n e^{-\lambda n x}$.

b. Gesucht ist $M_n(x)$ mit $P(X_{\max} \leq x) = \int_0^x M_n(t)dt$. Nun ist

$$\begin{aligned} P(X_{\max} \leq x) &= P(X_1 \leq x \wedge \ldots \wedge X_n \leq x) = \prod_{i=1}^n P(X_i \leq x) \\ &= \prod_{i=1}^n (1 - e^{-\lambda x}) = (1 - e^{-\lambda x})^n\,, \end{aligned}$$

und es folgt

$$M_n(x) = F_{X_{\max}}(x)' = n(1 - e^{-\lambda x})^{n-1} e^{-\lambda x} \lambda \,. \qquad \square$$

Folgerung 2.25. *Wir haben*

a. $\mathrm{E}[X_{\min}] = \frac{1}{n\lambda}$,

b. $\mathrm{E}[X_{\max}] = \frac{1}{\lambda}(1 + \frac{1}{2} + \cdots + \frac{1}{n})$, $\mathrm{E}[X_{\max}] \to \infty$ *für* $n \to \infty$.

Beweis. Teil a) wissen wir schon, da $X_{\min}$ exponentialverteilt ist nach $n\lambda$.
b. Wir wenden Lemma 2.16 an und erhalten

$$\mathrm{E}[X_{\max}] = \int_0^\infty (1 - (1 - e^{-\lambda x})^n) dx \,.$$

Mit der Substitution $y = 1 - e^{-\lambda x}$ haben wir $e^{-\lambda x} = 1 - y$, $-\lambda x = \log(1 - y)$, also $x = -\frac{1}{\lambda} \log(1 - y)$, $dx = \frac{1}{\lambda} \frac{1}{1-y} dy$. Dies ergibt

$$\begin{aligned} \mathrm{E}[X_{\max}] &= \frac{1}{\lambda} \int_0^1 (1 - y^n) \frac{1}{1 - y} dy = \frac{1}{\lambda} \int_0^1 (1 + y + \cdots + y^{n-1}) dy \\ &= \frac{1}{\lambda}(y + \frac{y^2}{2} + \cdots + \frac{y^n}{n})\Big|_0^1 = \frac{1}{\lambda}(1 + \frac{1}{2} + \cdots + \frac{1}{n}) \,. \end{aligned}$$

Insbesondere geht $\mathrm{E}[X_{\max}] \to \infty$, da die harmonische Reihe divergiert. $\square$

Beispiel. Wenn bei 5 Telefonzellen die mittlere Wartezeit 2 Minuten beträgt $(\lambda = \frac{1}{2})$, so ist

$$\begin{aligned} \mathrm{E}[X_{\min}] &= \frac{2}{5} = 0,4 \,, \\ \mathrm{E}[X_{\max}] &= 2(1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5}) = \frac{137}{30} = 4,57 \,. \end{aligned}$$

Bemerkung. Die diskrete Variante für gedächtnislose Wartezeiten wird genau von der geometrisch verteilten Zufallsvariablen beschrieben.