Kapitel 2

Zufallsvariablen

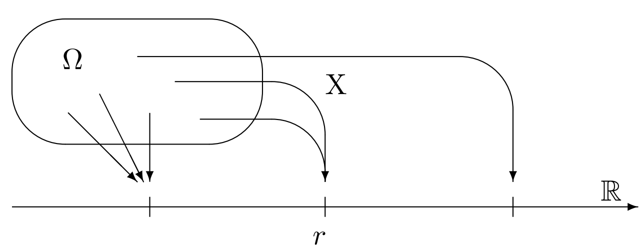
Motivation: Oft interessiert man sich nicht für den kompletten Stichprobenraum (den man häufig gar nicht genau kennt), sondern nur für Teilaspekte. ↓9.5.07
Beispiel: Wurf mehrerer Würfel. Die genaue Aufteilung der Ergebnisse auf die einzelnen Würfel ist irrelevant, interessant ist nur die Augensumme. Dieser Wert ist reell und verhält sich zufällig. Die Idee ist, statt des Gesamtexperiments (bei n Würfeln)

$$\Omega = \{\omega = (\omega_1, \ldots, \omega_n);\ \omega_i \in \{1, \ldots, 6\}\}$$

nur den Wert $X(\omega) = \sum_{i=1}^n \omega_i$ zu betrachen. Wir betrachten die Menge aller möglichen Werte

$$\tilde{\Omega} = X(\Omega) = \{n, n+1, \ldots, 6\,n\}$$

als neues (aber nicht mehr kombinatorisches) Zufallsexperiment, für das wir die Elementarw'keiten $P(X = r)$ für $r = n, \ldots, 6\,n$ durch Abzählen bestimmen können.
In diesem Fall ist $\tilde{\Omega}$ eine endliche Teilmenge von $\mathbb{R}$ (notwendigerweise endlich, da Ω schon endlich war). Wir wollen nun eine allgemeine Theorie aufbauen, in der wir beliebige Abbildungen $X : \Omega \mapsto \mathbb{R}$ über beliebigen Modellen betrachten und allgemein $\mathbb{R}$ mit einem „Bild-W'keitsmaß“ ausstatten, das von den W'keiten des ursprünglichen Experiments stammt:



$$P(X = r) = P(X^{-1}(r))$$

Wir haben mit dem Konzept kontinuierlicher W'räume 1.18 bereits eine Möglichkeit vorgestellt, den speziellen Stichprobenraum $\mathbb{R}$ über eine W'keitsdichte

$\rho(r)$ mit einem W'keitsmaß zu versehen:

$$P(r_1 < X \leq r_2) = \int_{r_1}^{r_2} \rho(r)\, dr \ .$$

Ist Ω diskret, so nimmt X nur diskrete Werte an. Leider können wir mit Hilfe einer Dichte keine W'keiten $P(X = r)$ für einzelne Punkte definieren, da ein Integral über ein Intervall der Länge 0 immer 0 ist. Hier hilft ein einfacher Trick: wir geben die W'keiten statt mit einer Dichte mit der Stammfunktion

$$F_X(r) = P(-\infty < X \leq r) \quad \Big(= \int_{-\infty}^{r} \rho(r)\, dr \Big)$$

der Dichte an. Also: wir versehen den Stichprobenraum $\mathbb{R}$ mit einem W'keitsmaß, indem wir die W'keiten $P(-\infty < X \leq r)$ vorgeben. Damit lassen sich diskrete Modelle (ohne Dichten) und kontinuierliche Modelle (mit Dichten) einheitlich behandeln.

2.1 Definitionen

Definition 2.1: (Zufallsvariable)

Sei $(\Omega, \mathcal{E}, P)$ ein beliebiges stochastisches Modell.

i) Eine (reelle) **Zufallsvariable** *ist eine Abbildung $X : \Omega \mapsto \mathbb{R}$, für welche die Urbilder aller Intervalle der Form $(-\infty, r]$ Ereignisse sind:*

$$X^{-1}((-\infty, r]) = \{\omega \in \Omega;\ X(\omega) \leq r\} \in \mathcal{E} \ .$$

ii) Die Abbildung

$$r \ \mapsto \ F_X(r) = P(X^{-1}((-\infty, r])) \equiv P(X \leq r)$$

heißt **(kumulative) Verteilungsfunktion** *von X (engl: cumulative distribution function = CDF)*

iii) X heißt **diskret***, wenn die Bildmenge $X(\Omega)$ abzählbar ist. X heißt* **kontinuierlich***, wenn $F_X(r)$ nach r differenzierbar ist. Die Ableitung $\rho(r) = F_X'(r)$ nennt man die* **Dichte** *der Verteilungsfunktion F_X (engl: probability density function = PDF)*

Bemerkung 2.2: *Ist das Modell $(\Omega, \mathcal{E}, P)$ diskret (speziell, kombinatorisch), so ist wegen $\mathcal{E} = \mathcal{P}(\Omega)$ = Potenzmenge von Ω trivialerweise jede Abbildung $X : \Omega \mapsto \mathbb{R}$ eine Zufallsvariable.*

Beispiel 2.3: Betrachten wir noch einmal Beispiel 1.16: werfe 2 mal mit einem fairen Würfel, betrachte die Augensumme X:

$$\Omega = \{(1,1),(1,2),\ldots,(6,5),(6,6)\}, \quad X : (\omega_1,\omega_2) \in \Omega \ \rightarrow \ \omega_1 + \omega_2.$$

Das Bild $X(\Omega) = \{2,3,\ldots,12\}$ ist eine diskrete Teilmenge von $\mathbb{R}$. Wir hatten in Beispiel 1.16 bereits einige der W'keiten $p_k = P(X = k) \equiv P(X^{-1}(\{k\}))$ berechnet, wir vervollständigen hier die Rechnung:

$$\begin{aligned}
p_2 &= P(X=2) = P(\{(1,1)\}) = \tfrac{1}{36}, \\
p_3 &= P(X=3) = P(\{(1,2),(2,1)\}) = \tfrac{2}{36} = \tfrac{1}{18}, \\
p_4 &= P(X=4) = P(\{(1,3),(2,2),(3,1)\}) = \tfrac{3}{36} = \tfrac{1}{12}, \\
p_5 &= P(X=5) = P(\{(1,4),(2,3),\ldots\}) = \tfrac{1}{9}, \\
p_6 &= P(X=6) = \ldots = \tfrac{5}{36}, \\
p_7 &= P(X=7) = \ldots = \tfrac{1}{6}, \\
p_8 &= p_6 = \tfrac{5}{36}, \\
p_9 &= p_5 = \tfrac{1}{9}, \\
p_{10} &= p_4 = \tfrac{1}{12}, \\
p_{11} &= p_3 = \tfrac{1}{18}, \\
p_{12} &= p_2 = \tfrac{1}{36}.
\end{aligned}$$

Offensichtlich gilt für beliebiges $r \in \mathbb{R}$:

$$\begin{aligned}
\boxed{F_X(r)} &= P(X \le r) \equiv P(X^{-1}((-\infty, r]) = P(X^{-1}((-\infty, r] \cap \{2,3,\ldots,12\})) \\
&= \sum_{\substack{k \in \{2,\ldots,12\} \\ k \le r}} P(X^{-1}(\{k\})) = \boxed{\sum_{\substack{k \in \{2,\ldots,12\} \\ k \le r}} P(X = k).}
\end{aligned}$$

Beispielsweise ergibt sich damit für jedes $-\infty < r < 2$ derselbe Wert

$$F_X(r) = P(X \le r) = 0.$$

Für jedes $2 \le r < 3$ ergibt sich derselbe Wert

$$F_X(r) = P(X \le r) = P(X^{-1}(\{2\})) = p_2 = \frac{1}{36}.$$

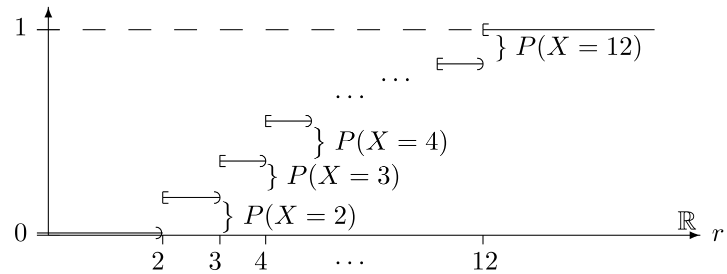
Für jedes $3 \le r < 4$ ergibt sich derselbe Wert

$$F_X(r) = P(X \le r) = P(X^{-1}(\{2,3\})) = p_2 + p_3 = \frac{1}{36} + \frac{1}{18} = \frac{1}{12}.$$

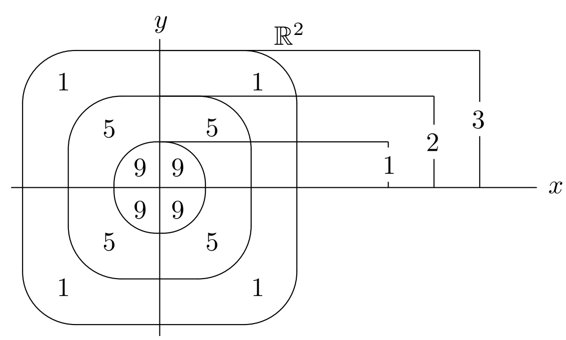
Usw. Für jedes $12 \le r < \infty$ ergibt sich derselbe Wert

$$F_X(r) = P(X \le r) = P(X^{-1}(\{2,\ldots,12\})) = p_2 + \cdots + p_{12} = 1.$$

Ergebnis: $F_X(r)$ ist eine Treppenfunktion, die bei $r = -\infty$ mit 0 startet und monoton bis 1 anwächst. Die Sprungstellen sind dort, wo X Werte annimmt. Die Sprunghöhen (der Zuwachs, wenn man eine Stelle $X = k$ überschreitet) sind jeweils $p_k = P(X = k)$:



Beispiel 2.4: (Vergleiche auch mit Beispiel 1.22) Ein Schütze schießt auf folgende im Ursprung des $\mathbb{R}^2$ zentrierte Zielscheibe vom Radius 3 mit 3 kreisförmigen Ringen mit den Punktzahlen 1, 5, 9:



Wir geben vor, dass ein Treffer mit der W'keitsdichte

$$\rho(t) = 2\,t\,e^{-t^2}$$

im Abstand t vom Ursprung landet. Betrachte $X =$ „Abstand des Treffers vom Zentrum“

$$X: \ t \in \Omega = [0, \infty) \quad \to \quad t \in \mathbb{R}$$

mit der Verteilungsfunktion

$$F_X(r) = P(X \le r) = \left\{ \begin{array}{cl} 0 & \text{für } r < 0, \\ \displaystyle\int_0^r 2\,t\,e^{-t^2}\,dt & \text{für } 0 \le r \end{array} \right\} = \left\{ \begin{array}{cl} 0 & \text{für } r < 0, \\ 1 - e^{-r^2} & \text{für } 0 \le r. \end{array} \right.$$

Ergebnis: Die Verteilungsfunktion $F_X(r)$ ist monoton steigend: sie beginnt bei $r = -\infty$ mit 0, bleibt konstant 0 bis $r = 0$, dann steigt sie monton in der Form $1 - e^{-r^2}$ bis 1 an.

Beispiel 2.5: Betrachte erneut die Zielscheibe des letzten Beispiels mit den Zufallsvariablen X = „Abstand des Treffers vom Zentrum“ und Y = „die geschossene Punktzahl“:

$$Y : t \in \Omega = [0, \infty) \to \begin{cases} 0 & \text{für } 3 < t, \\ 1 & \text{für } 2 < t \leq 3, \\ 5 & \text{für } 1 < t \leq 2, \\ 9 & \text{für } 0 \leq t \leq 1. \end{cases}$$

Obwohl das unterliegende Modell $\Omega = [0, \infty)$ kontinuierlich ist, ist die Zufallsvariable diskret. Die Variable nimmt nur 4 mögliche Wert an: $Y(\Omega) = \{0, 1, 5, 9\}$. Die Wahrscheinlichkeiten für die geschossenen Punktzahlen sind

$$\begin{aligned}
p_0 = P(Y = 0) = {}& P(3 < X) &= \textstyle\int_3^\infty \rho(t)\,dt &= \Big[-e^{-t^2}\Big]_{t=3}^{t=\infty} &= e^{-9} - 0 &\approx 0.000123..., \\
p_1 = P(Y = 1) = {}& P(2 < X \leq 3) &= \textstyle\int_2^3 \rho(t)\,dt &= \Big[-e^{-t^2}\Big]_{t=2}^{t=3} &= e^{-4} - e^{-9} &\approx 0.01819..., \\
p_5 = P(Y = 5) = {}& P(1 < X \leq 2) &= \textstyle\int_1^2 \rho(t)\,dt &= \Big[-e^{-t^2}\Big]_{t=1}^{t=2} &= e^{-1} - e^{-4} &\approx 0.3495..., \\
p_9 = P(Y = 9) = {}& P(0 \leq X \leq 1) &= \textstyle\int_0^1 \rho(t)\,dt &= \Big[-e^{-t^2}\Big]_{t=0}^{t=1} &= 1 - e^{-1} &\approx 0.6321... \\
&&&&& \overline{\ 1.0000}
\end{aligned}$$

Die Verteilungsfunktion von Y ist wieder eine Treppenfunktion:

$$F_Y(r) = \begin{cases} 0 & \text{für } r < 0, \\ p_0 & \text{für } 0 \leq r < 1, \\ p_0 + p_1 & \text{für } 1 \leq r < 5, \\ p_0 + p_1 + p_5 & \text{für } 5 \leq r < 9, \\ p_0 + p_1 + p_5 + p_9 = 1 & \text{für } 9 \leq r. \end{cases}$$

Notation 2.6:

Betrachte eine Zufallsvariable $X : \Omega \mapsto \mathbb{R}$ über einem beliebigen Modell $(\Omega, \mathcal{E}, P)$. Für $A \subset \mathbb{R}$ benutzen wir die vereinfachte Notation

$$P(X \in A) \equiv P(X^{-1}(A)) = P(\{\omega \in \Omega;\ X(\omega) \in A\}) .$$

Beispiele:

$$\begin{aligned}
P(X \leq r) &\equiv P(X \in (-\infty, r]) &\equiv P(X^{-1}((-\infty, r])) , \\
P(r_1 < X \leq r_2) &\equiv P(X \in (r_2, r_2]) &\equiv P(X^{-1}((r_1, r_2])) .
\end{aligned}$$

Rechenregeln 2.7:

Einige wichtige Rechenregeln in dieser Notation sind:

a) *für disjunkte Mengen $A_1, A_2 \subset \mathbb{R}$, $A_1 \cap A_2 = \emptyset$, gilt*

$$P(X \in A_1 \cup A_2) = P(X \in A_1) + P(X \in A_2)\ .$$

Z.B., für $r_1 \le r_2$, $[r_1, r_2] = \{r_1\} \cup (r_1, r_2] = A_1 \cup A_2$:

$$P(r_1 \le X \le r_2) = P(X = r_1) + P(r_1 < X \le r_2)\ .$$

(beachte Folgerung 1.6: die Urbilder $X^{-1}(A_1)$ und $X^{-1}(A_2)$ sind wieder disjunkt und es gilt $X^{-1}(A_1) \cup X^{-1}(A_2) = X^{-1}(A_1 \cup A_2)$.)

b) *Für $A_1 \subset A_2 \subset \mathbb{R}$ gilt die disjunkte Zerlegung $A_2 = A_1 \cup (A_2 \setminus A_1)$ und damit*

$$P(A_2 \setminus A_1) = P(A_2) - P(A_1)\ .$$

Z.B., für $r_1 < r_2$, $A_1 = (-\infty, r_1] \subset A_2 = (-\infty, r_2]$:

$$P(r_1 < X \le r_2) = P(X \le r_2) - P(X \le r_1)\ .$$

Folgerung 2.8:

Für $r_1 < r_2$ gilt: $\boxed{P(r_1 < X \le r_2) = F_X(r_2) - F_X(r_1)\ .}$

Also: W'keiten auf Intervallen (zunächst nur halboffene Intervalle) sind aus F_X konstruierbar! In der Tat kann aus F_X ein W'keitsmaß auf allen („vernünftigen") Teilmengen in $\mathbb{R}$ konstruiert werden.

14.5.07↓

Der nächste Satz beschreibt einige Eigenschaften, die jede Verteilungsfunktion über einem beliebigen Modell hat:

Satz 2.9: (einige technische Eigenschaften von Verteilungsfunktionen)

1) *Es gilt $0 \le F_X(r) \le 1$ für alle $r \in \mathbb{R}$.*
2) *$F_X(r)$ ist monoton wachsend: $F_X(r_1) \le F_X(r_2)$ für $r_1 < r_2$.*
3) $\lim\limits_{r \to -\infty} F_X(r) = 0$, $\lim\limits_{r \to \infty} F_X(r) = 1$.
4) *$F_X(r)$ ist rechtsseitig stetig: $F_X(r) = \lim\limits_{\substack{\epsilon \to 0 \\ \epsilon > 0}} F_X(r + \epsilon)$.*
5) *$F_X(r)$ hat höchstens abzählbar viele Unstetigkeitsstellen, diese sind Sprungstellen. An einer Sprungstelle r gilt für die Sprunghöhe:*

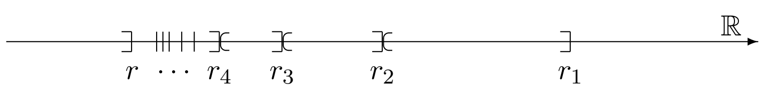
$$F_X(r) - \lim_{\substack{\epsilon \to 0 \\ \epsilon > 0}} F_X(r - \epsilon) = P(X = r)\ .$$

Beweis:
1) Klar, da $F_X(r)$ eine W'keit ist.
2) Klar, da $\{\omega \in \Omega;\ X(\omega) \leq r_1\} \subset \{\omega \in \Omega;\ X(\omega) \leq r_2\}$ für $r_1 < r_2$ und die W'keiten nach Folgerung 1.6.c) monoton wachsen, wenn die Mengen größer werden.
3) Betrachte die disjunkten Ereignisse $E_0 = \{\omega \in \Omega; -\infty < X(\omega) \leq 0\}$ und $E_i = \{\omega \in \Omega; i-1 < X(\omega) \leq i\}$ für $i = 1, 2, \ldots$. Es gilt $F_X(n) = P(\bigcup_{i=0}^n E_i) = \sum_{i=0}^n P(E_i)$. Für unendliche abzählbare Vereinigungen disjunkter Ereignisse gilt mit der „σ-Additivität“ in Definition 1.2.iii.2):

$$P(\bigcup_{i=0}^{\infty} E_i) = \sum_{i=0}^{\infty} P(E_i) = \lim_{n\to\infty} \sum_{i=0}^{n} P(E_i) = \lim_{n\to\infty} F_X(n) \ .$$

Es gilt $\bigcup_{i=0}^{\infty} E_i = \Omega$ und $P(\Omega) = 1$. Da F_X monoton ist, folgt der Grenzwert auch für beliebiges nichtganzzahliges $n \in \mathbb{R}$. Mit einem ähnlichen Argument zeigt man $\lim_{r\to-\infty} F_X(r) = 0$, indem man die leere Menge als Schnitt der Mengen $X^{-1}((-\infty, i])$ (mit negativem i) darstellt.
4) Sei (r_i) eine beliebige streng monoton fallende, gegen r konvergierende Folge. Es gilt:

$$X^{-1}((-\infty, r_n]) = X^{-1}((-\infty, r]) \cup \bigcup_{i=n}^{\infty} X^{-1}((r_{i+1}, r_i]) \ .$$

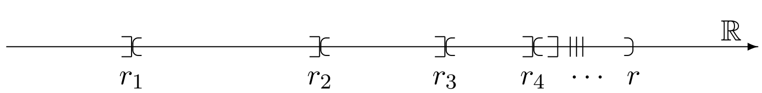


Mit der σ-Additivität folgt hieraus

$$\underbrace{P(X \leq r_n)}_{F_X(r_n)} = \underbrace{P(X \leq r)}_{F_X(r)} + \sum_{i=n}^{\infty} P(X^{-1}((r_{i+1}, r_i])).$$

Für jede konvergierende Reihe $\sum_{i=n}^{\infty} p_i$ gilt aber $\lim_{n\to\infty} \sum_{i=n}^{\infty} p_i = 0$. Damit folgt $\lim_{n\to\infty} F_X(r_n) = F_X(r)$.
5) Sei (r_i) eine beliebige streng monoton steigende, gegen r konvergierende Folge. Es gilt:

$$X^{-1}((-\infty, r)) = X^{-1}((-\infty, r_1]) \cup \bigcup_{i=1}^{\infty} X^{-1}((r_i, r_{i+1}]) \ .$$



Mit der σ-Additivität folgt hieraus

$$P(X < r) = \underbrace{P(X \le r_1)}_{F_X(r_1)} + \sum_{i=1}^{\infty} \underbrace{P(r_i < X \le r_{i+1})}_{F_X(r_{i+1}) - F_X(r_i)} .$$

Mit $\sum\limits_{i=1}^{\infty} .. = \lim\limits_{n\to\infty} \sum\limits_{i=1}^{n} ..$ folgt

$$\lim_{n\to\infty} \underbrace{F_X(r_1) + \sum_{i=1}^{n-1} \Big(F_X(r_{i+1}) - F_X(r_i) \Big)}_{F_X(r_n)} \; = \; P(X < r) \; ,$$

also $\lim_{n\to\infty} F_X(r_n) = P(X < r)$. Es folgt die Behauptung, dass die Sprunghöhe einer Unstetigkeitsstelle r die Interpretation $P(X = r)$ hat:

$$P(X = r) = P(X \le r) - P(X < r) = F_X(r) - \lim_{n\to\infty} F_X(r_n) \; .$$

Es kann höchstens abzählbar viele solcher Sprünge geben. Sei dazu $S \subset \mathbb{R}$ die Menge aller Unstetigkeitspunkte von F_X. Sei

$$S_n = \left\{ r \in \mathbb{R};\; F_X(r) - \lim_{\substack{\epsilon\to 0\\ \epsilon > 0}} F_X(r - \epsilon) \ge \frac{1}{n} \right\} \subset S$$

die Menge aller Unstetigkeitsstellen mit einer Sprunghöhe $\ge 1/n$. Wegen der Monotonie von F_X kann diese Menge maximal n Elemente enthalten, da $F_X(r) \in [0,1]$. Mit $S = S_1 \cup S_2 \cup \cdots$ ist S eine abzählbare Vereinigung endlicher Mengen und damit abzählbar.

Q.E.D.

$\boxed{15.5.07\downarrow}$ **Interpretation 2.10:**

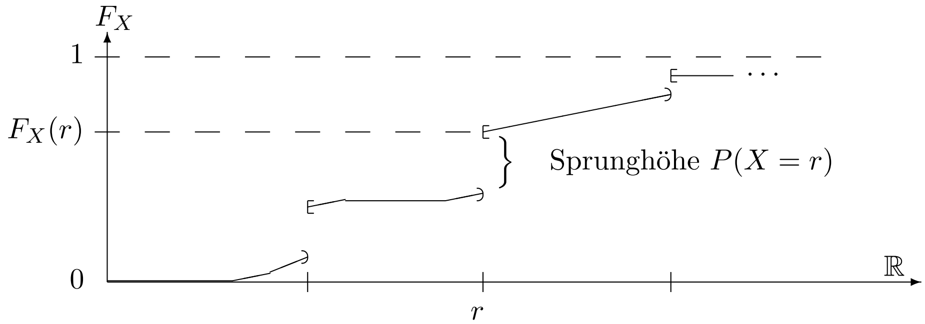
Anstatt sich die technischen Formulierungen des letzten Satzes einzuprägen, dürfte es einfacher sein, sich das folgende Bild zu merken. Der Satz besagt nur, dass dieses Bild typisch ist:

1-3) *F_X ist monoton (aber nicht unbedingt streng monoton) von 0 bis 1 wachsend.*

4) *Jeder Funktionswert stimmt mit dem Grenzwert von rechts überein (F_X ist „rechtsseitig stetig“).*

5) *Der Graph von F_X setzt sich aus Stetigkeitsintervallen und höchstens abzählbar vielen Sprungstellen zusammen.*

An Sprungstellen r ist die Sprunghöhe gleich $P(X = r)$.



In Verallgemeinerung von Folgerung 2.8 gilt:

Folgerung 2.11:

a) Für $r_1 < r_2$ gilt: $\boxed{P(r_1 < X \leq r_2) = F_X(r_2) - F_X(r_1)\ .}$

b) Für $r_1 \leq r_2$ gilt: $\boxed{P(r_1 \leq X \leq r_2) = F_X(r_2) - \lim_{\substack{\epsilon \to 0 \\ \epsilon > 0}} F_X(r_1 - \epsilon).}$

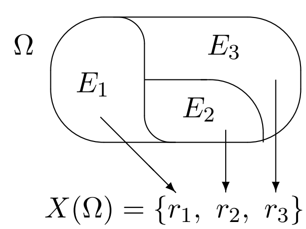
Beweis: a) war schon in Folgerung 2.8 gezeigt worden. b) folgt mit

$$P(r_1 \leq X \leq r_2) \;=\; P(X = r_1) + P(r_1 < X \leq r_2)$$

aus Teil 5) des Satzes 2.9 und a).

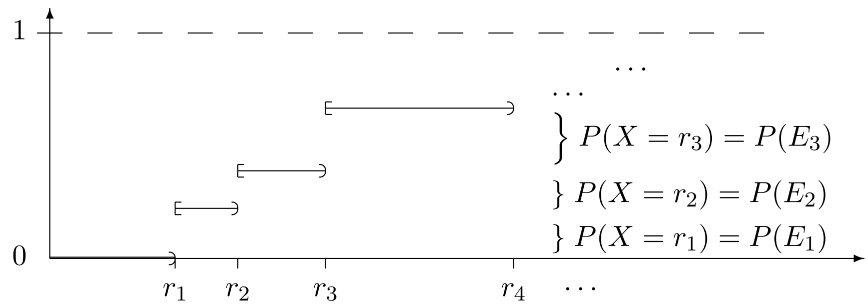
Q.E.D.

Bemerkung 2.12: *Für eine diskrete Variable $X(\Omega) = \{r_1, r_2, \dots\}$ zerlegt sich der Stichprobenraum $\Omega = \cup_i E_i$ in disjunkte Urbilder $E_i = X^{-1}(\{r_i\})$ der Werte, die X annimmt, z.B.:*

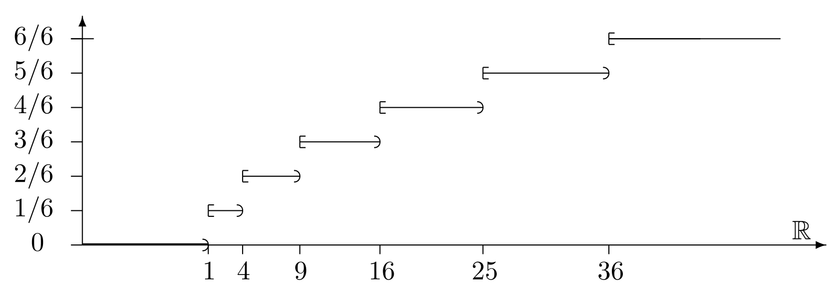


$\boxed{\text{\textit{Damit ist } } F_X \text{ \textit{für diskrete Variable immer eine Treppenfunktion:}}}$

$$F_X(r) = P(X \leq r) = \sum_{r_i \leq r} P(E_i) = \sum_{r_i \leq r} P(X = r_i)\ .$$



Beispiel 2.13: Ein Wurf mit einem fairen Würfel: $\Omega = \{1, \ldots, 6\}$. Betrachte $X : \omega \mapsto \omega^2$, also $X(\Omega) = \{1, 4, 9, 16, 25, 36\}$:



Zusammenfassung 2.14:

Es gibt mehrere Möglichkeiten, eine Zufallsvariable zu interpretieren:

i) Als Abbildung $X : \Omega \to \mathbb{R}$. Wenn man das unterliegende Experiment Ω kennt, hat man vollständiges Wissen über das zufällige Verhalten der Werte, die X annimmt.

ii) Man kann sich eine Zufallsvariable $X : \Omega \mapsto \mathbb{R}$ mit Verteilungsfunktion F_X auch selbst als ein stochastisches Modell $(\mathbb{R}, \mathcal{B}, P)$ vorstellen. Hierbei ist $\mathcal{B}$ die Menge aller Teilmengen von $\mathbb{R}$, über denen man eine Integrationstheorie aufbauen kann (die sogenannten „Borel"-Mengen). Speziell sind Intervalle $(r_1, r_2]$, $[r_1, r_2]$, $(-\infty, r]$, abzählbare Vereinigungen und Schnitte solcher Intervalle, die Komplemente etc. Ereignisse in $\mathcal{B}$. Die Angabe der Verteilungsfunktion F_X ist genausogut wie die Vorgabe der W'keiten auf den Ereignissen: das W'keitsmaß P auf den Teilmengen von $\mathbb{R}$ ist aus F_X konstruierbar! Z.B. gilt für die „Basistypen" von Ereignissen in $\mathcal{B}$:

$$\boxed{\begin{array}{ll} a) & P((-\infty, r]) = F_X(r), \\ b) & P((r_1, r_2]) = F_X(r_2) - F_X(r_1), \\ c) & P([r_1, r_2]) = F_X(r_2) - \lim\limits_{\substack{\epsilon \to 0 \\ \epsilon > 0}} F_X(r_1 - \epsilon), \\ d) & P(\{r\}) = P([r, r]) = F_X(r) - \lim\limits_{\substack{\epsilon \to 0 \\ \epsilon > 0}} F_X(r - \epsilon). \end{array}}$$

Hierbei ist d) der Spezialfall $r = r_1 = r_2$ von c). Für disjunkte Vereinigungen, Komplemente etc. von Mengen $E, E_i \in \mathcal{B}$ ergeben sich die W'keiten dann durch $P(\cup_i E_i) = \sum_i P(E_i)$, $P(\mathbb{R} \setminus E) = 1 - P(E)$ etc.

> *Mit den obigen Regeln kann $P(E)$ für alle „interessanten“ Teilmengen E von $\mathbb{R}$ aus F_X konstruiert werden.*

Beispielsweise:

$$\boxed{\begin{array}{l} P((r_1, r_2)) = \lim\limits_{\substack{\epsilon \to 0 \\ \epsilon > 0}} F_X(r_2 - \epsilon) - F_X(r_1), \\ P([r_1, r_2)) = \lim\limits_{\substack{\epsilon \to 0 \\ \epsilon > 0}} F_X(r_2 - \epsilon) - \lim\limits_{\substack{\epsilon \to 0 \\ \epsilon > 0}} F_X(r_1 - \epsilon), \end{array}}$$

In der Sichtweise ii) kann man das unterliegende Experiment Ω eigentlich vergessen: es diente nur dazu, die W'keiten vom Raum Ω auf die reelle Achse hinüberzuziehen und als Verteilungsfunktion F_X zu kodieren. Dies ergibt ein W'maß auf $\mathbb{R}$, welches sich – wie oben beschrieben – auf den geschlossenen/offenen/halboffenen Intervallen einfach ergibt. Man hat in dieser Sichtweise aber gewisse Informationen verloren, wenn man das unterliegende Ω vergisst. Zwar kann man alle gewünschten Aussagen über X machen, aber man kann nicht mehr mehrere Zufallsvariablen X_1, X_2 über demselben Ω vergleichen (z.B., weiss man nicht, wie sich die Summe $X_1 + X_2$ verhält, wenn man nur F_{X_1} und F_{X_2} kennt).

Bemerkung 2.15: *Im wichtigen Spezialfall diskreter Zufallsvariabler mit $X(\Omega) = \{r_1, r_2, \dots\} \subset \mathbb{R}$ ist die Interpretation als Zufallsexperiment auf dem Stichprobenraum $\mathbb{R}$ mit einem durch F_X (Treppenfunktion) gegebenen W'maß ziemlich gekünstelt und unnötig kompliziert. Alle Informationen stecken in den Sprunghöhen der Unstetigkeitsstellen, und es ist technisch wesentlich bequemer, sich die Zufallsvariable direkt als diskretes Modell $(\Omega_X, \mathcal{E}_X, P_X)$ mit $\Omega_X = X(\Omega) = \{r_1, r_2, \dots\}$ und $\mathcal{E}_X = \mathcal{P}(\Omega_X)$ vorzustellen, wobei das W'maß durch die Elementarw'keiten*

$$P_X(\{r_i\}) = P(X = r_i) = P(X^{-1}(\{r_i\}))$$

eindeutig festgelegt ist.

Der Vollständigkeit halber soll noch ein in der Literatur vielfach verwendeter Begriff erwähnt werden, der speziell in der Statistik z.B. beim Testen von Hypothesen eine wichtige Rolle spielt: die Quantilfunktion. Sie ist die Inverse der (kumulativen) Verteilungsfunktion F_X. Für nicht streng monotone Verteilungsfunktionen, die im mathematischen Sinne nicht invertierbar sind, vereinbart man:

Definition 2.16: (Die Quantilfunktion einer Verteilung)

Sei F_X die Verteilungsfunktion einer Zufallsvariable X. Die **Quantilfunktion** *der Verteilung ist:*

$$Q_X : \; p \in (0,1] \quad \rightarrow \quad \min\{r; \; p \leq F_X(r)\} \in \mathbb{R} \; .$$

Da F_X rechtsseitig stetig ist, existiert des Minimum für jedes $p \in (0,1)$ (für $p = 1$ setzt man $Q_X(1) = \infty$, falls $F_X(r) < 1$ für jedes endliche $r \in \mathbb{R}$ gilt).

Quantilfunktionen haben allgemein folgende Eigenschaften (Übungsaufgabe):

a) Sie sind monoton steigend in p.

b) Sie sind linksseitig stetig.

c) An allen Stellen gilt: $F_X(Q_X(p)) \geq p$ und $Q_X(F_X(r)) \leq r$.

d) An allen Stellen $r = Q_X(p)$, wo $F_X(r)$ beidseitig stetig ist, gilt $F_X(Q_X(p)) = p$.

e) An allen Stellen, wo $F_X(r)$ streng monoton ist, gilt $Q_X(F_X(r)) = r$.

In diesem Sinne ist Q_X die Inverse von F_X.

Bemerkung 2.17: *Im Computeralgebrasystem MuPAD existiert seit Version 2.5 eine umfangreiche Bibliothek* `stats`*, in der etliche Standardverteilungen mit ihren (kumulativen) Verteilungsfunktionen* `stats::nameCDF` *und den Quantilen* `stats::nameQuantile` *installiert sind. Weiterhin gibt es die (kontinuierlichen) Dichten* `stats::namePDF` *bzw. die diskreten W'keitswerte* `stats::namePF` *sowie Zufallsgeneratoren* `stats::nameRandom`.

Zahlenwerte der Binomialverteilung (Beispiel 1.64 auf Seite 36 sowie Seite 71) mit den Parametern n („Anzahl der Wiederholungen“) und p („Erfolgsw'keit“) sind beispielsweise folgendermaßen zu berechnen:

```
>> n:= 10: p:= 1/2:
>> F:= stats::binomialCDF(n, p):
>> F(0), F(1), F(2.5), F(3)

              1/1024, 11/1024, 0.0546875, 11/64

>> Q:= stats::binomialQuantile(n, p):
>> Q(0), Q(1/4), Q(0.3), Q(9/10), Q(1)

                       0, 4, 4, 7, 10
```

Weiterhin stehen z.B. die hypergeometrische Verteilung (Beispiel 1.10 auf Seite 7 und Seite 75) als

`stats::hypergeometicCDF`, `stats::hypergeometricQuantile`,

`stats::hypergeometicPF`, `stats::hypergeometricRandom`

oder die später ab Seite 78 diskutierte Normalverteilung als

`stats::normalCDF`, `stats::normalQuantile`,

`stats::normalPDF`, `stats::normalRandom`

etc. zur Verfügung.

2.2 Das Riemann-Stieltjes-Integral

Wir verallgemeinern den bekannten Begriff der Riemann-Integration, um kontinuierliche Zufallsvariablen (mit Dichten) und diskrete Zufallsvariablen (ohne Dichten, die Verteilungsfunktion ist eine Treppenfunktion) gemeinsam behandeln zu können:

Definition 2.18:

Sei $F : [a, b] \subset \mathbb{R} \mapsto \mathbb{R}$ eine monoton wachsende, rechtsseitig stetige, beschränkte Funktion. Zerlege das Interval $[a, b]$ in $n + 1$ Zwischenpunkte $r_i = a + i\,\frac{b-a}{n}$, $i = 0, \ldots, n$, und definiere für eine Funktion $f : [a, b] \to \mathbb{R}$ die **Riemann-Stieltjes-Summe** *(bzgl. F)*

$$R_n(f) = \sum_{i=0}^{n-1} f(r_i)\,\Big(F(r_{i+1}) - F(r_i)\Big)\ .$$

Wenn der Grenzwert existiert, definiert man als **Riemann-Stieltjes-Integral** *(bzgl. F)*

$$\int_{(a,b]} f(r)\ dF(r) = \lim_{n\to\infty} R_n(f)\ .$$

In Analogie zur Riemann-Integration kann man sich unschwer überlegen, dass zumindestens für (stückweise) stetiges f dieses Integral existiert.

Bemerkung 2.19: *Für $F(r) = r$ ergibt sich das übliche Riemann-Integral.*

Bemerkung 2.20: *Wie beim üblichen Riemann-Integral definiert man uneigentliche Stieltjes-Integrale (über unendlichen Intervallen) über*

$$\int_{(-\infty,b]} f(r)\ dF(r) = \lim_{a\to-\infty} \int_{(a,b]} f(r)\ dF(r) \ ,$$
$$\int_{\mathbb{R}} f(r)\ dF(r) = \lim_{b\to\infty} \int_{(-\infty,b]} f(r)\ dF(r) \ .$$

Satz 2.21: (einige Eigenschaften von Stieltjes-Integralen)

Es gilt:

1)

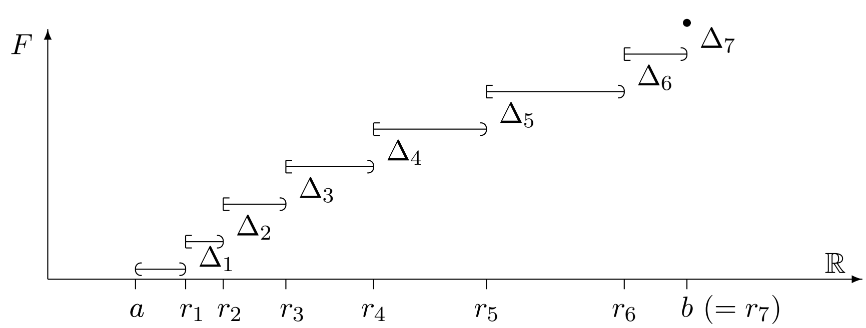
$$\int_{(a,b]} 1\ dF(r) = F(b) - F(a).$$

2) Falls F stetig differenzierbar ist:

$$\int_{(a,b]} f(r)\ dF(r) = \int_a^b f(r)\ F'(r)\ dr \ .$$

3) Ist F eine Treppenfunktion mit Sprungstellen $r_1, r_2, \dots \in (a,b]$ und Sprunghöhen

$$\Delta_i = F(r_i) - \lim_{\substack{\epsilon\to 0\\ \epsilon>0}} F(r_i - \epsilon) \ ,$$



so gilt für stetiges f:

$$\int_{(a,b]} f(r)\ dF(r) = \sum_i f(r_i)\ \Delta_i \ .$$

4) Ist f stetig differenzierbar, so gilt („partielle Integration“):

$$\int_{(a,b]} f(r)\ dF(r) = f(b)\, F(b) - f(a)\, F(a) - \int_a^b f'(r)\ F(r)\ dr\ .$$

Beweisskizze: $\boxed{\downarrow 21.5.07}$

i) Für $f(r) \equiv 1$ sind alle Stieltjes-Summen „Teleskopsummen“:

$$R_n(f) = \sum_{i=0}^{n-1} \Big(F(r_{i+1}) - F(r_i) \Big) = F(r_n) - F(r_0) = F(b) - F(a)\ .$$

2) Für stetig differenzierbares F gilt der Mittelwertsatz $F(r_{i+1}) - F(r_i) = F'(\xi_i)\,(r_{i+1} - r_i)$ mit einem Zwischenwert $\xi_i \in (r_i, r_{i+1})$. Die Stieltjes-Summe wird damit zu einer Riemann-Summe für den Integranden $f(r)\,F'(r)$:

$$R_n(f) = \sum_{i=0}^{n-1} f(r_i) \Big(F(r_{i+1}) - F(r_i) \Big) = \sum_{i=0}^{n-1} f(r_i)\, F'(\xi_i)\ (r_{i+1} - r_i)\ .$$

3) Analog zu Riemann-Integralen kann man sich leicht überlegen, dass für Stieltjes-Integrale

$$\int_{(a,b]} f(r)\ dF(r) = \int_{(a,c]} f(r)\ dF(r) + \int_{(c,b]} f(r)\ dF(r)$$

mit beliebigem $c \in (a,b)$ gilt. Seien $r_1 < r_2 < \dots$ die Sprungstellen von F. Mit der Setzung $r_0 = a$ folgt

$$\int_{(a,b]} f(r)\ dF(r) = \sum_i \int_{(r_{i-1},r_i]} f(r)\ dF(r)\ .$$

Da F auf $(r_{i-1}, r_i]$ konstant ist bis auf den Sprung am rechten Intervallende, besteht eine Stieltjes-Summe für $\int_{(r_{i-1},r_i]} f(r)\ dF(r)$ mit den Stützstellen $r_i^{(j)} = r_{i-1} + j\,(r_i - r_{i-1})/n$ nur aus einem einzigen Term:

$$R_n(f) = \sum_{j=0}^{n-1} f(r_i^{(j)}) \Big(F(r_i^{(j+1)}) - F(r_i^{(j)}) \Big) = f(r_i^{(n-1)}) \Big(\underbrace{F(r_i^{(n)})}_{F(r_i)} - \underbrace{F(r_i^{(n-1)})}_{F(r_{i-1})} \Big)$$

$$= f(r_i^{(n-1)})\,\Delta_i\ .$$

Mit $n \to \infty$, $r_i^{(n-1)} = r_i - (b-a)/n \to r_i$ und der Stetigkeit von f folgt

$$\int_{(r_{i-1},r_i]} f(r)\ dF(r) = f(r_i)\ \Delta_i\ .$$

4) Durch Umsummation ergibt sich die Darstellung

$$R_n(f) = \sum_{i=0}^{n-1} f(r_i) \Big(F(r_{i+1}) - F(r_i) \Big)$$

$$= f(r_n)\, F(r_n) - f(r_0)\, F(r_0) - \sum_{i=1}^{n} \Big(f(r_i) - f(r_{i-1}) \Big) F(r_i) \ ,$$

wobei $f(r_n)\, F(r_n) = f(b)\, F(b)$, $f(r_0)\, F(r_0) = f(a)\, F(a)$. Analog zu 2) lässt sich die verbleibende Summe als übliche Riemann-Summe für den Integranden $f'(r)\, F(r)$ interpretieren, die gegen das übliche Riemann-Integral $\int_a^b F(r)\, f'(r)\, dr$ konvergiert.

Q.E.D.

2.3 Erwartungswert und Streuung

Der Zusammenfassung 2.14 folgend können wir uns nun mit Hilfe des Stieltjes-Integrals eine Zufallsvariable X mit Verteilung F_X als ein Modell $(\mathbb{R}, \mathcal{E}, P)$ vorstellen, wo die Ereignisse („Borel-Mengen“) aus Intervallen zusammengesetzt sind und das W'keitsmaß P für ein Ereignis $E \subset \mathbb{R}$ durch

$$P(X \in E) = \int_{r \in E} dF_X(r)$$

definiert ist und hierdurch über die kumultative Verteilungsfunktion $F_X(r)$ kodiert ist. Speziell gilt nämlich für $E = (r_1, r_2]$ nach Satz 2.21.1) $P(r_1 < X \leq r_2]) = F_X(r_2) - F_X(r_1)$, was nach Folgerung 2.8 genau der Definition von W'keiten für Zufallsvariable entspricht:

- Für eine diskrete Zufallsvariable $X(\Omega) = \{r_1, r_2, \dots\}$ ergibt sich das W'maß
$$P(X \in E) = \sum_{r_i \in E} P(X = r_i)$$
gemäß der Definition 1.14 diskreter Modelle, sobald man $P(X = r_i)$ kennt.

- Für eine kontinuierliche Verteilungsfunktion mit einer Dichte $\rho(r) = F_X'(r)$ ergibt sich die Vorgabe von W'keiten
$$P(X \in (r_1, r_2]) = \int_{r_1}^{r_2} F_X'(r)\ dr = \int_{r_1}^{r_2} \rho(r)\ dr$$
als Integrale über Dichten gemäß der Definition 1.18 kontinuierlicher Modelle.

Was bringt uns nun die Tatsache, dass wir es hier mit Zufallswerten in $\mathbb{R}$ zu tun haben? Wir können Arithmetik betreiben! Speziell heißt dies, dass man Erwartungswerte definieren kann.
Die Stieltjes-Integrale fassen diskrete und kontinuierliche Modelle zusammen. Wir können damit für beliebige Modelle einheitlich definieren:

Definition 2.22: (Erwartungswert)

Sei $X : \Omega \mapsto \mathbb{R}$ eine Zufallsvariable mit Verteilungsfunktion F_X. Dann nennt man

$$\mathrm{E}\,(X) = \int_{-\infty}^{\infty} r\ dF_X(r)$$

den **Erwartungswert** *von X.*

Sprechweise 2.23: $\boxed{\downarrow 23.5.07}$

Fassen wir X als Zufallsexperiment $(\mathbb{R}, \mathcal{E}, P)$ auf, so nimmt X zufällige Werte in $\mathbb{R}$ an. Umgangssprachlich wird das Konzept des Erwartungswerts so formuliert:

„*X nimmt* im Mittel *den Wert $\mathrm{E}\,(X)$ an.*“

Im Rahmen dieser Vorlesung sind nur die folgenden Spezialfälle interessant (d.h., man braucht Stieltjes-Integrale nicht wirklich auszuwerten, sondern kann entweder einfach summieren oder ein übliches Riemann-Integral berechnen):

Spezialfälle 2.24:

a) Ist X diskret, also ist die Bildmenge $X(\Omega) = \{r_1, r_2, \ldots, \}$ abzählbar (F_X ist eine Treppenfunktion), so gilt mit den Sätzen 2.21.3) und 2.9.5):

$$\mathrm{E}\,(X) = \sum_{r \in X(\Omega)} r \cdot P(X = r)\ .$$

b) Ist X kontinuierlich mit der Dichte $\rho(r) = \frac{d}{dr}\ F_X(r)$, so gilt mit Satz 2.21.2):

$$\mathrm{E}\,(X) = \int_{-\infty}^{\infty} r \cdot \rho(r)\ dr.$$

Bemerkung 2.25: *Nach (Gegen-)Beispiel 2.5 muss für eine diskrete Zufallsvariable das unterliegende Modell Ω nicht notwendigerweise diskret sein. Umgekehrt gilt jedoch, dass jede Zufallsvariable über einem diskreten $\Omega = \{\omega_1, \omega_2, \ldots\}$ diskret ist. In diesem Fall ist die obige Definition äquivalent zu*

$$\mathrm{E}\,(X) = \sum_{\omega \in \Omega} X(\omega)\ P(\{\omega\})\ .$$

Dies ist eine alternative Berechnungsformel für den Erwartungswert, die sich anbietet, wenn man das unterliegende Modell Ω kennt und die Verteilungsdaten $P(X = r)$ nicht explizit ausgerechnet hat.

Beweisskizze: Siehe Übungsaufgabe 45.

Beispiel 2.26: (diskret) Wurf mit zwei fairen Würfeln:

$$\Omega = \{(\omega_1, \omega_2);\ \omega_1, \omega_2 \in \{1, \ldots, 6\}\}.$$

Sei X die Augensumme $X : (\omega_1, \omega_2) \mapsto \omega_1 + \omega_2$. Nach Definition 2.22 braucht man zur Berechnung des Erwartungswerts die (nichtkombinatorischen) W'keiten $P(X = r)$ mit $r = 2, 3, \ldots, 12$. Diese waren in Beispiel 2.3 angegeben worden:

$$\begin{aligned} \mathrm{E}\,(X) \;&=\; \sum_{r \in X(\Omega)} r\ P(X = r) \;=\; \sum_{r=2}^{12} r\ P(X = r) \\ &= 2 \cdot \frac{1}{36} + 3 \cdot \frac{1}{18} + 4 \cdot \frac{1}{12} + \cdots + 12 \cdot \frac{1}{36} = 7. \end{aligned}$$

Nach Bemerkung 2.25 kann der Erwartungswert auch direkt durch Summation über das unterliegende kombinatorische Modell berechnet werden:

$$\mathrm{E}\,(X) = \sum_{\omega_1=1}^{6} \sum_{\omega_2=1}^{6} (\omega_1 + \omega_2)\ \underbrace{P(\{(\omega_1, \omega_2)\})}_{1/36}$$

$$= \frac{1}{36} \left(\Big(\sum_{\omega_1=1}^{6} \omega_1 \Big) \cdot \Big(\sum_{\omega_2=1}^{6} 1 \Big) + \Big(\sum_{\omega_2=1}^{6} 1 \Big) \cdot \Big(\sum_{\omega_2=2}^{6} \omega_2 \Big) \right)$$

$$= \frac{1}{36} \cdot 2 \cdot 6 \cdot \sum_{\omega=1}^{6} \omega = \frac{1}{36} \cdot 2 \cdot 6 \cdot \frac{6 \cdot 7}{2} = 7\ .$$

Beispiel 2.27: (kontinuierlich) Ein zum Zeitpunkt $t_0 = 0$ existierendes radioaktives Atom zerfällt mit der W'keit $\lambda\, e^{-\lambda\, t} dt$ im Zeitintervall $(t, t + dt)$, $t > 0$. D.h., die Dichte der Zerfallsw'keit ist

$$\rho(t) = \begin{cases} 0 & \text{für } t < 0\ , \\ \lambda\, e^{-\lambda\, t} & \text{für } t \geq 0\ . \end{cases}$$

Sei $T \in [0, \infty]$ der Zerfallszeitpunkt. Die „mittlere Lebensdauer“ ist der Erwartungswert

$$\mathrm{E}\,(T) = \int_{-\infty}^{\infty} t\ \rho(t)\ dt = \int_0^{\infty} \underbrace{t}_{u}\ \underbrace{\lambda\, e^{-\lambda\, t}}_{v'} dt =$$

$$\Big[\underbrace{t}_{u} \cdot \underbrace{(-e^{-\lambda\, t})}_{v} \Big]_{t=0}^{t=\infty} - \int_0^{\infty} \underbrace{1}_{u'}\ \underbrace{(-e^{-\lambda\, t})}_{v} dt = \int_0^{\infty} e^{-\lambda\, t} dt = \Big[-\frac{1}{\lambda}\ e^{-\lambda\, t} \Big]_{t=0}^{t=\infty} = \frac{1}{\lambda}\ .$$

Bemerkung 2.28: (**Warnung**) *Nicht jede Zufallsvariable hat einen Erwartungswert. Betrachte z.B. $X : \Omega \mapsto [0, \infty)$ mit der Dichte $\rho(r) = 0$ für $r < 0$ und $\rho(r) = \frac{2}{\pi} \frac{1}{1+r^2}$ für $r \geq 0$. Das Integral*

$$\mathrm{E}\,(X) = \int_0^\infty \frac{2}{\pi} \frac{r\ dr}{1+r^2} = \lim_{R\to\infty} \int_0^R \frac{2}{\pi} \frac{r\ dr}{1+r^2} = \lim_{R\to\infty} \frac{1}{\pi}\ \ln(1+R^2) = \infty$$

existiert nicht.

Wir betrachten nun eine (glatte) Funktion $f : \mathbb{R} \mapsto \mathbb{R}$. Für eine gegebene Zufallsvariable $X : \Omega \mapsto \mathbb{R}$ kann man

$$\tilde{X} = f(X) : \omega \in \Omega \mapsto f(X(\omega)) \in \mathbb{R}$$

als neue Zufallsvariable auffassen, zu der sich eine neue Verteilungsfunktion $F_{\tilde{X}}(r) = P(\tilde{X} \leq r) = P(f(X) \leq r) = P(X \in f^{-1}((-\infty, r]))$ bestimmen lässt, mit der dann der Erwartungswert $\mathrm{E}\,(\tilde{X})$ berechnet werden kann. Aber: wir brauchen $F_{\tilde{X}}$ gar nicht! Man kann den Erwartungswert von $\tilde{X}$ direkt mittels der Verteilung $F_X(r)$ bestimmen. Im folgenden Satz ist F_X im diskreten Fall durch die Sprunghöhen $P(X = r)$ bzw. im kontinuierlichen Fall durch die Ableitung $\rho(r) = F'_X(r)$ kodiert:

Satz 2.29: (Transformation von Zufallsvariablen)

Sei $f : \mathbb{R} \mapsto \mathbb{R}$ eine glatte Funktion.

a) Für eine diskrete Variable X mit $X(\Omega) = \{r_1, r_2, \dots\}$ gilt

$$\mathrm{E}\,(f(X)) = \sum_{r \in X(\Omega)} f(r)\ P(X = r)\ .$$

Ist das unterliegende Experiment $\Omega = \{\omega_1, \omega_2, \dots\}$ selbst schon diskret, so gilt mit Bemerkung 2.25 auch die alternative Formel:

$$\mathrm{E}\,(f(X)) = \sum_{\omega \in \Omega} f(X(\omega))\ P(\{\omega\})\ .$$

b) Für eine kontinuierliche Variable X mit der Dichte $\rho(r) = F'_X(r)$ gilt

$$\mathrm{E}\,(f(X)) = \int_{r \in X(\Omega)} f(r)\ \rho(r)\ dr.$$

Beweisskizze:
Sei $\tilde{X}$ die Zufallsvariable $f(X)$. Im diskreten Fall gilt mit $X(\Omega) = \{r_1, r_2, \dots\}$ für $\tilde{r} \in \tilde{X}(\Omega) = \{f(r_1), f(r_2), \dots\}$:

$$P(\tilde{X} = \tilde{r}) \;=\; P(X \in f^{-1}(\{\tilde{r}\})) = \sum_{\substack{r \in X(\Omega) \\ f(r) = \tilde{r}}} P(X = r)$$

und damit

$$\begin{aligned}
\mathrm{E}\,(\tilde{X}) \;&=\; \sum_{\tilde{r} \in \tilde{X}(\Omega)} \tilde{r}\; P(\tilde{X} = \tilde{r}) \;=\; \sum_{\tilde{r} \in \tilde{X}(\Omega)} \tilde{r} \sum_{\substack{r \in X(\Omega) \\ f(r) = \tilde{r}}} P(X = r) \\
&= \sum_{\tilde{r} \in \tilde{X}(\Omega)} \sum_{\substack{r \in X(\Omega) \\ f(r) = \tilde{r}}} \tilde{r}\; P(X = r) \;=\; \sum_{\tilde{r} \in \tilde{X}(\Omega)} \sum_{\substack{r \in X(\Omega) \\ f(r) = \tilde{r}}} f(r)\; P(X = r) \\
&= \sum_{r \in X(\Omega)} f(r)\; P(X = r) \;=\; \mathrm{E}\,(f(X))\ .
\end{aligned}$$

Im kontinuierlichen Fall führen wir den Beweis nur für *streng monoton wachsendes* f (der allgemeine Fall ist mit den hier zur Verfügung stehenden Mitteln technisch zu aufwendig):

$$\begin{aligned}
\mathrm{E}\,(\tilde{X}) &= \int_{-\infty}^{\infty} \tilde{r}\; dF_{\tilde{X}}(\tilde{r}) = \int_{-\infty}^{\infty} \tilde{r}\; \frac{dF_{\tilde{X}}(\tilde{r})}{d\tilde{r}}\; d\tilde{r} \\
&\overset{\text{(Substitution: } \tilde{r} = f(r))}{=} \int_{-\infty}^{\infty} f(r)\; \frac{dF_{\tilde{X}}(f(r))}{d\tilde{r}}\; f'(r)\; dr \\
&\overset{\text{(Kettenregel)}}{=} \int_{-\infty}^{\infty} f(r)\; \frac{dF_{\tilde{X}}(f(r))}{dr}\; dr\ .
\end{aligned}$$

Mit

$$F_{\tilde{X}}(f(r)) = P(\tilde{X} \le f(r)) = P(f(X) \le f(r)) = P(X \le r) = F_X(r)$$

folgt

$$\mathrm{E}\,(\tilde{X}) = \int_{-\infty}^{\infty} f(r)\; \frac{dF_X(r)}{dr}\; dr = \int_{-\infty}^{\infty} f(r)\; dF_X(r) = \mathrm{E}\,(f(X))\ .$$

Q.E.D.

30.5.07↓ **Beispiel 2.30:** Betrachte einen Zufallszahlengenerator X, der (Gleitpunkt-)Zahlen zwischen -1 und 1 auswirft. Er sei gleichverteilt, d.h., für $-1 \leq r_1 \leq r_2 \leq 1$ gelte

$$P(r_1 < X \leq r_2) = \int_{r_1}^{r_2} \rho(r)\, dr = \frac{r_2 - r_1}{2}$$

mit der Dichte

$$\rho(r) = \begin{cases} \frac{1}{2} & \text{für } r \in [-1,1], \\ 0 & \text{für } r \notin [-1,1]. \end{cases}$$

Die Verteilungsfunktion ist

$$F_X(r) = \int_{-\infty}^{r} \rho(t)\, dt = \begin{cases} 0 & \text{für } r < -1, \\ \frac{r+1}{2} & \text{für } -1 \leq r \leq 1, \\ 1 & \text{für } r > 1. \end{cases}$$

Der Erwartungswert ist 0:

$$\mathrm{E}\,(X) = \int_{\infty}^{\infty} r\, \rho(r)\, dr = \int_{-1}^{1} \frac{r}{2}\, dr = 0.$$

Die Variable $\tilde{X} = X^2$ hat folgende Verteilung (für $\tilde{r} \geq 0$):

$$F_{\tilde{X}}(\tilde{r}) = P(X^2 \leq \tilde{r}) = P\Big(-\sqrt{\tilde{r}} \leq X \leq \sqrt{\tilde{r}}\Big) = P\Big(X \leq \sqrt{\tilde{r}}\,\Big) - P\Big(X < -\sqrt{\tilde{r}}\,\Big).$$

Da $\tilde{X} = X^2$ keine Werte $\tilde{r} < 0$ annehmen kann, ergibt sich insgesamt:

$$F_{\tilde{X}}(\tilde{r}) = F_X\Big(\sqrt{\tilde{r}}\,\Big) - F_X\Big(-\sqrt{\tilde{r}}\,\Big) = \begin{cases} 0 & \text{für } \tilde{r} < 0, \\ \sqrt{\tilde{r}} & \text{für } 0 \leq \tilde{r} \leq 1, \\ 1 & \text{für } \tilde{r} > 1. \end{cases}$$

Durch Ableiten erhalten wir die Dichtefunktion $\tilde{\rho}(\tilde{r})$ für $\tilde{X}$, mit der sich dann der Erwartungswert ergibt:

$$\tilde{\rho}(\tilde{r}) = \frac{d}{d\tilde{r}}\, F_{\tilde{X}}(\tilde{r}) = \begin{cases} \frac{1}{2}\,\frac{1}{\sqrt{\tilde{r}}} & \text{für } \tilde{r} \in [0,1], \\ 0 & \text{für } \tilde{r} \notin [0,1] \end{cases}$$

$$\Rightarrow \quad \mathrm{E}\,(\tilde{X}) = \int_{-\infty}^{\infty} \tilde{r}\, \tilde{\rho}(\tilde{r})\, d\tilde{r} = \int_{0}^{1} \tilde{r}\, \frac{1}{2\sqrt{\tilde{r}}}\, d\tilde{r} = \int_{0}^{1} \frac{\sqrt{\tilde{r}}}{2}\, d\tilde{r} = \frac{1}{3}.$$

Das mühselige Bestimmen der Verteilung von $\tilde{X}$ war völlig unnötig. Nach Satz 2.29 geht's einfacher direkt über die Verteilung von X:

$$\mathrm{E}\,(X^2) = \int_{-\infty}^{\infty} r^2\, \rho(r)\, dr = \int_{-1}^{1} \frac{r^2}{2}\, dr = \frac{1}{3}.$$

Bemerkung 2.31: *Mit Zufallsvariablen $X : \Omega \to \mathbb{R}$, $Y : \Omega \to \mathbb{R}$ kann man Arithmetik betreiben. Man kann z.B. Zufallsvariablen addieren oder multiplizieren, d.h., man betrachtet* die neuen Zufallsvariablen *$Z = X + Y : \Omega \mapsto \mathbb{R}$ oder $Z = X \cdot Y : \Omega \mapsto \mathbb{R}$, die punktweise durch*

$$\begin{aligned} X + Y \quad &: \quad \omega \in \Omega \quad \mapsto \quad X(\omega) + Y(\omega), \\ X \cdot Y \quad &: \quad \omega \in \Omega \quad \mapsto \quad X(\omega) \cdot Y(\omega) \end{aligned}$$

definiert sind. Es gilt die Faustregel:

> *Die Verteilung einer Funktion einer einzigen Zufallsvariablen X (z.B., $Z = X^2$) ist eindeutig durch die Verteilung F_X bestimmt.*
>
> *Die Verteilung einer Funktion mehrerer Zufallsvariablen X, Y (z.B., $Z = X + Y$ oder $Z = X \cdot Y$) lässt sich i.A.* <u>nicht</u> *allein aus den Verteilungen F_X, F_Y bestimmen (außer, wenn die Variablen X, Y unabhängig sind, siehe Abschnitt 2.5).*

Beispiel 2.32: Betrachte den zweifachen Wurf eines fairen Würfels

$$\Omega = \Big\{ (\omega_1, \omega_2); \omega_1, \omega_2 \in \{1, \ldots, 6\} \Big\}.$$

Sei

$$\begin{aligned} X_1 &= \text{Ergebnis des ersten Wurfs} : (\omega_1, \omega_2) \to \omega_1, \\ X_2 &= \text{Ergebnis des ersten Wurfs} : (\omega_1, \omega_2) \to \omega_2, \\ X = X_1 + X_2 &= \text{Augensumme} : (\omega_1, \omega_2) \to \omega_1 + \omega_2. \end{aligned}$$

Offensichtlich sind X_1, X_2 gleichverteilt:

$$P(X_1 = k_1) = P(X_2 = k_2) = \frac{1}{6} \quad \text{für alle } k_1, k_2 \in \{1, \ldots, 6\}.$$

Die Verteilung der Augensumme X war in Beispiel 2.3 berechnet worden:

$$P(X = 2) \ = \ \frac{1}{36}, \ P(X = 3) \ = \ \frac{1}{18}, \ P(X = 4) \ = \ \frac{1}{12}, \ \cdots \ , \ P(X = 12) \ = \ \frac{1}{36}.$$

> Es besteht keine offensichtliche Möglichkeit, aus den Verteilungswerten $P(X_1 = k_1)$, $P(X_2 = k_2)$ der Summanden direkt auf die Verteilung der Summe $X = X_1 + X_2$ zu schließen!

Trotzdem kann der Erwartungswert von X sehr leicht aus den Erwartungswerten der Summanden berechnet werden! Offensichtlich gilt

$$\mathrm{E}\,(X_1) = \mathrm{E}\,(X_2) = \frac{1 + 2 + \cdots + 6}{6} = \frac{7}{2},$$

der Erwartungswert von $X = X_1 + X_2$ war in Beispiel 2.26 als $\mathrm{E}\,(X) = 7$ bestimmt worden, also $\mathrm{E}\,(X_1 + X_2) = \mathrm{E}\,(X_1) + \mathrm{E}\,(X_2)$. Zufall? Nein! Siehe Satz 2.33.1).

Die wohl Wichtigste der folgenden Aussagen ist die Linearität 1) von Erwartungswerten:

Satz 2.33: (Rechenregeln für Erwartungswerte)

Seien X, Y reelle Zufallsvariable über demselbem $(\Omega, \mathcal{E}, P)$, sei $\alpha \in \mathbb{R}$.

1) Der Erwartungswert ist linear:

$$\mathrm{E}\,(\alpha\, X) = \alpha\, \mathrm{E}\,(X)\ , \quad \mathrm{E}\,(X+Y) = \mathrm{E}\,(X) + \mathrm{E}\,(Y)\ .$$

2) Sei $X = \alpha$ konstant. Dann gilt $\mathrm{E}\,(X) = \alpha$.

3) Sei $X(\omega) \geq Y(\omega)$ für alle $\omega \in \Omega$ (kurz: $X \geq Y$). Dann gilt

$$\mathrm{E}\,(X) \geq \mathrm{E}\,(Y)\ .$$

Speziell gilt $\mathrm{E}\,(X) \geq 0$ für $X \geq 0$. Für beliebiges X gilt:

$$\mathrm{E}\,(|X|) \geq |\mathrm{E}\,(X)|\ .$$

*4) Sei $A \subset \Omega$ ein Ereignis. Für die „**Indikatorfunktion**“*

$$1_A : \omega \in \Omega \ \longrightarrow \ \begin{cases} 1 & \text{für } \omega \in A\ , \\ 0 & \text{für } \omega \in \Omega \setminus A \end{cases}$$

gilt $\mathrm{E}\,(1_A) = P(A)$. Für unabhängige Ereignisse A, B gilt:

$$\mathrm{E}\,(1_A\, 1_B) = \mathrm{E}\,(1_{A\cap B}) = P(A \cap B) = P(A)\, P(B) = \mathrm{E}\,(1_A)\, \mathrm{E}\,(1_B)\ .$$

5) Sei $X \geq 0$. Für jedes $\alpha \in \mathbb{R}$ gilt:

$$\mathrm{E}\,(X) \ \geq \ \alpha \cdot P(X \geq \alpha)\ .$$

Beweisskizze:

Für den Spezialfall eines diskreten Stichprobenraums $\Omega = \{\omega_1, \omega_2, \dots\}$ folgend die Eigenschaften 1) – 4) unmittelbar aus der Darstellung 2.24.a):

$$E(X) = \sum_{\omega\in\Omega} X(\omega)\, P(\{\omega\}).$$

Für den allgemeinen Fall:

2) Aus der Definition der Integrale über Stieltjes-Summen folgt sofort

$$\mathrm{E}\,(X) = \int_{\mathbb{R}} \alpha\, dF_X(r) = \alpha \int_{\mathbb{R}} dF_X(r) = \alpha\, F_X(\infty) = \alpha\ .$$

4) $\mathrm{E}\,(1_A) = \int_{\mathbb{R}} 1_A \; dF_X(r) = \int_A \; dF_X(r) = P(A)$.
1) Sei $\alpha > 0$. Eine $\int_{(a,b]} r\, dF_{\alpha X}(r)$ approximierende Stieltjes-Summe hat die Form

$$\sum_{i=0}^{n-1} r_i \, \Big(F_{\alpha X}(r_{i+1}) - F_{\alpha X}(r_i)\Big) = \alpha \sum_{i=0}^{n-1} r_i \; P(r_i < \alpha X \le r_{i+1}) \; .$$

$$= \alpha \sum_{i=0}^{n-1} \frac{r_i}{\alpha} \; P(\frac{r_i}{\alpha} < X \le \frac{r_{i+1}}{\alpha}) \; .$$

Dies ist eine Stieltjes-Summe für $\alpha \int_{(a/\alpha, b/\alpha]} r\, dF_X(r)$, also gilt mit $n \to \infty$:

$$\int_{(a,b]} r\, dF_{\alpha X}(r) = \alpha \int_{(a/\alpha, b/\alpha]} r\, dF_X(r) \; .$$

Im Grenzwert $a \to -\infty$, $b \to -\infty$ folgt $\mathrm{E}\,(\alpha X) = \alpha\,\mathrm{E}\,(X)$.
Mit kleinen Variationen gelten dieselben Argumente auch für $\alpha < 0$. Für $\alpha = 0$ ist die Behauptung trivial.

Nun wird $\mathrm{E}\,(X+Y) = \mathrm{E}\,(X) + \mathrm{E}\,(Y)$ gezeigt. Seien zunächst X und Y und damit auch $Z = X + Y$ diskret: $X(\Omega) = \{x_1, x_2, \dots\}$, $Y(\Omega) = \{y_1, y_2, \dots\}$,

$$Z(\Omega) = \{z_1, z_2, \dots\} = \{x + y;\; x \in X(\Omega);\; y \in Y(\Omega)\} \; .$$

Mit der Notation

$$\begin{aligned} P(Z = z_k) \;&=\; \sum_{\substack{x_i, y_i \\ x_i + y_j = z_k}} P\Big(\{\omega \in \Omega;\; X(\omega) = x_i; Y(\omega) = y_j\}\Big) \\ &\equiv\; \sum_{x_i} \sum_{y_j} P(x_i + y_j = z_k) \end{aligned}$$

folgt

$$\begin{aligned} \mathrm{E}\,(Z) \;&=\; \sum_{z_k} z_k\, P(Z = z_k) \;=\; \sum_{z_k} \sum_{x_i} \sum_{y_j} (x_i + y_j)\, P(x_i + y_j = z_k) \\ =\; & \sum_{z_k} \sum_{x_i} \sum_{y_j} x_i\, P(x_i + y_j = z_k) \quad + \quad \sum_{z_k} \sum_{x_i} \sum_{y_j} y_j\, P(x_i + y_j = z_k) \\ =\; & \sum_{x_i} x_i \sum_{z_k} \sum_{y_j} P(x_i + y_j = z_k) \quad + \quad \sum_{y_j} y_j \sum_{z_k} \sum_{x_i} P(x_i + y_j = z_k) \\ =\; & \sum_{x_i} x_i \; P(X = x_i) \;+\; \sum_{y_j} y_j \; P(Y = y_j) \;=\; \mathrm{E}\,(X) + \mathrm{E}\,(Y) \; . \end{aligned}$$

Für den allgemeinen Fall deuten wir nur die Beweisidee an: zu gegebenem $n \in \mathbb{N}$ betrachte die Approximation von X durch die „Treppenfunktionen“

$$X_n = \sum_{k=-\infty}^{\infty} \frac{k}{n}\, 1_{A_{kn}}$$

mit der disjunkten Zerlegung

$$A_{kn} = \Big\{\omega \in \Omega;\, \frac{k}{n} < X(\omega) \le \frac{k+1}{n}\Big\}\ , \qquad \bigcup_{k=-\infty}^{\infty} A_{kn} = \Omega\ .$$

Beachte: jedes $\omega \in \Omega$ liegt in genau einem der A_{kn}, d.h., die Summe $X_n(\omega) = \sum_k (k/n)\, 1_{A_{kn}}(\omega)$ besteht aus genau einem Term. X_n ist eine diskrete Variable, die nur Werte der Form „ganze Zahl“$/n$ annimmt. Es gilt $X_n < X \le X_n + 1/n$. Es folgt mit den entsprechenden Approximation Y_n für Y, Z_n für Z:

$$Z_n \le Z = X + Y \le X_n + Y_n + \frac{2}{n} \le X + Y + \frac{2}{n} = Z + \frac{2}{n} \le Z_n + \frac{3}{n}\ ,$$

also

$$-\frac{2}{n} \le Z_n - X_n - Y_n \le \frac{1}{n}\ .$$

Die Rechenregel 3) gilt sicherlich für diskrete Variable (siehe die Argumente unten für 3), die Linearität des Erwartungswerts ist für die diskreten Variablen X_n, Y_n, Z_n sowie die Konstanten gezeigt). Es folgt

$$-\frac{2}{n} \le \mathrm{E}\,(Z_n - X_n - Y_n) = \mathrm{E}\,(Z_n) - \mathrm{E}\,(X_n) - \mathrm{E}\,(Y_n) \le \frac{1}{n}\ . \qquad (\#)$$

Mit wachsendem n folgt $\mathrm{E}\,(X_n) \to \mathrm{E}\,(X)$, $\mathrm{E}\,(Y_n) \to \mathrm{E}\,(Y)$, $\mathrm{E}\,(Z_n) \to \mathrm{E}\,(Z)$ (diesen wichtigen, aber sehr technischen Schritt unterschlagen wir hier). Für die Grenzwerte folgt mit $(\#)$ dann $0 \le \mathrm{E}\,(Z) - \mathrm{E}\,(X) - \mathrm{E}\,(Y) \le 0$, also $\mathrm{E}\,(Z) = \mathrm{E}\,(X) + \mathrm{E}\,(Y)$.

3) Sei $Z = X - Y \ge 0$. Alle den Erwartungswert $\mathrm{E}\,(Z)$ approximierenden endlichen Stieltjes-Summen sind ≥ 0, damit gilt auch im Grenzwert $\mathrm{E}\,(Z) \ge 0$. Mit der Linearität 1) folgt $\mathrm{E}\,(X) - \mathrm{E}\,(Y) \ge 0$. Mit $|X| \ge \pm X$ folgt $\mathrm{E}\,(|X|) \ge \pm\mathrm{E}\,(X)$.

5) Setze $A_\alpha = \{\omega \in \Omega;\ X(\omega) \ge \alpha\}$. Für $X \ge 0$ gilt dann $X \ge \alpha\, 1_{A_\alpha}$. Mit 1) bis 4) folgt sofort die Behauptung 5).

Q.E.D.

Beispiel 2.34: Gegeben sei eine Variable X mit der Verteilungsfunktion F_X. Betrachte die neue Zufallsvariable $Y = \alpha\, X + \beta$ mit konstanten $\alpha,\ \beta \in \mathbb{R}$. Der Erwartungswert von Y ist

$$\mathrm{E}\,(Y) = \mathrm{E}\,(\alpha\, X + \beta) = \mathrm{E}\,(\alpha\, X) + \mathrm{E}\,(\beta \cdot 1) = \alpha \cdot \mathrm{E}\,(X) + \beta.$$

In diesem einfachen Fall kann man sogar die Verteilung von Y explizit über die Verteilung von X ausdrücken. Für $\alpha > 0$ gilt

$$F_Y(r) = P(Y \leq r) = P(\alpha\, X + \beta \leq r) = P\Big(X \leq \frac{r-\beta}{\alpha}\Big) = F_X\Big(\frac{r-\beta}{\alpha}\Big).$$

Für $\alpha < 0$ gilt:

$$F_Y(r) = P(Y \leq r) = P(\alpha\, X + \beta \leq r) = P\Big(X \geq \frac{r-\beta}{\alpha}\Big) = 1 - P\Big(X < \frac{r-\beta}{\alpha}\Big)$$

$$= 1 + P\Big(X = \frac{r-\beta}{\alpha})\Big) - P\Big(X \leq \frac{r-\beta}{\alpha})\Big) = 1 + P\Big(X = \frac{r-\beta}{\alpha})\Big) - F_X\Big(\frac{r-\beta}{\alpha}\Big).$$

Für die Berechnung von $\mathrm{E}\,(Y)$ wird dies aber nicht benötigt: wegen der Linearität weiß man unmittelbar, dass $\mathrm{E}\,(Y) = \alpha \cdot \mathrm{E}\,(X) + \beta$ gilt.

Der folgende Begriff „Varianz“ ist genauso fundamental wie der Begriff „Erwartungswert“:

Definition 2.35: (Varianz und Streuung)
Die **Varianz** *einer Zufallsvariable* X *ist*

$$\mathrm{Var}\,(X) = \mathrm{E}\,(\ (X - \mathrm{E}\,(X))^2\ ) \stackrel{(*)}{=} \mathrm{E}\,(X^2) - \mathrm{E}\,(X)^2 \geq 0\ .$$

Die **Streuung** *oder auch* **Standardabweichung** *von* X *ist*

$$\sigma(X) = \sqrt{\mathrm{Var}\,(X)} = \sqrt{\mathrm{E}\,(\ (X - \mathrm{E}\,(X))^2\ )}\ .$$

Hierbei gilt die Identität $(*)$ wegen der Linearität des Erwartungswerts:

$$\mathrm{E}\,(\ (X - \mathrm{E}\,(X))^2\ ) = \mathrm{E}\,(X^2 - 2\,X\,\mathrm{E}\,(X) + \mathrm{E}\,(X)^2)$$

$$= \mathrm{E}\,(X^2) - 2\,\mathrm{E}\,(X)\,\mathrm{E}\,(X) + \mathrm{E}\,(X)^2 = \mathrm{E}\,(X^2) - \mathrm{E}\,(X)^2\ .$$

Die Interpretation der Varianz wird durch den folgenden Satz geliefert, dessen Stärke seine Allgemeinheit ist (keinerlei Voraussetzungen über X). In praktischen Anwendungen ist die Abschätzung allerdings recht grob:

Satz 2.36: (Chebyshevsche Ungleichung)
Sei X eine beliebige Zufallsvariable über einem beliebigen Modell $(\Omega, \mathcal{E}, P)$ mit Erwartungswert $\mathrm{E}\,(X)$ und Varianz $\mathrm{Var}\,(X)$. Dann gilt für jedes $\epsilon > 0$:

$$P(|X - \mathrm{E}\,(X)| \geq \epsilon) \quad \leq \quad \frac{\mathrm{Var}\,(X)}{\epsilon^2}$$

Äquivalenterweise gilt für die Komplementaussage:

$$P(|X - \mathrm{E}\,(X)| < \epsilon) \quad \geq \quad 1 - \frac{\mathrm{Var}\,(X)}{\epsilon^2} \ .$$

Beweis: Wende Satz 2.33.5) an auf $Y = (X - \mathrm{E}\,(X))^2$ mit $\alpha = \epsilon^2$. Es ergibt sich sofort

$$\mathrm{Var}\,(X) = \mathrm{E}\,(Y) \geq \alpha\ P(Y \geq \alpha) = \epsilon^2\ P((X - \mathrm{E}\,(X))^2 \geq \epsilon^2)$$
$$= \epsilon^2\ P(|X - \mathrm{E}\,(X)| \geq \epsilon) \ .$$

Q.E.D.

Bemerkung 2.37: *Für $\epsilon \leq \sigma(X)$ ergibt sich keinerlei Information, da dann $\mathrm{Var}\,(X)/\epsilon^2 \geq 1$ gilt und W'keiten trivialerweise immer zwischen 0 und 1 liegen. Wählt man $\epsilon = n\,\sigma(X)$ als kleines Vielfaches von $\sigma(X)$, so ergibt sich mit*

$$P\Big(\mathrm{E}\,(X) - n\,\sigma(X) < X < \mathrm{E}\,(X) + n\,\sigma(X)\Big) \ \geq \ 1 - \frac{1}{n^2}$$

die Interpretation:

Mit großer W'keit nimmt eine Zufallsvariable Werte an, die höchstens um ein kleines Vielfaches von σ vom Erwartungswert abweichen.

Beispielsweise gilt für $\epsilon = 3\,\sigma(X)$, dass X mit mindestens $8/9 \ \hat{=}\ 88.9\%$-iger W'keit Werte innerhalb des Intervalls $(\mathrm{E}\,(X) - 3\,\sigma(X), \mathrm{E}\,(X) + 3\,\sigma(X))$ annimmt.

Merke: *Die Standardabweichung σ gibt die Größenordnung an, um welche die Zufallswerte vom Erwartungswert abweichen („streuen“). Für kleines σ werden nur Werte dicht beim Erwartungswert beobachtet, für großes σ kommen mit großer W'keit auch Werte „weit weg“ vom Erwartungswert vor.*

Beispiel 2.38: Betrachte die Zufallsvariable $X : \Omega \mapsto \{0, 1, \dots, n\} \subset \mathbb{R}$ mit ↓4.6.07

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

wobei $p \in [0, 1]$ und $q = 1 - p$. Nach Satz 1.63 beschreibt X die Anzahl der Erfolge bei n-facher Wiederholung eines Bernoulli-Experiments mit Erfolgsw'keit p. Es gilt

$$\mathrm{E}\,(X) = \sum_{k=0}^{n} k \ P(X = k) \ = \sum_{k=0}^{n} k \binom{n}{k} p^k \, q^{n-k} = p \ \sum_{k=0}^{n} \binom{n}{k} \underbrace{k \, p^{k-1}}_{d\,p^k/dp} q^{n-k}$$

$$= p \ \frac{d}{dp} \ \sum_{k=0}^{n} \binom{n}{k} p^k \, q^{n-k} \overset{\text{(Binomi)}}{=} p \ \frac{d}{dp} \ (p + q)^n = p\,n \ (p+q)^{n-1} \overset{(p+q=1)}{=} p\,n \ .$$

Dies ist intuitiv: im Mittel führt der Bruchteil p aller n Versuche zum Erfolg, also sollte der Erwartungswert für die Gesamtanzahl der Erfolge $p\,n$ sein. Wie stark streuen die Erfolge um diesen Mittelwert?

$$\mathrm{Var}\,(X) = \mathrm{E}\,(X^2) - \mathrm{E}\,(X)^2 = \mathrm{E}\,(X^2) - (p\,n)^2 \ .$$

Es gilt

$$\mathrm{E}\,(X^2) - \mathrm{E}\,(X) = \mathrm{E}\,(X^2 - X) = \sum_{k=0}^{n} (k^2 - k) \ P(X = k) = \sum_{k=0}^{n} k\,(k-1) \binom{n}{k} p^k \, q^{n-k}$$

$$= p^2 \ \sum_{k=0}^{n} \binom{n}{k} \ \underbrace{k\,(k-1)\,p^{k-2}}_{d^2\,p^k/dp^2} \, q^{n-k} = p^2 \ \frac{d^2}{dp^2} \ \sum_{k=0}^{n} \binom{n}{k} p^k \, q^{n-k}$$

$$\overset{\text{(Binomi)}}{=} p^2 \ \frac{d^2}{dp^2} \ (p+q)^n = p^2\,n\ (n-1)\ (p+q)^{n-2} \overset{(p+q=1)}{=} p^2\,n\ (n-1) \ .$$

Es folgt

$$\mathrm{E}\,(X^2) = p^2\,n\ (n-1) + \mathrm{E}\,(X) = p^2\,n\ (n-1) + p\,n$$

und damit

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

Beispiel 2.39: Frage: mit welcher W'keit liegt bei 1000 Würfen einer fairen Münze die Anzahl der „Köpfe“ zwischen 400 und 600?
Antwort: dies ist eine $n = 1000$-fache Wiederholung des Bernoulli-Experiments „Münzwurf“ mit $p = P(K) = q = P(Z) = 1/2$. Nach Beispiel 2.38 hat die „Anzahl der Köpfe“ X den Erwartungswert $\mathrm{E}\,(X) = p\,n = 500$, die Varianz ist $\mathrm{Var}\,(X) = p\,q\,n = n/4 = 250$, die Streuung ist $\sigma(X) = \sqrt{\mathrm{Var}\,(X)} = \sqrt{250} \approx 15.8$. Nach Chebyshev gilt:

$$P(400 \le X \le 600) = P(|X - 500| \le 100) = P(|X - \mathrm{E}\,(X)| < 101)$$

$$\geq 1 - \frac{\mathrm{Var}\,(X)}{101^2} = 1 - \frac{250}{101^2} \approx 0.975 \ .$$

Für die W'keit, dass X zwischen 480 und 520 liegt, folgt analog

$$P(480 \leq X \leq 520) = P(|X - 500| \leq 20) = P(|X - \mathrm{E}\,(X)| < 21)$$

$$\geq 1 - \frac{\mathrm{Var}\,(X)}{21^2} = 1 - \frac{250}{21^2} \approx 0.433 \ .$$

Für die W'keit, dass X zwischen 490 und 510 liegt, folgt analog

$$P(490 \leq X \leq 510) = P(|X - 500| \leq 10) = P(|X - \mathrm{E}\,(X)| < 11)$$

$$\geq 1 - \frac{\mathrm{Var}\,(X)}{11^2} = 1 - \frac{250}{11^2} \approx -1.07$$

(hier hat Chebyshev keine Aussagekraft mehr). Die exakte Formel

$$P(a \leq X \leq b) = \sum_{k=a}^{b} \binom{n}{k} p^k \, q^{n-k}$$

liefert

$$\begin{array}{rcl} P(400 \leq X \leq 600) & \approx & 0.999999999820 \ , \\ P(480 \leq X \leq 520) & \approx & 0.805 \ , \\ P(490 \leq X \leq 510) & \approx & 0.493 \ . \end{array}$$

Man sieht an diesen Werten: die von Chebyshev gelieferten Abschätzungen für die W'keiten sind zwar korrekt, aber arg grob.

2.4 Einige Standardverteilungen

Hier werden einige wichtige Verteilungen mit den für Anwendungen wichtigen Interpretationen vorgestellt.

2.4.1 Die Binomial-Verteilung

Zu gegebenem $n \in \mathbb{N}$ und $p \in [0,1]$, $q = 1 - p$, betrachte $X : \Omega \to \{0, 1, \ldots, n\}$ mit

$$P(X = k) \ = \ \binom{n}{k} p^k \, q^{n-k} \ , \quad k = 0, 1, \ldots, n \ .$$

Nach Beispiel 2.38 gilt

$$\mathrm{E}\,(X) = p\,n \ , \quad \mathrm{Var}\,(X) = p\,q\,n \ , \quad \sigma(X) = \sqrt{p\,q\,n} \ .$$

Kürzel: Man sagt, X ist „Bi(n, p)-verteilt“.

MuPAD-Routinen: `stats::binomialCDF`, `stats::binomialQuantile`, `stats::binomialPF`, `stats::binomialRandom`.

Interpretation: Nach Satz 1.63 beschreibt X die Anzahl der Erfolge in einer n-fachen Wiederholung eines Bernoulli-Experiments mit Erfolgsw'keit p.

Beispiel: Anzahl der „Köpfe“ bei n-fachem Münzwurf (auch mit einer unfairen Münze), $p = P(\text{„Kopf“})$.

2.4.2 Die Poisson-Verteilung

6.6.07↓ Zu gegebenem $\lambda > 0$ betrachte $X : \Omega \to \{0, 1, 2, \dots\}$ mit

$$P(X = k) \ = \ \frac{\lambda^k}{k!}\, e^{-\lambda} \ ; \quad k = 0, 1, 2, \dots \ .$$

Man rechnet leicht nach:

$$\mathrm{E}\,(X) = \lambda \ ; \quad \mathrm{Var}\,(X) = \lambda \ ; \quad \sigma(X) = \sqrt{\lambda} \ .$$

Kürzel: Man sagt, X ist „Pois(λ)-verteilt“.

MuPAD-Routinen: `stats::poissonCDF`, `stats::poissonQuantile`, `stats::poissonPF`, `stats::poissonRandom`.

Interpretation: Im folgenden Sinne beschreibt die Poisson-Verteilung „seltene Bernoulli-Ereignisse“.

> Betrachte eine binomial-verteilte Variable X mit großem n und kleinem $p \ll 1/\sqrt{n} \ll 1$. Mit $\lambda = p\,n$ gilt dann
>
> $$P(X = k) \ = \ \binom{n}{k}\, p^k\, (1-p)^{n-k} \approx \frac{\lambda^k}{k!}\, e^{-\lambda} \quad \text{(„Poisson-Näherung“)}$$
>
> für alle ganzzahligen Werte $k \ll \sqrt{n}$.

Etwas genauer:

$$\binom{n}{k}\, p^k\, (1-p)^{n-k} = \frac{\lambda^k}{k!}\, e^{-\lambda} \left(1 + \frac{k}{2\,n} - \frac{k^2}{2\,n} - \frac{p^2\,n}{2} + k\,p + \cdots\right).$$

Beweis:
Mit den Forderungen $n \gg 1$, $p \ll 1/\sqrt{n}$ und $k \ll \sqrt{n}$ gelten folgende Bedingungen:

$$n \gg 1 \ , \ n - k \gg 1 \ , \ \frac{k}{n} \ll \frac{1}{\sqrt{n}} \ll 1 \ , \ \frac{k^2}{n} \ll 1 \ , \ \frac{p^2}{n} \ll 1 \ , \ k\,p \ll 1 \ .$$

Es gilt die Stirling-Formel

$$m! \approx \sqrt{2\,\pi\,m} \quad m^m\, e^{-m} \left(1 + \frac{1}{12\,m} + \frac{1}{288\,m^2} + \cdots\right) .$$

Schon ab $m = 5$ ist die Näherung $m! \approx \sqrt{2\,\pi\,m}\;\, m^m\, e^{-m}$ auf besser als 2% genau. Es folgt für $n \gg 1$, $n-k \gg 1$:

$$\begin{aligned}
\frac{\binom{n}{k}\ p^k\ (1-p)^{n-k}}{\frac{\lambda^k}{k!}\ e^{-\lambda}} &= \frac{n!}{(n-k)!}\ \frac{1}{n^k}\ e^{\lambda}\ e^{(n-k)\ \ln(1-p)} \\
&= \frac{n!}{(n-k)!}\ \frac{1}{n^k}\ e^{n\,p}\ e^{n\ \ln(1-p)}\ e^{-k\ \ln(1-p)} \\
&\overset{\text{(Stirling)}}{\approx} \frac{\sqrt{n}\ n^n\ e^{-n}}{\sqrt{n-k}\ (n-k)^{n-k}\ e^{-(n-k)}}\ \frac{1}{n^k}\ e^{n\,p+n\ \ln(1-p)}\ e^{-k\ \ln(1-p)} \\
&= \frac{1}{\sqrt{1-\frac{k}{n}}}\ \frac{e^{-k}}{(1-\frac{k}{n})^{n-k}}\ e^{n\,p+n\ \ln(1-p)}\ e^{-k\ \ln(1-p)} \\
&= \frac{1}{\sqrt{1-\frac{k}{n}}}\ e^{-k-(n-k)\ \ln(1-\frac{k}{n})}\ e^{n\,(p+\ln(1-p))}\ e^{-k\ \ln(1-p)} \\
&= \frac{1}{\sqrt{1-\frac{k}{n}}}\ e^{-k\,(\ 1+(\frac{n}{k}-1)\ \ln(1-\frac{k}{n})\ )}\ e^{n\,(p+\ln(1-p))}\ e^{-k\ \ln(1-p)}\ .
\end{aligned}$$

Beachte $\ln(1-\epsilon) = -\epsilon - \epsilon^2/2 - \epsilon^3/3 - \cdots$ für $\epsilon < 1$. Für $k/n \ll 1$ gelten damit die Approximationen

$$\frac{1}{\sqrt{1-\frac{k}{n}}} = 1 + \frac{1}{2}\ \frac{k}{n} + \frac{3}{8}\ \frac{k^2}{n^2} + \cdots \approx 1 + \frac{k}{2\,n},$$

$$-k\left(1+\left(\frac{n}{k}-1\right)\ln\left(1-\frac{k}{n}\right)\right) = -\frac{k^2}{2\,n}\left(1+\frac{1}{3}\ \frac{k}{n}+\frac{1}{6}\ \frac{k^2}{n^2}+\cdots\right) \approx -\frac{k^2}{2\,n}\ .$$

Analog mit $p \ll 1$:

$$n\,(p+\ln(1-p)) = -\frac{n\,p^2}{2}\ \left(1+\frac{2\,p}{3}+\frac{p^2}{2}+\cdots\right) \approx -\frac{n\,p^2}{2}\ ,$$

$$-k\ \ln(1-p) = k\,p\ \left(1+\frac{p}{2}+\frac{p^2}{3}+\cdots\right) \approx k\,p\ .$$

Mit $k^2/n \ll 1$, $p^2\,n \ll 1$, $k\,p \ll 1$ folgt dann

$$\frac{\binom{n}{k}\ p^k\ (1-p)^{n-k}}{\frac{\lambda^k}{k!}\ e^{-\lambda}} \approx \left(1+\frac{k}{2\,n}\right)\ e^{-\frac{k^2}{2\,n}}\ e^{-\frac{p^2\,n}{2}}\ e^{k\,p}$$

$$\approx 1 + \frac{k}{2\,n} - \frac{k^2}{2\,n} - \frac{p^2\,n}{2} + k\,p \approx 1 \ .$$

Q.E.D.

Merke: Die Poisson-Verteilung ist zuständig für häufige ($n \gg 1$) Wiederholungen von Bernoulli-Experimenten mit kleiner Erfolgsw'keit $p \ll 1/\sqrt{n} \ll 1$, deren Erwartungswert $\lambda = n\,p$ bekannt ist.

Beispiel 2.40: Man stellt fest, dass eine Sekretärin im Durchschnitt 1.5 Tippfehler pro Seite macht. Mit welcher W'keit hat eine betrachtete Seite keinen Tippfehler?
Lösung: Intuitiv ist das Auftauchen eines Tippfehlers ein „seltenes Ereignis", also verwende man die Poisson-Verteilung

$$P(\text{„genau } k \text{ Tippfehler auf einer Seite"}) = \frac{\lambda^k\,e^{-\lambda}}{k!}$$

mit dem Erwartungswert $\lambda = 1.5$. Damit folgt

$$P(\text{„kein Tippfehler"}) = \frac{\lambda^0\,e^{-\lambda}}{0!} = e^{-1.5} \approx 0.223 \ .$$

Genaueres Modell: Sei p die W'keit, beim Tippen eines Buchstabens die falsche Taste zu drücken. Voraussetzung: p sei für jeden Buchstaben gleich groß. Das Tippen einer Seite wird damit zu einer häufigen Wiederholung (n = Anzahl der Buchstaben auf einer Seite, Größenordnung 1000) eines Bernoulli-Experiments mit geringer „Erfolgs"-W'keit. Ein „Erfolg" = „Tippfehler" ist in der Tat unwahrscheinlich: der Erwartungswert $\lambda = np$ ist empirisch als 1.5 bestimmt, also ist $p = P(\text{„Tippfehler"})$ von der Größenordnung $1.5/1000$. Mit $p \approx 0.0015 \ll 1/\sqrt{n} \approx 0.03$ lässt sich die Poisson-Näherung anwenden. Für X = „Anzahl der Tippfehler auf einer Seite" ergibt sich

$$P(X = k) \ = \ \frac{\lambda^k\,e^{-\lambda}}{k!} \ , \quad \lambda = 1.5 \ .$$

Weitere Frage: Mit welcher W'keit wird ein 20 Seiten langes Skript weniger als 5 Tippfehler enthalten?
Lösung: Nun ist $\tilde{n} = 20\,n$, wobei n wie oben die Anzahl der Buchstaben auf einer Seite sei. Für p gilt (halbwegs) $p \approx 0.0015 \ll 1/\sqrt{\tilde{n}} \approx 0.007$, die Poisson-Näherung ist damit noch einigermaßen akzeptabel. Mit dem Erwartungswert von

$$\tilde{\lambda} = \tilde{n}\,p = 20\,n\,p = 20\,\lambda = 20 \cdot 1.5 = 30$$

Tippfehlern für 20 Seiten erhält man für die Variable $\tilde{X}$ = „Anzahl der Tippfehler im gesamten Manuskript":

$$P(\tilde{X} < 5) = \sum_{k=0}^{4} \frac{30^k\,e^{-30}}{k!} \approx 3.6 \cdot 10^{-9} \ .$$

Das ist realistisch: wie jeder passionierte Leser weiß, gibt es keine Bücher ohne Tippfehler.

2.4.3 Die geometrische Verteilung

Zu gegebenem $p \in (0,1]$, $q = 1 - p$, betrachte $X : \Omega \to \{1, 2, 3, \dots\}$ mit

$$P(X = k) \;=\; p \cdot q^{k-1} \;, \quad k = 1, 2, 3, \dots \;.$$

Es gilt

$$\mathrm{E}\,(X) = \frac{1}{p} \;, \quad \mathrm{Var}\,(X) = \frac{q}{p^2} \;, \quad \sigma(X) = \frac{\sqrt{q}}{p} \;.$$

Kürzel: Man sagt, X ist „Geom(p)-verteilt“ (oder auch Geo(p)).

MuPAD-Routinen: `stats::geometricCDF`, `stats::geometricQuantile`, `stats::geometricPF`, `stats::geometricRandom`.

Interpretation: Nach Blatt 6, Aufgabe 34, beschreibt X die Anzahl der benötigten Versuche, bis in der Wiederholung eines Bernoulli-Experiments mit Erfolgsw'keit p zum ersten Mal „Erfolg“ eintritt.

Beispiel: Man muss im Mittel 13 983 816 Mal Lotto spielen, bis man „6 Richtige“ erzielt (nach Beispiel 1.8 ist die Erfolgsw'keit des Bernoulli–Experiments „6 Richtige bei einem Lotto-Spiel“ $p = \frac{1}{13\,983\,816}$).

2.4.4 Die hypergeometrische Verteilung

Zu gegebenem $N, S, n \in \{0, 1, 2, \dots\}$ mit $0 \le S \le N$ und $0 \le n \le S$ betrachte $X : \Omega \to \{0, 1, \dots, \min(n, S)\}$ mit

$$P(X = s) = \frac{\binom{S}{s}\,\binom{N-S}{n-s}}{\binom{N}{n}} \;, \quad s = 0, 1, \dots, \min(n, S) \;.$$

Es gilt

$$\mathrm{E}\,(X) = \frac{S\,n}{N} \;, \quad \mathrm{Var}\,(X) = \sigma^2(X) = \frac{(N-S)\,(N-n)\,S\,n}{(N-1)\,N^2} \;.$$

Kürzel: Man sagt, X ist „Hypergeom(N, S, n)-verteilt“ (oder auch $HyG(N, S, n)$).

MuPAD-Routinen: `stats::hypergeometricCDF`, `stats::hypergeometricQuantile`, `stats::hypergeometricPF`, `stats::hypergeometricRandom`.

Interpretation: Nach Beispiel 1.10 beschreibt X die Anzahl der „Erfolge“, wenn man n Mal ohne Zurücklegen aus einer Urne mit N Objekten zieht, von denen S als „Erfolg“ gelten.

Beispiel: Die Anzahl der Asse pro Hand beim Skat sind HyperGeom$(32, 4, 10)$-verteilt. Man erhält im Mittel $\mathrm{E}\,(X) = 1.25$ Asse.

2.4.5 Die Exponentialverteilung

Zu gegebenem $\lambda > 0$ betrachte $X : \Omega \to [0, \infty)$ mit der Dichte

$$\rho(r) = \lambda\, e^{-\lambda\, r}\ , \quad r \geq 0$$

(und $\rho(r) = 0$ für $r < 0$), also für $0 \leq t$:

$$1 - F_X(t) = P(t < X) = \int_t^\infty \lambda\, e^{-\lambda\, r}\ dr = e^{-\lambda\, t}\ .$$

Man rechnet leicht nach:

$$\mathrm{E}\,(X) = \frac{1}{\lambda}\ ; \quad \mathrm{Var}\,(X) = \frac{1}{\lambda^2}\ ; \quad \sigma(X) = \frac{1}{\lambda}\ .$$

Kürzel: Man sagt, X ist „Exp(λ)-verteilt“.

MuPAD-Routinen: `stats::exponentialCDF`, `stats::exponentialQuantile`, `stats::exponentialPDF`, `stats::exponentialRandom`.

Interpretation: Ein Exp(λ)-verteiltes X beschreibt die Wartezeit bis zum ersten Eintreten eines Ereignisses, wenn die beobachteten Ereignisse unabhängig voneinander mit zeitlich konstanter Rate λ = „mittlere Anzahl von Ereignissen pro Zeiteinheit“ eintreten. Diese Verteilung hat „kein Gedächtnis“: für alle $t, t_0 \in [0, \infty)$ ergibt sich die bedingte W'keit

$$P(t_0 + t < X \mid t_0 < X) = \frac{P(t + t_0 < X)}{P(t_0 < X)} = \frac{e^{-\lambda\,(t+t_0)}}{e^{-\lambda\, t_0}} = P(t < X).$$

Im Klartext: zum Zeitpunkt 0 habe ich eine Beobachtung begonnen. $P(t < X)$ ist die W'keit, dass das Ereignis nicht innerhalb der ersten t Zeiteinheiten eintritt. Ich habe nun t_0 Zeiteinheiten gewartet, das Ereignis ist nicht eingetreten. Die bedingte W'keit, dass das Ereignis nicht in den nächsten t Zeiteinheiten eintritt, stimmt mit $P(t < X)$ überein. Also: es ist egal, zu welchem Zeitpunkt t_0 ich mit der Beobachtung beginne.

Beispiel: radioaktiver Zerfall, siehe Beispiel 2.27.

Beispiel 2.41: Ich beobachte im Mittel jede halbe Stunde einen Meteoriten am Himmel. Mit welcher W'keit beobachte ich einen Meteoriten innerhalb der ersten 15 Minuten meiner Beobachtung?
Lösung: Sei X = „der Zeitpunkt, an dem ich zum ersten Mal seit Beginn meiner Beobachtung einen Meteoriten sehe". Der Erwartungswert ist als $\mathrm{E}\,(X) = 1/\lambda = 1/2$ vorgegeben, also $\lambda = 2$ (Meteoriten pro Stunde). Damit ist die W'keit, in den ersten 15 Minuten (mindestens) einen Meteoriten zu sehen

$$P\Big(X \leq \frac{15}{60}\Big) = \int_0^{1/4} \lambda\, e^{-\lambda\, r}\; dr = 1 - e^{-\lambda/4} = 1 - e^{-1/2} \approx 0.39.$$

2.4.6 Die Gleichverteilung

$\boxed{\downarrow 11.6.07}$

Zu gegebenem Intervall $[a, b] \subset \mathbb{R}$ betrachte $X : \Omega \to [a, b]$ mit der konstanten Dichte

$$\rho(r) = \frac{1}{b-a}\ , \quad r \in [a, b]$$

(und $\rho(r) = 0$ für $r \not\in [a, b]$). Man rechnet leicht nach:

$$\mathrm{E}\,(X) = \frac{a+b}{2}\ ; \quad \mathrm{Var}\,(X) = \frac{(b-a)^2}{12}\ ; \quad \sigma(X) = \frac{b-a}{\sqrt{12}}\ .$$

Kürzel: Man sagt, X ist „UC(a, b)-verteilt" (uniform, continuous).

MuPAD-Routinen: `stats::uniformCDF`, `stats::uniformQuantile`, `stats::uniformPDF`, `stats::uniformRandom`.

Interpretation: Dies ist das kontinuierliche Analogon zur diskreten „Gleichwahrscheinlichkeit" der kombinatorischen Modelle. Jeder Punkt aus $[a, b]$ wird mit der gleichen W'keit gewählt.

Beispiel: Drehen eines Glücksrads. Die durch den Winkel $\in [0, 2\,\pi)$ gegen eine fixierte Markierung beschriebene Endstellung ist gleichverteilt.

Bemerkung 2.42: *In Softwaresystemen steht typischerweise ein Zufallszahlengenerator für gleichverteilte Gleitpunktzahlen zwischen* 0 *und* 1 *zur Verfügung. Sei* $Y : \Omega \mapsto [0, 1]$ *dieser UC(0,1)-Generator. Hiermit kann man sich leicht einen Zufallszahlengenerator bauen, dessen Werte* X *einer beliebigen vorgegebenen Verteilungsfunktion* F *genügen, die natürlich die in Satz 2.9 beschriebenen Eigenschaften haben muss. Man erzeugt dazu eine UC(0,1)-Zahl* Y *und berechnet* *dann (für kontinuierliches F) den Wert $X = F^{-1}(Y)$, indem man die Gleichung $F(X) = Y$ (typischerweise numerisch) löst. Damit hat X in der Tat die Verteilungsfunktion $F_X = F$:*

$$F_X(r) = P(X \leq r) = P(F^{-1}(Y) \leq r) \overset{(F \text{ streng monoton})}{=} P(Y \leq F(r))$$

$$\overset{UC(0,1)}{=} \int_0^{F(r)} d\tilde{r} = F(r) \ .$$

Anmerkung: Die Vorschrift „löse $F(X) = Y$“ gilt, wenn die Verteilung kontinuierlich ist, genauer, wenn F streng monoton wachsend ist und keine Sprünge macht. Dann hat $F(X) = Y$ genau eine Lösung. Für diskrete Verteilungen (F ist eine Treppenfunktion) hat $F(X) = Y$ i.A. keine Lösung, da Y nur mit W'keit 0 einen der diskreten Werte trifft. Die genaue Vorschrift für den allgemeinen Fall ist, X zu bestimmen als:

$$X = \text{Quantil}_F(Y) = \min\{x;\ Y \leq F(x)\},$$

wobei Quantil_F die in Definition 2.16 eingeführte Quantilfunktion der Verteilungsfunktion F ist. Das Minimum existiert (für $Y \in (0,1]$), da F rechtsseitig stetig ist. Wir zeigen nun allgemein, dass X die Verteilungsfunktion F hat. Für $r \in \mathbb{R}$ definiere:

$$A = \{\omega \in \Omega;\ X(\omega) \leq r\} = \{\omega \in \Omega;\ \min\{x; Y(\omega) \leq F(x)\} \leq r\} \ ,$$

$$B = \{\omega \in \Omega;\ Y(\omega) \leq F(r)\} \ .$$

Es gilt zu zeigen, dass $A = B$ gilt.
i) Sei $\omega \in A$. Damit gilt $x_0 \leq r$ für $x_0 = \min\{x; Y(\omega) \leq F(x)\}$ sowie $Y(\omega) \leq F(x_0)$. Da F monton ist, gilt mit $x_0 \leq r$ auch $Y(\omega) \leq F(x_0) \leq F(r)$. Damit gilt $\omega \in B$. Wir haben also $A \subset B$ gezeigt.
ii) Sei $\omega \in B$, also $Y(\omega) \leq F(r)$. Damit liegt r in der Menge $\{x; Y(\omega) \leq F(x)\}$ und es folgt $\min\{x; Y(\omega) \leq F(x)\} \leq r$. Damit gilt $\omega \in A$. Wir haben also $B \subset A$ gezeigt.
Aus $A \subset B$ und $B \subset A$ folgt $A = B$ und damit

$$P(A) = P(X \leq r) = F_X(r) = P(B) = P(Y \leq F(r)) \overset{\text{UC}(0,1)}{=} F(r) \ .$$

2.4.7 Die Normal-(Gauß-)Verteilung

13.6.07↓

Dies ist die wichtigste aller Verteilungen, weil nach den Grenzwertsätzen des Kapitels 4 die Mittelwerte von unabhängigen Wiederholungen *beliebiger* Zufallsexperimente gegen eine Normalverteilung konvergieren. Diese Verteilung kommt daher immer ins Spiel, wenn man das genaue Verhalten (das unterliegende Modell bzw. die Verteilung der Zufallsvariable) gar nicht kennt und durch häufige Wiederholungen statistische Daten ermittelt.

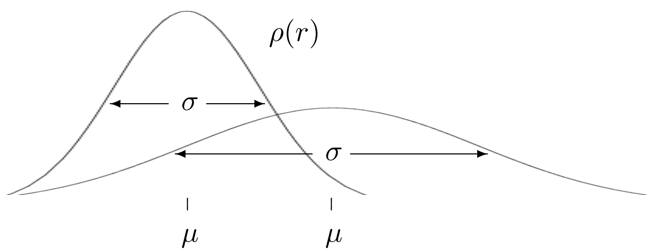
Zu gegebenem $\mu \in \mathbb{R}$ und $\sigma > 0$ betrachte $X : \Omega \to \mathbb{R}$ mit der Dichte

$$\rho(r) = \frac{1}{\sigma\,\sqrt{2\,\pi}}\; e^{-\frac{(r-\mu)^2}{2\,\sigma^2}} \;.$$

Man rechnet nach:

$$\mathrm{E}\,(X) = \mu\ ; \quad \mathrm{Var}\,(X) = \sigma^2\ ; \quad \sigma(X) = \sigma\ .$$

Der Parameter μ ist die Stelle, wo die W'keitsdichte ihr Maximum annimmt. Höhe und



Kürzel: Man sagt, X ist „$\mathrm{N}(\mu, \sigma^2)$-verteilt“.

MuPAD-Routinen: `stats::normalCDF`, `stats::normalQuantile`, `stats::normalPDF`, `stats::normalRandom`.

Interpretation und Beispiele: Die Mittelwerte unabhängiger Wiederholungen beliebiger Zufallsexperimente sind approximativ normalverteilt, wenn man nur genügend oft wiederholt. Details und Beispiele sind mit dem „Zentralen Grenzwertsatz“ (Abschnitt 4.3) beschrieben.

Bemerkung 2.43: *Die Funktion*

$$\mathrm{erf}\,(r) = \frac{2}{\sqrt{\pi}} \int_0^r e^{-x^2}\,dx$$

*heißt „**Fehlerfunktion**“ (engl.: „**error function**“). Sie ist eine Standardfunktion der Mathematik und in vielen Softwarepaketen (z.B. MuPAD) verfügbar. Es gilt der folgende Zusammenhang mit der $N(\mu, \sigma^2)$-Verteilungsfunktion:*

$$\boxed{F_X(r)} = \int_{-\infty}^{r} \frac{1}{\sigma\,\sqrt{2\,\pi}}\; e^{-\frac{(x-\mu)^2}{2\,\sigma^2}}\,dx \overset{(y=\frac{x-\mu}{\sigma\,\sqrt{2}})}{=} \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\frac{r-\mu}{\sigma\,\sqrt{2}}} e^{-y^2}\,dy$$

$$= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{0} e^{-y^2}\,dy + \frac{1}{2}\,\mathrm{erf}\Big(\frac{r-\mu}{\sigma\,\sqrt{2}}\Big) \;\; \boxed{= \frac{1}{2}\Big(1 + \mathrm{erf}\Big(\frac{r-\mu}{\sigma\,\sqrt{2}}\Big)\Big)}\ .$$

Eine weitere interessante Funktion ist

$$\Phi(r) = \frac{1}{\sqrt{2\,\pi}} \int_{-r}^{r} e^{-\frac{x^2}{2}}\, dx = \mathrm{erf}\Big(\frac{r}{\sqrt{2}}\Big) \ .$$

Für eine $N(0,1)$-verteilte Variable X gilt die Interpretation:

$$\Phi(r) = P(-r \le X \le r) \ .$$

Eine Wertetabelle für Φ ist auf Seite 81 beigefügt.

2.5 Unabhängigkeit von Zufallsvariablen

Für die Praxis ist die Frage nach der Unabhängigkeit von Zufallsvariablen sehr wichtig: erhöhen AKWs das Risiko, an Krebs zu erkranken? Besteht ein Zusammenhang zwischen der beruflichen Tätigkeit und den Vorlieben, wo und wie der Urlaub verbracht wird?

2.5.1 Definition und Folgerungen

Das Problem bei solchen Fragestellungen ist meist, dass man den unterliegenden Stichprobenraum nicht wirklich kennt und bestenfalls die Verteilung der interessierenden Zufallsvariablen einzeln durch empirische Messungen approximieren kann (Statistik). Leider ist aus den Verteilungen der Zufallsvariablen die Unabhängigkeit nicht abzulesen: man braucht zusätzliche Informationen, die nur dem den beobachteten Größen unterliegenden Modell $(\Omega, \mathcal{E}, P)$ zu entnehmen sind, die „gemeinsame Verteilung“:

Definition 2.44:

Seien $X_1, \ldots, X_n : \Omega \mapsto \mathbb{R}$ Zufallsvariablen über einem gemeinsamen Modell $(\Omega, \mathcal{E}, P)$. Die Funktion

$$F_{X_1,\ldots,X_n}(r_1, \ldots, r_n) \equiv P(X_1 \le r_1 \ ; \ \ldots \ ; \ X_n \le r_n)$$
$$= P\Big(X_1^{-1}((-\infty, r_1]) \cap \cdots \cap X_n^{-1}((-\infty, r_n])\Big)$$

heißt **gemeinsame Verteilungsfunktion** *der Variablen $X_1, \ldots, X_n$.*

Satz 2.45:

Gemeinsame Verteilungsfunktionen haben allgemein folgende Eigenschaften:

a) *$F_{X_1,\ldots,X_n}(r_1, \ldots, r_n)$ ist monoton steigend und rechtsseitig stetig in jedem r_i (bei fixierten $r_1, \ldots, r_{i-1}, r_{i+1}, \ldots, r_n$).*

Wertetabelle für die $N(0,1)$-W'keiten $P(|X| \leq r) = \Phi(r) = \dfrac{1}{\sqrt{2\,\pi}} \displaystyle\int_{-r}^{r} e^{-\frac{x^2}{2}}\,dx$:

r	$\Phi(r)$	r	$\Phi(r)$	r	$\Phi(r)$	r	$\Phi(r)$	r	$\Phi(r)$
0.00	0.000	0.60	0.451	1.20	0.770	1.80	0.928	2.40	0.9836
0.01	0.008	0.61	0.458	1.21	0.774	1.81	0.930	2.41	0.9840
0.02	0.016	0.62	0.465	1.22	0.778	1.82	0.931	2.42	0.9845
0.03	0.024	0.63	0.471	1.23	0.781	1.83	0.933	2.43	0.9849
0.04	0.032	0.64	0.478	1.24	0.785	1.84	0.934	2.44	0.9853
0.05	0.040	0.65	0.484	1.25	0.789	1.85	0.936	2.45	0.9857
0.06	0.048	0.66	0.491	1.26	0.792	1.86	0.937	2.46	0.9861
0.07	0.056	0.67	0.497	1.27	0.796	1.87	0.939	2.47	0.9865
0.08	0.064	0.68	0.503	1.28	0.799	1.88	0.940	2.48	0.9869
0.09	0.072	0.69	0.510	1.29	0.803	1.89	0.941	2.49	0.9872
0.10	0.080	0.70	0.516	1.30	0.806	1.90	0.943	2.50	0.9876
0.11	0.088	0.71	0.522	1.31	0.810	1.91	0.944	2.51	0.9879
0.12	0.096	0.72	0.528	1.32	0.813	1.92	0.945	2.52	0.9883
0.13	0.103	0.73	0.535	1.33	0.816	1.93	0.946	2.53	0.9886
0.14	0.111	0.74	0.541	1.34	0.820	1.94	0.948	2.54	0.9889
0.15	0.119	0.75	0.547	1.35	0.823	1.95	0.949	2.55	0.9892
0.16	0.127	0.76	0.553	1.36	0.826	1.96	0.950	2.56	0.9895
0.17	0.135	0.77	0.559	1.37	0.829	1.97	0.951	2.57	0.9898
0.18	0.143	0.78	0.565	1.38	0.832	1.98	0.952	2.58	0.9901
0.19	0.151	0.79	0.570	1.39	0.835	1.99	0.953	2.59	0.9904
0.20	0.159	0.80	0.576	1.40	0.838	2.00	$\boxed{0.9545}$	2.60	0.9907
0.21	0.166	0.81	0.582	1.41	0.841	2.01	0.9556	2.61	0.9909
0.22	0.174	0.82	0.588	1.42	0.844	2.02	0.9566	2.62	0.9912
0.23	0.182	0.83	0.593	1.43	0.847	2.03	0.9576	2.63	0.9915
0.24	0.190	0.84	0.599	1.44	0.850	2.04	0.9586	2.64	0.9917
0.25	0.197	0.85	0.605	1.45	0.853	2.05	0.9596	2.65	0.9920
0.26	0.205	0.86	0.610	1.46	0.856	2.06	0.9606	2.66	0.9922
0.27	0.213	0.87	0.616	1.47	0.858	2.07	0.9615	2.67	0.9924
0.28	0.221	0.88	0.621	1.48	0.861	2.08	0.9625	2.68	0.9926
0.29	0.228	0.89	0.627	1.49	0.864	2.09	0.9634	2.69	0.9929
0.30	0.236	0.90	0.632	1.50	0.866	2.10	0.9643	2.70	0.9931
0.31	0.243	0.91	0.637	1.51	0.869	2.11	0.9651	2.72	0.9935
0.32	0.251	0.92	0.642	1.52	0.871	2.12	0.9660	2.74	0.9939
0.33	0.259	0.93	0.648	1.53	0.874	2.13	0.9668	2.76	0.9942
0.34	0.266	0.94	0.653	1.54	0.876	2.14	0.9676	2.78	0.9946
0.35	0.274	0.95	0.658	1.55	0.879	2.15	0.9684	2.80	0.9949
0.36	0.281	0.96	0.663	1.56	0.881	2.16	0.9692	2.82	0.9952
0.37	0.289	0.97	0.668	1.57	0.884	2.17	0.9700	2.84	0.9955
0.38	0.296	0.98	0.673	1.58	0.886	2.18	0.9707	2.86	0.9958
0.39	0.303	0.99	0.678	1.59	0.888	2.19	0.9715	2.88	0.9960
0.40	0.311	1.00	$\boxed{0.683}$	1.60	0.890	2.20	0.9722	2.90	0.9963
0.41	0.318	1.01	0.688	1.61	0.893	2.21	0.9729	2.92	0.9965
0.42	0.326	1.02	0.692	1.62	0.895	2.22	0.9736	2.94	0.9967
0.43	0.333	1.03	0.697	1.63	0.897	2.23	0.9743	2.96	0.9969
0.44	0.340	1.04	0.702	1.64	0.899	2.24	0.9749	2.98	0.9971
0.45	0.347	1.05	0.706	1.65	0.901	2.25	0.9756	3.00	$\boxed{0.99730}$
0.46	0.354	1.06	0.711	1.66	0.903	2.26	0.9762	3.10	0.99806
0.47	0.362	1.07	0.715	1.67	0.905	2.27	0.9768	3.20	0.99863
0.48	0.369	1.08	0.720	1.68	0.907	2.28	0.9774	3.30	0.99903
0.49	0.376	1.09	0.724	1.69	0.909	2.29	0.9780	3.40	0.99933
0.50	0.383	1.10	0.729	1.70	0.911	2.30	0.9786	3.60	0.99968
0.51	0.390	1.11	0.733	1.71	0.913	2.31	0.9791	3.50	0.99953
0.52	0.397	1.12	0.737	1.72	0.915	2.32	0.9797	3.70	0.99978
0.53	0.404	1.13	0.742	1.73	0.916	2.33	0.9802	3.80	0.99986
0.54	0.411	1.14	0.746	1.74	0.918	2.34	0.9807	3.90	0.99990
0.55	0.418	1.15	0.750	1.75	0.920	2.35	0.9812	4.00	0.999937
0.56	0.425	1.16	0.754	1.76	0.922	2.36	0.9817	5.00	0.999999
0.57	0.431	1.17	0.758	1.77	0.923	2.37	0.9822		
0.58	0.438	1.18	0.762	1.78	0.925	2.38	0.9827		
0.59	0.445	1.19	0.766	1.79	0.927	2.39	0.9832		

b) $F_{X_1,\ldots,X_n}(r_1,\ldots,\underbrace{-\infty}_{r_i},\ldots r_n) = 0$.

c) $F_{X_1,\ldots,X_n}(r_1,\ldots,\overbrace{\infty},\ldots r_n) =$

$$F_{X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_n}(r_1,\ldots,r_{i-1},r_{i+1},\ldots r_n) \ .$$

Beweisskizze:
a) folgt analog zur Beweisskizze zu Satz 2.9.
b) ergibt sich aus

$$\lim_{r_i \to -\infty} P(X_1 \leq r_1 \ ; \ \ldots \ ; \ X_i \leq r_i \ ; \ \ldots \ ; \ X_n \leq r_n)$$

$$= P\Big(\bigcap_{r_i \in \mathbb{R}} \Big\{ \omega \in \Omega; \ X_1(\omega) \leq r_1 \ ; \ \ldots, X_i(\omega) \leq r_i \ ; \ \ldots, X_n(\omega) \leq r_n \Big\} \Big)$$

$$= P(\emptyset) = 0 \ .$$

c) ergibt sich aus

$$\lim_{r_i \to \infty} P(X_1 \leq r_1 \ ; \ \ldots \ ; \ X_i \leq r_i \ ; \ \ldots \ ; \ X_n \leq r_n)$$

$$= P\Big(\bigcup_{r_i \in \mathbb{R}} \Big\{ \omega \in \Omega; \ X_1(\omega) \leq r_1 \ ; \ \ldots \ ; \ X_i(\omega) \leq r_i \ ; \ \ldots \ ; \ X_n(\omega) \leq r_n \Big\} \Big)$$

$$= P\Big(\{ \omega \in \Omega; \ X_1(\omega) \leq r_1 \ ; \ \ldots \ ; \ \xcancel{X_i(\omega) \in \mathbb{R}} \ ; \ \ldots \ ; \ X_n(\omega) \leq r_n \} \Big) \ .$$

Q.E.D.

Definition 2.46:

Zufallsvariablen $X_1,\ldots,X_n : \Omega \mapsto \mathbb{R}$ über einem gemeinsamen Modell $(\Omega,\mathcal{E},P)$ heißen **unabhängig**, *wenn für jede Wahl von $r_1,\ldots,r_n$ die Ereignisse*

$$X_1^{-1}((-\infty,r_1]) \ , \ \ldots \ , \ X_n^{-1}((-\infty,r_n])$$

eine im Sinne von Definition 1.54 unabhängige Ereignisfamilie bilden.

Bemerkung 2.47: (Warnung) *Gemäß der Warnung 1.55 gilt wiederum: ist eine Familie von Variablen $X_1,\ldots,X_n$ unabhängig, so sind auch jeweils Paare von Variablen unabhängig. Es gilt aber nicht die Umkehrung: auch wenn die Zufallsvariablen paarweise unabhängig sind, so braucht die Gesamtfamilie nicht unabhängig zu sein.*

Intuitiv hätte man erwartet, dass Unabhängigkeit von Zufallsvariablen dadurch definiert wird, dass die Ereignisse $X_1 \in E_1$, $X_2 \in E_2$ etc. für *beliebiges* $E_1, \ldots, E_n \subset \mathbb{R}$ unabhängig sein sollen. Die obige Definition beschränkt sich auf die speziellen Ereignisse $E_i = (-\infty, r_i]$. Dies genügt, denn daraus folgt die Unabhängigkeit für beliebige E_i:

Satz 2.48:

Die Zufallsvariablen $X_1, \ldots, X_n$ sind genau dann unabhängig, wenn die gemeinsame Verteilung das Produkt der einzelnen Verteilungen ist:

$$F_{X_1,\ldots,X_n}(r_1, \ldots, r_n) = F_{X_1}(r_1) \cdot \ \ldots \ \cdot F_{X_n}(r_n) \ .$$

Dies ist genau dann der Fall, wenn für beliebige Ereignisse $E_1, \ldots, E_n \subset \mathbb{R}$ gilt:

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

Beweisskizze:
Aus der Definition 1.54 folgt unmittelbar, dass die gemeinsame Verteilung für unabhängige Variablen faktorisieren muss. Andererseits, wenn sie faktorisiert, so gilt mit Satz 2.45.c) und $F_{X_i}(\infty) = 1$ auch, dass die gemeinsame Verteilung jeder Teilauswahl von Variablen aus $X_1, \ldots, X_n$ faktorisiert, womit nach Definition auch für jede Teilfamilie der Ereignisse

$$\mathcal{A} = \Big\{ X_1^{-1}((-\infty, r_1]) \ , \ \ldots \ , X_n^{-1}((-\infty, r_n]) \Big\}$$

die Unabhängigkeit

$$P(X_{i_1} \le r_{i_1} \ ; \ \ldots \ ; \ X_{i_k} \le r_{i_k}) = P(X_{i_1} \le r_{i_1}) \cdot \ \ldots \ \cdot P(X_{i_k} \le r_{i_k})$$

gilt. Die Unabhängigkeit der Variablen sei nun vorausgesetzt, d.h., die Produktformel gilt für die speziellen Ereignisse „$X_1 \le r_1 \ ; \ \ldots \ ; \ X_n \le r_n$“. Wir zeigen hier nur, dass sich dies verallgemeinert auf die “typischen” Ereignisse $E_i = (a_i, b_i]$ (und Schnitte und Vereinigungen solcher Intervalle). Sei dazu zunächst

$$E_1 = (a_1, b_1] \ , E_2 = (-\infty, r_1] \ , \ \ldots \ , \ E_n = (-\infty, r_n] \ .$$

Aus der Definition der gemeinsamen Verteilung folgt unmittelbar:

$$\begin{aligned} &P(X_1 \in E_1 \ ; \ \ldots \ ; \ X_n \in E_n) \\ &\quad = F_{X_1,\ldots,X_n}(b_1, r_2, \ldots, r_n) - F_{X_1,\ldots,X_n}(a_1, r_2, \ldots, r_n) \\ &\quad = \Big(F_{X_1}(b_1) - F_{X_1}(a_1) \Big) \ \cdot \ F_{X_2}(r_2) \cdot \ \ldots \ \cdot F_{X_n}(r_n) \\ &\quad = P(a_1 < X_1 \le b_1) \ \cdot \ P(X_2 \le r_2 \ ; \ \cdots \ ; \ X_n \le r_n) \ . \end{aligned}$$

Es ist ziemlich offensichtlich, wie man mit analogen Argumenten die Produktformel auch für allgemeineres E_1 zeigen kann, das sich durch Schnitte oder Vereinigungen solcher Intervalle ergibt. Hat man erst einmal die Produktformel für allgemeine Ereignisse $E_1 \subset \mathbb{R}$ und die speziellen Ereignisse $E_2 = (-\infty, r_2]$, $\ldots$, $E_n = (-\infty, r_n]$ gezeigt, also

$$P(X \in E_1 \ ; \ X_2 \leq r_2 \ ; \ \ldots \ ; \ X_n \leq r_n)$$

$$= P(X \in E_1) \ \cdot \ P(X_2 \leq r_2 \ ; \ \ldots \ ; \ X_n \leq r_n) \ ,$$

so folgt rekursiv dann auch sofort die vollständige Produktzerlegung für beliebige Ereignisse $X_2 \in E_2 \subset \mathbb{R}_2, \ldots, X_n \in E_n \subset \mathbb{R}_n$.

Q.E.D.

Die folgende Bemerkung liefert ein leicht zu handhabendes Kriterium, mit dem bei diskreten Zufallsvariablen die Unabhängikeit nachgerechnet werden kann:

18.6.07↓ **Bemerkung 2.49:** *Nach Satz 2.48 bedeutet Unabhängigkeit bei diskreten Zufallsvariablen $X : \Omega \mapsto \{x_1, x_2, \ldots\}$, $Y : \Omega \mapsto \{y_1, y_2, \ldots\}$, dass*

$$P(X = x_i \ ; \ Y = y_j) \ = \ P(X = x_i) \cdot P(Y = y_j)$$

für alle i, j gilt. In der Tat ist dies ein notwendiges und hinreichendes Kriterum für die Unabhängigkeit diskreter Variabler.

Beispiel 2.50: Betrachte den 2-fachen Wurf $\Omega = \{(\omega_1, \omega_2); \ \omega_1, \omega_2 \in \{1, \ldots, 6\}\}$ eines fairen Würfels, sei X = Ergebnis des ersten Wurfs, Y = Ergebnis des zweiten Wurfs. Es gilt für alle Paare $i, j \in \{1, \ldots, 6\}$:

$$P(X = i \ ; \ Y = j) \ = \ P\Big(\{(i, j)\}\Big) = \frac{1}{36} \ ,$$

$$P(X = i) \ = \ P\Big(\{(i, \omega_2); \ \omega_2 \in \{1, \ldots, 6\}\}\Big) = \frac{1}{6} \ ,$$

$$P(Y = j) \ = \ P\Big(\{(\omega_1, j); \ \omega_1 \in \{1, \ldots, 6\}\}\Big) = \frac{1}{6} \ .$$

Für alle Paare i, j gilt damit $P(X = i \ ; \ Y = j) = 1/36 = P(X = i) \cdot P(Y = j)$, womit X und Y unabhängig sind.

Anmerkung: dies ist ein einfaches Beispiel für die in Bemerkung 2.54 formalisierte Situation, wo die Zufallsvariablen sich auf unabhängige Durchführungen/Wiederholungen eines Zufallsexperiments (hier: einfacher Wurf des Würfels) beziehen. Hätte sich hier Abhängigkeit zwischen X und Y ergeben, wäre die angegebene Definition von Unabhängigkeit von Zufallsvariablen wenig sinnvoll gewesen.

Beispiel 2.51:
Man zieht zufällig ein Wort aus der nebenstehenden Urne. Sei X die Anzahl der Buchstaben im gezogenen Wort, sei Y die Anzahl der Vokale im gezogenen Wort. Sind X und Y unabhängig?

Der Zufall regiert die Welt

Antwort: Für

$$\begin{array}{rlcl} X: & \{\text{'Der', 'Zufall', 'regiert', 'die', 'Welt'}\} & \mapsto & \{3,4,6,7\}, \\ Y: & \{\text{'Der', 'Zufall', 'regiert', 'die', 'Welt'}\} & \mapsto & \{1,2,3\} \end{array}$$

ergibt sich folgende gemeinsame Verteilung:

	$Y=1$	$Y=2$	$Y=3$	Σ
$X=3$	$\frac{1}{5}$	$\frac{1}{5}$	0	$\frac{2}{5}=P(X=3)$
$X=4$	$\frac{1}{5}$	0	0	$\frac{1}{5}=P(X=4)$
$X=6$	0	$\frac{1}{5}$	0	$\frac{1}{5}=P(X=6)$
$X=7$	0	0	$\frac{1}{5}$	$\frac{1}{5}=P(X=7)$
Σ	$\underbrace{\frac{2}{5}}_{P(Y=1)}$	$\underbrace{\frac{2}{5}}_{P(Y=2)}$	$\underbrace{\frac{1}{5}}_{P(Y=3)}$	1

In dieser Tabelle sind die W'keiten $P(X=i;Y=j)$ aufgelistet, z. B.:

$$P(X=3;Y=1)=P(\{\text{'Der'}\})=\frac{1}{5}, \quad P(X=3;Y=2)=P(\{\text{'die'}\})=\frac{1}{5}$$

usw. Es gilt z.B.

$$P(X=3;Y=3)=0 \quad \neq \quad P(X=3)\cdot P(Y=3)=\frac{2}{5}\cdot\frac{1}{5},$$

damit sind die Variablen abhängig.

Der folgende Sachverhalt liefert zwar kein handhabbares Kriterum, um Unabhängigkeit zu testen, ist aber wegen des folgenden Spezialfalls 2.53 interessant:

Bemerkung 2.52: *Die Zufallsvariablen $X_1,\ldots,X_n$ sind genau dann unabhängig, wenn für alle (glatten) Funktion $f_1,\ldots,f_n:\mathbb{R}\mapsto\mathbb{R}$ gilt:*

$$\mathrm{E}\Big(f_1(X_1)\cdot\ \ldots\ \cdot f_n(X_n)\Big)=\mathrm{E}\left(f_1(X_1)\right)\cdot\ \ldots\ \cdot\mathrm{E}\left(f_n(X_n)\right).$$

Beweisidee:
Man braucht hierzu das Konzept mehrdimensionaler Stieltjes-Integration

$$\mathrm{E}\left(f(X_1,\ldots,X_n)\right) = \int_{\mathbb{R}} \ldots \int_{\mathbb{R}} f(r_1,\ldots,r_n)\, dF_{X_1,\ldots,X_n}(r_1,\ldots,r_n)$$

für Funktionen $f : \mathbb{R}^n \mapsto \mathbb{R}$. Wenn $F_{X_1,\ldots,X_n}$ und f faktorisieren, so faktorisiert auch das Integral über den „Satz von Fubini“.

Q.E.D.

Als Spezialfall folgt:

Satz 2.53: (Unabhängigkeit ⇒ Erwartungswerte faktorisieren)
Für unabhängige Zufallsvariable $X_1,\ldots,X_n : \Omega \to \mathbb{R}$ gilt:

$$\mathrm{E}\left(X_1 \cdot\ \ldots\ \cdot X_n\right) = \mathrm{E}\left(X_1\right) \cdot\ \ldots\ \cdot \mathrm{E}\left(X_n\right) .$$

Beweisskizze:
Mangels Technik (mehrdimensionale Integration) führen wir den Beweis nur für diskrete Variable. Seien dazu $X_i(\Omega) = \{x_{i1}, x_{i2}, \ldots\}$. Die Produktvariable

$$Y = X_1 \cdot\ \ldots\ \cdot X_n :\ \Omega \to \{x_{1i_1} \cdot\ \ldots\ \cdot x_{ni_n};\ i_1,\ldots,i_n \in \{1,2,\ldots\}\}$$

ist wieder diskret. Für den diskreten Spezialfall 2.25 gilt

$$\begin{aligned}
\mathrm{E}\left(Y\right) &= \sum_{y \in Y(\Omega)} y\ P(Y = y) \\
&= \sum_{i_1} \cdots \sum_{i_n} x_{1i_1} \cdot\ \cdots\ \cdot x_{ni_n}\ P(X_1 = x_{1i_1};\ldots;X_n = x_{ni_n}) \\
&= \sum_{i_1} \cdots \sum_{i_n} x_{1i_1} \cdot\ \cdots\ \cdot x_{ni_n}\ P(X_1 = x_{1i_1}) \cdot\ \cdots\ \cdot P(X_n = x_{ni_n}) \\
&= \Big(\sum_{i_1} x_{1i_1} P(X_1 = x_{1i_1})\Big) \cdot\ \ldots\ \cdot \Big(\sum_{i_n} x_{ni_n} P(X_n = x_{ni_n})\Big) \\
&= \mathrm{E}\left(X_1\right) \cdot\ \cdots\ \cdot \mathrm{E}\left(X_n\right) .
\end{aligned}$$

Q.E.D.

20.6.07↓

Es gibt eine natürliche Situation, in der unabhängige Variablen auftreten, nämlich bei unabhängig durchgeführten Zufallsexperimenten, denen jeweils getrennt eine Zufallsvariable zugeordnet wird. Im entsprechenden Produktmodell 1.57 ergibt sich folgendes Bild:

Bemerkung 2.54: *Seien $X_i : \Omega_i \to \mathbb{R}$ $(i = 1, \ldots, n)$ Zufallsvariablen über (eventuell verschiedenen) Modellen $(\Omega_i, \mathcal{E}_i, P_i)$. Betrachte das Produktmodell*

$$\Omega = \Omega_1 \times \cdots \times \Omega_n = \{(\omega_1, \ldots, \omega_n);\ \omega_i \in \Omega_i\}$$

gemäß Definition 1.57 mit dem W'keitsmaß

$$P(E_1 \times \cdots \times E_n) = P_1(E_1) \cdot \ \ldots \ \cdot P_n(E_n)$$

für Ereignisse $E_i \in \mathcal{E}_i$. Die durch

$$\hat{X}_i : (\omega_1, \ldots, \omega_n) \ \to \ X_i(\omega_i)$$

definierten Zufallsvariablen sind unabhängig.

Beweis: Mit

$$\begin{aligned}
&P(\hat{X}_1 \leq r_1 \ ; \ \ldots \ ; \ \hat{X}_n \leq r_n) \\
&\quad = P\Big(\{(\omega_1, \ldots, \omega_n);\ X_1(\omega_1) \leq r_1 \ ; \ \ldots \ ; \ X_n(\omega_n) \leq r_n\}\Big) \\
&\quad = P\Big(X_1^{-1}((-\infty, r_1]) \times \cdots \times X_n^{-1}((-\infty, r_n])\Big) \\
&\quad = P_1(X_1 \leq r_1) \cdot \ \ldots \ \cdot P_n(X_n \leq r_n) \\
&\quad = P(\hat{X}_1 \leq r_1) \cdot \ \ldots \ \cdot P(\hat{X}_n \leq r_n)
\end{aligned}$$

faktorisiert die Verteilungsfunktion.

Q.E.D.

Beispiel 2.55: Betrachte unabhängige Würfe eines Würfels. Sei X_1 das Ergebnis des ersten Wurfs, X_2 das Quadrat des Ergebnisses des zweiten Wurfs, X_3 irgendeine Funktion des Ergebnisses des dritten Wurfs usw. In diesem Fall kann man sofort sagen, dass die X_i unabhängig sind.

2.5.2 Kovarianz

Fassen wir zusammen: für beliebige Zufallsvariable $X_1, \ldots, X_n$ gilt die Linearität des Erwartungswerts

$$\mathrm{E}\,(X_1 + \cdots + X_n) = \mathrm{E}\,(X_1) + \cdots + \mathrm{E}\,(X_n)$$

(egal, ob die X_i unabhängig sind oder nicht). Für die Varianz gilt jedoch

$$\sigma^2(X_1 + \cdots + X_n) = \mathrm{E}\left((\textstyle\sum_i X_i)^2\right) - \left(\mathrm{E}\,(\textstyle\sum_i X_i)\right)^2$$

$$= \mathrm{E}\left((\textstyle\sum_i X_i)(\textstyle\sum_j X_j)\right) - \left(\textstyle\sum_i \mathrm{E}\,(X_i)\right)\left(\textstyle\sum_j \mathrm{E}\,(X_j)\right)$$

$$= \sum_{i,j} \Big(\mathrm{E}\,(X_i\,X_j) - \mathrm{E}\,(X_i)\,\mathrm{E}\,(X_j) \Big)$$

$$= \sum_i \Big(\mathrm{E}\,(X_i^2) - \mathrm{E}\,(X_i)^2 \Big) + \sum_{\substack{i,j \\ i \neq j}} \Big(\mathrm{E}\,(X_i\,X_j) - \mathrm{E}\,(X_i)\,\mathrm{E}\,(X_j) \Big)$$

$$= \sum_i \sigma^2(X_i) + \sum_{\substack{i,j \\ i \neq j}} \Big(\mathrm{E}\,(X_i\,X_j) - \mathrm{E}\,(X_i)\,\mathrm{E}\,(X_j) \Big) \ .$$

Varianzen addieren sich also im allgemeinen nicht. Die obige Betrachtung führt zu:

Definition 2.56:

Für die Zufallsvariablen $X_1, \ldots, X_n$ nennt man die Matrix mit den Einträgen

$$C_{ij} = \mathrm{E}\,(X_i\,X_j) - \mathrm{E}\,(X_i)\,\mathrm{E}\,(X_j) = \mathrm{E}\,\Big((X_i - \mathrm{E}\,(X_i))\,(X_j - \mathrm{E}\,(X_j)) \Big)$$

„**Kovarianzmatrix**“ *der Variablen.*

Die Einträge $i \neq j$ der Kovarianzmatrix C_{ij} verschwinden für *paarweise unabhängige* Variablen. In diesem Fall gilt

$$\mathrm{E}\,\Big((X_i - \mathrm{E}\,(X_i))\,(X_j - \mathrm{E}\,(X_j)) \Big) = \mathrm{E}\,\Big(X_i - \mathrm{E}\,(X_i) \Big) \cdot \mathrm{E}\,\Big(X_j - \mathrm{E}\,(X_j) \Big)$$

$$= \Big(\mathrm{E}\,(X_i) - \mathrm{E}\,\Big(\mathrm{E}\,(X_i) \Big) \Big) \cdot \Big(\mathrm{E}\,(X_j) - \mathrm{E}\,\Big(\mathrm{E}\,(X_j) \Big) \Big)$$

$$= \Big(\mathrm{E}\,(X_i) - \mathrm{E}\,(X_i) \Big) \cdot \Big(\mathrm{E}\,(X_j) - \mathrm{E}\,(X_j) \Big) = 0.$$

Merke 2.57:

Für paarweise unabhängige Variablen addieren sich neben den Erwartungswerten auch die Varianzen:

$$\sigma^2(X_1 + \cdots + X_n) = \sigma^2(X_1) + \cdots + \sigma^2(X_n) \ .$$

2.6 Bedingte Erwartungswerte

Für Verteilungsfunktionen und Erwartungswerte gibt es das Analogon zu bedingten W'keiten:

Definition 2.58: (Bedingte Erwartung)

Sei $A \subset \mathcal{E}$ ein Ereignis in einem Modell $(\Omega, \mathcal{E}, P)$, sei $X : \Omega \mapsto \mathbb{R}$ eine Zufallsvariable. Dann heißt

$$F_X(r \,|\, A) = P(X^{-1}((-\infty, r]) \,|\, A) \equiv P(X \le r \,|\, A)$$

die **bedingte Verteilung** *von X* **bei gegebenem Ereignis** *A. Sei $f : \mathbb{R} \mapsto \mathbb{R}$ eine glatte Funktion. Man nennt*

$$\mathrm{E}\,(f(X) \,|\, A) = \int_{\mathbb{R}} f(r)\ dF_X(r \,|\, A)$$

den **bedingten Erwartungswert** *von $f(X)$* **bei gegebenem** *A.*

Im diskreten Fall $X(\Omega) = \{r_1, r_2, \dots\}$ ist dies:

$$\mathrm{E}\,(f(X) \,|\, A) = \sum_{r \in X(\Omega)} f(r)\ P(X = r \,|\, A)\ .$$

Im kontinuierlichen Fall ist dies:

$$\mathrm{E}\,(f(X) \,|\, A) = \int_{r \in X(\Omega)} f(r)\ \rho_A(r)\ dr,$$

wo $\rho_A(r) = \frac{d}{dr}\, F_X(r \,|\, A)$ die Dichte der bedingten Verteilung ist.

In Analogie zum Satz von der totalen W'keit 1.31 gilt:

Satz 2.59:

Sei $U_1, \dots, U_n$ eine disjunkte Zerlegung des Stichprobenraums $\Omega = \cup_i U_i$ eines beliebigen Modells $(\Omega, \mathcal{E}, P)$. Dann gilt für jede Zufallsvariable $X : \Omega \mapsto \mathbb{R}$ und jede glatte Funktion $f : \mathbb{R} \mapsto \mathbb{R}$:

$$\mathrm{E}\,(f(X)) = \sum_{i=1}^{n} \mathrm{E}\,(f(X) \,|\, U_i)\ P(U_i)\ .$$

Beweis: Die Formel von der totalen W'keit 1.31 liefert:

$$F_X(r) = P(X \le r) = \sum_{i=1}^{n} P(X \le r \,|\, U_i)\ P(U_i) = \sum_{i=1}^{n} F_X(r \,|\, U_i)\ P(U_i)\ .$$

Aus der Definition über Stieltjes-Summen 2.18 ist unmittelbar klar, dass Stieltjes-Integrale linear in der Maßfunktion sind. Mit $\alpha, \beta \in \mathbb{R}$ gilt für $F = \alpha\, F_1 + \beta\, F_2$:

$$\int_{(a,b]} f(r)\ dF(r) = \alpha \int_{(a,b]} f(r)\ dF_1(r) + \beta \int_{(a,b]} f(r)\ dF_2(r)\ .$$

Mit der Summendarstellung von $F_X(r)$ folgt unmittelbar die Behauptung.

Q.E.D.

Diese Formel eignet sich bei Hintereinanderschaltung von Experimenten. Ein erstes Experiment liefert das Ursachensystem U_i für ein zweites Experiment:

Beispiel 2.60: Gegeben sei ein Topf mit 70 fairen Münzen und 30 Münzen mit 2 Köpfen. Man entnimmt dem Topf eine Münze und wirft 1000 mal. Wieviele Köpfe wird man im Mittel werfen?
Lösung: Seien $U_1 =$ „ich habe eine faire Münze gezogen“, $U_2 =$ „ich habe eine unfaire Münze gezogen“. Nach der Auswahl erfolgt die $n = 1000$-fache Wiederholung eines Bernoulli-Experiments mit dem Erwartungswert $n \cdot P($„Kopf“$)$. Für $X =$ „die Anzahl der geworfenen Köpfe bei 1000 Würfen“ gilt

$$\mathrm{E}\,(X \,|\, U_1) = 500\ , \quad \mathrm{E}\,(X \,|\, U_2) = 1000\ , \quad P(U_1) = \frac{70}{100}\ , \quad P(U_2) = \frac{30}{100}$$

und damit

$$\mathrm{E}\,(X) = \mathrm{E}\,(X \,|\, U_1)\, P(U_1) + \mathrm{E}\,(X \,|\, U_1)\, P(U_2) = \frac{70}{100} \cdot 500 + \frac{30}{100} \cdot 1000 = 650\ .$$
