Stochastik II

Dr. Miriam Dieter

Sommersemester 2015

Stand: 8. Mai 2015

Inhaltsverzeichnis

1 Diskrete Zufallsvariablen

1.1 Diskrete Wahrscheinlichkeitsräume

Die Wahrscheinlichkeitstheorie liefert theoretische Grundlagen für die Statistik. Wir beginnen unsere Betrachtungen dieses Bereiches mit der Beschreibung von Experimenten, deren Ausgang wir als zufällig ansehen. Die grundlegende Idee ist, ein geeignetes Modell für ein solches Experiment zu finden.

Definition 1.1 *Ein* **Zufallsexperiment** *ist ein Experiment, dessen Ausgang nicht deterministisch ist und das daher bei wiederholter Durchführung verschiedene Ergebnisse liefern kann. Die Menge aller möglichen Ergebnisse bezeichnet man mit Ω. Ω heißt* **Ergebnismenge** *oder* **Ergebnisraum***. Seine Elemente $\omega \in \Omega$ heißen* **Elementarereignisse***, während die Teilmengen $A \subset \Omega$* **Ereignisse** *genannt werden. Die Gesamtheit aller Teilmengen von Ω nennt man* **Potenzmenge***; im Folgenden bezeichnen wir diese mit $\mathcal{P}(\Omega)$.*

Bemerkung 1.2 *Betrachtet man zwei Ereignisse A und B des Ergebnisraumes, so werden folgende neue Ereignisse entsprechend mengentheoretisch dargestellt und bezeichnet:*

- A *oder* B *treten ein:* $A \cup B$
- A *und* B *treten gleichzeitig ein:* $A \cap B$
- A *tritt* **nicht** *ein:* $\overline{A} = \Omega \setminus A$
- A *tritt ein,* **aber nicht** B*:* $A \setminus B$

Ω *und die leere Menge $\emptyset$ werden auch als Ereignisse betrachtet. Sie heißen dann:*

- Ω*: das* **sichere Ereignis**
- $\emptyset$*: das* **unmögliche Ereignis**

Zwei Ereignisse A und B **schließen sich gegenseitig aus***, wenn $A \cap B = \emptyset$, d.h. wenn sie disjunkt sind. Betrachtet man mehr als zwei Ereignisse $A_1, \ldots, A_k$, so* **schließen sich** *die Ereignisse* **paarweise aus***, wenn sie paarweise disjunkt sind, also $A_i \cap A_j = \emptyset$ für $i \neq j$.*

Definition 1.3 *Ein Wahrscheinlichkeitsraum ist ein Dreitupel $(\Omega, \mathcal{A}, P)$ bestehend aus*

(a) einem (nicht-leeren) Ergebnisraum Ω

(b) der Familie $\mathcal{A}$ aller möglichen Ereignisse und

(c) einem **Wahrscheinlichkeitsmaß** *P, d.h. einer Abbildung $P : \mathcal{A} \to [0, 1]$ mit*

$$P(\Omega) = 1 \text{ und}$$
$$P\left(\biguplus_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \text{ für } (A_i)_{i\in\mathbb{N}} \subset \mathcal{A} \text{ paarweise disjunkt}$$

Die Eigenschaften des Wahrscheinlichkeitsmaßes sind bekannt als **Kolmogorow-Axiome** *und bilden seit ihrer Veröffentlichung 1933 den theoretischen Grundstein der Wahrscheinlichkeitstheorie.*

Bemerkungen 1.4 *(a) Ist Ω endlich oder abzählbar unendlich, so ist die Familie aller Teilmengen von Ω durch die Potenzmenge gegeben, somit gilt $\mathcal{A} = \mathcal{P}(\Omega)$. Den Wahrscheinlichkeitsraum nennt man dann* **diskreten Wahrscheinlichkeitsraum***.*

(b) Ist Ω überabzählbar groß, so muss ein allgemeineres Mengensystem verwendet werden, die sogenannte σ-Algebra. Auch die Potenzmenge ist eine σ-Algebra. Darauf gehen wir in Kapitel 3 näher ein.

Beispiel 1.5 *(Einmaliger Würfelwurf). Das einfachste Modell für einen einmaligen Wurf mit einem fairen Würfel ist*

$$\Omega = \{1, 2, 3, 4, 5, 6\},$$

$$\mathcal{A} = \mathcal{P}(\Omega) = \{\emptyset, \{1\}, \{2\}, \dots, \{1, 2\}, \dots, \{1, 2, 3, 4, 5, 6\}\}.$$

Da der Würfel als fair vorausgesetzt wird, sollten alle Elementarereignisse gleich wahrscheinlich sein und somit gilt

$$P(\{1\}) = \ldots = P(\{6\}) = \frac{1}{6}.$$

Die Wahrscheinlichkeiten für die anderen Ereignisse lassen sich dann aus 1.3(c) gewinnen. Es sei z.B. A das Ereignis, dass die Augenzahl eine gerade Zahl ist. Dann gilt

$$A = \{2, 4, 6\} = \{2\} \uplus \{4\} \uplus \{6\}$$

und somit

$$P(A) = P(\{2\} \uplus \{4\} \uplus \{6\}) = P(\{2\}) + P(\{4\}) + P(\{6\}) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}.$$

Aus den Axiomen von Kolmogorov lassen sich weitere Rechenregeln für das Wahrscheinlichkeitsmaß herleiten.

Satz 1.6 *(Rechenregeln für Wahrscheinlichkeitsmaße) Ist $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, so gelten für Ereignisse $A, B \in \mathcal{A}$:*

(a) $P(\overline{A}) = 1 - P(A)$

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

(c) $P(B \setminus A) = P(B) - P(B \cap A)$
Ist $A \subset B$, so gilt: $P(B \setminus A) = P(B) - P(A)$

(d) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Beweis:
zu (a): Wegen 1.3(c) gilt

$$\begin{aligned} 1 &= P(\Omega) \\ &= P(A \uplus \overline{A}) \\ &= P(A) + P(\overline{A}) \\ \Rightarrow \quad P(\overline{A}) &= 1 - P(A) \end{aligned}$$

zu (b): Setzen wir in (a) $A = \Omega$, dann ist $\overline{A} = \overline{\Omega} = \emptyset$ und wir erhalten sofort

$$P(\emptyset) = 1 - P(\Omega) = 0.$$

zu (c): B lässt sich schreiben als

$$B = B \cap \Omega = B \cap (A \uplus \overline{A}) = (B \cap A) \uplus (B \cap \overline{A}).$$

Also ist

$$P(B) = P((B \cap A) \uplus (B \cap \overline{A})) = P(B \cap A) + P(B \cap \overline{A}).$$

Wegen $B \cap \overline{A} = B \setminus A$ folgt

$$P(B \setminus A) = P(B \cap \overline{A}) = P(B) - P(A \cap B).$$

zu (d): Dafür verwenden wir Teil (c) wie folgt:

$$\begin{aligned} P(A \cup B) &= P(A \uplus (B \setminus A)) \\ &= P(A) + P(B \setminus A) \\ &= P(A) + P(B) - P(A \cap B). \end{aligned}$$

Beispiel 1.7 *(Einmaliger Würfelwurf) Uns interessieren die Wahrscheinlichkeiten der folgenden Ereignisse beim einmaligen Würfelwurf (vgl. Beispiel 1.5):*

A_1: *Die geworfene Zahl ist gerade*

A_2: *Die geworfene Zahl ist ungerade*

A_3: *Die geworfene Zahl ist eine ungerade Zahl und größer als 1*

A_4: *Die geworfene Zahl ist eine gerade Zahl oder größer als 2*

Auch wenn es in diesem Fall einfachere Möglichkeiten gibt, die Wahrscheinlichkeiten dieser vier Ereignisse zu berechnen, wollen wir hier die Aussagen aus Satz 1.6 verwenden.

Das erste Ereignis A_1 "die geworfene Zahl ist gerade" wird beschrieben durch $A_1 = \{2,4,6\}$. Aus Beispiel 1.5 wissen wir bereits, dass gilt

$$P(A_1) = \frac{1}{2}.$$

Für das Ereignis A_2 "die geworfene Zahl ist ungerade" ist $A_2 = \{1,3,5\}$. Wir stellen fest, dass gilt: $A_2 = \overline{A_1}$. Somit können wir 1.6(a) benutzen und erhalten

$$P(A_2) = P(\overline{A_1}) = 1 - P(A_1) = 1 - \frac{1}{2} = \frac{1}{2}.$$

Für das Ereignis A_3 "die geworfene Zahl ist eine ungerade Zahl und größer als 1" ist $A_3 = \{3,5\}$. Zusätzlich benötigen wir noch $B_3 = \{1\}$, wodurch das Ereignis "die geworfene Zahl ist gleich 1" beschrieben wird. Dann gilt offensichtlich: $A_3 = A_2 \setminus B_3$. Zudem wissen wir, dass $B_3 \subset A_2$ und $P(B_3) = \frac{1}{6}$. Mit 1.6(c) ergibt sich:

$$P(A_3) = P(A_2 \setminus B_3) = P(A_2) - P(A_2 \cap B_3) = P(A_2) - P(B_3) = \frac{1}{2} - \frac{1}{6} = \frac{2}{6} = \frac{1}{3}.$$

Das letzte gesuchte Ereignis wird beschrieben durch $A_4 = \{2,3,4,5,6\}$. Als Hilfe benutzen wir das Ereignis B_4 "die geworfene Zahl ist größer als 2", d.h. $B_4 = \{3,4,5,6\}$. Dann ist offensichtlich $A_4 = A_1 \cup B_4$ und $A_1 \cap B_4 = \{4,6\}$. Mit 1.6(d) folgt

$$P(A_4) = P(A_1 \cup B_4) = P(A_1) + P(B_4) - P(A_1 \cap B_4) = \frac{1}{2} + \frac{4}{6} - \frac{2}{6} = \frac{5}{6}.$$

1.2 Zufallsvariablen

Oft ist man bei zufälligen Experimenten nicht ausschließlich an den einzelnen Ergebnissen interessiert, sondern an Größen, die von ihnen abhängen.

Beispiel 1.8 *(Vierfacher Münzwurf) In einem Spiel wird viermal nacheinander eine Münze geworfen. Als Modell haben wir somit den Wahrscheinlichkeitsraum* $(\Omega, \mathcal{A}, P)$ *mit*

$$\begin{aligned}\Omega =\{&(K,K,K,K),(Z,K,K,K),(K,Z,K,K),(K,K,Z,K),(K,K,K,Z),(Z,Z,K,K),\\&(K,Z,Z,K),(K,K,Z,Z),(Z,K,Z,K),(Z,K,K,Z),(K,Z,K,Z),(K,Z,Z,Z),\\&(Z,K,Z,Z),(Z,Z,K,Z),(Z,Z,Z,K),(Z,Z,Z,Z)\}\end{aligned}$$

und $\mathcal{A} = \mathcal{P}(\Omega)$. *Die Wahrscheinlichkeit für alle Elementarereignisse ist gleich und beträgt* $\frac{1}{16}$.

Jetzt interessieren wir uns aber nicht für den Ausgang jedes einzelnen Wurfes, sondern betrachten, wie viele der vier Würfe "Kopf" gezeigt haben. Da bei einem vierfachen Münzwurf entweder 0,1,2,3 oder 4-mal "Kopf" vorkommen kann, setzen wir $X(\omega) = k$, *falls die Wurfreihe* $\omega = (\omega_1, \omega_2, \omega_3, \omega_4)$ k*-mal "Kopf" enthält. Dann ist* X *eine Abbildung und ist wie folgt definiert:*

$$X : \Omega \to \{0, 1, 2, 3, 4\} \subset \mathbb{N}_0$$

Ganz allgemein definieren wir:

Definition 1.9 *Es sei* $(\Omega, \mathcal{A}, P)$ *ein Wahrscheinlichkeitsraum. Eine (reellwertige)* **Zufallsvariable** *ist eine Funktion* $X : \Omega \to \mathbb{R}$, *die jedem Elementarereignis* ω *der Ergebnismenge* Ω *eine reelle Zahl* $X(\omega)$ *zuordnet und bei der für jedes* $u \in \mathbb{R}$

$$\{\omega \in \Omega \ : \ X(\omega) \leq u\} \in \mathcal{A}$$

gilt. Das bedeutet, dass die Menge aller Elementarereignisse, deren Realisierung unterhalb eines bestimmten Wertes liegt, wieder ein Ereignis bilden muss.

Bemerkungen 1.10 *(a) Ist der Wertebereich* $W(X)$ *einer Zufallsvariable* X *endlich oder abzählbar unendlich, so bezeichnen wir diese Zufallsvariable als* **diskret**.

(b) Die Messbarkeitsbedingung in Definition 1.9 garantiert, dass man für jedes $u \in \mathbb{R}$ *eine Wahrscheinlichkeit bestimmen kann, dass* $X \leq u$. *Denn für diese Wahrscheinlichkeit, kurz* $P(X \leq u)$, *findet man demnach ein Ereignis* $A \in \mathcal{A}$, *so dass*

$$P(X \leq u) = P(\{\omega \in \Omega \mid X(\omega) \leq u\}) = P(A).$$

Dies werden wir am folgenden Beispiel veranschaulichen:

Beispiel 1.11 *(Vierfacher Münzwurf) Wir kehren zum Beispiel 1.8 zurück. Über die Zufallsvariable* X *hatten wir modelliert, wie oft bei einem vierfachen Münzwurf "Kopf" auftritt.*

Jetzt interessiert uns die Wahrscheinlichkeit dafür, dass höchstens zweimal "Kopf" geworfen wird. Enstprechend müssen wir alle Wurfreihen berücksichtigen, in denen höchstens zweimal "Kopf" auftritt. Das interessierende Ereignis ist also

$$\begin{aligned}A =\{&(Z,Z,K,K),(K,Z,Z,K),(K,K,Z,Z),(Z,K,Z,K),(Z,K,K,Z),(K,Z,K,Z),\\&(K,Z,Z,Z),(Z,K,Z,Z),(Z,Z,K,Z),(Z,Z,Z,K),(Z,Z,Z,Z)\}\end{aligned}$$

Da es insgesamt $2^4 = 16$ verschiedene Wurfreihen gibt (also $|\Omega| = 16$) und alle mit gleicher Wahrscheinlichkeit auftreten, erhält man für die gesuchte Wahrscheinlichkeit

$$P(X \leq 2) = P(\{\omega \in \Omega \mid X(\omega) \leq 2\}) = P(A) = \frac{11}{16}.$$

Ebenso können wir danach fragen, mit welcher Wahrscheinlichkeit X den Wert 2 annimmt. Offensichtlich nimmt die Zufallsvariable immer den Wert 2 an, wenn zweimal "Kopf" und zweimal "Zahl" geworfen werden. Somit ist

$$\begin{aligned}
P(X = 2) =& P(\{\omega \in \Omega \mid X(\omega) = 2\}) \\
=& P(\{(Z,Z,K,K),(K,Z,Z,K),(K,K,Z,Z),(Z,K,Z,K), \\
&(Z,K,K,Z),(K,Z,K,Z)\}) \\
=& \frac{6}{16} = \frac{3}{8}.
\end{aligned}$$

Abschließend können wir uns auch noch fragen, wie hoch die Wahrscheinlichkeit dafür ist, dass X einen Wert größer als 2 annimmt. Hierfür ergibt sich

$$\begin{aligned}
P(X > 2) =& P(\{\omega \in \Omega \mid X(\omega) > 2\}) \\
=& P(\{(K,K,K,K),(Z,K,K,K),(K,Z,K,K),(K,K,Z,K),(K,K,K,Z)\}) \\
=& \frac{5}{16}.
\end{aligned}$$

Da das Ereignis $X > 2$ das Gegenereignis zu $X \leq 2$ ist, hätte man die letzte Wahrscheinlichkeit auch mit Hilfe von Satz 1.6(a) berechnen können.

Allgemein betrachtet man für die Wahrscheinlichkeit, dass die Zufallsvariable X einen Wert aus der Menge $B \subset W(X)$ annimmt $P(\{\omega \in \Omega \mid X(\omega) \in B\})$. Daher definieren wir:

Definition 1.12 *Es sei X eine Zufallsvariable. Die* **Verteilung** *P_X von X ist ein Wahrscheinlichkeitsmaß auf dem Wertebereich $W(X)$ und wird definiert durch*

$$P_X(B) := P(\{\omega \in \Omega \mid X(\omega) \in B\}) \text{ für } B \subset W(X).$$

Bemerkung 1.13 *Es gibt verschiedene Möglichkeiten, Wahrscheinlichkeiten für Zufallsvariablen zu notieren:*

$$\begin{aligned}
P(X = u) &= P_X(\{u\}) = P(\{\omega \in \Omega \mid X(\omega) = u\}) \\
P(X \leq u) &= P_X(\{x \in W(X) \mid x \leq u\}) = P(\{\omega \in \Omega \mid X(\omega) \leq u\}) \\
P(X > u) &= P_X(\{x \in W(X) \mid x > u\}) = P(\{\omega \in \Omega \mid X(\omega) > u\})
\end{aligned}$$

Im Folgenden verwenden wir die Kurzschreibweisen $P(X \leq u)$ etc.

Bemerkung 1.14 *Mit Hilfe der Bezeichnungen aus Bemerkung 1.13 können wir uns die folgenden Rechenregeln für eine Zufallsvariable X sowie $a, b \in \mathbb{R}$ verdeutlichen:*

(a) (Gegenereignis) Es gilt $P(X > a) = 1 - P(X \leq a)$, denn

$$\begin{aligned}
P(X > a) =& P(\{\omega \in \Omega \mid X(\omega) > a\}) \\
=& P(\{\omega \in \Omega\} \setminus \{\omega \in \Omega \mid X(\omega) \leq a\}) \\
=& P(\Omega) - P(\{\omega \in \Omega \mid X(\omega) \leq a\}) \\
=& 1 - P(X \leq a).
\end{aligned}$$

(b) Es gilt $P(a < X \leq b) = P(X \leq b) - P(X \leq a)$, denn

$$\begin{aligned}
P(a < X \leq b) =& P(\{\omega \in \Omega \mid a < X(\omega) \leq b\}) \\
=& P(\{\omega \in \Omega \mid X(\omega) \leq b\} \setminus \{\omega \in \Omega \mid X(\omega) \leq a\}) \\
=& P(\{\omega \in \Omega \mid X(\omega) \leq b\}) - P(\{\omega \in \Omega \mid X(\omega) \leq a\}) \\
=& P(X \leq b) - P(X \leq a).
\end{aligned}$$

Für viele Fragestellungen ist es günstig neben der Verteilung auch die sogenannte Verteilungsfunktion einer Zufallsvariablen zu betrachten, die $P(X \leq u)$ als eine Funktion von $u \in \mathbb{R}$ betrachtet.

Definition 1.15 *Es sei X eine Zufallsvariable auf Ω. Die* **Verteilungsfunktion** *F_X von X ist definiert als*

$$F_X : \mathbb{R} \to [0,1], \; F_X(u) := P(X \leq u).$$

Satz 1.16 *Die Verteilungsfunktion F_X einer Zufallsvariablen X genügt den folgenden Eigenschaften:*

(a) F_X ist rechtsseitig stetig

(b) F_X ist monoton wachsend

(c) $F_X \in [0,1]$ für alle $u \in \mathbb{R}$

Bemerkung 1.17 *Die Beziehung aus Bemerkung 1.14(b) kann für $a, b \in \mathbb{R}$ mit $a \leq b$ auch geschrieben werden als $P(a < X \leq b) = F_X(b) - F_X(a)$.*

Wir kehren zum vierfachen Münzwurf aus Beispiel 1.8 zurück und werden dafür die Verteilungsfunktion bestimmen:

Beispiel 1.18 *Zunächst bestimmen wir die Verteilung von X. Diese sieht wie folgt aus:*

k	0	1	2	3	4
$P(X = k)$	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

Die Werte der Verteilungsfunktion $F_X(u)$ erhalten wir durch Aufsummieren der Einzelwahrscheinlichkeiten. Eine graphische Veranschaulichung dieser Verteilungsfunktion ist in Abbildung 1.1 dargestellt.

$$F_X(u) = \begin{cases} 0 & \text{für } \; u < 0 \\ \frac{1}{16} & \text{für } \; 0 \leq u < 1 \\ \frac{5}{16} & \text{für } \; 1 \leq u < 2 \\ \frac{11}{16} & \text{für } \; 2 \leq u < 3 \\ \frac{15}{16} & \text{für } \; 3 \leq u < 4 \\ 1 & \text{für } \; u \geq 4 \end{cases}$$

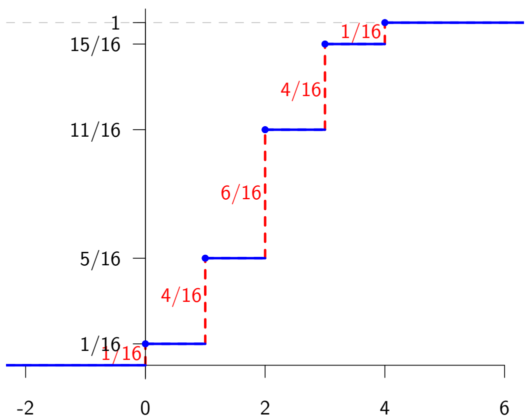


Abbildung 1.1: Verteilungsfunktion für die Anzahl an Kopfwürfen im vierfachen Münzwurf

Im vorherigen Beispiel haben wir gesehen, dass sich dort die Werte der Verteilungsfunktion durch das Aufsummieren der Einzelwahrscheinlichkeiten ergeben haben. Diese Eigenschaft spezifiziert eine wichtige Gruppe von Zufallsvariablen:

Definition 1.19 *Es sei $X : \Omega \to \mathbb{R}$ eine Zufallsvariable mit abzählbarem Wertebereich $W(X)$. X heißt* **diskret verteilt**, *wenn sich mit $f_X(x) = P(X = x)$ die Verteilungsfunktion $F_X(u)$ schreiben lässt als*

$$F_X(u) = \sum_{\{x \in W(X) | x \leq u\}} f_X(x) \text{ für alle } u \in \mathbb{R}$$

Man nennt f_X die **Dichtefunktion von** X.

Satz 1.20 *Es sei X eine diskrete Zufallsvariable mit abzählbarem Wertebereich $W(X) = \{x_i \mid i \in I\}$ (d.h. I ist abzählbar). Dann besitzt die Dichtefunktion die folgenen Eigenschaften:*

(a) $f_X(x) = P(X = x) = \begin{cases} \geq 0 & \text{für } x \in W(X) \\ = 0 & \text{sonst} \end{cases}$

(b) $\sum_{i \in I} f_X(x_i) = \sum_{i \in I} P(X = x_i) = P(\Omega) = 1$

Bemerkungen 1.21 (a) *Üblicherweise gibt man für eine diskrete Zufallsvariable Wertebereich und Dichtefunktion so an, dass für alle $x \in W(X)$ gilt $f_X(x) > 0$. Wie Satz 1.20(a) zeigt, ist dies aber nicht unbedingt notwendig. Manchmal ist es einfacher einen größeren Wertebereich anzugeben und in Kauf zu nehmen, dass die Dichtefunktion für einige Elemente den Wert Null annimmt.*

(b) *Jede Funktion f, die die Eigenschaften aus Satz 1.20 besitzt, ist eine Dichtefunktion. D.h. besitzt eine Funktion f diese Eigenschaften, so lässt sich eine Zufallsvariable finden, für die f die zugehörige Dichtefunktion ist.*

Die eben gewonnenen Einsichten werden wir an einem weiteren Beispiel veranschaulichen:

Beispiel 1.22 *(Augensumme zweier Würfel) Wir betrachten folgendes Zufallsexperiment: Wir würfeln mit zwei sechsseitigen Würfeln und notieren dann jeweils die Augensumme.*
Als Modell benutzen wir den Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit

$$\Omega = \{\omega = (\omega_1, \omega_2) \mid \omega_1, \omega_2 \in \{1, 2, 3, 4, 5, 6\}\}, \ \mathcal{A} = \mathcal{P}(\Omega), \ P(\{\omega\}) = \frac{1}{36}$$

Als Zufallsvariable X auf diesem Wahrscheinlichkeitsraum haben wir dann für unser Zufallsexperiment:

$$\begin{aligned} X : \{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\} &\longrightarrow \mathbb{N} \\ X((\omega_1, \omega_2)) &\longmapsto \omega_1 + \omega_2 \end{aligned}$$

Um die Verteilung der Zufallsvariablen X angeben zu können, benötigen wir zunächst den Wertebereich $W(X)$. In diesem Beispiel ermitteln wir $W(X)$ durch Erstellen einer Tabelle, in der alle möglichen Kombinationen, die beim Werfen zweier Würfel auftreten können, sowie die daraus resultierenden Augensummen erfasst werden.

X	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Hieraus ergibt sich für den Wertebereich $W(X) = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$.

Mit Hilfe der Tabelle können wir ebenfalls bestimmen, wie oft und bei welchen Wurfkombinationen die möglichen Augensummen vorkommen. Insgesamt gibt es 36 verschiedene Wurfmöglichkeiten und jede ist gleich wahrscheinlich; d.h. die Wahrscheinlichkeit beträgt jeweils $\frac{1}{36}$*. Daraus können wir sofort eine Tabelle für die Verteilung bzw. für die Dichtefunktion erstellen:*

u	2	3	4	5	6	7	8	9	10	11	12
$P(X=u)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Eine graphische Darstellung der Verteilung erfolgt mit Hilfe eines Stabdiagramms und ist im linken Bild in Abbildung 1.2 zu sehen.

Zu der Verteilung können wir nun auch noch die Verteilungsfunktion $F_X(u)$ *angeben. Dazu bilden wir die kumulierten Wahrscheinlichkeiten der Verteilung und erhalten die rechts stehende Funktion. Die graphische Darstellung der Verteilungsfunktion ist im rechten Bild der Abbildung 1.2 zu sehen.*

$$F_X(u) = \begin{cases} 0 & u < 2 \\ \frac{1}{36} & 2 \le u < 3 \\ \frac{3}{36} & 3 \le u < 4 \\ \frac{6}{36} & 4 \le u < 5 \\ \frac{10}{36} & 5 \le u < 6 \\ \frac{15}{36} & 6 \le u < 7 \\ \frac{21}{36} & 7 \le u < 8 \\ \frac{26}{36} & 8 \le u < 9 \\ \frac{30}{36} & 9 \le u < 10 \\ \frac{33}{36} & 10 \le u < 11 \\ \frac{35}{36} & 11 \le u < 12 \\ 1 & 12 \le u \end{cases}$$

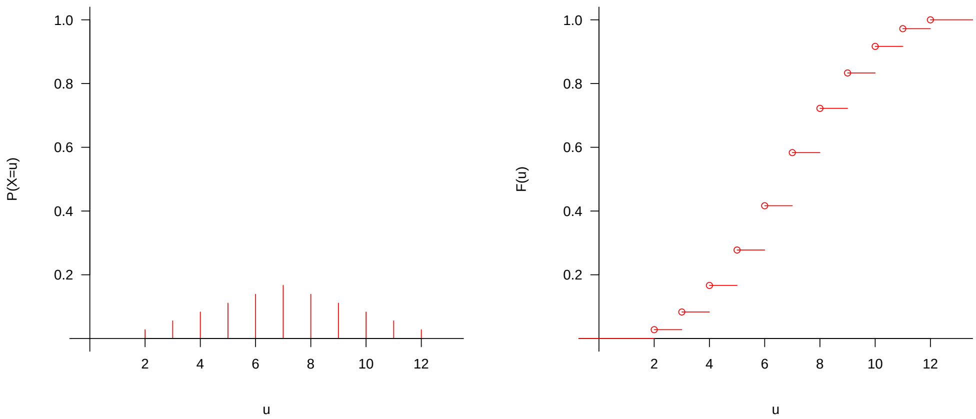


Abbildung 1.2: Verteilung und Verteilungsfunktion der Augensumme beim Wurf zweier Würfel

1.3 Erwartungswerte

Aus wahrscheinlichkeitstheoretischer Sicht kennt man eine Zufallsvariable, wenn man ihre Verteilung kennt. Wahrscheinlichkeitstheoretisch relevante Größen für Zufallsvariablen sind solche, die sich allein über deren Verteilung ausdrücken. Betrachten wir das Beispiel des n-fachen Münzwurfes. Wollen wir beispielsweise überprüfen, ob die Münze, die wir benutzen, fair ist, so sind wir daran interessiert, wie viele Kopfwürfe wir im Mittel bei einer fairen Münze zu erwarten hätten. Ist eine solche theoretische Größe bekannt, kann man eine Versuchsreihe durchführen und das hieraus errechnete arithmetische Mittel mit dem theoretischen Wert vergleichen. Die Beantwortung dieses Problems führt zu einer Größe, die über die Verteilung der Zufallsvariable definiert ist. (Eine Entscheidungsregel, wann man die Münze als fair bezeichnet und wann nicht, hat man damit allerdings noch nicht.)

Allgemein definieren wir:

Definition 1.23 *Es sei X eine diskrete Zufallsvariable mit abzählbarem Wertebereich $W(X) = \{x_i | i \in I\}$ und es gelte $\sum_{i \in I} x_i \cdot P(X = x_i) < \infty$. Der* **Erwartungswert** *von X ist definiert als*

$$E(X) := \sum_{i \in I} x_i \cdot P(X = x_i).$$

Bemerkung 1.24 *Der Erwartungswert einer Zufallsvariable ist ein theoretischer Wert, den man im Mittel erhält, wenn man unendlich viele unabhängige Durchführungen des Zufallsexperimentes vornimmt. Für eine diskrete Zufallsvariable mit endlichem Wertebereich lässt sich der Erwartungswert jedoch auch als eine Art Ausgleichspunkt sehen.*

Betrachten wir z.B. eine Zufallsvariable X mit $W(X) = \{0, 1, 2, 4\}$ und $P(X = 0) = 0,5$, $P(X = 1) = 0,25$, $P(X = 2) = 0,05$ und $P(X = 4) = 0,2$. Dann ergibt sich als Erwartungswert $E(X) = 0 \cdot 0,5 + 1 \cdot 0,25 + 2 \cdot 0,05 + 4 \cdot 0,2 = 1,15$. Dies ist genau der Punkt, den man unterstützen müsste damit der Zahlenstrahl ausbalanciert wäre, wenn man die Wahrscheinlichkeiten als Gewichte an den entsprechenden Stellen platziert.

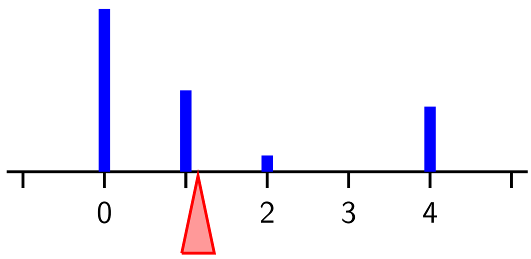


Abbildung 1.3: Erwartungswertwaage

Der Erwartungswert macht keine Aussage darüber, wie nah die einzelnen Werte der Zufallsvariable an diesem liegen. Dies veranschaulicht das nächste Beispiel.

Beispiel 1.25 *(Streuung) Es seien X und Y Zufallsvariablen auf Ω mit Wertebereich $W(X) = W(Y) = \{1, 2, 3\}$ mit den folgenden Verteilungen:*

$$P(X = 1) = P(X = 3) = \frac{1}{2},\; P(X = 2) = 0$$
$$P(Y = 1) = P(Y = 3) = \frac{1}{6},\; P(Y = 2) = \frac{2}{3}$$

Für die Erwartungswerte der beiden Zufallsvariablen ergibt sich:

$$E(X) = 1 \cdot \frac{1}{2} + 2 \cdot 0 + 3 \cdot \frac{1}{2} = 2$$
$$E(Y) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{2}{3} + 3 \cdot \frac{1}{6} = 2$$

X und Y sind also zwei Zufallsvariablen mit gleichem Wertebereich und gleichem Erwartungswert jedoch unterschiedlicher Verteilung. So liegt für X die gesamte Wahrscheinlichkeitsmasse symmetrisch links und rechts um den Erwartungswert, während für Y insgesamt $\frac{2}{3}$ der Wahrscheinlichkeitsmasse im Erwartungswert selbst liegen und jeweils $\frac{1}{6}$ links und rechts davon. Es ist also zu erwarten, dass die Werte von Y im Mittel näher am Erwartungswert liegen als die von X.

Aus diesem Grund führt man eine weitere Größere ein, die ein Maß für die Abweichung der Werte einer Zufallsvariable von ihrem Erwartungswert darstellt.

Definition 1.26 *Es sei X eine Zufallsvariable mit $E(X^2) < \infty$. Die* **Varianz** *$V(X)$ wird definiert durch*

$$V(X) := E\left((X - E(X))^2\right).$$

Die **Standardabweichung** *$SD(X)$ von X ist definiert als*

$$SD(X) := \sqrt{V(X)}.$$

Die Varianz misst also die erwartete (oder auch die mittlere) quadratische Abweichung der Werte vom Erwartungswert. Zur Berechnung der Varianz benötigen wir den folgenden allgemeingültigen Satz über die Berechnung von Erwartungwerten sogenannter transformierter Zufallsvariablen.

Satz 1.27 *Es sei X eine diskrete Zufallsvariable mit abzählbarem Wertebereich $W(X) = \{x_i | i \in I\}$ und es gelte $\sum_{i\in I} x_i \cdot P(X = x_i) < \infty$. Ferner sei $g : W(X) \to \mathbb{R}$ eine reellwertige Funktion. Gilt außerdem $\sum_{i\in I} |g(x_i)| \cdot P(X = x_i) < \infty$, dann ist*

$$E(g(X)) = \sum_{i\in I} g(x_i) \cdot P(X = x_i).$$

Wir benutzen die Aussage dieses Satzes, um für die beiden Zufallsvariablen aus Beispiel 1.25 Varianz und Standardabweichung zu ermitteln.

Beispiel 1.28 *Für die Zufallsvariablen X und Y aus Beispiel 1.25 ergibt sich mit $E(X) = E(Y) = 2$ und Satz 1.27*

$$\begin{aligned}
V(X) &= E\left((X-2)^2\right) = (1-2)^2 \cdot \frac{1}{2} + (2-2)^2 \cdot 0 + (3-2)^2 \cdot \frac{1}{2} = 1 \\
SD(X) &= \sqrt{V(X)} = 1 \\
V(Y) &= E\left((Y-2)^2\right) = (1-2)^2 \cdot \frac{1}{6} + (2-2)^2 \cdot \frac{2}{3} + (3-2)^2 \cdot \frac{1}{6} = \frac{1}{3} \\
SD(Y) &= \sqrt{V(Y)} \approx 0.5774
\end{aligned}$$

Die unterschiedlichen Werte für die Varianz bzw. für die Standardabweichung zeigen die unterschiedlichen Streuungen der jeweiligen Zufallsvariablen auf. Je kleiner der Wert, desto enger liegt die Wahrscheinlichkeitsmasse um den Erwartungswert.

Satz 1.27 lässt sich nicht nur zur Berechnung der Varianz verwenden, sondern auch in folgenden Zusammenhängen:

Beispiel 1.29 *Wir würfeln mit einem sechseitigen Würfel, der auf jeder Seite eine der Zahlen $-3, -2, -1, 1, 2, 3$ trägt.*

(a) Die Zufallsvariable X beschreibe die gewürfelte Zahl. Dann ist $W(X) = \{-3, -2, -1, 1, 2, 3\}$ und wir erhalten für den Erwartungswert

$$E(X) = \sum_{j\in W(X)} j \cdot P(X = j) = -3 \cdot \frac{1}{6} - 2 \cdot \frac{1}{6} - 1 \cdot \frac{1}{6} + 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} = 0.$$

(b) Die Zufallsvariable Y beschreibe nun nicht die gewürfelte Zahl, sondern ihr Quadrat. Dann können wir den Erwartungswert auf zwei verschiedene Arten berechnen:

(i) Es gilt $W(Y) = \{1, 4, 9\}$, sowie $P(Y = j^2) = P(Y = -j) + P(Y = j) = \frac{1}{3}$ für alle $j^2 \in W(Y)$. Auf die herkömmliche Weise erhalten wir

$$E(Y) = \sum_{j\in W(Y)} j \cdot P(Y = j) = 1 \cdot \frac{1}{3} + 4 \cdot \frac{1}{3} + 9 \cdot \frac{1}{3} = \frac{14}{3}.$$

(ii) Zum anderen haben wir $Y = X^2$. Bezogen auf Satz 1.27 ist $Y = g(X)$, wobei g die Quadratfunktion ist. Mit Hilfe des Satzes berechnet sich der Erwartungswert von Y als

$$\begin{aligned}
E(Y) &= \sum_{j\in W(X)} g(j) \cdot P(X = j) \\
&= \sum_{j\neq 0, j=-3}^{3} j^2 \cdot \frac{1}{6} = \frac{9+4+1+1+4+9}{6} = \frac{28}{6} = \frac{14}{3}.
\end{aligned}$$

In Satz 1.27 wurde bereits die Berechnung des Erwartungswertes von Zufallsvariablen der Form $g(X)$ mit $g : W(X) \to \mathbb{R}$ angegeben. Eine spezielle Form solcher Transformationen ist die lineare Transformation. Man spricht davon, dass die diskrete Zufallsvariable Y eine lineare Tansformation der Zufallsvariablen X ist, wenn es $a, b \in \mathbb{R}$ gibt, so dass $Y = aX + b$ gilt. Für Zufallsvariablen dieser Art lassen sich aus Satz 1.27 Rechenregeln für den Erwartungswert und die Varianz herleiten.

Satz 1.30 *Es seien X und Y diskrete Zufallsvariablen mit Erwartungswerten $E(X)$ und $E(Y)$ sowie $a, b \in \mathbb{R}$. Dann gilt*

(a) $E(aX + b) = a \cdot E(X) + b$

(b) $E(X + Y) = E(X) + E(Y)$

Beweis:

(a) Um das Vorgehen im Beweis deutlich zu machen, gehen wir hier von einem endlichen Wertebereich $W(X) = \{x_1, \dots, x_n\}$ aus. Für abzählbar unendliche Wertebereiche muss man die Regeln für die Reihenkonvergenz beachten.

$$\begin{aligned} E(aX+b) &\overset{\text{(Satz 1.27)}}{=} \sum_{i=1}^{n} (ax_i + b) \cdot P(X = x_i) \\ &= \underbrace{a \cdot \sum_{i=1}^{n} x_i \cdot P(X = x_i)}_{=E(X)} + b \cdot \underbrace{\sum_{i=1}^{n} P(X = x_i)}_{=1} \\ &= aE(X) + b \end{aligned}$$

(b) Die Rechenregel zur Bestimmung des Erwartungswertes der Summe zweier Zufallsvariablen lassen wir an dieser Stelle unbewiesen.

Neben der Rechenregel für die Varianz linearer Transformationen enthält der nachfolgende Satz auch eine allgemeine Rechenregel zur Bestimmung der Varianz. Manchmal lässt sich die Varianz unter Verwendung dieser Formel leichter berechnen als unter Verwendung der Definition der Varianz.

Satz 1.31 *Es seien X und Y diskrete Zufallsvariablen mit $E(X^2) < \infty$ und $E(Y^2) < \infty$. Dann gilt mit $a, b \in \mathbb{R}$:*

(a) $V(X) = E(X^2) - E(X)^2$

(b) $V(aX + b) = a^2 V(X)$

Beweis:

(a)
$$\begin{aligned} V(X) &= E\left((X - E(X))^2\right) = E\left(X^2 - 2 \cdot X \cdot E(X) + E(X)^2\right) \\ &\overset{\text{(Satz 1.30)}}{=} E(X^2) - 2 \cdot E(X)^2 + E(X)^2 \\ &= E(X^2) - E(X)^2 \end{aligned}$$

(b)
$$\begin{aligned} V(aX+b) &= E\left((aX + b - E(aX + b))^2\right) \\ &\overset{\text{(Satz 1.30)}}{=} E\left((aX + b - aE(X) - b)^2\right) \\ &= E\left(a^2 \cdot (X - E(X))^2\right) \\ &\overset{\text{(Satz 1.30)}}{=} a^2 E\left((X - E(X))^2\right) = a^2 V(X) \end{aligned}$$

Ebenso wie sich Datenreihen standardisieren lassen, so lassen sich auch Zufallsvariablen standardisieren.

Definition 1.32 *Es sei X eine diskrete Zufallsvariable mit Erwartungswert $E(X) < \infty$ und Varianz $V(X) < \infty$. Dann bezeichnet*

$$\widetilde{X} := \frac{X - E(X)}{SD(X)}$$

die zugehörige **standardisierte Zufallsvariable**. *Den Vorgang nennt man* **Standardisierung** *der Zufallsvariable.*

Bemerkung 1.33 *Für eine standardisierte Zufallsvariable $\widetilde{X}$ gilt*

$$E(\widetilde{X}) = 0 \text{ und } V(\widetilde{X}) = 1.$$

Beweis: Übungsaufgabe

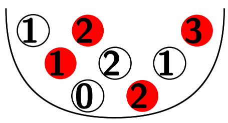
1.4 Unabhängigkeit von Zufallsvariablen

In Satz 1.30 haben wir gesehen, dass für zwei diskrete Zufallsvariablen stets $E(X+Y) = E(X)+E(Y)$ gilt. Nun drängt sich die Frage auf, ob eine solche Regel auch für das Produkt gilt, also ob

$$E(XY) = E(X) \cdot E(Y)?$$

Wir werden sehen, dass dies nur unter bestimmten Bedingungen erfüllt ist, nämlich dann, wenn X und Y unabhängig verteilt sind. Hierfür sind zunächst einige weitere Betrachtungen notwendig.

Beispiel 1.34 *Eine Urne enthält acht Kugeln, die mit den Zahlen 0, 1, 2 und 3 beschriftet sind. Zusätzlich ist die Hälfte der Kugeln rot und die andere weiß.*



Es wird eine Kugel entnommen. Die Zufallsvariable X_1 beschreibe die Anzahl der gezogenen roten Kugeln, d.h. $X_1 = 1$ für "rot" und $X_1 = 0$ für "weiß". Für die gezogene Zahl verwenden wir eine zweite Zufallsvariable X_2.

Die Wahrscheinlichkeiten für das Ziehungsergebnis, also die möglichen Werte von X_1 und X_2, lassen sich in einer Tabelle der sogenannten gemeinsamen Verteilung $P(X_1 = k, X_2 = j) = P(\{X_1 = k\} \cap \{X_2 = j\})$ darstellen, die die Wahrscheinlichkeit dafür beschreibt, dass gleichzeitig von X_1 der Wert k ($k \in W(X_1) = \{0, 1\}$) und von X_2 der Wert j ($j \in W(X_2) = \{0, 1, 2, 3\}$) angenommen wird.

$P(X_1 = k, X_2 = j)$		X_2			
		$j = 0$	$j = 1$	$j = 2$	$j = 3$
X_1	$k = 0$ (weiß)	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	0
	$k = 1$ (rot)	0	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$

Im vorherigen Beispiel haben wir uns Wahrscheinlichkeiten angesehen, mit denen Paare möglicher Werte der beiden Zufallsvariablen angenommen werden. Dies führt uns zu einer weiteren Begrifflichkeit. Wir definieren allgemein:

Definition 1.35 *Es sei $(\Omega, \mathcal{P}, P)$ ein diskreter Wahrscheinlichkeitsraum sowie $X_1, \ldots, X_n$ Zufallsvariablen auf $(\Omega, \mathcal{P}, P)$ mit eventuell verschiedenen Wertebereichen $W_1, \ldots, W_n$. Dann lassen sich diese Zufallsvariablen zu einer Zufallsvariable X mit Wertebereich $W = W_1 \times \ldots \times W_n$ zusammenfassen. Dafür setzt man $X(\omega) = (X_1(\omega), \ldots, X_n(\omega))$. Die Verteilung von X nennt man die* **gemeinsame Verteilung** *von $X_1, \ldots, X_n$. Sie ist durch die Angabe aller Wahrscheinlichkeiten $P(X_1 = x_1, \ldots, X_n = x_n)$ mit $(x_1, \ldots, x_n) \in W$ bestimmt.*

Wenn wir uns die Wahrscheinlichkeiten aus der Tabelle in Beispiel 1.34 ansehen, dann erkennen wir weitere Zusammenhänge.

Beispiel 1.36 *Offensichtlich kann man aus der gemeinsamen Verteilung von X_1 und X_2 aus Beispiel 1.34 auch die einzelnen Verteilungen von X_1 und X_2 ableiten. So ist die Wahrscheinlichkeit dafür, dass eine rote Kugel gezogen wird, die Summe der Wahrscheinlichkeiten, dass eine rote "0", eine rote "1", eine rote "2" oder eine rote "3" gezogen wird. Dies ist wiederum nichts anderes als die Summe der letzten Zeile. Das gleiche gilt für die Wahrscheinlichkeit, keine rote Kugel zu ziehen. Neben den Zeilen kann man diese Beobachtung auch für die Spalten tätigen. Die Spaltensummen geben jeweils die Wahrscheinlichkeiten dafür an, dass eine "0", eine "1", eine "2" oder eine "3" gezogen wird. Die so ermittelten Verteilungen heißen Randverteilungen und werden ebenfalls in die Tabelle integriert.*

$P(X_1 = k, X_2 = j)$		X_2				$P(X_1 = k)$
		$j = 0$	$j = 1$	$j = 2$	$j = 3$	
X_1	$k = 0$ *(weiß)*	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	*0*	$\frac{1}{2}$
	$k = 1$ *(rot)*	0	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$
$P(X_2 = j)$		$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	1

Die Summe der letzten Zeile und der letzten Spalte ergibt jeweils 1. Dies wird auch in der Tabelle notiert.

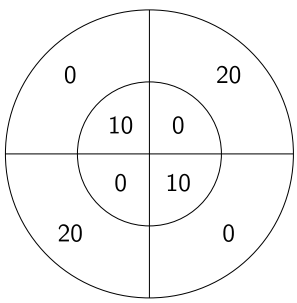
Definition 1.37 *Die sogenannten* **Randverteilungen** *der diskreten Zufallsvariablen X_1 und X_2 mit Wertebereichen $W(X_1)$ und $W(X_2)$ sind definiert durch*

(a) $P(X_1 = k) = \sum_{j \in W(X_2)} P(X_1 = k, X_2 = j)$

(b) $P(X_2 = j) = \sum_{k \in W(X_1)} P(X_1 = k, X_2 = j)$

Beispiel 1.38 *In den beiden folgenden Beispielen wird jeweils ein Glücksrad mit vier Ergebnismöglichkeiten einmal gedreht. Die Zufallsvariable X gibt die gedrehte Zahl im äußeren Ring und die Zufallsvariable Y die gedrehte Zahl im inneren Ring an.*

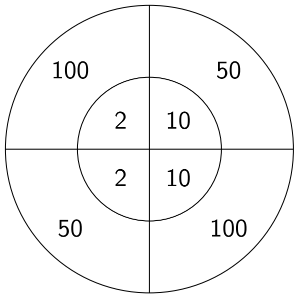
(a) Glücksrad 1:



Dann haben wir die Wertebereiche $W(X) = \{0, 20\}$ und $W(Y) = \{0, 10\}$ sowie die folgende Verteilung:

$P(X = k, Y = j)$		Y		$P(X = k)$
		$j = 0$	$j = 10$	
X	$k = 0$	0	$\frac{1}{2}$	$\frac{1}{2}$
	$k = 20$	$\frac{1}{2}$	0	$\frac{1}{2}$
$P(Y = j)$		$\frac{1}{2}$	$\frac{1}{2}$	1

(b) Glücksrad 2:



Dann haben wir die Wertebereiche $W(X) = \{50, 100\}$ und $W(Y) = \{2, 10\}$ sowie die folgende Verteilung:

$P(X = k, Y = j)$		Y		$P(X = k)$
		$j = 2$	$j = 10$	
X	$k = 50$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
	$k = 100$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
$P(Y = j)$		$\frac{1}{2}$	$\frac{1}{2}$	1

Wie wir am Beispiel der beiden Glücksräder gesehen haben, lassen sich aus der gemeinsamen Verteilung immer die Randverteilungen ableiten. Anders herum funktioniert dies nur in ganz bestimmten Fällen.

Definition 1.39 *Es sei $(\Omega, \mathcal{P}, P)$ ein diskreter Wahrscheinlichkeitsraum. Die Zufallsvariablen $(X_i)_{i \in I}$, $X_i : \Omega \to W_i$ heißen* **stochastisch unabhängig verteilt**, *wenn für jede Wahl von $A_i \subset W_i$ die Ereignisse $\{\omega_i : X_i(\omega_i) \in A_i\},\ i \in I$ unabhängig sind.*

Ohne Beweis haben wir den folgenden Satz:

Satz 1.40 *Es seien $X_1, \ldots, X_n$ diskrete Zufallsvariablen mit Verteilungen $P(X_i = x_i)$ $(i = 1, \ldots, n)$ und gemeinsamer Verteilung $P(X_1 = x_1, \ldots, X_n = x_n)$. Die Zufallsvariablen $X_1, \ldots, X_n$ sind genau dann stochastisch unabhängig verteilt, wenn für alle $(x_1, \ldots, x_n) \in W_1 \times \ldots \times W_n$ gilt*

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

n Ereignisse sind also stochastisch unabhängig, wenn die Wahrscheinlichkeit, dass alle n Ereignisse eintreten, gleich dem Produkt ihrer Einzelwahrscheinlichkeiten ist.

Beispiel 1.41 *Die Zufallsvariablen aus Beispiel 1.38(a) sind nicht unabhängig, denn es gilt z.B.*

$$P(X = 0, Y = 10) = \frac{1}{2} \neq \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(X = 0) \cdot P(Y = 10).$$

Hingegen sind die beiden Zufallsvariablen aus Beispiel 1.38(b) unabhängig, denn es gilt für alle $k \in W(X)$ und alle $j \in W(Y)$

$$P(X = k, Y = j) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(X = k) \cdot P(Y = j).$$

Mit Hilfe der gemeinsamen Verteilung lässt sich auch der Erwartungswert bestimmter Transformationen aus zwei Zufallsvariablen berechnen.

Satz 1.42 *Es seien X und Y zwei diskrete Zufallsvariablen und weiter sei h eine Abbildung mit $h : W(X) \times W(Y) \to \mathbb{R}$. Dann gilt*

$$\begin{aligned}
E(h(X,Y)) &= \sum_{(x_i, y_j) \in W(X) \times W(Y)} h(x_i, y_j) P(X = x_i, Y = y_j) \\
&= \sum_{x_i \in W(X)} \sum_{y_j \in W(Y)} h(x_i, y_j) P(X = x_i, Y = y_j) \\
&= \sum_{y_j \in W(Y)} \sum_{x_i \in W(X)} h(x_i, y_j) P(X = x_i, Y = y_j)
\end{aligned}$$

D.h. für Zufallsvariablen mit endlichen Wertebereichen, also $W(X) = \{x_1, \ldots, x_n\}$ und $W(Y) = \{y_1, \ldots, y_m\}$, gilt

$$E(h(X,Y)) = \sum_{i=1}^{n} \sum_{j=1}^{m} h(x_i, y_j) P(X = x_i, Y = y_j).$$

Nach wie vor beschäftigt uns die Frage, ob für die Erwartungswerte $E(X)$ und $E(Y)$ zweier diskreter Zufallsvariablen X und Y stets $E(XY) = E(X) \cdot E(Y)$ gilt oder ob dies nur unter bestimmten Bedingungen erfüllt ist. Wir haben den Begriff der stochastischen Unabhängigkeit kennengelernt und überprüfen nun an unserem Beispiel mit den Glücksrädern, welche Auswirkungen dies auf die uns interessierende Gleichung hat.

Beispiel 1.43 *Wir betrachten Beispiel 1.38 und wenden Satz 1.42 an, um jeweils $E(XY)$ zu berechnen. In diesem Fall ist dann $h(x_i, y_j) = x_i \cdot y_j$. Wir erhalten folgende Erwartungswerte:*

(a) Glücksrad 1:

$$\begin{aligned}
E(X) &= 0 \cdot \frac{1}{2} + 20 \cdot \frac{1}{2} = 10 \\
E(Y) &= 0 \cdot \frac{1}{2} + 10 \cdot \frac{1}{2} = 5 \\
E(XY) &= 0 \cdot 0 \cdot P(X = 0, Y = 0) + 0 \cdot 10 \cdot P(X = 0, Y = 10) \\
&\quad + 20 \cdot 0 \cdot P(X = 20, Y = 0) + 20 \cdot 10 \cdot P(X = 20, Y = 10) \\
&= 0 \cdot 0 \cdot 0 + 0 \cdot 10 \cdot \frac{1}{2} + 20 \cdot 0 \cdot \frac{1}{2} + 20 \cdot 10 \cdot 0 \\
&= 0
\end{aligned}$$

Somit ist für die nicht unabhängig verteilten Zufallsvariablen X und Y

$$E(XY) = 0 \neq 50 = 10 \cdot 5 = E(X) \cdot E(Y).$$

(b) Glücksrad 2:

$$\begin{aligned}
E(X) &= 50 \cdot \frac{1}{2} + 100 \cdot \frac{1}{2} = 75 \\
E(Y) &= 2 \cdot \frac{1}{2} + 10 \cdot \frac{1}{2} = 6 \\
E(XY) &= 50 \cdot 2 \cdot P(X = 50, Y = 2) + 50 \cdot 10 \cdot P(X = 50, Y = 10) \\
&\quad + 100 \cdot 2 \cdot P(X = 100, Y = 2) + 100 \cdot 10 \cdot P(X = 100, Y = 10) \\
&= 50 \cdot 2 \cdot \frac{1}{4} + 50 \cdot 10 \cdot \frac{1}{4} + 100 \cdot 2 \cdot \frac{1}{4} + 100 \cdot 10 \cdot \frac{1}{4} \\
&= 25 + 125 + 50 + 250 = 450
\end{aligned}$$

Somit ist für die unabhängig verteilten Zufallsvariablen X und Y

$$E(XY) = 450 = 75 \cdot 6 = E(X) \cdot E(Y).$$

Dass die Rechenregel für den Erwartungswert des Produkts zweier Zufallsvariablen ausschließlich für unabhängig verteilte Zufallsvariablen gilt, besagt der folgende Satz.

Satz 1.44 *Es seien X und Y unabhängig verteilte, diskrete Zufallsvariablen mit $E(|X|) < \infty$ und $E(|Y|) < \infty$. Dann gilt*

$$E(|XY|) < \infty \quad \textit{und} \quad E(XY) = E(X) \cdot E(Y).$$

Beweis: Es seien x_i $(i \in I)$ und y_j $(j \in J)$ die von X und Y angenommenen Werte. Dann gilt (aufgrund der Nichtnegativität aller Terme):

$$\begin{aligned} E(|XY|) &= \sum_{(i,j) \in I \times J} |x_i y_j| \cdot P(X = x_i, Y = y_j) \\ &= \sum_{i \in I} \sum_{j \in J} |x_i| \cdot |y_j| \cdot P(X = x_i) \cdot P(Y = y_j) \\ &= \left(\sum_{i \in I} |x_i| \cdot P(X = x_i) \right) \cdot \left(\sum_{j \in j} |y_j| \cdot P(Y = y_j) \right) \\ &= E(|X|) \cdot E(|Y|) \end{aligned}$$

Da $E(|X|) < \infty$ und $E(|Y|) < \infty$, ist auch $E(|XY|) < \infty$. Aus der gleichen Rechnung ohne Betragsstriche folgt dann auch $E(XY) = E(X) \cdot E(Y)$.

1.5 Kovarianz

Wir haben bereits die Varianz als Maß für die Abweichung der Werte einer Zufallsvariablen von ihrem Erwartungswert kennengelernt. Um den Zusammenhang zwischen zwei Zufallsvariablen beurteilen zu können, benötigen wir eine neue Kenngröße.

Definition 1.45 *Es seien X und Y zwei diskrete Zufallsvariablen mit $E(X^2) < \infty$ und $E(Y^2) < \infty$. Dann wird die* **Kovarianz** *zwischen X und Y definiert durch*

$$Cov(X, Y) = E\left((X - E(X))(Y - E(Y))\right).$$

Bemerkung 1.46 *Die Werte der Kovarianz sind wie folgt zu interpretieren:*

(a) Die Kovarianz ist positiv, wenn hohe (niedrige) Werte von X mit hohen (niedrigen) Werten von Y einhergehen.

(b) Die Kovarianz ist negativ, wenn hohe (niedrige) Werte von X mit niedrigen (hohen) Werten von Y einhergehen.

(c) Die Kovarianz ist null, wenn kein monotoner Zusammenhang zwischen X und Y besteht.