4 Elementare Wahrscheinlichkeitsrechnung

In der Stochastik geht es darum, den Zufall in den Griff zu bekommen, oder zumindest befriedigende Vorhersagen machen zu können, wenn immer das möglich ist. Man geht von einem konkreten Experiment oder naturwissenschaftlichen bzw. gesellschaftlichen Phänomen aus, welches sich genügend oft wiederholen lässt bzw. auftritt und einen nicht deterministischen Ausgang hat, d.h. es sind mehrere Ergebnisse möglich, welche dem Zufall unterliegen. Die Stochastiker entwickeln für solche Zufallsexperimente mathematische Modelle, welche dazu dienen, konkrete Aussagen bzw. Vorhersagen machen zu können. Solche Modelle tauchen in allen Naturwissenschaften auf, in der Wirtschaft, in der Technik, überall im Alltag, wo Vorhersagen benötigt werden. Mit stochastischen Modellen werden sicherheitskritische Systeme (Kraftwerke, Schutzsysteme, Betriebssysteme) entworfen und betrieben, Börsenaktivitäten, Energieproduktion und Energieverteilung gesteuert, Wetterprognosen, Wetter- und Katastrophenwarnungen ermöglicht, Versicherungsprämien kalkuliert, künstliche Intelligenz entwickelt, und vieles mehr.

4.1 Diskrete Wahrscheinlichkeitsräume

Wir wollen die elementaren Begriffe der Stochastik anhand des *Zufallsexperiments* des

Einmaligen Würfelns mit einem fairen Würfel

erklären. Mit *fair* meinen wir, dass jede der sechs Flächen des Würfels gleich wahrscheinlich ist. Dies ist ein theoretischer Zugang. Dabei handelt es sich um einen idealen Würfel. In der Praxis gibt es keine fairen Würfel. Dies grenzt den stochastischen Zugang gegen einen statistischen Zugang ab. In der Statistik werden Daten gesammelt, um mit einem Modell quantitative Aussagen über den Ausgang von Zufallsexperimenten machen zu können. In diesem Sinne würde der betreffende Würfel getestet werden. Die Aufgabe des Statistikers ist es dann, aus den Testergebnissen möglichst umfassende Aussagen zu machen. Allerdings ist ein statistischer Zugang nur möglich, wenn ausreichend viele Testergebnisse vorhanden sind. Ist dem nicht so, bleibt der stochastische Zugang, welcher unter gerechtfertigten Annahmen, Expertenmeinungen etc. ein abstraktes Modell für die quantitative Analyse des Experiments bereitstellt.

Als erstes ermitteln wir für das konkrete Zufallsexperiment alle möglichen *Ergebnisse*. Dabei muss immer sehr sorgfältig vorgegangen werden. Kein mögliches Resultat darf vergessen werden. Beim Würfeln mit einem Würfel sind die möglichen Ergebnisse die Zahlen 1 bis 6. Würde man zulassen, dass der Würfel auf einer Kante zu liegen kommen kann, so müsste man zusätzlich alle möglichen Kanten als Ergebnisse zulassen. Wir wollen nicht so spitzfindig sein und nehmen die Zahlen von 1 bis 6 als einzig mögliche Ergebnisse auf. Alle Ergebnisse werden in der Ergebnismenge, dem sogenannten *Ergebnisraum* Ω **(**Omega**)** zusammengefasst:

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

Weil wir einen fairen Würfel angenommen haben, ist jede mögliche Zahl gleich wahrscheinlich. Wahrscheinlichkeit kann man in Zahlen zwischen 0 und 1 oder auch in Prozenten ausdrücken. In unserem Fall hat jedes der Ergebnisse aus Omega die Wahrscheinlichkeit
1/6 100/6
…… bzw. ………… %.

Mathematisch wird diese Verteilung der möglichen Ergebnisse durch eine Funktion, die *Zähldichte* ρ (rho), auf Ω festgelegt:

$$\rho: \Omega \to [0,1],\ \rho(\omega) = \ldots\ldots 1/6, \omega \in \Omega$$

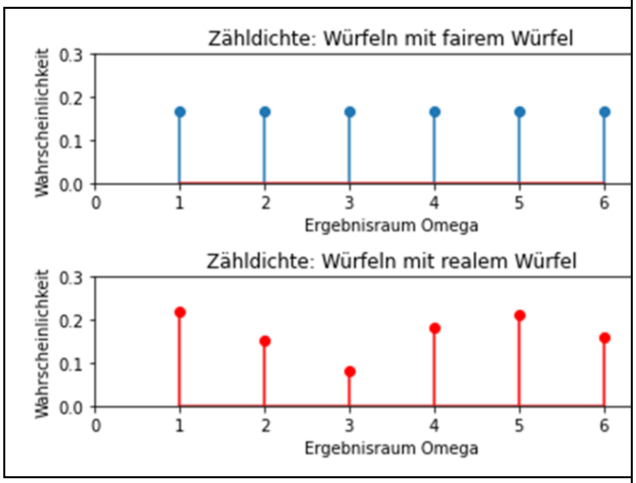
Mit $[0,1]$ ist das Intervall aller reellen Zahlen zwischen 0 und 1 gemeint. Hier handelt es sich um eine *Gleichverteilung*, denn alle Ergebnisse des Zufallsexperimentes sind gleich wahrscheinlich; man sagt, Ω sei ein *Laplace-Raum* (Definition s. S.5 unten).

Bei einem statistischen Zugang über Testergebnisse würden die verschiedenen Würfelflächen möglicherweise völlig andere Wahrscheinlichkeiten erhalten. Unten ist zum Vergleich ein statistischer Zugang zur Ermittlung einer Zähldichte für das Würfelexperiment mit einem beispielsweise von der Polizei in einem Spielkasino beschlagnahmten Würfel. In der Tabelle sind die absoluten Häufigkeiten h_i und die relativen Häufigkeiten f_i der Resulatet von 100 Würfelversuchen eingetragen. Beim statistischen Zugang entsprechen die relativen Häufigkeiten den

Werten der Zähldichte. Hier liegt ....keine Gleichverteilung vor.

ω	1	2	3	4	5	6
h_i	22	15	8	18	21	16
f_i	22/100	15/100	8/100	18/100	21/100	16/100

Häufig wird die Zähldichte als *Stabdiagramm* visualisiert. Unten sind die Zähldichten für den fairen Würfel und für den getesteten Würfel gegenübergestellt.




```
import numpy as np
import matplotlib.pyplot as plt

X=np.arange(1,7)
Zdichte=1/6*np.ones(6)
hx=np.array([22, 15, 8, 18, 21, 16])
fx=1/np.sum(hx)*hx

plt.figure(1)
plt.subplot(211)
plt.stem(X,Zdichte,use_line_collection='true')
plt.title('Zähldichte: Würfeln mit fairem Würfel')
plt.xlabel('Ergebnisraum Omega')
plt.ylabel('Wahrscheinlichkeit')
plt.xlim(0,7)
plt.ylim(0,0.3)

plt.subplot(212)
plt.stem(X,fx, use_line_collection='true',
linefmt='r',markerfmt='ro')
plt.title('Zähldichte: Würfeln mit realem Würfel')
plt.xlabel('Ergebnisraum Omega')
plt.ylabel('Wahrscheinlichkeit')
plt.xlim(0,7)
plt.ylim(0,0.3)
plt.tight_layout()
```

Gehen wir zurück zum stochastischen Modell mit der Gleichverteilung. Bei genauerer Betrachtung stellen wir fest, dass die Aussagekraft des Modells noch nicht ausreicht.

Wir wollen auch in der Lage sein, die Wahrscheinlichkeiten von Ereignissen wie den folgenden zu bestimmen:

A: Eine gerade Zahl wird gewürfelt.
B: Eine Zahl ≥ 3 wird gewürfelt.
C: Keine 2 wird gewürfelt.
D: 4 oder 6 wird gewürfelt.
E: Eine Zahl, welche nicht durch 3 teilbar ist, wird gewürfelt.

Solche Aussagen bezeichnen wir im Gegensatz zu den Ergebnissen (den Elementen des Ergebnisraumes Ω) als *Ereignisse*. Das Ereignis A tritt genau dann ein, wenn entweder

………………… 2, 4 oder 6

gewürfelt wird. Somit kann das Ereignis A mit der Menge $A =$……….. $\{2,4,6\}$ identifiziert werden. Es handelt sich dabei offensichtlich um eine Teilmenge des Ergebnisraums Ω. Für die restlichen Aussagen ergeben sich entsprechend die Mengen

$B = \{3,4,5,6\}$ $\quad C = \{1,3,4,5,6\}$ $\quad D = \{4,6\}$ $\quad E = \{1,2,4,5\}$

Jede solche Aussage über die gewürfelte Zahl kann mit einer bestimmten Teilmenge von Ω gleichgesetzt werden. Wir wollen im Folgenden – nicht ganz korrekt – keinen Unterschied zwischen Aussagen und Mengen machen. Damit können wir sagen: *Die Aussage A ist wahr* oder *Das Ereignis A ist eingetreten* oder *Eine Zahl aus der Menge* $A = \{2,4,6\}$ *wurde gewürfelt*. Wir erlauben uns auch, die Logik-Operatoren *und* $\wedge$, *oder* $\vee$ und *nicht* $\neg$ entsprechend den Mengenoperatoren *Durchschnitt* $\cap$, *Vereinigung* $\cup$ und *Komplement* $\overline{}$ zu verwenden und können damit Aussagen machen, wie $A \wedge B$ *ist wahr* oder *Eine gerade Zahl grösser als 2 wurde gewürfelt* oder *Das Ereignis* $A \cap B = \{4,6\}$ *ist eingetreten*, etc.

Die Menge aller Ereignisse wird als *Ereignisraum* des Zufallsexperiments bezeichnet. Wie wir eben gesehen haben, können wir jedes Ereignis durch eine Menge beschreiben. Entsprechend ist der Ereignisraum die Menge aller Teilmengen von Ω, die sogenannte *Potenzmenge* von Ω.
Die Kombinatorik sagt uns, dass die Anzahl aller Teilmengen (inklusive der leeren Menge) von Ω gleich $2^{|\Omega|} = 2^6 = 64$ ist. Deswegen bezeichnen wir den Ereignisraum mit dem Symbol 2^Ω.

Die Wahrscheinlichkeit, dass bei unserem Zufallsexperiment das Ereignis A eintritt, ist die Summe der Wahrscheinlichkeiten aller Elemente, welche zu A gehören, also $\frac{1}{2}$. Wir schreiben dafür: $P(A) = \frac{1}{2}$. Formal erhält man diese Wahrscheinlichkeit als die Summe der durch die Zähldichte ρ festgelegten Gewichte aller Ergebnisse, welche in A sind:

$$P(A) = \rho(2) + \rho(4) + \rho(6) = 1/6 + 1/6 + 1/6 = 1/2$$

So definieren wir das *Wahrscheinlichkeitsmass* (auch *Wahrscheinlichkeitsverteilung*) $P(..)$ als Funktion auf der Potenzmenge von Ω:

$$P: 2^\Omega \to [0,1],\ P(M) = \sum_{\omega \in M} \rho(\omega) \quad \text{mit } M \subseteq \Omega.$$

Da es sich um einen Laplace-Raum handelt, ist dies gleichbedeutend mit der Definition

$$P(M) = \frac{|M|}{|\Omega|}$$

Dabei bezeichnen wir mit $|M|$ die Anzahl der Elemente der Menge M. Das Wahrscheinlichkeitsmass $P(..)$ ordnet jedem möglichen Ereignis des Würfelexperiments eine Wahrscheinlichkeit zu. Der *diskrete Wahrscheinlichkeitsraum* (Ω, P) ist das stochastische Modell für das Würfelexperiment. Als *diskret* wird ein Wahrscheinlichkeitsraum genau dann bezeichnet, wenn der zugehörige Ergebnisraum aus endlich vielen (oder höchstens abzählbar vielen) Ergebnissen besteht.

Mit dem Wahrscheinlichkeitsmass $P(..)$ lassen sich die Wahrscheinlichkeiten der Ereignisse B-E bestimmen:

$P(B) = \ldots\ldots 2/3$ $\quad$ $P(C) = \ldots\ldots 5/6$ $\quad$ $P(D) = \ldots\ldots 1/3$ $\quad$ $P(E) = \ldots\ldots 2/3$

Man sieht, wie in diesem Zusammenhang mengentheoretische Begriffe eine entscheidende Rolle spielen. Die Begriffe *Menge*, *Element*, *Teilmenge*, *Durchschnitt*, *Vereinigung*, *Differenzmenge*, *Komplement* und *leere Menge* sollten geläufig sein. Mit den oben definierten Beispielmengen A-E rufen wir hier das Wichtigste in Erinnerung:

- Die *Elemente* der Menge $A = \{2,4,6\}$ sind 2, 4 und 6.
- Die Menge $B = \{3,4,5,6\}$ ist eine *Teilmenge* von $C = \{1,3,4,5,6\}$, in Zeichen $B \subseteq C$, da jedes Element von B ein Element von C ist.
- Der *Durchschnitt* von A und C, in Zeichen $A \cap C$, ist die Menge aller Elemente, die sowohl in A als auch in C liegen: $A \cap C = \ldots\ldots\ldots\ldots \{4,6\}$.
- Die *Vereinigung* von $D = \{4,6\}$ und $E = \{1,2,4,5\}$, in Zeichen $D \cup E$, ist die Menge aller Elemente, die in D oder in E liegen: $D \cup E = \ldots\ldots\ldots\ldots \{1,2,4,5,6\}$.
- Die *Differenzmenge* A ohne B, in Zeichen $A \setminus B$, ist die Menge aller Elemente, die in A, aber nicht in B liegen: $A \setminus B = \ldots\ldots\ldots \{2\}$.
- Das *Komplement* von A in Ω, in Zeichen $\bar{A}$, ist die Menge aller Elemente aus der Grundmenge Ω, die nicht in A liegen: $\bar{A} = \Omega \setminus A = \ldots\ldots\ldots\ldots \{1,3,5\}$.
- Die *leere Menge* ist die eindeutig bestimmte Menge, welche kein Element enthält, in Zeichen $\{\}$ oder auch $\emptyset$
- Zwei Mengen heissen *disjunkt*, falls ihr Durchschnitt leer ist. Eine Anzahl von Mengen heisst paarweise disjunkt, wenn jede Auswahl von zwei Mengen daraus disjunkt ist.

Soweit unser Würfel-Beispiel.

Hier nun noch die wichtigsten Definitionen, denen wir begegnet sind, in allgemeiner Form:

Definitionen
Ein *Zufallsexperiment* ist ein Vorgang, bei dem folgende Bedingungen erfüllt sind:

- Der Vorgang lässt sich unter den gleichen äusseren Bedingungen beliebig oft wiederholen.
- Es sind mehrere sich gegenseitig ausschliessende *Ergebnisse* möglich; diese sind vor der Durchführung des Zufallsexperiments bekannt.
- Das Ergebnis lässt sich nicht mit Sicherheit voraussagen, sondern ist zufallsbedingt.

Ergebnisraum
Der ………………………………………… Ω ist die Menge aller möglichen, sich gegenseitig ausschliessenden Ergebnisse eines Zufallsexperiments. Der Ergebnisraum heisst *diskret*, falls er eine endliche oder abzählbar unendliche Anzahl von Ergebnissen enthält.

Zähldichte
Eine ……………………… ρ ist eine Funktion $\rho: \Omega \to [0,1]$, für die gilt: $\sum_{\omega \in \Omega} \rho(\omega) = 1$. Sie gibt an, mit welcher Wahrscheinlichkeit die möglichen Ergebnisse des Zufallsexperiments auftreten.

Ereignis
Ein ……………………. A ist die Zusammenfassung von gewissen Ergebnissen zu einer Menge, also einer Teilmenge von Ω. Man sagt, ein Ereignis A sei eingetroffen, wenn das Ergebnis des Zufallsexperiments in der Menge A enthalten ist. Die leere Menge $\{\}$ entspricht dem *unmögliche Ereignis*, das niemals eintritt; die Menge Ω selbst steht für das *sichere Ereignis*, das immer eintritt.

Ereignisraum
Der ……………………………………….. 2^{Ω} des Zufallsexperiments ist die Menge aller möglichen Ereignisse, also die Menge aller Teilmengen des Ergebnisraums Ω.

Wahrscheinlichkeitsmass
Das ……………………………………………………………… P ist eine Funktion $P: 2^{\Omega} \to [0,1]$ mit $P(M) = \sum_{\omega \in M} \rho(\omega)$ für jede Teilmenge $M \subseteq \Omega$. Sie gibt an, mit welcher Wahrscheinlichkeit das Ereignis, das durch die Menge M beschrieben wird, eintritt.

Ein diskreter Ergebnisraum Ω in Kombination mit einem Wahrscheinlichkeitsmass P wird als *diskreter Wahrscheinlichkeitsraum* (Ω, P) bezeichnet. Wenn die Zähldichte ρ für jedes mögliche Ergebnis denselben Wert hat, wenn also alle Ergebnisse gleich wahrscheinlich sind, dann nennt man (Ω, P) einen *Laplace-Raum*.

Der folgende Satz listet die wichtigsten Eigenschaften von diskreten Wahrscheinlichkeitsräumen auf:

Diskreter Wahrscheinlichkeitsraum
Jeder diskrete Wahrscheinlichkeitsraum (Ω, P) hat die folgenden Eigenschaften:

(A1) (Unmögliches Ereignis) $P(\{\}) = 0$
(A2) (Sicheres Ereignis) $P(\Omega) = 1$
(A3) (Komplementäres Ereignis) $P(\Omega \backslash A) = 1 - P(A)$
(A4) (Vereinigung) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
(A5) (Sigma-Additivität) $P(A_1 \cup A_2 \cup A_3 \cup \ldots) = P(A_1) + P(A_2) + P(A_3) + \ldots$
falls die Ereignisse $A_1, A_2, A_3, \ldots$ paarweise disjunkt sind.

Ist (Ω, P) ein Laplace-Raum, so gilt: $P(M) = \frac{|M|}{|\Omega|}$.

Der russische Mathematiker *Andrei Kolmogorov* hat in seiner axiomatischen Begründung der Wahrscheinlichkeitstheorie in der ersten Hälfte des 20. Jahrhunderts entdeckt, dass die Aussagen (A2) und (A5) (oder gleichbedeutend (A1) und (A5)) bereits ausreichen, um die restlichen Aussagen nachweisen zu können, und hat diese zur abstrakten Definition des Begriffs eines Wahrscheinlichkeitsmasses benutzt.

Im nächsten Beispiel wollen wir die eben besprochenen Begriffe an einem diskreten Wahrscheinlichkeitsraum mit einem *unendlichen* Ergebnisraum vertiefen. Dabei handelt es sich nun nicht mehr um einen *Laplace-Raum*.

Beispiel (**Münzwurf bis Kopf kommt**)
Eine Münze wird solange geworfen, bis Kopf 'K' kommt. Die möglichen *Ergebnisse* sind K, ZK, ZZK, $ZZZK$, usw. Der *Ergebnisraum* Ω ist bei diesem Zufallsexperiment unendlich:

$$\Omega = \{K, ZK, ZZK, ZZZK, \ldots\} \cup \{ZZZZ\ldots\}$$

Wir haben die nie abbrechende Folge $ZZZZ\ldots$, bei der niemals Kopf erscheint, dazugenommen, weil dieses Ergebnis theoretisch möglich ist. Für die *Zähldichte* entscheiden wir uns für einen fairen Münzwurf, bei dem jede Seite gleich wahrscheinlich ist. Die Zähldichte ist mit dieser Annahme festgelegt:

$$\rho(K) = \ldots\ldots 1/2,\ \rho(ZK) = \ldots\ldots 1/4,\ \rho(ZZK) = \ldots\ldots 1/8, \text{ usw.}$$

Ist dies auch wirklich ein Zählmass? Dazu müssen wir über alle Gewichte summieren und die Eigenschaft der Zähldichte nachweisen.
Es gilt nach der Summenformel für eine geometrische Reihe:

$$\rho(K) + \rho(ZK) + \rho(ZZK) + \ldots\ldots = 1/2 + 1/4 + 1/8 + \ldots = \sum_{k=1}^{\infty} \frac{1}{2^k} = \frac{1}{1-1/2} - 1 = 1$$

Wir haben einem Ergebnis noch kein Gewicht zugeordnet: Nachdem sich bereits alle Gewichte zu 1 aufsummieren, muss $\rho(ZZZZ\ldots) = 0$ gelten, sonst wäre die Funktion $\rho(..)$ keine Zähldichte. Das Wahrscheinlichkeitsmass ist somit festgelegt:

$$P: 2^{\Omega} \to [0,1],\ P(M) = \sum_{\omega \in M} \rho(\omega) \text{ und } M \subseteq \Omega.$$

Dies erlaubt uns, Aussagen über die Wahrscheinlichkeit beliebiger Ereignisse zu machen. Wir zeigen dies am Beispiel des Ereignisses

A: mindestens 3 Würfe lang erscheint kein Kopf

Zunächst drücken wir das Ereignis A als Menge aus: $A = \ldots$ {ZZZK, ZZZZK, ZZZZZK, …}

Um die Wahrscheinlichkeit von A zu bestimmen, gehen wir über zum Komplementärereignis
$\Omega \backslash A = \ldots$ {K, ZK, ZZK}. Für die Wahrscheinlichkeit des Ereignisses A gilt:

$$P(A) = 1 - P(\Omega \backslash A) = 1 - P(\{K, ZK, ZZK\}) = 1 - 1/2 - 1/4 - 1/8 = 1/8$$

Dies ist gleichbedeutend mit der Wahrscheinlichkeit, dass bei 3 Münzwürfen kein Kopf kommt. Diese beträgt für eine faire Münze $(\frac{1}{2})^3 = \frac{1}{8}$.

4.2 Zufallsvariablen

Um die Wahrscheinlichkeiten im stochastischen Modell elegant ausdrücken zu können, bedient man sich häufig sogenannter *Zufallsvariablen*. Dabei handelt es sich um reellwertige Funktionen auf dem Ergebnisraum, welche vom Ausgang des zugrundeliegenden Zufallsexperiments abhängen. Am besten erklären wir dieses zentrale Konzept der Stochastik im nächsten Beispiel, der Augensumme beim zweimaligen Würfeln mit einem fairen Würfel.

Beispiel 1: Augensumme beim zweimaligen Würfeln eines fairen Würfels

Zunächst legen wir den Ergebnisraum Ω fest. Es ist plausibel, in diesem Fall die Menge aller möglichen Zahlenpaare für den ersten und den zweiten Wurf zu nehmen:

$$\Omega = \ldots\ldots\ldots\ldots\ldots\ldots \{(1,1),(1,2),(1,3),\ldots,(6,6)\}$$

Der Ergebnisraum Ω hat ………………… $6 \cdot 6 = 36$ Elemente, die den möglichen Ergebnissen beim zweimaligen Würfeln entsprechen. Die Zähldichte definiert, da es sich um einen fairen Würfel handelt, eine Gleichverteilung (Laplace-Verteilung). Folglich setzen wir

$$\rho(1,1) = \rho(1,2) = \ldots = \rho(6,6) = \ldots\ldots 1/36.$$

Nun wollen wir aber ein spezielles Resultat dieses Experiments untersuchen, und zwar geht es uns um die Augensumme. Dazu definieren wir eine Funktion (eine sogenannte *Zufallsvariable*) X, die variabel ist, d.h. vom jeweiligen geworfenen Augenpaar abhängt. Diese Funktion gibt uns bei einem bestimmten gewürfelten Zahlenpaar dessen Augensumme an. Formal wird diese Zufallsvariable folgendermassen definiert

$$X: \Omega \to \{2,3,4,5,6,7,8,9,10,11,12\} \qquad X(\omega_1,\omega_2) = \omega_1 + \omega_2$$

Dieses Vorgehen hat den Vorteil, dass wir sehr genau ausdrücken können, worauf es uns in diesem Zufallsexperiment ankommt. Wir wollen wissen, wie wahrscheinlich es ist, dass eine bestimmte Augensumme, zum Beispiel die Augensumme 2, gewürfelt wird, d.h. dass X den Wert 2 annimmt, in Zeichen:

$$\ldots\ldots\ldots\ldots P(X = 2).$$

Dabei steht $X = 2$ verkürzt für das Ereignis bzw. die Menge $\{\omega \in \Omega | X(\omega) = 2\}$.

Um diese Frage zu beantworten, müssen wir die sogenannte *Verteilung der Zufallsvariablen X* bestimmen. Jede Verteilung einer Zufallsvariablen ist eine Wahrscheinlichkeitsverteilung. Umgekehrt kann jede Wahrscheinlichkeitsverteilung als Verteilung einer nicht näher präzisierten Zufallsvariablen angesehen werden.
Mithilfe der Zufallsvariablen X können dann viele Wahrscheinlichkeiten elegant aufgeschrieben und berechnet werden.

Die Wahrscheinlichkeit, eine Augensumme grösser als 5 zu würfeln: ……………… $P(X > 5)$

Die Wahrscheinlichkeit, eine Augensumme zwischen 7 und 10 zu würfeln: ……………… $P(7 \le X \le 10)$

Kehren wir zur *Verteilung* der Zufallsvariablen X zurück. Bei Zufallsvariablen mit endlich vielen möglichen Werten ermittelt man die Verteilung am besten anhand einer Tabelle, wie sie unten zu sehen ist.
In der ersten Zeile sind alle möglichen Werte der Zufallsvariablen aufgeführt. In der zweiten Zeile stehen alle für den jeweiligen Wert aus der ersten Zeile infrage kommenden Zahlenpaare aus Ω. Darunter steht die sogenannte (*Wahrscheinlichkeits-)Dichtefunktion* $f(x)$ (*probability density function*, kurz *PMF* oder *Zähldichte*) der Zufallsvariablen X, welche definiert ist durch:

$$f(x) = P(X = x)$$

(in Worten: Die Wahrscheinlichkeit, dass die Zufallsvariable den Wert x annimmt), wobei x für alle möglichen Werte zwischen 2 und 12 stehen kann.

x	2	3	4	5	6	7	8	9	10	11	12
Ergebnisse aus Ω	(1,1)	(1,2) (2,1)	(1,3) (2,2) (3,1)	(1,4) (2,3) (3,2) (4,1)	(1,5) (2,4) (3,3) (4,2) (5,1)	(1,6) (2,5) (3,4) (4,3) (5,2) (6,1)	(2,6) (3,5) (4,4) (5,3) (6,2)	(3,6) (4,5) (5,4) (6,3)	(4,6) (5,5) (6,4)	(5,6) (6,5)	(6,6)
$f(x) = P(X = x)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36
$F(x) = P(X \leq x)$	1/36	3/36	6/36	10/36	15/36	21/36	26/36	30/36	33/36	35/36	1

Wir können die Dichtefunktion für alle reellen Zahlen definieren, indem wir $f(x)$ für alle Werte x, welche nicht in der Tabelle auftauchen, Null setzen. Zum Beispiel:

$$f(2) = P(X = 2) = 1/36 \qquad f(2.2) = P(X = 2.2) = 0 \qquad f(-101) = P(X = -101) = 0$$

So kann die Dichtefunktion der Zufallsvariablen als Funktion auf den reellen Zahlen aufgefasst werden:

$$f: \mathbb{R} \to [0,1], f(x) = P(X = x)$$

In der letzten Zeile der Tabelle stehen die Werte einer weiteren, für die Zufallsvariable sehr wichtigen Funktion, der *kumulativen Verteilungsfunktion* $F(x)$ (*cumulative distribution function*, kurz *CDF*) von X, formal definiert durch

$$F(x) = P(X \leq x)$$

Um die Werte der *CDF* von X zu bestimmen, werden Wahrscheinlichkeiten kumuliert. Zum Beispiel ist $F(4) = P(X \leq 4)$ die Wahrscheinlichkeit, dass eine Augensumme von höchstens 4 gewürfelt wird. Sie ergibt sich aus der Summe aller vorangehenden *PMF*-Werte:

$$F(4) = P(X \leq 4) = P(X = 4) + P(X \leq 3) = P(X = 4) + P(X = 3) + P(X \leq 2)$$
$$= P(X = 4) + P(X = 3) + P(X = 2)$$

Mit der *PMF* $f(x)$ und insbesondere der *CDF* $F(x)$ der Zufallsvariablen X lassen sich Wahrscheinlichkeiten von Aussagen systematisch berechnen.
Die Wahrscheinlichkeit, eine Augensumme zwischen 7 und 10 zu würfeln:

$$P(7 \leq X \leq 10) = F(10) - F(6)$$
$$= P(X \leq 10) - P(X \leq 6) = 33/36 - 15/36 = 18/36 = 1/2$$

Die Wahrscheinlichkeit, eine Augensumme grösser als 10 zu würfeln:

$$P(X > 10) = 1 - F(10) = 1 - P(X \leq 10) = 1 - 33/36 = 3/36 = 1/12$$

Die Wahrscheinlichkeit, eine Augensumme 3 oder 7 zu würfeln:

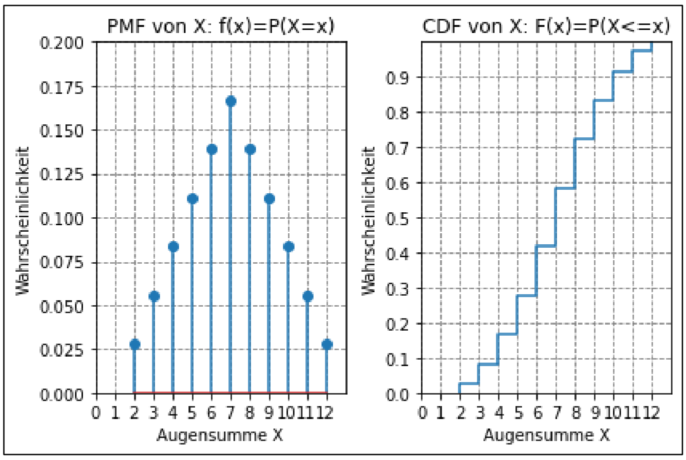
$$P(X = 3 \text{ oder } X = 7) = f(3) + f(7)$$
$$= P(X = 3) + P(X = 7) = 2/36 + 6/36 = 8/36 = 2/9$$

Die Wahrscheinlichkeit, die Augensumme 4 nicht zu würfeln:

$$P(X \neq 4) = 1 - P(X = 4) = 1 - f(4) = 1 - 3/36 = 33/36 = 11/12$$

Beide Funktionen, die *PMF* und die *CDF*, sind in ihrer Aussagekraft gleichwertig. Ist nur eine von beiden gegeben, so kann die andere daraus abgeleitet werden. Die *CDF* wird zur Berechnung der Wahrscheinlichkeit von Aussagen wie *Die Zufallsvariable nimmt Werte zwischen a und b an* oder *Die Zufallsvariable nimmt Werte kleiner (oder gleich) a an* etc., sogenannter *Intervallwahrscheinlichkeiten*, benutzt.

Häufig werden *PMF* und *CDF* in Diagrammen visualisiert. Nachfolgend sieht man das Stabdiagramm der *PMF* (links) und daneben die Treppenfunktion der *CDF* unseres Beispiels. Man beachte, dass die *CDF* eine rechtsseitig stetige Funktion ist.




```
import numpy as np
import matplotlib.pyplot as plt

X=np.arange(2,13)
PMFx=1/36*np.concatenate((np.arange(1,6),np.array
([6]),np.arange(5,0,-1)),axis=0)
CDFx=np.cumsum(PMFx)

plt.figure(1)
plt.subplot(121)
plt.stem(X,PMFx,use_line_collection='true')
plt.title('PMF von X: f(x)=P(X=x)')
plt.xlabel('Augensumme X')
plt.ylabel('Wahrscheinlichkeit')
plt.xticks([i for i in range(0,13)])
plt.grid(b=True, color='gray', linestyle='--')
plt.xlim(0,13)
plt.ylim(0,0.2)
plt.subplot(122)
plt.step(np.append(X,[12]),np.append([0],CDFx))
plt.title('CDF von X: F(x)=P(X<=x)')
plt.xlabel('Augensumme X')
plt.ylabel('Wahrscheinlichkeit')
plt.xticks([i for i in range(0,13)])
plt.yticks([i/10 for i in range(0,10)])
plt.grid(b=True, color='gray', linestyle='--')
plt.xlim(0,13)
plt.ylim(0,1)
plt.tight_layout()
```

Hier die Definitionen der wichtigsten Begriffe dieses Abschnitts im Überblick:

Definitionen
Eine *Zufallsvariable* X auf einem diskreten Ergebnisraum Ω ist eine Funktion $X: \Omega \to \mathbb{R}$.
Die reelle Funktion $f: \mathbb{R} \to [0,1], f(x) = P(X = x)$ heisst *Dichtefunktion* (*PMF*) von X.
Die reelle Funktion $F: \mathbb{R} \to [0,1], F(x) = P(X \leq x)$ heisst *kumulative Verteilungsfunktion* (*CDF*) von X.

Beispiele von Zufallsvariablen

Zufallsexperiment	**Ergebnis**	**Wert der Zufallsvariablen**
Lotto-Tipp	6 Zahlen aus 45	Gewinn
Entnahme einer Stichprobe für die Qualitätskontrolle	Menge von Objekten	Anzahl defekter Objekte in der Stichprobe
Übertragung eines 10-Bit-Signals	10er-Folge von 1 und 0, wobei 1 für ein korrekt und 0 für ein falsch übertragenes Bit steht	Anzahl korrekt übertragener Bits

Satz (Zufallsvariable, *PMF*, *CDF*)
Gegeben sind ein diskreter Wahrscheinlichkeitsraum (Ω, P) sowie eine Zufallsvariable $X: \Omega \to \mathbb{R}$. Dann haben *PMF* und *CDF* von X folgende Eigenschaften:
(1) $\sum_{x=-\infty}^{\infty} f(x) = 1$ und $F(z) = \sum_{x=-\infty}^{z} f(x)$
(2) $\lim\limits_{x\to\infty} F(x) = 1$ und $\lim\limits_{x\to-\infty} F(x) = 0$
(3) Monotonie: Aus $x \leq y$ folgt $F(x) \leq F(y)$
(4) $f(x) = F(x) - \lim\limits_{y\to x^-} F(y)$
(5) $P(a < X \leq b) = F(b) - F(a)$ und $P(a \leq X \leq b) = F(b) - \lim\limits_{x\to a^-} F(x)$
(6) $P(X > b) = 1 - F(b)$ und $P(X \geq b) = 1 - \lim\limits_{x\to b^-} F(x)$

4.3 Kenngrössen

Um verschiedene Verteilungen vergleichen zu können, bedient man sich sogenannter *Kenngrössen* wie zum Beispiel Erwartungswert, Varianz und Standardabweichung. Man teilt Kenngrössen in *Lagemasse* und *Streumasse* ein und verwendet sie, um die Verteilung eines Merkmals innerhalb einer Stichprobe (Statistik) oder die Verteilung von Zufallsvariablen (Stochastik) zu charakterisieren. Dabei beschreiben Lagemasse das Zentrum der Verteilung und Streumasse die Abweichungen vom Zentrum. Im folgenden Beispiel werden Kenngrössen benutzt, um die Verteilung der Prüfungsnoten bei zwei Klassen zu vergleichen.

Beispiel 2:
Es soll verglichen werden, wie gut zwei Klassen eine bestimmte Aufgabe gelöst haben. Folgende Tabelle zeigt die Punkteverteilung beider Klassen.

Punkte	0	1	2	3	4	5
Klasse 1	1	3	2	7	3	4
Klasse 2	0	1	5	9	5	0

Üblicherweise nimmt man nicht die Einzelergebnisse für den Vergleich: Man berechnet für jede Verteilung zwei Werte, einen für das Zentrum und einen für die Abweichung vom Zentrum. Das arithmetische Mittel oder kurz der *Mittelwert* der Noten ist unser Mass für das Zentrum. Für die Klasse 1 wird dieser so berechnet:

$$\mu_1 = \frac{1}{20}(1\cdot 0 + 3\cdot 1 + 2\cdot 2 + 7\cdot 3 + 3\cdot 4 + 4\cdot 5) = \frac{60}{20} = 3$$

Für die Klasse 2 erhalten wir den Mittelwert

$$\mu_2 = \frac{1}{20}(0\cdot 0 + 1\cdot 1 + 5\cdot 2 + 9\cdot 3 + 5\cdot 4 + 0\cdot 5) = \frac{58}{20} = 2.9$$

Man sieht, dass die beiden Klassen in etwa denselben Punktedurchschnitt haben. Allerdings gibt es einen markanten Unterschied bei der Verteilung der Punkte um den Mittelwert. Ein Mass dafür ist die durchschnittliche quadrierte Abweichung vom Mittelwert, die sogenannte (empirische) *Varianz*.

Für Klasse 1 erhalten wir die Varianz

$$\sigma_1^2 = \frac{1}{20}\left(1(0-3)^2 + 3(1-3)^2 + 2(2-3)^2 + 7(3-3)^2 + 3(4-3)^2 + 4(5-3)^2\right)$$

$$= 42/20 = 2.1$$

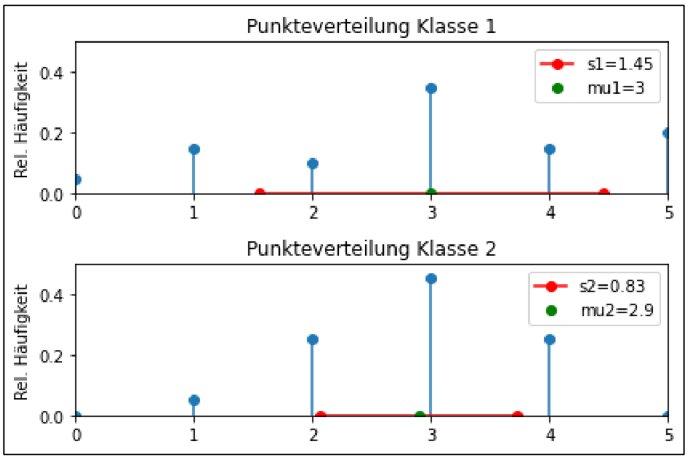
und für Klasse 2 erhalten wir die Varianz

$$\sigma_2^2 = \frac{1}{20}\left(1\cdot(1-2.9)^2 + 5\cdot(2-2.9)^2 + 9\cdot(3-2.9)^2 + 5\cdot(4-2.9)^2\right) = \frac{13.8}{20} = 0.69$$

Die Einheit der Varianz ist Punkte^2 .... Zieht man die Wurzel aus der Varianz, so erhält man ein weiteres Mass für die Streuung um den Mittelwert, die sogenannte (empirische) *Standardabweichung*, welche als Einheit diejenige des Mittelwerts hat, also hier Punkte ....

Die Punkteverteilungen der beiden Klassen weisen folgende Standardabweichungen auf:

$$\sigma_1 = \sqrt{\sigma_1^2} = \sqrt{2.1} \approx 1.45 \qquad \sigma_2 = \sqrt{\sigma_2^2} = \sqrt{0.69} \approx 0.83$$



Rechts sind die Punkteverteilungen der beiden Klassen abgebildet. Deutlich ist die grosse Streuung der Punkte bei der ersten Klasse. Die Standardabweichung ist ein Mass dafür. Bei fast gleichem Notenschnitt ist die Standardabweichung bei der ersten Klasse deutlich grösser als bei der zweiten Klasse.

An diesem Beispiel haben wir die Verwendung und Bedeutung von Kenngrössen in der deskriptiven Statistik kennengelernt, wo es darum geht, die Verteilung eines Merkmals in einer Stichprobe mithilfe bestimmter Parameter zu charakterisieren. Bei Zufallsvariablen werden diese Parameter entsprechend definiert. Dazu ein einfaches Beispiel.

Beispiel 3: Kenngrössen einer Zufallsvariablen
Folgende Tabelle gibt die Dichtefunktion (*PMF*) der Zufallsvariable X an.

x	-1	0	1	2
$f(x) = P(X = x)$	0.2	0.1	0.4	0.3

Man beachte, dass die *PMF*-Werte einer Zufallsvariablen den relativen Häufigkeiten bei einer Stichprobe entsprechen. So berechnet sich das arithmetische Mittel (bei Zufallsvariablen ist die Bezeichnung *Erwartungswert* üblich) der Zufallsvariable X als

$$\mu = E(X) = 0.2 \cdot (-1) + 0.1 \cdot 0 + 0.4 \cdot 1 + 0.3 \cdot 2 = 0.8$$

Die durchschnittliche quadrierte Abweichung vom Erwartungswert, die sogenannte *Varianz* der Zufallsvariablen X berechnet sich als

$$\sigma^2 = V(X) = 0.2 \cdot (-1 - 0.8)^2 + 0.1 \cdot (0 - 0.8)^2 + 0.4 \cdot (1 - 0.8)^2 + 0.3 \cdot (2 - 0.8)^2 = 1.16$$

Durch Wurzelziehen erhält man aus der Varianz die *Standardabweichung*:

$$\sigma = S(X) = \sqrt{V(X)} = \sqrt{\sigma^2} = \sqrt{1.16} \approx 1.077$$

Definitionen
Wir betrachten einen diskreten Wahrscheinlichkeitsraum (Ω, P) und eine Zufallsvariable $X: \Omega \to \mathbb{R}$. Dann definieren wir für X folgende Kenngrössen:

Der *Erwartungswert* $E(X) = \sum_{x\in\mathbb{R}} P(X = x) \cdot x$ ist ein *Lagemass* der Verteilung von X.
Die *Varianz* $V(X) = \sum_{x\in\mathbb{R}} P(X = x) \cdot (x - E(X))^2$ und die *Standardabweichung* $S(X) = \sqrt{V(X)}$ sind *Streumasse* der Verteilung von X.

Bemerkung
Diese Kenngrössen existieren nicht für jede Verteilung.

Satz (Kenngrössen)
Gegeben sind ein diskreter Wahrscheinlichkeitsraum (Ω, P) sowie eine Zufallsvariable $X: \Omega \to \mathbb{R}$. Dann haben die Kenngrössen von X folgende Eigenschaften:

(1) **Linearität des Erwartungswertes**:
$E(X+Y) = E(X) + E(Y)$ und $E(\alpha X) = \alpha E(X)$, mit $\alpha \in \mathbb{R}$.

(2) **Verschiebungssatz für die Varianz**
$$V(X) = E(X^2) - (E(X))^2 = \Big(\sum_{x \in \mathbb{R}} P(X = x) \cdot x^2\Big) - E(X)^2$$

(3) $V(\alpha X + \beta) = \alpha^2 \cdot V(X)$ mit $\alpha, \beta \in \mathbb{R}$.

Bemerkung
Im Allgemeinen gilt für die Verteilung einer Zufallsvariable X nach der *Tschebyscheff'schen Ungleichung*, dass mindestens 75% der Werte im Bereich $E(X) \pm 2 \cdot S(X)$ liegen. Man kann dies bei den beiden vorangehenden Beispielen verifizieren.
Die Tschebyscheff'sche Ungleichung lautet allgemein (für $k > 0$):
$$P(|X - E(X)| \geq k) \leq \frac{S(X)^2}{k^2}$$
bzw. für das komplementäre Ereignis
$$P(|X - E(X)| < k) \geq 1 - \frac{S(X)^2}{k^2}$$
Setzen wir $k^2 = 2 \cdot S(X)^2$ finden wir, dass mind. 50% der Werte im Bereich $E(X) \pm \sqrt{2} \cdot S(X)$ liegen.

In der Praxis sind viele Zufallsvariablen annähernd nach einer Gauss'schen Normalverteilung verteilt. Bei einer normalverteilten Zufallsvariable liegen etwa 68% der Werte im Bereich $E(X) \pm S(X)$ und etwa 95% im Bereich $E(X) \pm 2 \cdot S(X)$.

Beispiel 4: Gewinnspiel
Ein Spieler zahlt den Einsatz von 1 Franken und würfelt mit 5 (fairen) Würfel. Der Gewinn (abzüglich des Einsatzes) wird nach der folgenden Tabelle bestimmt.

Resultat	3er-Pasch	4er-Pasch	5er-Pasch	Full House	Grosse Strasse	Sonstiges
Gewinn	0	10	20	4	8	-1

Unter *Grosser Strasse* wollen wir 5 Zahlen in einer Reihe verstehen, also entweder 1,2,3,4,5 oder 2,3,4,5,6. *Full House* steht für 3 gleiche und 2 davon abweichende gleiche Zahlen. Die anderen in der Tabelle vorkommenden Würfelergebnisse werden unten noch erklärt. Wir fragen uns nach dem zu erwartenden Gewinn bei diesem Spiel. Eine weitere Frage wäre, wenn wir das Spiel 100 Mal spielen, was für ein Gewinn bzw. Verlust zu erwarten ist. Lohnt es, sich auf dieses Spiel einzulassen?

Um diese Fragen zu beantworten, fassen wir den Gewinn als Zufallsvariable G auf und bestimmen deren Dichte (*PMF*) und Verteilungsfunktion (*CDF*). Für jede Gewinnkategorie werden mithilfe von Kombinatorik die zutreffenden Möglichkeiten abgezählt. Für den *3er-Pasch* (also genau drei identische, und zwei davon abweichende, aber unterschiedliche Augenzahlen) geht man so vor:

$$\binom{5}{3} = \frac{5 \cdot 4 \cdot 3}{3 \cdot 2} = 10$$

- Aus den 5 Würfeln kann man auf ………………………… Arten 3 Würfel auswählen.
- Diese 3 Würfel können (für einen 3er-Pasch) mit ……6 verschiedenen Zahlen besetzt werden;
- Für den 4. Würfel bleiben ……5 mögliche Zahlen (ohne dass es einen 4er-Pasch gibt);
- Für den 5. Würfel bleiben ……4 mögliche Zahlen (ohne dass es einen 4er-Pasch oder ein Full House gibt).

Folglich gilt:

Anzahl Möglichkeiten für einen **3er-Pasch** mit 5 Würfeln: $\binom{5}{3} \cdot 6 \cdot 5 \cdot 4 = 1200$

Entsprechend lässt sich berechnen:

Anzahl Möglichkeiten für einen **4er-Pasch** mit 5 Würfeln: $\binom{5}{4} \cdot 6 \cdot 5 = 150$

Anzahl Möglichkeiten für einen **5er-Pasch** mit 5 Würfeln: $\binom{5}{5} \cdot 6 = 6$

Für ein **Full House** müssen es genau 3 gleiche und 2 davon abweichende gleiche Zahlen sein:

$$\binom{5}{3} = 10$$

- Aus den 5 Würfeln kann man auf ……………… Arten 3 Würfel auswählen.
- Diese 3 Würfel können (für einen 3er-Pasch) mit ……6 verschiedenen Zahlen besetzt werden.
- Die verbleibenden 2 Würfel können mit ….5 verschiedenen Zahlen besetzt werden.

Somit gilt:

Anzahl Möglichkeiten für ein **Full House** mit 5 Würfeln: $\binom{5}{3} \cdot 6 \cdot 5 = 300$

Für die **Grosse Strasse** sind die Zahlenreihen 1,2,3,4,5 und 2,3,4,5,6 möglich. Jede dieser beiden Reihen kann mit 5 Würfeln auf ……………5! = 120 Arten gewürfelt werden; also ergibt sich:

Anzahl Möglichkeiten für eine **Grosse Strasse** mit 5 Würfeln: $2 \cdot 5! = 240$

Insgesamt gibt es mit 5 Würfeln $6^5 = 7776$ verschiedene Möglichkeiten.

Unten in der Tabelle sind die Würfelergebnisse nach aufsteigendem Gewinn geordnet angegeben. Dies erleichtert die Berechnung der *CDF*, da dann fortschreitend die vorhergehenden Spalteneinträge kumuliert werden können.

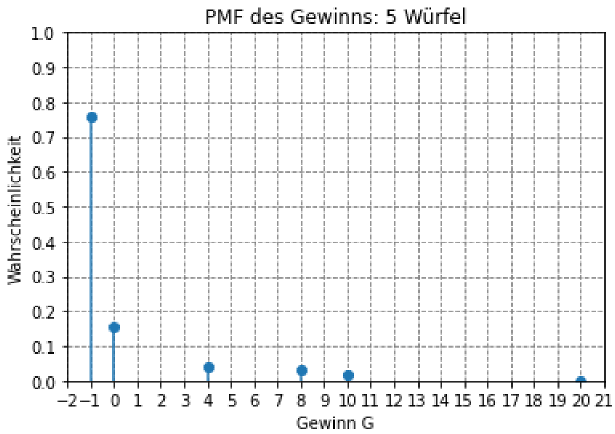
Resultat	Sonstiges	3er-Pasch	Full House	Grosse Strasse	4er-Pasch	5er-Pasch
Gewinn x	-1	0	4	8	10	20
Anzahl Möglich-keiten	7776-1896 =5880	1200	300	240	150	6
$P(G=x)$	$=\frac{5880}{6^5}$ $=\frac{245}{324}$ ≈ 0.7562	$=\frac{1200}{6^5}$ $=\frac{25}{162}$ ≈ 0.1543	$=\frac{300}{6^5}$ $=\frac{25}{648}$ ≈ 0.0386	$=\frac{240}{6^5}$ $=\frac{5}{162}$ ≈ 0.0309	$=\frac{150}{6^5}$ $=\frac{25}{1296}$ ≈ 0.0193	$=\frac{6}{6^5}$ $=\frac{1}{1296}$ ≈ 0.00077
$P(G \leq x)$	$=\frac{5880}{6^5}$	$=\frac{7080}{6^5}$	$=\frac{7380}{6^5}$	$=\frac{7620}{6^5}$	$=\frac{7770}{6^5}$	1

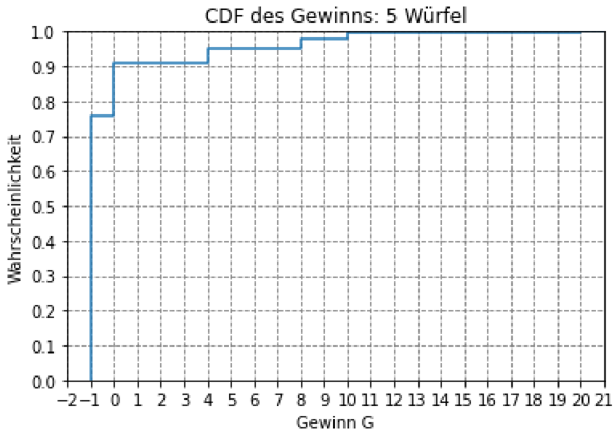
Unter $6^5 = 7776$ Würfelexperimenten mit 5 fairen Würfeln gibt es, da es sich um eine Laplace-Verteilung handelt, durchschnittlich 1200-mal einen 3er-Pasch, 300-mal Full House, 240-mal Grosse Strasse, 150-mal 4er Pasch, 6-mal 5er Pasch und 5880-mal Sonstiges. Werden diese absoluten Häufigkeiten noch mit dem zugehörigen Gewinn G multipliziert, dann summiert und durch $6^5 = 7776$ geteilt, erhält man das arithmetische Mittel des Gewinns, also den erwarteten Gewinn pro Spiel. In der Stochastik heisst dieser Wert *Erwartungswert* der Zufallsvariable G, in Zeichen $E(G)$.

$$E(G) = \frac{1}{6^5}[G(Sonstiges)\cdot 5880 + G(3erPasch)\cdot 1200 + G(FullHouse)\cdot 300$$
$$+G(GrosseStrasse)\cdot 240 + G(4erPasch)\cdot 150 + G(5erPasch)\cdot 6]$$
$$= \frac{1}{6^5}[-1\cdot 5880 + 0\cdot 1200 + 4\cdot 300 + 8\cdot 240 + 10\cdot 150 + 20\cdot 6]$$
$$= \frac{-95}{648} \approx -0.1466$$

Pro Spiel wird demnach ein Gewinn von Fr. -0.15 erwartet. Bei 100 Spielen ergibt das einen durchschnittlichen Gewinn von Fr. -15.

Auf der folgenden Seite sind die Dichtefunktion (*PMF*) und die Verteilungsfunktion (*CDF*) dargestellt.





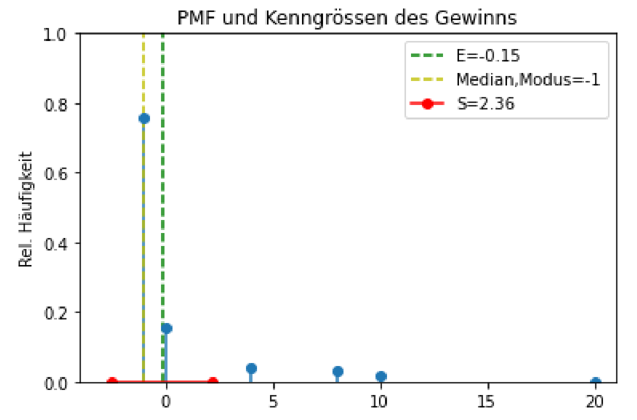

```
import numpy as np
import matplotlib.pyplot as plt
from scipy import special as sp

G=np.array([-1,0,4,8,10,20])
Total=6**5 #Anzahl Möglichkeiten
M=np.zeros(6)
M[1]=sp.binom(5,3)*6*5*4
M[2]=sp.binom(5,3)*6*5
M[3]=2*sp.factorial(5)
M[4]=sp.binom(5,4)*6*5
M[5]=6
M[0]=Total-np.sum(M)
PMF=1/Total*M #PMF des Gewinns
CDF=np.cumsum(PMF) #CDF des Gewinns
E=np.sum(G*PMF) #Erwartungswert
S2=np.sum(G**2*PMF)-E**2 #Varianz
S=np.sqrt(S2) #Standardabweichung

plt.figure(1)
plt.stem(G,PMF,use_line_collection='true',basefmt='None')
plt.title('PMF des Gewinns: 5 Würfel')
plt.xlabel('Gewinn G')
plt.ylabel('Wahrscheinlichkeit')
plt.xticks([i for i in range(-2,22)])
plt.yticks([i/10 for i in range(0,11)])
plt.grid(b=True, color='gray', linestyle='--')
plt.xlim(-2,21)
plt.ylim(0,1)

plt.figure(2)
plt.step(np.append(G,[21]),np.append([0],CDF))
plt.title('CDF des Gewinns: 5 Würfel')
plt.xlabel('Gewinn G')
plt.ylabel('Wahrscheinlichkeit')
plt.xticks([i for i in range(-2,22)])
plt.yticks([i/10 for i in range(0,11)])
plt.grid(b=True, color='gray', linestyle='--')
plt.xlim(-2,21)
plt.ylim(0,1)
```

Wir berechnen die Varianz des Gewinns G mithilfe des Verschiebungssatzes:

$$V(G) = E(G^2) - E(G)^2 = G(Sonstiges)^2 \cdot \frac{5880}{6^5} + G(3erPasch)^2 \cdot \frac{1200}{6^5}$$
$$+G(FullHouse)^2 \cdot \frac{300}{6^5} + G(GrosseStrasse)^2 \cdot \frac{240}{6^5}$$
$$+G(4erPasch)^2 \cdot \frac{150}{6^5} + G(5erPasch)^2 \cdot \frac{6}{6^5} - E(G)^2$$
$$= (-1)^2 \cdot \frac{245}{324} + 0^2 \cdot \frac{25}{162} + 4^2 \cdot \frac{25}{648} + 8^2 \cdot \frac{5}{162} + 10^2 \cdot \frac{25}{1296} + 20^2 \cdot \frac{1}{1296} - [\frac{-95}{648}]^2$$
$$\approx 5.5649$$

Die Zufallsvariable G hat die Einheit Fr., und der Erwartungswert ebenfalls. Die Varianz allerdings hat als Einheit das Quadrat der zugrundeliegenden Einheit, also hier Fr^2. Zieht man die Wurzel aus der Varianz, gelangt man zur sogenannten *Standardabweichung* $S(G)$, mit der gleichen Einheit Fr. wie die Zufallsvariable G selbst:

Im untenstehenden Bild sind die Kenngrössen für die gegebene Verteilung von G zusammen mit der *PMF* visualisiert.




```
E=np.sum(G*PMF) #Erwartungswert
S2=np.sum(G**2*PMF)-E**2 #Varianz
S=np.sqrt(S2) #Standardabweichung

plt.figure(3)
plt.plot([E,E],[0,1],'g--',label='E=-0.15')
plt.plot([-1,-1],[0,1],'y--',label='Median,Modus=-1')
plt.plot([E-S,E+S],[0,0],'ro-', label='S=2.36')
plt.stem(G,PMF,use_line_collection='true',basefmt='None'
)
plt.legend()
plt.title('PMF und Kenngrössen des Gewinns')
plt.ylabel('Rel. Häufigkeit')
plt.xlim(-4,21)
plt.ylim(0,1)
```

4.4 Bedingte Wahrscheinlichkeiten

Die Wahrscheinlichkeit des Eintretens eines Ereignisses kann durch das Eintreten eines anderen Ereignisses erheblich beeinflusst werden. Das Risiko an Lungenkrebs zu erkranken, ist davon beeinflusst, ob man ein starker bzw. schwacher bzw. kein Raucher ist oder ob man für längere Zeit bestimmten Strahlungen ausgesetzt ist, beispielsweise ob man in unmittelbarer Nähe eines AKWs wohnt, etc. Die Wahrscheinlichkeit, unter einer bestimmten Anzahl von Losen einen der n Hauptgewinne zu ziehen, verändert sich, wenn man bereits ein Los gezogen hat. Bestimmt man die Wahrscheinlichkeiten von Ereignissen, so ist die Kenntnis solcher Abhängigkeiten von entscheidender Bedeutung.

Beispiel 5

Die Tabelle zeigt Frauen und Männer einer Firma, unterteilt nach Rauchern und Nichtrauchern. Beantworten Sie aufgrund dieser Tabelle (einer sogenannten *Vierfeldertafel*) die untenstehenden Fragen.

	F (Frauen)	M (Männer)	Summe
R (Raucher/in)	100	400	500
N (Nichtraucher/in)	200	700	900
Summe	300	1100	1400

Sie gehen durch die Firma und treffen zufällig auf eine Person (wir gehen davon aus, dass jede/r Angestellte dieselbe Chance hat, Ihnen zu begegnen).

(a) Wie gross ist die Wahrscheinlichkeit, eine/n Raucher/in anzutreffen?
500/1400 = 5/14 = P(R)
……………………………………………

(b) Wie gross ist die Wahrscheinlichkeit, eine Frau anzutreffen?
300/1400 = 3/14 = P(F)
……………………………………………

(c) Wie gross ist die Wahrscheinlichkeit, eine Raucherin anzutreffen?
100/1400 = 1/14 = P(F ∩ R)
……………………………………………

(d) Sie treffen eine Frau an. Mit welcher Wahrscheinlichkeit ist sie Raucherin?
100/300 = 1/3 = P(R | F) (Bezeichnungen für bedingte W. nach Definition hinzufügen)
……………………………………………

(e) Mit welcher Wahrscheinlichkeit ist eine Person, die raucht, ein Mann?
400/500 = 4/5 = P(M | R)
……………………………………………
In den letzten beiden Fällen betrachten wir nicht mehr alle 1400 Mitarbeitenden der Firma, sondern nur noch eine bestimmte Untergruppe. Entsprechend steht im Nenner nicht 1400, sondern die Anzahl Mitglieder dieser Untergruppe, also die Anzahl derer, die eine bestimmte 'Bedingung' erfüllen. Untergruppen für (d) und (e) je farbig umrahmen.

Mithilfe des klassischen Wahrscheinlichkeitsbegriffs können wir aus der obigen Tabelle die Wahrscheinlichkeiten für die verschiedenen Ereignisse ableiten, indem wir je die Anzahl „günstiger" Fälle durch die Anzahl „möglicher Fälle" (also das Total von 1400) teilen:

	F (Frauen)	M (Männer)	Summe
R (Raucher/in)	$P(F \cap R) = \frac{100}{1400}$	$P(M \cap R) = \frac{400}{1400}$	$P(R) = \frac{500}{1400}$
N (Nichtraucher/in)	$P(F \cap N) = \frac{200}{1400}$	$P(M \cap N) = \frac{700}{1400}$	$P(N) = \frac{900}{1400}$
Summe	$P(F) = \frac{300}{1400}$	$P(M) = \frac{1100}{1400}$	$P(\Omega) = \frac{1400}{1400} = 1$

Die Antworten auf Fragen (d) und (e) können wir auch mithilfe von diesen Wahrscheinlichkeiten ausdrücken: (Bezeichnungen für bedingte W. nach Definition hinzufügen)

(d) $\frac{P(F \cap R)}{P(F)} = \frac{\frac{100}{1400}}{\frac{300}{1400}} = \frac{1}{3} = P(R|F)$

(e) $\frac{P(M \cap R)}{P(R)} = \frac{\frac{400}{1400}}{\frac{500}{1400}} = \frac{4}{5} = P(M|R)$

Allgemeine Fragestellung
Wie gross ist die Wahrscheinlichkeit für ein Ereignis B, wenn wir nur die Versuchsergebnisse berücksichtigen, bei denen ein Ereignis A eintritt?

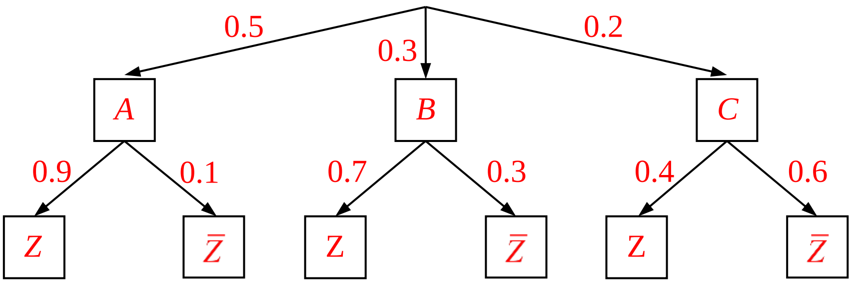
Definition
Die Wahrscheinlichkeit für das Eintreten des Ereignisses B unter der Bedingung oder Voraussetzung, dass das Ereignis A eintritt, heisst *bedingte Wahrscheinlichkeit* von B unter der Bedingung A. Schreibweise: $P(B|A)$.

Sie wird durch die folgende Gleichung definiert:

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

Ein anschauliches Hilfsmittel für die Berechnung von bedingten Wahrscheinlichkeiten sind sogenannte *Ereignisbäume*. Wir erklären diese an folgendem Beispiel:

Beispiel 6
Eine Firma beschäftigt drei Angestellte, die telefonische Anfragen von Kunden beantworten sollen. Herr Alleskönner kann 90% aller Fragen zur Zufriedenheit der Kunden beantworten, Frau Besserwisser 70% und Herr Chancenlos nur gerade 40%. 50% der Kunden werden von Herrn Alleskönner betreut, 30% von Frau Besserwisser und 20% von Herrn Chancenlos.

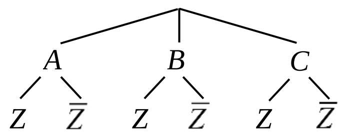


Bei einem Ereignisbaum müssen folgende Bedingungen erfüllt sein:

- Alle von einem Blatt (Verzweigungspunkt) aus mit Pfeilen erreichbaren Ereignisse sind paarweise disjunkt (d.h. sie schliessen sich gegenseitig aus).
- Die Summe der Übergangswahrscheinlichkeiten aller von einem Blatt ausgehenden Pfeile ist 1.

Mithilfe des Ereignisbaumes wollen wir nun folgende Fragen beantworten:

- Wie gross ist die Wahrscheinlichkeit, dass ein Kunde an Herrn Alleskönner gerät und eine zufrieden stellende Antwort bekommt? (Relevanten Pfad einzeichnen)

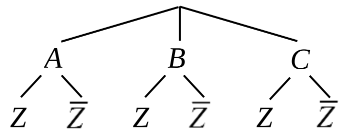


$$P(A \cap Z) = P(A) \cdot P(Z|A) = 0.5 \cdot 0.9 = 0.45$$

Allgemein gilt:
Die Wahrscheinlichkeiten längs eines Pfades werden miteinander multipliziert;
…………………………………………………………………………………………..
so erhält man die sogenannte Pfadwahrscheinlichkeit.
…………………………………………………………………………………………..

…………………………………………………………………………………………..

- Wie gross ist die Wahrscheinlichkeit, dass ein Kunde mit der Antwort, die er erhält, nicht zufrieden ist? (Relevante Pfade einzeichnen)



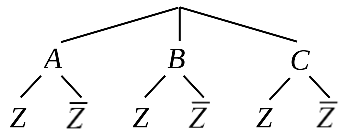
$$P(\bar{Z}) = P(A)\cdot P(\bar{Z}|A) + P(B)\cdot P(\bar{Z}|B) + P(C)\cdot P(\bar{Z}|C)$$

$$= 0.5\cdot 0.1 + 0.3\cdot 0.3 + 0.2\cdot 0.6 = 0.05 + 0.09 + 0.12 = 0.26$$

Allgemein gilt:
Um die Wahrscheinlichkeit für ein bestimmtes Endergebnis zu berechnen,
……………………………………………………………………………………………..
addiert man die Pfadwahrscheinlichkeiten aller Pfade, die zu diesem Ergebnis führen.
……………………………………………………………………………………………..

……………………………………………………………………………………………...

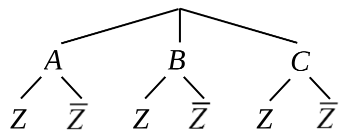
- Wie gross ist die Wahrscheinlichkeit, dass ein zufriedener Kunde von Herrn Chancenlos beraten wurde? (Pfad für Zähler in einer Farbe, Pfade für Nenner in einer anderen Farbe einzeichnen)



$$P(C|Z) = \frac{P(C\cap Z)}{P(Z)} = \frac{0.2\cdot 0.4}{0.5\cdot 0.9+0.3\cdot 0.7+0.2\cdot 0.4}$$

$$= \frac{0.08}{0.45+0.21+0.08} = \frac{8}{74} = \frac{4}{37} \approx 0.11$$

- Wie gross ist die Wahrscheinlichkeit, dass ein unzufriedener Kunde von Frau Besserwisser beraten wurde? (Pfad für Zähler in einer Farbe, Pfade für Nenner in einer anderen Farbe einzeichnen)



$$P(B|\bar{Z}) = \frac{P(B\cap \bar{Z})}{P(\bar{Z})} = \frac{0.3\cdot 0.3}{0.5\cdot 0.1+0.3\cdot 0.3+0.2\cdot 0.6}$$

$$= \frac{0.09}{0.05+0.09+0.12} = \frac{9}{26} \approx 0.35$$

Satz (Bedingte Wahrscheinlichkeit)
Seien (Ω, P) ein diskreter Wahrscheinlichkeitsraum und A und B zwei Ereignisse, d.h. Teilmengen von Ω. Dann gilt:

(1) **Multiplikationssatz** – *Pfadwahrscheinlichkeit*

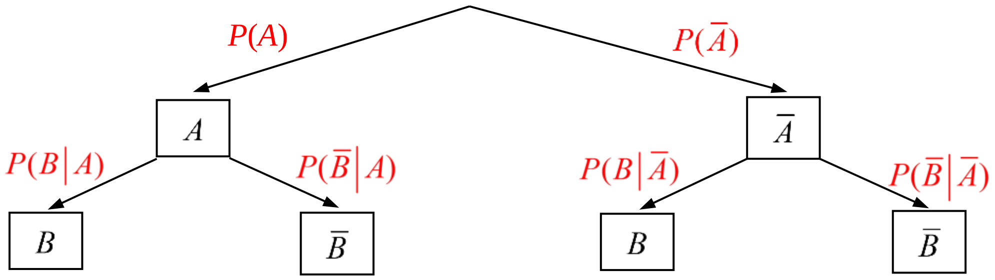
$$P(A\cap B) = P(A)\cdot P(B|A) = P(B)\cdot P(A|B)$$

(2) **Satz von der Totalen Wahrscheinlichkeit**

$$P(B) = P(A)\cdot P(B|A) + P(\bar{A})\cdot P(B|\bar{A})$$

(3) **Satz von Bayes**

$$P(A|B) = \frac{P(A)\cdot P(B|A)}{P(B)}$$



Beispiel 7
Die in einem Warenhaus angebotenen Steckdosen stammen aus den beiden Fabriken $F1$ und $F2$. Dabei liefert $F1$ 80% des Angebots. Man weiss, dass Lieferungen aus $F1$ zu 5% und Lieferungen von $F2$ zu 10% fehlerhaft sind. Wir benützen die folgenden Abkürzungen:

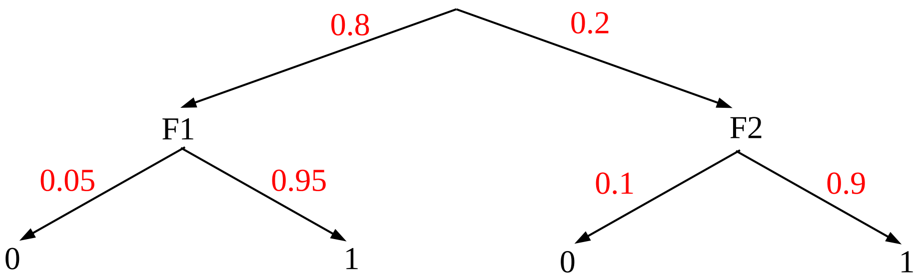
$F1$: Steckdose stammt aus $F1$.

$F2$: Steckdose stammt aus $F2$.

0: Steckdose ist fehlerhaft.

1: Steckdose ist nicht fehlerhaft.

Der unten abgebildete Wahrscheinlichkeitsbaum zeigt die Zusammenhänge für diese vier Ereignisse:



- Wie gross ist die Wahrscheinlichkeit, dass eine bestimmte Steckdose nicht fehlerhaft ist und aus Fabrik $F1$ stammt?

$$P(F1 \cap 1) = P(F1) \cdot P(1|F1) = 0.8 \cdot 0.95 = 0.76$$

- Wie gross ist die Wahrscheinlichkeit, dass eine bestimmte Steckdose defekt ist?

$$P(0) = P(F1 \cap 0) + P(F2 \cap 0) = P(F1) \cdot P(0|F1) + P(F2) \cdot P(0|F2)$$

$$= 0.8 \cdot 0.05 + 0.2 \cdot 0.1 = 0.06$$

- Wie gross ist die Wahrscheinlichkeit, dass eine bestimmte Steckdose aus der Fabrik $F1$ stammt oder fehlerhaft ist?

$$P(F1 \cup 0) = P(F1 \cap 1) + P(0) = 0.76 + 0.06 = 0.82$$

- Wie gross ist die Wahrscheinlichkeit, dass eine defekte Steckdose aus Fabrik $F1$ stammt?

$$P(F1|0) = \frac{P(F1 \cap 0)}{P(0)} = \frac{P(F1) \cdot P(0|F1)}{P(0)} = \frac{0.8 \cdot 0.05}{0.06} = \frac{2}{3}$$

- Wie gross ist die Wahrscheinlichkeit, dass eine defekte Steckdose aus Fabrik $F2$ stammt?

$$P(F2|0) = 1 - P(F1|0) = 1 - \frac{2}{3} = \frac{1}{3}$$

4.5 Stochastische Unabhängigkeit

Man sagt, zwei Ereignisse sind *stochastisch unabhängig*, wenn das Eintreten des einen Ereignisses keinen Einfluss auf das Eintreten des anderen hat (weder im negativen noch im positiven Sinn). Die Steuer- und Schutzsysteme bei sicherheitskritischen Anwendungen, beispielsweise im Kraftwerk oder in Flugzeugen, sind so implementiert, dass es mehrere redundante Systeme gibt, die nach dem Ausfall eines oder mehrerer Teilsysteme unabhängig voneinander alle nötigen Aufgaben durchführen können. Es ist erstrebenswert, dass der Ausfall des einen Teilsystems den Ausfall des anderen Teilsystems nicht beeinflusst. Dass zwei parallele Schutzsysteme stochastisch unabhängig voneinander sind, ist in der Praxis allerdings kaum zu erreichen. Dazu braucht es komplette Diversität, d.h. Unabhängigkeit und Verschiedenheit im Design und in der Implementierung.

Wenn A und B stochastisch unabhängig sind, beeinflusst das Eintreten des einen Ereignisses also das Eintreten des anderen Ereignisses nicht.
Falls $P(B) \neq 0$, so gilt dann: $P(A|B) = P(A)$
und falls $P(A) \neq 0$, so gilt: $P(B|A) = P(B)$.

Definition (Stochastische Unabhängigkeit)
Sei (Ω, P) ein diskreter Wahrscheinlichkeitsraum.

- Zwei Ereignisse A und B heissen *stochastisch unabhängig*, falls gilt:

  $$P(A \cap B) = P(A) \cdot P(B)$$

  Andernfalls heissen A und B *stochastisch abhängig*.

- Zwei Zufallsvariablen $X: \Omega \to \mathbb{R}$ und $Y: \Omega \to \mathbb{R}$ heissen *stochastisch unabhängig*, falls

  $$P(X = x \wedge Y = y) = P(X = x) \cdot P(Y = y) \text{ für alle } x, y \in \mathbb{R}$$

  Andernfalls heissen die Zufallsvariablen *stochastisch abhängig*.
  Dabei bezeichnet $P(X = x \wedge Y = y)$ die Wahrscheinlichkeit, dass die beiden Ereignisse $\{X = x\}$ und $\{Y = x\}$ gleichzeitig eintreten, d.h. dass die Zufallsvariable X den Wert x annimmt und gleichzeitig die Zufallsvariable Y den Wert y annimmt. Die Funktion $f(x, y) = P(X = x \wedge Y = y)$ heisst die *gemeinsame Verteilung* (*Verbundverteilung*) von X und Y.

Beispiel 8
Beim einmaligen Würfeln mit einem fairen Würfel betrachten wir die drei Ereignisse:

$gerade$: Die gewürfelte Zahl ist gerade.

≥ 3: Die gewürfelte Zahl ist grösser oder gleich 3.

$3teilt$: Die gewürfelte Zahl ist durch 3 teilbar.

Es gilt:

$P(gerade \cap \geq 3) = \frac{|\{4,6\}|}{|\Omega|} = \frac{1}{3}$ $\qquad$ $P(gerade) \cdot P(\geq 3) = \frac{1}{2} \cdot \frac{2}{3} = \frac{1}{3}$

Also sind die Ereignisse $gerade$ und ≥ 3 stochastisch unabhängig ……
In anderen Worten: Das Eintreten des einen der beiden Ereignisse beeinflusst das Eintreten des anderen Ereignisses in keiner Weise. Dieser Sachverhalt lässt sich alternativ sehr prägnant mit bedingten Wahrscheinlichkeiten ausdrücken:

$$P(gerade| \geq 3) = \frac{P(gerade \cap \geq 3)}{P(\geq 3)} = \frac{|\{4,6\}|}{|\{3,4,5,6\}|} = \frac{1}{2} = P(gerade)$$

und

$$P(\geq 3|gerade) = \frac{P(gerade \cap \geq 3)}{P(gerade)} = \frac{|\{4,6\}|}{|\{2,4,6\}|} = \frac{2}{3} = P(\geq 3)$$

Es gilt:

$P(gerade \cap 3teilt) = \frac{|\{6\}|}{|\Omega|} = \frac{1}{6}$ $\qquad$ $P(gerade) \cdot P(3teilt) = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$

Also sind die Ereignisse $gerade$ und $3teilt$ stochastisch unabhängig ……

Es gilt:

$P(\geq 3 \cap 3teilt) = \frac{|\{3,6\}|}{|\Omega|} = \frac{1}{3}$ $\qquad$ $P(\geq 3) \cdot P(3teilt) = \frac{2}{3} \cdot \frac{1}{3} \neq \frac{1}{3}$

Also sind die Ereignisse ≥ 3 und $3teilt$ stochastisch abhängig ……
Das bedeutet, dass das Eintreten des einen der beiden Ereignisse das Eintreten des anderen beeinflusst
……

$$P(\geq 3|3teilt) = \frac{P(3teilt \wedge \geq 3)}{P(3teilt)} = \frac{|\{3,6\}|}{|\{3,6\}|} = 1 > \frac{2}{3} = P(\geq 3)$$

erhöht
Das Eintreten des Ereignisses $3teilt$ .... die Wahrscheinlichkeit für das Ereignis ≥ 3.

$$P(3teilt| \geq 3) = \frac{P(3teilt \wedge \geq 3)}{P(\geq 3)} = \frac{|\{3,6\}|}{|\{3,4,5,6\}|} = \frac{1}{2} > \frac{1}{3} = P(3teilt)$$

erhöht
Das Eintreten des Ereignisses ≥ 3 .... die Wahrscheinlichkeit für das Ereignis $3teilt$.

Beispiel 9: Glücksräder
Wir betrachten die beiden unten abgebildeten, aufeinander befestigten Glücksräder (klein und gross). Die Räder sind aufeinander fixiert und werden (gemeinsam) im Uhrzeigersinn gedreht. Der schwarze Pfeil zeigt auf die Gewinnfarbe für das äussere und das innere Rad, sobald das Rad zum Stillstand kommt. Die Zufallsvariable X nimmt den Wert 1 an, wenn der schwarze Pfeil auf den roten Bereich des äusseren (grossen) Rads zeigt, und sonst 0. Die Zufallsvariable Y nimmt den Wert 1 an, wenn der schwarze Pfeil auf den roten Bereich des inneren (kleinen) Rads zeigt, und sonst 0. In der Tabelle unten links ist die gemeinsame Verteilung

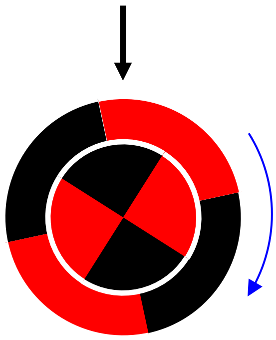
$$f(x,y) = P(X = x \wedge Y = y)$$

von X und Y zu sehen. Zum Beispiel tritt der Fall, dass der Pfeil auf einen roten äusseren Bereich und einen schwarzen inneren Bereich zeigt, in einem Viertel aller Fälle auf, da die gemeinsame rote äussere und schwarze innere Fläche der beiden Räder einen Viertel der Kreisfläche ausmacht. Somit folgt

$$f(1,0) = P(X = 1 \wedge Y = 0) = \text{.........}\ 1/4$$

Dies gilt ebenso für alle anderen Kombinationen.

y \ x	0 (schwarz)	1 (rot)	$P(Y = y)$
0 (schwarz)	1/4	1/4	1/2
1 (rot)	1/4	1/4	1/2
$P(X = x)$	1/2	1/2	1 (Spalten- bzw. Zeilensumme)



In der letzten Zeile der Tabelle stehen die Werte der Dichtefunktion (*PMF*) von X. Die jeweiligen Werte ergeben sich als Summe aller darüber liegenden Werte (Spaltensumme) der gemeinsamen Verteilung. In der vierten Spalte der Tabelle stehen die Werte der Dichtefunktion (*PMF*) von Y. Die jeweiligen Werte ergeben sich als Summe der links daneben liegenden Werte (Zeilensumme) der gemeinsamen Verteilung. Weil sie am Tabellenrand stehen, nennt man die Dichtefunktionen (*PMF*) von X und von Y auch *Randverteilungen*. Die Summe aller Wahrscheinlichkeiten der gemeinsamen Verteilung muss 1 ergeben, genauso wie es für die Dichtefunktionen (*PMF*) von X und Y gilt, daher der Eintrag 1 im Feld rechts unten.

Sind die Zufallsvariablen X und Y in diesem Beispiel stochastisch unabhängig? Nach Definition müssen wir überprüfen, ob gilt:

$$P(X = x \wedge Y = y) = P(X = x) \cdot P(Y = y) \text{ für alle möglichen Werte } x \text{ und } y.$$

Wir überprüfen dies exemplarisch für $x = 1$ und $y = 0$:

$P(X = 1 \wedge Y = 0) = \frac{1}{4}$ $\qquad$ $P(X = 1) \cdot P(Y = 0) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$

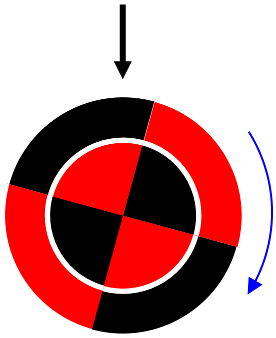
Die Rechnungen für die anderen Werte x und $y \in \{0,1\}$ gehen genau gleich, da alle Wahrscheinlichkeiten in den inneren Feldern und alle Wahrscheinlichkeiten in den äusseren Feldern jeweils gleich sind. Wir können somit schliessen:

stochastisch unabhängig

Die Zufallsvariablen X und Y sind ………………………………………………………………..

Betrachten wir zum Vergleich noch ein Glücksrad, bei dem das innere und das äussere Rad in einer anderen Position zueinander sind:

y \ x	0 (schwarz)	1 (rot)	$P(Y = y)$
0 (schwarz)	0	1/2	1/2
1 (rot)	1/2	0	1/2
$P(X = x)$	1/2	1/2	1 (Spalten- bzw. Zeilensumme)



Wie steht es nun um die stochastische Unabhängigkeit von X und Y? Wir betrachten wieder exemplarisch den Fall $x = 1$ und $y = 0$:

$P(X = 1 \wedge Y = 0) = \frac{1}{2}$ $\qquad$ $P(X = 1) \cdot P(Y = 0) = \frac{1}{2} \cdot \frac{1}{2} \neq \frac{1}{2}$

Wir schliessen daraus:

stochastisch abhängig

Die Zufallsvariablen X und Y sind ………………………………………………………………..

Bemerkungen

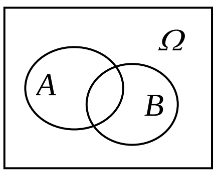
- Für zwei Ereignisse A und B sind die folgenden drei Aussagen gleichbedeutend:

  (i) A und B sind stochastisch unabhängig

  (ii) A und $\Omega\backslash B$ (das Komplement von B) sind stochastisch unabhängig

  (iii) $\Omega\backslash A$ und $\Omega\backslash B$ sind stochastisch unabhängig.

  Wir zeigen dazu: Wenn A und B stochastisch unabhängig sind, so folgt die stochastische Unabhängigkeit von A und $\Omega\backslash B$. (In Venndiagramm einzeichnen)

  $P(A \cap (\Omega\backslash B)) = P(A) - P(A \cap B) = P(A) - P(A) \cdot P(B)$

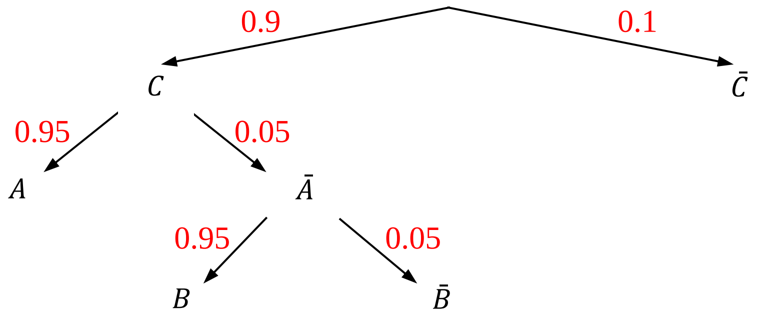
  $= P(A) \cdot (1 - P(B)) = P(A) \cdot P(\Omega\backslash B)$

- Wenn die Ereignisse A und B stochastisch unabhängig sind, beeinflusst das Eintreten des einen Ereignisses das Eintreten des anderen Ereignisses nicht.

  Denn falls $P(B) \neq 0$, so gilt: $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A) \cdot P(B)}{P(B)} = P(A)$

  und falls $P(A) \neq 0$, so gilt: $P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A) \cdot P(B)}{P(A)} = P(B)$

Beispiel 10:
Eine Maschinenanlage besteht aus den Blöcken A, B und C. Die drei Blöcke können (stochastisch) unabhängig voneinander ausfallen. Der Ausfall von Block A geschieht mit Wahrscheinlichkeit $P(\bar{A}) = 0.05$, der Ausfall von Block B mit $P(\bar{B}) = 0.05$ und der Ausfall von Block C mit $P(\bar{C}) = 0.10$. Die Anlage als Ganzes kann nicht mehr arbeiten, wenn A und B zugleich ausfallen oder wenn C ausfällt. Wie wahrscheinlich ist es, dass die gesamte Anlage ausfällt?
Im folgenden Wahrscheinlichkeitsbaum haben wir alle für einen
Ausfall relevanten Pfade visualisiert. Bei den Übergangswahrscheinlichkeiten ist zu beachten, dass die Vorgeschichte keine Rolle spielt, da die Ereignisse stochastisch unabhängig sind.



Der komplette Ausfall der Anlage wird durch die beiden folgenden Pfade beschrieben:

- Block A und Block B fallen aus und Block C arbeitet korrekt ……………; und
- Block C fällt aus ……………

Die Wahrscheinlichkeit, dass die Anlage ausfällt, entspricht der Summe der Pfadwahrscheinlichkeiten der beiden Pfade:

$$P(\bar{A} \cap \bar{B} \cap C) + P(\bar{C}) \underbrace{=}_{A,B,C\ st.unabhängig} P(\bar{A}) \cdot P(\bar{B}) \cdot P(C) + P(\bar{C})$$

$$= 0.9 \cdot 0.05 \cdot 0.05 + 0.1 = 0.10225$$

Satz (Stochastische Unabhängigkeit – Ereignisse)
Wir betrachten einen diskreten Wahrscheinlichkeitsraum (Ω, P) und zwei Ereignisse A und $B \subseteq \Omega$.

Folgende Eigenschaften sind äquivalent:

(i) A und B sind stochastisch unabhängig.
(ii) A und $\Omega \backslash B$ sind stochastisch unabhängig.
(iii) $\Omega \backslash A$ und $\Omega \backslash B$ sind stochastisch unabhängig.

Liegen mehr als zwei Ereignisse $A_1, A_2, \dots, A_n$ vor, dann heissen diese (vollständig) stochastisch unabhängig, wenn für jede natürliche Zahl k mit $1 < k \leq n$ und für jede Teilmenge von k Ereignissen $\{A_{i1}, A_{i2}, \dots, A_{ik}\}$ der Menge $\{A_1, A_2, \dots, A_n\}$ gilt:

$$P(A_{i1} \cap A_{i2} \cap \dots \cap A_{ik}) = P(A_{i1}) \cdot P(A_{i2}) \cdot \ldots \cdot P(A_{ik})$$

Anmerkung: Es genügt dafür dann also nicht nur jeweils die paarweise Unabhängigkeit zu prüfen. Zum Nachweis der Unabhängigkeit von n Ereignissen sind $2^n - n - 1$ Gleichungen zu überprüfen.

Für drei Ereignisse A_1, A_2, A_3 bedeutet das, dass sie stochastisch unabhängig sind, wenn die folgenden vier Gleichungen erfüllt sind:

(1) $P(A_1 \cap A_2) = P(A_1) \cdot P(A_2)$
(2) $P(A_1 \cap A_3) = P(A_1) \cdot P(A_3)$
(3) $P(A_2 \cap A_3) = P(A_2) \cdot P(A_3)$
(4) $P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2) \cdot P(A_3)$

Mit der *stochastischen Unabhängigkeit* erleichtert sich die Berechnung der *Erwartungswerte* und *Varianzen* bei *Zufallsvariablen*.
Für stochastisch unabhängige Zufallsvariablen X und Y gilt:

$$E(X \cdot Y) = \sum_{x,y \in \mathbb{R}} x \cdot y \cdot \underbrace{P(X = x \wedge Y = y)}_{=P(X=x) \cdot P(Y=y)} = \sum_{x \in \mathbb{R}} \sum_{y \in \mathbb{R}} x \cdot y \cdot P(X = x) \cdot P(Y = y)$$
$$= \sum_{x \in \mathbb{R}} x \cdot P(X = x) \cdot \sum_{y \in \mathbb{R}} y \cdot P(Y = y) = E(X) \cdot E(Y)$$

Für die Varianz der Summe folgt:

$$V(X+Y) = E(\underbrace{(X+Y)^2}_{=X^2+2XY+Y^2}) - \underbrace{E(X+Y)^2}_{=E(X)^2+2E(X)E(Y)+E(Y)^2}$$
$$= \underbrace{2E(X \cdot Y) - 2E(X)E(Y)}_{=0} + \underbrace{E(X^2) - E(X)^2}_{=V(X)} + \underbrace{E(Y^2) - E(Y)^2}_{=V(Y)}$$
$$= V(X) + V(Y)$$

Fassen wir diese Erkenntnisse in einem Satz zusammen:

Satz (Stochastische Unabhängigkeit – Zufallsvariablen)
Wir betrachten einen diskreten Wahrscheinlichkeitsraum (Ω, P) sowie zwei stochastisch unabhängige Zufallsvariablen $X, Y : \Omega \rightarrow \mathbb{R}$.

- Falls die Erwartungswerte $E(X)$ und $E(Y)$ existieren, so gilt:

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

- Falls die Varianzen $V(X)$ und $V(Y)$ existieren, so gilt:

$$V(X + Y) = V(X) + V(Y)$$

Bemerkung
Entsprechendes gilt, wenn wir mehr als zwei Zufallsvariablen haben.

4.6 Lernziele für dieses Kapitel

- ☐ Ich kann *Mengenkalkül* und vom *Zahlenkalkül* unterscheiden und anwenden.
- ☐ Ich kann die Rechenregeln des *Wahrscheinlichkeitsmasses P* unter Berücksichtigung des Mengenkalküls korrekt anwenden.
- ☐ Ich kann stochastische Modelle aufstellen, d.h. ich kann den *Ergebnisraum*, den *Ereignisraum*, die *Zähldichte* und das *Wahrscheinlichkeitsmass* für einfache *diskrete Wahrscheinlichkeitsräume* definieren.
- ☐ Ich kenne die grundlegenden Eigenschaften eines *Wahrscheinlichkeitsmasses* und kann Wahrscheinlichkeiten von *Durchschnitt, Vereinigung* und *Komplement* von Ereignissen berechnen.
- ☐ Ich kann beurteilen, ob es sich bei einem diskreten Wahrscheinlichkeitsraum um eine *Gleichverteilung* (*Laplace-Verteilung*) handelt.
- ☐ Ich kann den Begriff einer *Zufallsvariable* anhand eines Beispiels erläutern.
- ☐ Ich kann *Zufallsvariablen* definieren und deren *Wahrscheinlichkeitsdichte (PMF)* und *kumulative Verteilungsfunktion (CDF)* mithilfe einer Tabelle bestimmen.
- ☐ Ich kann die Graphen von *PMF* bzw. *CDF* als Stabdiagramm bzw. Treppenfunktion zeichnen.
- ☐ Ich kann die *CDF* aus der *PMF* berechnen und umgekehrt.
- ☐ Ich kenne die Bedeutung der *CDF* und kann diese zur Berechnung von *Intervallwahrscheinlichkeiten* einsetzen.
- ☐ Ich kann die Bedeutung von *Kenngröss*en (*Lagemasse, Streuungsmasse*) für Zufallsvariablen beschreiben.
- ☐ Ich kann die Definitionen von *Erwartungswert*, *Varianz* und *Standardabweichung* auswendig formulieren.
- ☐ Ich kann den *Erwartungswert*, die *Varianz* und die *Standardabweichung* diskreter Zufallsvariablen bestimmen.
- ☐ Ich kenne die elementaren Rechenregeln für den *Erwartungswert* von Summen und skalaren Vielfachen von Zufallsvariablen.
- ☐ Ich kann die *Varianz* von Zufallsvariablen mit dem Verschiebungssatz berechnen.
- ☐ Ich kenne die *Rechenregel* für die Varianz von skalaren Vielfachen von Zufallsvariablen.
- ☐ Ich kann die Definition der *bedingten Wahrscheinlichkeit* für Ereignisse anhand von Beispielen erläutern.
- ☐ Ich kann *bedingte Wahrscheinlichkeiten* berechnen.
- ☐ Ich kann *Vierfeldertafeln* anfertigen und für die Berechnung von bedingten Wahrscheinlichkeiten einsetzen.

- ☐ Ich kann *Ereignisbäume* anfertigen.
- ☐ Ich kann den *Multiplikationssatz* am Ereignisbaum erläutern und anwenden.
- ☐ Ich kann den *Satz von der Totalen Wahrscheinlichkeit* am Ereignisbaum erläutern und anwenden.
- ☐ Ich kann den *Satz von Bayes* am Ereignisbaum erläutern und anwenden.
- ☐ Ich kann nachprüfen, ob Ereignisse *stochastisch unabhängig* bzw. *abhängig* sind.
- ☐ Ich kann die *gemeinsame Verteilung (Verbundverteilung)* von Zufallsvariablen in einer Tabelle aufstellen und die *Randverteilungen* berechnen.
- ☐ Ich kann beurteilen, ob Zufallsvariablen *stochastisch abhängig* bzw. *unabhängig* sind.