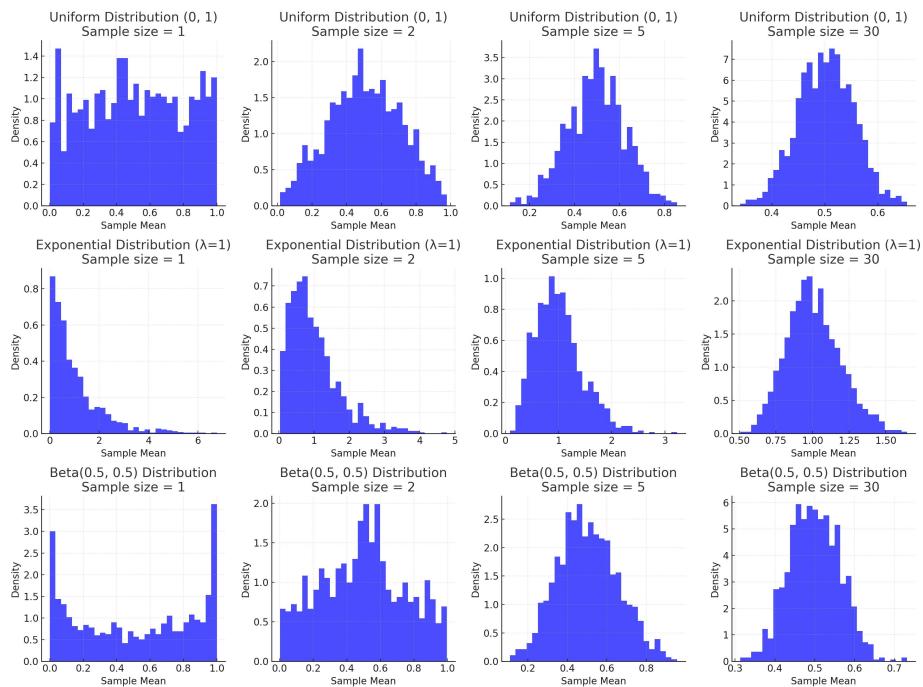


Grundlagen der Stochastik: Wahrscheinlichkeitsräume, Zufallsvariablen und zentrale Sätze

Max Renesse & ChatGPT

Stand 22. Januar 2025



Inhaltsverzeichnis

1 Einführung in die Stochastik	3
2 Wahrscheinlichkeitsräume	4
2.1 Wahrscheinlichkeitsräume – Definition	4
2.1.1 Beispiele	5
2.2 Exkurs: Elementare Kombinatorik	6
2.2.1 Bedeutung der Kombinatorik für Laplace-Experimente	6
2.2.2 Grundformeln der elementaren Kombinatorik	7
3 Bedingte Wahrscheinlichkeiten etc.	10
3.1 Unabhängige Ereignisse	10
3.1.1 Unabhängigkeit von zwei Ereignissen	10
3.1.2 Unabhängigkeit von mehr als zwei Ereignissen	10
3.2 Bedingte Wahrscheinlichkeit	12
3.3 Satz von der totalen Wahrscheinlichkeit	12
3.4 Satz von Bayes	13
3.5 Zweistufige Zufallsexperimente, Baumdiagramme und der Bayes'sche Satz .	14
4 Zufallsvariablen und ihre Verteilungen	16
4.1 Verteilung einer (diskreten) Zufallsvariablen	17
4.2 Beispiel: Zufallsexperiment mit Kugeln in rhombischer Anordnung	18
4.3 Beispiel: Binomialverteilung	22
4.4 Tschebyschev-Ungleichung	22
4.4.1 Markov-Ungleichung	23
4.4.2 Beweis der Tschebyschev-Ungleichung	24
5 Unabhängigkeit von Zufallsvariablen	25
5.1 Faktorisierung des Erwartungswerts bei Unabhängigkeit	25
5.2 Kovarianz und Korrelation	26
5.3 Satz von Bienaym� zur Additivit�t der Varianz	29
6 Gesetz der gro�en Zahlen (GGZ)	30
6.1 Satz (Schwaches Gesetz der gro�en Zahlen)	30
6.2 Anwendungen	31
6.2.1 Monte-Carlo-Integration	31
6.2.2 Beispiel: Berechnung von π durch wiederholten Münzwurf	33

7 Mehr zu Verteilungen von Zufallsvariablen	35
7.1 Vektorwertige Zufallsvariablen	35
7.2 Gemeinsame Verteilung	36
7.2.1 Spezialfall $k = 2$: Kreuztabellen	36
7.2.2 Produktverteilung und Stochastische Unabhängigkeit	38
7.3 Diskrete Verteilungen als Wahrscheinlichkeitsmaße auf \mathbb{R}^d	39
7.4 Verteilungsfunktion einer Zufallsvariablen	39
7.5 Verteilungsfunktion einer diskreten vektorwertigen Zufallsvariablen	41
7.6 Beispiel: Empirische Verteilung von Stichproben	43
7.7 Zufallsvariablen mit Dichtefunktion	43
7.7.1 Kontinuierliche vs. Diskrete Zufallsvariablen	43
7.7.2 Beispiel: Gleichverteilung auf $[0, 1] \subset \mathbb{R}$	43
7.7.3 Kontinuierliche Zufallsvariablen mit Dichtefunktion	44
7.7.4 Die Verteilungsfunktion und ihr Zusammenhang mit der Dichtefunktion	45
7.7.5 Beispiele	46
7.8 Konstruktion bzw. Simulation einer Zufallsvariablen zu gegebener Verteilungsfunktion	49
8 Der Zentrale Grenzwertsatz	52
8.1 Aussage	52
8.2 Anschauliche Bedeutung	52
8.3 Simulationsbeispiele zur Illustration des Zentralen Grenzwertsatzes	53
8.4 Präzise Formulierung vom Zentralen Grenzwertsatz	55
8.5 Beweis vom Zentralen Grenzwertsatz (Stein'sche Methode)	55
8.5.1 1. Schritt: Konvergenz von Erwartungswerten	55
8.5.2 2. Schritt: Konvergenz von Wahrscheinlichkeiten	57
8.6 Anwendungsbeispiel: Rückstellungssumme bei einer Kfz-Versicherung	58
8.7 Nachtrag: Poisson'scher Grenzwertsatz und Anwendung im Mobilfunk	60
9 Elemente der schließenden Statistik	63
9.1 Wahrscheinlichkeitstheorie und Statistik als Teilgebiete der Stochastik	63
9.2 Beispiel: Das Schätzproblem für den Erfolgsparameter einer Münze	64
9.2.1 Verbindung zum Schwachen Gesetz der Großen Zahlen	65
9.2.2 Berechnung der benötigten Anzahl von Würfen bei gegebener Fehlertoleranz und maximal zugelassener Irrtumswahrscheinlichkeit	65
9.3 Allgemeine Schätzprobleme	68
9.3.1 Kenngrößen einer Verteilung	68
9.3.2 Schätzer für eine Kenngröße	68
9.3.3 Erwartungstreue Schätzer für Mittelwert, Varianz und Schiefe	68
9.3.4 Mittlerer quadratischer Fehler eines Schätzers	70
9.3.5 Fehlerzerlegung des mittleren quadratischen Fehlers	70
9.3.6 Beispiel: Vergleich zweier Schätzer im 'Gameshow'-Problem	71

Kapitel 1

Einführung in die Stochastik

Die Stochastik ist die Wissenschaft vom Zufall und beschäftigt sich mit der Modellierung und Analyse zufälliger Ereignisse. Sie liefert die Grundlage für viele Anwendungen, von Statistik über Versicherungsmathematik bis hin zur Datenanalyse. Die zentralen Begriffe sind Wahrscheinlichkeitsräume, Zufallsvariablen und Wahrscheinlichkeitsverteilungen.

Kapitel 2

Wahrscheinlichkeitsräume

2.1 Wahrscheinlichkeitsräume – Definition

Ein Wahrscheinlichkeitsraum (Ω, \mathcal{F}, P) besteht aus:

- Ω : Ergebnismenge, die alle möglichen Ausgänge eines Zufallsexperiments umfasst bzw. parametrisiert¹.
- $\mathcal{F} \subset 2^\Omega$ Menge der *Ereignisse* $E \in \mathcal{F}$
- P : Wahrscheinlichkeitsmaß, das jedem Ereignis $A \in \mathcal{F}$ eine Wahrscheinlichkeit $P(A) \in [0, 1]$ zuordnet, wobei $P(\Omega) = 1$ gilt.

Die Menge der Ereignisse \mathcal{F} soll dabei die leere Menge enthalten und stabil unter Komplementbildung und abzählbaren Vereinigung sein, d.h. es soll gelten, dass

- $\emptyset \in \mathcal{F}$
- $E^c = \Omega \setminus E \in \mathcal{F}$, falls $E \in \mathcal{F}$
- $\bigcup_{k=1}^{\infty} E_k \in \mathcal{F}$ falls $E_k \in \mathcal{F}$ für alle $k \in \mathbb{N}$

Man sagt, dass \mathcal{F} eine σ -Algebra ist. Unter einem Wahrscheinlichkeitsmaß P auf einer σ -Algebra verstehen wir dabei ist eine *abzählbar-additive* Funktion $P : \mathcal{F} \mapsto [0, 1]$ auf \mathcal{F} in dem Sinne, dass gelten soll

$$P\left(\bigcup_{k=1}^{\infty} E_k\right) = \sum_{k=1}^{\infty} P(E_k)$$

falls $E_k \cap E_l = \emptyset$ für alle $k \neq l$. Zudem wird als Normierung verlangt, dass $P(\Omega) = 1$.

Im Rahmen der *elementaren Stochastik* behandelt man meistens nur den Fall von *diskreten, d.h. abzählbaren Wahrscheinlichkeitsräumen*, d.h. wenn o.B.d.A. nach Wahl einer Nummerierung von Ω gilt, dass $\Omega \subset \mathbb{N}$. In diesem Fall wählt man man typischer Weise einfach $\mathcal{F} = 2^\Omega$ und verzichtet auf die explizite Nennung von \mathcal{F} . In diesem Fall ist

¹D.h. jedes ω symbolisiert einen grundsätzlich möglichen Ausgang des Zufallsexperiments, Ω ist die Zusammenfassung aller möglichen Ausgänge. Mit 'Ausgang' bzw. synonym 'Ergebnis' bezeichnen wir die resultierende Beobachtung bei einmaliger Durchführung des zugrundegelegten Zufallsexperiments.

P festgelegt durch Angabe der *Elementarwahrscheinlichkeiten* $\{P(\{\omega\}) \mid \omega \in \Omega\}$, denn für beliebige Ereignisse $E \in \mathcal{F}$ gilt dann wegen der abzählbaren Additivität von P , dass

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

Auf der rechten Seite haben wir von der vereinfachten Schreibweise $P(\omega)$ für die Wahrscheinlichkeit $P(E)$ von so genannten *Elementarereignissen* der Form $E = \{\omega\}$ gemacht.

2.1.1 Beispiele

1. Endlicher Laplace'scher Wahrscheinlichkeitsraum

Ein endlicher Laplace-Raum besteht aus einer endlichen Menge Ω mit $P(\omega) = \frac{1}{|\Omega|}$ für alle $\omega \in \Omega$. Er beschreibt ein Zufallsexperiment mit endlich vielen Ausgängen, die alle mit gleicher Wahrscheinlichkeit auftreten.

Beispiel:

Beim Würfeln ist $\Omega = \{1, 2, 3, 4, 5, 6\}$, und für einen fairen Würfel gilt $P(\{i\}) = \frac{1}{6}$ für alle $i \in \{1, \dots, 6\}$. Wir sprechen von einem Laplace'schen Wahrscheinlichkeitsraum mit Kardinalität $|\Omega| = 6$, bzw. von einem Laplace'schen Zufallsexperiment mit 6 gleichwahrscheinlichen Elementar-Ergebnissen.

2. Endlicher nicht-Laplace'scher Wahrscheinlichkeitsraum

Ein Wahrscheinlichkeitsraum ist *nicht-Laplace'sch*, wenn die Wahrscheinlichkeiten der Ergebnisse nicht gleichverteilt sind.

Beispiel:

Betrachten wir einen Beutel mit Kugeln in den Farben Rot, Blau und Grün, wobei die Wahrscheinlichkeiten ungleich verteilt sind:

$$\Omega = \{\text{Rot}, \text{Blau}, \text{Grün}\}.$$

Sei das Wahrscheinlichkeitsmaß P wie folgt definiert:

$$P(\text{Rot}) = 0.5, \quad P(\text{Blau}) = 0.3, \quad P(\text{Grün}) = 0.2.$$

Dieser Wahrscheinlichkeitsraum ist nicht-Laplace'sch, da die Ergebnisse unterschiedliche Wahrscheinlichkeiten haben.

Eigenschaften:

- Ω ist endlich.
- Die Wahrscheinlichkeiten summieren sich zu 1:

$$P(\text{Rot}) + P(\text{Blau}) + P(\text{Grün}) = 0.5 + 0.3 + 0.2 = 1.$$

3. Abzählbar unendlicher Wahrscheinlichkeitsraum

Ein Wahrscheinlichkeitsraum ist *abzählbar unendlich*, wenn die Ergebnismenge Ω abzählbar unendlich ist.

Beispiel:

Werfen wir eine faire Münze beliebig oft und zählen die Anzahl der Würfe bis zum ersten Mal Kopf erscheint. Die Ergebnismenge ist:

$$\Omega = \{1, 2, 3, \dots\},$$

wobei $k \in \Omega$ die Anzahl der Würfe bis zum ersten Kopf darstellt.

Das Wahrscheinlichkeitsmaß P ist definiert durch:

$$P(k) = \left(\frac{1}{2}\right)^k, \quad \text{für } k \in \mathbb{N}.$$

Eigenschaften:

- Ω ist abzählbar unendlich, da \mathbb{N} abzählbar unendlich ist.
- Die Wahrscheinlichkeiten summieren sich zu 1 (geometrische Reihe):

$$\sum_{k=1}^{\infty} P(k) = \sum_{k=1}^{\infty} \left(\frac{1}{2}\right)^k = \frac{\frac{1}{2}}{1 - \frac{1}{2}} = 1.$$

2.2 Exkurs: Elementare Kombinatorik

2.2.1 Bedeutung der Kombinatorik für Laplace-Experimente

In Laplace'schen Wahrscheinlichkeitsräumen, bei denen alle elementaren Ereignisse gleich wahrscheinlich sind, ist die Wahrscheinlichkeit eines Ereignisses A definiert als:

$$P(A) = \frac{|A|}{|\Omega|},$$

wobei $|A|$ die Anzahl der für A günstigen Ereignisse und $|\Omega|$ die Gesamtanzahl der möglichen Ereignisse ist. Die Kombinatorik ist hier essenziell, da sie ermöglicht, $|A|$ und $|\Omega|$ effizient zu berechnen.

Beispiele

1. Wurf eines Würfels: Beim Wurf eines Würfels beträgt die Wahrscheinlichkeit, eine gerade Zahl zu würfeln ($A = \{2, 4, 6\}$):

$$P(A) = \frac{|A|}{|\Omega|} = \frac{3}{6} = \frac{1}{2}.$$

2. Kombinationen beim Kartenziehen: Aus einem Kartendeck mit 52 Karten beträgt die Wahrscheinlichkeit, eine beliebige Herzkarde zu ziehen ($A = \{13\text{ Herzkarten}\}$):

$$P(A) = \frac{|A|}{|\Omega|} = \frac{13}{52} = \frac{1}{4}.$$

3. Kombinatorik bei Kombinationen: Beim Zufallsexperiment 'Auswahl von 3 Kugeln aus einer Urne mit 10 Kugeln' beträgt die Gesamtanzahl möglicher Kombinationen ($|\Omega| = C(10, 3)$):

$$C(10, 3) = \frac{10!}{3! \cdot (10-3)!} = 120.$$

Die Wahrscheinlichkeit, 3 blaue Kugeln zu ziehen, falls 4 von 10 Kugeln blau sind, beträgt ($A = \{3 \text{ aus } 4 \text{ blauen Kugeln}\}$):

$$P(A) = \frac{|A|}{|\Omega|} = \frac{C(4, 3)}{C(10, 3)} = \frac{\frac{4!}{3!(4-3)!}}{120} = \frac{4}{120} = \frac{1}{30}.$$

Fazit

Die Kombinatorik ist ein unverzichtbares Werkzeug in Laplace'schen Wahrscheinlichkeitsräumen, da sie die effiziente Bestimmung der Anzahl günstiger und möglicher Ereignisse erlaubt. Sie bildet die Grundlage für viele Wahrscheinlichkeitsberechnungen in theoretischen und praktischen Kontexten.

2.2.2 Grundformeln der elementaren Kombinatorik

Die Kombinatorik beschäftigt sich mit der Anzahl von Möglichkeiten, Objekte unter bestimmten Bedingungen anzurufen oder auszuwählen. Die wichtigsten Grundformeln sind:

1. Permutationen (Anordnung ohne Wiederholung)

Die Anzahl der Möglichkeiten, n verschiedene Objekte in einer bestimmten Reihenfolge anzurufen, ist:

$$P(n) = n! = n \cdot (n - 1) \cdot (n - 2) \cdots \cdot 1.$$

Beispiel:

Wie viele Möglichkeiten gibt es, 4 verschiedene Bücher in einer Reihe anzurufen?

$$P(4) = 4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24.$$

2. Permutationen mit Wiederholung

Wenn einige Objekte identisch sind, wird die Anzahl der Permutationen durch die Formel angepasst:

$$P(n; k_1, k_2, \dots, k_r) = \frac{n!}{k_1! \cdot k_2! \cdots k_r!},$$

wobei k_1, k_2, \dots, k_r die Häufigkeiten der identischen Objekte sind.

Beispiel:

Wie viele unterschiedliche Wörter können mit den Buchstaben von AABB gebildet werden?

$$P(4; 2, 2) = \frac{4!}{2! \cdot 2!} = \frac{24}{4} = 6.$$

3. Kombinationen ohne Wiederholung

Die Anzahl der Möglichkeiten, k Objekte aus n verschiedenen Objekten auszuwählen, wobei die Reihenfolge keine Rolle spielt, ist:

$$C(n, k) = \binom{n}{k} = \frac{n!}{k! \cdot (n - k)!}.$$

Beispiel:

Wie viele Möglichkeiten gibt es, 3 Kugeln aus einem Beutel mit 5 verschiedenen Kugeln auszuwählen?

$$C(5, 3) = \binom{5}{3} = \frac{5!}{3! \cdot (5-3)!} = \frac{120}{6 \cdot 2} = 10.$$

4. Kombinationen mit Wiederholung

Die Anzahl der Möglichkeiten, k Objekte aus n verschiedenen Objekten auszuwählen, wenn jedes Objekt mehrfach gewählt werden kann, ist:

$$C(n + k - 1, k) = \binom{n + k - 1}{k}.$$

Beispiel:

Wie viele Möglichkeiten gibt es, 3 Kugeln aus einem Beutel mit 2 verschiedenen Kugeln auszuwählen, wenn jede Kugel mehrfach gezogen werden kann?

$$C(2 + 3 - 1, 3) = \binom{4}{3} = \frac{4!}{3! \cdot (4-3)!} = 4.$$

5. Variationen (Anordnung mit Wiederholung)

Die Anzahl der Möglichkeiten, k Objekte aus n verschiedenen Objekten auszuwählen, wobei die Reihenfolge eine Rolle spielt und jedes Objekt mehrfach gewählt werden kann, ist:

$$V(n, k) = n^k.$$

Beispiel:

Wie viele dreistellige Zahlen können aus den Ziffern 0–9 gebildet werden, wenn Ziffern wiederholt werden dürfen?

$$V(10, 3) = 10^3 = 1000.$$

6. Variationen ohne Wiederholung

Die Anzahl der Möglichkeiten, k Objekte aus n verschiedenen Objekten auszuwählen, wobei die Reihenfolge eine Rolle spielt und kein Objekt mehrfach gewählt wird, ist:

$$V(n, k) = \frac{n!}{(n - k)!}.$$

Beispiel:

Wie viele Möglichkeiten gibt es, die ersten drei Plätze in einem Wettbewerb mit 8 Teilnehmern zu vergeben?

$$V(8, 3) = \frac{8!}{(8 - 3)!} = \frac{40320}{120} = 336.$$

Zusammenfassung der wichtigsten Formeln

- Permutationen ohne Wiederholung: $P(n) = n!$
- Permutationen mit Wiederholung: $P(n; k_1, k_2, \dots, k_r) = \frac{n!}{k_1! \cdot k_2! \cdots k_r!}$
- Kombinationen ohne Wiederholung: $C(n, k) = \binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}$
- Kombinationen mit Wiederholung: $C(n + k - 1, k) = \binom{n+k-1}{k}$
- Variationen ohne Wiederholung: $V(n, k) = \frac{n!}{(n-k)!}$
- Variationen mit Wiederholung: $V(n, k) = n^k$

Kapitel 3

Bedingte Wahrscheinlichkeiten etc.

3.1 Unabhängige Ereignisse

3.1.1 Unabhängigkeit von zwei Ereignissen

Zwei Ereignisse $A, B \subseteq \Omega$ in einem diskreten Wahrscheinlichkeitsraum (Ω, \mathcal{F}, P) heißen *unabhängig*, wenn gilt:

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

Unabhängigkeit bedeutet, dass das Eintreten von A keinen Einfluss auf die Wahrscheinlichkeit von B hat.

Beispiel:

Beim zweimaligen Würfeln eines fairen Würfels seien die Ereignisse:

$$A = \{\text{Erster Wurf zeigt eine } 6\}, \quad B = \{\text{Zweiter Wurf zeigt eine gerade Zahl}\}.$$

Da die Würfe unabhängig sind, gilt:

$$P(A) = \frac{1}{6}, \quad P(B) = \frac{3}{6} = \frac{1}{2}, \quad P(A \cap B) = P(A) \cdot P(B) = \frac{1}{6} \cdot \frac{1}{2} = \frac{1}{12}.$$

3.1.2 Unabhängigkeit von mehr als zwei Ereignissen

Eine Menge von Ereignissen $\{A_1, A_2, \dots, A_n\}$ heißt *stochastisch unabhängig*, wenn für jede Teilmenge $I \subseteq \{1, 2, \dots, n\}$ gilt:

$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i).$$

Beispiele: 1. Drei Münzwürfe mit $A_1 = \{\text{Erster Wurf zeigt Kopf}\}$, $A_2 = \{\text{Zweiter Wurf zeigt Kopf}\}$, $A_3 = \{\text{Dritter Wurf zeigt Kopf}\}$. Hier gilt:

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2) \cdot P(A_3) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}.$$

2. Sind $A_1 = \{\text{Erster Würfel zeigt eine } 6\}$, $A_2 = \{\text{Zweiter Würfel zeigt eine gerade Zahl}\}$, so sind diese Ereignisse unabhängig, da:

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2) = \frac{1}{6} \cdot \frac{1}{2} = \frac{1}{12}.$$

Für die Unabhängigkeit mehrerer Ereignisse muss dies für jede mögliche Kombination von Teilmengen gelten.

Bemerkung:

Unabhängigkeit von Paaren von Ereignissen (z. B. $P(A_1 \cap A_2) = P(A_1) \cdot P(A_2)$) impliziert nicht automatisch die Unabhängigkeit von mehr als zwei Ereignissen. Die Definition verlangt, dass jede Kombination unabhängig ist.

Beispiel:

Betrachten wir den Wahrscheinlichkeitsraum (Ω, P) mit:

$$\Omega = \{(0, 0), (0, 1), (1, 0), (1, 1)\},$$

und definiere ein Wahrscheinlichkeitsmaß:

$$P((0, 0)) = P((1, 1)) = \frac{1}{4}, \quad P((0, 1)) = P((1, 0)) = \frac{1}{4}.$$

Definiere die Ereignisse:

$$A = \{(0, 0), (0, 1)\}, \quad B = \{(0, 0), (1, 0)\}, \quad C = \{(0, 0), (1, 1)\}.$$

Prüfung der paarweisen Unabhängigkeit

Berechne zunächst die Wahrscheinlichkeiten der einzelnen Ereignisse:

$$\begin{aligned} P(A) &= P(\{(0, 0), (0, 1)\}) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}, \\ P(B) &= P(\{(0, 0), (1, 0)\}) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}, \\ P(C) &= P(\{(0, 0), (1, 1)\}) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}. \end{aligned}$$

Berechne die paarweisen Schnittmengen:

$$\begin{aligned} P(A \cap B) &= P(\{(0, 0)\}) = \frac{1}{4}, \quad P(A) \cdot P(B) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}, \\ P(A \cap C) &= P(\{(0, 0)\}) = \frac{1}{4}, \quad P(A) \cdot P(C) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}, \\ P(B \cap C) &= P(\{(0, 0)\}) = \frac{1}{4}, \quad P(B) \cdot P(C) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}. \end{aligned}$$

Da die Schnittwahrscheinlichkeiten den Produkten der Einzelwahrscheinlichkeiten entsprechen, sind A, B und C paarweise unabhängig.

Prüfung der gemeinsamen Unabhängigkeit

Berechne die Wahrscheinlichkeit der gemeinsamen Schnittmenge:

$$P(A \cap B \cap C) = P(\{(0, 0)\}) = \frac{1}{4}.$$

Vergleiche mit dem Produkt der Wahrscheinlichkeiten:

$$P(A) \cdot P(B) \cdot P(C) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}.$$

Da $P(A \cap B \cap C) = \frac{1}{4} \neq \frac{1}{8}$, sind A, B, C nicht unabhängig.

Fazit

Die Ereignisse A, B, C sind paarweise unabhängig, da die Schnittwahrscheinlichkeiten der Paare den Produkten der Wahrscheinlichkeiten entsprechen. Sie sind jedoch nicht gemeinsam unabhängig, da $P(A \cap B \cap C) \neq P(A) \cdot P(B) \cdot P(C)$.

3.2 Bedingte Wahrscheinlichkeit

Die bedingte Wahrscheinlichkeit eines Ereignisses B , gegeben A , ist definiert als:

$$P(B | A) = \frac{P(A \cap B)}{P(A)}, \quad \text{falls } P(A) > 0.$$

Die bedingte Wahrscheinlichkeit beschreibt, wie sich die Wahrscheinlichkeit von B ändert, wenn bekannt ist, dass A eingetreten ist.

Beispiel:

Beim Ziehen einer Karte aus einem Kartendeck sei:

$$A = \{\text{gezogene Karte ist rot}\}, \quad B = \{\text{gezogene Karte ist ein Herz}\}.$$

Dann gilt:

$$P(A) = \frac{26}{52} = \frac{1}{2}, \quad P(A \cap B) = \frac{13}{52} = \frac{1}{4}, \quad P(B | A) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2}.$$

Bemerkung:

Allgemein gilt für A und B mit $P(A) > 0$

$$P(B|A) = P(B)$$

genau dann, wenn A und B unabhängig.

3.3 Satz von der totalen Wahrscheinlichkeit

Der Satz von der totalen Wahrscheinlichkeit beschreibt, wie die Wahrscheinlichkeit eines Ereignisses B berechnet werden kann, wenn eine Zerlegung des Wahrscheinlichkeitsraums durch paarweise disjunkte Ereignisse A_1, A_2, \dots, A_n mit $P(A_i) > 0$ und $\bigcup_{i=1}^n A_i = \Omega$ vorliegt:

$$P(B) = \sum_{i=1}^n P(B | A_i) \cdot P(A_i).$$

Beweis:

- Das Ereignis B kann in disjunkte Teilereignisse $B \cap A_i$ zerlegt werden:

$$B = \bigcup_{i=1}^n (B \cap A_i), \quad \text{wobei } (B \cap A_i) \cap (B \cap A_j) = \emptyset \text{ für } i \neq j.$$

- Nach der Additivität der Wahrscheinlichkeiten gilt:

$$P(B) = \sum_{i=1}^n P(B \cap A_i).$$

- Mit $P(B \cap A_i) = P(B | A_i) \cdot P(A_i)$:

$$P(B) = \sum_{i=1}^n P(B | A_i) \cdot P(A_i).$$

Beispiel:

In einem Beutel befinden sich 3 rote und 2 grüne Kugeln (Beutel 1) sowie 1 rote und 4 grüne Kugeln (Beutel 2). Ein Beutel wird zufällig gewählt ($P(A_1) = P(A_2) = 0.5$), und daraus eine Kugel gezogen. Sei $B = \{\text{gezogene Kugel ist rot}\}$. Mit $P(B | A_1) = \frac{3}{5}$ und $P(B | A_2) = \frac{1}{5}$ ergibt sich:

$$P(B) = P(B | A_1) \cdot P(A_1) + P(B | A_2) \cdot P(A_2) = \frac{3}{5} \cdot 0.5 + \frac{1}{5} \cdot 0.5 = \frac{2}{5}.$$

3.4 Satz von Bayes

Der Satz von Bayes ermöglicht es, bedingte Wahrscheinlichkeiten umzukehren:

$$P(A | B) = \frac{P(B | A) \cdot P(A)}{P(B)}, \quad \text{falls } P(B) > 0.$$

Beispiel:

Eine Maschine produziert 90% fehlerfreie und 10% fehlerhafte Produkte. Ein Test detektiert Fehler mit 95% Wahrscheinlichkeit ($P(\text{Test positiv} | \text{fehlerhaft}) = 0.95$) und liefert bei fehlerfreien Produkten zu 5% ein falsches positives Ergebnis ($P(\text{Test positiv} | \text{fehlerfrei}) = 0.05$).

Wir berechnen die Wahrscheinlichkeit, dass ein Produkt fehlerhaft ist, gegeben der Test ist positiv:

$$A = \{\text{fehlerhaft}\}, \quad B = \{\text{Test positiv}\}.$$

Mit:

$$P(A) = 0.1, \quad P(B | A) = 0.95, \quad P(B | A^c) = 0.05, \quad P(A^c) = 0.9,$$

ist die Gesamtwahrscheinlichkeit von B (Gesetz der totalen Wahrscheinlichkeit):

$$P(B) = P(B | A) \cdot P(A) + P(B | A^c) \cdot P(A^c) = 0.95 \cdot 0.1 + 0.05 \cdot 0.9 = 0.14.$$

Mit dem Satz von Bayes:

$$P(A | B) = \frac{P(B | A) \cdot P(A)}{P(B)} = \frac{0.95 \cdot 0.1}{0.14} \approx 0.679.$$

3.5 Zweistufige Zufallsexperimente, Baumdiagramme und der Bayes'sche Satz

Zweistufige Zufallsexperimente und Baumdiagramme

Ein zweistufiges Zufallsexperiment besteht aus zwei hintereinander ausgeführten Teilprozessen. Die Ergebnisse der ersten Stufe beeinflussen möglicherweise die Wahrscheinlichkeiten der Ereignisse in der zweiten Stufe. Solche Experimente lassen sich gut durch Baumdiagramme darstellen, die alle möglichen Pfade und deren Wahrscheinlichkeiten visualisieren.

- Jeder Pfad im Baum entspricht einer Kombination von Ereignissen.
 - Die Wahrscheinlichkeiten entlang eines Pfads werden multipliziert, um die Wahrscheinlichkeit des Gesamtereignisses zu berechnen.
 - Die Wahrscheinlichkeit eines Ereignisses auf der zweiten Stufe wird oft durch *bedingte Wahrscheinlichkeiten* beschrieben.

Das Baumdiagramm hilft auch, den Bayes'schen Satz zu verstehen und anzuwenden. Der Satz erlaubt es, eine bedingte Wahrscheinlichkeit $P(A | B)$ umzukehren:

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

wobei $P(B)$ durch den Satz von der totalen Wahrscheinlichkeit berechnet wird:

$$P(B) = \sum_i P(B | A_i) \cdot P(A_i),$$

wenn A_i eine Zerlegung des Wahrscheinlichkeitsraums darstellt.

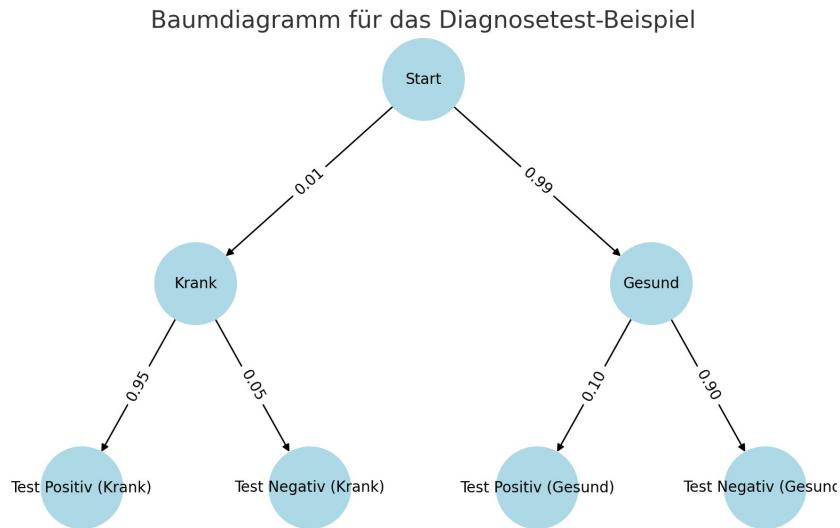
Praktisches Beispiel: Diagnosetest

Ein Diagnosetest für eine Krankheit hat folgende Eigenschaften:
 - Sensitivität (Wahrscheinlichkeit, dass der Test positiv ist, gegeben die Person ist krank): $P(\text{Test positiv} | \text{krank}) = 0.95$.
 - Spezifität (Wahrscheinlichkeit, dass der Test negativ ist, gegeben die Person ist gesund): $P(\text{Test negativ} | \text{gesund}) = 0.90$.

Die Prävalenz der Krankheit (Wahrscheinlichkeit, dass eine zufällige Person krank ist) ist:

$$P(\text{krank}) = 0.01, \quad P(\text{gesund}) = 0.99.$$

Frage: Wie groß ist die Wahrscheinlichkeit, dass eine Person tatsächlich krank ist, wenn der Test positiv ausfällt?



Mit dem Bayes'schen Satz berechnen wir:

$$P(\text{krank} \mid \text{Test positiv}) = \frac{P(\text{Test positiv} \mid \text{krank}) \cdot P(\text{krank})}{P(\text{Test positiv})}.$$

Um die Wahrscheinlichkeit eines positiven Tests ($P(\text{Test positiv})$) ergibt sich noch aus dem Satz von der totalen Wahrscheinlichkeit $P(\text{Test positiv})$ als Summe der Wahrscheinlichkeiten für alle möglichen Ursachen (krank und gesund):

$$P(\text{Test positiv}) = P(\text{Test positiv} \mid \text{krank}) \cdot P(\text{krank}) + P(\text{Test positiv} \mid \text{gesund}) \cdot P(\text{gesund}),$$

$$P(\text{Test positiv}) = (0.95 \cdot 0.01) + (0.10 \cdot 0.99) = 0.1085$$

Diese Wahrscheinlichkeit $P(\text{Test positiv})$ wird nun im Bayes'schen Satz eingesetzt, um die Wahrscheinlichkeit zu berechnen, tatsächlich krank zu sein, gegeben ein positiver Test:

$$P(\text{krank} \mid \text{Test positiv}) = \frac{0.95 \cdot 0.01}{0.1085} \approx 0.0876.$$

Fazit

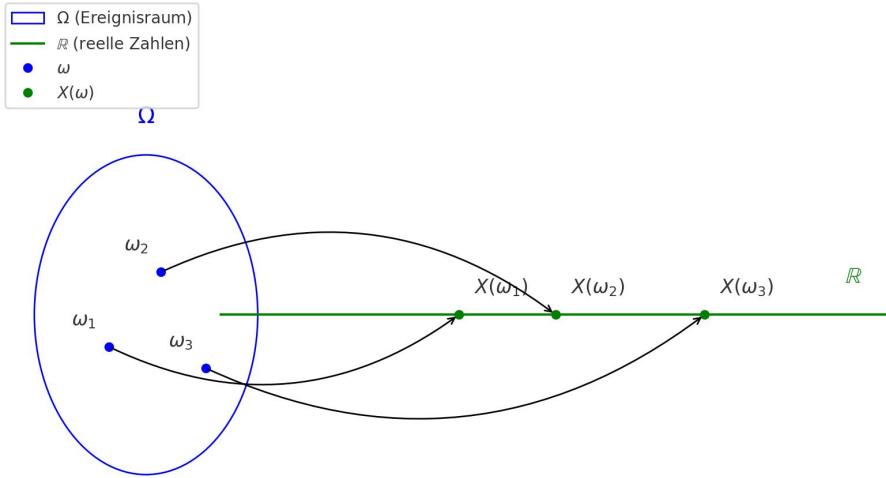
Das Baumdiagramm veranschaulicht die Wahrscheinlichkeiten in zweistufigen Zufallsexperimenten und unterstützt die Anwendung des Bayes'schen Satzes. Im Beispiel zeigt sich, dass trotz eines positiven Tests die Wahrscheinlichkeit, tatsächlich krank zu sein, nur etwa 8.76% beträgt. Dies liegt an der geringen Prävalenz der Krankheit, was die Bedeutung des Satzes von Bayes und der bedingten Wahrscheinlichkeit unterstreicht.

Kapitel 4

Zufallsvariablen und ihre Verteilungen

Eine Zufallsvariable X ist eine Funktion, die jedem Ergebnis eines Zufallsexperiments eine Zahl zuordnet: $X : \Omega \rightarrow \mathbb{R}$.

Schematische Darstellung einer reellwertigen Zufallsvariablen mit alternierenden gebogenen Pfeilen



Erwartungswert: Der Erwartungswert einer Zufallsvariable X ist definiert als:

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \cdot P(\omega)$$

Ordnet man die Summation nach den auftretenden Bildwerten $\{X(\omega) \mid \omega \in \Omega\} = X(\Omega)$ ergibt sich die alternative Formel zur Berechnung des Erwartungswertes wie folgt

$$\mathbb{E}[X] = \sum_{x \in X(\Omega)} x \cdot P(X = x).$$

In der ersten Formel ist über alle Elemente von Ω zu summieren, in der zweiten über alle möglichen Bildwerte von X , was im Regelfall deutlich weniger Summationen erfordert.

Varianz und Standardabweichung: Die Varianz misst die Streuung einer Zufallsvariablen um ihren Erwartungswert:

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

Die Standardabweichung ist die Quadratwurzel der Varianz,

$$\sigma(X) = \sqrt{\text{Var}(X)}.$$

4.1 Verteilung einer (diskreten) Zufallsvariablen

Sei X eine diskrete Zufallsvariable auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{F}, P) . Die **Verteilung** von X ist definiert als die Menge:

$$\mathcal{D}_X = \{(x, p_x) \mid x \in \text{Wertebereich von } X, p_x = P(X = x)\},$$

wobei:

- x die möglichen Werte von X (auch *Realisationen* oder *Träger der Verteilung*) sind,
- $p_x = P(X = x)$ die Wahrscheinlichkeit ist, dass X den Wert x annimmt,
- die Wahrscheinlichkeiten die Eigenschaften erfüllen:

$$p_x \geq 0 \quad \text{für alle } x, \quad \text{und} \quad \sum_x p_x = 1.$$

Beispiel 1: Würfelwurf

Das Zufallsexperiment des Würfelwurfs wird mathematisch beschrieben durch einen Laplace'schen Wahrscheinlichkeitsraum $\Omega = \{\omega_1, \omega_2, \dots, \omega_6\}$ mit 6 Elementen, so dass mit $P(\{\omega_i\}) = \frac{1}{6}$ für $i = 1, \dots, 6$. Für die Zufallsvariable $X : \Omega \mapsto \mathbb{R}$, $X(\omega_i) = i$, die die Augenzahl beschreibt bzw.e auszahlt, ist der Wertebereich von X $\{1, 2, 3, 4, 5, 6\}$. Die Wahrscheinlichkeiten sind $P(X = x) = \frac{1}{6}$ für jedes x . Die Verteilung von X ist daher:

$$\mathcal{D}_X = \left\{ \left(1, \frac{1}{6}\right), \left(2, \frac{1}{6}\right), \left(3, \frac{1}{6}\right), \left(4, \frac{1}{6}\right), \left(5, \frac{1}{6}\right), \left(6, \frac{1}{6}\right) \right\}.$$

Beispiel 2: Augensumme beim zweimaligen Münzwurf

Beschreibung des Zufallsexperiments

Betrachten wir ein Experiment, bei dem eine faire Münze zweimal geworfen wird. Die Ergebnisse der beiden Würfe seien K (Kopf) oder Z (Zahl). Wir definieren eine Zufallsvariable X , die die Anzahl der Köpfe (**Augensumme**) in diesen beiden Würfen beschreibt.

- **Ergebnismenge (Ω):** Die möglichen Ergebnisse des Experiments sind:

$$\Omega = \{(K, K), (K, Z), (Z, K), (Z, Z)\}.$$

- **Zufallsvariable X :** Die Zufallsvariable X beschreibt die Anzahl der Köpfe in einem Ergebnis:

$$X((K, K)) = 2, \quad X((K, Z)) = 1, \quad X((Z, K)) = 1, \quad X((Z, Z)) = 0.$$

- **Wertebereich von X :** Die möglichen Werte, die X annehmen kann, sind:

$$\text{Wertebereich von } X = \{0, 1, 2\}.$$

Wahrscheinlichkeiten

Da die Münze fair ist und die beiden Würfe unabhängig sind, hat jedes Ergebnis in Ω die gleiche Wahrscheinlichkeit von:

$$P(\text{jeweiliges Ergebnis}) = \frac{1}{4}.$$

Für die Zufallsvariable X ergibt sich die Wahrscheinlichkeitsverteilung durch Aggregation der Wahrscheinlichkeiten der Ereignisse, die zu demselben Wert von X führen:

- $P(X = 0)$: Nur das Ergebnis (Z, Z) führt zu $X = 0$:

$$P(X = 0) = P((Z, Z)) = \frac{1}{4}.$$

- $P(X = 1)$: Die Ergebnisse (K, Z) und (Z, K) führen zu $X = 1$:

$$P(X = 1) = P((K, Z)) + P((Z, K)) = \frac{1}{4} + \frac{1}{4} = \frac{2}{4} = \frac{1}{2}.$$

- $P(X = 2)$: Nur das Ergebnis (K, K) führt zu $X = 2$:

$$P(X = 2) = P((K, K)) = \frac{1}{4}.$$

Verteilung der Zufallsvariablen

Die Verteilung der Zufallsvariablen X , beschrieben als Menge von 2-Tupeln, ist:

$$\mathcal{D}_X = \left\{ (0, \frac{1}{4}), (1, \frac{1}{2}), (2, \frac{1}{4}) \right\}.$$

4.2 Beispiel: Zufallsexperiment mit Kugeln in rhombischer Anordnung

1. Wahrscheinlichkeitsraum

Das Zufallsexperiment wird wie folgt beschrieben:

- **Ergebnismenge (Ω)**: Die Ergebnismenge besteht aus den 9 Kugeln, nummeriert von 1 bis 9. Jede Kugel hat eine eindeutige Position (x_i, y_i) :

$$\Omega = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}.$$

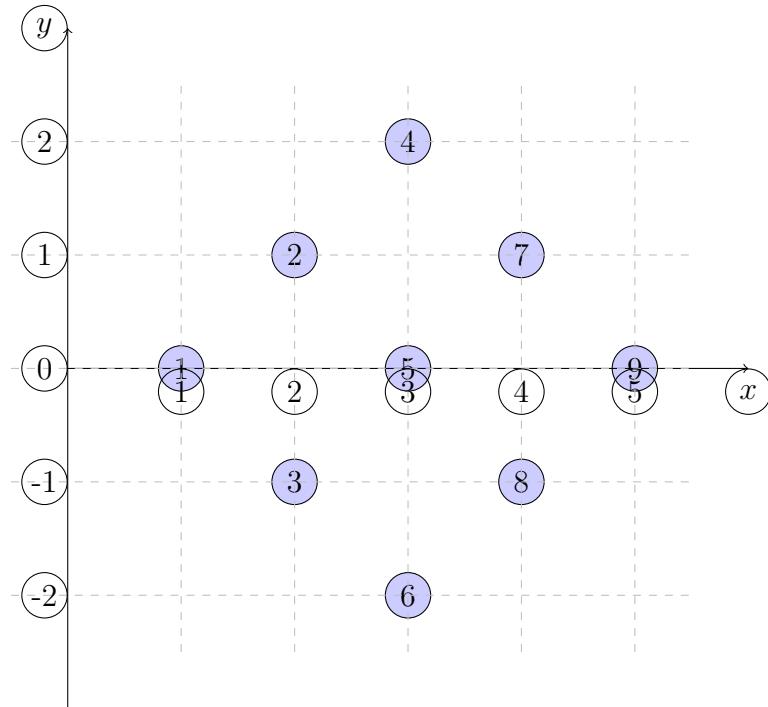
- **Positionen der Kugeln**: Die Positionen der Kugeln auf der Ebene sind:

$$\begin{aligned} 1 &\mapsto (1, 0), \\ 2 &\mapsto (2, 1), \quad 3 \mapsto (2, -1), \\ 4 &\mapsto (3, 2), \quad 5 \mapsto (3, 0), \quad 6 \mapsto (3, -2), \\ 7 &\mapsto (4, 1), \quad 8 \mapsto (4, -1), \\ 9 &\mapsto (5, 0). \end{aligned}$$

- **Ereignisalgebra (\mathcal{F}):** Die Ereignisalgebra ist die Potenzmenge von Ω , d.h., die Menge aller Teilmengen von Ω .
- **Wahrscheinlichkeitsmaß (P):** Jede Kugel wird mit gleicher Wahrscheinlichkeit gezogen:

$$P(\{i\}) = \frac{1}{9}, \quad \text{für } i = 1, 2, \dots, 9.$$

Skizze der Kugelanordnung



2. Eine Zufallsvariable

Die Zufallsvariable $X : \Omega \rightarrow \mathbb{R}$ gibt die x -Koordinate der gezogenen Kugel an. Sie ist definiert durch:

$$X(i) = x_i, \quad \text{wobei } x_i \text{ die } x\text{-Koordinate der Kugel } i \text{ ist.}$$

Die x -Koordinaten der Kugeln sind:

$$x_1 = 1, x_2 = x_3 = 2, x_4 = x_5 = x_6 = 3, x_7 = x_8 = 4, x_9 = 5.$$

3. Wahrscheinlichkeitsverteilung von X

Die Wahrscheinlichkeitsverteilung von X wird durch die Häufigkeit der x -Koordinaten bestimmt:

$$P(X = x) = \frac{\text{Anzahl der Kugeln mit } x \text{ als } x\text{-Koordinate}}{9}.$$

Somit ergibt sich:

$$\begin{aligned} P(X = 1) &= \frac{1}{9}, \\ P(X = 2) &= \frac{2}{9}, \\ P(X = 3) &= \frac{3}{9} = \frac{1}{3}, \\ P(X = 4) &= \frac{2}{9}, \\ P(X = 5) &= \frac{1}{9}. \end{aligned}$$

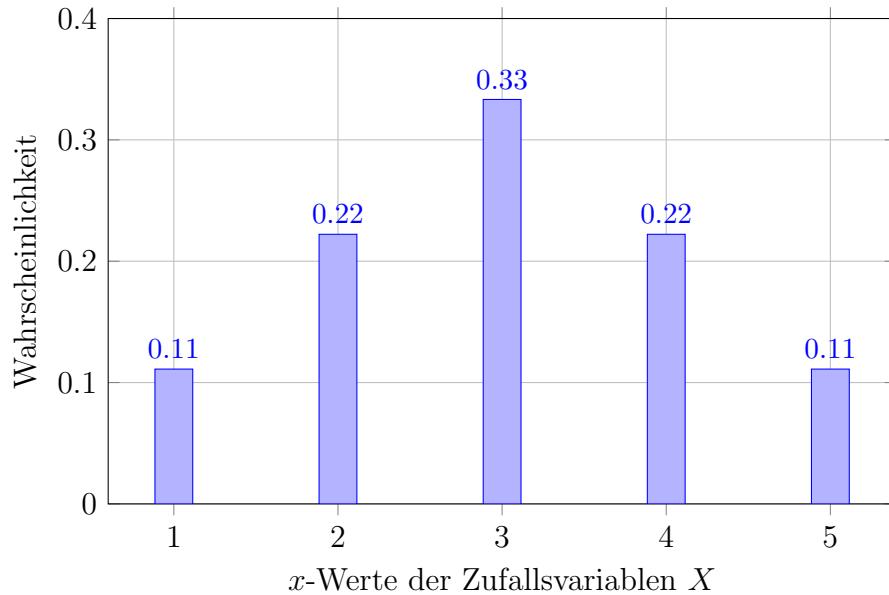
Die Wahrscheinlichkeitsverteilung ist:

$$P(X = x) = \begin{cases} \frac{1}{9}, & \text{für } x = 1 \text{ oder } x = 5, \\ \frac{2}{9}, & \text{für } x = 2 \text{ oder } x = 4, \\ \frac{1}{3}, & \text{für } x = 3. \end{cases}$$

Somit ergibt sich als Verteilung von X

$$\mathcal{D}_X = \left\{ (1, \frac{1}{9}), (2, \frac{2}{9}), (3, \frac{1}{3}), (4, \frac{2}{9}), (5, \frac{1}{9}) \right\}.$$

Histogramm der Verteilung von X



4. Erwartungswert von X

Der Erwartungswert von X ist definiert als:

$$\mathbb{E}[X] = \sum_{x \in \text{Wertebereich von } X} x \cdot P(X = x).$$

Einsetzen der Wahrscheinlichkeiten:

$$\begin{aligned}\mathbb{E}[X] &= 1 \cdot \frac{1}{9} + 2 \cdot \frac{2}{9} + 3 \cdot \frac{1}{3} + 4 \cdot \frac{2}{9} + 5 \cdot \frac{1}{9} \\ &= \frac{1}{9} + \frac{4}{9} + \frac{9}{9} + \frac{8}{9} + \frac{5}{9} \\ &= \frac{1+4+9+8+5}{9} \\ &= \frac{27}{9} = 3.\end{aligned}$$

Der Erwartungswert ist:

$$\mathbb{E}[X] = 3.$$

5. Varianz von X

Die Varianz von X ist definiert als:

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

Berechnung von $\mathbb{E}[X^2]$:

$$\mathbb{E}[X^2] = \sum_{x \in \text{Wertebereich von } X} x^2 \cdot P(X = x).$$

Einsetzen der Wahrscheinlichkeiten:

$$\begin{aligned}\mathbb{E}[X^2] &= 1^2 \cdot \frac{1}{9} + 2^2 \cdot \frac{2}{9} + 3^2 \cdot \frac{1}{3} + 4^2 \cdot \frac{2}{9} + 5^2 \cdot \frac{1}{9} \\ &= \frac{1}{9} + \frac{8}{9} + \frac{27}{9} + \frac{32}{9} + \frac{25}{9} \\ &= \frac{1+8+27+32+25}{9} \\ &= \frac{93}{9} = 10.3333.\end{aligned}$$

Berechnung der Varianz:

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

Einsetzen:

$$\text{Var}(X) = 10.3333 - 3^2 = 10.3333 - 9 = 1.3333.$$

6. Ergebnisse

- Wahrscheinlichkeitsverteilung von X :

$$P(X = x) = \begin{cases} \frac{1}{9}, & \text{für } x = 1 \text{ oder } x = 5, \\ \frac{2}{9}, & \text{für } x = 2 \text{ oder } x = 4, \\ \frac{1}{3}, & \text{für } x = 3. \end{cases}$$

- Erwartungswert:

$$\mathbb{E}[X] = 3.$$

- Varianz:

$$\text{Var}(X) = 1.3333.$$

4.3 Beispiel: Binomialverteilung

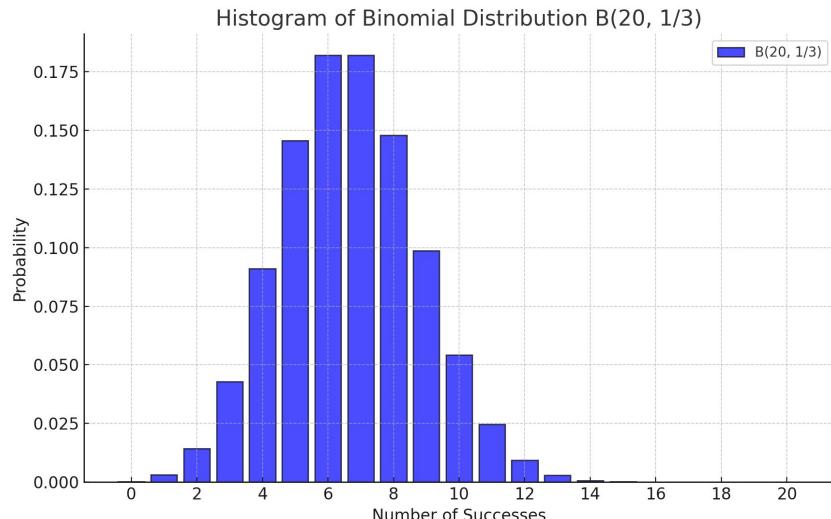
Die Binomialverteilung $B(n, p)$ beschreibt die Anzahl der Erfolge in n unabhängigen Bernoulli-Experimenten, bei denen die Erfolgswahrscheinlichkeit p konstant ist. Die Zufallsvariable X gibt die Anzahl der Erfolge an und hat die Wahrscheinlichkeitsfunktion:

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

wobei $\binom{n}{k}$ der Binomialkoeffizient ist. In der Tat, beim n -maligen Münzwurf gibt es $\binom{n}{k}$ viele Möglichkeiten, die k Erfolge auf die n Spielrunden zu verteilen. Jede der hierbei gezählten Sequenzen von Erfolgen und Misserfolgen hat die Wahrscheinlichkeit $p^k(1-p)^{n-k}$.

Eigenschaften der Binomialverteilung: 1. $E[X] = n \cdot p$: Der Erwartungswert ist proportional zur Anzahl der Versuche und der Erfolgswahrscheinlichkeit. 2. $\text{Var}(X) = n \cdot p \cdot (1-p)$: Die Varianz hängt von n , p und der Misserfolgswahrscheinlichkeit $1 - p$ ab. Dieser Aussagen ergeben sich leicht aus der Darstellung $X = \sum_{i=1}^n \xi_i$, wobei (ξ_i) eine Folge von unabhängigen Bernoulli 1/0-Zufallsvariablen mit Erfolgsparameter p ist, dh. mit $P(\xi_i = 1) = 1 - P(\xi_i = 0) = p$.

Ein Beispiel ist das Werfen einer fairen Münze $n = 20$ Mal, wobei $p = \frac{1}{3}$ die Wahrscheinlichkeit für Erfolg ist. Die Verteilung der Erfolge X wird durch das Histogramm illustriert.



4.4 Tschebyschev-Ungleichung

Formulierung der Ungleichung

Sei X eine Zufallsvariable mit Erwartungswert $\mathbb{E}[X] = \mu$ und Varianz $\text{Var}(X) = \sigma^2 < \infty$. Für jedes $\varepsilon > 0$ gilt:

$$P(|X - \mu| \geq \varepsilon) \leq \frac{\text{Var}(X)}{\varepsilon^2}.$$

Interpretation

Die Tschebyschev-Ungleichung liefert eine obere Schranke für die Wahrscheinlichkeit, dass eine Zufallsvariable X mehr als ε vom Erwartungswert μ abweicht. Sie ist besonders nützlich, da sie für beliebige Verteilungen gilt.

Die Tschebyschev-Ungleichung ist ein Spezialfall der folgenden allgemeinen Abschätzung.

4.4.1 Markov-Ungleichung

Satz (Markov-Ungleichung)

Sei $X \geq 0$ eine nichtnegative Zufallsvariable auf einem diskreten Wahrscheinlichkeitsraum (Ω, \mathcal{F}, P) . Für jedes $\varepsilon > 0$ gilt:

$$P(X \geq \varepsilon) \leq \frac{\mathbb{E}[X]}{\varepsilon}.$$

Beweis

1. Definition der Wahrscheinlichkeit $P(X \geq \varepsilon)$: Die Wahrscheinlichkeit, dass $X \geq \varepsilon$, ist:

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

2. Aufteilung der Summe der Erwartungswertdefinition: Der Erwartungswert von X ist definiert als:

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \cdot P(\omega).$$

Wir schreiben die Summe über Ω auf und trennen sie in zwei Teile:

$$\mathbb{E}[X] = \sum_{\omega \in \Omega, X(\omega) \geq \varepsilon} X(\omega) \cdot P(\omega) + \sum_{\omega \in \Omega, X(\omega) < \varepsilon} X(\omega) \cdot P(\omega).$$

3. Abschätzung des ersten Summenanteils: Für ω , bei denen $X(\omega) \geq \varepsilon$, gilt $X(\omega) \geq \varepsilon$. Daraus folgt:

$$\sum_{\omega \in \Omega, X(\omega) \geq \varepsilon} X(\omega) \cdot P(\omega) \geq \sum_{\omega \in \Omega, X(\omega) \geq \varepsilon} \varepsilon \cdot P(\omega).$$

4. Faktor ε ausklammern: Da ε konstant ist, kann es aus der Summe ausgeklammert werden:

$$\sum_{\omega \in \Omega, X(\omega) \geq \varepsilon} \varepsilon \cdot P(\omega) = \varepsilon \cdot \sum_{\omega \in \Omega, X(\omega) \geq \varepsilon} P(\omega).$$

5. Verbindung mit $P(X \geq \varepsilon)$: Die Summe $\sum_{\omega \in \Omega, X(\omega) \geq \varepsilon} P(\omega)$ ist per Definition $P(X \geq \varepsilon)$. Also gilt:

$$\sum_{\omega \in \Omega, X(\omega) \geq \varepsilon} X(\omega) \cdot P(\omega) \geq \varepsilon \cdot P(X \geq \varepsilon).$$

6. Abschließende Ungleichung: Setze dies in die Erwartungswertformel ein:

$$\mathbb{E}[X] \geq \varepsilon \cdot P(X \geq \varepsilon).$$

Umstellen nach $P(X \geq \varepsilon)$ ergibt:

$$P(X \geq \varepsilon) \leq \frac{\mathbb{E}[X]}{\varepsilon}.$$

Fazit

Die Markov-Ungleichung ist bewiesen. Sie zeigt, dass für eine nichtnegative Zufallsvariable die Wahrscheinlichkeit, dass sie einen Schwellenwert ε überschreitet, durch den Erwartungswert von X und ε begrenzt wird.

4.4.2 Beweis der Tschebyschev-Ungleichung

Sei X eine Zufallsvariable mit Erwartungswert $\mathbb{E}[X] = \mu$ und Varianz $\text{Var}(X) = \sigma^2 < \infty$. Die Tschebyschev-Ungleichung lautet:

$$P(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}.$$

Wir beweisen diese Aussage, indem wir die Markov-Ungleichung auf den Ausdruck $(X - \mu)^2$ anwenden.

1. Definiere die Zufallsvariable $Z = (X - \mu)^2$: Die Zufallsvariable Z ist nichtnegativ, da $(X - \mu)^2 \geq 0$. Weiterhin gilt:

$$\mathbb{E}[Z] = \mathbb{E}[(X - \mu)^2] = \text{Var}(X) = \sigma^2.$$

2. Wende die Markov-Ungleichung auf Z an: Für jedes $\varepsilon > 0$ gilt nach der Markov-Ungleichung:

$$P(Z \geq \varepsilon^2) \leq \frac{\mathbb{E}[Z]}{\varepsilon^2}.$$

3. Ersetze $Z = (X - \mu)^2$:

$$P((X - \mu)^2 \geq \varepsilon^2) \leq \frac{\text{Var}(X)}{\varepsilon^2}.$$

4. Äquivalenz der Ereignisse $(X - \mu)^2 \geq \varepsilon^2$ und $|X - \mu| \geq \varepsilon$: Da $(X - \mu)^2 \geq \varepsilon^2$ genau dann gilt, wenn $|X - \mu| \geq \varepsilon$, folgt:

$$P(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}.$$

Fazit

Die Tschebyschev-Ungleichung ist bewiesen durch Anwendung der Markov-Ungleichung auf $(X - \mu)^2$. Sie zeigt, dass die Wahrscheinlichkeit großer Abweichungen einer Zufallsvariablen vom Erwartungswert durch ihre Varianz begrenzt wird.

Kapitel 5

Unabhängigkeit von Zufallsvariablen

Zwei Zufallsvariablen X und Y heißen unabhängig, wenn:

$$P(X \in A \text{ und } Y \in B) = P(X \in A) \cdot P(Y \in B) \quad \text{für alle } A, B \subseteq \mathbb{R}.$$

Unabhängigkeit bedeutet, dass die Verteilung von X keinen Einfluss auf die Verteilung von Y hat.

Beispiel:

Beim zweimaligen Würfeln sind die Ergebnisse der beiden Würfe unabhängig. Die Wahrscheinlichkeit, zweimal eine 6 zu würfeln, ist:

$$P(X_1 = 6 \text{ und } X_2 = 6) = P(X_1 = 6) \cdot P(X_2 = 6) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}.$$

5.1 Faktorisierung des Erwartungswerts bei Unabhängigkeit

Satz

Seien X und Y zwei unabhängige Zufallsvariablen auf einem diskreten Wahrscheinlichkeitsraum. Dann gilt:

$$\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y].$$

Beweis durch Summation über die Verteilung der Produktzufallsvariablen

1. Erwartungswertdefinition: Der Erwartungswert einer Funktion $f(X, Y)$ zweier Zufallsvariablen ist definiert als:

$$\mathbb{E}[f(X, Y)] = \sum_{x \in \text{Wertebereich von } X} \sum_{y \in \text{Wertebereich von } Y} f(x, y) \cdot P(X = x, Y = y).$$

Setze $f(X, Y) = X \cdot Y$, dann:

$$\mathbb{E}[X \cdot Y] = \sum_x \sum_y x \cdot y \cdot P(X = x, Y = y).$$

2. Unabhängigkeit der Zufallsvariablen: Da X und Y unabhängig sind, gilt:

$$P(X = x, Y = y) = P(X = x) \cdot P(Y = y).$$

Einsetzen ergibt:

$$\mathbb{E}[X \cdot Y] = \sum_x \sum_y x \cdot y \cdot P(X = x) \cdot P(Y = y).$$

3. Aufteilen der Summation: Schreibe die Summation als Produkt zweier unabhängiger Summen:

$$\mathbb{E}[X \cdot Y] = \sum_x x \cdot P(X = x) \cdot \sum_y y \cdot P(Y = y).$$

Beachte, dass der Faktor $x \cdot P(X = x)$ unabhängig von y ist und aus der inneren Summe ausgeklammert werden kann.

4. Erkennung der Erwartungswerte: Die einzelnen Summen sind die Definitionen der Erwartungswerte von X und Y :

$$\sum_x x \cdot P(X = x) = \mathbb{E}[X], \quad \sum_y y \cdot P(Y = y) = \mathbb{E}[Y].$$

5. Ergebnis: Setze dies zusammen:

$$\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y].$$

Fazit

Durch Summation über die Verteilung der Produktzufallsvariablen haben wir gezeigt, dass der Erwartungswert einer Produktzufallsvariablen $X \cdot Y$ sich bei Unabhängigkeit von X und Y in das Produkt der Erwartungswerte faktorisieren lässt:

$$\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y].$$

5.2 Kovarianz und Korrelation

Definition der Korrelation

Die Korrelation beschreibt den linearen Zusammenhang zwischen zwei Zufallsvariablen X und Y . Sie wird durch den Korrelationskoeffizienten $\rho(X, Y)$ definiert:

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

Hierbei ist:

- $\text{Cov}(X, Y)$ die Kovarianz von X und Y :

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X] \cdot \mathbb{E}[Y].$$

- $\rho(X, Y)$ ist dimensionslos und liegt im Bereich:

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

Interpretation der Korrelation

- $\rho(X, Y) = 0$: X und Y sind unkorreliert (kein linearer Zusammenhang).
- $\rho(X, Y) = 1$: Perfekte positive lineare Korrelation.
- $\rho(X, Y) = -1$: Perfekte negative lineare Korrelation.

Zusammenhang mit Unabhängigkeit

- Sind X und Y unabhängig, so ist $\text{Cov}(X, Y) = 0$ und somit $\rho(X, Y) = 0$.
- Unkorreliertheit ($\rho(X, Y) = 0$) bedeutet jedoch nicht zwangsläufig Unabhängigkeit, wie durch Beispiele (z. B. quadratische Abhängigkeiten) gezeigt werden kann.

Beispiel (Forts.) Rhombische Kugelanordnung

Wir betrachten das Experiment der rhombischen Kugelanordnung. Jede Kugel hat eine eindeutige Position (x_i, y_i) , wie folgt:

$$\begin{aligned} 1 &\mapsto (1, 0), \\ 2 &\mapsto (2, 1), \quad 3 \mapsto (2, -1), \\ 4 &\mapsto (3, 2), \quad 5 \mapsto (3, 0), \quad 6 \mapsto (3, -2), \\ 7 &\mapsto (4, 1), \quad 8 \mapsto (4, -1), \\ 9 &\mapsto (5, 0). \end{aligned}$$

Die Zufallsvariablen sind definiert als:

- X : Die x -Koordinate der gezogenen Kugel.
- Y : Die y -Koordinate der gezogenen Kugel.

1. Wertebereich und Wahrscheinlichkeiten

Die Verteilungen von X und Y sind:

- **Verteilung von X :**

$$P(X = 1) = \frac{1}{9}, \quad P(X = 2) = \frac{2}{9}, \quad P(X = 3) = \frac{3}{9}, \quad P(X = 4) = \frac{2}{9}, \quad P(X = 5) = \frac{1}{9}.$$

- **Verteilung von Y :**

$$P(Y = -2) = \frac{1}{9}, \quad P(Y = -1) = \frac{2}{9}, \quad P(Y = 0) = \frac{3}{9}, \quad P(Y = 1) = \frac{2}{9}, \quad P(Y = 2) = \frac{1}{9}.$$

- **Gemeinsame Verteilung von X und Y :** Die Wahrscheinlichkeiten für die Kombinationen von X und Y sind durch die Positionen der Kugeln gegeben. Zum Beispiel:

$$\begin{aligned} P(X = 1, Y = 0) &= \frac{1}{9}, \quad P(X = 2, Y = 1) = \frac{1}{9}, \quad P(X = 2, Y = -1) = \frac{1}{9}, \\ P(X = 3, Y = 2) &= \frac{1}{9}, \quad P(X = 3, Y = 0) = \frac{1}{9}, \quad P(X = 3, Y = -2) = \frac{1}{9}, \\ P(X = 4, Y = 1) &= \frac{1}{9}, \quad P(X = 4, Y = -1) = \frac{1}{9}, \quad P(X = 5, Y = 0) = \frac{1}{9}. \end{aligned}$$

2. Erwartungswerte

Die Erwartungswerte von X und Y sind:

$$\mathbb{E}[X] = \sum_x x \cdot P(X = x) = 1 \cdot \frac{1}{9} + 2 \cdot \frac{2}{9} + 3 \cdot \frac{3}{9} + 4 \cdot \frac{2}{9} + 5 \cdot \frac{1}{9} = 3.$$

$$\mathbb{E}[Y] = \sum_y y \cdot P(Y = y) = (-2) \cdot \frac{1}{9} + (-1) \cdot \frac{2}{9} + 0 \cdot \frac{3}{9} + 1 \cdot \frac{2}{9} + 2 \cdot \frac{1}{9} = 0.$$

Das Produkt der Erwartungswerte ist:

$$\mathbb{E}[X] \cdot \mathbb{E}[Y] = 3 \cdot 0 = 0.$$

3. Kovarianz und Korrelation

Die Kovarianz von X und Y ist:

$$\text{Cov}(X, Y) = \mathbb{E}[X \cdot Y] - \mathbb{E}[X] \cdot \mathbb{E}[Y].$$

Zunächst berechnen wir $\mathbb{E}[X \cdot Y]$:

$$\mathbb{E}[X \cdot Y] = \sum_{x,y} x \cdot y \cdot P(X = x, Y = y).$$

Einsetzen der gemeinsamen Wahrscheinlichkeiten:

$$\begin{aligned} \mathbb{E}[X \cdot Y] &= 1 \cdot 0 \cdot \frac{1}{9} + 2 \cdot 1 \cdot \frac{1}{9} + 2 \cdot (-1) \cdot \frac{1}{9} \\ &\quad + 3 \cdot 2 \cdot \frac{1}{9} + 3 \cdot 0 \cdot \frac{1}{9} + 3 \cdot (-2) \cdot \frac{1}{9} \\ &\quad + 4 \cdot 1 \cdot \frac{1}{9} + 4 \cdot (-1) \cdot \frac{1}{9} + 5 \cdot 0 \cdot \frac{1}{9} \\ &= 0 + \frac{2}{9} - \frac{2}{9} + \frac{6}{9} + 0 - \frac{6}{9} + \frac{4}{9} - \frac{4}{9} + 0 \\ &= 0. \end{aligned}$$

Damit ist:

$$\text{Cov}(X, Y) = \mathbb{E}[X \cdot Y] - \mathbb{E}[X] \cdot \mathbb{E}[Y] = 0 - 0 = 0.$$

Da $\text{Cov}(X, Y) = 0$, sind X und Y unkorreliert.

Bemerkung:

Man prüft leicht nach, dass in diesem Beispiel die Zufallsvariablen $U = X + Y$ und $V = X - Y$ stochastisch unabhängig sind.

4. Abhängigkeit

Die Zufallsvariablen X und Y sind jedoch nicht unabhängig, da die gemeinsame Verteilung nicht als Produkt der Randverteilungen darstellbar ist. Zum Beispiel:

$$P(X = 3, Y = 2) = \frac{1}{9}, \quad P(X = 3) \cdot P(Y = 2) = \frac{1}{3} \cdot \frac{1}{9} = \frac{1}{27} \neq \frac{1}{9}.$$

Fazit

Die Zufallsvariablen X (die x -Koordinate) und Y (die y -Koordinate) aus dem Experiment sind **nicht unabhängig**, da ihre gemeinsame Verteilung nicht das Produkt der Randverteilungen ist. Sie sind jedoch **unkorreliert**, da ihre Kovarianz $\text{Cov}(X, Y) = 0$ ist. Dieses Beispiel illustriert, dass Unkorreliertheit nicht notwendigerweise Unabhängigkeit bedeutet.

5.3 Satz von Bienaym  zur Additivit t der Varianz

Satz von Bienaym 

Seien X und Y zwei unabhängige Zufallsvariablen auf einem Wahrscheinlichkeitsraum. Dann gilt:

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).$$

Verallgemeinerung

Für n paarweise unabhängige Zufallsvariablen X_1, X_2, \dots, X_n gilt:

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

Beweis

1. Die Varianz einer Summe von Zufallsvariablen ist:

$$\text{Var}(X + Y) = \mathbb{E}[(X + Y)^2] - \mathbb{E}[X + Y]^2.$$

2. Verwenden der Linearit t des Erwartungswerts:

$$\mathbb{E}[X + Y]^2 = (\mathbb{E}[X] + \mathbb{E}[Y])^2.$$

3. Der entscheidende Schritt: Bei Unabhängigkeit gilt:

$$\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y],$$

wodurch sich gemischte Terme in $\mathbb{E}[(X + Y)^2]$ eliminieren.

4. Nach Vereinfachung bleibt:

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).$$

Kapitel 6

Gesetz der großen Zahlen (GGZ)

Das Gesetz der großen Zahlen besagt, dass der arithmetische Mittelwert einer Folge von unabhängigen, identisch verteilten Zufallsvariablen $(X_i)_{i \in \mathbb{N}}$ mit $\mathbb{E}[X_i] = \mu$ gegen den Erwartungswert μ konvergiert:

$$\frac{1}{n} \sum_{i=1}^n X_i \longrightarrow \mu \quad \text{für } n \rightarrow \infty.$$

Wie man diese Konvergenz hier verstehen muss, wird weiter unten präzise formuliert.

Beispiel:

Bei n -maligem Würfeln nähert sich der Durchschnitt der Augenzahlen $\frac{1}{n} \sum_{i=1}^n X_i$ dem Erwartungswert 3.5.

6.1 Satz (Schwaches Gesetz der großen Zahlen)

Sei $(X_i)_{i \in \mathbb{N}}$ eine Folge von unabhängigen und identisch verteilten Zufallsvariablen mit Erwartungswert $\mathbb{E}[X_i] = \mu$ und Varianz $\text{Var}(X_i) = \sigma^2 < \infty$. Dann gilt für alle $\epsilon > 0$

$$\mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - \mu\right|\right) \longrightarrow 0 \quad \text{für } n \rightarrow \infty.$$

Bemerkung

Man sagt hierfür auch, dass der Stichprobenmittelwert $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ in Wahrscheinlichkeit gegen den Erwartungswert μ konvergiert.

Beweis

1. Definition des Stichprobenmittelwerts: Sei $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$. Der Erwartungswert und die Varianz von \bar{X}_n sind:

$$\mathbb{E}[\bar{X}_n] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] = \mu,$$

$$\text{Var}(\bar{X}_n) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{\sigma^2}{n}.$$

2. Nach der Tschebyschev-Ungleichung gilt für jedes $\varepsilon > 0$:

$$P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{\text{Var}(\bar{X}_n)}{\varepsilon^2}.$$

3. Einsetzen der Varianz von \bar{X}_n : Da $\text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$, folgt:

$$P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}.$$

4. Grenzwertbetrachtung: Wenn $n \rightarrow \infty$, dann konvergiert $\frac{\sigma^2}{n\varepsilon^2} \rightarrow 0$. Somit:

$$P(|\bar{X}_n - \mu| \geq \varepsilon) \rightarrow 0 \quad \text{für } n \rightarrow \infty.$$

Fazit

Nach Definition der Konvergenz in Wahrscheinlichkeit folgt:

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P} \mu.$$

Das schwache Gesetz der großen Zahlen ist bewiesen.

Bemerkung

Wir haben hier das Gesetz der Großen Zahlen in der sogenannten 'schwachen Version' formuliert und bewiesen. Das schwache Gesetz der großen Zahlen besagt, dass der Mittelwert einer Folge unabhängiger Zufallsvariablen mit wachsender Stichprobengröße in Wahrscheinlichkeit gegen den Erwartungswert konvergiert, während das starke Gesetz der großen Zahlen die fast sichere (d.h. fast überall) Konvergenz des Mittelwerts gegen den Erwartungswert garantiert. Das starke Gesetz ist die weitergehende Aussage, der Beweis ist etwas (aber nicht viel) aufwändiger und beruht auf einer Anwendung des Borel-Cantelli-Lemmas. Er soll hier nicht besprochen werden.

6.2 Anwendungen

6.2.1 Monte-Carlo-Integration

Die Monte-Carlo-Integration ist eine numerische Methode zur Approximation von Integralen, insbesondere in hohen Dimensionen. Sie basiert auf dem Schwachen Gesetz der großen Zahlen, das garantiert, dass der Mittelwert einer Folge unabhängiger Zufallsvariablen in Wahrscheinlichkeit gegen ihren Erwartungswert konvergiert.

Gegeben ein Integral einer Funktion $f(x)$ über ein Intervall $[a, b]$:

$$I = \int_a^b f(x) dx,$$

kann dieses als Erwartungswert interpretiert werden, wenn X eine Zufallsvariable ist, die gleichverteilt auf $[a, b]$ ist:

$$I = (b - a) \cdot \mathbb{E}[f(X)].$$

Die Monte-Carlo-Methode approximiert $\mathbb{E}[f(X)]$ durch den Mittelwert:

$$\hat{I}_n = \frac{b-a}{n} \sum_{i=1}^n f(X_i),$$

wobei X_1, X_2, \dots, X_n unabhängige, gleichverteilte Stichproben auf $[a, b]$ sind.

Nach dem Schwachen Gesetz der großen Zahlen gilt:

$$\hat{I}_n \xrightarrow{P} I \quad \text{für } n \rightarrow \infty.$$

Numerische Beispiele

Beispiel 1: Integral einer linearen Funktion

Berechne $\int_0^1 x \, dx = \frac{1}{2}$ mittels Monte-Carlo-Integration:

$$I = \int_0^1 x \, dx = 1 \cdot \mathbb{E}[X],$$

wobei $X \sim U(0, 1)$. Ziehe $n = 1000$ Zufallsstichproben $X_i \sim U(0, 1)$ und berechne:

$$\hat{I}_n = \frac{1}{1000} \sum_{i=1}^{1000} X_i.$$

Angenommen, die Stichproben ergeben einen Mittelwert von 0.4987, so ist:

$$\hat{I}_{1000} \approx 0.4987.$$

Die Schätzung nähert sich dem wahren Wert $I = 0.5$ mit zunehmendem n .

Beispiel 2: Integral einer nichtlinearen Funktion

Berechne $\int_0^1 x^2 \, dx = \frac{1}{3}$:

$$I = \int_0^1 x^2 \, dx = \mathbb{E}[X^2], \quad X \sim U(0, 1).$$

Für $n = 1000$ Zufallsstichproben $X_i \sim U(0, 1)$, berechne den Mittelwert der Funktion $f(X) = X^2$:

$$\hat{I}_n = \frac{1}{1000} \sum_{i=1}^{1000} X_i^2.$$

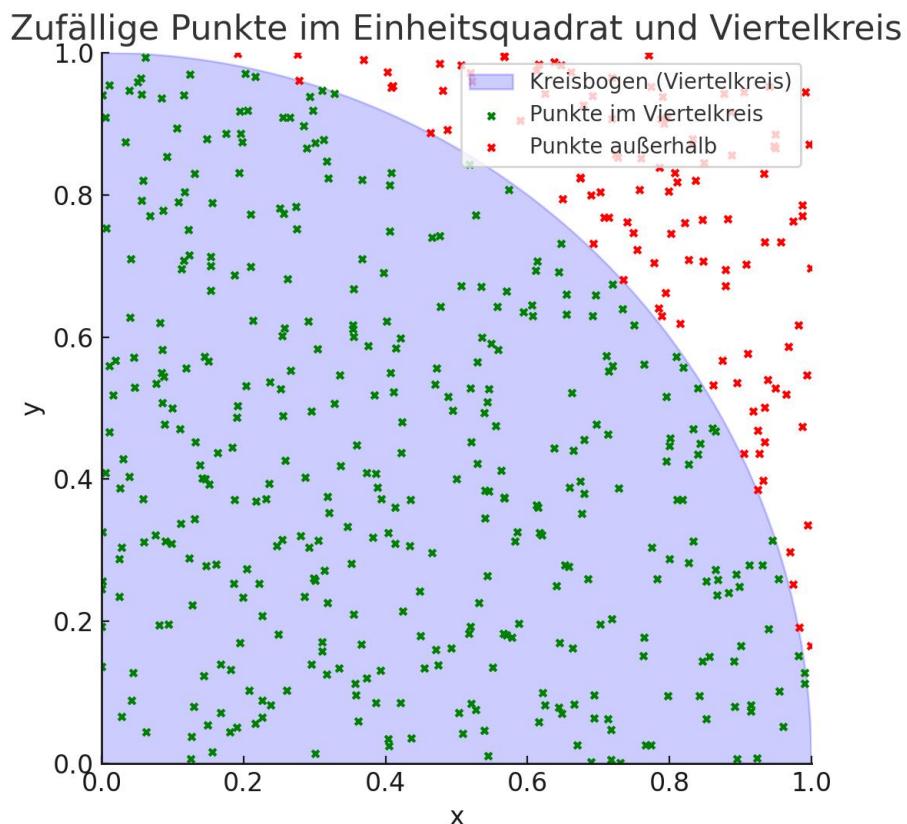
Angenommen, die Berechnung liefert $\hat{I}_{1000} \approx 0.3332$, so nähert sich diese Schätzung dem wahren Wert $I = \frac{1}{3} \approx 0.3333$ mit wachsendem n .

Fazit

Die Monte-Carlo-Integration ist eine flexible Methode zur numerischen Berechnung von Integralen, insbesondere in Fällen, in denen analytische Lösungen schwierig oder unmöglich sind. Dank des Schwachen Gesetzes der großen Zahlen wird der Fehler mit wachsendem n systematisch reduziert. Ihre Effizienz macht die Methode besonders nützlich in hohen Dimensionen und bei komplexen Integranden.

6.2.2 Beispiel: Berechnung von π durch wiederholten Münzwurf

Zur Approximation der Kreiszahl π kann man zufällig verteilte Punkte im Einheitsquadrat $[0, 1] \times [0, 1]$ verwenden. Sei (X, Y) ein Punkt, dessen Koordinaten unabhängig und gleichverteilt auf $[0, 1]$ sind. Berechnet man die Distanz des Punktes zum Ursprung, gehört der Punkt in den Viertelkreis (innerhalb des Kreises mit Radius 1 und Zentrum im Ursprung), wenn $X^2 + Y^2 \leq 1$. Wiederholt man dieses Experiment n -mal, dann ist der Anteil der Punkte, die im Viertelkreis liegen, näherungsweise gleich der Fläche des Viertelkreises ($\pi/4$). Somit kann man π approximieren durch:



Mit wachsendem n wird die Schätzung präziser, da die Monte-Carlo-Methode auf dem Gesetz der großen Zahlen beruht.

$$\pi \approx 4 \cdot \frac{\text{Anzahl der Punkte im Viertelkreis}}{\text{Gesamtanzahl der Punkte}}.$$

Erzeugung zufälliger Punkte auf dem Einheitsquadrat durch Münzwürfe

Man kann zufällige Punkte (X, Y) im Einheitsquadrat $[0, 1] \times [0, 1]$ mithilfe einer Folge unabhängiger Würfe einer fairen Münze erzeugen, indem man die Binärdarstellung der x - und y -Koordinaten aus den Münzwürfen konstruiert. Dies basiert auf der dyadischen Darstellung.

Schrittweises Vorgehen

1. Dyadische Darstellung: Jeder Punkt $x \in [0, 1)$ kann als eine unendliche Reihe in der Form

$$x = \sum_{i=1}^{\infty} B_i \cdot 2^{-i},$$

dargestellt werden, wobei $B_i \in \{0, 1\}$ die Binärziffern von x sind.

2. Erzeugung der x -Koordinate: - Verwende die Ergebnisse einer unendlichen Folge von Münzwürfen B_1, B_2, B_3, \dots , um die Binärziffern für die x -Koordinate zu erzeugen. - Falls die Münze 'Kopf' zeigt, setze $B_i = 1$; andernfalls $B_i = 0$.

3. Erzeugung der y -Koordinate: - Nutze eine unabhängige Folge von Münzwürfen C_1, C_2, C_3, \dots , um die Binärziffern für die y -Koordinate zu erzeugen. - Auch hier entspricht 'Kopf' der Ziffer 1 und 'Zahl' der Ziffer 0.

4. Kombinieren der Koordinaten: - Die x -Koordinate wird aus der Folge (B_1, B_2, \dots) gebildet:

$$x = B_1 \cdot 2^{-1} + B_2 \cdot 2^{-2} + \dots$$

- Die y -Koordinate wird aus der Folge (C_1, C_2, \dots) gebildet:

$$y = C_1 \cdot 2^{-1} + C_2 \cdot 2^{-2} + \dots$$

5. Ergebnis: - Der zufällige Punkt im Einheitsquadrat ist dann:

$$(X, Y) = \left(\sum_{i=1}^{\infty} B_i \cdot 2^{-i}, \sum_{i=1}^{\infty} C_i \cdot 2^{-i} \right).$$

Eigenschaften

Da die Münzwürfe unabhängig und gleichverteilt sind, sind die resultierenden x - und y -Koordinaten ebenfalls unabhängig und gleichverteilt auf $[0, 1)$. - Die Punkte (X, Y) sind somit gleichverteilt auf dem Einheitsquadrat $[0, 1] \times [0, 1]$.

Anwendung

Dieses Verfahren verdeutlicht, wie man durch eine Sequenz von diskreten Zufallsentscheidungen (Münzwürfen) kontinuierlich verteilte Zufallsvariablen und mehrdimensionale Zufallsvektoren erzeugen kann. Es wird oft als Grundlage für Simulationen in Monte-Carlo-Methoden verwendet.

Kapitel 7

Mehr zu Verteilungen von Zufallsvariablen

7.1 Vektorwertige Zufallsvariablen

Der Begriff der Zufallsvariablen verallgemeinert sich auf vektorwertige Abbildungen

$$Z : \Omega \mapsto \mathbb{R}^d$$

mit $d \geq 1$. Entsprechend redet man auch hier von Verteilungen: Sei Z eine diskrete Zufallsvariable auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{F}, P) , die Werte im zweidimensionalen Raum \mathbb{R}^2 annimmt. Die **Verteilung** von Z ist definiert als die Menge:

$$\mathcal{D}_Z = \{(z, p_z) \mid z \in Z(\Omega), p_z = P(Z = z)\},$$

wobei:

- $z = (x, y)$ die möglichen Werte von Z (auch *Realisationen* oder *Träger der Verteilung*) sind,
- $p_z = P(Z = z)$ die Wahrscheinlichkeit ist, dass Z den Wert z annimmt,
- die Wahrscheinlichkeiten die Eigenschaften erfüllen:

$$p_z \geq 0 \quad \text{für alle } z, \quad \text{und} \quad \sum_z p_z = 1.$$

Analog zum reellwertigen Fall $d = 1$ definiert man Erwartungswert und Varianz einer Vektorwertigen Zufallsvariablen

$$\mathbb{E}(Z) = \sum_{z \in Z(\Omega)} z \cdot p_z = m_Z \in \mathbb{R}^d,$$

$$\mathbb{V}(Z) = \sum_{z \in Z(\Omega)} |z|^2 \cdot p_z - |m_Z|^2,$$

wobei $|z| = \sqrt{\sum_{i=1}^d z_i^2}$ die euklidische Länge des Vektors $z \in \mathbb{R}^d$ bezeichnet.

Bemerkung.

Sind die Koordinaten von X , jeweils aufgefasst als reellwertige Zufallsvariablen, paarweise unkorreliert, so ist die Varianz von Z gleich der Summe der Varianzen der Koordinaten. Im Allgemeinen gilt diese Gleichheit jedoch nicht.

Beispiel:

Es werden zwei faire Würfel gleichzeitig geworfen, die Zufallsvariable Z bezeichne das Paar aus Augensumme und Augendifferenz (Absolutwert) der beiden Würfel.

Punkt z	Wahrscheinlichkeit
(2, 0)	$\frac{1}{36}$
(3, 1)	$\frac{2}{36}$
(4, 0)	$\frac{1}{36}$
(4, 2)	$\frac{2}{36}$
(5, 1)	$\frac{2}{36}$
(5, 3)	$\frac{2}{36}$
(6, 0)	$\frac{1}{36}$
(6, 2)	$\frac{2}{36}$
(6, 4)	$\frac{2}{36}$
(7, 1)	$\frac{2}{36}$
(7, 3)	$\frac{2}{36}$
(7, 5)	$\frac{2}{36}$
(8, 0)	$\frac{1}{36}$
(8, 2)	$\frac{2}{36}$
(8, 4)	$\frac{2}{36}$
(9, 1)	$\frac{2}{36}$
(9, 3)	$\frac{2}{36}$
(10, 0)	$\frac{1}{36}$
(10, 2)	$\frac{2}{36}$
(11, 1)	$\frac{2}{36}$
(12, 0)	$\frac{1}{36}$

$$\mathbb{E}[Z] = \left(7, \frac{35}{18} \right),$$

$$\mathbb{V}(Z) = \frac{2555}{324}.$$

7.2 Gemeinsame Verteilung

Es seien X_1, \dots, X_k Zufallsvariablen auf einem diskreten Wahrscheinlichkeitsraum (Ω, P) gegeben. Dann heißt die Verteilung \mathcal{D}_Z der vektorwertigen Zufallsvariable $Z = (X_1, \dots, X_k) : \Omega \rightarrow \mathbb{R}^k$ die **gemeinsame Verteilung** von X_1, \dots, X_k .

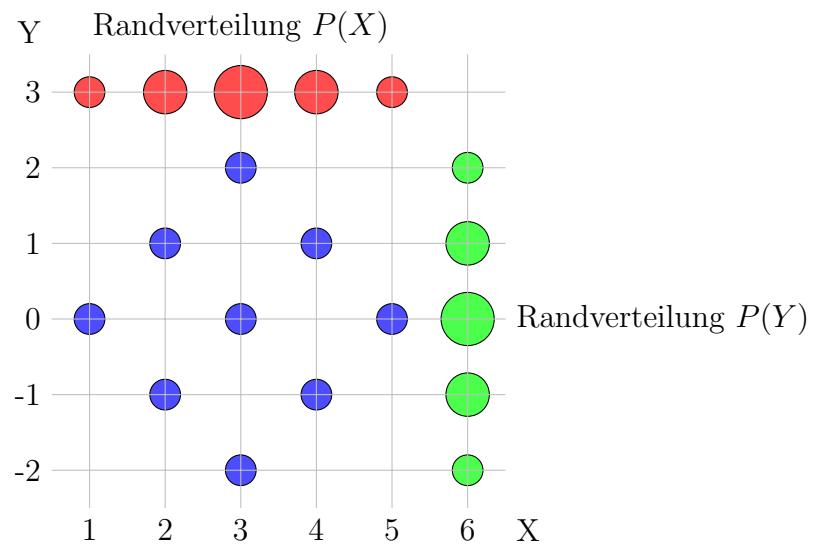
7.2.1 Spezialfall $k = 2$: Kreuztabellen

Im Fall $d = 2$ ist es häufig praktisch, die gemeinsame Verteilung in Form von Kreuztabelen anzugeben. Durch Summation entlang den Zeilen bzw. Spalten lassen sich hieraus Verteilungen der beiden eingehenden Zufallsvariablen bei isolierter Betrachtung ermitteln.

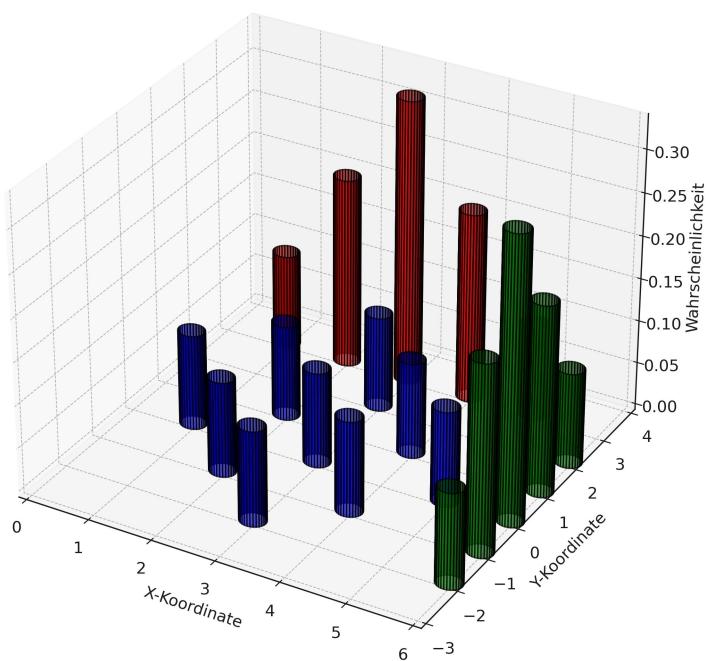
Beispiel: Rhombische Kugelanordnung (Forts.)

Für X als x -Koordinate und Y als y -Koordinate einer zufällig aus der rhombischen Anordnung gezogenen Kugel ergibt sich die folgende Kreuztabelle.

$X \setminus Y$	$y = -2$	$y = -1$	$y = 0$	$y = 1$	$y = 2$	$P(X = x)$
$x = 1$	0	0	$\frac{1}{9}$	0	0	$\frac{1}{9}$
$x = 2$	0	$\frac{1}{9}$	0	$\frac{1}{9}$	0	$\frac{1}{9}$
$x = 3$	$\frac{1}{9}$	0	$\frac{1}{9}$	0	$\frac{1}{9}$	$\frac{1}{9}$
$x = 4$	0	$\frac{1}{9}$	0	$\frac{1}{9}$	0	$\frac{1}{9}$
$x = 5$	0	0	$\frac{1}{9}$	0	0	$\frac{1}{9}$
$P(Y = y)$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{1}{3}$	$\frac{2}{9}$	$\frac{1}{9}$	1



3D-Zylinderdarstellung der gemeinsamen Verteilung von X und Y mit Randverteilungen



7.2.2 Produktverteilung und Stochastische Unabhängigkeit

Die stochastische Unabhängigkeit von X_1, \dots, X_k lässt sich mithilfe der gemeinsamen Verteilung durch die Eigenschaft

$$p_z = p_{z_1}^1 \cdots p_{z_k}^k$$

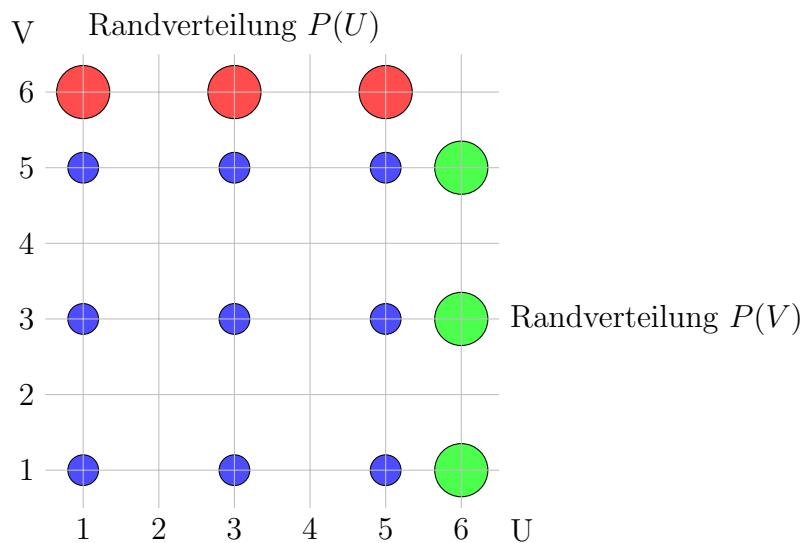
für alle $z = (z_1, \dots, z_k) \in Z(\Omega)$ ausdrücken, wobei $p_{z_i}^i = P(X_i = z_i)$. Die Wahrscheinlichkeiten der gemeinsamen Verteilung in \mathbb{R}^k ergeben sich also einfach aus der Multiplikation der zugehörigen Wahrscheinlichkeiten der betrachteten reellwertigen Zufallsvariablen, man spricht in einem solchen Fall von einer **Produktverteilung**.

Im Fall $k = 2$ heißt das, dass sich bei Unabhängigkeit die Einträge in der Kreuztabelle gerade durch Produktbildung der entsprechenden ('Rand')-Wahrscheinlichkeiten ergeben.

Beispiel: Rhombische Kugelanordnung (Forts.)

Bei der Rhombischen Kugelanordnung sind die x - und y -Koordinate als Zufallsvariablen nicht stochastisch unabhängig, wie etwa das Beispiel $z = (2, 1)$ zeigt in der Tabelle oben zeigt. Betrachtet man hingegen die Zufallsvariablen $U = X + Y$ und $V = X - Y$ so sind diese beiden stochastisch unabhängig:

$U \setminus V$	$V = 1$	$V = 3$	$V = 5$	$P(U = u)$
$U = 1$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{3}$
$U = 3$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{3}$
$U = 5$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{3}$
$P(V = v)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1



7.3 Diskrete Verteilungen als Wahrscheinlichkeitsmaße auf \mathbb{R}^d

Äquivalenz von diskreten Verteilungen und diskreten Zufallsvariablen

Wir definieren eine diskrete Verteilung \mathcal{D} auf \mathbb{R}^d als eine abzählbare Familie von Punkten im Raum, versehen mit nichtnegativen Gewichten, welche sich zu eins summieren, d.h.

$$\mathcal{D} = \{(z_i, p_i) | z_i \in \mathbb{R}^d, p_i \in [0, 1], i \in \mathbb{N}, \sum_{i=1}^{\infty} p_i = 1\}.$$

\mathcal{D} kann aufgefasst werden als eine Beschreibung eines Zufallsmechanismus, welcher einen Punkt im Raum zufällig wählt. Hierbei werden nur Punkte aus der Menge $\{p_i\}$ gewählt, und zwar jeweils mit Wahrscheinlichkeit $p_i \in [0, 1]$.

Wir haben gesehen, dass eine vektorwertige Zufallsvariable $Z : \Omega \rightarrow \mathbb{R}^d$ eine diskrete Verteilung $D = D_Z$ induziert. Umgekehrt lässt sich jede diskrete Verteilung so gewinnen. Dazu setzt man zu gegebener diskreter Verteilung \mathcal{D} als $\Omega = \{z_i | (z_i, p_i) \in \mathcal{D}\} \subset \mathbb{R}^d$ und definiert ein Wahrscheinlichkeitsmaß auf Ω durch

$$P_{\mathcal{D}}(A) = \sum_{i: z_i \in A} p_i \text{ für } A \subset \mathbb{R}^d.$$

Die identische Abbildung $Z : \Omega \rightarrow \mathcal{R}^d$, $\omega = x \rightarrow Z(\omega) = x$, aufgefasst als vektorwertige Zufallsvariable hat dann $\mathcal{D} = \mathcal{D}_Z$ als Verteilung. Für diese Zufallsvariable gilt dann

$$\mathbb{P}(Z \in A) = \sum_{i: z_i \in A} p_i = P_{\mathcal{D}}(A).$$

In der Tat definiert diese Vorschrift ein Wahrscheinlichkeitsmaß auf $\Omega' = \mathbb{R}^d$ versehen mit der σ -Algebra $\mathcal{F}' = 2^{\mathbb{R}^d}$, welches auf einer abzählbaren Menge konzentriert ist. Diskrete Zufallsvariablen lassen sich also als diskrete Verteilungen bzw. Wahrscheinlichkeitsmaße auf \mathbb{R}^d interpretieren und umgekehrt, beides sind äquivalente Konzepte zur mathematischen Beschreibung der Wahl eines zufälligen Punktes im Raum, sofern die Menge der möglichen Punkte höchstens abzählbar ist.

Insbesondere kann man jeder Verteilung kann man einen Erwartungswert, die Varianz etc. nach den bekannten Formeln zuweisen. Anschaulich ist der Erwartungswert der Schwerpunkt des Ensembles $\{(z_i, p_i)\}$ von Masseteilchen, die in den Punkten z_i mit Gewicht p_i platziert sind.

7.4 Verteilungsfunktion einer Zufallsvariablen

Die **Verteilungsfunktion** $F_X : \mathbb{R} \rightarrow [0, 1]$ einer diskreten Zufallsvariablen X beschreibt die kumulative Wahrscheinlichkeit, dass X einen Wert kleiner oder gleich x annimmt:

$$F_X(x) = P(X \leq x) = \sum_{x_i \leq x} p_i,$$

wobei $\{x_i\}$ die möglichen Werte von X und $p_i = P(X = x_i)$ deren zugehörige Wahrscheinlichkeiten sind.

Eigenschaften

- $F_X(x)$ ist monoton wachsend: $F_X(x_1) \leq F_X(x_2)$ für $x_1 \leq x_2$.
- $F_X(x)$ ist rechtsseitig stetig.
- $\lim_{x \rightarrow -\infty} F_X(x) = 0$ und $\lim_{x \rightarrow \infty} F_X(x) = 1$.

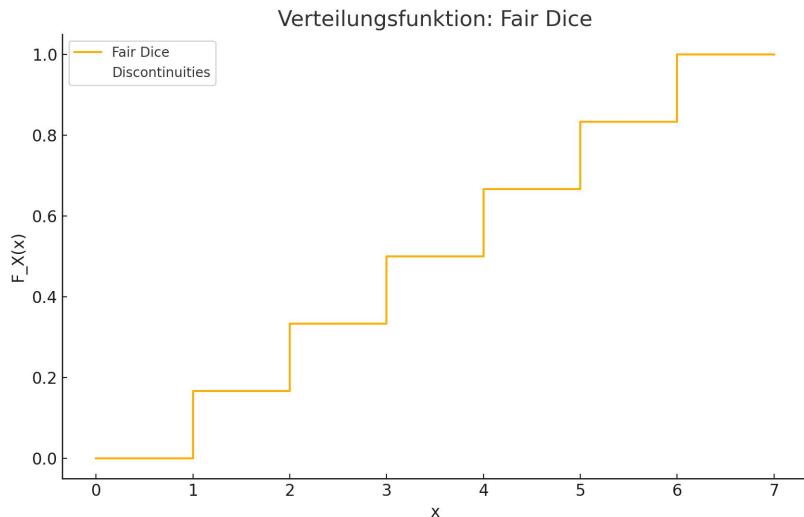
Beispiele

1. Endlicher Wertebereich (Wurf eines fairen Würfels):

$$X \in \{1, 2, 3, 4, 5, 6\}, \quad P(X = x) = \frac{1}{6}.$$

Die Verteilungsfunktion ist:

$$F_X(x) = \begin{cases} 0, & x < 1, \\ \frac{k}{6}, & k \leq x < k+1, \quad k \in \{1, 2, \dots, 6\}, \\ 1, & x \geq 6. \end{cases}$$

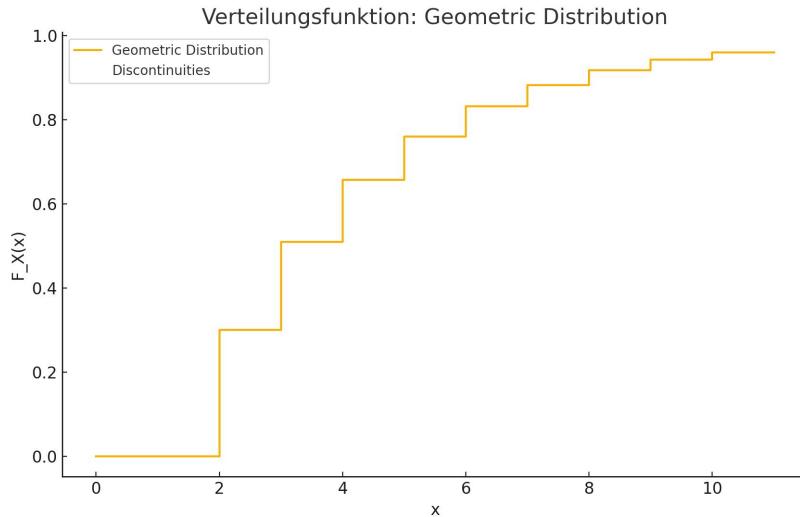


2. Unendlicher Wertebereich (Geometrische Verteilung):

$$X \in \mathbb{N}, \quad P(X = k) = (1-p)^{k-1}p, \quad 0 < p \leq 1.$$

Die Verteilungsfunktion ist:

$$F_X(x) = 1 - (1-p)^{\lfloor x \rfloor}, \quad x \geq 1.$$

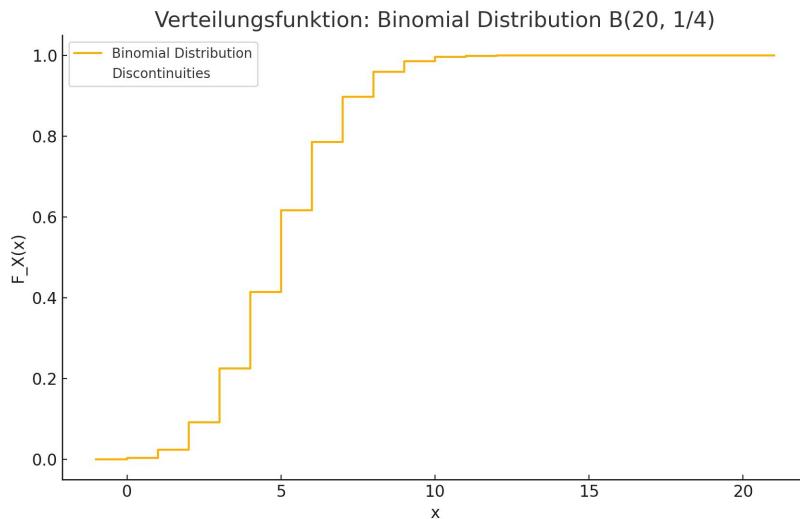


3. Binomialverteilung $B(20, \frac{1}{4})$:

$$X \in \{0, 1, \dots, 20\}, \quad P(X = k) = \binom{20}{k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{20-k}.$$

Die Verteilungsfunktion summiert die Wahrscheinlichkeiten bis x :

$$F_X(x) = \sum_{k \leq x} P(X = k).$$



7.5 Verteilungsfunktion einer diskreten vektorwertigen Zufallsvariablen

Sei $\mathbf{X} = (X_1, X_2, \dots, X_d)$ eine d -dimensionale diskrete Zufallsvariable auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{F}, P) . Die Verteilungsfunktion von \mathbf{X} ist definiert als:

$$F_{\mathbf{X}}(\mathbf{x}) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_d \leq x_d),$$

wobei $\mathbf{x} = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$. Sie gibt die Wahrscheinlichkeit an, dass jede Komponente von \mathbf{X} den entsprechenden Wert aus \mathbf{x} nicht überschreitet.

Eigenschaften

- $F_{\mathbf{X}}(\mathbf{x})$ ist monoton wachsend in jedem Argument x_i .
- $\lim_{\mathbf{x} \rightarrow -\infty} F_{\mathbf{X}}(\mathbf{x}) = 0$.
- $\lim_{\mathbf{x} \rightarrow \infty} F_{\mathbf{X}}(\mathbf{x}) = 1$, wobei die Grenze in jedem Argument genommen wird.
- Es gilt $F_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^d F_{\mathbf{X}_i}(x_i)$ für alle $\mathbf{x} = (x_1, \dots, x_d)$ genau dann, wenn X_1, \dots, X_n stochastisch unabhängig sind.

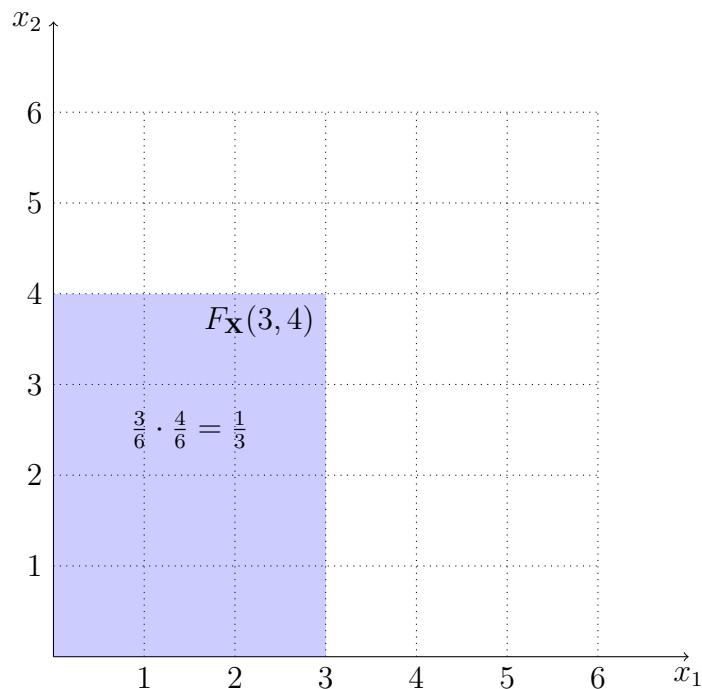
Beispiel

Betrachten wir die gemeinsame Verteilung von X_1 und X_2 , den Ergebnissen zweier Würfe eines fairen Würfels. Die Zufallsvariable $\mathbf{X} = (X_1, X_2)$ hat folgende Verteilungsfunktion:

$$F_{\mathbf{X}}(x_1, x_2) = P(X_1 \leq x_1, X_2 \leq x_2).$$

Da $X_1, X_2 \in \{1, 2, 3, 4, 5, 6\}$ gleichverteilt und unabhängig sind, ergibt sich:

$$F_{\mathbf{X}}(x_1, x_2) = \begin{cases} 0, & \text{falls } x_1 < 1 \text{ oder } x_2 < 1, \\ \frac{\lfloor x_1 \rfloor}{6} \cdot \frac{\lfloor x_2 \rfloor}{6}, & \text{falls } 1 \leq x_1, x_2 \leq 6, \\ 1, & \text{falls } x_1 \geq 6 \text{ und } x_2 \geq 6. \end{cases}$$



7.6 Beispiel: Empirische Verteilung von Stichproben

Die **empirische Verteilung** einer Stichprobe lässt sich in direkter Analogie zur Verteilung einer diskreten Zufallsvariablen definieren. Sei $\{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n\}$ eine Stichprobe, wobei jede Beobachtung $\mathbf{X}_i = (X_{i1}, X_{i2}, \dots, X_{id}) \in \mathbb{R}^d$ ein d -dimensionaler Vektor ist. Hieraus leitet sich die empirische Verteilung

$$\mathcal{D}_X = \{(\mathbf{X}_i, \frac{1}{n}) \mid i = 1, \dots, n\}$$

mit zugehöriger Verteilungsfunktion F_X .

$$F_X(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n 1_{\{\mathbf{X}_i \leq \mathbf{x}\}},$$

wobei $1_{\{\mathbf{X}_i \leq \mathbf{x}\}}$ die Indikatorfunktion ist, die 1 annimmt, wenn $X_{ij} \leq x_j$ für alle $j = 1, \dots, d$, und 0 sonst.

Erwartungswert und Varianz dieser Verteilung entsprechen dann dem empirischen Mittel bzw. der mittleren quadratischen Streuung der Stichprobe.

7.7 Zufallsvariablen mit Dichtefunktion

7.7.1 Kontinuierliche vs. Diskrete Zufallsvariablen

Zufallsvariablen lassen sich in zwei Hauptkategorien einteilen: *diskrete* und *kontinuierliche* Zufallsvariablen. Allgemein nennen wir eine Zufallsvariable diskret, wenn es eine abzählbare Menge $Z = \{z_1, z_2, \dots\} = \{z_i \mid i \in \mathbb{N}\} \subset \mathbb{R}$ gibt, so dass $P(X \in Z) = 1$. Können wir eine solche abzählbare Menge nicht finden, so nennen wir X eine kontinuierliche Zufallsvariable.

7.7.2 Beispiel: Gleichverteilung auf $[0, 1] \subset \mathbb{R}$

Die natürliche kontinuierliche Erweiterung etwa eines Laplace-Experimentes auf den Fall eines Zufallsexperimentes mit überabzählbar vielen 'gleichwahrscheinlichen' Ausgängen ist im einfachsten Fall die sogenannte **Gleichverteilung auf dem Einheitsintervall**. Demnach ist eine Zufallsvariable Z **uniform auf $[0, 1]$ verteilt**, falls

$$\mathbb{P}(Z \in [\alpha, \beta]) = \beta - \alpha$$

für $0 \leq \alpha \leq \beta \leq 1$. Aus dieser Forderung ergibt sich dann sofort, dass allgemein

$$\mathbb{E}(f(Z)) = \int_0^1 f(t)dt,$$

für jede stetige Funktion f gelten muss. – Intuitiv stellen uns Z als 'Grenzwert' von Zufallsvariablen Z_n vor, die jeweils uniform, d.h. Laplace'sch auf der Menge $\Omega_n = \{\frac{l}{n} \mid l = 1, \dots, n\}$ verteilt sind. Und in der Tat zeigt die elementare Theorie des Integrals, dass

$$\mathbb{E}(f(Z_n)) = \sum_{l=1}^n f\left(\frac{l}{n}\right) \longrightarrow \int_0^1 f(t)dt$$

für alle stetigen Funktionen f . In der Sprache der zugehörigen Verteilungen von Z_n

$$\mathcal{D}_n = \left\{ \left(\frac{l}{n}, \frac{1}{n} \right) \mid l = 1, \dots, n \right\}$$

sagen wir hierfür, dass

$$\mathcal{D}_n \implies \mathcal{U},$$

wobei \mathcal{U} das *uniforme Wahrscheinlichkeitsmaß* auf $[0, 1]$ bezeichnet und das auf Intervallen der Form $[a, b]$ durch

$$\mathcal{U}([a, b]) = b - a$$

festgelegt ist. In der Maßtheorie lernen wir \mathcal{U} als das 1-dimesionale Lebesgue-Mass auf dem Einheitsintervall kennen, das den naiven Längen- bzw. Volumenbegriff auf von Intervallen bzw. Quadern auf allgemeinere 'messbare' Mengen erweitert.

Um eine Zufallsvariable mit Verteilung \mathcal{U} direkt zu konstruieren, wählen wir

$$Z = \sum_{k=1}^{\infty} \xi_k \frac{1}{2^k},$$

wobei $(\xi^k)_{k \in \mathbb{N}}$ eine Folge von stochastisch unabhängigen *Bernoulli- $\frac{1}{2}$ -Zufallsvariablen* ist, d.h. mit $P(\xi^k = 1) = P(\xi^k = -1) = \frac{1}{2}$. Die Zufallsvariable Z nimmt überabzählbar viele Werte an, gleichzeitig gilt $P(Z = t) = 0$ für jedes $t \in [a, b]$. Die Eigenschaft

$$\mathbb{P}(Z \in [\alpha, \beta]) = \beta - \alpha$$

lässt sich mit elementairen Überlegungen leicht überprüfen.

7.7.3 Kontinuierliche Zufallsvariablen mit Dichtefunktion

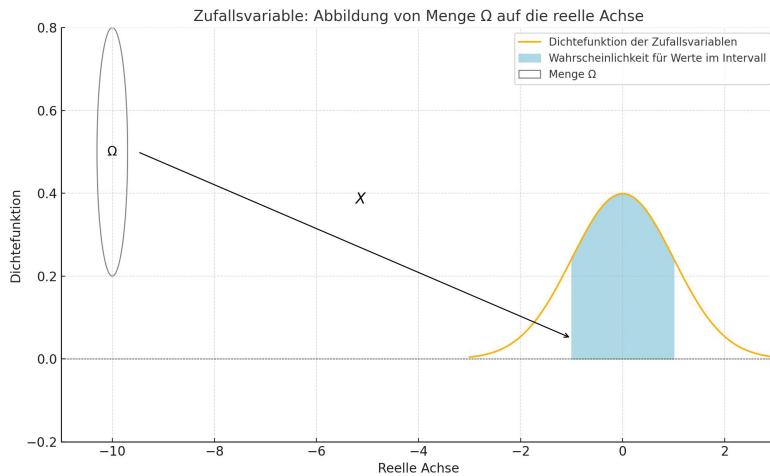
Bisher haben wir ausschließlich diskrete Zufallsvariablen behandelt. Diskrete Zufallsvariablen nehmen nur abzählbare Werte an, wie z. B. ganze Zahlen. Ihre Verteilung wird durch eine *Wahrscheinlichkeitsmassefunktion* (PMF, probability mass function) beschrieben, die jedem möglichen Wert der Zufallsvariablen eine Wahrscheinlichkeit zuordnet, z. B. $P(X = k)$. Ein klassisches Beispiel ist die Anzahl geworfener Sechsen beim Würfeln.

Kontinuierliche Zufallsvariablen hingegen können grundsätzlich jeden Wert aus einem Intervall annehmen, z. B. $X \in [a, b]$ oder $X \in \mathbb{R}$. Da in diesem Fall die Wahrscheinlichkeit eines einzelnen Werts typischer Weise $P(X = x)$ null ist, kann die Verteilung nicht durch eine Massefunktion beschrieben werden. In den meisten Fällen kann man dann die Wahrscheinlichkeit über Intervalle mithilfe einer *Wahrscheinlichkeitsdichtefunktion* (PDF, probability density function) angegeben. Diese Dichtefunktion $f(x)$ ist so definiert, dass die Wahrscheinlichkeit, dass die Zufallsvariable in einem Intervall liegt, durch das Integral berechnet wird:

$$P(a \leq X \leq b) = \int_a^b f(x) dx.$$

Falls die Beschreibung der Zufallsvariablen X mithilfe einer solchen Funktion möglich ist, sprechen wir von einer Zufallsvariablen mit Dichte.

Eine *Wahrscheinlichkeitsdichtefunktion* (PDF, probability density function) beschreibt die relative Wahrscheinlichkeit, dass eine Zufallsvariable einen bestimmten Wert annimmt. Für stetige Zufallsvariablen ist die PDF so definiert, dass das Integral über einen Bereich die Wahrscheinlichkeit liefert, dass die Zufallsvariable in diesen Bereich fällt.



7.7.4 Die Verteilungsfunktion und ihr Zusammenhang mit der Dichtefunktion

Die Verteilung einer kontinuierlichen Zufallsvariablen kann auch durch die *Verteilungsfunktion* (CDF, cumulative distribution function) $F(x)$ beschrieben werden. Diese ist definiert als die Wahrscheinlichkeit, dass die Zufallsvariable X höchstens den Wert x annimmt:

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt.$$

Hierbei ist $f(t)$ die Wahrscheinlichkeitsdichtefunktion der Zufallsvariablen X .

Der Zusammenhang zwischen der Dichtefunktion $f(x)$ und der Verteilungsfunktion $F(x)$ ergibt sich aus dem *Hauptsatz der Integral- und Differentialrechnung*:

$$f(x) = \frac{d}{dx} F(x).$$

Das bedeutet, dass die Dichtefunktion die Ableitung der Verteilungsfunktion ist, während die Verteilungsfunktion das Integral der Dichtefunktion darstellt.

Anschaulich gesprochen beschreibt die Dichtefunktion die 'lokale Änderungsrate' der Verteilungsfunktion. Die Verteilungsfunktion selbst gibt die kumulierte Wahrscheinlichkeit bis zu einem bestimmten Wert x an, während die Dichtefunktion die Wahrscheinlichkeit beschreibt, dass die Zufallsvariable X in der unmittelbaren Nähe dieses Werts liegt.

Durch diese Beziehung können sowohl die Wahrscheinlichkeit für bestimmte Intervalle als auch die gesamte Verteilung der Zufallsvariablen einheitlich beschrieben werden.

Bemerkung zur Allgemeinheit der Verteilungsfunktion

Es ist hervorzuheben, dass die Verteilungsfunktion $F(x)$ für alle Zufallsvariablen existiert, unabhängig davon, ob sie diskret, kontinuierlich oder eine Mischung aus beiden sind. Auch diskrete Zufallsvariablen, die keine Dichtefunktion besitzen, haben eine Verteilungsfunktion, welche die kumulierte Wahrscheinlichkeit $F(x) = P(X \leq x)$ beschreibt.

Im Folgenden betrachten wir einige klassische Beispiele für reelle Zufallsvariablen mit verschiedenen Wahrscheinlichkeitsdichtefunktionen: die *Standardnormalverteilung*, die *Exponentialverteilung* und die *Beta-Verteilung*.

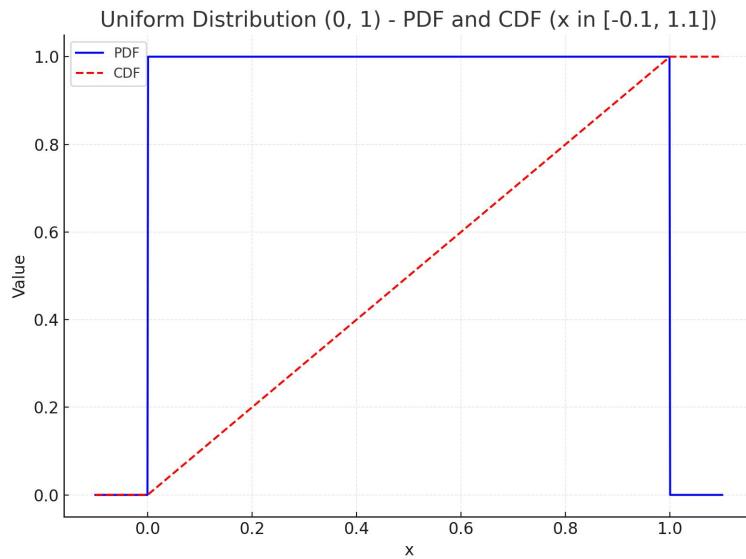
7.7.5 Beispiele

Gleichverteilung auf dem Einheitsintervall

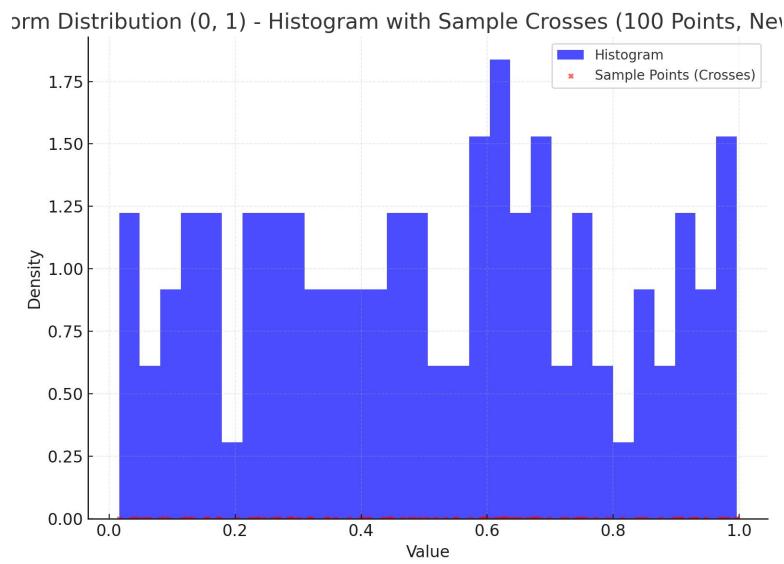
Die Gleichverteilung auf dem Intervall $[0, 1]$ hat die folgende Wahrscheinlichkeitsdichtefunktion:

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

Diese Verteilung beschreibt den Fall, dass alle Werte im Intervall $[0, 1]$ mit gleicher Wahrscheinlichkeit auftreten. Die Graphen von Dichte- und Verteilungsfunktion sehen wie folgt aus.



Wir können 100 Punkte zufällig entsprechend dieser Verteilung als Stichprobe ziehen und das Histogramm der zugehörigen empirischen Verteilung erstellen. Das ergibt zum Beispiel das folgende Resultat

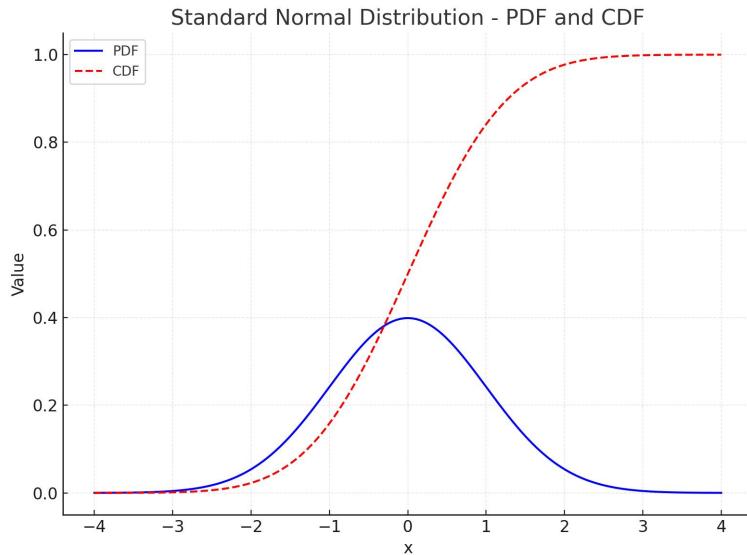


Standardnormalverteilung

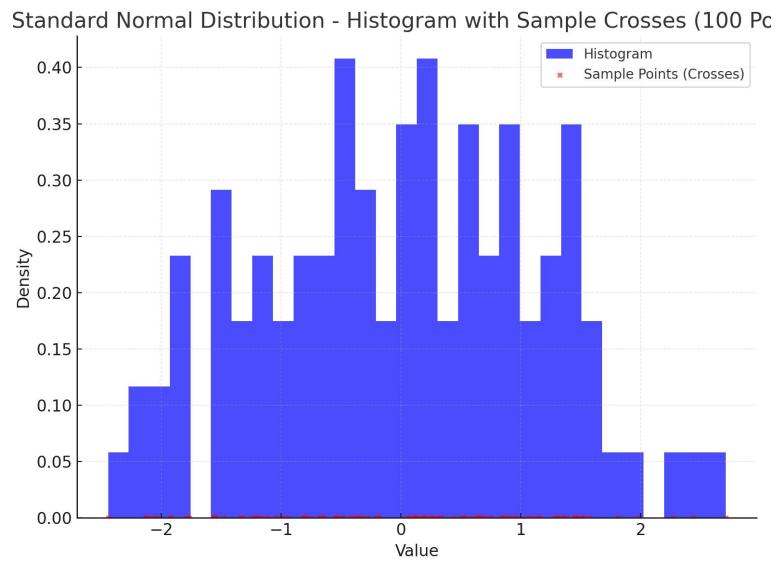
Die Standardnormalverteilung hat die folgende Wahrscheinlichkeitsdichtefunktion:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad x \in \mathbb{R}.$$

Die Zufallsvariable X folgt der Verteilung $X \sim N(0, 1)$, wobei der Mittelwert 0 und die Varianz 1 ist. Die Graphen der Dichte- und Verteilungsfunktion sehen wie folgt aus.



Wir können 100 Punkte zufällig entsprechend dieser Verteilung als Stichprobe ziehen und das Histogramm der zugehörigen empirischen Verteilung erstellen. Das ergibt zum Beispiel das folgende Resultat



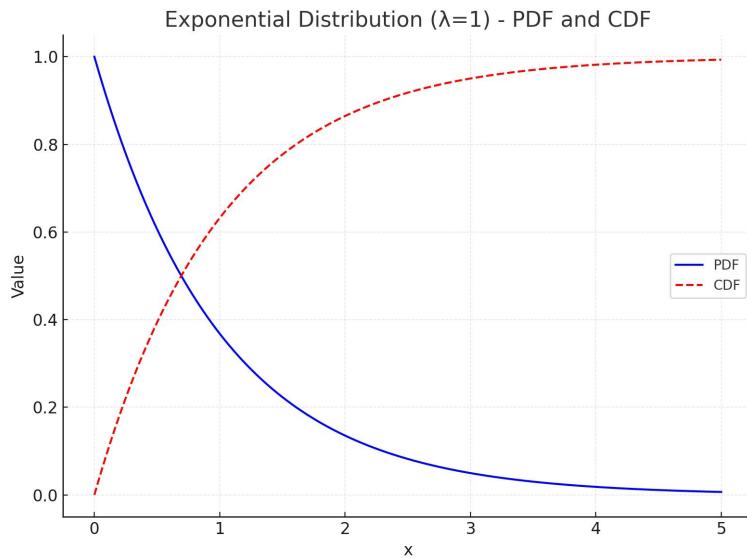
Exponentialverteilung

Die Exponentialverteilung hat die folgende Wahrscheinlichkeitsdichtefunktion:

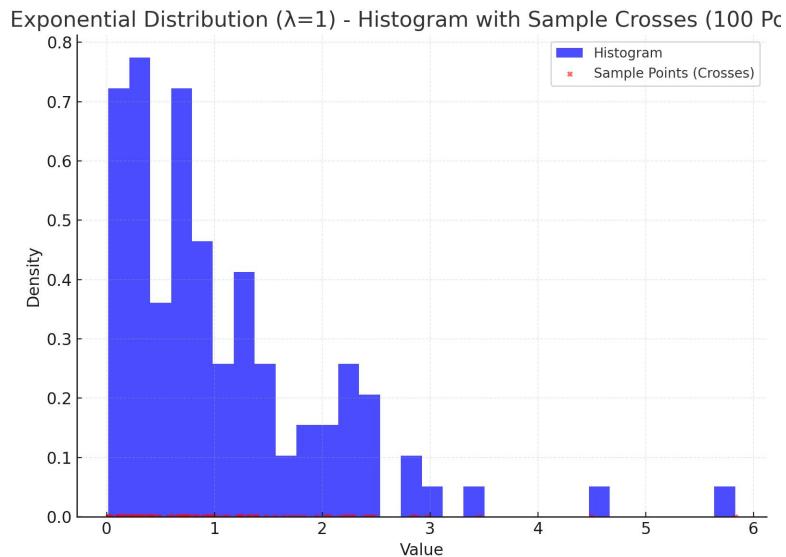
$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0, \end{cases}$$

wobei $\lambda > 0$ der Rate-Parameter ist. Die Zufallsvariable X folgt der Verteilung $X \sim \text{Exp}(\lambda)$. Im Beispiel verwenden wir $\lambda = 1$.

Die Graphen der Dichte- und Verteilungsfunktion sehen wie folgt aus.



Wir können 100 Punkte zufällig entsprechend dieser Verteilung als Stichprobe ziehen und das Histogramm der zugehörigen empirischen Verteilung erstellen. Das ergibt zum Beispiel das folgende Resultat



Beta-Verteilung

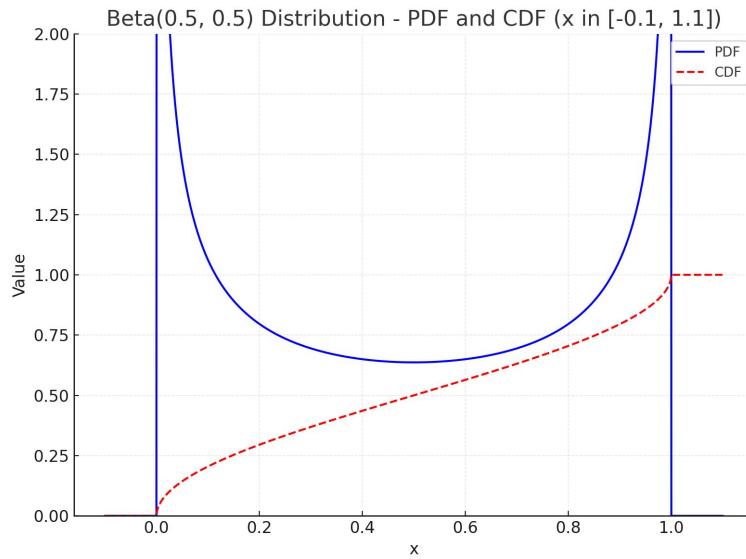
Die Beta-Verteilung hat die folgende Wahrscheinlichkeitsdichtefunktion:

$$f(x) = \begin{cases} \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}, & 0 \leq x \leq 1, \\ 0, & \text{sonst,} \end{cases}$$

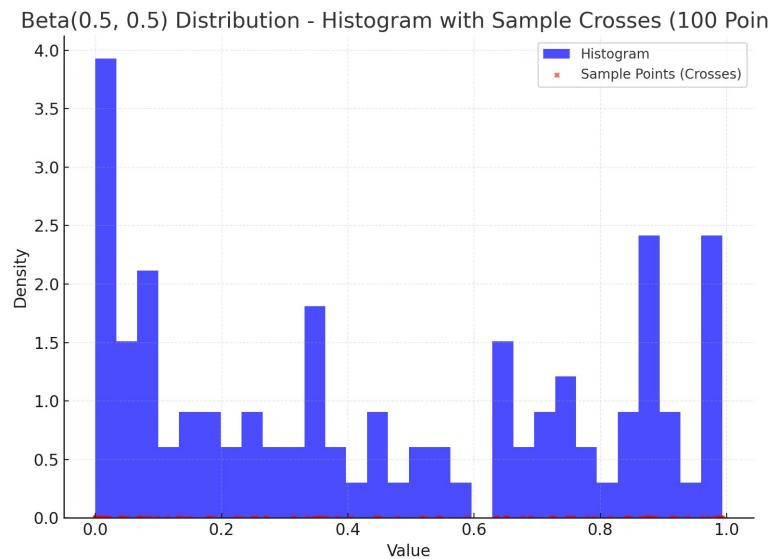
wobei $B(\alpha, \beta)$ die Beta-Funktion ist und $\alpha, \beta > 0$. Für $\alpha = \beta = 0.5$ lautet die Dichtefunktion:

$$f(x) = \frac{1}{\pi \sqrt{x(1-x)}}, \quad 0 < x < 1.$$

Die Graphen der Dichte- und Verteilungsfunktion sehen wie folgt aus.



Wir können 100 Punkte zufällig entsprechend dieser Verteilung als Stichprobe ziehen und das Histogramm der zugehörigen empirischen Verteilung erstellen. Das ergibt zum Beispiel das folgende Resultat



Aus dem Gesetz der Großen Zahlen folgt, dass sich die Histogramme von immer größeren Stichproben von wiederholten Ziehungen zu gegebener Dichte dem Graphen der Dichtefunktion asymptotisch angleichen.

7.8 Konstruktion bzw. Simulation einer Zufallsvariablen zu gegebener Verteilungsfunktion

Sei U eine Zufallsvariable mit einer gleichmäßigen Verteilung auf dem Intervall $[0, 1]$, also $U \sim \mathcal{U}(0, 1)$. Um eine Zufallsvariable X mit einer gegebenen Verteilungsfunktion $F_X(x)$

zu konstruieren, verwendet man die verallgemeinerte rechtsstetige Inverse der Verteilungsfunktion, definiert als:

$$F_X^{-1}(u) = \inf\{x \in \mathbb{R} \mid F_X(x) \geq u\}.$$

Dann setzt man:

$$X = F_X^{-1}(U).$$

Diese Konstruktion garantiert, dass X die Verteilungsfunktion $F_X(x)$ besitzt, da für jede $x \in \mathbb{R}$ gilt:

$$P(X \leq x) = P(F_X^{-1}(U) \leq x) = P(U \leq F_X(x)) = F_X(x),$$

weil U gleichverteilt ist. Diese Methode ist insbesondere in der Simulation von Zufallsvariablen nützlich, da sie eine einfache Transformation einer leicht generierbaren uniformen Zufallsvariable erlaubt.

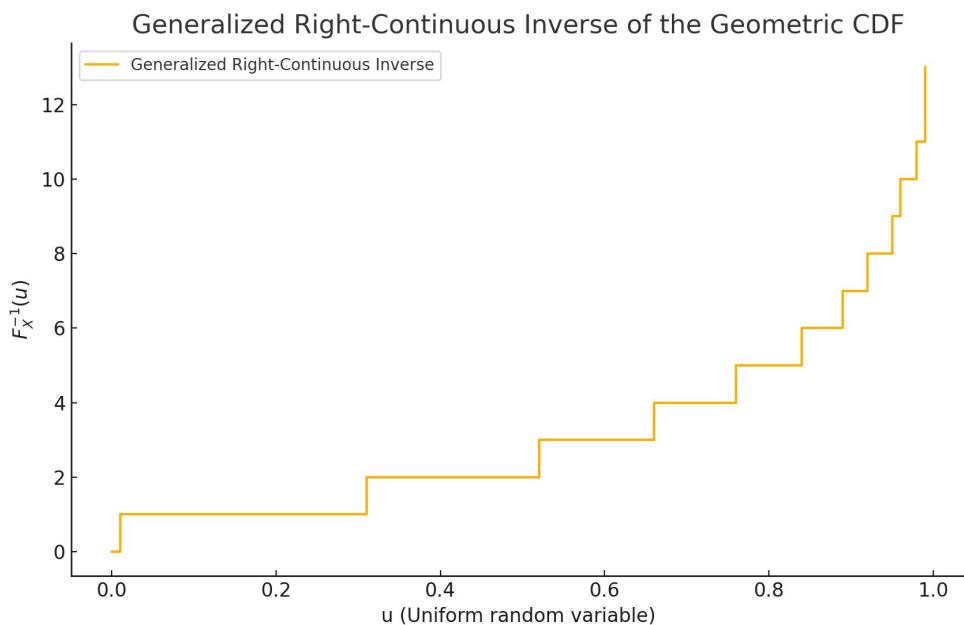
Beispiel 1. Verteilung mit endlichem Wertebereich

Für eine diskrete Zufallsvariable mit Werten $\{x_1, x_2, \dots, x_n\}$ und Wahrscheinlichkeiten $P(X = x_i) = p_i$, ergibt sich:

$$F_X^{-1}(u) = \begin{cases} x_1, & 0 \leq u < p_1, \\ x_2, & p_1 \leq u < p_1 + p_2, \\ \vdots \\ x_n, & \sum_{i=1}^{n-1} p_i \leq u \leq 1. \end{cases}$$

Beispiel 2: Erzeugung einer geometrisch verteilten Zufallsvariablen

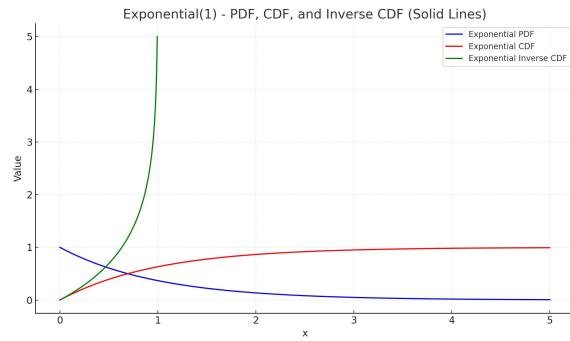
Der graph der verallgemeinerten Rechtsimversen einer nicht-fallenden Funktion ergibt sich wie bei invertierbaren Funktionen durch Spiegelung an der Winkelhalbierenden. Im Fall der Verteilungsfunktion der Geometrischen Verteilung ergibt sich der folgende Graph für F_X^{-1} .



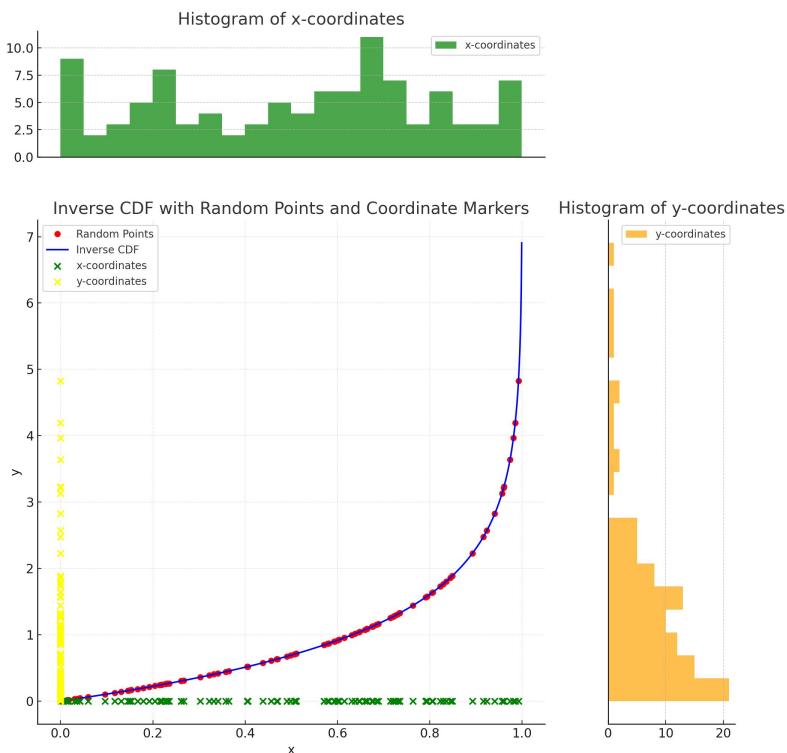
Falls U uniform auf $[0, 1]$ verteilt ist, ergibt sich für die neue Zufallsvariable $Z = F_X^{-1}(U)$ dann eine geometrische Verteilung.

Beispiel 3: Exponentiell Verteilte Zufallsvariable

Im Fall von Verteilungen mit strikt positiver Dichte ist die Verteilungsfunktion invertierbar. Im Fall der Exponentialverteilung können wir pdf, cdf und inverse cdf in einen gemeinsamen Plot eintragen.



Wir erhalten eine exponentiell verteilte Zufallsgröße $Z = F_X^{-1}(U)$, wenn U gleichverteilt auf dem Einheitsintervall gewählt wird. Die untenstehende Grafik verdeutlicht das am Beispiel von 100 uniform auf $[0, 1]$ gewählten Punkten (grün) und den zugehörigen Bildpunkten unter F_X^{-1} (gelb). Für die mithilfe der Funktion F_X^{-1} aus den grünen konstruierten gelben Punkten ergibt sich eine Exponentialverteilung.



Kapitel 8

Der Zentrale Grenzwertsatz

8.1 Aussage

Der Zentrale Grenzwertsatz (ZGS, engl. 'Central Limit Theorem') ist eines der fundamentalen Ergebnisse der Wahrscheinlichkeitstheorie. Er besagt, dass die Summe (oder der Mittelwert) einer großen Anzahl unabhängiger und identisch verteilter Zufallsvariablen, unabhängig von ihrer ursprünglichen Verteilung, approximativ einer Normalverteilung folgt.

Formalisiert gilt: Seien X_1, X_2, \dots, X_n unabhängige und identisch verteilte Zufallsvariablen mit Erwartungswert μ und Varianz σ^2 . Dann konvergiert die Verteilung des normierten Mittelwerts

$$Z_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}}$$

gegen die Standardnormalverteilung $N(0, 1)$, wenn $n \rightarrow \infty$.

Die genaue Bedeutung dieser Aussage sowie die zum Verständnis erforderlichen Details werden etwas später erläutert. Wir beginnen mit einer Reihe von Vorbemerkungen und praktischen Illustrationen, um die Tragweite dieses sehr grundlegenden Resultates zu beleuchten.

8.2 Anschauliche Bedeutung

Die zentrale Aussage des Satzes ist, dass sich durch Mittelung einer großen Anzahl unabhängiger Zufallsvariablen eine Normalverteilung ergibt, selbst wenn die ursprünglichen Zufallsvariablen nicht normalverteilt sind. Dies erklärt die allgegenwärtige Bedeutung der Normalverteilung in der Statistik, Naturwissenschaft und Technik.

Unabhängigkeit von der Ursprungsverteilung

Eine bemerkenswerte Eigenschaft des zentralen Grenzwertsatzes ist seine Unabhängigkeit von den genauen Eigenschaften der Ausgangsverteilung. Die einzige Voraussetzung ist, dass die Zufallsvariablen eine endliche Varianz besitzen. Beispielsweise konvergieren die Mittelwerte von Zufallsvariablen aus einer Gleichverteilung, einer Exponentialverteilung oder einer Beta-Verteilung gleichermaßen gegen eine Normalverteilung, wie es in Simulationen und empirischen Daten zu beobachten ist.

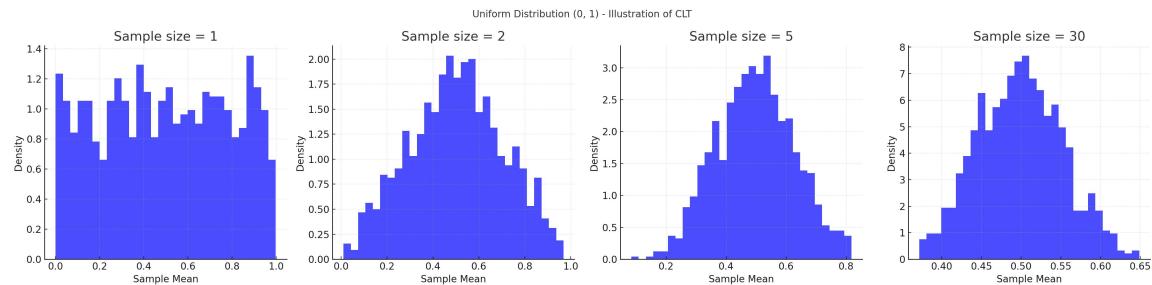
Diese Eigenschaft macht den Zentralen Grenzwertsatz zu einem universellen Prinzip, das eine zentrale Rolle bei der Modellierung und Analyse von Daten spielt, insbesondere bei großen Stichproben.

8.3 Simulationsbeispiele zur Illustration des Zentralen Grenzwertsatzes

Um die Aussage des Zentralen Grenzwertsatzes anschaulich zu verdeutlichen, betrachten wir drei Beispiele, die jeweils eine unterschiedliche Ausgangsverteilung der Zufallsvariablen verwenden: die Gleichverteilung, die Exponentialverteilung und die Beta(0.5, 0.5)-Verteilung. Für jede dieser Verteilungen betrachten wir die Mittelwerte von Stichproben unterschiedlicher Größe und analysieren deren Verteilung.

1. Gleichverteilung auf $[0, 1]$

Die Gleichverteilung beschreibt eine Zufallsvariable, bei der alle Werte im Intervall $[0, 1]$ die gleiche Wahrscheinlichkeit besitzen. Sie ist eine typische Verteilung mit konstanter Dichte. Bei kleinen Stichprobengrößen sind die Mittelwerte relativ gleichmäßig verteilt und spiegeln die Form der ursprünglichen Verteilung wider. Mit zunehmender Stichprobengröße wird die Verteilung der Mittelwerte jedoch schmäler und nähert sich der Glockenform einer Normalverteilung an.

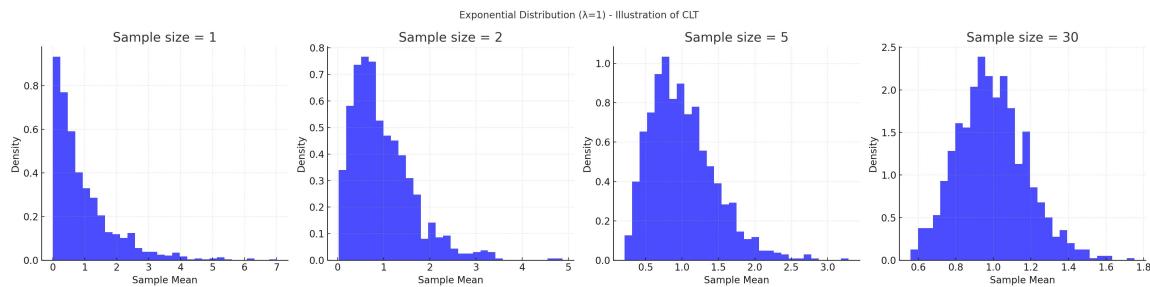


Erläuterung am Beispiel $n = 5$.

Im Fall $n = 5$ (dritter Plot oben) wurden 1000 unabhängige Stichproben vom Umfang $n = 5$ aus der Gleichverteilung $[0, 1]$ erzeugt. D.h. jede Stichprobe besteht hierbei aus 5 unabhängigen, gleichverteilten Zufallszahlen. (Insgesamt wurden also $5 * 1000$ Zufallszahlen unabhängig gemäß der Gleichverteilung auf $[0, 1]$ erzeugt.) Für jede der 1000 Stichproben vom Umfang 5 wurde der Mittelwert der 5 zugehörigen Einzelbeobachtungen berechnet. Dadurch ergibt sich eine neue Datenreihe mit 1000 Mittelwerten. Die Mittelwerte wurden in einem Histogramm dargestellt. Dabei wurde die Anzahl der Bins (Intervalle) auf 30 gesetzt, um die Verteilung der Mittelwerte übersichtlich zu visualisieren. Das Histogramm wurde normiert, sodass die Höhe der Balken eine Wahrscheinlichkeitsdichte repräsentiert. Analog gehen wir in den anderen Fällen $n = 1, n = 2$ bzw. $n = 30$ vor, um die drei restlichen Histogramme der empirischen Stichprobenmittelwerte zu erstellen.

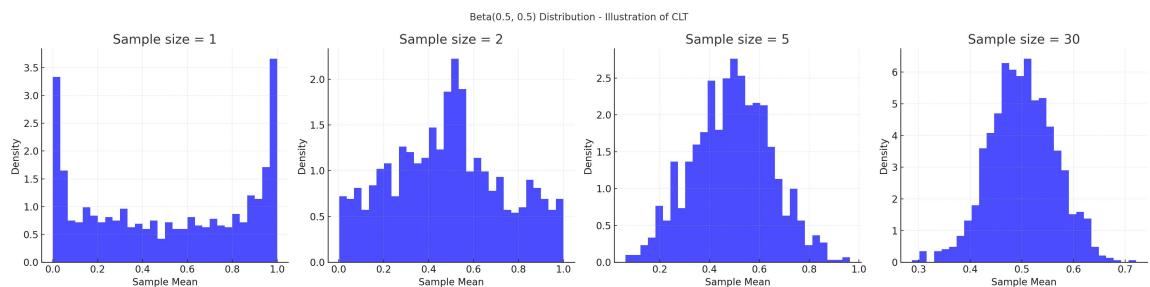
2. Exponentialverteilung ($\lambda = 1$)

Die Exponentialverteilung wird häufig zur Modellierung von Wartezeiten verwendet und hat eine stark asymmetrische Dichtefunktion, die bei $x = 0$ ihren Höhepunkt erreicht und exponentiell abfällt. Wir gehen vor wie im vorausgehenden Beispiel, wobei die Zufallszahlen zu jeder Stichprobe jetzt gemäß der Exponentialverteilung erzeugt sind. Trotz dieser asymmetrischen Ausgangsverteilung zeigt sich, dass die Verteilung der Mittelwerte mit wachsender Stichprobengröße symmetrisch wird und sich ebenfalls der Normalverteilung annähert. Dies unterstreicht die Unabhängigkeit des Zentralen Grenzwertsatzes von der ursprünglichen Form der Verteilung.



3. Beta(0.5, 0.5)-Verteilung

Die Beta($0.5, 0.5$)-Verteilung ist eine Verteilung, die ihre Dichte an den Rändern $x = 0$ und $x = 1$ konzentriert. Sie ist daher eine Verteilung mit sehr spezifischen Eigenschaften und stark ausgeprägten Extremen. Die Mittelwerte kleiner Stichproben reflektieren zunächst diese Eigenschaft, da sie häufiger Werte nahe 0 oder 1 annehmen. Mit wachsender Stichprobengröße wird die Verteilung der Mittelwerte jedoch immer glatter und nähert sich schließlich der Normalverteilung an.



Fazit

Diese drei Beispiele zeigen eindrucksvoll, dass die Ausgangsverteilung der Zufallsvariablen keinen Einfluss auf das asymptotische Verhalten der Verteilung der Mittelwerte hat, solange die Grundvoraussetzungen des Zentralen Grenzwertsatzes erfüllt sind. Die Form der Normalverteilung tritt unabhängig davon auf, ob die ursprüngliche Verteilung symmetrisch, asymmetrisch oder extrem ist. Dies verdeutlicht die universelle Anwendbarkeit des Zentralen Grenzwertsatzes in der Praxis.

8.4 Präzise Formulierung vom Zentralen Grenzwertsatz

Seien X_1, X_2, \dots unabhängige und identisch verteilte Zufallsvariablen mit $\mathbb{E}[X_1] = \mu$ und $\text{Var}(X_1) = \sigma^2 > 0$. Sei $S_n = \sum_{i=1}^n X_i$ die Summe der ersten n Zufallsvariablen. Dann gilt

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1),$$

im *Sinne der Konvergenz in Verteilung*, was bedeutet, dass für jedes abgeschlossene Intervall $[a, b] \subset \mathbb{R}$ gilt:

$$P\left(a \leq \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq b\right) \rightarrow \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx,$$

wenn $n \rightarrow \infty$. Die Verteilung der normierten Summe nähert sich also der Standardnormalverteilung $\mathcal{N}(0, 1)$.

8.5 Beweis vom Zentralen Grenzwertsatz (Stein'sche Methode)

Die historisch ersten Fassungen vom ZGS gehen auf de Moivre (1733) und Laplace (1812) zurück und behandeln den Spezialfall einer Folge von binomialverteilten Zufallsvariablen. Wir führen hier den Beweis im allgemeinen Fall nach der heutzutage sehr beliebten Stein'schen Methode (Charles Stein, 1970). Der Beweis verwendet neben der Unabhängigkeit nur elementare analytische Argumente.

Wir gehen dabei in zwei Schritten vor, wobei wir uns zum Zweck einer kompakten Darstellung auf den Fall beschränkter Zufallsvariablen (d.h. $|X_k| \leq C < \infty$ für alle k) konzentrieren wollen.

8.5.1 1. Schritt: Konvergenz von Erwartungswerten

Wir zeigen im ersten Schritt, dass für jede Funktion $f : \mathbb{R} \mapsto \mathbb{R}$, die einschließlich aller ihrer Ableitungen stetig und beschränkt ist, dass

$$\mathbb{E}(f(\xi_n)) \rightarrow \int_{\mathbb{R}} f(t)\gamma(t)dt$$

für $n \rightarrow \infty$, wobei,

$$\xi_n := \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

und

$$\gamma(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Ohne Beschränkung der Allgemeinheit können wir dabei im Folgenden davon ausgehen, dass $\mu = 0$ und $\sigma = 1$, denn andernfalls könnten wir einfach übergehen zu der Folge der zentrierten und standardisierten Zufallsgrößen $\hat{X}_k = (X_k - \mu)/\sigma$ und den Beweis hiermit führen. Auch können wir annehmen, dass $\int_{\mathbb{R}} f(t)\gamma(t)dt = 0$, denn andernfalls argumentieren wir einfach mit $\bar{f} = f - \int_{\mathbb{R}} f(t)\gamma(t)dt$ anstelle von f .

Mit dem gegebenen f definieren wir die neue Funktion h

$$h(x) := \frac{1}{\gamma(x)} \int_{-\infty}^x f(t)\gamma(t) dt = -\frac{1}{\gamma(x)} \int_x^\infty f(t)\gamma(t) dt.$$

Dann erfüllt h die Differentialgleichung

$$h'(x) = f(x) + xh(x),$$

da $\gamma'(x) = -x\gamma(x)$. Weiterhin macht man sich mit elementaren Argumenten klar, dass h und h' beschränkt und gleichmäßig stetig auf \mathbb{R} sind, d.h. insbesondere gilt, dass

$$\limsup_{\delta \rightarrow 0} \sup_{x \in \mathbb{R}} |h'(x + \delta) - h'(x)| = 0.$$

Unter Verwendung der oben stehenden Differentialgleichung für h können wir nun schreiben

$$\mathbb{E} \left[f \left(\frac{S_n}{\sqrt{n}} \right) \right] = \mathbb{E} \left[h' \left(\frac{S_n}{\sqrt{n}} \right) \right] - \mathbb{E} \left[\frac{S_n}{\sqrt{n}} h \left(\frac{S_n}{\sqrt{n}} \right) \right].$$

Aufgrund der Symmetrie der Summe $S_n = \sum_{i=1}^n X_i$ in den Größen X_1, X_2, \dots, X_n schreiben wir den letzten Term auf der rechten Seite als

$$\mathbb{E} \left[\frac{S_n}{\sqrt{n}} h \left(\frac{S_n}{\sqrt{n}} \right) \right] = \frac{n}{\sqrt{n}} \mathbb{E} \left[X_1 h \left(\frac{S_n}{\sqrt{n}} \right) \right]$$

und weiter mit $S_n = \tilde{S}_n + X_1$, wobei $\tilde{S}_n = \sum_{i=2}^n X_i$, und dem Hauptsatz der Differentialrechnung

$$= \mathbb{E} \left[X_1^2 \int_0^1 h' \left(\frac{\tilde{S}_n + sX_1}{\sqrt{n}} \right) ds \right] - \sqrt{n} \mathbb{E} \left[X_1 h \left(\frac{\tilde{S}_n}{\sqrt{n}} \right) \right],$$

wobei der zweite Term infolge der Unabhängigkeit von X_1 und \tilde{S}_n und da nach Voraussetzung $\mathbb{E}(X_1) = m = 0$ verschwindet, denn

$$\mathbb{E} \left[X_1 h \left(\frac{\tilde{S}_n}{\sqrt{n}} \right) \right] = \mathbb{E}[X_1] \cdot \mathbb{E} \left[h \left(\frac{\tilde{S}_n}{\sqrt{n}} \right) \right] = 0.$$

Insgesamt erhalten wir als Zwischenergebnis, dass

$$\mathbb{E} \left[f \left(\frac{S_n}{\sqrt{n}} \right) \right] = \mathbb{E} \left[h' \left(\frac{S_n}{\sqrt{n}} \right) \right] - \mathbb{E} \left[X_1^2 \int_0^1 h' \left(\frac{\tilde{S}_n + sX_1}{\sqrt{n}} \right) ds \right].$$

Wegen der Unabhängigkeit von X_1 und \tilde{S}_n und $\sigma^2 = \mathbb{E}[X_1^2] = 1$ gilt ferner

$$0 = \mathbb{E} \left[h' \left(\frac{\tilde{S}_n}{\sqrt{n}} \right) \right] - \mathbb{E} \left[X_1^2 h' \left(\frac{\tilde{S}_n}{\sqrt{n}} \right) \right]$$

so dass wir nach Ergänzung der Null zu der Darstellung gelangen

$$\mathbb{E} \left[f \left(\frac{S_n}{\sqrt{n}} \right) \right] = \mathbb{E} \left[h' \left(\frac{S_n}{\sqrt{n}} \right) - h' \left(\frac{\tilde{S}_n}{\sqrt{n}} \right) \right] - \mathbb{E} \left[X_1^2 \int_0^1 \left(h' \left(\frac{\tilde{S}_n + sX_1}{\sqrt{n}} \right) - h' \left(\frac{\tilde{S}_n}{\sqrt{n}} \right) \right) ds \right].$$

Aus der Gleichmäßigen Stetigkeit von h' und der Beschränktheit von X_1 folgt aus dieser Darstellung, dass in der Tat $\mathbb{E} \left[f \left(\frac{S_n}{\sqrt{n}} \right) \right] \rightarrow 0$ für $n \rightarrow \infty$ wie behauptet.

8.5.2 2. Schritt: Konvergenz von Wahrscheinlichkeiten

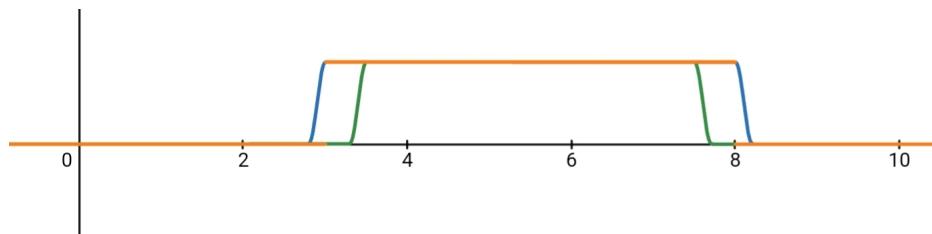
Die Aussage des ZGS im Hinblick auf die Wahrscheinlichkeiten ergäbe sich formal direkt aus der Konvergenz der Erwartungswerte, wenn man als Funktion f die Indikatorfunktion $f = \chi_{[a,b]}$ für das Intervall $[a, b]$ einsetzen könnte, d.h.

$$\chi_{[a,b]}(x) = \begin{cases} 1 & \text{falls } x \in [a, b], \\ 0 & \text{sonst,} \end{cases}$$

doch das ist aufgrund mangelnder Stetigkeit nicht zulässig.

Um diese Beschränkung zu überwinden, wählt man sich zu gegebenen $\epsilon > 0$ geeignete Approximationen χ_ϵ und χ^ϵ von unten bzw. von oben

$$\chi_\epsilon \leq \chi_{[a,b]} \leq \chi^\epsilon,$$



die jeweils die Voraussetzungen vom ersten Schritt erfüllen mit und so dass

$$\int_{\mathbb{R}} |\chi_\epsilon(t) - \chi^\epsilon(t)| \gamma(t) dt \leq \epsilon.$$

Man überzeugt sich leicht, dass solche Funktionen existieren. (Hierbei geht insbesondere ein, dass das Integral über immer kleinere Intervalle der Gauss'schen Glockenfunktion als Dichte gegen Null konvergiert.)

Sei nun ξ eine standardnormalverteilte Zuvallsvariable, d.h. es gelte $\mathbb{P}(\xi \in [u, v]) = \int_u^v \gamma(t) dt$ für alle $u \leq v$, so können wir unter Verwendung des Ergebnisses aus dem ersten Schritt folgern, dass

$$\begin{aligned} \limsup_{n \rightarrow \infty} \mathbb{P}(\xi_n \in [a, b]) - P(\xi \in [a, b]) &\leq \limsup_{n \rightarrow \infty} \mathbb{E}(\chi^\epsilon(\xi_n)) - \mathbb{E}(\chi_\epsilon(\xi)) \\ &= \mathbb{E}(\chi^\epsilon(\xi)) - \mathbb{E}(\chi_\epsilon(\xi)) \leq \int_{\mathbb{R}} |\chi_\epsilon(t) - \chi^\epsilon(t)| \gamma(t) dt \leq \epsilon \end{aligned}$$

und analog

$$\begin{aligned} \limsup_{n \rightarrow \infty} (\mathbb{P}(\xi \in [a, b]) - P(\xi_n \in [a, b])) &\leq \limsup_{n \rightarrow \infty} (\mathbb{E}(\chi^\epsilon(\xi)) - \mathbb{E}(\chi_\epsilon(\xi_n))) \\ &= \mathbb{E}(\chi^\epsilon(\xi)) - \mathbb{E}(\chi_\epsilon(\xi)) \leq \int_{\mathbb{R}} |\chi_\epsilon(t) - \chi^\epsilon(t)| \gamma(t) dt \leq \epsilon \end{aligned}$$

so dass also insgesamt

$$\limsup_{n \rightarrow \infty} |\mathbb{P}(\xi \in [a, b]) - P(\xi_n \in [a, b])| \leq \epsilon,$$

wobei $\epsilon > 0$ beliebig ist, d.h. also in der Tat erhalten wir

$$\lim_{n \rightarrow \infty} \mathbb{P}(\xi_n \in [a, b]) = \mathbb{P}(\xi \in [a, b]).$$

8.6 Anwendungsbeispiel: Rückstellungssumme bei einer Kfz-Versicherung

Problemstellung

Eine Versicherung möchte den Betrag bestimmen, den sie vorhalten muss, um mit einer Wahrscheinlichkeit von 95 % die jährlich anfallenden Schäden regulieren zu können.

Annahmen

- Anzahl der Kunden: $n = 10,000$.
- Erwartungswert des Schadens pro Kunde: $E(X_1) = \mu = 1,500$ Euro.
- Varianz des Schadens pro Kunde: $V(X_1) = \sigma^2 = 250,000$ Euro², also $\sigma = 500$ Euro.
- Maximaler Schaden pro Kunde: $X_{\max} = 10,000$ Euro.
- Schäden der Kunden sind unabhängig.

Berechnung des Betrags für 95-%-Deckung mit dem ZGS

Der Gesamtschaden $S_n = \sum_{i=1}^n X_i$ ist nach Verschiebung und Normierung gemäß ZGS approximativ normalverteilt:

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} \sim \mathcal{N}(0, 1).$$

- Erwartungswert:

$$\mathbb{E}[S_n] = n\mu = 10,000 \cdot 1,500 = 15,000,000 \text{ Euro.}$$

- Standardabweichung:

$$\sqrt{\text{Var}(S_n)} = \sqrt{n}\sigma = \sqrt{10,000} \cdot 500 = 50,000 \text{ Euro.}$$

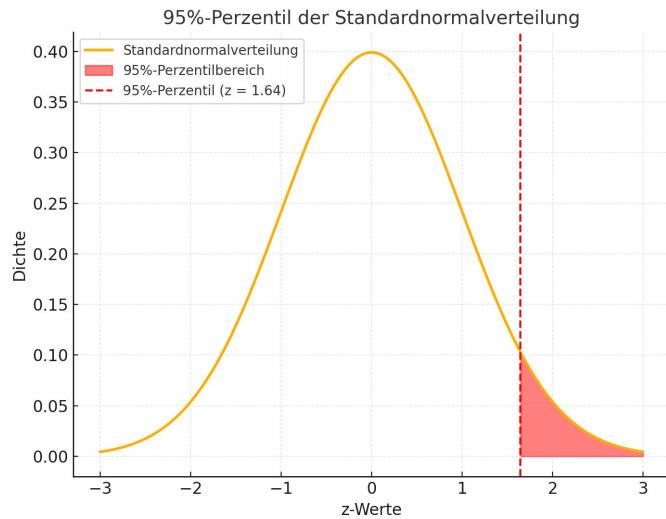
- Der Betrag Kritische Bertag L , der mit 95 % Wahrscheinlichkeit nicht überschritten wird, ergibt sich dann aus der Standardnormalverteilung in der Weise

$$L = n\mu + z_{0.95} \cdot \sqrt{n}\sigma,$$

wobei $z_{0.95} = 1.645$ der 95-%-Quantilwert, der Standardnormalverteilung ist, d.h.

$$z_{0.95} = \sup\{z \in \mathbb{R} \mid P(\xi \leq z) \leq 0.95\},$$

wobei ξ eine standardnormalverteilte Zufallsvariable sei.



Einsetzen ergibt:

$$L = 15,000,000 + 1.645 \cdot 50,000 = 15,082,250 \text{ Euro.}$$

Die Versicherung muss also 15,082,250 Euro vorhalten, um mit einer Wahrscheinlichkeit von 95% sämtliche in Verlaufe eines Jahres evtl. auftretenden Schäden regulieren zu können.

Vergleich mit Worst-Case-Szenario

Die theoretische obere Schranke S_{\max} für den Gesamtschaden tritt ein, wenn jeder Kunde den maximal möglichen Schaden verursacht:

$$S_{\max} = n \cdot X_{\max} = 10,000 \cdot 10,000 = 100,000,000 \text{ Euro.}$$

Ergebnisse

- Erwartungswert der Gesamtschäden: 15,000,000 Euro.
- Betrag für 95 %-Deckung: 15,082,250 Euro.
- Obere Schranke(Worst Case Szenario): 100,000,000 Euro.

Fazit

Die Versicherung sollte 15,082,250 Euro vorhalten, um mit 95 % Wahrscheinlichkeit alle jährlichen Schäden regulieren zu können. Die theoretische obere Schranke von 100,000,000 Euro ist jedoch ein Worst-Case-Szenario, das praktisch nicht auftritt.

Alternative Berechnung einer Rückstellungssumme mit der Tschebyschev-Ungleichung

Die Tschebyschev-Ungleichung garantiert, dass mindestens $1 - \frac{1}{k^2}$ der Werte innerhalb von k -Standardabweichungen vom Mittelwert liegen.

Für ein 95%-Konfidenzniveau setzen wir:

$$1 - \frac{1}{k^2} = 0.95 \implies k = \sqrt{\frac{1}{1 - 0.95}} = \sqrt{20} \approx 4.472$$

Die Rücklage wird berechnet als:

$$R = \mathbb{E}[S] + k \cdot \sigma_S$$

Einsetzen der Werte:

$$R = 15,000,000 + 4.472 \cdot 5,000,000 \approx 15,223,607 \text{ Euro}$$

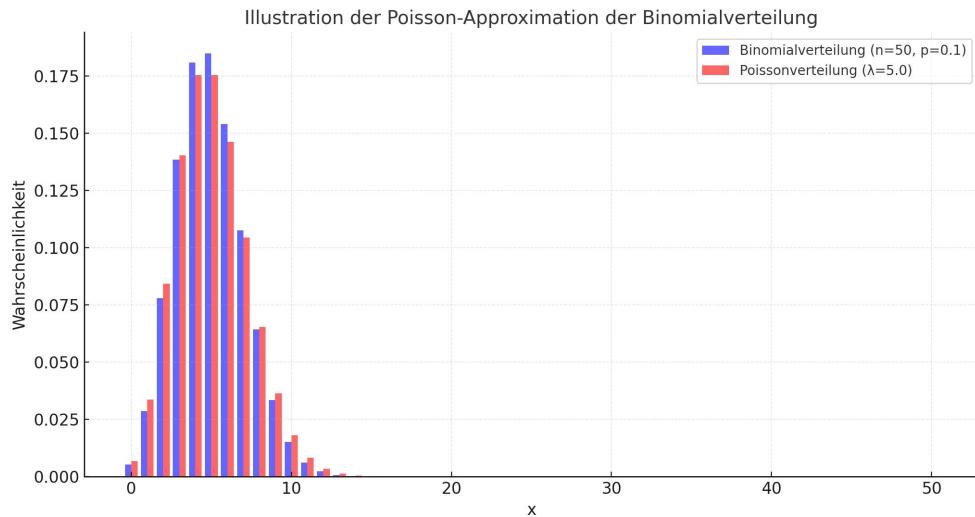
Die Rücklage basierend auf der Tschebyschev-Ungleichung ist höher, da sie weniger spezifische Annahmen macht (z. B. keine Normalverteilung voraussetzt) und für beliebige Verteilungen gilt. Diese Methode ist allgemeiner, aber weniger präzise als die Normalverteilungsannahme.

8.7 Nachtrag: Poisson'scher Grenzwertsatz und Anwendung im Mobilfunk

Der Poisson'sche Grenzwertsatz ist eine andere aber weniger tiefliegende Aussage zur Asymptotik von Summen von Zufallsvariablen, hier allerdings im Spezialfall von Binomialverteilungen im Fall von sehr großen n und sehr kleiner Erfolgswahrscheinlichkeit p : Er besagt, dass in diesem Parameterbereich die Binomialverteilung $B(n, p)$ approximativ durch die Poisson-Verteilung mit Parameter $\lambda = n \cdot p$ beschrieben werden kann:

$$P(X = k) \approx \frac{\lambda^k e^{-\lambda}}{k!}, \quad k \in \mathbb{N}_0.$$

Dies Näherung ergibt sich direkt aus der expliziten Formel für die Wahrscheinlichkeiten einer Binomialverteilung und einer Anwendung der Stirling'schen Näherung für die auftretenden Binomialkoeffizienten. Wir verzichten auf den Beweis und gehen direkt zu einer praktischen Anwendung über.



Praktische Anwendung:

Ein Mobilfunkanbieter möchte die Wahrscheinlichkeit berechnen, dass in einer Minute genau k Anrufe in einer kleinen Zelle eingehen. Gegeben sei, dass im Durchschnitt $\lambda = 2$ Anrufe pro Minute eintreffen. Nach der Poisson-Verteilung ergibt sich:

$$P(X = k) = \frac{2^k e^{-2}}{k!}.$$

Für $k = 3$ (genau 3 Anrufe) ist die Wahrscheinlichkeit:

$$P(X = 3) = \frac{2^3 e^{-2}}{3!} = \frac{8 \cdot e^{-2}}{6} \approx 0.180.$$

Dies ermöglicht eine einfache Planung von Kapazitäten und Ressourcen im Netzwerk.

Vergleich zur exakten Berechnung im Binomialmodell

Im ursprünglichen Beispiel wird angenommen, dass Anrufe in einer Mobilfunkzelle eintreffen. Für eine exakte Berechnung der Wahrscheinlichkeit, dass genau $k = 3$ Anrufe in einer Minute eintreffen, kann ein Binomialmodell verwendet werden.

Annahmen

- n : Maximale Anzahl von Anrufen, die in einer Minute möglich sind.
- p : Wahrscheinlichkeit, dass ein einzelner Kunde in einer Minute einen Anruf tätigt.
- Die Anzahl der Anrufe X folgt der Binomialverteilung:

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

- Erwartungswert: $\mathbb{E}[X] = n \cdot p$.
- Damit die Poisson-Approximation gilt, setzen wir $\lambda = n \cdot p = 2$.

Exakte Berechnung

Für die Wahrscheinlichkeit $P(X = 3)$ setzen wir die Binomialformel ein:

$$P(X = 3) = \binom{n}{3} p^3 (1 - p)^{n-3}.$$

Um die Poisson-Bedingung zu erfüllen (p klein und n groß), wählen wir $n = 1000$ und $p = 0.002$, sodass $\lambda = n \cdot p = 2$. Einsetzen ergibt:

$$P(X = 3) = \binom{1000}{3} (0.002)^3 (1 - 0.002)^{997}.$$

Die Binomialkoeffizienten und Wahrscheinlichkeiten berechnen wir:

$$\binom{1000}{3} = \frac{1000 \cdot 999 \cdot 998}{3 \cdot 2 \cdot 1} = 166,167,000,$$

$$(0.002)^3 = 0.000000008, \quad (1 - 0.002)^{997} \approx e^{-2}.$$

Daraus folgt:

$$P(X = 3) \approx 166,167,000 \cdot 0.000000008 \cdot 0.1353 \approx 0.180.$$

Vergleich mit der Poisson-Approximation

Die exakte Berechnung im Binomialmodell liefert nahezu denselben Wert wie die Poisson-Approximation ($P(X = 3) \approx 0.180$). Der Unterschied ist minimal, da die Annahmen der Poisson-Näherung hier gut erfüllt sind.

Kapitel 9

Elemente der schließenden Statistik

9.1 *Wahrscheinlichkeitstheorie und Statistik als Teilgebiete der Stochastik*

Die **Stochastik** ist die mathematische Lehre vom Zufall, die in zwei zentrale Teildisziplinen untergliedert wird: die **Wahrscheinlichkeitstheorie** und die **Statistik**. Während die Wahrscheinlichkeitstheorie die Grundlage bildet, auf der statistische Verfahren aufbauen, unterscheidet sich der Fokus beider Disziplinen deutlich.

Wahrscheinlichkeitstheorie: Die Grundlage der Stochastik

Die Wahrscheinlichkeitstheorie befasst sich mit der (mathematisch idealisierten) Modellierung zufälliger Phänomene, beginnend mit den bereits behandelten Konzepten von Wahrscheinlichkeitsraum, Zufallsvariablen, Verteilungen etc.

Ein **Münzwurf** wird z. B. durch $\Omega = \{\text{Kopf, Zahl}\}$ beschrieben, mit $P(\text{Kopf}) = p$ und $P(\text{Zahl}) = 1 - p$. Die **Zufallsvariablen** X , die den Ausgang eines Münzwurfs (z. B. $X = 1$ für Kopf, $X = 0$ für Zahl) oder die Anzahl der Köpfe in mehreren Würfen modellieren, erlauben eine präzise mathematische Beschreibung. Im Beispiel des wiederholten Münzwurfs (einem **Bernoulli-Experiment**) interessiert man sich z. B. für:

- die Wahrscheinlichkeit, dass bei $n = 10$ Würfen genau $k = 7$ -mal Kopf erscheint,
- die Verteilung der Anzahl der Köpfe nach n Würfen (Binomialverteilung).

Die Wahrscheinlichkeitstheorie liefert hier die Grundlagen für die Berechnung von Wahrscheinlichkeiten und weitergehende Analysen im Rahmen der eingeführten mathematischen Modelle, bei denen praktische Erfahrungen 'nur' idalisierend den Hintergrund als Motivation bilden.

Statistik: Rückschluss aus Daten

Die Statistik baut auf den Konzepten der Wahrscheinlichkeitstheorie auf, wendet diese jedoch an, um aus beobachteten Daten Rückschlüsse auf unbekannte Wahrscheinlichkeiten oder Verteilungen zu ziehen. Dabei wird von empirischen Beobachtungen in Form von Daten bzw. 'Stichproben' ausgegangen und im ersten Schritt die Frage gestellt, welches wahrscheinlichkeitstheoretische Modell zu Erklärung der gegebenen Bobachtung am besten geeignet ist. Im Münzwurf-Beispiel bedeutet das, dass es mit der Beobachtung eines

wiederholten Münzwurfs zu tun hat, deren Erfolgsparameter p – die Wahrscheinlichkeit für Kopf – **nicht bekannt** ist. Typische grundlegende Fragestellungen der Statistik im Münzbeispiel etwa sind:

- Wie kann p geschätzt werden, wenn bei 100 Würfen 62-mal Kopf beobachtet wurde?
- Ist die Hypothese $H_0 : p = 0.5$ (die Münze ist fair) mit den Daten vereinbar?

Die Statistik verwendet Werkzeuge wie die **Maximum-Likelihood-Schätzung**, um Parameter wie p zu schätzen, oder **Hypothesentests**, um Annahmen über p zu überprüfen. Hierbei spielen die von der Wahrscheinlichkeitstheorie bereitgestellten Konzepte – wie Wahrscheinlichkeitsräume, Zufallsvariablen und Verteilungen – eine zentrale Rolle.

Zusammenhang der Disziplinen

Die Wahrscheinlichkeitstheorie bietet das mathematische Fundament der Stochastik, indem sie zufällige Prozesse präzise modelliert. Die Statistik nutzt diese Modelle, um aus Daten Schlüsse zu ziehen und Ungewissheit zu quantifizieren. Während die Wahrscheinlichkeitstheorie von idealisierten Wahrscheinlichkeitsmaßen ausgeht, beschäftigt sich die Statistik mit der umgekehrten Aufgabe: den zugrunde liegenden Wahrscheinlichkeitsraum aus empirischen Beobachtungen zu rekonstruieren. Gemeinsam ermöglichen beide Disziplinen eine umfassende mathematische Beschreibung und Analyse zufälliger Phänomene.

9.2 Beispiel: Das Schätzproblem für den Erfolgsparameter einer Münze

Beim Schätzproblem für den Erfolgsparameter p eines wiederholten Münzwurfs wird untersucht, wie der unbekannte Wert $p = P(\text{Kopf})$ aus Beobachtungen geschätzt werden kann. Dabei wird die Münze n -mal geworfen, und die Zufallsvariable X , die die Anzahl der Köpfe zählt, spielt eine zentrale Rolle. X kann als Summe von n unabhängigen Bernoulli-Zufallsvariablen X_1, X_2, \dots, X_n aufgefasst werden, wobei jede Zufallsvariable X_i den Wert 1 annimmt, wenn der i -te Wurf Kopf ergibt, und 0, wenn er Zahl zeigt. Formal ist:

$$X = \sum_{i=1}^n X_i,$$

wobei $P(X_i = 1) = p$ und $P(X_i = 0) = 1 - p$.

Die Verteilung von X ist dann eine **Binomialverteilung** mit Parametern n und p , d. h.:

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

Der Schätzer für den Erfolgsparameter p ist der relative Anteil der Köpfe:

$$\hat{p} = \frac{X}{n}.$$

Dieser Schätzer entspricht der empirischen Häufigkeit von Kopf und ist intuitiv, da er den beobachteten Anteil der Erfolge an der Gesamtzahl der Würfe angibt.

9.2.1 Verbindung zum Schwachen Gesetz der Großen Zahlen

Das schwache Gesetz der großen Zahlen besagt, dass \hat{p} mit wachsendem n gegen den wahren Wert p konvergiert. Da X als Summe unabhängiger Zufallsvariablen geschrieben werden kann, gilt für den Erwartungswert und die Varianz:

$$\mathbb{E}[X] = \sum_{i=1}^n \mathbb{E}[X_i] = np, \quad \text{und} \quad \text{Var}(X) = \sum_{i=1}^n \text{Var}(X_i) = np(1-p).$$

Der Erwartungswert von \hat{p} ist:

$$\mathbb{E}[\hat{p}] = \mathbb{E}\left[\frac{X}{n}\right] = \frac{\mathbb{E}[X]}{n} = p,$$

und die Varianz von \hat{p} ist:

$$\text{Var}(\hat{p}) = \text{Var}\left(\frac{X}{n}\right) = \frac{\text{Var}(X)}{n^2} = \frac{p(1-p)}{n}.$$

Da die Varianz mit zunehmendem n gegen 0 geht, konzentriert sich \hat{p} immer stärker um p , und wir erhalten:

$$P(|\hat{p} - p| > \epsilon) \rightarrow 0 \quad \text{für } n \rightarrow \infty.$$

Illustration am Beispiel

Angenommen, wir werfen eine Münze $n = 10$ -mal, und die wahre Wahrscheinlichkeit für Kopf ist $p = 0.6$. Jeder Wurf ist ein unabhängiges Bernoulli-Experiment mit $P(X_i = 1) = 0.6$ und $P(X_i = 0) = 0.4$. Eine mögliche Beobachtung könnte $X = 7$ Köpfe sein. Der Schätzer wäre dann:

$$\hat{p} = \frac{X}{n} = \frac{7}{10} = 0.7.$$

Da n klein ist, weicht \hat{p} vom wahren p ab, was eine hohe Varianz des Schätzers wider- spiegelt. Wenn jedoch $n = 1000$ Würfe durchgeführt werden und z. B. $X = 598$ Köpfe beobachtet werden, ist:

$$\hat{p} = \frac{598}{1000} = 0.598,$$

und \hat{p} liegt sehr nahe an $p = 0.6$. Dies illustriert die Konvergenz von \hat{p} gegen p gemäß dem schwachen Gesetz der großen Zahlen, da die relative Häufigkeit stabiler wird, wenn die Anzahl der Versuche wächst.

9.2.2 Berechnung der benötigten Anzahl von Würfen bei gegebener Fehlertoleranz und maximal zugelassener Irrtumswahrscheinlichkeit

In der Praxis ist eine vollständig präzise Schätzung von p unmöglich, da der wahre Wert von p nur bei unendlich vielen Würfen exakt bestimmt werden könnte. Stattdessen wird der Schätzer \hat{p} verwendet, der mit einer gewissen Unsicherheit behaftet ist. Diese Unsi- cherheit wird durch zwei Größen charakterisiert: die **Fehlertoleranz** Δ , die angibt, wie stark \hat{p} vom wahren p maximal abweichen darf, und die **Irrtumswahrscheinlichkeit** ϵ ,

die die Wahrscheinlichkeit beschreibt, dass diese Abweichung Δ überschreitet. Praktisch wird n so gewählt, dass die Wahrscheinlichkeit

$$P(|\hat{p} - p| \geq \Delta) \leq \epsilon$$

erfüllt ist, wodurch eine kontrollierte, jedoch niemals perfekte Annäherung an p erreicht wird. In der praktischen Durchführung gibt man die Fehlertoleranz Δ und die Irrtumswahrscheinlichkeit ϵ vor, und fragt, ob man einen Schätzer \hat{p} konstruieren kann, der die Genauigkeit im Sinne der obigen Forderung hat. Im Falle der Münzwurfs läuft das auf die Bestimmung einer minimalen Anzahl n von Wiederholungen hinaus.

Variante 1: Tschebyschev-Ungleichung

Um die notwendige Anzahl von Würfen n zu bestimmen, damit der Schätzer \hat{p} den Erfolgsparameter p mit einer Fehlertoleranz von $\Delta = 0.1$ und einer Irrtumswahrscheinlichkeit von maximal 5% (also $\epsilon = 0.05$) abschätzen kann, verwenden wir die **Tschebyscheff-Ungleichung**. Diese lautet:

$$P(|\hat{p} - p| \geq \Delta) \leq \frac{\text{Var}(\hat{p})}{\Delta^2}.$$

Die Varianz von \hat{p} ist gegeben durch:

$$\text{Var}(\hat{p}) = \frac{p(1-p)}{n}.$$

Setzen wir dies in die Tschebyscheff-Ungleichung ein:

$$P(|\hat{p} - p| \geq \Delta) \leq \frac{p(1-p)}{n\Delta^2}.$$

Damit die Irrtumswahrscheinlichkeit $\leq \epsilon = 0.05$ bleibt, muss gelten:

$$\frac{p(1-p)}{n\Delta^2} \leq 0.05.$$

Um die Berechnung unabhängig von der Kenntnis des wahren Werts von p zu machen, nutzen wir die Ungleichung:

$$p(1-p) \leq \frac{1}{4},$$

die für alle $p \in [0, 1]$ gilt (das Maximum von $p(1-p)$ wird bei $p = 0.5$ erreicht). Damit können wir die Abschätzung für n vereinfachen:

$$\frac{\frac{1}{4}}{n\Delta^2} \leq 0.05 \quad \Rightarrow \quad n \geq \frac{1}{4 \cdot \epsilon \cdot \Delta^2}.$$

Numerisches Beispiel

Setzen wir die Werte $\epsilon = 0.05$ und $\Delta = 0.1$ ein, ergibt sich:

$$n \geq \frac{1}{4 \cdot 0.05 \cdot 0.1^2} = \frac{1}{4 \cdot 0.05 \cdot 0.01} = \frac{1}{0.002} = 500.$$

Somit sind mindestens $n = 500$ Würfe erforderlich, um sicherzustellen, dass der Schätzer \hat{p} den Erfolgsparameter p mit einer Fehlertoleranz von $\Delta = 0.1$ und einer Irrtumswahrscheinlichkeit von höchstens 5% approximiert. Dieser Wert ist unabhängig vom tatsächlichen p , da die Abschätzung für $p(1-p) \leq 1/4$ allgemeingültig ist.

Variante 2: Zentraler Grenzwertsatz

Alternativ zur Tschebyscheff-Ungleichung kann der **Zentrale Grenzwertsatz (ZGS)** verwendet werden, um die benötigte Anzahl von Würfen n zu bestimmen. Der ZGS besagt, dass für große n die standardisierte Zufallsvariable

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

approximativ standardnormalverteilt ist, d. h. $Z \sim \mathcal{N}(0, 1)$. Wir können damit die Wahrscheinlichkeit abschätzen, dass \hat{p} vom wahren p um nicht mehr als Δ abweicht:

$$P(|\hat{p} - p| < \Delta) \approx P(-z_{\alpha/2} < Z < z_{\alpha/2}),$$

wobei $z_{\alpha/2}$ der kritische Wert der Standardnormalverteilung ist, der $\alpha = 0.05$ beschreibt (für $\alpha/2 = 0.025$ ist $z_{0.025} \approx 1.96$).

Um diese Wahrscheinlichkeit zu garantieren, fordern wir:

$$\frac{\Delta}{\sqrt{\frac{p(1-p)}{n}}} \geq z_{0.025},$$

was nach Umstellen zu:

$$n \geq \frac{z_{0.025}^2 \cdot p(1-p)}{\Delta^2}.$$

Verwendung der konservativen Schranke für $p(1-p)$

Wie im vorherigen Abschnitt nehmen wir die maximale Schranke $p(1-p) \leq 1/4$ an, um keine Vorabinformation über p zu benötigen. Einsetzen liefert:

$$n \geq \frac{z_{0.025}^2 \cdot \frac{1}{4}}{\Delta^2}.$$

Numerisches Beispiel

Für $\Delta = 0.1$ und $z_{0.025} \approx 1.96$ ergibt sich:

$$n \geq \frac{1.96^2 \cdot \frac{1}{4}}{0.1^2} = \frac{3.8416 \cdot 0.25}{0.01} = \frac{0.9604}{0.01} = 96.04.$$

Da n ganzzahlig sein muss, benötigen wir mindestens $n = 97$ Würfe. Dies ist deutlich weniger als bei der Tschebyscheff-Ungleichung, da der ZGS präzisere Annahmen über die Verteilung von \hat{p} für große n erlaubt.

Vergleich der Ansätze

Während die Tschebyscheff-Ungleichung allgemeingültig ist und keine Annahmen über die Verteilung von \hat{p} macht, ist der Zentrale Grenzwertsatz für große n genauer, da er die Normalverteilungsnäherung nutzt. Daher liefert der ZGS in diesem Fall eine deutlich kleinere Anzahl von Würfen, ohne die Zuverlässigkeit der Schätzung zu beeinträchtigen.

9.3 Allgemeine Schätzprobleme

Im Falle von Beobachtungen eines zufallsbehafteten Vorgangs spricht man auch von einer **Stichprobe** \mathcal{X} . Grundaufgabe der Statistik ist, aus den beobachteten Daten \mathcal{X} auf eine unbekannte Eigenschaft des zugrundeliegenden Zufallsmechanismus zu schließen. Wird die gesuchte Eigenschaft durch einen unbekannten Parameter beschrieben, der möglichst genau ermittelt werden soll, spricht man von einem **Schätzproblem**.

9.3.1 Kenngrößen einer Verteilung

Der interessierende Parameter wird dabei auch als **Kenngröße** ϑ bezeichnet. Eine Kenngröße beschreibt eine charakteristische Eigenschaft der Verteilung der Zufallsvariablen.

Zu den wichtigsten Beispielen von zu schätzenden Kenngrößen gehören:

- Erwartungswert $\vartheta = \mu$: Gibt die mittlere Lage der Verteilung an und beschreibt den durchschnittlichen Wert der Zufallsvariablen.
- Standardabweichung $\vartheta = \sigma$ bzw. Varianz $\vartheta = \sigma^2$: Misst die Streuung der Werte um den Erwartungswert und gibt an, wie stark die Beobachtungen voneinander abweichen.
- Schiefe $\vartheta = \gamma_1 = \frac{\mathbb{E}[(X-\mu)^3]}{\sigma^3}$: Charakterisiert die Asymmetrie der Verteilung. Eine positive Schiefe bedeutet, dass die Verteilung einen längeren rechten Schwanz hat, während eine negative Schiefe auf eine Linksverzerrung hindeutet.

Ziel des Schätzproblems ist es, diese unbekannten Kenngrößen auf Basis der Stichprobewerte möglichst genau zu bestimmen.

9.3.2 Schätzer für eine Kenngröße

Ziel ist es, aus den gemachten Beobachtungen \mathbf{X} einen Schätzwert $\hat{\vartheta} = \hat{\vartheta}(\mathcal{X})$ für die gesuchte Kenngröße θ zu konstruieren, die den gesuchten Parameter, wie beispielsweise den Erwartungswert oder die Varianz, möglichst genau approximiert.

Der **systematische Fehler** (Bias) des Schätzers ist definiert als:

$$\text{Bias}(\hat{\theta}) = \mathbb{E}[\hat{\theta}] - \theta$$

Er gibt an, um wie viel der erwartete Wert des Schätzers systematisch vom wahren Parameterwert abweicht. Häufig interessiert man sich für Schätzer ohne systematischen Fehler, in dem Fall spricht man von einem **erwartungstreuen Schätzer**.

9.3.3 Erwartungstreue Schätzer für Mittelwert, Varianz und Schiefe

Gegeben sei eine Stichprobe $\mathcal{X} = \{X_1, X_2, \dots, X_n\}$ von unabhängigen und identisch verteilten Realisierungen einer reellen Zufallsvariablen X mit Erwartungswert $\mu = \mathbb{E}[X]$ und Varianz $\sigma^2 = \text{Var}(X)$. Erwartungstreue Schätzer für die wichtigsten Kenngrößen einer Verteilung lauten wie folgt:

Schätzer für den Mittelwert

Der erwartungstreue Schätzer für den Erwartungswert μ ist das Stichprobenmittel:

$$\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

Dieser Schätzer ist erwartungstreu, da:

$$\mathbb{E}[\bar{X}] = \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] = \mu.$$

Schätzer für die Varianz

Ein erwartungstreuer Schätzer für die Varianz σ^2 ist die korrigierte Stichprobenvarianz:

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Dieser Schätzer ist erwartungstreu, da:

$$\mathbb{E}[\hat{\sigma}^2] = \sigma^2.$$

Schätzer für die Schiefe

Ein erwartungstreuer Schätzer für die Schiefe γ_1 ist:

$$\hat{\gamma}_1 = \frac{n}{(n-1)(n-2)} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\hat{\sigma}} \right)^3$$

Dieser Schätzer kompensiert die Verzerrung, die bei kleinen Stichproben durch die Schätzung des Mittelwerts und der Varianz entsteht.

Schätzung der Kenngröße $\vartheta = \frac{1}{\lambda}$ einer Exponentialverteilung

Sei X_1, X_2, \dots, X_n eine Stichprobe unabhängiger und identisch verteilter Zufallsvariablen aus der Exponentialverteilung mit Parameter $\lambda > 0$. Die Kenngröße $\vartheta = \frac{1}{\lambda}$ entspricht dem Erwartungswert der Verteilung, also $\mathbb{E}[X] = \vartheta$. Ein natürlicher erwartungstreuer Schätzer für ϑ ist das Stichprobenmittel:

$$\hat{\vartheta} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

Zusammenfassung

- Mittelwert: $\hat{\mu} = \frac{1}{n} \sum X_i$
- Varianz: $\hat{\sigma}^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$
- Schiefe: $\hat{\gamma}_1 = \frac{n}{(n-1)(n-2)} \sum \left(\frac{X_i - \bar{X}}{\hat{\sigma}} \right)^3$

Diese Schätzer sind erwartungstreu und liefern konsistente Schätzwerte für die zugrunde liegenden Kenngrößen der Verteilung.

9.3.4 Mittlerer quadratischer Fehler eines Schätzers

Der **mittlere quadratische Fehler** (Mean Squared Error, MSE) eines Schätzers $\hat{\theta}$ für einen unbekannten Parameter θ ist definiert als:

$$\text{MSE}(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \theta)^2]$$

Er misst die durchschnittliche quadratische Abweichung des Schätzers von dem wahren Wert θ .

Die Tschebyschev-Ungleichung ermöglicht eine allgemeine Fehlerabschätzung für beliebige Verteilungen, indem sie eine obere Schranke für die Wahrscheinlichkeit liefert, dass der Schätzfehler $|\hat{\theta} - \theta|$ eine gegebene Schwelle überschreitet:

$$P(|\hat{\theta} - \theta| \geq \varepsilon) \leq \frac{\text{MSE}(\hat{\theta})}{\varepsilon^2}.$$

Diese Ungleichung zeigt, dass ein kleiner MSE sicherstellt, dass der Schätzfehler mit hoher Wahrscheinlichkeit klein bleibt. Dies ist besonders nützlich in Situationen, in denen keine genauen Verteilungsannahmen getroffen werden können, aber dennoch eine verlässliche Fehlerkontrolle erforderlich ist.

9.3.5 Fehlerzerlegung des mittleren quadratischen Fehlers

Die **Streuung** (Varianz) eines Schätzers $\hat{\theta}$ ist gegeben durch:

$$\text{Var}(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \mathbb{E}[\hat{\theta}])^2]$$

Sie misst die Variabilität des Schätzers um seinen Erwartungswert. Der MSE lässt sich in zwei Fehleranteile (Streuung und Varianz Zerlegen):

$$\text{MSE}(\hat{\theta}) = \text{Var}(\hat{\theta}) + (\text{Bias}(\hat{\theta}))^2$$

Beweis

Durch Einsetzen der Definition des MSE:

$$\text{MSE}(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \theta)^2]$$

Umformen durch Einfügen und Subtrahieren von $\mathbb{E}[\hat{\theta}]$:

$$\mathbb{E}[(\hat{\theta} - \mathbb{E}[\hat{\theta}] + \mathbb{E}[\hat{\theta}] - \theta)^2]$$

Anwendung der Quadratförmel $(a + b)^2 = a^2 + 2ab + b^2$:

$$\mathbb{E}[(\hat{\theta} - \mathbb{E}[\hat{\theta}])^2 + 2(\hat{\theta} - \mathbb{E}[\hat{\theta}])(\mathbb{E}[\hat{\theta}] - \theta) + (\mathbb{E}[\hat{\theta}] - \theta)^2]$$

Da der Erwartungswert des mittleren Abweichungsterms null ist, folgt:

$$\mathbb{E}[(\hat{\theta} - \mathbb{E}[\hat{\theta}])^2] + (\mathbb{E}[\hat{\theta}] - \theta)^2$$

Dies entspricht genau:

$$\text{Var}(\hat{\theta}) + (\text{Bias}(\hat{\theta}))^2$$

□

9.3.6 Beispiel: Vergleich zweier Schätzer im 'Gameshow'-Problem

Gegeben sei eine Stichprobe von $n = 10$ unabhängigen Beobachtungen einer Zufallsvariablen, die gleichverteilt auf der Menge $\{1, 2, \dots, L\}$ ist. Wir betrachten die folgenden beiden Schätzer für L :

1. Schätzer 1:

$$\hat{L}_1 = \frac{2}{n} \sum_{i=1}^n X_i - 1$$

2. Schätzer 2:

$$\hat{L}_2 = \max(X_1, X_2, \dots, X_n)$$

Für $L = 10$ ergeben sich die folgenden Werte für Bias, Varianz und den mittleren quadratischen Fehler (MSE):

Kenngröße	\hat{L}_1	\hat{L}_2
Bias	0.00	-0.91
Varianz	3.30	0.69
MSE	3.30	1.52

Interpretation

- Bias: Der Schätzer \hat{L}_1 ist erwartungstreu ($\text{Bias} = 0$), während \hat{L}_2 eine leichte Unterschätzung von L aufweist. - Varianz: Der Schätzer \hat{L}_2 besitzt eine deutlich geringere Streuung als \hat{L}_1 . - Mittlerer quadratischer Fehler (MSE): Da \hat{L}_2 eine kleinere Varianz hat und der Bias nur gering ist, führt dies zu einem insgesamt besseren MSE-Wert im Vergleich zu \hat{L}_1 .

Fazit

Obwohl \hat{L}_1 erwartungstreu ist, zeigt \hat{L}_2 aufgrund seiner geringeren Varianz einen besseren MSE-Wert. Damit ist \hat{L}_2 insgesamt der bevorzugte Schätzer für L .