

RAFAEL PEÑALOZA

WAHRSCHEINLICHKEITS- THEORIE UND STATISTIK

FREIE UNIVERSITÄT BOZEN

Disclaimer

DIESES DOKUMENT ENTHÄLT das Vorlesungsskript für den Kurs über Wahrscheinlichkeitstheorie und Statistik in den Bachelor-Studiengängen an der Fakultät für Informatik der Freien Universität Bozen. Es wurde von Rafaele Peñaloza auf Englisch verfasst und von Werner Nutt ins Deutsche übersetzt. Der größte Teil des Inhalts basiert auf dem Buch von Sheldon M. Ross.¹ Es gibt derzeit fünf Ausgaben dieses Buches; jede von ihnen sollte für die Zwecke des Kurses geeignet sein.²

Dieses Skriptum ersetzt nicht das Material, das während des Kurses präsentiert wird. Es kann sachliche Fehler und Tippfehler enthalten und relevante Details, die in der Vorlesung besprochen wurden können fehlen.³ Verwenden Sie es mit Vorsicht. Am besten nehmen Sie an der Vorlesung teil, machen sich selbst Notizen und verwenden dieses Dokument als Hilfsmittel.

Hintergrundinformationen und zusätzliches Material wie Kopien der Vorlesungsnotizen, Handouts, Übungen und Code werden auf der OLE-Seite für den Kurs bereitgestellt.

¹ Sheldon M. Ross, Introduction to Probability and Statistics for Engineers and Scientists, Academic Press, 2014

² Die fünfte Ausgabe kann von Angehörigen der Universität Bozen kostenlos **hier** heruntergeladen werden.

³ Für alle Fehler in diesem Skriptum tragen Autor und Übersetzer die Verantwortung. Wenn Sie einen Fehler bemerken, informieren Sie bitte den Dozenten Ihrer Vorlesung, damit er behoben werden kann.

Inhaltsverzeichnis

Einführung in die Wahrscheinlichkeitstheorie 5

Zufallsvariablen 19

Spezielle Zufallsvariablen 38

Einführung in die Wahrscheinlichkeitstheorie

DIE WAHRSCHEINLICHKEITSTHEORIE IST ein Mittel, um die Ungewissheit über das Eintreten von Ereignissen zu messen und darüber zu sprechen. Das Ereignis wird eintreten oder nicht, aber wir wissen nicht, was von beiden der Fall sein wird. Nehmen wir zum Beispiel an, dass wir einen Würfel werfen. Dann wird auf jeden Fall eine Seite oben liegen, aber wir wissen nicht, welche: das Ereignis „die 6 wird oben liegen“ ist ungewiss.

Wahrscheinlichkeiten quantifizieren diese Unsicherheit. In diesem Fall würden wir etwa sagen: „Die Wahrscheinlichkeit, dass der Würfel eine 6 zeigt, beträgt $1/6$ “; denn es gibt sechs mögliche Ergebnisse, die wir als gleich wahrscheinlich ansehen. Wir können diese Aussage auf zwei Arten verstehen:

- Wenn das Experiment oft genug wiederholt wird, dann ist das Ergebnis in $1/6$ der Fälle eine 6.
- Wir glauben, es gibt bei diesem speziellen Wurf eine Möglichkeit von 1 aus 6, dass die 6 oben landet.

Die erste ist die frequentistische Sichtweise. Sie geht davon aus, dass eine Wahrscheinlichkeit dem Ereignis inhärent ist und durch die Wiederholung eines Experiments untersucht und bestimmt werden kann. Die zweite ist die subjektive (oder Bayes'sche) Sichtweise, bei der sich die Wahrscheinlichkeit auf die Überzeugung desjenigen bezieht, der die Wahrscheinlichkeit angibt. Obwohl diese beiden Ansichten einige tiefgreifende philosophische Unterschiede aufweisen, haben sie keine Auswirkungen auf die Untersuchung der mathematischen Eigenschaften von Wahrscheinlichkeiten. Daher lassen wir diese Unterschiede bei der formalen Untersuchung der Wahrscheinlichkeitstheorie oft außer Acht.

Ereignisse

WIR WERDEN UNS mit *Experimenten* befassen, deren Ausgang ungewiss ist, bei denen aber die Menge aller möglichen *Ergebnisse* (engl. *outcomes*) bekannt ist.

Diese Menge aller möglichen Ergebnisse nennen wir *Stichprobenraum* (engl. *sample space*), und wir schreiben ihn als \mathcal{S} . Wenn wir einen Würfel werfen, ist der Stichprobenraum $\mathcal{S} = \{1, \dots, 6\}$. Stichprobenräume können sehr einfach oder sehr komplex sein. Bei einem Wettlauf beispielsweise

kann der Stichprobenraum aus allen möglichen Reihenfolgen bestehen, in denen die Teilnehmer ins Ziel kommen können. Im manchen Fällen ist \mathcal{S} unendlich.

Jede Teilmenge des Stichprobenraums ist ein *Ereignis* (engl. *event*). Wenn das Ergebnis des Experiments in einem gegebenen Ereignis \mathcal{E} als Element enthalten ist, dann sagen wir, dass \mathcal{E} *eingetreten* ist (engl. *occurred*). Zum Beispiel ist das Ereignis, dass der Würfel eine gerade Zahl zeigt, $\mathcal{E} = \{2, 4, 6\}$.⁴

Da Ereignisse Mengen sind, können wir neue Ereignisse durch die Anwendung von Mengenoperationen auf andere Ereignisse definieren. Für zwei Ereignisse \mathcal{E} und \mathcal{F} sind ihre *Vereinigung* $\mathcal{E} \cup \mathcal{F}$ (engl. *union*), ihr *Durchschnitt* $\mathcal{E} \cap \mathcal{F}$ (engl. *intersection*), und ihr *Komplement* $\overline{\mathcal{E}}$ (engl. *complement*) ebenfalls Ereignisse.⁵

Wenn in unserem Würfelbeispiel $\mathcal{E} = \{2, 4, 6\}$ das Ereignis „Würfeln einer geraden Zahl“ ist und $\mathcal{F} = \{2, 3, 5\}$ das Ereignis „Würfeln einer Primzahl“, dann können wir daraus weitere Ereignisse bilden:

- $\mathcal{E} \cup \mathcal{F} = \{2, 3, 4, 5, 6\}$: Würfeln einer geraden Zahl oder einer Primzahl,
- $\mathcal{E} \cap \mathcal{F} = \{2\}$: Würfeln einer ungeraden Primzahl, und
- $\overline{\mathcal{E}} = \{1, 3, 5\}$: Würfeln einer ungeraden Zahl.

Es kann sein, dass ein Ereignis leer ist, zum Beispiel, wenn man den Durchschnitt zweier Ereignisse nimmt, die keine gemeinsamen Ergebnisse haben, wie $\{2\}$ und $\{3\}$. Dieses Ereignis wird das *leere Ereignis* oder das *unmögliche Ereignis* genannt und als \emptyset geschrieben, wie in $\{2\} \cap \{3\} = \emptyset$. Wenn $\mathcal{E} \cap \mathcal{F} = \emptyset$, dann *schließen sich \mathcal{E} und \mathcal{F} gegenseitig aus* (engl. *\mathcal{E} and \mathcal{F} are mutually exclusive*). Man beachte, dass der gesamte Stichprobenraum ebenfalls ein Ereignis ist, und dass $\overline{\emptyset} = \mathcal{S}$.⁶

Genau wie zwischen Mengen können wir auch Inklusionen zwischen Ereignissen betrachten. Das Ereignis \mathcal{E} ist in \mathcal{F} *enthalten* (engl. *contained*), wenn alle Ergebnisse von \mathcal{E} auch Ergebnisse von \mathcal{F} sind. In diesem Fall schreiben wir $\mathcal{E} \subseteq \mathcal{F}$ oder $\mathcal{F} \supseteq \mathcal{E}$. Wenn $\mathcal{E} \subseteq \mathcal{F}$ und $\mathcal{F} \subseteq \mathcal{E}$, dann haben sie dieselbe Bedeutung und wir sagen \mathcal{E} und \mathcal{F} sind *äquivalent* (engl. *equivalent*) und schreiben $\mathcal{E} \equiv \mathcal{F}$. Zum Beispiel ist das Ereignis „Würfeln einer Zwei“ im Ereignis „Würfeln einer geraden Zahl“ enthalten und ist äquivalent zum Würfeln einer geraden Primzahl.

Axiome der Wahrscheinlichkeitstheorie

WIE SIE SICH ERINNERN, besagt die frequentistische Sichtweise, dass bei wiederholter Durchführung eines Experiments unter identischen Bedingungen der Anteil der Ergebnisse, die zu einem bestimmten Ereignis \mathcal{E} gehören, gegen eine Konstante konvergiert, die die Wahrscheinlichkeit von \mathcal{E} ausdrückt.

Aus mathematischer Sicht gibt es für jedes Ereignis \mathcal{E} in einem Stichprobenraum \mathcal{S} eine Zahl, die wir die *Wahrscheinlichkeit von \mathcal{E}* (engl. *probability of \mathcal{E}*) nennen und als $P(\mathcal{E})$ schreiben, so dass die folgenden drei Axiome erfüllt sind:

Axiom 1 $0 \leq P(\mathcal{E}) \leq 1$

⁴ Für andere Räume sind natürlich komplexere Ereignisse möglich.

⁵ Eine kurze Einführung in die Mengenlehre und Venn-Diagramme finden Sie im entsprechenden Handout.

⁶ Sprachlich kann man die Vereinigung als ein (nicht ausschließendes) *oder*, den Durchschnitt als ein *und* und das Komplement als ein *nicht* verstehen.

Axiom 2 $P(\mathcal{S}) = 1$ und

Axiom 3 für jede Folge von sich gegenseitig ausschließenden Ereignissen $\mathcal{E}_1, \mathcal{E}_2, \dots,$ ⁷ und für jedes $n \in \mathbb{N}$ gilt, dass $P(\bigcup_{i=1}^n \mathcal{E}_i) = \sum_{i=1}^n P(\mathcal{E}_i)$.

⁷ Das heißt Ereignisse, so dass $\mathcal{E}_i \mathcal{E}_j = \emptyset$ wenn $i \neq j$.

Das erste Axiom besagt, dass die Wahrscheinlichkeit eines beliebigen Ereignisses immer eine Zahl zwischen 0 und 1 ist. Axiom 2 macht eine Aussage über die Vollständigkeit von Wahrscheinlichkeiten: Die Wahrscheinlichkeit, ein beliebiges Ergebnis aus dem Stichprobenraum zu beobachten, ist immer 1. Das letzte Axiom zeigt, wie man die Wahrscheinlichkeiten von sich gegenseitig ausschließenden Ereignissen zusammenfasst: Die Wahrscheinlichkeit ihrer Vereinigung ist die Summe ihrer einzelnen Wahrscheinlichkeiten.

Die frequentistische Sichtweise der Wahrscheinlichkeiten erfüllt eindeutig diese drei Axiome: Für jedes Ereignis \mathcal{E} liegt der Anteil (oder die Häufigkeit) der Fälle, in denen bei einem wiederholten Experiment ein Ergebnis aus \mathcal{E} eintritt, notwendigerweise zwischen 0 und 1; bei jeder Wiederholung des Experiments gehört das Ergebnis zum Stichprobenraum S ; und wenn zwei Ereignisse \mathcal{E} und \mathcal{F} kein gemeinsames Ergebnis haben, dann ist der Anteil der Fälle, in denen ein Ergebnis aus \mathcal{E} oder \mathcal{F} eintritt, die Summe der jeweiligen Häufigkeiten. Als Beispiel für dieses letzte Axiom nehmen wir an, dass \mathcal{E} das Ereignis ist, bei dem der Würfel eine gerade Zahl zeigt, und \mathcal{F} , dass er die Zahl 1 zeigt. Dann erwarten wir, dass \mathcal{E} in der Hälfte (50 Prozent) der Fälle eintritt und \mathcal{F} in $1/6$ der Fälle. Somit sollte $\mathcal{E} \cup \mathcal{F}$ in $2/3$ der Fälle beobachtet werden.

Aus diesen einfachen Axiomen lassen sich viele Eigenschaften von Wahrscheinlichkeiten ableiten und formal verstehen. Wir beginnen mit zwei einfachen Sätzen, die bereits einige interessante Aspekte der Theorie aufzeigen.

Satz 1. Für jedes Ereignis \mathcal{E} gilt $P(\overline{\mathcal{E}}) = 1 - P(\mathcal{E})$.

Beweis. Nach Definition sind \mathcal{E} und $\overline{\mathcal{E}}$ disjunkt, und $\mathcal{S} = \mathcal{E} \cup \overline{\mathcal{E}}$. Aus Axiom 3 folgt, dass $P(\mathcal{S}) = P(\mathcal{E} \cup \overline{\mathcal{E}}) = P(\mathcal{E}) + P(\overline{\mathcal{E}})$. Aus Axiom 2 folgt, dass $P(\mathcal{S}) = 1$ ist, und damit $P(\overline{\mathcal{E}}) = 1 - P(\mathcal{E})$. \square

Mit anderen Worten: Die Wahrscheinlichkeit, dass ein Ereignis nicht eintritt, ist immer eins minus der Wahrscheinlichkeit, dass es eintritt. Wenn zum Beispiel bei einem manipulierten Würfel die Wahrscheinlichkeit, eine gerade Zahl zu würfeln, 0,6 ist, dann muss die Wahrscheinlichkeit, eine ungerade Zahl zu würfeln, 0,4 sein.

Beachte, dass Axiom 3 eine Möglichkeit bietet, die Wahrscheinlichkeit der Vereinigung von zwei oder mehr Ereignissen zu berechnen, allerdings nur, wenn sie sich gegenseitig ausschließen. Oft wollen wir solche Wahrscheinlichkeiten für Ereignisse berechnen, die nicht unbedingt disjunkt sind. Das ist in der Tat möglich, wenn wir die Wahrscheinlichkeit für den Durchschnitt kennen.

Satz 2. Für jeweils zwei Ereignisse \mathcal{E}, \mathcal{F} gilt

$$P(\mathcal{E} \cup \mathcal{F}) = P(\mathcal{E}) + P(\mathcal{F}) - P(\mathcal{E} \mathcal{F}).$$

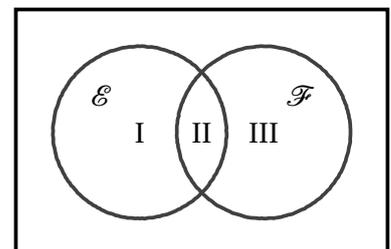


Abbildung 1: Drei relevante Regionen in einem Venn-Diagramm.

Beweis. Das Venn-Diagramm in Abbildung 1 erleichtert den Beweis. Die drei Bereiche I, II und III schließen sich gegenseitig aus. Daraus folgt, dass

$$P(\mathcal{E} \cup \mathcal{F}) = P(\text{I}) + P(\text{II}) + P(\text{III});$$

$$P(\mathcal{E}) = P(\text{I}) + P(\text{II});$$

$$P(\mathcal{F}) = P(\text{II}) + P(\text{III}).$$

Somit ist $P(\mathcal{E} \cup \mathcal{F}) = P(\mathcal{E}) + P(\mathcal{F}) - P(\text{II})$. Da $\text{II} \equiv \mathcal{E}\mathcal{F}$ ist, schließt dies den Beweis ab. \square

Beispiel 3. Von allen Erwachsenen trinken 23 Prozent Bier, 7 Prozent Wein, und 2 Prozent beides, Bier und Wein. Wieviel Prozent der Erwachsenen trinken weder Bier noch Wein?

Lösung. Um zu wissen, wie viele Personen weder Bier noch Wein trinken, reicht es aus, zu wissen, wie viele Personen mindestens eines dieser Getränke trinken (das gesuchte Ereignis ist das Komplement dieses Ereignisses). Sei E das Ereignis, dass eine Person Bier trinkt, und F das Ereignis, dass sie Wein trinkt. Die Wahrscheinlichkeit, eines der beiden Getränke zu trinken, ist⁸

$$P(\mathcal{E} \cup \mathcal{F}) = P(\mathcal{E}) + P(\mathcal{F}) - P(\mathcal{E}\mathcal{F}) = 0.23 + 0.07 - 0.02 = 0.28.$$

Die Wahrscheinlichkeit, dass eine Person keinen Alkohol trinkt, ist $1 - 0.28 = 0.72$. \triangle

EIN WEITERER WICHTIGER BEGRIFF, wenn es um Unsicherheit und Wahrscheinlichkeit geht, ist der Begriff der Chance (engl. *odds*).

Definition 4. Die *Chance*, dass ein Ereignis \mathcal{E} eintritt ist definiert als $\frac{P(\mathcal{E})}{1 - P(\mathcal{E})}$.

Intuitiv sagt die Chance für ein Ereignis \mathcal{E} aus, wie viel wahrscheinlicher es ist, dass \mathcal{E} eintritt als dass es nicht eintritt. Wenn die Chance für \mathcal{E} größer als 1 ist, ist es wahrscheinlicher, dass \mathcal{E} eintritt, als dass es nicht eintritt, und umgekehrt, wenn sie kleiner als 1 ist. Wenn zum Beispiel der Wurf einer 2 mit einem Würfel die Wahrscheinlichkeit $P(\mathcal{E}) = \frac{1}{6}$ hat, dann ist die Chance für dieses Ereignis $\frac{1/6}{5/6} = \frac{1}{5}$. Folglich ist es fünfmal wahrscheinlicher, keine 2 zu sehen, als eine 2 zu würfeln. In diesem Fall sagen wir, dass „die Chancen 1 zu 5 gegen das Ereignis \mathcal{E} stehen“. ⁹

⁸ Man kann über dies Problem auch auf anderem Wege nachdenken: Wie viele Menschen trinken *nur* Bier? Das sind diejenigen, die Bier trinken, abzüglich derer, die Bier und Wein trinken: $P(\mathcal{E}) - P(\mathcal{E}\mathcal{F})$. Diejenigen, die *nur* Wein trinken, sind $P(\mathcal{F}) - P(\mathcal{E}\mathcal{F})$. Die Personen, die Bier *oder* Wein trinken, sind also diejenigen, die nur Bier trinken, plus diejenigen, die nur Wein trinken, plus diejenigen, die beides trinken: $P(\mathcal{E}) - P(\mathcal{E}\mathcal{F}) + P(\mathcal{F}) - P(\mathcal{E}\mathcal{F}) + P(\mathcal{E}\mathcal{F})$.

⁹ Oder, umgekehrt, *zugunsten von* $\bar{\mathcal{E}}$.

Gleichverteilte Wahrscheinlichkeiten

ES GIBT VIELE FÄLLE, in denen man davon ausgehen kann, dass jeder Punkt im Stichprobenraum \mathcal{S} mit der gleichen Wahrscheinlichkeit auftritt. Wenn der Stichprobenraum *endlich* ist (zum Beispiel, wenn $\mathcal{S} = \{1, \dots, n\}$ für irgendein $n \in \mathbb{N}$), dann bedeutet dies, dass

$$P(\{1\}) = \dots = P(\{n\}) = p.$$

Aus den Axiomen 2 und 3, folgt, dass

Axiome der Wahrscheinlichkeitstheorie:

1. $0 \leq P(\mathcal{E}) \leq 1$
2. $P(\mathcal{S}) = 1$
3. $P(\bigcup_{i=1}^n \mathcal{E}_i) = \sum_{i=1}^n P(\mathcal{E}_i)$ wenn sich alle \mathcal{E}_i s gegenseitig ausschließen.

$$1 = P(\mathcal{S}) = P(\{1\}) + \dots + P(\{n\}) = np,$$

und damit gilt für all i , dass $P(\{i\}) = p = \frac{1}{n}$. Das heißt, jedes Ergebnis hat die gleiche Eintrittswahrscheinlichkeit, die durch den Gesamtumfang des Stichprobenraums bestimmt wird.

Wir können Axiom 3 erneut verwenden, um dieses Ergebnis auf beliebige Ereignisse zu verallgemeinern. Für ein Ereignis \mathcal{E} sei $\#\mathcal{E}$ die Anzahl der Elemente (Ergebnisse) in \mathcal{E} . Wenn alle Ergebnisse mit gleicher Wahrscheinlichkeit in einem Stichprobenraum mit n Elementen auftreten,¹⁰ dann ist

$$P(\mathcal{E}) = \frac{\#\mathcal{E}}{n}.$$

Das heißt, die Wahrscheinlichkeit von \mathcal{E} ist gleich dem Anteil, den die Elemente von \mathcal{E} am Stichprobenraum haben. Aus diesem Grund ist es oft wichtig, die verschiedenen Arten abzählen zu können, auf die ein Ereignis eintreten kann. Ein grundlegendes Werkzeug für das Abzählen von Möglichkeiten ist die Produktregel der Kombinatorik.

Prinzip (Produktregel der Kombinatorik). Wenn zwei Experimente so angelegt sind, dass Experiment 1 zu einem von m verschiedenen Ergebnissen führt, und für jedes von ihnen Experiment 2 zu einem von n Ergebnissen führt, dann gibt es insgesamt mn verschiedene mögliche Ergebnisse für die beiden Experimente.

Um zu zeigen, dass dies der Fall ist, reicht es aus, alle möglichen Ergebnisse der beiden Experimente aufzuzählen und sie in einer Matrix anzuordnen, wobei jede Zeile dem Ergebnis von Experiment 1 und jede Spalte dem Ergebnis von Experiment 2 entspricht. Diese Matrix hat insgesamt mn Elemente (siehe Abbildung 2).

(1, 1)	(1, 2)	...	(1, n)
(2, 1)	(2, 2)	...	(2, n)
⋮	⋮	⋱	⋮
(m, 1)	(m, 2)	...	(m, n)

Abbildung 2: Die Produktregel der Kombinatorik erfasst alle möglichen Kombinationen.

Beispiel 5. Eine Schublade enthält 8 schwarze und 6 weiße Socken. Wir nehmen „zufällig“ zwei Socken aus der Schublade. Wie groß ist die Wahrscheinlichkeit, dass sie kein Paar bilden?¹¹

Lösung. Zunächst ermitteln wir die Größe des Stichprobenraums: die möglichen Ergebnisse sind die Kombinationen von jeweils zwei Socken, die wir nacheinander aus der Schublade nehmen. Anfangs befinden sich 14 Socken darin. Zuerst können wir eine beliebige dieser 14 Socken wählen, und dann eine beliebige der 13 verbleibenden. Insgesamt gibt es also $14 \cdot 13 = 182$ mögliche Kombinationen. Damit eine Kombination kein Paar bildet, müssen wir erst eine schwarze und dann eine weiße Socke herausnehmen ($8 \cdot 6 = 48$ Kombinationen) oder erst eine weiße und dann eine schwarze Socke ($6 \cdot 8 = 48$ Kombinationen). Unter der Annahme, dass jede Kombination des Stichprobenraums mit gleicher Wahrscheinlichkeit auftritt,¹² ergibt sich somit eine Wahrscheinlichkeit von $\frac{48+48}{182} = \frac{96}{182} \approx 0.53$. \triangle

Die Produktregel der Kombinatorik lässt sich natürlich auf Fälle verallgemeinern, in denen mehr als zwei Experimente durchgeführt werden. Wenn r Experimente durchgeführt werden, so dass Experiment 1 zu einem von n_1 verschiedenen Ergebnissen führt, und für jedes dieser Ergebnisse Experiment 2 zu einem von n_2 Ergebnissen führt, und so weiter, dann gibt es insgesamt $n_1 \cdot n_2 \cdots n_r$ verschiedene mögliche Ergebnisse für die r Experimente.

¹⁰ Wir sprechen in diesem Fall häufig von einer *uniformen* oder *gleichförmigen Verteilung* der Wahrscheinlichkeiten.

¹¹ Das heißt, wir suchen die Wahrscheinlichkeit dafür, dass die beiden ausgewählten Socken unterschiedliche Farben haben.

¹² Das ist es, was wir oft meinen, wenn wir sagen, dass ein Experiment „zufällig“ erfolgt.

Betrachten wir zum Beispiel die Frage, wie viele Möglichkeiten es gibt, eine Menge von n verschiedenen Objekten auf einer Linie anzuordnen. Für ein kleines n können wir diese Zahl durch Aufzählen bestimmen: die möglichen Reihenfolgen der Buchstaben a , b und c sind abc , acb , bac , bca , cab , und cba ; das heißt, es gibt 6 solcher Anordnungen. Wir bezeichnen diese Anordnungen als *Permutationen* (engl. *permutations*). Es gibt also sechs Permutationen für eine Menge von drei Objekten. Anstatt sie alle aufzuzählen, was unmöglich (oder zumindest mühsam) wird, wenn die Anzahl der Objekte wächst, kann man die Anzahl der Permutationen auch nach der Produktregel der Kombinatorik ermitteln: Das erste Objekt in der Permutation kann ein beliebiges von drei möglichen sein, das zweite eines der verbleibenden zwei, und das letzte wird durch die vorherigen Auswahlen bestimmt (das heißt, es bleibt nur eine Möglichkeit). Mit anderen Worten: Es gibt $3 \cdot 2 \cdot 1 = 6$ Permutationen.

Wenn wir n Objekte anstatt 3 haben, können wir mit demselben Argument herleiten, dass es von ihnen $n(n-1)(n-2)\cdots 3 \cdot 2 \cdot 1$ verschiedene Permutationen gibt. Dieser Ausdruck wird *n-Fakultät* (engl. *n factorial*) genannt und $n!$ geschrieben.¹³

Beispiel 6. Wir wollen in einem Bücherregal 10 Bücher aufstellen, die vier verschiedene Themengebiete behandeln: 4 sind aus der Informatik, 3 aus der Mathematik, 2 aus der Statistik und 1 aus der Geschichte. Wenn wir wollen, dass alle Bücher desselben Gebiets zusammen stehen, wie viele verschiedene Anordnungen sind dann möglich?

Lösung. Zunächst bestimmen wir die Anzahl der möglichen Permutationen der Themen: Da es 4 verschiedene Themen gibt, können wir eine von $4!$ verschiedenen Anordnungen wählen. Sobald wir die Reihenfolge der Themen gewählt haben, wählen wir die Reihenfolge der Bücher innerhalb des Themas. Die Informatikbücher können auf $4!$ verschiedene Arten angeordnet werden, die Mathematikbücher auf $3!$, die Statistikbücher auf $2!$ und das Geschichtsbuch auf $1!$ Insgesamt gibt es also $4!4!3!2!1! = 6.912$ Möglichkeiten, die Bücher so aufzustellen, dass Bücher zum selben Thema zusammen stehen. \triangle

Beispiel 7. Ein Kurs hat 5 männliche und 3 weibliche Studenten. Aufgrund einer Prüfung werden alle Studenten entsprechend ihrer Leistung eingestuft. Am Ende des Kurses findet eine Prüfung statt. Es ergibt sich, dass alle Studenten unterschiedliche Bewertungen erhalten. (a) Wie viele verschiedene Rangfolgen sind möglich? (b) Wenn alle Bewertungen gleich wahrscheinlich sind, wie groß ist die Wahrscheinlichkeit, dass die 3 Studentinnen die 3 höchsten Bewertungen erhalten?

Lösung. (a) Am Test nehmen 8 Personen teil, also gibt es $8! = 40320$ mögliche Rangfolgen. (b) Es gibt $3!$ mögliche Rangfolgen unter den weiblichen Studenten und $5!$ mögliche Rangfolgen unter den männlichen. Nach der Produktregel der Kombinatorik sind dies insgesamt $5!3!$ Rangfolgen, bei denen die höchstplatzierten Studenten weiblich sind. Die gesuchte Wahrscheinlichkeit dafür ist dann $\frac{5!3!}{8!} = \frac{3 \cdot 2 \cdot 1}{8 \cdot 7 \cdot 6} = 1/56$. \triangle

ANSTATT DIE ELEMENTE einer Menge von n Objekten zu anzuordnen, möchten wir manchmal wissen, wie viele Möglichkeiten es gibt, Gruppen

¹³ Das heißt, $n! = n(n-1)(n-2)\cdots 3 \cdot 2 \cdot 1$. Der Vollständigkeit halber definieren wir $0! := 1$.

von r Objekten zu bilden. Zum Beispiel: Wie viele Gruppen von 3 Objekten können mit den fünf Elementen A, B, C, D und E gebildet werden? Ein Weg, diese Frage zu beantworten, ist der folgende. Wir haben 5 Möglichkeiten, das erste Objekt auszuwählen, 4, um das zweite auszuwählen, und 3, um das dritte auszuwählen. [◦] Es gibt also $5 \cdot 4 \cdot 3 = 60$ Möglichkeiten, 3 Objekte auszuwählen, wenn die Reihenfolge, in der sie gezogen werden, von Bedeutung ist. Wenn wir uns jedoch nur für die Gruppen von Objekten interessieren, ohne die Reihenfolge zu beachten, dann wird bei dieser Methode jede Gruppe 6 Mal gezählt.¹⁴ Die Gesamtzahl der Gruppen, die gebildet werden können (ohne Berücksichtigung der Reihenfolge), beträgt also $\frac{5 \cdot 4 \cdot 3}{3 \cdot 2 \cdot 1} = 10$.

[◦] Ist Ihnen klar warum?

¹⁴ Wenn zum Beispiel die ausgewählte Gruppe A, B, C ist, werden der Fall, dass wir A, B, C ziehen, und der Fall, dass wir C, B, A ziehen, als verschieden gezählt.

Im Allgemeinen existieren $n(n-1) \cdots (n-r+1)$ verschiedene Arten, Gruppen von r Elementen aus einer Menge von n Objekten zu auswählen, wenn die Reihenfolge, in der sie gezogen werden, von Bedeutung ist. Da jede Gruppe bei dieser Methode $r!$ mal gezählt wird, gibt es, wenn die Reihenfolge nicht relevant ist,

$$\frac{n(n-1) \cdots (n-r+1)}{r!} = \frac{n!}{(n-r)!r!}$$

Gruppen von r Elementen, die aus einer Menge von n Objekten ausgewählt werden können.

Definition 8. Sei $r \leq n$. Die Anzahl der *Kombinationen* (engl. *combinations*) von r Objekten, die aus einer Menge von n Objekten gewählt werden können, ist

$$\binom{n}{r} := \frac{n!}{(n-r)!r!}.$$

Auch hier ist $\binom{n}{r}$ die Anzahl der Gruppen der Größe r , die aus einer Menge von n Elementen ausgewählt werden können, wenn ihre Reihenfolge irrelevant ist; zum Beispiel gibt es $\binom{7}{2} = \frac{7 \cdot 6}{2 \cdot 1} = 21$ verschiedene Paare, die man aus einer Menge von 7 Personen auswählen kann. Es sei daran erinnert, dass $0! = 1$ ist. Dann gilt $\binom{n}{0} = \binom{n}{n} = 1$.

Beispiel 9. Aus einer Gruppe von 5 Männern und 8 Frauen werden fünf Personen zufällig ausgewählt. Wie hoch ist die Wahrscheinlichkeit, dass 3 Frauen und 2 Männer ausgewählt werden?

Lösung. Unter einer *zufälligen Auswahl* verstehen wir, dass jede der $\binom{13}{5}$ möglichen Kombinationen gleich wahrscheinlich ist. Es gibt $\binom{5}{2}$ mögliche Kombinationen von 2 Männern und $\binom{8}{3}$ von drei Frauen. Aus der Produktregel der Kombinatorik ergibt sich die Wahrscheinlichkeit solch einer Auswahl als

$$\frac{\binom{5}{2} \binom{8}{3}}{\binom{13}{5}} = \frac{5 \cdot 4 \cdot 8 \cdot 7 \cdot 6 \cdot 5!}{13 \cdot 12 \cdot 11 \cdot 10 \cdot 9 \cdot 2 \cdot 3 \cdot 2} = \frac{8 \cdot 7 \cdot 5 \cdot 2}{13 \cdot 11 \cdot 9} = \frac{560}{1287}. \quad \triangle$$

Beispiel 10. Wenn wir aus einer Menge von n Objekten k Elemente zufällig auswählen, wie groß ist dann die Wahrscheinlichkeit, dass ein gegebenes Objekt unter den k ausgewählten ist?

Lösung. Es gibt eine Möglichkeit, das gegebene Objekt auszuwählen, und $\binom{n-1}{k-1}$ Möglichkeiten, $k-1$ Elemente aus den übrigen Objekten der Menge auszuwählen. Nach der Produktregel der Kombinatorik existieren $1 \cdot \binom{n-1}{k-1}$ verschiedene Teilmengen aus k der n Elemente, die das ausgewählte Element enthalten. Da die Gesamtzahl der möglichen Auswahlen $\binom{n}{k}$ beträgt, lautet die Wahrscheinlichkeit, dass ein gegebenes Objekt unter den k ausgewählten ist, ^o

$$\frac{\binom{n-1}{k-1}}{\binom{n}{k}} = \frac{(n-1)!(n-k)!k!}{(n-k)!(k-1)!n!} = \frac{k}{n}. \quad \Delta$$

^oWie groß ist die Wahrscheinlichkeit, dass in einer Gruppe von 10 Personen alle einen unterschiedlichen Geburtstag haben?

Bedingte Wahrscheinlichkeiten

DER BEGRIFF DER BEDINGTEN WAHRSCHEINLICHKEIT ist einer der grundlegenden Begriffe in der Wahrscheinlichkeitstheorie. Bedingte Wahrscheinlichkeiten geben uns die Möglichkeit, Unbestimmtheit zu messen, wenn zusätzliche Informationen (oder Evidenzen) verfügbar sind: Sie geben uns die Möglichkeit darzustellen, wie sich Wahrscheinlichkeiten ändern, wenn sich unsere Informationen über die Welt ändern. Sie sind auch ein technisches Mittel, mit dem man die Berechnung von Wahrscheinlichkeiten vereinfachen kann, indem man ein zusätzliches Ereignis betrachtet, und prüft, wie Wahrscheinlichkeiten davon beeinflusst werden, dass dieses Ereignis eintritt oder nicht.

Um bedingte Wahrscheinlichkeiten zu veranschaulichen, nehmen wir an, wir werfen zwei Würfel. Der Stichprobenraum dieses Experiments ist die Menge der 36 Ergebnisse $\mathcal{S} = \{(i, j) \mid 1 \leq i, j \leq 6\}$, wobei (i, j) für das Ergebnis steht, dass der erste Würfel i und der zweite j zeigt. Wenn jedes Ergebnis gleich wahrscheinlich ist, hat jedes von ihnen die Wahrscheinlichkeit $1/36$.¹⁵ Nehmen wir an, wir stellen fest, dass beim ersten Würfel die Zahl 3 oben ist. Wie hoch ist die Wahrscheinlichkeit dafür, dass die Summe der beiden Würfel genau 8 ergibt?

Da wir bereits wissen, dass der erste Würfel eine 3 zeigt, gibt es nur noch 6 mögliche Ergebnisse für unser Experiment: $(3, 1)$, $(3, 2)$, $(3, 3)$, $(3, 4)$, $(3, 5)$, und $(3, 6)$. Da ursprünglich jedes dieser Ergebnisse gleich wahrscheinlich war, gilt dies auch für diese eingeschränkte Situation. Das heißt, da der erste Würfel eine 3 zeigt, hat jedes dieser Ergebnisse nun eine Wahrscheinlichkeit von $\frac{1}{6}$.¹⁶ Die Wahrscheinlichkeit, dass die Summe 8 ergibt, ist dann die Wahrscheinlichkeit von $(3, 5)$, welche $\frac{1}{6}$ beträgt.

Bezeichnen wir mit \mathcal{E} und \mathcal{F} die Ereignisse, dass die Summe der Würfel 8 ist, bzw. das Ereignis, dass der erste Würfel eine 3 zeigt. Was wir soeben berechnet haben, nennen wir die *Wahrscheinlichkeit von \mathcal{E} unter der Bedingung \mathcal{F}* , die wir als $P(\mathcal{E}|\mathcal{F})$ schreiben. Wir lesen den Ausdruck $P(\mathcal{E}|\mathcal{F})$ auch als *die Wahrscheinlichkeit von \mathcal{E} , wenn \mathcal{F} eingetreten ist* oder *von \mathcal{E} wenn \mathcal{F}* oder *von \mathcal{E} unter der Bedingung \mathcal{F}* .

Um den allgemeinen Begriff der bedingten Wahrscheinlichkeit zu definieren, folgen wir der gleichen Intuition wie in diesem Beispiel. Unter der Voraussetzung, dass das Ereignis \mathcal{F} eintritt, müssen wir, um \mathcal{E} zu beobachten, in der Tat beide Ereignisse gleichzeitig beobachten, also $\mathcal{E}\mathcal{F}$.

¹⁵ Wir sagen dann, der Würfel ist *ideal* oder *fair*.

¹⁶ In der Tat ist die (bedingte) Wahrscheinlichkeit aller anderen 30 Ereignisse jetzt 0, da sie unter der gegebenen Bedingung nicht eintreten können.

Da wir jedoch wissen, dass \mathcal{F} eingetreten ist, können wir es als den neuen Stichprobenraum betrachten (alle Ergebnisse außerhalb dieses Raums sind nun unmöglich), und daher wird die Wahrscheinlichkeit, \mathcal{E} zu beobachten, zur Wahrscheinlichkeit von $\mathcal{E}\mathcal{F}$ relativ zu der von \mathcal{F} .

Definition 11. Seien \mathcal{E} und \mathcal{F} zwei Ereignisse im gleichen Stichprobenraum, so dass $P(\mathcal{F}) > 0$ ist. Die Wahrscheinlichkeit von \mathcal{E} unter der Bedingung \mathcal{F} ist

$$P(\mathcal{E} | \mathcal{F}) = \frac{P(\mathcal{E}\mathcal{F})}{P(\mathcal{F})}$$

Man beachte, dass die Definition nur dann Sinn ergibt, wenn $P(\mathcal{F}) > 0$ ist. Ausgehend von der Intuition, dass wir \mathcal{F} beobachtet haben, ist diese Annahme vernünftig, da wir nicht erwarten, dass ein Ereignis mit der Wahrscheinlichkeit 0 stattfinden kann.

Obwohl wir ursprünglich bedingte Wahrscheinlichkeiten motiviert haben als eine Möglichkeit zur Aktualisierung von Annahmen bei Vorliegen von Evidenz,¹⁷ ist Definition 11 auch mit der frequentistischen Sichtweise vereinbar. Angenommen, ein Experiment wird n -mal wiederholt. Da $P(\mathcal{F})$ den Anteil der Fälle ausdrückt, in denen \mathcal{F} beobachtet wird, wird dieses Ereignis ungefähr $nP(\mathcal{F})$ Mal eintreten. Ebenso werden $nP(\mathcal{E}\mathcal{F})$ mal die Ereignisse \mathcal{E} und \mathcal{F} zusammen auftreten. Von den ungefähr $nP(\mathcal{F})$ Experimenten, deren Ergebnis in \mathcal{F} liegt, gehören also $nP(\mathcal{E}\mathcal{F})$ auch zu \mathcal{E} . Mit anderen Worten: Unter den Ergebnissen in \mathcal{F} ist der Anteil derer, die auch in \mathcal{E} liegen, näherungsweise $\frac{nP(\mathcal{E}\mathcal{F})}{nP(\mathcal{F})} = \frac{P(\mathcal{E}\mathcal{F})}{P(\mathcal{F})}$. Diese Näherung wird umso genauer, je öfter wir das Experiment wiederholen.

¹⁷ Der subjektiven Sichtweise folgend.

Beispiel 12. Wir haben eine Schachtel mit Transistoren. Angenommen, 4 davon sind defekt (funktionieren nicht), 8 sind teilweise defekt (funktionieren nur kurz) und 20 sind intakt (funktionieren gut). Wir wählen einen von ihnen zufällig aus und benutzen ihn. Wenn er nicht sofort ausfällt, wie hoch ist dann die Wahrscheinlichkeit, dass es sich um einen intakten Transistor handelt?

Lösung. Da der Transistor nicht sofort ausgefallen ist, wissen wir, dass er nicht zu den 4 defekten Transistoren gehört. Wir rechnen also

$$P(\text{intakt} | \overline{\text{defekt}}) = \frac{P(\text{intakt}, \overline{\text{defekt}})}{P(\overline{\text{defekt}})} = \frac{P(\text{intakt})}{P(\overline{\text{defekt}})},$$

wobei die letzte Gleichung gilt, weil jeder intakte Transistor notwendigerweise nicht defekt ist. Da der Transistor zufällig ausgewählt wurde,¹⁸ gilt

$$P(\text{intakt} | \overline{\text{defekt}}) = \frac{\frac{20}{32}}{\frac{32}{32}} = \frac{5}{7}. \quad \triangle$$

¹⁸ Jeder Transistor wird mit gleicher Wahrscheinlichkeit ausgewählt.

Beispiel 13. Wir werfen zweimal eine faire Münze und wetten bei jedem Wurf um 5€, dass sie auf „Kopf“ fällt. Angenommen, wir gewinnen mindestens eine dieser Wetten, wie hoch ist die Wahrscheinlichkeit, dass wir 10€ gewonnen haben?

Lösung. Der Stichprobenraum für die beiden Münzwürfe kann dargestellt werden als $\mathcal{S} = \{(k, k), (k, z), (z, k), (z, z)\}$.²⁰ Wir gewinnen 10€, wenn bei-

¹⁹ Man beachte, dass dieselbe Wahrscheinlichkeit durch Abzählen ermittelt werden kann: Da der Transistor nicht defekt ist, reduziert sich das Problem auf die Berechnung der Wahrscheinlichkeit, dass ein Transistor, der zufällig aus einer Schachtel mit 20 intakten und 8 teilweise defekten Transistoren ausgewählt wird, intakt ist. Diese ist natürlich $\frac{20}{28}$.

²⁰ Hier bedeutet (k, z) , dass der erste Wurf Kopf war und der zweite Zahl.

de Würfe Kopf zeigen; wir nennen dieses Ereignis \mathcal{E} . Wenn \mathcal{F} für das Ereignis steht, dass mindestens ein Wurf Kopf ist, dann ist die gesuchte Wahrscheinlichkeit

$$P(\mathcal{E} | \mathcal{F}) = \frac{P(\mathcal{E}\mathcal{F})}{P(\mathcal{F})} = \frac{P(\{(k, k)\})}{P(\{(k, k), (k, z), (z, k)\})} = \frac{1/4}{3/4} = \frac{1}{3}. \quad \triangle$$

Die Gleichung für bedingte Wahrscheinlichkeiten aus Definition 11 kann umgeschrieben werden zu $P(\mathcal{E}\mathcal{F}) = P(\mathcal{E} | \mathcal{F})P(\mathcal{F})$.²¹ Diese Variante ist nützlich für die Berechnung von Wahrscheinlichkeiten von Durchschnitten von Ereignissen, wenn die bedingten Wahrscheinlichkeiten bekannt sind.

Beispiel 14. Ihre Lieblings-Fußballmannschaft hat eine 20-prozentige Wahrscheinlichkeit, das Endspiel zu erreichen, und wenn sie es schafft, dann hat sie eine 50-prozentige Wahrscheinlichkeit zu gewinnen. Wie hoch ist die Wahrscheinlichkeit dafür, dass die Mannschaft Meister wird?

Lösung. Sei \mathcal{F} das Ereignis, dass die Mannschaft das Endspiel erreicht, und \mathcal{E} das Ereignis, dass sie es gewinnt. Dann ist die Wahrscheinlichkeit dafür, dass sie Meister wird, $P(\mathcal{E}\mathcal{F}) = P(\mathcal{E} | \mathcal{F})P(\mathcal{F}) = 0.5 \cdot 0.2 = 0.1$. \triangle

Der Satz von Bayes

FÜR JE ZWEI EREIGNISSE \mathcal{E} und \mathcal{F} kann \mathcal{E} äquivalent als Vereinigung $\mathcal{E}\mathcal{F} \cup \mathcal{E}\overline{\mathcal{F}}$ dargestellt werden (siehe Abbildung 33). In der Tat muss jedes Ergebnis aus \mathcal{E} entweder in \mathcal{F} oder in $\overline{\mathcal{F}}$ liegen, und damit auch im Durchschnitt dieser Ereignisse mit \mathcal{E} . Da $\mathcal{E}\mathcal{F}$ und $\mathcal{E}\overline{\mathcal{F}}$ sich gegenseitig ausschließen, folgt aus Axiom 3, dass

$$P(\mathcal{E}) = P(\mathcal{E}\mathcal{F}) + P(\mathcal{E}\overline{\mathcal{F}}) = P(\mathcal{E} | \mathcal{F})P(\mathcal{F}) + P(\mathcal{E}\overline{\mathcal{F}})P(\overline{\mathcal{F}}). \quad (1)$$

Das bedeutet, dass die Wahrscheinlichkeit von \mathcal{E} eine gewichtete Summe der Wahrscheinlichkeiten für \mathcal{E} unter der Bedingung eines andere Ereignisses \mathcal{F} und seines Komplements $\overline{\mathcal{F}}$ ist, wobei die Gewichte die jeweiligen Wahrscheinlichkeiten des Eintretens der bedingenden Ereignisse sind. Dies ist praktisch, denn es ermöglicht, die Wahrscheinlichkeit von \mathcal{E} zu ermitteln, indem man ein zusätzliches Ereignis \mathcal{F} einführt und eine Fallunterscheidung danach vornimmt, ob \mathcal{E} zusammen mit \mathcal{F} oder ohne \mathcal{F} eintritt. Dieser Ansatz ermöglicht es uns, Wahrscheinlichkeiten in Fällen zu bestimmen, in denen ein direkter Weg nicht möglich ist.

Beispiel 15. Es gibt zwei Arten von Menschen: vorsichtige und risikofreudige. Im Laufe eines Jahres verunglücken vorsichtige Menschen mit einer Wahrscheinlichkeit von 0,2, während diese Wahrscheinlichkeit bei risikofreudigen Menschen 0,4 beträgt. Wenn 30% der Bevölkerung risikofreudig sind, wie hoch ist dann die Wahrscheinlichkeit, dass eine zufällig ausgewählte Person im Zeitraum von einem Jahr einen Unfall hat?

Lösung. Wir unterscheiden die Fälle, dass die ausgewählte Person risikofreudig ist oder nicht. Es sei \mathcal{E} das Ereignis, dass die Person in dem Jahreszeitraum einen Unfall hat, und \mathcal{F} das Ereignis, dass sie risikofreudig ist.²² Dann gilt

²¹ Damit man nicht verlangen muss, dass $P(\mathcal{F}) > 0$ ist, definiert man bedingte Wahrscheinlichkeiten oft mit dieser Variante.

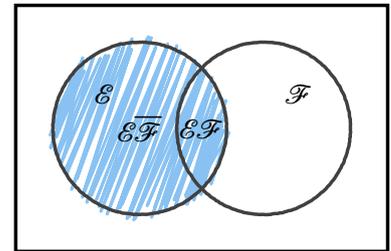


Abbildung 3: $\mathcal{E} = \mathcal{E}\mathcal{F} \cup \mathcal{E}\overline{\mathcal{F}}$.

²² $\overline{\mathcal{F}}$ bedeutet, dass sie vorsichtig ist.

$$P(\mathcal{E}) = P(\mathcal{E} | \mathcal{F})P(\mathcal{F}) + P(\mathcal{E} | \overline{\mathcal{F}})P(\overline{\mathcal{F}}) = 0.4 \cdot 0.3 + 0.2 \cdot 0.7 = 0.26 \quad \Delta$$

DIE FORMEL IN GLEICHUNG (1) ist nicht nur für die Berechnung der Wahrscheinlichkeiten von Ereignissen selbst nützlich, sondern auch für die Aktualisierung einer früheren Einschätzungen der Wahrscheinlichkeit eines Ereignisses bei Vorliegen zusätzlicher oder neuer Informationen. Wir zeigen anhand eines Beispiels, wie dies funktioniert.

Beispiel 16. Betrachten wir noch einmal Beispiel 15 und nehmen an, dass die zufällig ausgewählte Person einen Unfall hat. Wie hoch ist die Wahrscheinlichkeit, dass sie zur Gruppe der risikofreudigen Personen gehört?

Lösung. Denken wir daran, dass es für die Person ursprünglich (ohne weitere Anhaltspunkte) eine Wahrscheinlichkeit von 30% gab, risikofreudig zu sein ($P(\mathcal{F}) = 0.3$). Angesichts der neuen Information, dass sie einen Unfall hatte, können wir diese Wahrscheinlichkeit jedoch neu evaluieren wie folgt:

$$P(\mathcal{F} | \mathcal{E}) = \frac{P(\mathcal{E} | \mathcal{F})P(\mathcal{F})}{P(\mathcal{E})} = \frac{0.3 \cdot 0.4}{0.26} = \frac{6}{13} \approx 0.4615. \quad \Delta$$

DIESE FORMEL²³ können wir verallgemeinern, indem wir den Stichprobenraum feiner unterteilen. Sei $\mathcal{F}_1, \dots, \mathcal{F}_n$ eine *Partition* des Stichprobenraums \mathcal{S} .²⁴ Bezogen auf die Ereignisse bedeutet dies, dass bei jedem Experiment genau eines dieser \mathcal{F}_i eintritt, und wie zuvor gilt $\mathcal{E} \equiv \bigcup_{i=1}^n \mathcal{E} \mathcal{F}_i$. Da diese Ereignisse sich gegenseitig ausschließen, erhalten wir

$$P(\mathcal{E}) = \sum_{i=1}^n P(\mathcal{E} \mathcal{F}_i) = \sum_{i=1}^n P(\mathcal{E} | \mathcal{F}_i)P(\mathcal{F}_i). \quad (2)$$

Das heißt, wir können $P(\mathcal{E})$ berechnen, indem wir mittels einer Partition eine Fallunterscheidung über die möglichen Ergebnisse im Stichprobenraum vornehmen, und dann einen gewichteten Durchschnitt der Wahrscheinlichkeit der einzelnen Fälle formen, wobei das Gewicht jedes Mal der Wahrscheinlichkeit des Ereignisses \mathcal{F}_i entspricht. Beachten Sie, dass die ursprüngliche Formel nur ein Spezialfall davon ist: \mathcal{G} und $\overline{\mathcal{G}}$ bilden eine Partition.

Angenommen, wir wissen, dass \mathcal{E} eingetreten ist, und wir möchten wissen, welches der Ereignisse in der Partition beobachtet wurde. Mit Gleichung (2) erhalten wir

$$P(\mathcal{F}_j | \mathcal{E}) = \frac{P(\mathcal{E} \mathcal{F}_j)}{P(\mathcal{E})} = \frac{P(\mathcal{E} | \mathcal{F}_j)P(\mathcal{F}_j)}{\sum_{i=1}^n P(\mathcal{E} | \mathcal{F}_i)P(\mathcal{F}_i)}. \quad (3)$$

Diese Gleichung ist als *Bayes'sche Formel* bekannt.²⁵ Wenn wir uns die Ereignisse \mathcal{F}_j als mögliche Hypothesen über einen Forschungsgegenstand denken, kann die Bayessche Formel dahingehend interpretiert werden, dass sie aufzeigt, wie sich unsere Ansichten über diese Hypothesen (das heißt die ihnen zugeschriebenen Wahrscheinlichkeiten $P(\mathcal{F}_j)$) auf Grundlage der Evidenz eines Experiments ändern sollten.²⁶

Beispiel 17. Betrachten wir einen Fehler in einem Programm, der sich mit gleicher Wahrscheinlichkeit in jedem von drei verschiedenen Codeteile

²³ In Gleichung (1)

²⁴ $\mathcal{F}_1, \dots, \mathcal{F}_n$ bilden eine Partition des Raums \mathcal{S} , wenn sie sich gegenseitig ausschließende Ereignisse sind und $\bigcup_{i=1}^n \mathcal{F}_i = \mathcal{S}$.

²⁵ Nach Thomas Bayes.

²⁶ Dabei wird $P(\mathcal{F}_j)$ als *A-priori-Wahrscheinlichkeit* und $P(\mathcal{F}_j | \mathcal{E})$ als *A-posteriori-Wahrscheinlichkeit* des Ereignisses \mathcal{F}_j bezeichnet.

befinden kann. Für jeden der drei Teile i sei $1 - \alpha_i$ die Wahrscheinlichkeit, den Fehler bei der Suche in diesem Teil zu finden, vorausgesetzt, er befindet sich tatsächlich dort.²⁷ Wenn wir im ersten Teil suchen und keinen Fehler finden, wie hoch ist dann für jeden der drei Teile $i \in \{1, 2, 3\}$ die Wahrscheinlichkeit, dass sich der Fehler im Teil i befindet?

Lösung. Sei $\mathcal{F}_i, 1 \leq i \leq 3$ das Ereignis, dass sich der Fehler im i -ten Codestück befindet, und \mathcal{E} das Ereignis, dass die Suche im ersten Stück erfolglos war. Dann haben wir

$$P(\mathcal{F}_1 | \mathcal{E}) = \frac{P(\mathcal{E} | \mathcal{F}_1)P(\mathcal{F}_1)}{\sum_{i=1}^3 P(\mathcal{E} | \mathcal{F}_i)P(\mathcal{F}_i)} = \frac{\alpha_1 \cdot (1/3)}{\alpha_1 \cdot (1/3) + 1 \cdot (1/3) + 1 \cdot (1/3)} = \frac{\alpha_1}{\alpha_1 + 2},$$

und für $j, 2 \leq i \leq 3$ gilt

$$P(\mathcal{F}_j | \mathcal{E}) = \frac{P(\mathcal{E} | \mathcal{F}_j)P(\mathcal{F}_j)}{\sum_{i=1}^3 P(\mathcal{E} | \mathcal{F}_i)P(\mathcal{F}_i)} = \frac{1 \cdot (1/3)}{\alpha_1 \cdot (1/3) + 1 \cdot (1/3) + 1 \cdot (1/3)} = \frac{1}{\alpha_1 + 2}.$$

Ist zum Beispiel $\alpha_1 = 0.4$, dann ist die bedingte Wahrscheinlichkeit, dass sich der Fehler im ersten Stück befindet, nachdem wir ihn dort nicht gefunden haben, $P(\mathcal{F}_1 | \mathcal{E}) = \frac{1}{6}$.[◦] △

²⁷ Die Zahl α_i ist die *Übersehenswahrscheinlichkeit* (engl. *overlook probability*): Wie groß ist die Wahrscheinlichkeit, den Fehler zu übersehen, wenn man an der richtigen Stelle sucht? Manche Fehler sind schwer zu finden.

[◦]What about the other pieces?

Unabhängige Ereignisse

WIE IN DEN VORIGEN BEISPIELEN gesehen, ist die bedingte Wahrscheinlichkeit $P(\mathcal{E} | \mathcal{F})$ eines Ereignisses \mathcal{E} , gegeben \mathcal{F} , im Allgemeinen nicht gleich der (unbedingten) Wahrscheinlichkeit $P(\mathcal{E})$. Das heißt, Wissen über \mathcal{F} ändert im Allgemeinen die Wahrscheinlichkeit, \mathcal{E} zu beobachten. Im speziellen Fall, dass $P(\mathcal{E} | \mathcal{F}) = P(\mathcal{E})$ ist, sagen wir, dass \mathcal{E} und \mathcal{F} *unabhängig* sind. Mit anderen Worten: \mathcal{E} und \mathcal{F} sind unabhängig, wenn Wissen über das Eintreten von \mathcal{F} die Wahrscheinlichkeit von \mathcal{E} nicht verändert.

Wir erinnern uns, es gilt $P(\mathcal{E}\mathcal{F}) = P(\mathcal{E} | \mathcal{F})P(\mathcal{F})$. Wenn also \mathcal{E} und \mathcal{F} unabhängig sind, dann ist $P(\mathcal{E}\mathcal{F}) = P(\mathcal{E})P(\mathcal{F})$. Die umgekehrte Implikation ist ebenfalls wahr.

Definition 18. Zwei Ereignisse \mathcal{E}, \mathcal{F} sind *unabhängig* wenn

$$P(\mathcal{E}\mathcal{F}) = P(\mathcal{E})P(\mathcal{F}).$$

Sind \mathcal{E} and \mathcal{F} nicht unabhängig, dann sind sie *abhängig*.

Beispiel 19. Eine Karte wird zufällig aus einem Standardspiel von 52 Karten gezogen.²⁸ Sei \mathcal{E} das Ereignis, dass ein As gezogen, und \mathcal{F} , dass eine rote Karte gezogen wurde. Dann sind \mathcal{E} und \mathcal{F} unabhängig, denn $P(\mathcal{E}\mathcal{F}) = \frac{1}{26}$, $P(\mathcal{E}) = \frac{4}{52}$, and $P(\mathcal{F}) = \frac{26}{52}$. △

²⁸ Französische Spielkarten.

Satz 20. Sind \mathcal{E} und \mathcal{F} unabhängig, dann auch \mathcal{E} und $\overline{\mathcal{F}}$.²⁹

Beweis. Angenommen \mathcal{E} und \mathcal{F} sind unabhängig. Dann ist

$$P(\mathcal{E}) = P(\mathcal{E}\mathcal{F}) + P(\mathcal{E}\overline{\mathcal{F}}) = P(\mathcal{E})P(\mathcal{F}) + P(\mathcal{E}\overline{\mathcal{F}}),$$

wobei die letzte Gleichheit wegen der Unabhängigkeit von \mathcal{E} und \mathcal{F} gilt. Weiter folgt

$$P(\mathcal{E}\overline{\mathcal{F}}) = P(\mathcal{E})(1 - P(\mathcal{F})) = P(\mathcal{E})P(\overline{\mathcal{F}}). \quad \square$$

²⁹ Das heißt, wenn \mathcal{E} und \mathcal{F} unabhängig sind, wird die Wahrscheinlichkeit von \mathcal{E} nicht verändert durch Information über \mathcal{F} , weder über das Eintreten noch über das Ausbleiben.

Wir haben jetzt gesehen, dass Unabhängigkeit eine symmetrische Beziehung ist und dass sie abgeschlossen ist unter Komplementbildung. Überraschenderweise ist sie jedoch *nicht* abgeschlossen unter Durchschnitten. Das heißt, wenn \mathcal{E} unabhängig von \mathcal{F} und von \mathcal{G} ist, kann es dennoch sein, dass \mathcal{E} und $\mathcal{F}\mathcal{G}$ abhängig sind.

Beispiel 21. Wir werfen zwei faire Würfel. Sei \mathcal{E} das Ereignis, dass die Summe der beiden Würfel 7 ist, \mathcal{F} das Ereignis, dass beim ersten Würfel 1 oben liegt, und \mathcal{G} das Ereignis, dass beim zweiten Würfel 6 oben liegt. Man kann zeigen, dass \mathcal{E} unabhängig von \mathcal{F} und \mathcal{G} ist.^o Offensichtlich ist \mathcal{E} jedoch nicht unabhängig von $\mathcal{F}\mathcal{G}$; in der Tat ist $P(\mathcal{E} | \mathcal{F}\mathcal{G}) = 1$.

^oSie werden dies in der Übung zeigen.

Dieses Beispiel zeigt, dass es komplexer ist, die Unabhängigkeit von mehr als zwei Ereignissen zu definieren, als nur zu verlangen, dass alle Ereignisse paarweise unabhängig sind. Wir kommen damit zu folgender Definition:

Definition 22. Drei Ereignisse \mathcal{E} , \mathcal{F} , und \mathcal{G} sind *unabhängig*, falls³⁰

$$\begin{aligned} P(\mathcal{E}\mathcal{F}\mathcal{G}) &= P(\mathcal{E})P(\mathcal{F})P(\mathcal{G}), \\ P(\mathcal{E}\mathcal{F}) &= P(\mathcal{E})P(\mathcal{F}), \\ P(\mathcal{E}\mathcal{G}) &= P(\mathcal{E})P(\mathcal{G}), & \text{und} \\ P(\mathcal{F}\mathcal{G}) &= P(\mathcal{F})P(\mathcal{G}). \end{aligned}$$

³⁰ Informell kann dies so verstanden werden, dass für jede Teilmenge von Ereignissen alle Elemente voneinander unabhängig sind.

Wenn \mathcal{E} , \mathcal{F} und \mathcal{G} unabhängig sind, dann ist \mathcal{E} zwangsläufig unabhängig von jedem Ereignis, das durch Kombination von \mathcal{F} und \mathcal{G} gebildet wird. Zum Beispiel ist \mathcal{E} unabhängig von $\mathcal{F} \cup \mathcal{G}$ weil

$$\begin{aligned} P(\mathcal{E}(\mathcal{F} \cup \mathcal{G})) &= P(\mathcal{E}\mathcal{F} \cup \mathcal{E}\mathcal{G}) = P(\mathcal{E}\mathcal{F}) + P(\mathcal{E}\mathcal{G}) - P(\mathcal{E}\mathcal{F}\mathcal{G}) \\ &= P(\mathcal{E})P(\mathcal{F}) + P(\mathcal{E})P(\mathcal{G}) - P(\mathcal{E})P(\mathcal{F})P(\mathcal{G}) \\ &= P(\mathcal{E})(P(\mathcal{F}) + P(\mathcal{G}) - P(\mathcal{F}\mathcal{G})) = P(\mathcal{E})P(\mathcal{F} \cup \mathcal{G}). \end{aligned}$$

Definition 22 lässt sich auf offensichtliche Art für mehr als drei Ereignisse verallgemeinern. Die Ereignisse $\mathcal{E}_1, \dots, \mathcal{E}_n$ sind *unabhängig*, wenn für jede Teilmenge $\mathcal{F}_1, \dots, \mathcal{F}_m$, $m \leq n$, dieser Ereignisse gilt, dass $P(\mathcal{F}_1 \dots \mathcal{F}_m) = P(\mathcal{F}_1) \dots P(\mathcal{F}_m)$.

Viele voneinander unabhängige Ereignisse treten auf, wenn wir eine Reihe von Wiederholungen eines Experiments untersuchen. Wenn ein Experiment beispielsweise darin besteht, einen Würfel mehrmals zu werfen, können wir jeden Wurf als ein Teilexperiment betrachten, und es ist sinnvoll anzunehmen, dass die Ergebnisse früherer (oder zukünftiger) Würfe keinen Einfluss auf das Ergebnis des aktuellen Wurfs haben.

Beispiel 23. Betrachten Sie die parallele Schaltung in Abbildung 4, in der Strom fließt, wenn mindestens einer der n Schalter geschlossen ist. Jeder Schalter i ist, unabhängig von allen anderen, geschlossen mit der Wahrscheinlichkeit p_i , $1 \leq i \leq n$. Wie groß ist die Wahrscheinlichkeit, dass durch die Schaltung Strom fließt?

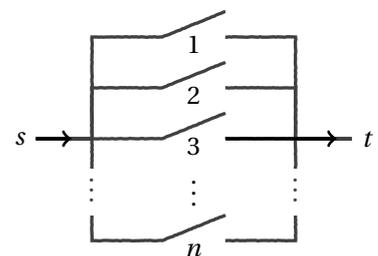


Abbildung 4: Eine parallele Schaltung, durch die Strom von s nach t fließen kann.

Lösung. Sei \mathcal{E} das Ereignis, dass Strom fließt, und \mathcal{F}_i , $1 \leq i \leq n$, das Ereignis, dass die Komponente i geschlossen ist. Dann ist $P(\mathcal{E})$ exakt die Wahrscheinlichkeit, dass alle Komponenten geöffnet sind, das heißt,

$$P(\mathcal{E}) = 1 - P(\overline{\mathcal{E}}) = 1 - P(\overline{\mathcal{F}_1} \dots \overline{\mathcal{F}_n}) = 1 - \prod_{i=1}^n P(\overline{\mathcal{F}_i}),$$

wobei sich die letzte Gleichheit aus der Unabhängigkeit der Ereignisse ergibt.³¹

³¹ Falls nötig, kann der letzte Ausdruck so umgeschrieben werden, dass die Wahrscheinlichkeiten der Ereignisse verwendet werden, und nicht die ihrer Komplemente.

Zufallsvariablen

BEI EINEM EXPERIMENT sind wir in der Regel nicht an den Ergebnissen selbst interessiert, sondern nur an bestimmten numerischen Größen, die durch die Ergebnisse determiniert sind. Wenn wir zum Beispiel Monopoly spielen und zwei Würfel fallen lassen, interessieren wir uns zwar für deren Summe, aber nicht für die Werte der einzelnen Würfel, die zu dieser Summe geführt haben.³² Diese numerischen Größen bezeichnen wir als *Zufallsvariablen*.³³ Da die Werte solcher Zufallsvariablen Konsequenzen der Ergebnisse eines Experiments sind, haben sie, wie die Ergebnisse, einen Wahrscheinlichkeitsgrad.

Betrachten wir zum Beispiel die Zufallsvariable \mathcal{X} , gebildet als die Summe zweier fairer Würfel. Die Variable \mathcal{X} nimmt ausschließlich Werte zwischen 2 und 12 an, und die Wahrscheinlichkeiten für jeden dieser Werte sind wie folgt:

$$P(\{\mathcal{X} = 2\}) = P(\{(1, 1)\}) = \frac{1}{36}$$

$$P(\{\mathcal{X} = 3\}) = P(\{(1, 2), (2, 1)\}) = \frac{2}{36}$$

$$P(\{\mathcal{X} = 4\}) = P(\{(1, 3), (2, 2), (3, 1)\}) = \frac{3}{36}$$

$$P(\{\mathcal{X} = 5\}) = P(\{(1, 4), (2, 3), (3, 2), (4, 1)\}) = \frac{4}{36}$$

$$P(\{\mathcal{X} = 6\}) = P(\{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\}) = \frac{5}{36}$$

$$P(\{\mathcal{X} = 7\}) = P(\{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}) = \frac{6}{36}$$

$$P(\{\mathcal{X} = 8\}) = P(\{(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)\}) = \frac{5}{36}$$

$$P(\{\mathcal{X} = 9\}) = P(\{(3, 6), (4, 5), (5, 4), (6, 3)\}) = \frac{4}{36}$$

$$P(\{\mathcal{X} = 10\}) = P(\{(4, 6), (5, 5), (6, 4)\}) = \frac{3}{36}$$

$$P(\{\mathcal{X} = 11\}) = P(\{(5, 6), (6, 5)\}) = \frac{2}{36}$$

$$P(\{\mathcal{X} = 12\}) = P(\{(6, 6)\}) = \frac{1}{36}.$$

Da \mathcal{X} immer einen dieser Werte annimmt, ist die Summe aller dieser Wahrscheinlichkeiten 1. Dies kann man leicht für diese Verteilung überprüfen.

Eine andere in diesem Zusammenhang interessante Variable \mathcal{Y} ist der Wert des ersten Würfels. In diesem Fall ist es gleich wahrscheinlich, dass \mathcal{Y} einen der Werte von 1 bis 6 annimmt.

³² Wenn die Summe 7 ist, kümmert uns nicht, welche der 6 Möglichkeiten, diese Summe zu erhalten, wir tatsächlich beobachtet haben.

³³ Dies ist eine schlechte Namenswahl, denn sie sind weder *zufällig* noch *Variablen*.

Die Beispiele bis jetzt waren Zufallsvariablen, die endlich viele Werte annehmen. Zufallsvariablen, die endlich oder abzählbar unendlich viele Werte annehmen werden *diskret* (engl. *discrete*) genannt.

Es gibt auch Zufallsvariablen, die ein Kontinuum von von möglichen Werten annehmen können. Diese nennt man *stetig* (engl. *continuous*). Als Gewicht einer Person zum Beispiel ist jeder Wert aus einem bestimmten Intervall (a, b) möglich.³⁴

Definition 24. Die *kumulative Verteilungsfunktion*, oft auch einfach *Verteilungsfunktion*, (engl. *cumulative distribution function*, kurz *cdf*) einer Zufallsvariablen \mathcal{X} ist die Funktion F , die für jede reelle Zahl x definiert ist durch:³⁵

$$F(x) := P[\mathcal{X} \leq x].$$

Die Schreibweise $\mathcal{X} \sim F$ drückt aus, dass F die Verteilungsfunktion von \mathcal{X} ist.

Mit Hilfe der Verteilungsfunktion F können wir alle Fragen zur Wahrscheinlichkeit von \mathcal{X} beantworten.

Um beispielsweise $P[a < \mathcal{X} \leq b]$ zu berechnen, nutzen wir die Tatsache, dass $[\mathcal{X} \leq b]$ in die beiden sich gegenseitig ausschließenden Ereignisse $[\mathcal{X} \leq a]$ und $[a < \mathcal{X} \leq b]$ zerlegt werden kann. Folglich ist $P[\mathcal{X} \leq b] = P[\mathcal{X} \leq a] + P[a < \mathcal{X} \leq b]$. Daraus können wir ableiten, dass

$$P[a < \mathcal{X} \leq b] = F(b) - F(a).$$

Beispiel 25. Wir betrachten die stetige Zufallsvariable \mathcal{X} mit der Verteilungsfunktion

$$F(x) = \begin{cases} 0 & x \leq 0 \\ 1 - \exp(-x^2) & x > 0. \end{cases}$$

Die Wahrscheinlichkeit dafür, dass \mathcal{X} größer als 1 ist, beträgt

$$P[\mathcal{X} > 1] = 1 - P[\mathcal{X} \leq 1] = 1 - F(1) = e^{-1} = 0.368. \quad \triangle$$

Arten von Zufallsvariablen

WENN EINE ZUFALLSVARIABLE (ZV) \mathcal{X} diskret ist,^o dann können wir ihre *Wahrscheinlichkeitsfunktion* (engl. *probability mass function*) durch $p(x) := P[\mathcal{X} = x]$ definieren. Da die Zufallsvariable diskret ist und \mathcal{X} für jedes Element des Stichprobenraums einen Wert annehmen muss, wissen wir, dass $p(x)$ für höchstens abzählbar viele Elemente positiv ist³⁶ und dass $\sum_{i=1}^{\infty} p(x_i) = 1$.

Beispiel 26. Betrachten wir eine ZV \mathcal{X} , die Werte aus $\{1, 2, 3\}$ annehmen kann. Wenn wir wissen, dass $p(2) = \frac{1}{3}$ und $p(3) = \frac{1}{6}$ ist, dann können wir ableiten, dass gilt

$$p(1) = 1 - p(2) - p(3) = 1 - \frac{1}{3} - \frac{1}{6} = \frac{1}{2}.$$

Diese Funktion ist in Abbildung 5 dargestellt. △

³⁴ Es gibt auch *hybride* Zufallsvariablen. In dieser Vorlesung, jedoch, werden wir sie nicht weiter untersuchen.

³⁵ In diesem Kurs befassen wir uns nur mit Zufallsvariablen, die als Werte reelle Zahlen haben.

^oDas heißt, dass sie nur die Werte einer Folge x_1, x_2, \dots annehmen kann.

³⁶ Höchstens die, welche von \mathcal{X} angenommen werden.

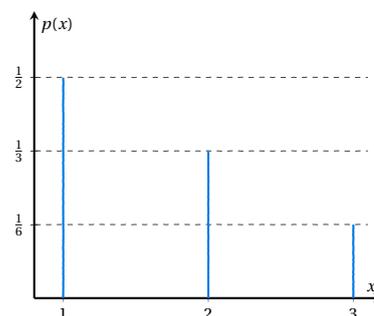


Abbildung 5: Wahrscheinlichkeitsfunktion aus Beispiel 26.

Die Verteilungsfunktion F kann in diesen Fällen aus p durch Addition aller relevanten Werte berechnet werden: $F(x) = \sum_{y \leq x} p(y)$.

Wenn \mathcal{X} eine diskrete ZV ist, deren mögliche Werte $x_1 < x_2 < \dots$ sind, ist die Verteilungsfunktion F eine *Treppenfunktion* (engl. *step function*): Sie bleibt auf dem Intervall $[x_{i-1}, x_i)$ konstant und macht dann bei x_i einen Sprung der Größe $p(x_i)$. Wenn \mathcal{X} zum Beispiel die in Beispiel 26 angegebene Wahrscheinlichkeitsfunktion hat, dann ist die kumulative Verteilungsfunktion von \mathcal{X}

$$F(x) = \begin{cases} 0 & x < 1 \\ \frac{1}{2} & 1 \leq x < 2 \\ \frac{5}{6} & 2 \leq x < 3 \\ 1 & 3 \leq x, \end{cases}$$

wie in Abbildung 6 dargestellt.

OFT HABEN WIR MIT Zufallsvariablen zu tun, die alle möglichen Werte aus einem (reellen) Intervall annehmen können.

Definition 27. Die ZV \mathcal{X} ist *stetig* (engl. *continuous*), wenn es eine nicht-negative Funktion f gibt, die über allen reellen Zahlen $x \in \mathbb{R}$ definiert ist, so dass für jede Menge $B \subseteq \mathbb{R}$ gilt, dass

$$P[\mathcal{X} \in B] = \int_B f(x) dx.$$

Diese Funktion f wird die *Wahrscheinlichkeitsdichtefunktion* oder kurz *Dichtefunktion*³⁷ von \mathcal{X} genannt (engl. *probability density function*) und mit *WDF* (engl. *pdf*) abgekürzt.

Was die Wahrscheinlichkeitsfunktion für diskrete Variablen, ist die Dichtefunktion für kontinuierliche Variablen. Sei \mathcal{X} stetig und sei B eine Menge von Werten. Wenn wir die Wahrscheinlichkeit des Ereignisses, dass der Wert von \mathcal{X} in B liegt, bestimmen wollen, ermitteln wir das Integral der Dichtefunktion über B . Da \mathcal{X} für jedes Element des Stichprobenraums einen Wert in $(-\infty, \infty)$ annimmt, muss f die folgende Bedingung erfüllen:

$$1 = P[\mathcal{X} \in (-\infty, \infty)] = \int_{-\infty}^{\infty} f(x) dx.$$

Alle Tatsachen über die Wahrscheinlichkeit von \mathcal{X} können durch Integration von f hergeleitet werden; zum Beispiel gilt für die Wertemenge $B = [a, b]$, dass

$$P[a \leq \mathcal{X} \leq b] = \int_a^b f(x) dx.$$

Als Spezialfall erhalten wir, mit $a = b$, dass $P[\mathcal{X} = a] = \int_a^a f(x) dx = 0$.³⁸ Als Beispiel zeigt Abbildung 7 die Wahrscheinlichkeitsdichte

$$f(x) = \begin{cases} e^{-x} & x \geq 0 \\ 0 & x < 0, \end{cases}$$

und die Fläche, von der $P[a < \mathcal{X} < b]$ repräsentiert wird.

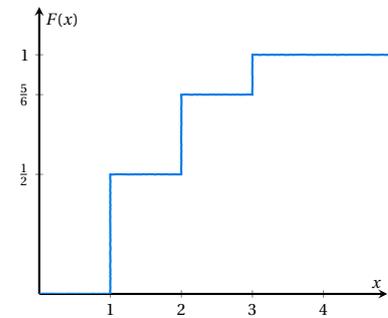


Abbildung 6: Verteilungsfunktion der ZV aus Beispiel 26

³⁷ Andere Bezeichnungen sind *Wahrscheinlichkeitsdichte*, *Verteilungsdichte* oder nur *Dichte*.

³⁸ Ist \mathcal{X} stetig, dann ist die Wahrscheinlichkeit, einen bestimmten Wert zu beobachten, immer 0.

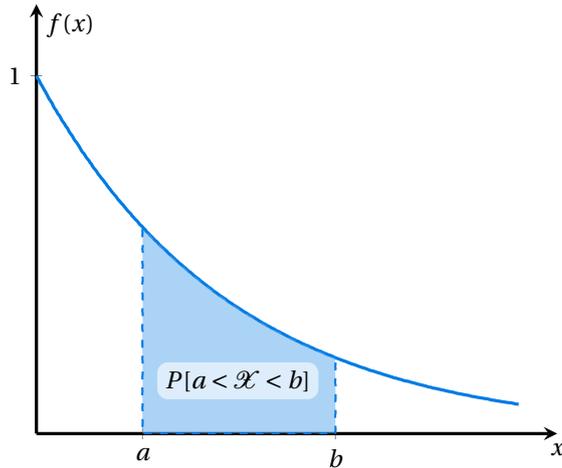


Abbildung 7: Eine Wahrscheinlichkeitsdichte und die Fläche, die $P[a < \mathcal{X} < b]$ repräsentiert.

Die Wahrscheinlichkeitsdichte f und die Verteilungsfunktion F stehen zueinander in Beziehung durch die Gleichung

$$F(a) = P[\mathcal{X} \leq a] = \int_{-\infty}^a f(x) dx,$$

oder, äquivalent dazu, $\frac{d}{dx}F(x) = f(x)$. Mit anderen Worten: Die Dichte ist die Ableitung der Verteilungsfunktion.

Ist die Dichte stetig, so ergibt sich aus Definition 5, durch die Wahl eines kleinen ε , dass

$$P[a - \frac{\varepsilon}{2} \leq \mathcal{X} \leq a + \frac{\varepsilon}{2}] = \int_{a - \frac{\varepsilon}{2}}^{a + \frac{\varepsilon}{2}} f(x) dx \approx \varepsilon f(a).$$

Mit anderen Worten, f sagt uns, wie wahrscheinlich es ist, einen Wert zu sehen, der nahe bei a ist.

Beispiel 28. Wir betrachten die stetige Zufallsvariable \mathcal{X} mit der Dichtefunktion

$$f(x) = \begin{cases} c(4x - 2x^2) & 0 < x < 2 \\ 0 & \text{sonst,} \end{cases}$$

für eine Konstante c .

1. Welchen Wert hat c ?
2. Was ist $P[\mathcal{X} > 1]$.

Lösung. Da f eine Dichte ist, gilt

$$1 = \int_{-\infty}^{\infty} f(x) dx = \int_0^2 c(4x - 2x^2) dx = c \left(2x^2 - \frac{2x^3}{3} \right) \Big|_0^2,$$

und somit ist $c = \frac{3}{8}$.

Hieraus folgt

$$P[\mathcal{X} > 1] = \int_1^{\infty} f(x) dx = \frac{3}{8} \int_1^2 (4x - 2x^2) dx = \frac{1}{2}. \quad \triangle$$

Gemeinsame Verteilungen

BEI EXPERIMENTEN interessieren wir uns oft für die Beziehung zwischen zwei oder mehr Zufallsvariablen und nicht nur für die Werte einer einzigen Variablen. In der Gesundheitspolitik beschäftigt uns zum Beispiel den Zusammenhang zwischen der Zeit, die wir im Sitzen verbringen, und dem Auftreten von Rückenschmerzen, oder in der Wirtschaft wollen wir den Zusammenhang zwischen Arbeitszeit und Produktivität der Beschäftigten verstehen.

Um zwei Zufallsvariablen \mathcal{X} and \mathcal{Y} gleichzeitig zu untersuchen, führen wir die *gemeinsame Verteilungsfunktion* F ein (engl. *joint cumulative probability distribution function*), die die Wahrscheinlichkeit dafür angibt, dass \mathcal{X} and \mathcal{Y} unter einem bestimmten Wert liegen; das heißt,

$$F(x, y) = P[\mathcal{X} \leq x, \mathcal{Y} \leq y].$$

Wenn wir die gemeinsame Verteilung kennen, sind wir in der Lage, die Wahrscheinlichkeiten für verschiedene Aussagen über die Variablen \mathcal{X} and \mathcal{Y} zu ermitteln. So erhalten wir zum Beispiel die Verteilungsfunktion $F_{\mathcal{X}}$ von \mathcal{X} , wenn wir für die Werte von \mathcal{Y} keine Schranken vorgeben, das heißt

$$F_{\mathcal{X}}(x) = P[\mathcal{X} \leq x] = P[\mathcal{X} \leq x, \mathcal{Y} < \infty] = F(x, \infty).$$

Analog dazu ist die Verteilungsfunktion von \mathcal{Y} gleich $F_{\mathcal{Y}}(y) = F(\infty, y)$.³⁹

Für diskrete Zufallsvariablen \mathcal{X} und \mathcal{Y} , die die Werte x_1, x_2, \dots bzw. y_1, y_2, \dots annehmen, ist die *gemeinsame Wahrscheinlichkeitsfunktion* p von \mathcal{X} und \mathcal{Y} auf offensichtliche Weise definiert:

$$p(x, y) := P[\mathcal{X} = x, \mathcal{Y} = y].$$

Die Wahrscheinlichkeitsfunktionen der einzelnen Zufallsvariablen lassen sich leicht ermitteln, indem man die Variable, die nicht von Interesse ist, eliminiert. Da z.B. \mathcal{Y} als Wert stets eins der y_j annimmt, ist das Ereignis $[\mathcal{X} = x]$ äquivalent zu der Vereinigung der paarweise disjunkten Ereignisse $[\mathcal{X} = x, \mathcal{Y} = y_j]$. Mit Axiom 3 der Wahrscheinlichkeitstheorie erhalten wir

$$P[\mathcal{X} = x] = P\left(\bigcup_j [\mathcal{X} = x, \mathcal{Y} = y_j]\right) = \sum_j P[\mathcal{X} = x, \mathcal{Y} = y_j] = \sum_j p(x, y_j).$$

Analog gilt $P[\mathcal{Y} = y] = \sum_i p(x_i, y)$.

Dies bedeutet, dass die gemeinsame Wahrscheinlichkeitsfunktion die individuellen Wahrscheinlichkeitsfunktionen für jede der beteiligten Zufallsvariablen vollständig bestimmt. Der Umkehrschluss gilt jedoch nicht: Es reicht nicht aus, $P[\mathcal{X} = x]$ und $P[\mathcal{Y} = y]$ zu kennen, um $P[\mathcal{X} = x, \mathcal{Y} = y]$ zu bestimmen.

Beispiel 29. Wir haben eine Schachtel mit Batterien, von denen 2 neu, 3 teilweise geladen und 4 völlig leer sind. Wir wählen zufällig 3 Batterien aus dieser Schachtel aus. Mit \mathcal{X} und \mathcal{Y} bezeichnen wir die Anzahl der

³⁹ Formal sind dies Grenzwerte (d.h. $F_{\mathcal{X}}(x) = \lim_{y \rightarrow \infty} F(x, y)$), aber die Intuition im Text passt besser für unseren Zweck.

Axiome der Wahrscheinlichkeitstheorie:

1. $0 \leq P(\mathcal{E}) \leq 1$
2. $P(\mathcal{S}) = 1$
3. $P(\bigcup_{i=1}^n \mathcal{E}_i) = \sum_{i=1}^n P(\mathcal{E}_i)$ falls die \mathcal{E}_i paarweise disjunkt sind.

neuen bzw. die Anzahl der teilweise geladenen Batterien. Die gemeinsame Wahrscheinlichkeitsfunktion ist:[◦]

$$\begin{aligned}
 p(0,0) &= \frac{\binom{4}{3}}{\binom{9}{3}} = \frac{4}{84} & p(0,1) &= \frac{\binom{3}{1}\binom{4}{2}}{\binom{9}{3}} = \frac{18}{84} \\
 p(0,2) &= \frac{\binom{3}{2}\binom{4}{1}}{\binom{9}{3}} = \frac{12}{84} & p(0,3) &= \frac{\binom{3}{3}}{\binom{9}{3}} = \frac{1}{84} \\
 p(1,0) &= \frac{\binom{2}{1}\binom{4}{2}}{\binom{9}{3}} = \frac{12}{84} & p(1,1) &= \frac{\binom{2}{1}\binom{3}{1}\binom{4}{1}}{\binom{9}{3}} = \frac{24}{84} \\
 p(1,2) &= \frac{\binom{2}{1}\binom{3}{2}}{\binom{9}{3}} = \frac{6}{84} & p(2,0) &= \frac{\binom{2}{2}\binom{4}{1}}{\binom{9}{3}} = \frac{4}{84} \\
 p(2,1) &= \frac{\binom{2}{2}\binom{3}{1}}{\binom{9}{3}} = \frac{3}{84}.
 \end{aligned}$$

[◦]Ist es klar, warum?

Tabelle 1 fasst diese Wahrscheinlichkeiten in einfacher Form zusammen. Δ

Man beachte, dass die Wahrscheinlichkeitsfunktion von \mathcal{X} durch Summieren der Elemente in einer Zeile ermittelt wird, während die Wahrscheinlichkeitsfunktion von \mathcal{Y} durch Summieren der Spalten entsteht. Diese Wahrscheinlichkeiten werden häufig als *marginale Wahrscheinlichkeitsfunktionen* (engl. *marginal probability mass functions*) von \mathcal{X} bzw. \mathcal{Y} bezeichnet.⁴⁰ Um die Korrektheit einer solchen Wahrscheinlichkeitstabelle zu überprüfen, sollte man kontrollieren, ob die Summe der Randzeile und der Randspalte jeweils 1 ist.[◦]

Tabelle 1: Die gemeinsame Wahrscheinlichkeitsfunktion $p(x, y)$ aus Beispiel 29. Der Nenner 84 in den Zahlen ist nicht enthalten.

$i \backslash j$	0	1	2	3	Sum
0	4	18	12	1	35
1	12	24	6	0	42
2	4	3	0	0	7
Sum	20	45	18	1	

Beispiel 30. In einer Gemeinde haben 15% der Familien keine Kinder, 20% haben 1 Kind, 35% haben 2 Kinder und 30% haben 3 Kinder. Nehmen wir an, dass jedes Kind mit gleicher Wahrscheinlichkeit ein Junge oder ein Mädchen ist. Für eine zufällig ausgewählte Familie sei \mathcal{X} die Zufallsvariable, die die Anzahl der Jungen angibt, und \mathcal{Y} die Anzahl der Mädchen. Tabelle 2 zeigt die gemeinsame Wahrscheinlichkeitsfunktion.

⁴⁰ Um sich diesen Namen zu merken, kann man daran denken, dass diese Funktionen am *Rand* (= englisch *margin*) der gemeinsamen Wahrscheinlichkeitstabelle aufscheinen.

[◦]Verstehen Sie, warum? Im Fall von Tabelle 1 muss die Summe 84 sein, weil der Nenner 84 nicht ausgeschrieben wurde.

$i \backslash j$	0	1	2	3	Summe
0	0,1500	0,1000	0,0875	0,0375	0,3750
1	0,1000	0,1750	0,1125	0	0,3875
2	0,0875	0,1125	0	0	0,2000
3	0,0375	0	0	0	0,0375
Summe	0,3750	0,3875	0,2000	0,0375	

Tabelle 2: Die gemeinsame Wahrscheinlichkeitsfunktion $p(x, y)$ aus Beispiel 30.

Im Folgenden rechnen wir die Werte in der ersten Zeile aus:[◦]

$$\begin{aligned}
 P[\mathcal{X} = 0, \mathcal{Y} = 0] &= P[\text{keine Kinder}] \\
 &= 0,15 \\
 P[\mathcal{X} = 0, \mathcal{Y} = 1] &= P[1 \text{ Kind, das ein Mädchen ist}] \\
 &= P[1 \text{ Kind}] \cdot P[1 \text{ Mädchen} \mid 1 \text{ Kind}] \\
 &= 0,2 \cdot 0,5 = 0,1 \\
 P[\mathcal{X} = 0, \mathcal{Y} = 2] &= P[2 \text{ Kinder, die Mädchen sind}]
 \end{aligned}$$

[◦]Verifiziere die Werte in den anderen Zeilen!

$$\begin{aligned}
&= P[2 \text{ Kinder}] \cdot P[2 \text{ Mädchen} \mid 2 \text{ Kinder}] \\
&= 0,35 \cdot (0,5)^2 = 0,0875 \\
P[\mathcal{X} = 0, \mathcal{Y} = 3] &= P[3 \text{ Kinder}] \cdot P[3 \text{ Mädchen} \mid 3 \text{ Kinder}] \\
&= 0,3 \cdot (0,5)^3 = 0,0375.
\end{aligned}$$

Wie Tabelle 2 zeigt, ist die Wahrscheinlichkeit, mindestens ein Mädchen zu haben, 0,625.⁴¹ \triangle

Definition 31. Die ZVen \mathcal{X} und \mathcal{Y} sind *gemeinsam stetig* (engl. *jointly continuous*), wenn es eine Funktion $f(x, y)$ gibt, definiert für alle $x, y \in \mathbb{R}$, so dass für jede Menge $C \subseteq \mathbb{R} \times \mathbb{R}$ gilt⁴²

$$P[(\mathcal{X}, \mathcal{Y}) \in C] = \iint_{(x,y) \in C} f(x, y) dx dy.$$

Die Funktion $f(x, y)$ wird als *gemeinsame Wahrscheinlichkeitsdichte* (engl. *joint probability density function*) von \mathcal{X} und \mathcal{Y} bezeichnet, oder kurz als *gemeinsame Dichte*.

Für je zwei Mengen reeller Zahlen $A, B \subseteq \mathbb{R}$ folgt aus dieser Definition, dass

$$P[\mathcal{X} \in A, \mathcal{Y} \in B] = \int_B \int_A f(x, y) dx dy.$$

Aus $F(a, b) = P[\mathcal{X} \leq a, \mathcal{Y} \leq b] = \int_{-\infty}^b \int_{-\infty}^a f(x, y) dx dy$ folgt (durch Differenzieren), dass immer dann, wenn die partiellen Ableitungen definiert sind, gilt

$$f(a, b) = \frac{\partial^2}{\partial a \partial b} F(a, b).$$

Wie im Fall einer einzelnen Zufallsvariablen ist $f(a, b)$ ein Maß dafür, wie wahrscheinlich es ist, dass der Zufallsvektor $(\mathcal{X}, \mathcal{Y})$ in der Nähe von (a, b) vorkommt; die Wahrscheinlichkeit, dass er exakt gleich (a, b) ist, bleibt jedoch 0.

Man beachte, wenn \mathcal{X} und \mathcal{Y} gemeinsam stetig sind, dass auch jede von ihnen einzeln stetig ist. Außerdem ist die Dichtefunktion von \mathcal{X} gleich $f_{\mathcal{X}}(x) := \int_{-\infty}^{\infty} f(x, y) dy$. Daraus folgt,

$$P[\mathcal{X} \in A] = P[\mathcal{X} \in A, \mathcal{Y} \in \mathbb{R}] = \int_A \int_{-\infty}^{\infty} f(x, y) dy dx = \int_A f_{\mathcal{X}}(x) dx.$$

Beispiel 32. Die gemeinsame Dichtefunktion von \mathcal{X} und \mathcal{Y} sei

$$f(x, y) = \begin{cases} 2e^{-x}e^{-2y} & x, y > 0 \\ 0 & \text{sonst.} \end{cases}$$

Berechne $P[\mathcal{X} > 1, \mathcal{Y} < 1]$, $P[\mathcal{X} < \mathcal{Y}]$ und $P[\mathcal{X} < a]$.⁴³

Lösung.

$$\begin{aligned}
P[\mathcal{X} > 1, \mathcal{Y} < 1] &= \int_0^1 \int_1^{\infty} 2e^{-x}e^{-2y} dx dy \\
&= \int_0^1 2e^{-2y} (-e^{-x}|_1^{\infty}) dy = e^{-1} \int_0^1 2e^{-2y} dy \\
&= e^{-1} (-e^{-2y}|_0^1) = e^{-1}(1 - e^{-2}) \\
P[\mathcal{X} < \mathcal{Y}] &= \iint_{(x,y)|x < y} 2e^{-x}e^{-2y} dx dy = \int_0^{\infty} \int_0^y 2e^{-x}e^{-2y} dx dy
\end{aligned}$$

⁴¹ Die Tabelle ist symmetrisch. Daher ist in diesem Fall die Wahrscheinlichkeit, mindestens einen Jungen zu haben, genauso groß.

⁴² Das heißt, für jede Menge in der zweidimensionalen Ebene.

⁴³ Zur Erinnerung: $\int e^{kx} dx = (1/k)e^{kx} + c$.

$$\begin{aligned}
&= \int_0^\infty 2e^{-2y}(1-e^{-y})dy = \int_0^\infty 2e^{-2y}dy - \int_0^\infty 2e^{-3y}dy \\
&= 1 - \frac{2}{3} = \frac{1}{3} \\
P[\mathcal{X} < a] &= \int_0^a \int_0^\infty 2e^{-x}e^{-2y}dydx = \int_0^a e^{-x}dx = 1 - e^{-a} \quad \triangle
\end{aligned}$$

Unabhängige Zufallsvariablen

DIE ZUFALLSVARIABLEN \mathcal{X} und \mathcal{Y} sind *unabhängig*, wenn für je zwei beliebige Mengen $A, B \subseteq \mathbb{R}$ gilt:⁴⁴

$$P[\mathcal{X} \in A, \mathcal{Y} \in B] = P[\mathcal{X} \in A] \cdot P[\mathcal{Y} \in B].$$

Das heißt, sie sind unabhängig, wenn für je zwei Mengen A, B die Ereignisse $\mathcal{E}_A = [\mathcal{X} \in A]$ und $\mathcal{F}_B = [\mathcal{Y} \in B]$ unabhängig sind. Diese Definition ist gleichbedeutend mit der Forderung, dass für alle $a, b \in \mathbb{R}$ gilt

$$P[\mathcal{X} \leq a, \mathcal{Y} \leq b] = P[\mathcal{X} \leq a] \cdot P[\mathcal{Y} \leq b],$$

oder, alternativ, dass $F(a, b) = F_{\mathcal{X}}(a) \cdot F_{\mathcal{Y}}(b)$.⁴⁵

Für diskrete ZVen \mathcal{X} und \mathcal{Y} ist Unabhängigkeit gleichbedeutend damit, dass für alle $x, y \in \mathbb{R}$ gilt

$$p(x, y) = p_{\mathcal{X}}(x) \cdot p_{\mathcal{Y}}(y),$$

wobei $p_{\mathcal{X}}$ und $p_{\mathcal{Y}}$ die Wahrscheinlichkeitsfunktionen von \mathcal{X} und \mathcal{Y} sind. Für stetige ZVen gilt die analoge Bedingung

$$f(x, y) = f_{\mathcal{X}}(x) \cdot f_{\mathcal{Y}}(y),$$

bei der wir die Wahrscheinlichkeitsfunktionen durch Dichten ersetzt haben.⁴⁶

Beispiel 33. Seien \mathcal{X} und \mathcal{Y} zwei unabhängige Zufallsvariablen, jede mit der Dichtefunktion

$$f(x) = \begin{cases} e^{-x} & x > 0 \\ 0 & \text{sonst.} \end{cases}$$

Was ist die Dichtefunktion von \mathcal{X}/\mathcal{Y} ?

Lösung. Wir berechnen zunächst die Verteilungsfunktion $F_{\mathcal{X}/\mathcal{Y}}$ von \mathcal{X}/\mathcal{Y} . Gegeben $a > 0$, haben wir

$$\begin{aligned}
F_{\mathcal{X}/\mathcal{Y}}(a) &= P[\mathcal{X}/\mathcal{Y} \leq a] = \iint_{x/y \leq a} f(x, y) dx dy \\
&= \iint_{x/y \leq a} e^{-x} e^{-y} dx dy = \int_0^\infty \int_0^{ay} e^{-x} e^{-y} dx dy \\
&= \int_0^\infty (1 - e^{-ay}) e^{-y} dy = \left(-e^{-y} + \frac{e^{-(a+1)y}}{a+1} \right) \Big|_0^\infty = 1 - \frac{1}{a+1}.
\end{aligned}$$

Leiten wir diese Funktion ab nach der Variablen a , erhalten wir die Dichte $f_{\mathcal{X}/\mathcal{Y}}(a) = 1/(a+1)^2$, für $0 < a < \infty$. \triangle

⁴⁴ Dies ist mathematisch nicht ganz richtig. Die korrekte Definition verlangt diese Eigenschaft nur von sogenannten „messbaren“ Mengen. In diesem Skriptum stehen die Ideen und Konzepte der Wahrscheinlichkeitstheorie im Vordergrund. Für eine rigorose Darstellung der mathematischen Theorie verweisen wir auf entsprechende Lehrbücher.

⁴⁵ Wir erinnern an Definition 18.

⁴⁶ Die zweite Formulierung besagt, dass zwei ZV schon unabhängig sind, wenn alle Ereignisse unabhängig sind, die durch Einschränkung der möglichen Werte auf Intervalle der Form $(-\infty, a)$ beschrieben werden können. Dass dies gilt, muss bewiesen werden. Eine Richtung ist offensichtlich, denn sie stellt einen Spezialfall der Definition dar. Die andere gilt für sogenannte messbare Mengen A, B und benutzt die Axiome der Wahrscheinlichkeitstheorie.

⁴⁷ F ist die gemeinsame Verteilung.

⁴⁸ Wie bei Ereignissen bezeichnen wir Zufallsvariablen, die nicht unabhängig sind, als *abhängig*.

WIR KÖNNEN DEN BEGRIFF der gemeinsamen Verteilung von 2 auf n Zufallsvariablen verallgemeinern, indem wir die zu den gerade eingeführten Begriffen analoge Begriffe definieren für die gemeinsame (kumulative) Wahrscheinlichkeitsverteilung F , die gemeinsame Wahrscheinlichkeitsfunktion p für diskrete Variablen, die gemeinsame Dichte f für stetige Variablen, und für die Unabhängigkeit von n Zufallsvariablen, die vorliegt, wenn jede Teilfamilie von ZVen unabhängig ist.

Beispiel 34. Wir betrachten eine Aktie, deren Kurs sich jeden Tag ändert, wobei die Änderungen an allen Tagen unabhängig voneinander seien. Der Kurs ändere sich gemäß der folgenden Wahrscheinlichkeitsfunktion p :

$$p(x) = \begin{cases} 0,05 & x \in \{-3, 3\} \\ 0,10 & x \in \{-2, 2\} \\ 0,15 & x \in \{-1, 1\} \\ 0,40 & x = 0. \end{cases}$$

Wenn \mathcal{X}_i für die Veränderung am Tag i steht, dann ist die Wahrscheinlichkeit, dass die Aktie an den nächsten Tagen um 1, 2 und 0 zunimmt

$$P[\mathcal{X}_1 = 1, \mathcal{X}_2 = 2, \mathcal{X}_3 = 0] = p(1)p(2)p(0) = 0,15 \cdot 0,1 \cdot 0,4 = 0,006. \quad \Delta$$

Der Erwartungswert

EIN GRUNDLEGENDES KONZEPT in der Wahrscheinlichkeitstheorie ist der Erwartungswert einer Zufallsvariablen. Wenn \mathcal{X} eine diskrete Zufallsvariable ist, die die Werte x_1, x_2, \dots annimmt, dann definieren wir den *Erwartungswert* (engl. *expected value* oder *expectation*) von \mathcal{X} als

$$E[\mathcal{X}] := \sum_i x_i P[\mathcal{X} = x_i] = \sum_i x_i p(x_i).$$

Das heißt, $E[\mathcal{X}]$ ist der gewichtete Durchschnitt der Werte von \mathcal{X} , wobei die Gewichte die Wahrscheinlichkeiten des Auftretens der Werte sind.

Nehmen wir zum Beispiel an, die Wahrscheinlichkeitsfunktion p von \mathcal{X} erfüllt $p(0) = p(1) = \frac{1}{2}$, dann ist $E[\mathcal{X}] = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2}$.⁴⁷ Wenn aber $p(0) = \frac{1}{3}$ und $p(1) = \frac{2}{3}$ ist, dann gilt $E[\mathcal{X}] = 0 \cdot \frac{1}{3} + 1 \cdot \frac{2}{3} = \frac{2}{3}$. Dies folgt daraus, dass der Wert 1 doppelt so wahrscheinlich ist wie der Wert 0.

Eine intuitive Motivation für den Konzept des Erwartungswertes ergibt sich aus der frequentistischen Sichtweise. Wenn wir ein Experiment viele Male durchführen und alle Wiederholungen unabhängig voneinander sind, dann ist die Wahrscheinlichkeit, mit der wir ein bestimmtes Ereignis \mathcal{E} beobachten, $P(\mathcal{E})$.

Betrachten wir eine Zufallsvariable \mathcal{X} mit den Werten x_1, x_2, \dots . Nehmen wir an, dass \mathcal{X} den Gewinn einer Wette darstellt, gemessen in gewissen Einheiten, und wir diese Wette n -mal wiederholen, wobei n hinreichend groß ist. Die Wahrscheinlichkeitsfunktion von \mathcal{X} sei p .⁴⁸ Dann gewinnen wir in etwa $np(x_i)$ dieser Male x_i Einheiten. Da dies für alle x_i gilt, ist der gesammte Gewinn $\sum_i x_i \cdot np(x_i)$. Im Durchschnitt gewinnen wir also bei jeder Wette

$$\sum_i \frac{x_i \cdot np(x_i)}{n} = \sum_i x_i p(x_i) = E[\mathcal{X}].$$

⁴⁷ Der übliche Durchschnitt der Werte von \mathcal{X} .

⁴⁸ Das heißt, mit Wahrscheinlichkeit $p(x_i)$ gewinnen wir x_i Einheiten.

Beispiel 35. Ist \mathcal{X} das Ergebnis des Würfels mit einem fairen Würfel, dann ist $p(i) = 1/6$ für alle i mit $1 \leq i \leq 6$. Wir erhalten somit

$$E[\mathcal{X}] = 1 \left(\frac{1}{6}\right) + 2 \left(\frac{1}{6}\right) + 3 \left(\frac{1}{6}\right) + 4 \left(\frac{1}{6}\right) + 5 \left(\frac{1}{6}\right) + 6 \left(\frac{1}{6}\right) = \frac{21}{6} = \frac{7}{2}. \quad \triangle$$

Dies Beispiel macht deutlich, dass der Erwartungswert von \mathcal{X} ein Wert sein kann, den \mathcal{X} niemals annimmt. Der Erwartungswert ist also nicht der Wert, den wir mit der größten Wahrscheinlichkeit beobachten werden, sondern vielmehr der Durchschnitt der Resultate eines Experiments über eine lange Reihe von Wiederholungen.⁴⁹

Die *Indikatorvariable* (engl. *indicator variable*) für das Ereignis \mathcal{E} ist die Zufallsvariable, die den Wert 1 annimmt, wenn \mathcal{E} eintritt, und ansonsten 0 ist. Ist \mathcal{X} die Indikatorvariable für \mathcal{E} , dann ist $E[\mathcal{X}] = 1 \cdot P(\mathcal{E}) + 0 \cdot P(\overline{\mathcal{E}}) = P(\mathcal{E})$. Das heißt, der Erwartungswert einer Indikatorvariablen ist genau die Wahrscheinlichkeit des Ereignisses, das sie anzeigt.

MAN KANN AUCH DEN ERWARTUNGSWERT für stetige Zufallsvariablen definieren. Ähnlich wie im diskreten Fall legen wir für eine stetige Zufallsvariable \mathcal{X} mit der Dichte f fest:

$$E[\mathcal{X}] := \int_{-\infty}^{\infty} x f(x) dx.$$

Beispiel 36. Unsere IKEA-Bestellung wird irgendwann nach 17:00 Uhr eintreffen. Wir wissen aus Erfahrung, dass die Anzahl der Stunden \mathcal{X} , die wir auf die Ankunft warten müssen, eine Zufallsvariable ist mit der Dichte

$$f(x) = \begin{cases} \frac{1}{2} & 0 < x < 2 \\ 0 & \text{sonst.} \end{cases}$$

Die voraussichtliche Dauer der Wartezeit beträgt dann

$$E[\mathcal{X}] = \int_0^2 \frac{x}{2} dx = 1.$$

Das heißt, im Durchschnitt müssen wir eine Stunde warten. △

Der Erwartungswert einer Verteilung kann intuitiv als ihr Schwerpunkt verstanden werden. Nehmen wir für die diskrete Zufallsvariable \mathcal{X} an, dass sich an jedem Punkt x_i ein materielles Objekt mit einer Masse proportional zu $p(x_i)$ befindet. Dann ist $E[\mathcal{X}]$ der Punkt, an dem man diese Struktur ausbalancieren kann. Siehe Abbildung 8. Man beachte auch, dass für jede Zufallsvariable \mathcal{X} der Erwartungswert $E[\mathcal{X}]$ in der gleichen Größeneinheit gemessen wird wie \mathcal{X} .

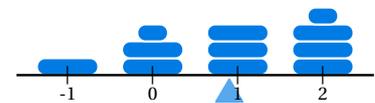


Abbildung 8: Erwartungswert als Schwerpunkt.

Eigenschaften des Erwartungswerts

WIR BETRACHTEN EINE ZUFALLSVARIABLE \mathcal{X} mit ihrer Wahrscheinlichkeitsverteilung. Oft sind wir nicht an $E[\mathcal{X}]$ interessiert, sondern an einer *Funktion* von \mathcal{X} , zum Beispiel $g(\mathcal{X})$, und ihrem Erwartungswert. Man beachte, dass $g(\mathcal{X})$ selbst eine Zufallsvariable ist und daher eine Wahrscheinlichkeitsverteilung hat, die von der Verteilung von \mathcal{X} abhängt. Gelingt es uns, diese Verteilung zu bestimmen, dann können wir $E[g(\mathcal{X})]$ aus ihr ableiten.

⁴⁹ Wenn wir viele Male würfeln, konvergiert der Durchschnitt der Ergebnisse gegen 3.5.

Beispiel 37. Sei \mathcal{X} eine Zufallsvariable mit der Wahrscheinlichkeitsfunktion $p(0) = 0,3$, $p(1) = 0,5$, und $p(2) = 0,2$. Wir wollen $E[\mathcal{X}^2]$ berechnen.

Lösung. Wenn $\mathcal{Y} = \mathcal{X}^2$ ist, dann nimmt \mathcal{Y} die Werte $0 = 0^2$, $1 = 1^2$ und $4 = 2^2$ an mit den Wahrscheinlichkeiten $0,3$ sowie $0,5$ und $0,2$. Daher gilt

$$E[\mathcal{X}^2] = E[\mathcal{Y}] = 0 \cdot 0,3 + 1 \cdot 0,5 + 4 \cdot 0,2 = 1,3. \quad \triangle$$

Beispiel 38. Die Zeit, die für die Korrektur eines Fehlers in einem Softwaresystem benötigt wird, sei eine Zufallsvariable \mathcal{X} mit der Dichte

$$f(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{sonst.} \end{cases}$$

Wenn das Vorhandensein eines Fehlers während einer Zeit x Kosten in Höhe von x^3 mit sich bringt, wie hoch sind dann die zu erwartenden Kosten eines solchen Fehlers?

Lösung. Sei $\mathcal{Y} = \mathcal{X}^3$ die Zufallsvariable für die Kosten. Die Verteilung $F_{\mathcal{Y}}$ von \mathcal{Y} ist gegeben, für alle $0 \leq a \leq 1$, durch

$$F_{\mathcal{Y}}(a) = P[\mathcal{Y} \leq a] = P[\mathcal{X}^3 \leq a] = P[\mathcal{X} \leq a^{1/3}] = \int_0^{a^{1/3}} 1 dx = a^{1/3}.$$

Durch Differenzieren erhalten wir die Dichte $f_{\mathcal{Y}}(a) = \frac{1}{3} a^{-2/3}$ für $0 \leq a < 1$. Daraus folgt,

$$\begin{aligned} E[\mathcal{X}^3] &= E[\mathcal{Y}] = \int_{-\infty}^{\infty} a f_{\mathcal{Y}}(a) da \\ &= \int_0^1 a \frac{1}{3} a^{-2/3} da = \frac{1}{3} \int_0^1 a^{1/3} da \\ &= \frac{1}{3} \frac{3}{4} a^{4/3} \Big|_0^1 = 1/4 \end{aligned} \quad \triangle$$

Wie man sieht, können wir mit diesem Ansatz immer den Erwartungswert einer beliebigen Funktion von \mathcal{X} bestimmen, wenn wir die Verteilung von \mathcal{X} kennen. Das Verfahren ist allerdings nicht gerade einfach. Ein leichter Weg, diesen Erwartungswert zu berechnen, beruht auf der folgenden Intuition: Da $g(\mathcal{X})$ immer dann den Wert $g(x)$ annimmt, wenn $\mathcal{X} = x$ ist, sollte $E[g(\mathcal{X})]$ der gewichtete Durchschnitt der Werte von $g(x)$ sein, wobei die Gewichte durch $p(x)$ für jeden möglichen Wert x gegeben sind.⁵⁰

Satz 39 (Erwartungswert von Funktionen/LOTUS-Theorem).⁵¹ *Seien \mathcal{X} eine Zufallsvariable und g eine reellwertige Funktion.*

1. *Ist \mathcal{X} diskret, mit der Wahrscheinlichkeitsfunktion $p(x)$, dann ist*

$$E[g(\mathcal{X})] = \sum_x g(x)p(x);$$

2. *Ist \mathcal{X} stetig, mit der Dichte $f(x)$, dann ist*

$$E[g(\mathcal{X})] = \int_{-\infty}^{\infty} g(x)p(x) dx.$$

⁵⁰ Diese Intuition ist allgemein zutreffend. Wir können nachprüfen, dass der folgende Satz für die beiden vorherigen Beispiele gilt.

⁵¹ Auf Englisch ist dieser Satz bekannt als *Law of the Unconscious Statistician*, kurz *LOTUS*, weil die Aussage selbstverständlich erscheint, obwohl der Beweis nicht ganz offensichtlich ist.

Korollar 40. Für alle Konstanten $a, b \in \mathbb{R}$ gilt $E[a\mathcal{X} + b] = aE[\mathcal{X}] + b$.

Beweis. Wir geben hier nur einen Beweis für den diskreten Fall.⁵²

$$E[a\mathcal{X} + b] = \sum_x (ax + b)p(x) = a \sum_x xp(x) + b \sum_x p(x) = aE[\mathcal{X}] + b \quad \square$$

Man beachte insbesondere, dass $E[b] = b$ and $E[a\mathcal{X}] = aE[\mathcal{X}]$ ist.

Der Erwartungswert von \mathcal{X} wird auch *Mittelwert* oder einfach *Mittel* (engl. *mean*) von \mathcal{X} genannt. Manchmal wird der Erwartungswert auch als *erstes Moment* (engl. *first moment*) von \mathcal{X} bezeichnet. Im Allgemeinen ist das *n-te Moment* von \mathcal{X} , für ein beliebiges $n \in \mathbb{N}$, definiert als der Erwartungswert $E[\mathcal{X}^n]$, den wir nach dem LOTUS-Theorem (Satz 39) bestimmen können.

DAS LOTUS-THEOREM in Satz 39 kann auf mehrere Zufallsvariablen erweitert werden. Im Fall von zwei Zufallsvariablen \mathcal{X} and \mathcal{Y} und wenn g eine Funktion von Paaren reeller Zahlen ist, haben wir

$$E[g(\mathcal{X}, \mathcal{Y})] = \begin{cases} \sum_y \sum_x g(x, y)p(x, y) & \mathcal{X}, \mathcal{Y} \text{ diskret} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)f(x, y) dx dy & \mathcal{X}, \mathcal{Y} \text{ gemeinsam stetig.} \end{cases}$$

Insbesondere wenn $g(\mathcal{X}, \mathcal{Y}) = \mathcal{X} + \mathcal{Y}$ ist und \mathcal{X}, \mathcal{Y} gemeinsam stetig sind, erhalten wir

$$\begin{aligned} E[\mathcal{X} + \mathcal{Y}] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x + y)f(x, y) dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xf(x, y) dx dy + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} yf(x, y) dx dy \\ &= E[\mathcal{X}] + E[\mathcal{Y}], \end{aligned} \quad (\dagger)$$

wobei (\dagger) durch Anwendung von $g_{\mathcal{X}}(\mathcal{X}, \mathcal{Y}) = \mathcal{X}$ folgt; das heißt

$$E[\mathcal{X}] = E[g_{\mathcal{X}}(\mathcal{X}, \mathcal{Y})] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xf(x, y) dx dy,$$

und genauso für $g_{\mathcal{Y}}(\mathcal{X}, \mathcal{Y}) = \mathcal{Y}$.[◦]

Da die Addition assoziativ ist, können wir dieses Argument einfach wiederholen, um eine allgemeine Aussage über Summen von Zufallsvariablen zu erhalten: für alle $n \in \mathbb{N}$ gilt

$$E \left[\sum_{i=1}^n \mathcal{X}_i \right] = \sum_{i=1}^n E[\mathcal{X}_i].$$

Beispiel 41. Es werden zwei faire Würfel geworfen. Was ist der Erwartungswert ihrer Summe?

Lösung. Ist \mathcal{X} die Zufallsvariable, die die Summe beschreibt, so können wir den Erwartungswert berechnen können als $E[\mathcal{X}] = \sum_{i=1}^{12} iP[\mathcal{X} = i]$. Einfacher ist es jedoch, zwei Variablen \mathcal{Y}_1 und \mathcal{Y}_2 einzuführen, die jede das Ergebnis eines einzelnen Würfels darstellen. Dann ist $\mathcal{X} = \mathcal{Y}_1 + \mathcal{Y}_2$ und $E[\mathcal{X}] = E[\mathcal{Y}_1] + E[\mathcal{Y}_2] = 7$.⁵³

Beispiel 42. Wir haben n verschiedene Paar Schuhe, die alle durcheinander im Schrank liegen. Wenn wir sie völlig zufällig kombinieren (indem wir einen linken und einen rechten Schuh auswählen), wie viele richtige Paare erhalten wir dann voraussichtlich?

⁵² Der stetige Fall bleibt dem interessierten Leser als Übung belassen.

[◦]Versuchen Sie zu beweisen, dass dies auch für diskrete Variablen gilt.

⁵³ Siehe Beispiel 35.

Lösung. Sei \mathcal{X} die Anzahl der richtigen Paare. Dann ist \mathcal{X} die Summe der Zufallsvariablen $\mathcal{X}_1, \dots, \mathcal{X}_n$, die für jedes i , $1 \leq i \leq n$, anzeigen, ob das i -te Paar korrekt gebildet wurde.⁵⁴ Wir definieren also

$$\mathcal{X}_i = \begin{cases} 1 & \text{das } i\text{-te Paar ist korrekt} \\ 0 & \text{sonst.} \end{cases}$$

Dann gilt $\mathcal{X} = \sum_{i=1}^n \mathcal{X}_i$. Für jeden rechten Schuh ist die Wahrscheinlichkeit, dass er mit dem i -ten linken Schuh zusammengestellt wird, gleich groß. Daraus folgt $P[\mathcal{X}_i = 1] = 1/n$ ⁵⁵ und somit $E[\mathcal{X}_i] = 1/n$. Daraus ergibt sich $E[\mathcal{X}] = \sum_{i=1}^n E[\mathcal{X}_i] = n(1/n) = 1$.⁵⁶ \triangle

AUF EINE WICHTIGE EIGENSCHAFT des Mittelwerts stoßen wir, wenn wir versuchen, den Wert einer Zufallsvariablen \mathcal{X} vorherzusagen. Angenommen, wir sagen voraus, dass \mathcal{X} den Wert c annimmt (z.B. beim Abschluss einer Wette), dann interessiert uns, wie weit wir mit unserer Vorhersage neben dem tatsächlichen Wert liegen. Diese Abweichung wird oft gemessen durch das *Quadrat des Vorhersagefehlers* $(\mathcal{X} - c)^2$ (engl. *square of the prediction error*). Setzen wir $\mu = E[\mathcal{X}]$, dann gilt

$$\begin{aligned} E[(\mathcal{X} - c)^2] &= E[(\mathcal{X} - \mu + \mu - c)^2] = E[(\mathcal{X} - \mu)^2 + 2(\mathcal{X} - \mu)(\mu - c) + (\mu - c)^2] \\ &= E[(\mathcal{X} - \mu)^2] + 2(\mu - c)E[\mathcal{X} - \mu] + (\mu - c)^2 \\ &= E[(\mathcal{X} - \mu)^2] + (\mu - c)^2 \quad (\dagger) \\ &\geq E[(\mathcal{X} - \mu)^2], \end{aligned}$$

wobei (\dagger) aus $E[\mathcal{X} - \mu] = E[\mathcal{X}] - \mu = 0$ folgt. Der durchschnittliche quadratische Fehler wird also minimiert, wenn wir vorhersagen, dass \mathcal{X} gleich seinem Mittelwert μ ist.⁵⁷

Die Varianz

DER ERWARTUNGSWERT einer Zufallsvariablen \mathcal{X} stellt das gewichtete Mittel der Werte dar, die sie annehmen kann, aber diese Information sagt nicht viel über die *Verteilung* der Werte aus, etwa wie weit die Werte voneinander entfernt sind.⁵⁸ Zum Beispiel haben die Zufallsvariablen \mathcal{X} und \mathcal{Y} mit $p_{\mathcal{X}}(0) = 1$ bzw. $p_{\mathcal{Y}}(-100) = p_{\mathcal{Y}}(100) = 0,5$ beide den gleichen Erwartungswert, nämlich 0. Die Werte von \mathcal{Y} liegen jedoch viel weiter auseinander als die von \mathcal{X} , das konstant ist.

Um die Streuung der Werte von \mathcal{X} zu messen, können wir versuchen zu bestimmen, wie weit \mathcal{X} im Durchschnitt von seinem Mittelwert μ entfernt ist; das heißt, wir könnten $E[|\mathcal{X} - \mu|]$ als Streuungsmaß wählen. Absolute Werte sind jedoch oft mathematisch schwierig zu handhaben.⁵⁹ Deshalb betrachten wir die erwartete quadratische Differenz zwischen \mathcal{X} und μ .

Definition 43. Für eine Zufallsvariable \mathcal{X} mit Erwartungswert $E[\mathcal{X}] = \mu$ definieren wir die *Varianz* (engl. *variance*) von \mathcal{X} als

$$\text{Var}(\mathcal{X}) := E[(\mathcal{X} - \mu)^2].$$

⁵⁴ Die \mathcal{X}_i sind Indikatorvariablen.

⁵⁵ Dies ist die Wahrscheinlichkeit, dass der i -te linke Schuh zum rechten Schuh passt.

⁵⁶ Also unabhängig davon, mit wie vielen Paaren wir beginnen, wird im Schnitt nur ein Paar richtig sein.

⁵⁷ Der beste Prädiktor für eine Zufallsvariable im Sinne einer Minimierung des erwarteten Fehlerquadrats ist einfach der Mittelwert.

⁵⁸ Die *Streuung* oder *Variation* der Werte.

⁵⁹ Zum Beispiel ist die Betragsfunktion nicht differenzierbar.

Man beachte, dass

$$\begin{aligned} \text{Var}(\mathcal{X}) &= E[(\mathcal{X} - \mu)^2] = E[\mathcal{X}^2 - 2\mu\mathcal{X} + \mu^2] = E[\mathcal{X}^2] - 2\mu E[\mathcal{X}] + \mu^2 \\ &= E[\mathcal{X}^2] - \mu^2 = E[\mathcal{X}^2] - E[\mathcal{X}]^2. \end{aligned}$$

Dies ist oft ein einfacherer Weg zur Berechnung der Varianz.

Beispiel 44. Sei \mathcal{X} das Ergebnis eines Wurfs mit einem fairen Würfel. Wir wissen, dass $P[\mathcal{X} = i] = 1/6$ ist für $1 \leq i \leq 6$. Dann gilt

$$E[\mathcal{X}^2] = \frac{1}{6} \sum_{i=1}^6 i^2 = \frac{1}{6}(1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2) = \frac{91}{6}.$$

Wir haben früher gesehen, dass $E[\mathcal{X}] = 7/2$ ist. Somit haben wir

$$\text{Var}(\mathcal{X}) = E[\mathcal{X}^2] - (E[\mathcal{X}])^2 = \frac{91}{6} - \frac{49}{4} = \frac{35}{12}. \quad \triangle$$

Die Varianz einer linearen Transformation von \mathcal{X} ist ebenfalls leicht zu berechnen. Sei $\mu = E[\mathcal{X}]$.⁶⁰ Dann gilt

$$\begin{aligned} \text{Var}(a\mathcal{X} + b) &= E[(a\mathcal{X} + b - E[a\mathcal{X} + b])^2] = E[(a\mathcal{X} + b - a\mu - b)^2] \\ &= E[(a\mathcal{X} - a\mu)^2] = E[a^2(\mathcal{X} - \mu)^2] = a^2 E[(\mathcal{X} - \mu)^2] \\ &= a^2 \text{Var}(\mathcal{X}). \end{aligned}$$

⁶⁰ Zur Erinnerung, $E[a\mathcal{X} + b] = aE[\mathcal{X}] + b$.

Insbesondere bedeutet dies, dass $\text{Var}(b) = 0$ und $\text{Var}(\mathcal{X} + b) = \text{Var}(\mathcal{X})$ ist für jede beliebige Konstante b .⁶¹ Mit anderen Worten, Konstanten haben eine Varianz von 0, und die Verschiebung der Werte von \mathcal{X} um eine Konstante hat keinen Einfluss auf die Varianz.^o Die Skalierung von \mathcal{X} mit einer Konstanten skaliert jedoch die Varianz quadratisch, d.h. $\text{Var}(a\mathcal{X}) = a^2 \text{Var}(\mathcal{X})$.

⁶¹ Indem wir $a = 0$ bzw. $a = 1$ setzen.

^o Überlegen Sie, warum dies so ist.

Der Wert $\sqrt{\text{Var}(\mathcal{X})}$, die so genannte *Standardabweichung* (engl. *standard deviation*) von \mathcal{X} , hat die gleiche Größeneinheit wie der Mittelwert.

Kovarianz

WIR HABEN GESEHEN, DASS der Mittelwert einer Summe von Zufallsvariablen gleich der Summe ihrer Mittelwerte ist. Dieses Resultat lässt sich im Allgemeinen nicht auf die Varianzen übertragen. In der Tat wissen wir bereits, dass

$$\text{Var}(\mathcal{X} + \mathcal{X}) = \text{Var}(2\mathcal{X}) = 4\text{Var}(\mathcal{X}) \neq \text{Var}(\mathcal{X}) + \text{Var}(\mathcal{X}).$$

Wenn jedoch zwei Zufallsvariablen unabhängig sind, dann stimmt die Varianz ihrer Summe mit der Summe ihrer Varianzen überein. Wir werden dies mit Hilfe der *Kovarianz* zeigen.

Definition 45. Es seien \mathcal{X} und \mathcal{Y} zwei Zufallsvariablen und $\mu_{\mathcal{X}} = E[\mathcal{X}]$, $\mu_{\mathcal{Y}} = E[\mathcal{Y}]$. Die *Kovarianz* (engl. *covariance*) von \mathcal{X} und \mathcal{Y} ist dann definiert als

$$\text{Cov}(\mathcal{X}, \mathcal{Y}) := E[(\mathcal{X} - \mu_{\mathcal{X}})(\mathcal{Y} - \mu_{\mathcal{Y}})].$$

Expandieren wir diese Definition, erhalten wir

$$\begin{aligned} \text{Cov}(\mathcal{X}, \mathcal{Y}) &= E[\mathcal{X}\mathcal{Y} - \mu_{\mathcal{X}}\mathcal{Y} - \mu_{\mathcal{Y}}\mathcal{X} + \mu_{\mathcal{X}}\mu_{\mathcal{Y}}] \\ &= E[\mathcal{X}\mathcal{Y}] - \mu_{\mathcal{X}}E[\mathcal{Y}] - \mu_{\mathcal{Y}}E[\mathcal{X}] + \mu_{\mathcal{X}}\mu_{\mathcal{Y}} \\ &= E[\mathcal{X}\mathcal{Y}] - \mu_{\mathcal{X}}\mu_{\mathcal{Y}} - \mu_{\mathcal{Y}}\mu_{\mathcal{X}} + \mu_{\mathcal{X}}\mu_{\mathcal{Y}} = E[\mathcal{X}\mathcal{Y}] - E[\mathcal{X}]E[\mathcal{Y}]. \end{aligned}$$

Man beachte, dass die Kovarianz symmetrisch ist, und $\text{Cov}(\mathcal{X}, \mathcal{X}) = \text{Var}(\mathcal{X})$. Darüber hinaus gilt für jede Konstante a , dass $\text{Cov}(a\mathcal{X}, \mathcal{Y}) = a\text{Cov}(\mathcal{X}, \mathcal{Y})$. Die Kovarianz ist außerdem additiv.

Satz 46. $\text{Cov}(\mathcal{X}_1 + \mathcal{X}_2, \mathcal{Y}) = \text{Cov}(\mathcal{X}_1, \mathcal{Y}) + \text{Cov}(\mathcal{X}_2, \mathcal{Y})$.

Beweis.

$$\begin{aligned} \text{Cov}(\mathcal{X}_1 + \mathcal{X}_2, \mathcal{Y}) &= E[(\mathcal{X}_1 + \mathcal{X}_2)\mathcal{Y}] - E[\mathcal{X}_1 + \mathcal{X}_2]E[\mathcal{Y}] \\ &= E[\mathcal{X}_1\mathcal{Y}] + E[\mathcal{X}_2\mathcal{Y}] - (E[\mathcal{X}_1] + E[\mathcal{X}_2])E[\mathcal{Y}] \\ &= E[\mathcal{X}_1\mathcal{Y}] - E[\mathcal{X}_1]E[\mathcal{Y}] + E[\mathcal{X}_2\mathcal{Y}] - E[\mathcal{X}_2]E[\mathcal{Y}] \\ &= \text{Cov}(\mathcal{X}_1, \mathcal{Y}) + \text{Cov}(\mathcal{X}_2, \mathcal{Y}). \quad \square \end{aligned}$$

Dies kann leicht auf beliebige Summen verallgemeinert werden und unter Verwendung der Symmetrie der Kovarianz erhalten wir das folgende Theorem.

Theorem 47. $\text{Cov}(\sum_{i=1}^n \mathcal{X}_i, \sum_{j=1}^m \mathcal{Y}_j) = \sum_{i=1}^n \sum_{j=1}^m \text{Cov}(\mathcal{X}_i, \mathcal{Y}_j)$.

Wir können nun die Varianz einer Summe von Zufallsvariablen bestimmen als⁶²

$$\begin{aligned} \text{Var}\left(\sum_{i=1}^n \mathcal{X}_i\right) &= \text{Cov}\left(\sum_{i=1}^n \mathcal{X}_i, \sum_{i=1}^n \mathcal{X}_i\right) \\ &= \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(\mathcal{X}_i, \mathcal{X}_j) \\ &= \sum_{i=1}^n \left[\sum_{j \neq i} \text{Cov}(\mathcal{X}_i, \mathcal{X}_j) + \text{Cov}(\mathcal{X}_i, \mathcal{X}_i) \right] \\ &= \sum_{i=1}^n \sum_{j \neq i} \text{Cov}(\mathcal{X}_i, \mathcal{X}_j) + \sum_{i=1}^n \text{Cov}(\mathcal{X}_i, \mathcal{X}_i) \\ &= \sum_{i=1}^n \sum_{j \neq i} \text{Cov}(\mathcal{X}_i, \mathcal{X}_j) + \sum_{i=1}^n \text{Var}(\mathcal{X}_i). \end{aligned}$$

Für $n = 2$ bedeutet dies, dass

$$\begin{aligned} \text{Var}(\mathcal{X} + \mathcal{Y}) &= \text{Var}(\mathcal{X}) + \text{Var}(\mathcal{Y}) + \text{Cov}(\mathcal{X}, \mathcal{Y}) + \text{Cov}(\mathcal{Y}, \mathcal{X}) \\ &= \text{Var}(\mathcal{X}) + \text{Var}(\mathcal{Y}) + 2\text{Cov}(\mathcal{X}, \mathcal{Y}). \end{aligned}$$

Wir kommen nun zu dem Resultat, das wir bereits angekündigt haben.

Theorem 48. Sind \mathcal{X} und \mathcal{Y} unabhängig, dann ist $\text{Cov}(\mathcal{X}, \mathcal{Y}) = 0$.⁶³

Beweis. Es reicht zu zeigen, dass $E[\mathcal{X}\mathcal{Y}] = E[\mathcal{X}]E[\mathcal{Y}]$ ist. Wir zeigen dies nur für den diskreten Fall.⁶⁴

$$E[\mathcal{X}\mathcal{Y}] = \sum_j \sum_i x_i y_j P[\mathcal{X} = x_i, \mathcal{Y} = y_j]$$

⁶² Zur Erinnerung, $\text{Cov}(\mathcal{X}, \mathcal{X}) = \text{Var}(\mathcal{X})$.

⁶³ Dies ist keine Äquivalenz: Zwei Variablen können die Kovarianz 0 haben und trotzdem nicht unabhängig sein.

⁶⁴ Für andere Fälle ist der Beweis analog.

$$\begin{aligned}
 &= \sum_j \sum_i x_i y_j P[\mathcal{X} = x_i] P[\mathcal{Y} = y_j] & (\dagger) \\
 &= \sum_j y_j P[\mathcal{Y} = y_j] \sum_i x_i P[\mathcal{X} = x_i] = E[\mathcal{X}] E[\mathcal{Y}],
 \end{aligned}$$

wobei (†) aus der Unabhängigkeit folgt. □

Insbesondere, wenn $\mathcal{X}_1, \dots, \mathcal{X}_n$ unabhängig sind, gilt

$$\text{Var} \left(\sum_{i=1}^n \mathcal{X}_i \right) = \sum_{i=1}^n \text{Var}(\mathcal{X}_i).$$

Beispiel 49. Was ist die Varianz der Summe von 10 unabhängigen Würfeln eines fairen Würfels?

Lösung. Wenn \mathcal{X}_i für den i -ten Wurf steht, erhalten wir[◦]

[◦]Wir erinnern an Beispiel 44.

$$\text{Var} \left(\sum_{i=1}^{10} \mathcal{X}_i \right) = \sum_{i=1}^{10} \text{Var}(\mathcal{X}_i) = 10 \cdot \left(\frac{35}{12} \right) = \frac{175}{6}. \quad \triangle$$

Intuitiv beschreibt die Kovarianz die Beziehung zwischen zwei Variablen. Betrachten wir zum Beispiel die Indikatorvariablen \mathcal{X}, \mathcal{Y} für die Ereignisse \mathcal{E} bzw. \mathcal{F} . Dann folgt, dass

$$\begin{aligned}
 \text{Cov}(\mathcal{X}, \mathcal{Y}) &= E[\mathcal{X}\mathcal{Y}] - E[\mathcal{X}] \cdot E[\mathcal{Y}] \\
 &= P[\mathcal{X} = 1, \mathcal{Y} = 1] - P[\mathcal{X} = 1] \cdot P[\mathcal{Y} = 1] \\
 &= P(\mathcal{E}\mathcal{F}) - P(\mathcal{E}) \cdot P(\mathcal{F}).
 \end{aligned}$$

Es ist also $\text{Cov}(\mathcal{X}, \mathcal{Y}) > 0$ genau dann, wenn $P(\mathcal{E}\mathcal{F}) > P(\mathcal{E}) \cdot P(\mathcal{F})$ gilt, oder, was gleichbedeutend ist, wenn $P(\mathcal{F} | \mathcal{E}) > P(\mathcal{F})$ gilt. In einfachen Worten: Die Kovarianz sagt uns, ob es wahrscheinlicher oder unwahrscheinlicher ist, \mathcal{F} zu beobachten, wenn wir wissen, dass \mathcal{E} eingetreten ist.⁶⁵

Im Allgemeinen drückt eine positive Kovarianz zwischen zwei ZVen aus, dass beide Variablen gemeinsam wachsen,⁶⁶ während eine negative Kovarianz besagt, dass die eine abnimmt, während die andere zunimmt. Die Stärke der Beziehung zwischen den beiden Variablen wird durch ihre *Korrelation* (engl. *correlation*) angezeigt, definiert als⁶⁷

$$\text{Corr}(\mathcal{X}, \mathcal{Y}) = \frac{\text{Cov}(\mathcal{X}, \mathcal{Y})}{\sqrt{\text{Var}(\mathcal{X}) \text{Var}(\mathcal{Y})}}.$$

Dieser dimensionslose Wert liegt immer zwischen -1 und 1.

⁶⁵ Und aufgrund der Symmetrie bedingter Wahrscheinlichkeiten auch die Umkehrung.

⁶⁶ Das heißt, \mathcal{Y} wächst in dem Maße, wie \mathcal{X} wächst.

⁶⁷ Dieser Wert wird auch als *Pearson-scher Korrelationskoeffizient* bezeichnet (engl. *Pearson's correlation coefficient*), benannt nach dem englischen Statistiker Karl Pearson, der ihn mathematisch begründete.

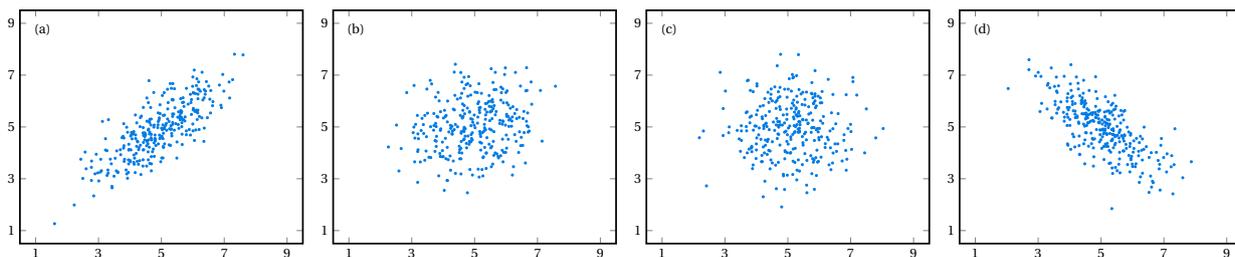


Abbildung 9: Zufallsvariablen \mathcal{X} und \mathcal{Y} mit Korrelationskoeffizienten (a) 0,75; (b) 0,2; (c) 0; und (d) -0,75.

Momenterzeugende Funktionen

JEDE ZUFALLSVARIABLE \mathcal{X} hat eine *momenterzeugende Funktion* (engl. *moment generating function*), kurz *MEF*, die für jedes $t \in \mathbb{R}$ definiert ist durch

$$\phi(t) := E[e^{t\mathcal{X}}] = \begin{cases} \sum_x e^{tx} p(x) & \text{wenn } \mathcal{X} \text{ diskret ist} \\ \int_x e^{tx} f(x) dx & \text{wenn } \mathcal{X} \text{ stetig ist.} \end{cases}$$

Diese Funktion wird momenterzeugend genannt, weil man alle Momente von \mathcal{X} durch Differenzieren und Auswerten an der Stelle Null erhalten kann.⁶⁸ Ein Beispiel

$$\phi'(t) = \frac{d}{dt} E[e^{t\mathcal{X}}] = E\left[\frac{d}{dt} e^{t\mathcal{X}}\right] = E[\mathcal{X} e^{t\mathcal{X}}].$$

Folglich ist $\phi'(0) = E[\mathcal{X}]$. Analog dazu ist $\phi''(t) = E[\mathcal{X}^2 e^{t\mathcal{X}}]$ und somit $\phi''(0) = E[\mathcal{X}^2]$. Im Allgemeinen ist das n -te Moment von \mathcal{X} die n -te Ableitung von ϕ , ausgewertet an der Stelle $t = 0$.

Eine wichtige Eigenschaft momenterzeugender Funktionen ist, dass die MEF einer Summe unabhängiger Zufallsvariablen das Produkt der einzelnen MEFen ist. Für zwei Zufallsvariablen \mathcal{X}, \mathcal{Y} ist die MEF von $\mathcal{X} + \mathcal{Y}$

$$\begin{aligned} \phi_{\mathcal{X}+\mathcal{Y}}(t) &= E\left[e^{t(\mathcal{X}+\mathcal{Y})}\right] \\ &= E\left[e^{t\mathcal{X}} e^{t\mathcal{Y}}\right] \\ &= E\left[e^{t\mathcal{X}}\right] E\left[e^{t\mathcal{Y}}\right] = \phi_{\mathcal{X}}(t) \phi_{\mathcal{Y}}(t), \end{aligned} \quad (\dagger)$$

wobei (\dagger) folgt, weil \mathcal{X} and \mathcal{Y} , und somit auch $e^{t\mathcal{X}}$ und $e^{t\mathcal{Y}}$ unabhängig sind.⁶⁹

Interessanterweise bestimmen MEFen *eindeutig* die Verteilung von Zufallsvariablen, d.h. es besteht eine Eins-zu-Eins-Entsprechung zwischen Verteilungen und ihren momenterzeugenden Funktionen.

Das schwache Gesetz der großen Zahlen

WIR BEWEISEN NUN zwei wichtige Theoreme der Wahrscheinlichkeitstheorie.

Theorem 50 (Markovsche Ungleichung). *Sei \mathcal{X} eine Zufallsvariable, die nur nicht-negative Werte annimmt. Dann gilt für jedes $a > 0$*

$$P[\mathcal{X} \geq a] \leq \frac{E[\mathcal{X}]}{a}.$$

Beweis. Wir beweisen die Aussage für eine stetige Zufallsvariable mit der Dichte f :

$$\begin{aligned} E[\mathcal{X}] &= \int_0^\infty xf(x) dx = \int_0^a xf(x) dx + \int_a^\infty xf(x) dx \\ &\geq \int_a^\infty xf(x) dx \geq \int_a^\infty af(x) dx = a \int_a^\infty f(x) dx = aP[\mathcal{X} \geq a]. \quad \square \end{aligned}$$

⁶⁸ Das n -te Moment von \mathcal{X} ist $E[\mathcal{X}^n]$.

⁶⁹ Im Beweis von Theorem 48 haben wir gezeigt, dass $E[\mathcal{X}\mathcal{Y}] = E[\mathcal{X}] \cdot E[\mathcal{Y}]$ ist, wenn die zwei Zufallsvariablen \mathcal{X} und \mathcal{Y} unabhängig sind.

Korollar 51 (Tschebyscheffsche Ungleichung). *Ist \mathcal{X} eine Zufallsvariable mit Mittelwert μ und Varianz σ^2 , dann gilt für jedes $k > 0$*

$$P[|\mathcal{X} - \mu| \geq k] \leq \frac{\sigma^2}{k^2}.$$

Beweis. Beachte, dass $(\mathcal{X} - \mu)^2$ eine nicht-negative Zufallsvariable ist. Daher können wir die Markovsche Ungleichung mit $a = k^2$ anwenden und erhalten

$$P[(\mathcal{X} - \mu)^2 \geq k^2] \leq \frac{E[(\mathcal{X} - \mu)^2]}{k^2}.$$

Da $(\mathcal{X} - \mu)^2 \geq k^2$ genau dann gilt, wenn $|\mathcal{X} - \mu| \geq k$ ist, folgt hieraus

$$P[|\mathcal{X} - \mu| \geq k] \leq \frac{E[(\mathcal{X} - \mu)^2]}{k^2} \leq \frac{\sigma^2}{k^2}. \quad \square$$

Mit den Ungleichungen von Markov und Tschebyscheff können wir Schranken für Wahrscheinlichkeiten von \mathcal{X} angeben, auch wenn wir die Verteilung von \mathcal{X} nicht kennen, solange der Mittelwert und möglicherweise die Varianz bekannt sind.⁷⁰

Beispiel 52. Nehmen wir an, dass die Anzahl der Stunden, die eine Person pro Woche arbeitet, eine Zufallsvariable mit Mittelwert 40 ist.

1. Was können wir über die Wahrscheinlichkeit sagen, dass die Person in dieser Woche mehr als 60 Stunden arbeiten wird?
2. Wenn die Varianz der Arbeitsstunden pro Woche 16 beträgt, wie hoch ist dann die Wahrscheinlichkeit, dass ihre Arbeitszeit in dieser Woche zwischen 32 und 48 Stunden liegt?

Lösung. Sei \mathcal{X} die Anzahl der Arbeitsstunden in einer Woche.

1. Durch Anwendung der Markovschen Ungleichung erhalten wir

$$P[\mathcal{X} \geq 60] \leq \frac{E[\mathcal{X}]}{60} = \frac{40}{60} = 2/3.$$

2. Mit Hilfe der Tschebyscheffschen Ungleichung schließen wir

$$P[|\mathcal{X} - 40| \geq 8] \leq \frac{16}{64} = 1/4.$$

Daraus folgern wir, dass $P[|\mathcal{X} - 40| \leq 8] = 1 - P[|\mathcal{X} - 40| \geq 8] \geq 0.75$. \triangle

Wenn wir die Tschebyscheffsche Ungleichung mit der Distanz $k\sigma$ verwenden, erhalten wir

$$P[|\mathcal{X} - \mu| \geq k\sigma] \leq \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2}.$$

Das heißt, die Wahrscheinlichkeit, dass \mathcal{X} um mindestens k Standardabweichungen von seinem Mittelwert differiert, ist beschränkt durch $\frac{1}{k^2}$.⁷¹

Eine Folge dieser Ungleichung ist, dass bei mehreren unabhängigen und identisch verteilten Zufallsvariablen ihr Durchschnitt zu ihrem Mittelwert tendiert (mit der Wahrscheinlichkeit 1).⁷²

Theorem 53 (Schwaches Gesetz der großen Zahlen). *Sei $\mathcal{X}_1, \mathcal{X}_2, \dots$ eine Folge unabhängiger identisch verteilter ZVen mit Mittelwert $E[\mathcal{X}_i] = \mu$. Dann gilt für jedes $\varepsilon > 0$*

$$\lim_{n \rightarrow \infty} P \left[\left| \frac{\sum_{i=1}^n \mathcal{X}_i}{n} - \mu \right| > \varepsilon \right] = 0.$$

⁷⁰ Wenn die Verteilung bekannt ist, können die Wahrscheinlichkeiten in der Regel genau berechnet werden, ohne dass eine Näherung nötig wäre.

⁷¹ Diese Größe nimmt quadratisch ab.

⁷² Dies ist die Grundlage vieler statistischer Analysen, die auf wiederholten Experimenten beruhen.

Beweis. Wir nehmen der Einfachheit halber an, die Zufallsvariablen haben die endliche Varianz σ^2 . Da alle Variablen unabhängig sind, folgt aus $\mathcal{Y} = \frac{\sum_{i=1}^n \mathcal{X}_i}{n}$, dass

$$E[\mathcal{Y}] = \mu \qquad \text{Var}(\mathcal{Y}) = \frac{\sigma^2}{n}.$$

Die Anwendung der Tschebyscheffsche Ungleichung ergibt nun

$$P \left[\left| \frac{\sum_{i=1}^n \mathcal{X}_i}{n} - \mu \right| > \varepsilon \right] \leq \frac{\sigma^2}{n\varepsilon^2}. \quad \square$$

Nehmen wir zum Beispiel an, wir wiederholen unabhängig voneinander eine Reihe klinischer Versuche. Sei \mathcal{E} ein bestimmtes Ereignis, das bei jedem Versuch mit der Wahrscheinlichkeit $P(\mathcal{E})$ eintritt. Ist \mathcal{X}_i die Indikatorvariable für das Auftreten von \mathcal{E} in Versuch i , dann ist $\sum_{i=1}^n \mathcal{X}_i$ die Anzahl der Beobachtungen von \mathcal{E} in den ersten n Versuchen. Da $E[\mathcal{X}_i] = P(\mathcal{E})$ ist, folgt aus dem schwachen Gesetz der großen Zahlen, dass die Wahrscheinlichkeit dafür, dass der Anteil der Versuche, in denen \mathcal{E} beobachtet wird, um mehr als ε von $P(\mathcal{E})$ abweicht, mit wachsendem n gegen 0 tendiert.[◦]

[◦]Man beachte die Namensgebung „große Zahlen“.

Spezielle Zufallsvariablen

WIR STELLEN EINIGE DER am häufigsten vorkommenden und verwendeten Verteilungen von ZVen vor.

Bernoulli- und binomialverteilte Variablen

WIR BETRACHTEN EIN EXPERIMENT, dessen Ergebnis ein *Erfolg* oder ein *Misserfolg* sein kann. Sei \mathcal{X} die Indikatorvariable für den Erfolg.⁷³ Damit hat die Wahrscheinlichkeitsfunktion von \mathcal{X} die Eigenschaft, dass $P[\mathcal{X} = 1] = p$ und $P[\mathcal{X} = 0] = 1 - p$ ist. Da die ZV \mathcal{X} nur zwei Werte annehmen kann, ist \mathcal{X} durch p , die Wahrscheinlichkeit für den Erfolg, bereits vollständig bestimmt. Eine solche Zufallsvariable wird als *Bernoulli-Variable*⁷⁴ (engl. *Bernoulli variable*) bezeichnet. Als Indikatorvariable hat sie als Erwartungswert die Erfolgswahrscheinlichkeit p .

Wenn wir das Experiment n -mal wiederholen, so dass die Durchführungen unabhängig voneinander sind, und wenn \mathcal{X} nun die Anzahl der Erfolge darstellt, dann ist \mathcal{X} eine *binomiale* Zufallsvariable (engl. *binomial RV*), verteilt nach der Binomialverteilung mit den Parametern (n, p) . Die Wahrscheinlichkeitsfunktion dieser Zufallsvariablen hat für alle i mit $0 \leq i \leq n$ die Eigenschaft, dass

$$P[\mathcal{X} = i] = \binom{n}{i} p^i (1 - p)^{n-i}.$$

In Abbildung 10 sind die Wahrscheinlichkeitsfunktionen von drei binomialverteilten Zufallsvariablen mit den Parametern $(10, 0.5)$, $(10, 0.4)$ und $(10, 0.75)$ wiedergegeben. Es fällt auf, dass zwei von ihnen vom Zentrum weg geneigt (oder schief) sind.

Beispiel 54. Ein System mit n Komponenten funktioniert, wenn mindestens die Hälfte der Komponenten funktioniert. Angenommen, jede Komponente funktioniert unabhängig von den anderen mit der Wahrscheinlichkeit p :

1. Für welche Werte von p ist ein 5-Komponenten-System zuverlässiger als ein 3-Komponenten-System?
2. Allgemein, wann ist ein System mit $2k + 1$ Komponenten zuverlässiger als ein System mit $2k - 1$ Komponenten?

⁷³ That is, $\mathcal{X} = 1$ if the experiment succeeds, and $\mathcal{X} = 0$ otherwise.

⁷⁴ Nach Jakob Bernoulli (1654–1705).

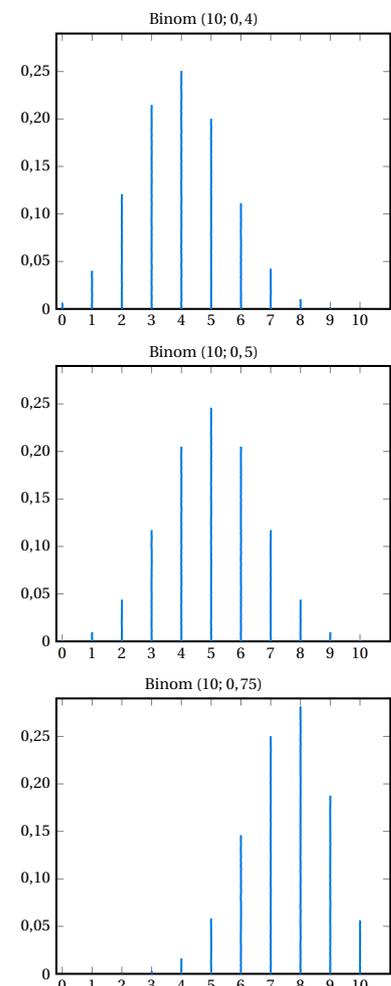


Abbildung 10: Drei binomiale Wahrscheinlichkeitsfunktionen.

Lösung. In Bezug auf die erste Frage wissen wir, dass ein 5-Komponenten-System bzw. ein 3-Komponenten-System mit den folgenden Wahrscheinlichkeiten funktionieren:

$$\binom{5}{3}p^3(1-p)^2 + \binom{5}{4}p^4(1-p) + \binom{5}{5}p^5,$$

$$\binom{3}{2}p^2(1-p) + \binom{3}{3}p^3.$$

Wir vereinfachen $\binom{5}{5}$ bzw. $\binom{3}{3}$ zu 1 und sehen, dass das 5-Komponenten-System zuverlässiger ist, wenn gilt

$$\binom{5}{3}p^3(1-p)^2 + \binom{5}{4}p^4(1-p) + p^5 \geq \binom{3}{2}p^2(1-p) + p^3$$

oder, anders ausgedrückt, wenn $3(p-1)^2(2p-1) \geq 0$ ist, was das Gleiche bedeutet wie $p \geq 1/2$.

Wir betrachten nun den allgemeinen Fall eines Systems mit $2k+1$ Komponenten. Dazu schauen wir auf die ersten $2k-1$ Komponenten und bezeichnen mit \mathcal{X} die Anzahl derjenigen Komponenten unter diesen ersten $2k-1$, die funktionieren. Das gesamte System funktioniert, wenn eine der folgenden drei Bedingungen erfüllt ist:

1. $\mathcal{X} \geq k+1$,
2. $\mathcal{X} = k$ und mindestens eine der beiden letzten Komponenten funktioniert, oder
3. $\mathcal{X} = k-1$ und die letzten zwei Komponenten funktionieren.

Die Wahrscheinlichkeit, dass das System funktioniert, ist damit gleich

$$P[\mathcal{X} \geq k+1] + P[\mathcal{X} = k](1 - (1-p)^2) + P[\mathcal{X} = k-1]p^2.$$

Entsprechend ist die Wahrscheinlichkeit, dass das kleinere System funktioniert,^o

$$P[\mathcal{X} = k] + P[\mathcal{X} \geq k+1].$$

Dies bedeutet, dass das größere System zuverlässiger ist, wenn⁷⁵

$$\begin{aligned} 0 &\leq P_{2k+1}(\text{funktioniert}) - P_{2k-1}(\text{funktioniert}) \\ &= P[\mathcal{X} = k-1]p^2 - P[\mathcal{X} = k](1-p)^2 \\ &= \binom{2k-1}{k-1}p^{k-1}(1-p)^k p^2 - \binom{2k-1}{k}p^k(1-p)^{k-1}(1-p)^2 \\ &= \binom{2k-1}{k}p^k(1-p)^k(p - (1-p)) = \binom{2k-1}{k}p^k(1-p)^k(2p-1), \quad (\dagger) \end{aligned}$$

wobei (\dagger) aus $\binom{2k-1}{k} = \binom{2k-1}{k-1}$ folgt. Das größere System ist also genau dann zuverlässiger, wenn $p \geq 1/2$ ist.⁷⁶ \triangle

Zur Erinnerung, eine binomiale Zufallsvariable \mathcal{X} gibt die Anzahl der Erfolge bei n unabhängigen Versuchen an, deren Erfolgswahrscheinlichkeit jeweils p ist. Dann ist $\mathcal{X} = \sum_{i=1}^n \mathcal{X}_i$, wobei die \mathcal{X}_i unabhängige Bernoulli-Variablen sind. Insbesondere gilt⁷⁷

^oWegen der Unabhängigkeit.

⁷⁵ ... die Wahrscheinlichkeit für das Funktionieren des größeren Systems die Wahrscheinlichkeit für das Funktionieren des kleineren Systems übersteigt, also wenn ...

⁷⁶ Für jede Zahl N gilt $N(2p-1) \geq 0$ genau dann, wenn $2Np \geq N$ ist.

⁷⁷ Man beachte, dass $\mathcal{X}^2 = \mathcal{X}$ ist, weil \mathcal{X} nur die Werte 0 oder 1 annehmen kann.

$$E[\mathcal{X}_i] = P[\mathcal{X}_i = 1] = p$$

$$\text{Var}(\mathcal{X}_i) = E[\mathcal{X}_i^2] - (E[\mathcal{X}_i])^2 = p - p^2 = p(1 - p).$$

Hieraus folgt unmittelbar, dass

$$E[\mathcal{X}] = \sum_{i=1}^n E[\mathcal{X}_i] = np$$

$$\text{Var}(\mathcal{X}) = \sum_{i=1}^n \text{Var}(\mathcal{X}_i) = np(1 - p).$$

Außerdem gilt für zwei unabhängige binomiale ZVen \mathcal{X}_1 and \mathcal{X}_2 mit Parametern (n_i, p) für $i = 1, 2$, dass ihre Summe ebenfalls binomial verteilt ist, mit Parametern $(n_1 + n_2, p)$.⁷⁸

ZUR BERECHNUNG DER VERTEILUNGSFUNKTION einer binomialen Zufallsvariablen \mathcal{X} mit den Parametern (n, p) ist es praktisch, die folgende Gleichung zu verwenden:

$$P[\mathcal{X} = k + 1] = \frac{p}{1 - p} \frac{n - k}{k + 1} P[\mathcal{X} = k].$$

Beispiel 55. Sei \mathcal{X} binomialverteilt mit $n = 5$ und $p = 0,3$. Dann erhalten wir, ausgehend von $P[\mathcal{X} = 0] = (0,7)^5 = 0,168$,

$$P[\mathcal{X} = 1] = \frac{3}{7} \frac{5}{1} P[\mathcal{X} = 0]$$

$$P[\mathcal{X} = 2] = \frac{3}{7} \frac{4}{2} P[\mathcal{X} = 1]$$

$$P[\mathcal{X} = 3] = \frac{3}{7} \frac{3}{3} P[\mathcal{X} = 2]$$

$$P[\mathcal{X} = 4] = \frac{3}{7} \frac{2}{4} P[\mathcal{X} = 3]$$

$$P[\mathcal{X} = 5] = \frac{3}{7} \frac{1}{5} P[\mathcal{X} = 4].$$

△

Hypergeometrisch verteilte Zufallsvariablen

BETRACHTEN WIR EINEN BEHÄLTER mit N funktionierenden und M defekten Batterien. Wir ziehen eine Zufallsstichprobe der Größe n .⁷⁹ Wenn \mathcal{X} die Anzahl der funktionierenden Batterien in der Stichprobe ist, dann gilt für alle $i, 0 \leq i \leq \min(N, m)$ ⁸⁰

$$P[\mathcal{X} = i] = \frac{\binom{N}{i} \binom{M}{n-i}}{\binom{N+M}{n}}.$$

Von einer Zufallsvariablen mit solch einer Wahrscheinlichkeitsfunktion sagen wir, sie ist *hypergeometrisch* (engl. *hypergeometric*) mit den Parametern N, M, n .

Intuitiv kann man sich eine hypergeometrische Zufallsvariable \mathcal{X} so vorstellen, dass die Stichprobe sequentiell gezogen wird und die Zufallsvariablen \mathcal{X}_i wie folgt definiert sind:

$$\mathcal{X}_i = \begin{cases} 1 & \text{die } i\text{-te Auswahl funktioniert} \\ 0 & \text{sonst.} \end{cases}$$

⁷⁸ Wichtig ist, dass die Erfolgswahrscheinlichkeit für beide binomialen Variablen gleich groß ist. Jedes \mathcal{X}_i gibt die Anzahl der Erfolge bei n_i unabhängigen Versuchen an. Die Summe $\mathcal{X}_1 + \mathcal{X}_2$ gibt also die Anzahl der Erfolge in der Gesamtheit der (unabhängig durchgeführten) Versuche an.

Es ist natürlich möglich, Bernoulli- und binomiale Zufallsvariablen so zu verallgemeinern, dass mehr als zwei Werte angenommen werden können. In diesem Fall würde man von einer *multinomialen* ZVen (engl. *multinomial* RV) sprechen. Sich die Eigenschaften solcher ZVen klar zu machen, bleibt dem interessierten Leser als Übung überlassen.

⁷⁹ Jede der $\binom{N+M}{n}$ Stichproben dieser Art ist gleich wahrscheinlich.

⁸⁰ Zur Vereinfachung nehmen wir $\binom{m}{r} = 0$ an, wenn $r > m$ oder $r < 0$.

Jede Auswahl hat die gleiche Wahrscheinlichkeit, eine der $N + M$ Batterien zu sein; daher ist $P[\mathcal{X}_i = 1] = \frac{N}{N+M}$. Darüber hinaus gilt für jedes $i \neq j$,⁸¹

$$P[\mathcal{X}_i = 1, \mathcal{X}_j = 1] = P[\mathcal{X}_i = 1 \mid \mathcal{X}_j = 1]P[\mathcal{X}_j = 1] = \frac{N-1}{N+M-1} \frac{N}{N+M}.$$

Da $\mathcal{X} = \sum_{i=1}^n \mathcal{X}_i$ ist, folgt

$$\begin{aligned} E[\mathcal{X}] &= \sum_{i=1}^n E[\mathcal{X}_i] = \sum_{i=1}^n P[\mathcal{X}_i = 1] = \frac{nN}{N+M}, \\ \text{Var}(\mathcal{X}) &= \sum_{i=1}^n \text{Var}(\mathcal{X}_i) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(\mathcal{X}_i, \mathcal{X}_j). \end{aligned}$$

Da jedes \mathcal{X}_i eine Bernoulli-ZV ist, gilt

$$\text{Var}(\mathcal{X}_i) = P[\mathcal{X}_i = 1](1 - P[\mathcal{X}_i = 1]) = \frac{NM}{(N+M)^2},$$

und für $i < j$ ist⁸²

$$E[\mathcal{X}_i \mathcal{X}_j] = P[\mathcal{X}_i = 1, \mathcal{X}_j = 1] = \frac{N(N-1)}{(N+M)(N+M-1)}.$$

Somit haben wir

$$\begin{aligned} \text{Cov}(\mathcal{X}_i, \mathcal{X}_j) &= E[\mathcal{X}_i \mathcal{X}_j] - E[\mathcal{X}_i]E[\mathcal{X}_j] \\ &= \frac{N(N-1)}{(N+M)(N+M-1)} - \left(\frac{N}{N+M} \right)^2 \\ &= \frac{-NM}{(N+M)^2(N+M-1)}. \end{aligned}$$

Da die Summe der Kovarianzen $\binom{n}{2}$ Terme enthält, folgt daraus, dass

$$\begin{aligned} \text{Var}(\mathcal{X}) &= \frac{nNM}{(N+M)^2} - \frac{n(n-1)NM}{(N+M)^2(N+M-1)} \\ &= \frac{nNM}{(N+M)^2} \left[1 - \frac{n-1}{N+M-1} \right]. \end{aligned}$$

Es sei nun $p = N/(N+M)$ der Anteil der Batterien, die funktionieren. Dann folgt $E[\mathcal{X}] = np$ und $\text{Var}(\mathcal{X}) = np(1-p) \left[1 - \frac{n-1}{N+M-1} \right]$.

Wenn $N + M$ gegen unendlich strebt und dabei der Anteil p der funktionierenden Batterien gleich bleibt, konvergiert $\text{Var}(\mathcal{X})$ gegen $np(1-p)$, was die Varianz einer Binomialverteilung mit den Parametern (n, p) ist.

Seien \mathcal{X} und \mathcal{Y} zwei unabhängige binomiale Zufallsvariablen mit den Parametern (n, p) bzw. (m, p) . Die bedingte Wahrscheinlichkeitsfunktion von \mathcal{X} mit der Bedingung $\mathcal{X} + \mathcal{Y} = k$ ist⁸³

$$\begin{aligned} P[\mathcal{X} = i \mid \mathcal{X} + \mathcal{Y} = k] &= \frac{P[\mathcal{X} = i, \mathcal{Y} = k-i]}{P[\mathcal{X} + \mathcal{Y} = k]} = \frac{P[\mathcal{X} = i]P[\mathcal{Y} = k-i]}{P[\mathcal{X} + \mathcal{Y} = k]} \\ &= \frac{\binom{n}{i} p^i (1-p)^{n-i} \binom{m}{k-i} p^{k-i} (1-p)^{m-(k-i)}}{\binom{n+m}{k} p^k (1-p)^{n+m-k}} \\ &= \frac{\binom{n}{i} \binom{m}{k-i}}{\binom{n+m}{k}}, \end{aligned}$$

das heißt, diese bedingte Verteilung ist hypergeometrisch.

⁸¹ Sobald wir eine funktionierende Batterie gefunden haben, kann jede weitere Wahl nur noch eine der verbleibenden $N - 1$ funktionierenden Batterien sein.

⁸² $\mathcal{X}_i \mathcal{X}_j = 1$ gdw $\mathcal{X}_i = 1 = \mathcal{X}_j$.

⁸³ Recall that the sum of two binomials (n, p) and (m, p) is a binomial $(n+m, p)$.

Stetige gleichverteilte Zufallsvariablen

EINE STETIGE ZUFALLSVARIABLE auf dem Intervall $[\alpha, \beta]$ gleichverteilt, wenn ihre Dichtefunktion die folgende Form hat (siehe Abbildung 11):

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha} & \alpha \leq x \leq \beta \\ 0 & \text{otherwise.} \end{cases}$$

Ist zum Beispiel \mathcal{X} gleichmäßig auf dem Intervall $[0, 10]$ verteilt, dann ist $P[\mathcal{X} > 6] = 0.4$ und $P[2 < \mathcal{X} < 5] = 0.3$.

Intuitiv sollte der Mittelwert einer gleichförmigen Zufallsvariablen in der Mitte ihres Intervalls liegen.[◦] Die Berechnung des Erwartungswerts entsprechend der Definition bestätigt, dass dies der Fall ist:

$$E[\mathcal{X}] = \int_{\alpha}^{\beta} \frac{x}{\beta - \alpha} dx = \frac{\beta^2 - \alpha^2}{2(\beta - \alpha)} = \frac{(\beta + \alpha)(\beta - \alpha)}{2(\beta - \alpha)} = \frac{\alpha + \beta}{2}.$$

Wir berechnen die Varianz mittels des LOTUS-Theorems:⁸⁴

$$E[\mathcal{X}^2] = \frac{1}{\beta - \alpha} \int_{\alpha}^{\beta} x^2 dx = \frac{\beta^3 - \alpha^3}{3(\beta - \alpha)} = \frac{\beta^2 + \alpha\beta + \alpha^2}{3},$$

woraus folgt

$$\text{Var}(\mathcal{X}) = \frac{\beta^2 + \alpha\beta + \alpha^2}{3} - \left(\frac{\alpha + \beta}{2}\right)^2 = \frac{\alpha^2 + \beta^2 - 2\alpha\beta}{12} = \frac{(\beta - \alpha)^2}{12}.$$

Der Wert einer auf $(0, 1)$ gleichverteilten Zufallsvariablen wird als *Zufallszahl* (engl. *random number*) bezeichnet. Es gibt mathematische Methoden, mit denen man auf dem Computer Folgen unabhängiger (Pseudo-)Zufallszahlen erzeugen kann.⁸⁵ Zufallszahlen werden häufig in klinischen Studien verwendet, insbesondere in so genannten Doppelblindtests (engl. *double-blind tests*).⁸⁶

DAS KONZEPT DER GLEICHVERTEILTEN stetigen Zufallsvariablen lässt sich auf gemeinsame Verteilungen ausweiten. Die gemeinsame Wahrscheinlichkeitsverteilung von \mathcal{X}, \mathcal{Y} ist gleichförmig über einem Gebiet G mit der Fläche a , wenn $f(x, y) = 1/a$ ist für $(x, y) \in G$ und 0 sonst. Ist G zum Beispiel der rechteckige Bereich zwischen (α_1, α_2) und (β_1, β_2) , dann kann man zeigen, dass \mathcal{X} und \mathcal{Y} unabhängige Gleichverteilungen über $[\alpha_1, \beta_1]$ und $[\alpha_2, \beta_2]$ sind.⁸⁷

Exponentiell verteilte Zufallsvariablen

EINE STETIG VERTEILTE ZUFALLSVARIABLE mit der Dichtefunktion

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

für eine bestimmte Konstante λ wird als *exponentiell* (oder *exponentiell verteilt*) bezeichnet, und λ wird oft die *Rate* der Verteilung genannt. Die kumulative Verteilungsfunktion einer solchen Variablen ist für jedes $x \geq 0$

$$F(x) = P[\mathcal{X} \leq x] = \int_0^x \lambda e^{-\lambda y} dy = 1 - e^{-\lambda x}.$$

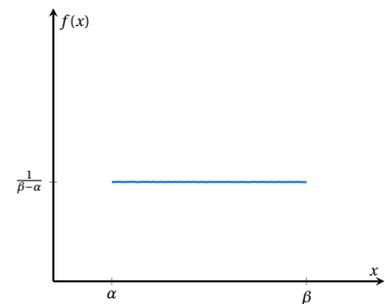


Abbildung 11: Dichte der Gleichverteilung über $[\alpha, \beta]$.

[◦]Ist Ihnen klar, warum?

⁸⁴ $x^3 - y^3 = (x^2 + xy + y^2)(x - y)$

⁸⁵ Einen *echten* Zufallszahlengenerator finden Sie unter <http://www.random.org>.

⁸⁶ Bei einem Doppelblindversuch wird eine Gruppe von Freiwilligen für eine Studie in zwei (gleich große) Untergruppen aufgeteilt; eine Gruppe erhält die Behandlung, die andere ein Placebo. Auf diese Weise lässt sich feststellen, ob (und in welchem Umfang) die Behandlung wirksam ist. Durch die Verwendung von Zufallszahlen wird sichergestellt, dass die Einteilung der Gruppen wirklich zufällig ist und nicht durch einen möglicherweise verborgenen Faktor verzerrt wird.

⁸⁷ Dazu startet man mit den jeweiligen Verteilungen von $f_{\mathcal{X}}(x)$ und $f_{\mathcal{Y}}(y)$ und zeigt, dass das Produkt $f_{\mathcal{X}}(x) \cdot f_{\mathcal{Y}}(y)$ die Gleichverteilung auf dem kartesischen Produkt $G = [\alpha_1, \beta_1] \times [\alpha_2, \beta_2]$ ist.

Diese Funktion tritt oft auf als die Verteilung der Zeitspanne bis zum Eintreten eines Ereignisses.⁸⁸

Die momenterzeugende Funktion einer Exponentialverteilung ist

$$\phi(t) = E[e^{t\mathcal{X}}] = \int_0^\infty e^{tx} \lambda e^{-\lambda x} dx = \lambda \int_0^\infty e^{-(\lambda-t)x} dx = \frac{\lambda}{\lambda-t}, \quad t < \lambda.$$

Durch Differenzieren erhält man

$$\begin{aligned} \phi'(t) &= \frac{\lambda}{(\lambda-t)^2} \\ \phi''(t) &= \frac{2\lambda}{(\lambda-t)^3}, \end{aligned}$$

was impliziert, dass $E[\mathcal{X}] = \phi'(0) = 1/\lambda$, und $Var(\mathcal{X}) = \phi''(0) - 1/\lambda^2 = 1/\lambda^2$.

Die Schlüsseleigenschaft der Exponentialverteilung ist, dass sie *gedächtnislos* ist; d.h. für alle $s, t > 0$ gilt, dass $P[\mathcal{X} > s+t | \mathcal{X} > t] = P[\mathcal{X} > s]$.⁸⁹ In dem wir die Definition der bedingten Wahrscheinlichkeit anwenden, können wir die obige Gleichung umschreiben als $\frac{P[\mathcal{X} > s+t, \mathcal{X} > t]}{P[\mathcal{X} > t]} = P[\mathcal{X} > s]$, oder äquivalent dazu als

$$P[\mathcal{X} > s+t] = P[\mathcal{X} > t]P[\mathcal{X} > s].$$

Für eine exponentiell verteilte Zufallsvariable \mathcal{X} ist diese Gleichung erfüllt, weil $e^{-\lambda(s+t)} = e^{-\lambda s} e^{-\lambda t}$ ist. Umgekehrt ist jede stetige Verteilung, die diese Eigenschaft hat, eine Exponentialverteilung.

Beispiel 56. In einer Werkstatt stehen drei Maschinen, deren Laufzeit bis zum nächsten Ausfall exponentiell verteilt ist mit der Rate λ . Zwei von ihnen werden gleichzeitig benutzt, bis eine ausfällt, die dann durch die ungenutzte Maschine ersetzt wird. Bezeichnen wir diese Maschine als U und die noch laufende Maschine als L. Wie groß ist die Wahrscheinlichkeit, dass die nächste Maschine, die ausfällt, U ist?

Lösung. Die Laufzeit bis zum nächsten Ausfall ist exponentiell verteilt mit derselben Rate. Daher ist die Laufzeit von U genauso verteilt ist wie die Laufzeit von L seit dem Start von U. Daher ist die Wahrscheinlichkeit, dass U vor L ausfällt genauso groß wie die, dass L vor U ausfällt. Also sind beide Wahrscheinlichkeiten genau 0,5.

Angenommen ein System besteht aus n Komponenten und für jede ist die Zeit bis zum Ausfall exponentiell verteilt mit Rate λ_i .

Satz 57. Seien $\mathcal{X}_1, \dots, \mathcal{X}_n$ unabhängige exponentielle Zufallsvariablen mit Parametern $\lambda_1, \dots, \lambda_n$. Dann ist die Variable $\min\{\mathcal{X}_1, \dots, \mathcal{X}_n\}$ exponentiell verteilt mit Rate $\sum_{i=1}^n \lambda_i$.

Beweis. Der Satz folgt, weil⁹⁰

$$\begin{aligned} P[\min\{\mathcal{X}_1, \dots, \mathcal{X}_n\} > x] &= P[\mathcal{X}_1 > x, \dots, \mathcal{X}_n > x] = \prod_{i=1}^n P[\mathcal{X}_i > x] \\ &= \prod_{i=1}^n e^{-\lambda_i x} = e^{-\sum_{i=1}^n \lambda_i x}. \quad \square \end{aligned}$$

Ist \mathcal{X} exponentiell verteilt mit Rate λ , dann ist die Verteilungsfunktion

$$F(x) = P[\mathcal{X} \leq x] = 1 - e^{-\lambda x}.$$

⁸⁸ Zum Beispiel ist die Zeit, bis es regnet oder bis Sie eine neue E-Mail erhalten, näherungsweise exponentiell verteilt.

⁸⁹ Dies bedeutet, dass, falls das Ereignis zum Zeitpunkt t nicht eingetreten ist, die Wahrscheinlichkeit, mindestens s Zeiteinheiten länger warten zu müssen, genau dieselbe ist, wie die, von Anfang an s Zeiteinheiten warten zu müssen. Intuitiv startet die bedingte Wahrscheinlichkeit die Uhr neu.

⁹⁰ Ist \mathcal{X} exponentiell verteilt mit Rate λ , dann ist die Verteilungsfunktion $F(x) = P[\mathcal{X} \leq x] = 1 - e^{-\lambda x}$.

Zum Beispiel ist diese Tatsache nützlich, wenn wir wissen wollen, wann die erste Komponente eines Systems ausfallen wird. Des weiteren können wir damit für eine ZV \mathcal{X} , die mit der Rate λ verteilt ist, die Verteilung von $c\mathcal{X}$ herleiten. Es gilt nämlich

$$P[c\mathcal{X} > x] = P[\mathcal{X} > x/c] = e^{-\lambda x/c} = e^{-(\lambda/c)x},$$

was bedeutet, dass $c\mathcal{X}$ exponentiell verteilt ist mit der Rate λ/c .

Die Poisson-Verteilung

DIE MÖGLICHEN WERTE einer Zufallsvariablen \mathcal{X} , die Poisson-verteilt ist mit dem Parameter $\lambda > 0$, sind natürliche Zahlen, wobei für die Wahrscheinlichkeit, dass \mathcal{X} den Wert $i \in \mathbb{N}$ annimmt, gilt:⁹¹

$$P[\mathcal{X} = i] = e^{-\lambda} \frac{\lambda^i}{i!}.$$

Abbildung 12 zeigt die Wahrscheinlichkeitsfunktion der Poisson-Verteilung für $\lambda = 4$.⁹²

Wir bestimmen den Mittelwert und die Varianz einer Poisson-Zufallsvariablen, indem wir ihre momenterzeugende Funktion und ihre beiden ersten Ableitungen berechnen:

$$\begin{aligned} \phi(t) &= E[e^{t\mathcal{X}}] = \sum_{i=0}^{\infty} e^{ti} e^{-\lambda} (\lambda^i / i!) \\ &= e^{-\lambda} \sum_{i=0}^{\infty} (\lambda e^t)^i / i! \\ &= e^{-\lambda} e^{\lambda e^t} = \exp[\lambda(e^t - 1)] \\ \phi'(t) &= \lambda e^t \exp[\lambda(e^t - 1)] \\ \phi''(t) &= (\lambda e^t)^2 \exp[\lambda(e^t - 1)] + \lambda e^t \exp[\lambda(e^t - 1)] \end{aligned}$$

Von diesen Gleichungen leiten wir her, dass

$$\begin{aligned} E[\mathcal{X}] &= \phi'(0) = \lambda \\ \text{Var}(\mathcal{X}) &= \phi''(0) - (E[\mathcal{X}])^2 = \lambda^2 + \lambda - \lambda^2 = \lambda. \end{aligned}$$

Das heißt, der Mittelwert und die Varianz einer Poisson-Zufallsvariablen sind beide gleich dem Parameter λ .

Poisson-Zufallsvariablen bieten gute Näherungen für binomialverteilte Variablen mit den Parametern (n, p) , wenn n groß und p klein ist. Nehmen wir an, dass \mathcal{X} eine solche binomiale Zufallsvariable ist und sei $\lambda = np$. Dann ist

$$\begin{aligned} P[\mathcal{X} = i] &= \frac{n!}{(n-i)!i!} p^i (1-p)^{n-i} = \frac{n!}{(n-i)!i!} \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i} \\ &= \frac{n(n-1)\cdots(n-i+1)}{n^i} \frac{\lambda^i (1-\lambda/n)^n}{i! (1-\lambda/n)^i}. \end{aligned}$$

Unter der Voraussetzung, dass n groß und p klein ist, können wir folgern, dass⁹³

$$\left(1 - \frac{\lambda}{n}\right)^n \approx e^{-\lambda}; \quad \left(1 - \frac{\lambda}{n}\right)^i \approx 1; \quad \frac{n(n-1)\cdots(n-i+1)}{n^i} \approx 1.$$

Und somit folgt, dass $P[\mathcal{X} = i] \approx e^{-\lambda} \frac{\lambda^i}{i!}$.⁹⁴

⁹¹ Zuerst definiert von S. D. Poisson.

⁹² Dies ist tatsächlich eine Wahrscheinlichkeitsfunktion, weil für alle $x \in \mathbb{R}$ gilt $\sum_{i=1}^{\infty} \frac{x^i}{i!} = e^x$.

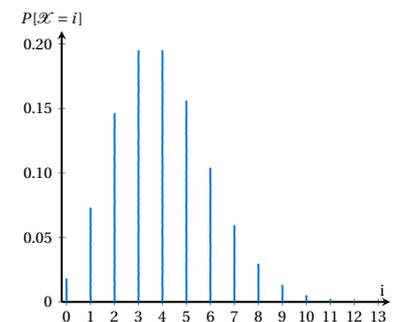


Abbildung 12: Wahrscheinlichkeitsfunktion der Poisson-Verteilung mit $\lambda = 4$.

⁹³ Wir erinnern uns, dass $\lim_{n \rightarrow \infty} (1 + x/n)^n = e^x$.

⁹⁴ Wenn man eine sehr große Anzahl n von unabhängigen Versuchen mit einer sehr geringen Erfolgswahrscheinlichkeit p durchführt, dann ist die Anzahl der Erfolge annähernd Poisson-verteilt mit $\lambda = np$.

Beispiel 58. Angenommen, auf der Autobahn zwischen Trient und Bozen ereignen sich im Durchschnitt drei Unfälle pro Woche. Wie hoch ist die Wahrscheinlichkeit, dass es in dieser Woche mindestens einen Unfall gibt?

Lösung. Sei \mathcal{X} die Anzahl der Unfälle in dieser Woche. Wir können ausgehen, dass eine große Zahl von Autos vorbeifährt, bei denen für jedes einzelne nur eine geringe Wahrscheinlichkeit besteht, einen Unfall zu haben. Daher sollte \mathcal{X} annähernd Poisson-verteilt sein. Daraus folgt,

$$P[\mathcal{X} \geq 1] = 1 - P[\mathcal{X} = 0] \approx 1 - e^{-3} \frac{3^0}{0!} = 1 - e^{-3} \approx 0.95. \quad \triangle$$

DIE POISSON-APPROXIMATION BLEIBT auch unter allgemeineren Bedingungen gültig. Wenn n unabhängige Versuche durchgeführt werden, jeder mit einer Erfolgswahrscheinlichkeit p_i , $1 \leq i \leq n$, dann ist, wenn n groß und jedes p_i klein ist, die Anzahl der erfolgreichen Versuche annähernd Poisson-verteilt mit Mittelwert $\sum_{i=1}^n p_i$. Dies gilt auch ohne die Unabhängigkeitsannahme, solange die Abhängigkeit nicht sehr stark ist.

Beispiel 59. Nehmen wir an, es gibt n Personen, die ihren Regenschirm am Eingang einer Bar zurücklassen und beim Verlassen zufällig einen der Schirme mitnehmen. Wenn \mathcal{X} die Anzahl der Personen bezeichnet, die ihren eigenen Schirm mitnehmen, dann ist \mathcal{X} für große n annähernd Poisson-verteilt mit Mittelwert 1. Intuitiv ist dies der Fall, weil wir \mathcal{X} als $\mathcal{X} = \sum_{i=1}^n \mathcal{X}_i$ ausdrücken können, wobei jedes \mathcal{X}_i die Indikatorvariable dafür ist, dass Person i ihren eigenen Schirm mitnimmt. Da es für jede Person gleich wahrscheinlich ist, einen der n Schirme zu wählen, ist $P[\mathcal{X} = i] = 1/n$.⁹⁵ Daraus schließen wir, dass $E[\mathcal{X}] = \sum_{i=1}^n E[\mathcal{X}_i] = n(1/n) = 1$.

Poisson-Zufallsvariablen sind außerdem *reproduktiv* (engl. *reproductive*), d.h. die Summe von zwei unabhängigen Poisson-Zufallsvariablen \mathcal{X} und \mathcal{Y} ist ebenfalls Poisson-verteilt. Um dies nachzuweisen, betrachten wir die momenterzeugende Funktion von $\mathcal{X} + \mathcal{Y}$ ⁹⁶

$$\begin{aligned} E[e^{t(\mathcal{X} + \mathcal{Y})}] &= E[e^{t\mathcal{X}} e^{t\mathcal{Y}}] = E[e^{t\mathcal{X}}] E[e^{t\mathcal{Y}}] \\ &= \exp(\lambda_{\mathcal{X}}(e^t - 1)) \exp(\lambda_{\mathcal{Y}}(e^t - 1)) = \exp((\lambda_{\mathcal{X}} + \lambda_{\mathcal{Y}})(e^t - 1)). \end{aligned}$$

Da dies die MEF einer Poisson-Variablen mit Mittelwert $\lambda_{\mathcal{X}} + \lambda_{\mathcal{Y}}$ ist, folgt daraus, dass $\mathcal{X} + \mathcal{Y}$ tatsächlich eine Poisson-Variable ist.⁹⁷

Beispiel 60. Die Anzahl der Kunden, die pro Stunde in eine Bar kommen, sei Poisson-verteilt mit einem Mittelwert von 4. Wie groß ist die Wahrscheinlichkeit, dass in einem Zeitraum von 2 Stunden nicht mehr als 3 Kunden kommen?

Lösung. Sei \mathcal{X}_i , $i = 1, 2$ die Anzahl der Kunden in der i -ten Stunde. Unter der Annahme, dass \mathcal{X}_1 and \mathcal{X}_2 unabhängig sind, ist $\mathcal{X} + \mathcal{X}_2$ Poisson-verteilt mit Mittelwert 8. Also ist

$$P[\mathcal{X}_1 + \mathcal{X}_2 \leq 3] = \sum_{n=0}^3 e^{-8} \frac{8^n}{n!} = 0.423. \quad \triangle$$

⁹⁵ Das heißt, wir können das Experiment näherungsweise durch eine Binomialverteilung mit der Wahrscheinlichkeit $1/n$ modellieren.

⁹⁶ Wir nehmen an, dass die Mittelwerte von \mathcal{X} und \mathcal{Y} gleich $\lambda_{\mathcal{X}}$ und $\lambda_{\mathcal{Y}}$ sind.

⁹⁷ Es sei daran erinnert, dass es eine Eins-zu-Eins-Beziehung zwischen momenterzeugenden Funktionen und Verteilungen gibt.

BETRACHTEN WIR NUN EIN SZENARIO, in dem eine zufällige Anzahl N von Ereignissen eintritt, von denen jedes, unabhängig von anderen, vom Typ 1 oder vom Typ 2 ist, mit Wahrscheinlichkeiten p bzw. $1 - p$.⁹⁸ Sei N_i , $i = 1, 2$ die Anzahl der beobachteten Ereignisse des Typs i . Wenn N Poisson-verteilt ist mit Mittelwert λ , dann ist die gemeinsame Wahrscheinlichkeitsfunktion von N_1 und N_2

$$\begin{aligned} P[N_1 = n, N_2 = m] &= P[N_1 = n, N_2 = m, N = n + m] \\ &= P[N_1 = n, N_2 = m \mid N = n + m] P[N = n + m] \\ &= P[N_1 = n, N_2 = m \mid N = n + m] e^{-\lambda} \frac{\lambda^{n+m}}{(n+m)!}. \end{aligned}$$

Nach Annahme ist jedes der $n + m$ Ereignisse unabhängig von den anderen mit der Wahrscheinlichkeit p vom Typ 1. Daher ist die Wahrscheinlichkeit, dass es genau n Ereignisse vom Typ 1 gibt, gleich der Wahrscheinlichkeit, dass eine mit Parametern $(n + m, p)$ binomialverteilte ZV den Wert n annimmt, das heißt,

$$\begin{aligned} P[N_1 = n, N_2 = m] &= \frac{(n+m)!}{n!m!} p^n (1-p)^m e^{-\lambda} \frac{\lambda^{n+m}}{(n+m)!} \\ &= e^{-\lambda p} \frac{(\lambda p)^n}{n!} e^{-\lambda(1-p)} \frac{(\lambda(1-p))^m}{m!}. \end{aligned}$$

Die Wahrscheinlichkeitsfunktion N_1 ist^o

$$\begin{aligned} P[N_1 = n] &= \sum_{m=0}^{\infty} P[N_1 = n, N_2 = m] \\ &= e^{-\lambda p} \frac{(\lambda p)^n}{n!} \sum_{m=0}^{\infty} e^{-\lambda(1-p)} \frac{(\lambda(1-p))^m}{m!} = e^{-\lambda p} \frac{(\lambda p)^n}{n!}. \end{aligned}$$

Daher ist N_1 eine Poisson-ZV mit dem Mittelwert λp und N_2 ist ebenfalls eine Poisson-ZV mit dem Mittelwert $\lambda(1-p)$. Außerdem sind diese beiden Variablen unabhängig.

Dieses Resultat kann auf den Fall verallgemeinert werden, dass es r verschiedene Kategorien gibt und jedes Ergebnis mit Wahrscheinlichkeit p_1, \dots, p_r in eine dieser r Kategorien fällt. Das heißt, die Anzahlen der Ergebnisse vom Typ i , $1 \leq i \leq r$, sind unabhängige Poisson-Zufallsvariablen mit Mittelwert λp_i .

UM DIE WAHRSCHEINLICHKEITSFUNKTION einer Poisson-Zufallsvariablen \mathcal{X} mit Mittelwert λ zu berechnen, nutzen wir aus, dass gilt

$$\frac{P[\mathcal{X} = i + 1]}{P[\mathcal{X} = i]} = \frac{e^{-\lambda} \lambda^{i+1} / (i+1)!}{e^{-\lambda} \lambda^i / i!} = \frac{\lambda}{i+1}.$$

Ausgehend von $P[\mathcal{X} = 0] = e^{-\lambda}$, können wir diese Gleichung verwenden, um nacheinander für alle aufeinanderfolgenden Werte i die Wahrscheinlichkeiten $P[\mathcal{X} = i]$ zu berechnen.⁹⁹

⁹⁸ Ein Beispiel: Die Anzahl der Kunden N in einer Bar ist eine Zufallsvariable; ein Kunde ist mit der Wahrscheinlichkeit p männlich und mit der Wahrscheinlichkeit $1 - p$ weiblich.

^o Es sei daran erinnert, dass $\sum_{i=1}^{\infty} \frac{x^i}{i!} = e^x$.

⁹⁹ $P[\mathcal{X} = i + 1] = (\lambda / (i + 1)) P[\mathcal{X} = i]$.

Normalverteilung

EINE ZUFALLSVARIABLE IST *normalverteilt* (engl. *normally distributed*) mit den Parametern μ und σ^2 (geschrieben als $\mathcal{X} \sim \mathcal{N}(\mu, \sigma^2)$), wenn ihre Dichte die Form hat:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}.$$

Der Graph dieser Funktion ist eine glockenförmige symmetrische Kurve, die ihren maximalen Wert $1/\sqrt{2\pi}\sigma$ an der Stelle μ hat (siehe Abbildung 13). Figure 13).

Die Bedeutung der Normalverteilung und ihre Verwendung in der Statistik ergibt sich aus dem *zentralen Grenzwertsatz* (*central limit theorem*), der kurz gesagt besagt, dass Zufallsphänomene bei einer ausreichend großen Anzahl von Beobachtungen dazu neigen, sich einer Normalverteilung anzunähern.¹⁰⁰ Beispiele, bei denen dies empirisch beobachtet werden kann, sind die Körpergröße von Menschen oder der Messfehler bei physikalischen Größen.

Die Parameter μ und σ^2 einer Verteilung $\mathcal{N}(\mu, \sigma^2)$ sind der Mittelwert beziehungsweise die Varianz. Wir zeigen zunächst, dass $E[\mathcal{X} - \mu] = 0$,¹⁰¹ indem wir eine neue Variable $y = (x - \mu)/\sigma$ definieren.¹⁰²

$$\begin{aligned} E[\mathcal{X} - \mu] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{(x - \mu)}{\sigma} e^{-(x-\mu)^2/2\sigma^2} dx \\ &= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y e^{-y^2/2} dy \quad (*) \\ &= \frac{\sigma}{\sqrt{2\pi}} \left[-e^{-y^2/2} \right]_{-\infty}^{\infty} = 0. \end{aligned}$$

Um die Varianz zu berechnen, verwenden wir $u = y$ und $\frac{dv}{dy} = y e^{-y^2/2}$ und sehen, dass¹⁰³

$$\int_{-\infty}^{\infty} y^2 e^{-y^2/2} dy = \left[-y e^{-y^2/2} \right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-y^2/2} dy = \int_{-\infty}^{\infty} e^{-y^2/2} dy.$$

Dann folgt

$$\begin{aligned} \text{Var}(\mathcal{X}) &= E[(\mathcal{X} - \mu)^2] = \int_{-\infty}^{\infty} \frac{(x - \mu)^2}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} dx \\ &= \sigma^2 \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} y^2 e^{-y^2/2} dy = \sigma^2 \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy = \sigma^2, \end{aligned}$$

wobei die letzte Gleichung sich daraus ergibt, dass $\frac{1}{\sqrt{2\pi}} e^{-y^2/2}$ die Dichtefunktion einer Normalverteilung (mit den Parametern $\mu = 0$ und $\sigma = 1$) ist; daher muss das Integral über alle reellen Zahlen gleich 1 sein.

Eine wichtige Eigenschaft der Normalverteilung ist, dass eine lineare Transformation $a\mathcal{X} + b$ einer normalverteilten Zufallsvariablen \mathcal{X} ebenfalls normalverteilt ist.¹⁰⁴

Das bedeutet, dass die Variable $Z = \frac{\mathcal{X} - \mu}{\sigma}$ in der Tat eine normalverteilte Zufallsvariable mit Mittelwert 0 und Varianz 1 ist. Diese wird standardnormalverteilte ZV genannt und ihre Verteilung wird mit Φ bezeichnet.¹⁰⁵ Wir können Aussagen über \mathcal{X} übersetzen in Aussagen über Z . Wissen wir

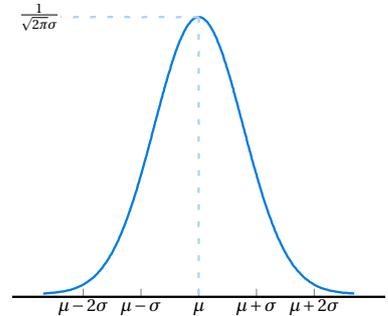


Abbildung 13: Die Dichte der Normalverteilung.

¹⁰⁰ Dies wird im nächsten Kapitel ausführlicher behandelt.

¹⁰¹ Es sei an die Regel $E[\mathcal{X} + b] = E[\mathcal{X}] + b$ erinnert.

¹⁰² Wir wenden die Substitutionsregel an, die $\sigma dy = dx$ liefert.

¹⁰³ Wir erinnern an die Formel der *partiellen Integration*: $\int u \frac{dv}{dx} dx = uv - \int v \frac{du}{dx} dx$.

¹⁰⁴ Wie schon bekannt, mit Mittelwert $a\mu + b$ und Varianz $a^2\sigma^2$.

¹⁰⁵ Das heißt, $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy$.

zum Beispiel, dass $\mathcal{X} < b$ genau dann gilt, wenn $\frac{\mathcal{X}-\mu}{\sigma} < \frac{b-\mu}{\sigma}$ ist, erhalten wir, dass

$$P[\mathcal{X} < b] = P\left[\frac{\mathcal{X}-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right] = \Phi\left(\frac{b-\mu}{\sigma}\right).$$

Mit anderen Worten: Für den Umgang mit Normalverteilungen reicht es aus, die Werte von Φ zu kennen. Es gibt viele Softwaresysteme und Tabellen, die diese Werte liefern¹⁰⁶. Diese Tabellen enthalten oft nur Φ -Werte für positive Werte von x , aber da Z symmetrisch ist, können negative Werte von x leicht berücksichtigt werden. Zum Beispiel ist $P[Z < -x] = P[X > x] = 1 - \Phi(x)$. Wenn $x = 1$ ist, ergibt sich also $P[Z < -1] = 1 - \Phi(1) = 1 - 0.8413 = 0.1587$.

Beispiel 61. Die Übertragung von Daten ist in der Regel Störungen durch Kanalrauschen unterworfen. Um die Möglichkeit von Fehlern zu verringern, kodieren wir binäre Nachrichten mit den Werten $-2, 2$, die für die ursprüngliche 0 bzw. 1 stehen. Wenn wir eine Nachricht $x \in \{-2, 2\}$ übermitteln, erreicht sie aufgrund des Rauschens ihr Ziel als $R = x + N$. Der Empfänger dekodiert die Nachricht, indem er, wenn $R \geq 0.5$ ist, folgert, dass die gesendete Nachricht 1 war, andernfalls 0 . Wir gehen davon aus, dass N standardnormalverteilt ist.

Die Wahrscheinlichkeiten dafür, dass ein falsches Signal empfangen wird, sind aufgrund des Rauschens

$$\begin{aligned} P[R < 0.5 \mid \text{Signal} = 1] &= P[N < -1.5] = 1 - \Phi(1.5) = 0.0668 \\ P[R \geq 0.5 \mid \text{Signal} = 0] &= P[N \geq 2.5] = 1 - \Phi(2.5) = 0.0062. \quad \triangle \end{aligned}$$

DIE MOMENTERZEUGENDE FUNKTION einer standardnormalverteilten Zufallsvariablen Z ist

$$\begin{aligned} E[e^{tZ}] &= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(x^2-2tx)/2} dx \\ &= e^{-t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(x-t)^2/2} dx = e^{-t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy = e^{-t^2/2}. \end{aligned}$$

Eine normalverteilte Zufallsvariable \mathcal{X} mit Mittelwert μ und Varianz σ^2 kann als $\sigma Z + \mu$ dargestellt werden. Ihre momenterzeugende Funktion ist also

$$E[e^{t\mathcal{X}}] = E[e^{t\mu + t\sigma Z}] = e^{t\mu} E[e^{t\sigma Z}] = e^{t\mu} e^{-(t\sigma)^2/2} = e^{t\mu - \sigma^2 t^2/2}.$$

Wir können diese momenterzeugende Funktion verwenden, um zu beweisen, dass die Summe von unabhängigen normalverteilten Zufallsvariablen eine normalverteilte Zufallsvariable ist.^o

Beispiel 62. Gegeben, eine Person hat zwei (erwachsene) Söhne, und angenommen, die Größen der einzelnen Kinder sind unabhängig, wie groß ist dann die Wahrscheinlichkeit, dass der ältere Sohn mindestens 2 cm größer ist als der jüngere?

Lösung. Seien \mathcal{X}_1 und \mathcal{X}_2 die Größen des ersten bzw. zweiten Kindes. Da $-\mathcal{X}_2$ eine normalverteilte ZV ist mit Mittelwert $-177,6$ und Varianz $4^2 = 16$, ist auch $\mathcal{X}_1 - \mathcal{X}_2$ normalverteilt mit Mittelwert 0 und Varianz 32 .

¹⁰⁶ Siehe OLE für eine solche Tabelle. Wir werden auch erfahren, wie man sie aus R erhalten kann.

^oVersuchen Sie es selbst. Was sind der Mittelwert und die Varianz der Summe von zwei unabhängigen standardnormalverteilten ZVen?

Dann gilt:

$$\begin{aligned}
 P[\mathcal{X}_1 > \mathcal{X}_2 + 2] &= P[\mathcal{X}_1 - \mathcal{X}_2 > 2] = P\left[\frac{\mathcal{X}_1 - \mathcal{X}_2}{\sqrt{32}} > \frac{2}{\sqrt{32}}\right] \\
 &= P[Z > 0.3536] \approx 1 - 0.6368 = 0.3632. \quad \triangle
 \end{aligned}$$

Für eine beliebige Zahl $\alpha \in (0, 1)$, sei z_α der Wert, für den $P[Z \geq z_\alpha] = \alpha$ gilt (siehe Abbildung 14). Zum Beispiel ist $z_{0.05} = 1.645$ und $z_{0.01} = 2.33$. Der Wert z_α ist das $100(1 - \alpha)\%$ Perzentil (engl. *percentile*) von Z .¹⁰⁷

Chi-Quadrat-verteilte Variablen

Die Summe von n unabhängigen standardnormalverteilten ZVen $\mathcal{X} = \sum_{i=1}^n Z_i^2$ ist eine *Chi-Quadrat-verteilte ZV mit n Freiheitsgraden* (engl. *chi-square RV with n degrees of freedom*). Dies wird als $\mathcal{X} \sim \chi_n^2$ notiert.

Die Chi-Quadrat-Verteilung ist *additiv* (engl. *additive*) in dem Sinne, dass für unabhängige Chi-Quadrat-verteilte ZVen \mathcal{X} und \mathcal{Y} auch $\mathcal{X} + \mathcal{Y}$ Chi-Quadrat-verteilt ist mit $n+m$ Freiheitsgraden. Dies gilt, weil $\mathcal{X} + \mathcal{Y}$ die Summe von $n + m$ quadrierten unabhängigen standardnormalverteilten ZVen ist.

Wenn \mathcal{X} Chi-Quadrat-verteilt ist mit n Freiheitsgraden, dann ist für jedes $\alpha \in (0, 1)$ die Zahl $\chi_{\alpha,n}^2$ der Wert, bei dem $P[\mathcal{X} \geq \chi_{\alpha,n}^2] = \alpha$ ist.

Eine Anwendung, bei der die Chi-Quadrat-Verteilung von Nutzen sein kann, ist der Umgang mit Messfehlern in mehreren Dimensionen. Wenn Sensoren die Position eines Objekts in jeder Dimension messen, dann ist das Quadrat des Abstands zwischen der gemessenen und der tatsächlichen Position Chi-Quadrat-verteilt, falls die Messfehler in jeder Dimension normalverteilt und die Fehler in den verschiedenen Dimensionen unabhängig sind.¹⁰⁸

The t -Verteilung

WENN EINE STANDARDNORMALVERTEILTE Zufallsvariable Z und eine Chi-Quadrat-Zufallsvariable χ_n^2 mit n Freiheitsgraden voneinander unabhängig sind, dann hat die ZV

$$T_n := \frac{Z}{\sqrt{\chi_n^2/n}}$$

eine *t -Verteilung mit n Freiheitsgraden* (engl. *t -distribution with n degrees of freedom*). Abbildung 15 zeigt die Dichtefunktion von T_n für verschiedene Freiheitsgrade.

Die t -Dichte ist symmetrisch zur Achse durch 0 und nähert sich der Dichte der Standardnormalverteilung mit wachsendem n an. Es sei daran erinnert, dass χ_n^2 die Summe der Quadrate von n unabhängigen standardnormalverteilten ZVen ist. Aus dem schwachen Gesetz der großen Zahlen folgt, dass χ_n^2/n mit wachsendem n mit der Wahrscheinlichkeit 1 gegen $E[Z_i^2] = 1$ konvergiert. Für ein ausreichend großes n hat $T_n = Z/\sqrt{\chi_n^2/n}$ also ungefähr die gleiche Verteilung wie Z (siehe Abbildung 16).

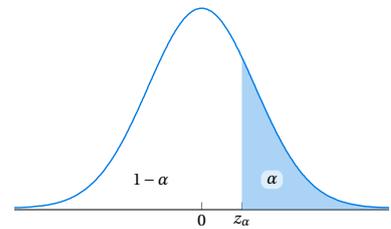


Abbildung 14: $P[Z > z_\alpha] = \alpha$.

¹⁰⁷ Der Name beschreibt, dass in $100(1 - \alpha)$ Prozent der Fälle der Wert von Z unter z_α fallen wird.

¹⁰⁸ Wir werden die Chi-Quadrat-Verteilung später verwenden, um Parameter einer anderen Verteilung zu schätzen.

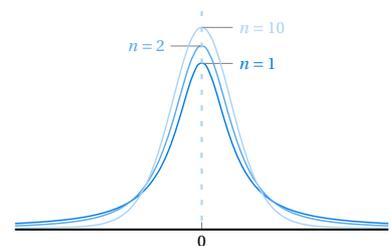


Abbildung 15: Die Dichtefunktion von T_n für $n = 1, 2, 10$.

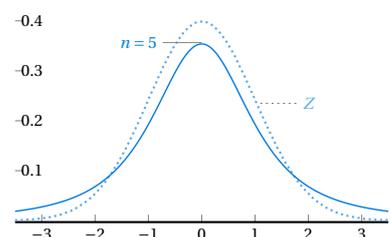


Abbildung 16: Die Dichtefunktionen von T_5 (durchgezogen) und Z (gepunktet).

Der Mittelwert und die Varianz von T_n sind für $n > 1$ and $m > 2$,

$$E[T_n] = 0$$

$$\text{Var}(T_m) = \frac{m}{m-2}.$$

Man beachte, dass die Varianz von T_n von oben gegen 1 konvergiert, wenn n gegen unendlich geht.

Gegeben ein $\alpha \in (0, 1)$, sei $t_{\alpha,n}$ die Zahl, für die $P[T_n \geq t_{\alpha,n}] = \alpha$ ist. Da t symmetrisch um 0 ist, folgt daraus, dass

$$\alpha = P[-T_n \geq t_{\alpha,n}] = P[T_n \leq -t_{\alpha,n}] = 1 - P[T_n > t_{\alpha,n}].$$

Dies bedeutet, dass $P[T_n \geq t_{\alpha,n}] = 1 - \alpha$ ist, oder, äquivalent, $-t_{\alpha,n} = t_{1-\alpha,n}$.
Siehe Abbildung 17.

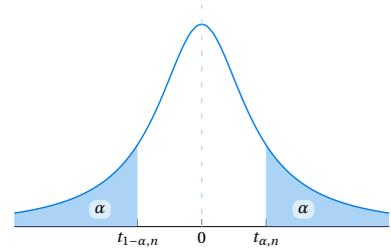


Abbildung 17: Die Flächen vor $-t_{\alpha,n}$ und nach $t_{\alpha,n}$ haben die Größe α .