

Statistik nach der Grundausbildung

Andreas Handl

Sommersemester 2000

Inhaltsverzeichnis

| | | |
|----------|---|------------|
| 1 | Einführung | 5 |
| 2 | Wahrscheinlichkeitsrechnung | 7 |
| 2.1 | Zufallsvorgänge, Ereignisse, Meßräume | 7 |
| 2.1.1 | Definitionen und Beispiele | 7 |
| 2.1.2 | Operationen zwischen Ereignissen | 10 |
| 2.1.3 | σ -Algebren und Meßräume | 16 |
| 2.2 | Wahrscheinlichkeit | 19 |
| 2.2.1 | Vorbemerkungen | 19 |
| 2.2.2 | Klassischer Ansatz | 19 |
| 2.2.3 | Frequentistischer Ansatz | 20 |
| 2.2.4 | Axiomatische Definition | 21 |
| 2.2.5 | Kombinatorik | 27 |
| 2.2.6 | Bedingte Wahrscheinlichkeit | 44 |
| 2.2.7 | Multiplikationssätze | 48 |
| 2.2.8 | Satz von der totalen Wahrscheinlichkeit | 51 |
| 2.2.9 | Satz von Bayes | 55 |
| 2.2.10 | Unabhängigkeit | 58 |
| 3 | Univariate Verteilungen | 63 |
| 3.1 | Zufallsvariablen und ihre Verteilungen | 63 |
| 3.2 | Diskrete Verteilungsmodelle | 70 |
| 3.3 | Stetige Verteilungsmodelle | 88 |
| 3.4 | Die Exponentialfamilie | 100 |
| 4 | Funktionen von Zufallsvariablen | 109 |
| 5 | Parameter univariater Verteilungen | 123 |

| | | |
|----------|---|------------|
| 5.1 | Erwartungswerte | 123 |
| 5.2 | Erwartungswerte von Funktionen von Zufallsvariablen | 136 |
| 5.3 | Die Varianz | 143 |
| 5.4 | Momente und momenterzeugende Funktion | 150 |
| 5.5 | Momente spezieller Verteilungen | 157 |
| 5.6 | Score und Fisher-Information | 161 |
| 5.7 | Momente der Exponentialfamilie | 167 |
| 6 | Mehrdimensionale Verteilungen | 173 |
| 6.1 | Mehrdimensionale Zufallsvariablen | 173 |
| 6.2 | Randverteilungen | 181 |
| 6.3 | Bedingte Verteilungen | 186 |
| 6.4 | Unabhängigkeit | 194 |
| 6.5 | Die bivariate Normalverteilung | 200 |
| 7 | Funktionen | 207 |
| 7.1 | Funktionen von diskreten Zufallsvariablen | 207 |
| 7.2 | Funktionen von stetigen Zufallsvariablen | 218 |
| 7.3 | Verteilung des Maximums und des Minimums | 231 |
| 7.4 | Funktionen von unabhängigen Zufallsvariablen | 237 |
| 8 | Parameter | 239 |
| 8.1 | Erwartungswerte | 239 |
| 8.2 | Kovarianz und Korrelationskoeffizient | 245 |
| 8.3 | Kovarianzen in der Zeitreihenanalyse | 262 |
| 8.4 | Kovarianzen in der Multivariaten Analyse | 272 |
| 8.5 | Bedingte Erwartungswerte | 277 |
| 8.6 | Prognose | 291 |

Kapitel 1

Einführung

Womit beschäftigt sich Statistik?

Eine schöne Beschreibung des Aufgabengebietes der Statistik gibt Efron im SIAM-Review. Hier versucht er Mathematikern klarzumachen, was Statistik ist und womit sie sich beschäftigt.

Statistics concerns the comparison of sets of numbers

- with each other
- with theoretical models
- with past experience

Im ersten und letzten Fall kann man diesen Vergleich rein beschreibend durchführen. In diesem Fall befindet man sich in der deskriptiven Statistik. In allen drei Fällen kann man diesen Vergleich auch induktiv mit Hilfe der Wahrscheinlichkeitsrechnung durchführen.

Ziel dieses Skriptes ist es, eine Einführung in die Theorie der Wahrscheinlichkeitsrechnung zu geben.

Kapitel 2

Wahrscheinlichkeitsrechnung

2.1 Zufallsvorgänge, Ereignisse, Meßräume

2.1.1 Definitionen und Beispiele

Der Ausgang der meisten Vorgänge ist unbestimmt.

So sind die Brötchen beim morgendlichen Besuch des Bäckerladens manchmal knusperig, manchmal aber leider weich. Manchmal wird man sofort bedient, manchmal muß man warten. Wenn man warten muß, ist die Anzahl der Personen, die vor einem steht, nicht immer gleich. Auch die Wartezeit variiert. Es ist zwar in der Regel unmöglich, den Ausgang eines Vorgangs vorherzusagen, wir haben aber oft eine Vorstellung über die Chance der einzelnen Ausgänge. Wir werden in diesem Kapitel sehen, wie wir diese zahlenmäßig erfassen können.

Hierzu beginnen wir mit der folgenden Definition.

Definition 2.1.1 *Ein Experiment oder Vorgang, der unter Konstanz eines bestimmten Bedingungskomplexes beliebig oft wiederholbar ist, und dessen Ausgang nicht vorhersagbar ist, wird als **Zufallsexperiment** oder **Zufallsvorgang** bezeichnet.*

Die Menge aller möglichen Ergebnisse des Zufallsvorgangs nennt man **Ergebnismenge** Ω .

Schauen wir uns einige Beispiele an.

Beispiel 2.1.1 *Eine Münze wird einmal geworfen, wobei entweder Kopf K oder Zahl Z auftritt.*

Die Ergebnismenge ist somit

$$\Omega = \{K, Z\}$$

Beispiel 2.1.2 *Eine Münze wird zweimal geworfen, wobei bei einem Wurf entweder Kopf oder Zahl auftritt.*

Die Ergebnismenge ist somit

$$\Omega = \{KK, KZ, ZK, ZZ\}$$

Beispiel 2.1.3 *Ein Würfel wird einmal geworfen.*

Die Ergebnismenge ist

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

Beispiel 2.1.4 *Eine Münze wird so oft geworfen, bis zum ersten Mal Kopf auftritt.*

Die Ergebnismenge ist

$$\Omega = \{K, ZK, ZZK, ZZZK, \dots\}$$

Beispiel 2.1.5 *Ein Kunde zählt die Anzahl der Kunden, die vor ihm im Bäckerladen stehen.*

Die Ergebnismenge ist

$$\Omega = \{0, 1, 2, 3, \dots\}$$

Beispiel 2.1.6 *Wir bestimmen die Wartezeit im Bäckerladen, wobei ein Kunde nicht bereit ist, länger als 10 Minuten zu warten.*

Die Ergebnismenge ist

$$\Omega = \{x | x \in \mathfrak{R}, 0 \leq x \leq 10\} = [0, 10]$$

Die Ergebnismengen der ersten drei Beispiele enthalten endlich viele Ergebnisse, die restlichen unendlich viele. Die unendlich vielen Ergebnisse der Beispiele 2.1.4 und 2.1.5 können abgezählt werden während dies bei der Ergebnismenge des Beispiels 2.1.6 nicht möglich ist.

Ergebnismengen heißen **diskret**, wenn sie endlich viele oder abzählbar unendlich viele Ergebnisse enthalten. Ansonsten heißen sie **stetig**.

Bei Zufallsvorgängen ist man oft an Zusammenfassungen von Ergebnissen interessiert, zum Beispiel beim Warten im Bäckerladen daran, daß höchstens zwei Leute vor einem stehen, oder man höchstens zwei Minuten warten muß.

Eine Zusammenfassung von Ergebnissen aus Ω heißt **Ereignis** A .

Man sagt, daß das Ereignis A eingetreten ist, wenn ein Ergebnis ω aus A beobachtet wurde.

Beispiel 2.1.7 *Beim einmaligen Wurf eines Würfels ist man interessiert an den geraden Augenzahlen.*

Das zugehörige Ereignis ist

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

Beispiel 2.1.8 *Beim zweimaligen Wurf eines Würfels ist man daran interessiert, daß genau einmal Kopf eintritt.*

Das zugehörige Ereignis ist

$$A = \{KZ, ZK\}$$

Beispiel 2.1.9 *Beim Besuch im Bäckerladen ist man daran interessiert, daß höchstens zwei Leute vor einem stehen.*

Das zugehörige Ereignis ist

$$A = \{0, 1, 2\}$$

Beispiel 2.1.10 *Beim Besuch im Bäckerladen ist man daran interessiert, daß man höchstens zwei Minuten warten muß.*

Das zugehörige Ereignis ist

$$A = [0, 2] = \{x | x \in \mathbb{R}, 0 \leq x \leq 2\}$$

Die Ergebnismenge Ω heißt **sicheres Ereignis**, da sie immer eintritt.

Die leere Menge \emptyset heißt **unmögliches Ereignis**, da sie niemals eintritt.

Die einelementigen Ereignisse heißen **Elementarereignisse**.

So gibt es beim einmaligen Wurf eines Würfels die Elementarereignisse

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

2.1.2 Operationen zwischen Ereignissen

Da Ereignisse Mengen sind, können wir mit ihnen die üblichen Mengenoperationen durchführen.

Wir schauen uns die für folgendes Beispiel an:

Beispiel 2.1.11 *Einmaliger Wurf eines Würfels.*

Wir betrachten die Ereignisse:

- *A: Wurf einer geraden Augenzahl.*
 $A = \{2, 4, 6\}$
- *B: Wurf einer ungeraden Augenzahl.*
 $B = \{1, 3, 5\}$
- *C: Wurf einer 6*
 $C = \{6\}$

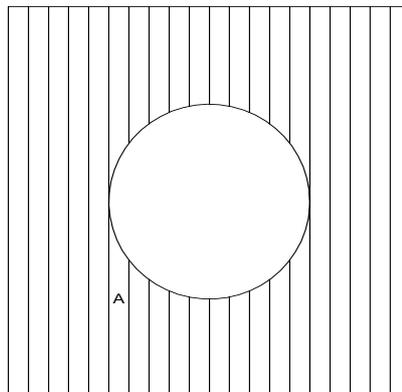
Das **Komplementärereignis** \bar{A} tritt ein, wenn das Ereignis A nicht eintritt:

$$\bar{A} = \{x | x \in \Omega, x \notin A\} \quad (2.1)$$

Man sagt auch:

A tritt nicht ein

Das folgende Bild veranschaulicht den Sachverhalt. Dabei ist der schraffierte Teil gleich \bar{A} .



Für das Beispiel 2.1.11 gilt

$$\begin{aligned}\bar{A} &= \{1, 3, 5\} \\ \bar{B} &= \{2, 4, 6\} \\ \bar{C} &= \{1, 2, 3, 4, 5\}\end{aligned}$$

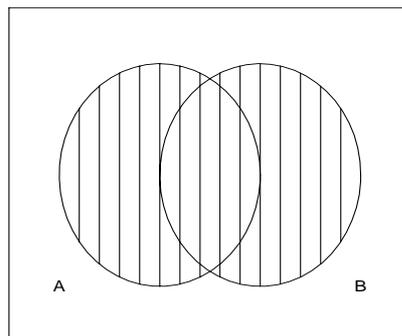
Sind A und B Ereignisse, dann ist die **Vereinigung** $A \cup B$ der beiden definiert durch

$$A \cup B = \{\omega \mid \omega \in \Omega, \omega \in A \text{ oder } \omega \in B\} \quad (2.2)$$

Man sagt auch:

mindestens eines der beiden Ereignisse tritt ein

Das folgende Bild veranschaulicht den Sachverhalt:



Für das Beispiel 2.1.11 gilt

$$\begin{aligned}A \cup B &= \{1, 2, 3, 4, 5, 6\} \\ A \cup C &= \{2, 4, 6\} \\ B \cup C &= \{1, 3, 5, 6\}\end{aligned}$$

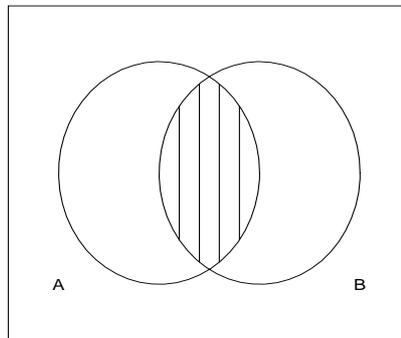
Sind A und B Ereignisse, dann ist der **Schnitt** $A \cap B$ der beiden definiert durch

$$A \cap B = \{\omega \mid \omega \in \Omega, \omega \in A \text{ und } \omega \in B\} \quad (2.3)$$

Man sagt auch:

beide Ereignisse treten gleichzeitig ein

Das folgende Bild veranschaulicht den Sachverhalt:



Für das Beispiel 2.1.11 gilt

$$\begin{aligned} A \cap B &= \emptyset \\ A \cap C &= \{6\} \\ B \cap C &= \emptyset \end{aligned}$$

Definition 2.1.2 *Gilt*

$$A \cap B = \emptyset$$

für zwei Ereignisse A und B , dann heißen A und B **disjunkt**.
Inhaltlich heißt dies, daß A und B nicht gleichzeitig eintreten können.

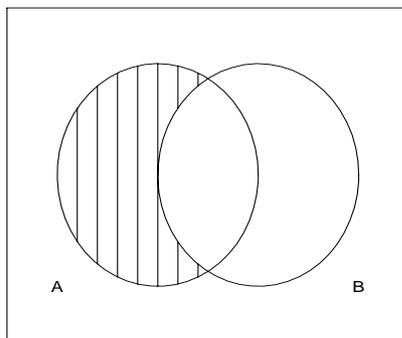
Sind A und B Ereignisse, dann ist die **Differenz** A/B der beiden definiert durch

$$A/B = \{\omega \mid \omega \in \Omega, \omega \in A \text{ und } \omega \notin B\} \quad (2.4)$$

Man sagt auch:

Nur A tritt ein

Das folgende Venndiagramm veranschaulicht den Sachverhalt:



Für das Beispiel 2.1.11 gilt

$$A/B = \{2, 4, 6\}$$

$$A/C = \{2, 4\}$$

$$B/C = \{1, 3, 5\}$$

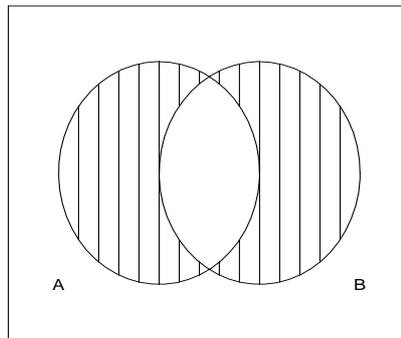
Sind A und B Ereignisse, dann ist die **symmetrische Differenz** $A\Delta B$ der beiden definiert durch

$$A\Delta B = (A \cap \bar{B}) \cup (B \cap \bar{A}) \quad (2.5)$$

Man sagt auch:

Genau eines der beiden Ereignisse tritt ein

Das folgende Venndiagramm veranschaulicht den Sachverhalt:



Für das Beispiel 2.1.11 gilt

$$A\Delta B = \Omega$$

$$A\Delta C = \{2, 4\}$$

$$B\Delta C = \{1, 3, 5, 6\}$$

Für zwei Ereignisse A und B gelten die de Morganschen Regeln

$$\overline{A \cap B} = \bar{A} \cup \bar{B} \quad (2.6)$$

und

$$\overline{A \cup B} = \bar{A} \cap \bar{B} \quad (2.7)$$

Für eine Folge A_1, A_2, \dots von Ereignissen gilt entsprechend

$$\overline{\bigcup_{i=1}^{\infty} A_i} = \bigcap_{i=1}^{\infty} \bar{A}_i \quad (2.8)$$

und

$$\overline{\bigcap_{i=1}^{\infty} A_i} = \bigcup_{i=1}^{\infty} \bar{A}_i \quad (2.9)$$

Für drei Ereignisse A , B und C gelten die Distributivgesetze

- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

und für eine Folge A_1, A_2, \dots von Ereignissen und ein Ereignis B gilt entsprechend

$$B \cap \left(\bigcup_{i=1}^{\infty} A_i \right) = \bigcup_{i=1}^{\infty} (A_i \cap B) \quad (2.10)$$

und

$$B \cup \left(\bigcap_{i=1}^{\infty} A_i \right) = \bigcap_{i=1}^{\infty} (A_i \cup B) \quad (2.11)$$

2.1.3 σ -Algebren und Meßräume

Unser Ziel ist es, die Chance des Eintretens von Ereignissen bei Zufallsvorgängen zu bewerten. Wünschenswert ist es, jedem möglichen Ereignis eine reelle Zahl zuzuordnen, die die Chance seines Eintretens angibt.

Definition 2.1.3 *Sei M eine Menge.*

Dann heißt die Menge aller Teilmengen von M Potenzmenge $\mathcal{P}(M)$.

Beispiel 2.1.12 *Sei*

$$M = \emptyset.$$

Dann gilt

$$\mathcal{P}(M) = \{\emptyset\}$$

Beispiel 2.1.13 *Sei*

$$M = \{1\}$$

Dann gilt

$$\mathcal{P}(M) = \{\emptyset, M\}$$

Beispiel 2.1.14 *Sei*

$$M = \{K, Z\}.$$

Dann gilt

$$\mathcal{P}(M) = \{\emptyset, \{K\}, \{Z\}, \{K, Z\}\}$$

Sei M eine endliche Menge vom Umfang n . Dann enthält $\mathcal{P}(M)$ genau 2^n Elemente. Wir werden dies im Kapitel über Kombinatorik beweisen.

Es liegt nahe, als Definitionsbereich der Bewertungsfunktion der Chance des Eintretens die Potenzmenge zu wählen. Dies ist im Falle diskreter Ergebnismengen problemlos möglich. Bei stetigen Ergebnismengen stößt man jedoch auf Probleme. Es ist nicht möglich, eine Bewertungsfunktion anzugeben, die sinnvolle Bedingungen einhält und gleichzeitig alle Elemente der Potenzmenge widerspruchsfrei bewertet.

Aus diesem Grunde betrachtet man nicht die Potenzmenge, sondern die Menge von Ereignissen, die neben den im Zusammenhang mit einem Zufallsexperiment unmittelbar interessierenden Ereignissen auch alle diejenigen enthält, die sich durch Anwendung der Komplementärbildung, Vereinigung und Durchschnitt ergeben.

Definition 2.1.4 Sei Ω eine Ergebnismenge.

Ein Menge \mathcal{A} von Ereignissen aus Ω heißt σ -Algebra über Ω , wenn gilt

$$\emptyset \in \mathcal{A} \quad (2.12)$$

$$A \in \mathcal{A} \Rightarrow \bar{A} \in \mathcal{A} \quad (2.13)$$

$$A_1, A_2, \dots \in \mathcal{A} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A} \quad (2.14)$$

Satz 2.1.1 Sei Ω eine Ergebnismenge und \mathcal{A} eine σ -Algebra über Ω dann gilt

1. $\Omega \in \mathcal{A}$
2. $A_1, A_2, \dots \in \mathcal{A} \Rightarrow \bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$

Beweis:

1. Wegen (2.12) und (2.13) gilt

$$\bar{\emptyset} \in \mathcal{A}$$

Wegen $\Omega = \bar{\emptyset}$ gilt die Behauptung.

2. Sind

$$A_1, A_2, \dots, A_n \in \mathcal{A},$$

so sind wegen (2.13) auch

$$\bar{A}_1, \bar{A}_2, \dots \in \mathcal{A},$$

und wegen (2.14) auch

$$\bigcup_{i=1}^{\infty} \bar{A}_i \in \mathcal{A}$$

Wegen (2.13) gilt dann auch

$$\overline{\bigcup_{i=1}^{\infty} \bar{A}_i} \in \mathcal{A}$$

Es gilt

$$\overline{\bigcup_{i=1}^{\infty} \bar{A}_i} = \bigcap_{i=1}^{\infty} \overline{\bar{A}_i} = \bigcap_{i=1}^{\infty} A_i$$

Also gilt

$$\bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$$

Schauen wir uns ein Beispiel an:

Beispiel 2.1.15 *Einmaliger Wurf eines Würfels.*

Es gilt

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

Wir sind interessiert am Ereignis $A = \{6\}$, daß wir eine 6 werfen.

Dann ist

$$\mathcal{A} = \{\emptyset, \Omega, A, \bar{A}\}$$

eine σ -Algebra über Ω .

Wir prüfen die drei Bedingungen von Definition 2.1.3.

1. $\emptyset \in \mathcal{A}$ ist erfüllt.

2. *Es gilt*

$$\begin{aligned} \bar{\emptyset} &= \Omega \in \mathcal{A} \\ \bar{\Omega} &= \emptyset \in \mathcal{A} \\ \bar{A} &\in \mathcal{A} \\ \bar{\bar{A}} &= A \in \mathcal{A} \end{aligned}$$

3. *Es gilt*

$$\begin{aligned} \emptyset \cup \Omega &= \Omega \in \mathcal{A} \\ \emptyset \cup A &= A \in \mathcal{A} \\ \emptyset \cup \bar{A} &= \bar{A} \in \mathcal{A} \\ \Omega \cup A &= \Omega \in \mathcal{A} \\ \Omega \cup \bar{A} &= \Omega \in \mathcal{A} \\ A \cup \bar{A} &= \Omega \in \mathcal{A} \\ \emptyset \cup \Omega \cup A &= \Omega \in \mathcal{A} \\ \emptyset \cup \Omega \cup \bar{A} &= \Omega \in \mathcal{A} \\ \Omega \cup A \cup \bar{A} &= \Omega \in \mathcal{A} \\ \emptyset \cup A \cup \bar{A} &= \Omega \in \mathcal{A} \end{aligned}$$

Bisher haben wir die bei Zufallsvorgängen interessierenden Ereignisse mit Hilfe von Mengen mathematisch beschrieben. Im täglichen Leben bewerten wir Ereignisse hinsichtlich ihrer Realisierungsmöglichkeit unterschiedlich.

Wie können wir dies quantifizieren?

2.2 Wahrscheinlichkeit

2.2.1 Vorbemerkungen

Formal sind Wahrscheinlichkeiten Zahlenwerte, die wir den Ereignissen eines Zufallsvorgangs zuordnen. Die Wahrscheinlichkeitsrechnung nimmt an, daß diese Zahlen irgendwie gegeben sind und bestimmten trivialen Konsistenzbedingungen (Axiomen) genügen. Es ist der Wahrscheinlichkeitstheorie jedoch völlig gleichgültig, woher man in einem konkreten Fall die Zahlen bekommt. Intuitiv bedeuten Wahrscheinlichkeiten Grade der Realisierungschance von Ereignissen oder Grade des subjektiven Glaubens an Ereignisse.

Wir wollen zunächst einige Ansätze zur Formalisierung der intuitiven Vorstellung angeben.

2.2.2 Klassischer Ansatz

Ausgangspunkt ist ein Zufallsvorgang mit endlich vielen Ergebnissen.

Die Wahrscheinlichkeit eines Ereignisses A ist definiert durch

$$P(A) = \frac{\text{Anzahl der Ergebnisse in } A}{\text{Anzahl der Ergebnisse in } \Omega}.$$

Beispiel 2.2.1 *Wurf eines Würfels. Wir unterstellen, daß der Würfel fair ist.*

Es gilt also für die Elementarereignisse

$$\begin{aligned} P(\{1\}) &= \frac{1}{6} \\ P(\{2\}) &= \frac{1}{6} \\ P(\{3\}) &= \frac{1}{6} \\ P(\{4\}) &= \frac{1}{6} \\ P(\{5\}) &= \frac{1}{6} \\ P(\{6\}) &= \frac{1}{6}. \end{aligned}$$

Für $A = \{1, 3, 5\}$ zum Beispiel gilt

$$P(A) = \frac{3}{6}$$

Den klassischen Ansatz kann man nur verwenden, wenn die Ergebnismenge endlich ist und alle Elementarereignisse gleichwahrscheinlich sind.

2.2.3 Frequentistischer Ansatz

Eine Möglichkeit, zu einer Bewertung eines Ereignisses im Rahmen eines Zufallsvorgangs zu gelangen, besteht darin, den Zufallsvorgang mehrmals unter identischen Bedingungen zu beobachten und zu zählen, wie oft das interessierende Ereignis eingetreten ist. Man sammelt also Erfahrungen über die Realisierungsmöglichkeiten eines Ereignisses durch Beobachten und Zählen. Dies ist der klassische Ansatz.

Das Ereignis A sei bei den n Durchführungen des Zufallsvorgangs $H_n(A)$ -mal eingetreten. Wir nennen $H_n(A)$ die absolute Häufigkeit des Ereignisses A . Setzen wir $H_n(A)$ in Beziehung zur Anzahl n der Durchführungen, so erhalten wir die relative Häufigkeit

$$h_n(A) = \frac{H_n(A)}{n}$$

Für die relative Häufigkeit $h_n(A)$ gilt

1. $h_n(\Omega) = 1$
2. $0 \leq h_n(A) \leq 1$ für jedes Ereignis A
3. $h_n(A \cup B) = h_n(A) + h_n(B)$ für disjunkte Ereignisse A und B .

In der Realität ist nun folgendes zu beobachten:

Wiederholen wir einen Zufallsvorgang immer wieder unter den gleichen Bedingungen und bilden für ein beliebiges Ereignis A die Folge der relativen Häufigkeiten $h_n(A)$, $n = 1, 2, \dots$, dann schwanken die $h_n(A)$ mit wachsendem n immer weniger und scheinen einem Grenzwert zuzustreben. Die Folge der relativen Häufigkeiten zeigt ein konvergenzartiges Verhalten.

Es scheint so, als ob gilt

$$\lim_{n \rightarrow \infty} h_n(A) = p$$

Es liegt nun nahe, die Chance der Realisierung von Ereignissen über diesen Grenzwert zu definieren. Die Konvergenz läßt sich aber in der Realität weder verifizieren noch falsifizieren, da wir nur ein Anfangsstück der Reihe beobachten können.

Eine Bewertung der Chance des Eintretens von Ereignissen über den Grenzwert der Folge der relativen Häufigkeiten ist also nicht möglich, da über die Existenz dieses Grenzwerts nichts bekannt ist.

2.2.4 Axiomatische Definition

Bei vielen Zufallsvorgängen ist zu beobachten, daß die relative Häufigkeit eines Ereignisses mit wachsender Zahl der Wiederholungen immer weniger schwankt. Es liegt also nahe, den Grad der Unbestimmtheit des Eintretens eines Ereignisses durch eine Zahl zu charakterisieren. Dabei sollte eine zahlenmäßige Bewertung von Ereignissen die gleichen Eigenschaften wie die relative Häufigkeit besitzen. Sie wird **Wahrscheinlichkeit** genannt.

Kolmogoroff hat 1933 folgende Definition der Wahrscheinlichkeit gegeben:

Definition 2.2.1 Sei Ω eine Ergebnismenge eines Zufallsvorgangs und \mathcal{A} eine σ -Algebra über Ω .

Die Abbildung $P : \mathcal{A} \rightarrow \mathfrak{R}$ heißt Wahrscheinlichkeitsmaß auf (Ω, \mathcal{A}) , wenn P jedem $A \in \mathcal{A}$ eine reelle Zahl $P(A)$ zuordnet mit

$$0 \leq P(A) \quad \text{für alle } A \in \mathcal{A} \quad (2.15)$$

$$P(\Omega) = 1 \quad (2.16)$$

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \quad (2.17)$$

für Ereignisse A_i , $i = 1, 2, \dots$ mit $A_i \cap A_j = \emptyset$ für $i \neq j$.

Man spricht auch von den drei Axiomen der Wahrscheinlichkeitsrechnung.

Das erste Axiom stellt sicher, daß Wahrscheinlichkeiten nichtnegative Zahlen sind.

Das zweite Axiom normiert Wahrscheinlichkeiten, so daß man sagen kann, ob eine Wahrscheinlichkeit groß oder klein ist.

Das Tripel (Ω, \mathcal{A}, P) heißt Wahrscheinlichkeitsraum.

Bevor wir uns ein Beispiel anschauen, schauen wir uns noch an, welche Konsequenzen wir aus den Axiomen ziehen können.

Satz 2.2.1 Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum.

Dann gilt

$$P(\emptyset) = 0. \quad (2.18)$$

Beweis:

Setze $A_i = \emptyset$ für $i = 1, 2, \dots$

Es gilt

$$\emptyset = \bigcup_{i=1}^{\infty} A_i \quad \text{mit } A_i \cap A_j = \emptyset \quad \text{für } i \neq j.$$

Aufgrund von (2.17) gilt also

$$P(\emptyset) = P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) = \sum_{i=1}^{\infty} P(\emptyset) = P(\emptyset) + \sum_{i=2}^{\infty} P(\emptyset)$$

Da Wahrscheinlichkeiten nichtnegativ sind, folgt hieraus

$$P(\emptyset) = 0$$

Satz 2.2.2 Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und A_1, \dots, A_n paarweise disjunkte Ereignisse in \mathcal{A} .

Dann gilt

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) \quad (2.19)$$

Beweis:

Sei

$$A_{n+1} = A_{n+2} = \dots = \emptyset.$$

Dann gilt wegen $A_i \cap A_j = \emptyset$ für $i \neq j$ und $A_i \cup \emptyset = A_i$:

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &= P\left(\bigcup_{i=1}^{\infty} A_i\right) \\ &= \sum_{i=1}^{\infty} P(A_i) \\ &= \sum_{i=1}^n P(A_i) + \sum_{i=n+1}^{\infty} P(\emptyset) \\ &= \sum_{i=1}^n P(A_i) \end{aligned}$$

da wegen (2.18) $P(\emptyset) = 0$ gilt.

Satz 2.2.3 Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum.
Dann gilt

$$P(\bar{A}) = 1 - P(A) \quad (2.20)$$

für alle $A \in \mathcal{A}$.

Beweis:

Es gilt

$$A \cup \bar{A} = \Omega$$

und

$$A \cap \bar{A} = \emptyset$$

Dann gilt wegen (2.19):

$$1 = P(\Omega) = P(A \cup \bar{A}) = P(A) + P(\bar{A})$$

Hieraus folgt die Behauptung.

Satz 2.2.4 Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum.
Dann gilt

$$P(A \cap \bar{B}) = P(A) - P(A \cap B) \quad (2.21)$$

für alle $A, B \in \mathcal{A}$.

Beweis:

Es gilt

$$A = (A \cap B) \cup (A \cap \bar{B})$$

Wegen

$$(A \cap B) \cap (A \cap \bar{B}) = \emptyset$$

folgt wegen (2.17)

$$P(A) = P((A \cap B) \cup (A \cap \bar{B})) = P((A \cap B)) + P((A \cap \bar{B}))$$

Also gilt

$$P(A \cap \bar{B}) = P(A) - P(A \cap B)$$

Satz 2.2.5 Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum.

Dann gilt

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad (2.22)$$

für alle $A, B \in \mathcal{A}$.

Beweis:

Es gilt

$$A \cup B = (A \cap \bar{B}) \cup B$$

Wegen

$$(A \cap \bar{B}) \cap B = \emptyset$$

folgt dann wegen (2.17) und (2.21)

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

Satz 2.2.6 Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $A, B \in \mathcal{A}$ mit $A \subset B$.

Dann gilt

$$P(A) \leq P(B) \quad (2.23)$$

Beweis:

Aus

$$A \subset B$$

folgt

$$A \cap B = A$$

Es gilt

$$\begin{aligned} P(B) &= P(B \cap A) + P(B \cap \bar{A}) \\ &= P(A) + P(B \cap \bar{A}) \end{aligned}$$

Wegen

$$P(B \cap \bar{A}) \geq 0$$

folgt also

$$P(B) - P(A) \geq 0$$

und somit

$$P(A) \leq P(B)$$

Satz 2.2.7 Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum.
Dann gilt für jedes $A \in \mathcal{A}$:

$$P(A) \leq 1 \quad (2.24)$$

Beweis:

Für jedes $A \in \mathcal{A}$ gilt

$$A \subset \Omega$$

Aus (2.16) und (2.23) folgt

$$P(A) \leq 1$$

Schauen wir uns nun ein Beispiel für ein Wahrscheinlichkeitsmaß an:

Beispiel 2.2.2 Einmaliger Wurf einer Münze.

Es gilt

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

Wir wählen als σ -Algebra $\mathcal{P}(\Omega)$.

Es gilt

$$\mathcal{P}(\Omega) = \{\emptyset, \{K\}, \{Z\}, \{K, Z\}\}$$

Wir setzen

$$P(\{K\}) = p$$

mit $0 \leq p \leq 1$.

Hierdurch ist ein Wahrscheinlichkeitsmaß auf (Ω, \mathcal{A}, P) definiert, denn jedem Element aus \mathcal{A} wird eine Zahl zwischen 0 und 1 zugeordnet.

Wegen (2.18) gilt

$$P(\emptyset) = 0$$

Wegen (2.16) gilt

$$P(\{K, Z\}) = 1$$

Wegen (2.20) gilt

$$P(\{Z\}) = 1 - P(\{K\}) = 1 - p$$

Das Beispiel zeigt, wie man bei einer abzählbaren Ergebnismenge ein Wahrscheinlichkeitsmaß konstruieren kann.

Sei

$$\Omega = \{\omega_1, \omega_2, \dots, \}$$

Wir legen Wahrscheinlichkeiten $P(\{\omega_i\}) = p_i$ für die Elementarereignisse $\{\omega_i\}$, $i = 1, 2, \dots$, fest.

Dabei muß gelten

$$0 \leq p_i \leq 1$$

und

$$\sum_i p_i = 1.$$

Für jedes $A \in \mathcal{P}(\Omega)$ setzen wir dann

$$P(A) = \sum_{\omega_i \in A} p_i$$

Hierdurch erhalten wir ein Wahrscheinlichkeitsmaß über $(\Omega, \mathcal{P}(\Omega))$.

Wir zeigen die Gültigkeit der drei Axiome:

1. $P(A) = \sum_{\omega_i \in A} p_i \geq 0$, da $p_i \geq 0$.
2. $P(\Omega) = \sum_{\omega_i \in \Omega} p_i = \sum_i p_i = 1$.
3. Seien $A_1, A_2, \dots, \in \mathcal{A}$ mit $A_i \cap A_j = \emptyset$ für $i \neq j$.

Dann gilt

$$\begin{aligned} P(\cup_i A_i) &= \sum_{\omega_i \in \cup_i A_i} p_i \\ &= \sum_{\omega_i \in A_1} p_i + \sum_{\omega_i \in A_2} p_i + \dots \\ &= P(A_1) + P(A_2) + \dots \end{aligned}$$

2.2.5 Kombinatorik

Ausgangspunkt der Kombinatorik ist das Gleichmöglichkeitsmodell.

Wir gehen also von einer endlichen Ergebnismenge $\Omega = \{\omega_1, \dots, \omega_N\}$ aus und nehmen an, daß jedes der Elementarereignisse gleichwahrscheinlich ist.

Es gilt also

$$P(\{\omega_i\}) = \frac{1}{N}.$$

Die Wahrscheinlichkeit eines beliebigen Ereignisses $A \subset \Omega$ ist dann

$$P(A) = \sum_{\omega_i \in A} P(\{\omega_i\}) = \frac{|A|}{|\Omega|}$$

wobei $|M|$ die Mächtigkeit, d.h. die Anzahl der Elemente einer Menge M , angibt.

Wir müssen also die Mächtigkeit von Mengen bestimmen.

Oft ist es gar nicht so einfach, die Anzahl der möglichen Ergebnisse zu bestimmen.

Dem folgenden liegt ein Grundprinzip zugrunde, das wir uns zunächst an einem Beispiel anschauen.

Beispiel 2.2.3 *Ein Mann besitzt drei Pullover und zwei Hosen.*

Die Farben der Pullover sind rot, gelb und blau, während die Farben der Hosen schwarz und weiß sind.

Auf wieviele Arten kann sich der Mann kleiden?

Er kann den roten Pullover anziehen und zu diesem entweder die schwarze oder die weiße Hose.

Er kann den gelben Pullover anziehen und zu diesem entweder die schwarze oder die weiße Hose.

Er kann den blauen Pullover anziehen und zu diesem entweder die schwarze oder die weiße Hose.

Es gibt also die folgenden 6 Möglichkeiten:

- 1. roten Pullover und schwarze Hose*
- 2. roten Pullover und weiße Hose*
- 3. gelber Pullover und schwarze Hose*
- 4. gelber Pullover und weiße Hose*
- 5. blauer Pullover und schwarze Hose*

6. blauer Pullover und weiße Hose

Zu jedem der drei Pullover kann man jede der beiden Hosen anziehen.

Allgemein heißt dies

Satz 2.2.8 Seien $A = \{a_1, \dots, a_m\}$ und $B = \{b_1, \dots, b_n\}$ Mengen.
Für das kartesische Produkt

$$A \times B = \{(a, b) | a \in A, b \in B\} \quad (2.25)$$

gilt

$$|A \times B| = |A| \cdot |B| = m \cdot n \quad (2.26)$$

Beweis:

Der Beweis ist offensichtlich, wenn man sich die möglichen Ergebnisse folgendermaßen hinschreibt.

| | b_1 | b_2 | \dots | b_j | \dots | b_n |
|----------|--------------|--------------|----------|--------------|----------|--------------|
| a_1 | (a_1, b_1) | (a_1, b_2) | \dots | (a_1, b_j) | \dots | (a_1, b_n) |
| a_2 | (a_2, b_1) | (a_2, b_2) | \dots | (a_2, b_j) | \dots | (a_2, b_n) |
| \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots |
| a_i | (a_i, b_1) | (a_i, b_2) | \dots | (a_i, b_j) | \dots | (a_i, b_n) |
| \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots |
| a_m | (a_m, b_1) | (a_m, b_2) | \dots | (a_m, b_j) | \dots | (a_m, b_n) |

Hieraus folgt sofort die Verallgemeinerung

Satz 2.2.9 Seien A_1, \dots, A_r endliche Mengen.
Für das kartesische Produkt

$$A_1 \times \dots \times A_r = \{(a_1, \dots, a_r) | a_i \in A_i, i = 1, \dots, r\} \quad (2.27)$$

gilt

$$|A_1 \times A_2 \times \dots \times A_r| = |A_1| \cdot |A_2| \cdot \dots \cdot |A_r| \quad (2.28)$$

Schauen wir uns ein Beispiel an.

Beispiel 2.2.4 Aus $n_1 = 5$ Wegen von A nach B und $n_2 = 4$ Wegen von B nach C kann man $n_1 \cdot n_2 = 5 \cdot 4 = 20$ Reisen von A nach C über B zusammenstellen.

Schauen wir uns nun die klassische Fragestellung der Kombinatorik an. Dessen liegt folgendes Urnenmodell zugrunde:

Ein Urne enthält n Kugeln, die von 1 bis n durchnummeriert sind. Der Urne werden nacheinander k Kugeln entnommen. Nach jedem Zug notiert man die Nummer der gezogenen Kugel und legt die Kugel entweder zurück (Ziehen mit Zurücklegen) oder legt die Kugel zur Seite (Ziehen ohne Zurücklegen).

Neben der Art der Ziehung interessiert noch die Reihenfolge, in der die Kugeln gezogen werden.

Beispiel 2.2.5 *Wir gehen davon aus, daß die Urne vier Kugeln enthält, die von 1 bis 4 durchnummeriert sind.*

$$U = \{1, 2, 3, 4\}$$

Wir fangen mit dem Ziehen mit Zurücklegen unter Berücksichtigung der Anordnung an.

Schauen wir uns dies für das Beispiel an:

Beispiel 2.2.6 *Vor dem ersten Zug ist der Zustand der Urne*

$$U = \{1, 2, 3, 4\}$$

und, da wir mit Zurücklegen ziehen, ist er vor dem zweiten Zug ebenfalls

$$U = \{1, 2, 3, 4\}$$

Die Zahl auf der ersten gezogenen Kugel ist entweder eine 1, 2, 3 oder 4.

Nehmen wir an, es ist eine 1. Dann kann die Zahl auf der zweiten gezogenen Kugel entweder eine 1, 2, 3 oder 4 sein.

Wir erhalten also folgende 4 Möglichkeiten

$$(1, 1) \quad (1, 2) \quad (1, 3) \quad (1, 4)$$

Zu jeder der anderen 3 ersten Zahlen beim ersten Zug gibt es ebenfalls wieder 4 Möglichkeiten beim zweiten Zug.

Somit gibt es folgende $4^2 = 16$ Möglichkeiten:

$$\begin{array}{cccc} (1, 1) & (1, 2) & (1, 3) & (1, 4) \\ (2, 1) & (2, 2) & (2, 3) & (2, 4) \\ (3, 1) & (3, 2) & (3, 3) & (3, 4) \\ (4, 1) & (4, 2) & (4, 3) & (4, 4) \end{array}$$

Der folgende Satz verallgemeinert das Beispiel.

Satz 2.2.10 *Die Anzahl der geordneten Stichproben vom Umfang k aus einer Menge vom Umfang n beträgt beim Ziehen mit Zurücklegen*

$$n^k \tag{2.29}$$

Beweis:

Sei B eine Menge mit n Elementen.

Die Menge aller möglichen geordneten Stichproben ist beim Ziehen mit Zurücklegen

$$A = \{(a_1, \dots, a_k) \mid a_i \in B, i = 1, \dots, k\} = B \times B \times \dots \times B$$

Vor jeder Ziehung enthält die Menge, aus der gezogen wird, also n Elemente. Aufgrund von Satz 2.2.9 gilt also

$$|A| = |B \times B \times \dots \times B| = \underbrace{|B| \cdot \dots \cdot |B|}_{k \text{ Faktoren}} = |B|^k = n^k$$

Schauen wir uns einige Beispiele an:

Beispiel 2.2.7 *Mit einer Base lassen sich 4, mit zwei Basen $4 \cdot 4 = 16$ Aminosäuren kodieren. Für 3 Basen haben wir $4 \cdot 4 \cdot 4 = 64 > 20$ Möglichkeiten. Tatsächlich bilden jeweils ein Tripel von Basen die Grundbausteine des genetischen Codes.*

Beispiel 2.2.8 *Ein Byte besteht aus 8 Bits, wobei ein Bit ein Informationseinheit ist, die zwei unterschiedliche Zustände, zum Beispiel 0 oder 1 annehmen kann. In einem Byte können also $2 \cdot 2 = 2^8 = 256$ unterschiedliche Informationen gespeichert werden.*

Beispiel 2.2.9 *Eine Menge M vom Umfang n hat 2^n Teilmengen.*

Diese sieht man folgendermaßen:

Ohne Beschränkung der Allgemeinheit nehmen wir an, daß gilt

$$M = \{1, 2, \dots, n-1, n\}.$$

Wir definieren einen Vektor $v = (v_1, v_2, \dots, v_n)$ mit

$$v_i = \begin{cases} 1 & \text{falls } i \text{ in der Teilmenge} \\ 0 & \text{sonst} \end{cases}$$

Ist zum Beispiel

$$M = \{1, 2, 3\},$$

so liefert der Vektor

$$v = \{1, 0, 1\}$$

die Teilmenge

$$\{1, 3\},$$

während der Vektor

$$v = \{0, 0, 0\}$$

die leere Menge \emptyset liefert.

Zu jeder Teilmenge gibt es genau einen Vektor v .

Da es 2^n Bitvektoren der Länge n gibt, hat eine Menge vom Umfang n genau 2^n Teilmengen.

Beispiel 2.2.10 Wir betrachten ein Beispiel aus Tversky und Kahnemann.

Personen wurde jedes der folgenden beiden Muster gezeigt:

Muster 1

```

X X X X X X X X
X X X X X X X X
X X X X X X X X

```

Muster 2

```

X X
X X
X X
X X
X X
X X
X X
X X
X X
X X

```

Sie sollten sagen, bei welchem der beiden Muster es mehr Pfade von der ersten Zeile bis zur letzten Zeile gibt, wobei man auf einem beliebigen X der ersten Zeile startet und in jeder der darunterliegenden Zeilen auf genau ein Symbol geht.

Für Muster 1 gibt es $8^3 = 512$ unterschiedliche Pfade, da man drei Symbole trifft, und es für jedes dieser Symbole genau 8 Möglichkeiten gibt.

Für Muster 2 gibt es $2^9 = 512$ unterschiedliche Pfade, da man neun Symbole trifft, und es für jedes dieser Symbole genau 2 Möglichkeiten gibt.

Bei beiden Mustern ist die Anzahl der Pfade also gleich.

Von den befragten Personen fanden 85 Prozent, daß es bei Muster 1 mehr Pfade gibt.

Dies liegt daran, daß man sich die unterschiedlichen Pfade bei Muster 1 leichter vorstellen kann.

Wenden wir uns nun dem Ziehen ohne Zurücklegen mit Berücksichtigung der Anordnung zu.

Schauen wir uns dies für das Beispiel an:

Beispiel 2.2.11 *Vor dem ersten Zug ist der Zustand der Urne*

$$U = \{1, 2, 3, 4\}$$

Da wir ohne Zurücklegen ziehen, hängt er vor dem zweiten Zug vom Ergebnis des ersten Zuges ab.

Nehmen wir an, daß die erste gezogene Kugel eine 1 ist.

Dann kann die Zahl auf der zweiten gezogenen Kugel entweder eine 2, 3 oder 4 sein.

Wir erhalten also folgende 3 Möglichkeiten

$$(1, 2) \quad (1, 3) \quad (1, 4)$$

Zu jeder der anderen 3 ersten Zahlen beim ersten Zug gibt es ebenfalls wieder 3 Möglichkeiten beim zweiten Zug.

Somit gibt es folgende $4 \cdot 3 = 12$ Möglichkeiten:

$$\begin{array}{ccc} & (1, 2) & (1, 3) & (1, 4) \\ (2, 1) & & (2, 3) & (2, 4) \\ (3, 1) & (3, 2) & & (3, 4) \\ (4, 1) & (4, 2) & (4, 3) & \end{array}$$

Der folgende Satz verallgemeinert das Beispiel.

Satz 2.2.11 *Die Anzahl der geordneten Stichproben vom Umfang k aus einer Menge vom Umfang n beträgt beim Ziehen ohne Zurücklegen*

$$(n)_k = n \cdot (n - 1) \cdots (n - k + 1) \tag{2.30}$$

Beweis:

Sei B eine Menge mit n Elementen.

Die Menge aller möglichen geordneten Stichproben ist beim Ziehen ohne Zurücklegen

$$\begin{aligned} A &= \{(a_1, \dots, a_k) \mid a_i \in A_i, |A_i| = n - i + 1, i = 1, \dots, k\} \\ &= A_1 \times A_2 \times \dots \times A_k \end{aligned}$$

Bei jeder Ziehung enthält die Menge, aus der gezogen wird, also ein Element weniger, bei der ersten Ziehung n Elemente, bei der zweiten $n - 1$ Elemente, u.s.w..

Aufgrund von Satz 2.2.9 gilt also

$$|A| = |A_1 \times A_2 \times \dots \times A_k| = |A_1| \cdot |A_2| \cdot \dots \cdot |A_k| = n \cdot (n - 1) \cdot \dots \cdot (n - k + 1)$$

$(n)_k$ wird gelesen als 'n sub k'.

Satz 2.2.12 *Es gilt*

$$(n)_k = \frac{n!}{(n - k)!}$$

Beweis:

Es gilt

$$\begin{aligned} (n)_k &= n \cdot (n - 1) \cdot \dots \cdot (n - k + 1) \\ &= \frac{n \cdot (n - 1) \cdot \dots \cdot (n - k + 1) \cdot (n - k) \cdot (n - k - 1) \cdot \dots \cdot 2 \cdot 1}{(n - k) \cdot (n - k - 1) \cdot \dots \cdot 2 \cdot 1} \\ &= \frac{n!}{(n - k)!} \end{aligned}$$

Die Menge aller möglichen geordneten Stichproben vom Umfang n aus einer Menge, die n Elemente enthält, ist also:

$$n! = n \cdot (n - 1) \cdot \dots \cdot 2 \cdot 1$$

Man spricht auch von den Permutationen einer Menge.

Wir lesen $n!$ als 'n Fakultät'.

Wir setzen $0! = 1$.

Die folgende Tabelle gibt die Werte von $n!$ für $n \leq 10$ an:

| n | $n!$ |
|-----|---------|
| 0 | 1 |
| 1 | 1 |
| 2 | 2 |
| 3 | 6 |
| 4 | 24 |
| 5 | 120 |
| 6 | 720 |
| 7 | 5040 |
| 8 | 40320 |
| 9 | 362880 |
| 10 | 3628800 |

Beispiel 2.2.12 *Wir betrachten die Menge*

$$U = \{1, 2, 3, 4\}$$

Es gibt $4! = 24$ Permutationen der Elemente der Menge U .

$$\begin{array}{cccccc} (1, 2, 3, 4) & (1, 2, 4, 3) & (1, 3, 2, 4) & (1, 3, 4, 2) & (1, 4, 2, 3) & (1, 4, 3, 2) \\ (2, 1, 3, 4) & (2, 1, 4, 3) & (2, 3, 1, 4) & (2, 3, 4, 1) & (2, 4, 1, 3) & (2, 4, 3, 1) \\ (3, 1, 2, 4) & (3, 1, 4, 2) & (3, 2, 1, 4) & (3, 2, 4, 1) & (3, 4, 1, 3) & (3, 4, 3, 1) \\ (4, 1, 2, 3) & (4, 1, 3, 2) & (4, 2, 1, 3) & (4, 2, 3, 1) & (4, 3, 1, 2) & (4, 3, 2, 1) \end{array}$$

Schauen wir uns noch ein Beispiel an.

Beispiel 2.2.13 *6 Personen können auf $6! = 720$ unterschiedliche Arten auf 6 Stühlen nebeneinander sitzen.*

Wir bringen nun die beiden Situationen zusammen.

Beispiel 2.2.14 *Ein fairer Würfel wird sechsmal hintereinander geworfen. Wie groß ist die Wahrscheinlichkeit, daß alle sechs Augenzahlen auftreten? Wir benutzen das Gleichmöglichkeitsmodell. Die Anzahl der möglichen Ergebnisse ist*

$$6^6 = 46656,$$

da wir mit Zurücklegen 6 Zahlen aus der Menge $\{1, 2, 3, 4, 5, 6\}$ ziehen. Die Anzahl der günstigen Fälle ist

$$6! = 720,$$

da wir ohne Zurücklegen 6 Zahlen aus der Menge $\{1, 2, 3, 4, 5, 6\}$ ziehen. Die gesuchte Wahrscheinlichkeit ist also

$$\frac{6!}{6^6} = 0.0154321$$

Beispiel 2.2.15 *In einem Zimmer befinden sich drei Personen. Wie groß ist die Wahrscheinlichkeit, daß mindestens zwei von ihnen am gleichen Tag des Jahres Geburtstag haben?*

Sei A_3 : mindestens zwei der drei Personen haben am gleichen Tag des Jahres Geburtstag.

Es ist einfacher, die Wahrscheinlichkeit des Gegenereignisses zu bestimmen.

Es ist $\overline{A_3}$: jede der drei Personen hat an einem anderen Tag des Jahres Geburtstag.

Wir betrachten wieder das Gleichmöglichkeitsmodell.

Die Anzahl der möglichen Fälle ist

$$365^3 = 48627125,$$

da wir mit Zurücklegen drei Zahlen aus der Menge $\{1, 2, 3, \dots, 364, 365\}$ ziehen.

Die Anzahl der günstigen Fälle ist

$$(365)_3 = 48228180,$$

da wir ohne Zurücklegen drei Zahlen aus der Menge $\{1, 2, 3, \dots, 364, 365\}$ ziehen.

Es gilt also

$$P(\overline{A_3}) = \frac{(365)_3}{365^3} = 0.9917958$$

Also gilt

$$P(A_3) = 1 - P(\overline{A_3}) = 0.0082042$$

Wie nicht anders zu erwarten war, ist diese Wahrscheinlichkeit ziemlich klein.

Ab wievielen Leuten lohnt es sich, darauf zu wetten, daß mindesten zwei am gleichen Tag des Jahres Geburtstag haben.

Sei A_k : mindestens zwei von k Leuten haben am gleichen Tag des Jahres Geburtstag.

Es gilt

$$P(A_k) = 1 - \frac{(365)_k}{365^k}$$

Wir erhalten speziell

$$P(A_{20}) = 0.4114384$$

$$P(A_{21}) = 0.4436883$$

$$P(A_{22}) = 0.4756953$$

$$P(A_{23}) = 0.5072972$$

Ab 23 Leuten lohnt es sich, darauf zu wetten, daß mindesten zwei am gleichen Tag des Jahres Geburtstag haben.

Diese Anzahl ist überraschend klein.

In einer amerikanischen Talkshow trug ein Teilnehmer dieses Beispiel vor. Der Moderator wollte es gleich überprüfen und fragte:

Ich habe am 23. Juli Geburtstag.

Wer hat noch am 23. Juli Geburtstag?

Keiner der 74 Zuschauer im Raum meldete sich.

Die Wahrscheinlichkeit, daß von 75 Personen in einem Raum mindestens zwei am gleichen Tag des Jahres Geburtstag haben, beträgt aber

$$P(A_{75}) = 0.9997199$$

Also spricht sehr viel dafür, daß etwas an der Frage des Moderators falsch sein muß.

Der Moderator hat einen Fehler gemacht.

Im Beispiel wurde nach der Wahrscheinlichkeit gefragt, daß zwei Personen an irgendeinem Tag des Jahres Geburtstag haben. Der Moderator wollte wissen, ob zwei Personen an einem bestimmten Tag des Jahres Geburtstag haben.

Sei B_k : mindestens einer von k Personen hat am gleichen Tag wie der Moderator Geburtstag.

Wir betrachten wiederum das Gegenereignis $\overline{B_k}$: keiner von k Personen hat am gleichen Tag wie der Moderator Geburtstag.

Es gilt

$$P(\overline{B_k}) = \frac{364^k}{365^k}$$

Es gilt

$$P(B_{75}) = 1 - \frac{364^{75}}{365^{75}} = 0.1859728$$

Für $k = 252$ gilt

$$P(B_{252}) = 1 - \frac{364^{252}}{365^{252}} = 0.499$$

und für $k = 253$ gilt

$$P(B_{253}) = 1 - \frac{364^{253}}{365^{253}} = 0.5005.$$

Erst ab 253 Personen im Raum lohnt es sich zu wetten, daß mindestens einer am gleichen Tag Geburtstag hat wie der Moderator.

Nun soll die Reihenfolge keine Rolle spielen.

Beginnen wir hier mit dem Ziehen ohne Zurücklegen.

Beispiel 2.2.16 Wir betrachten wieder

$$U = \{1, 2, 3, 4\}$$

Wir suchen die Anzahl der Möglichkeiten, aus U zwei Zahlen ohne Zurücklegen zu ziehen, wobei uns die Reihenfolge der Ziehung egal ist.

Die Ziehungen $(1, 2)$ und $(2, 1)$ sind für uns also identisch.

Wir schauen uns also alle Möglichkeiten beim Ziehen ohne Zurücklegen ohne Berücksichtigung der Anordnung an und behalten nur die, bei denen die Zahlen der Größe nach geordnet sind.

Die folgende Tabelle zeigt die Vorgehensweise.

| | | |
|----------|----------|------------|
| $(1, 2)$ | $(2, 1)$ | $\{1, 2\}$ |
| $(1, 3)$ | $(3, 1)$ | $\{1, 3\}$ |
| $(1, 4)$ | $(4, 1)$ | $\{1, 4\}$ |
| $(2, 3)$ | $(3, 2)$ | $\{2, 3\}$ |
| $(2, 4)$ | $(4, 2)$ | $\{2, 4\}$ |
| $(3, 4)$ | $(4, 3)$ | $\{3, 4\}$ |

Der folgende Satz verallgemeinert das Beispiel.

Satz 2.2.13 *Eine Menge vom Umfang n hat*

$$\frac{(n)_k}{k!} \tag{2.31}$$

Teilmenge vom Umfang k .

Beweis:

Aus einer Menge vom Umfang n lassen sich $(n)_k$ k -Tupel ohne Wiederholung bilden.

Jedes solche k -Tupel entsteht aus einer Teilmenge vom Umfang k durch Permutation der k Elemente dieser Teilmenge. Da es $k!$ Permutationen einer Menge gibt, gibt es $k!$ -mal so viele r -Tupel wie Teilmengen vom Umfang k .

Sei x die Anzahl der Teilmengen vom Umfang k .

Es gilt also

$$(n)_k = x \cdot k!$$

und somit

$$x = \frac{(n)_k}{k!}$$

Wir setzen

$$\binom{n}{k} = \frac{(n)_k}{k!} \tag{2.32}$$

und lesen dies als 'n über k'.

Man spricht auch vom Binomialkoeffizienten.

Schauen wir uns einige Eigenschaften der Binomialkoeffizienten an.

Satz 2.2.14 *Es gilt*

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Beweis:

$$\binom{n}{k} = \frac{(n)_k}{k!} = \frac{n!}{k!(n-k)!}$$

Satz 2.2.15 *Es gilt*

1.

$$\binom{n}{0} = 1$$

2.

$$\binom{n}{1} = n$$

3.

$$\binom{n}{k} = \binom{n}{n-k}$$

4.

$$\binom{n}{k} = \frac{n}{k} \binom{n-1}{k-1}$$

5.

$$\binom{n}{k} = \frac{n}{n-k} \binom{n-1}{k}$$

6.

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$$

Beweis:

1. *Es gibt eine Teilmenge vom Umfang 0 aus einer Menge M . Dies ist die leere Menge.*
2. *Es gibt n Teilmengen vom Umfang 1 aus einer Menge M vom Umfang n .
Dies sind die einelementigen Mengen.*
3. *Es gibt $\binom{n}{k}$ Möglichkeiten, aus n Elementen k ohne Zurücklegen zu ziehen, wobei die Reihenfolge keine Rolle spielt. Anstatt die gezogenen Elemente zu betrachten, kann man auch die nicht gezogenen betrachten.*

4.

$$\begin{aligned}
 \binom{n}{k} &= \frac{n!}{k!(n-k)!} \\
 &= \frac{n \cdot (n-1)!}{k \cdot (k-1)!(n-k)!} \\
 &= \frac{n}{k} \binom{n-1}{k-1}
 \end{aligned}$$

5.

$$\begin{aligned}
 \binom{n}{k} &= \binom{n}{n-k} \\
 &= \frac{n}{n-k} \binom{n-1}{n-k-1} \\
 &= \frac{n}{n-k} \binom{n-1}{k}
 \end{aligned}$$

6. Die Auswahl von k Elementen aus einer Menge von n Elementen kann so erfolgen:

Wir färben eines der Elemente weiß, die anderen schwarz.

Dann gibt es zwei Sorten von Teilmengen vom Umfang k - solche, die das weiße Element enthalten, und solche, die es nicht enthalten.

Von der ersten Sorte gibt es $\binom{n-1}{k-1}$, da wir $k-1$ aus den $n-1$ schwarzen aus wählen müssen, von der zweiten Sorte gibt es $\binom{n-1}{k}$, da wir k aus den $n-1$ schwarzen aus wählen müssen.

Schauen wir uns einige Beispiele an.

Beispiel 2.2.17 *Beim Lotto werden 6 Kugeln aus 49 Kugeln ohne Zurücklegen gezogen, wobei die Reihenfolge der Ziehung nicht interessiert.*

Es gibt also

$$\binom{49}{6} = \frac{(49)_6}{6!} = \frac{49 \cdot 48 \cdot 47 \cdot 46 \cdot 45 \cdot 44}{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 13983816$$

unterschiedliche Ziehungen.

Beispiel 2.2.18 *Wieviele Folgen aus n Kugeln, von denen k schwarz und $n - k$ weiß sind, kann man bilden?*

Eine Folge ist eindeutig festgelegt, wenn man die Positionen der schwarzen Kugeln kennt.

Die Positionen der schwarzen Kugeln erhält man also, wenn man aus den n Positionen k ohne Zurücklegen auswählt, wobei die Reihenfolge der Ziehung irrelevant ist.

Es gibt also

$$\binom{n}{k}$$

mögliche Folgen.

Für den Fall $n = 4$ und $k = 2$ sind die $\binom{4}{2} = 6$ möglichen Folgen zusammengestellt.

| | |
|--------|------|
| (1, 2) | ssww |
| (1, 3) | swws |
| (1, 4) | swws |
| (2, 3) | wssw |
| (2, 4) | wsws |
| (3, 4) | wsws |

Wir betrachten noch ein sehr wichtiges Beispiel, auf das wir in diesem Skript sehr oft zurückkommen werden.

Beispiel 2.2.19 *Eine Urne enthält N Kugeln, die von 1 bis N durchnummeriert sind. Von den Kugeln sind K schwarz und die restlichen $N - K$ weiß.*

Es werden n Kugeln aus der Urne gezogen.

Wie groß ist die Wahrscheinlichkeit, daß k der gezogenen Kugeln schwarz sind, wenn

1. **mit** Zurücklegen
2. **ohne** Zurücklegen

gezogen wird?

Sei A_k das Ereignis, daß k der gezogenen Kugeln schwarz sind.

Es gilt

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

Wir bestimmen zuerst $|\Omega|$. Es gilt

1. *Da aus den N Kugeln n mit Zurücklegen gezogen werden, gilt*

$$|\omega| = N^n$$

2. *Da aus den N Kugeln n ohne Zurücklegen gezogen werden, gilt*

$$|\omega| = (N)_n$$

Die k schwarzen und $n - k$ weißen Kugeln können auf $\binom{n}{k}$ unterschiedliche Arten angeordnet werden.

Zu jeder dieser $\binom{n}{k}$ unterschiedlichen Positionen gibt es beim Ziehen

1. **mit** Zurücklegen

$$K^k (N - K)^{(n-k)}$$

2. **ohne** Zurücklegen

$$(K)_k (N - K)_{(n-k)}$$

unterscheidbare $n -$ Tupel.

Es gilt also beim Ziehen

1. **mit** Zurücklegen

$$\begin{aligned} P(A_k) &= \frac{\binom{n}{k} K^k (N - K)^{n-k}}{N^n} \\ &= \binom{n}{k} p^k (1 - p)^{n-k} \end{aligned}$$

mit

$$p = \frac{K}{N}$$

2. **ohne** Zurücklegen

$$\begin{aligned} P(A_k) &= \binom{n}{k} \frac{K_k (N - K)_{n-k}}{N_n} \\ &= \frac{n!}{k!(n-k)!} \frac{K_k (N - K)_{n-k}}{N_n} \\ &= \frac{\frac{(K)_k}{k!} \frac{(N - K)_{n-k}}{(n-k)!}}{\frac{(N)_n}{n!}} \\ &= \frac{\binom{K}{k} \binom{N - K}{n - k}}{\binom{N}{n}} \end{aligned}$$

2.2.6 Bedingte Wahrscheinlichkeit

Vielfach ist bei Zufallsvorgängen bekannt, daß ein Ereignis B eingetreten ist. Gesucht ist dann die Wahrscheinlichkeit, daß auch das Ereignis A eintritt. So weiß man, daß eine Person weiblich ist, und sucht die Wahrscheinlichkeit, daß sie eine bestimmte Partei wählt.

Wie kann man die Wahrscheinlichkeit eines Ereignisses A bestimmen, wenn man weiß, daß das Ereignis B eingetreten ist?

Schauen wir uns ein motivierendes Beispiel an.

Beispiel 2.2.20 *Ein fairer Würfel wird einmal geworfen.*

Die Ergebnismenge ist also

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

Als σ -Algebra wählen wir die Potenzmenge $\mathcal{P}(\Omega)$.

Als Wahrscheinlichkeitsmaß ergibt sich

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

für jedes $A \in \mathcal{P}(\Omega)$.

Sei nun speziell $A = \{1, 2, 3\}$, dh. die gewürfelte Augenzahl beträgt höchstens 3.

Es gilt

$$P(A) = \frac{|\{1, 2, 3\}|}{|\{1, 2, 3, 4, 5, 6\}|} = \frac{3}{6}.$$

Nun seien die Seiten des Würfels, auf denen die geraden Augenzahlen stehen, rot gefärbt.

Sei B das Ereignis, daß eine rote Seite obenliegt, also $B = \{2, 4, 6\}$.

Es gilt

$$P(B) = \frac{|\{2, 4, 6\}|}{|\{1, 2, 3, 4, 5, 6\}|} = \frac{3}{6}.$$

Der Würfel werde einmal geworfen. Man kann erkennen, daß eine rote Seite obenliegt, die Augenzahl ist jedoch nicht zu erkennen.

Wie groß ist die Wahrscheinlichkeit, daß die Augenzahl höchstens 3 beträgt?

Wir wissen also, daß das Ereignis B eingetreten ist und suchen die Wahrscheinlichkeit, daß das Ereignis A eintritt.

Man spricht von der bedingten Wahrscheinlichkeit von A unter der Bedingung, daß das Ereignis B eingetreten ist, und schreibt hierfür

$$P(A|B).$$

Wie kann man diese Wahrscheinlichkeit in unserem konkreten Fall bestimmen?

Dadurch, daß man weiß, daß das Ereignis B eingetreten ist, ist die Menge der möglichen Ergebnisse nicht mehr Ω , sondern B .

Die Wahrscheinlichkeit, daß nun A eintritt, ergibt sich aufgrund des Gleichmöglichkeitsmodells zu

$$P(A|B) = \frac{|A \cap B|}{|B|} = \frac{|\{2\}|}{|\{2, 4, 6\}|} = \frac{1}{3}$$

Wegen

$$\begin{aligned} \frac{|A \cap B|}{|B|} &= \frac{\frac{|A \cap B|}{|\Omega|}}{\frac{|B|}{|\Omega|}} \\ &= \frac{P(A \cap B)}{P(B)} \end{aligned}$$

hätten wir die bedingte Wahrscheinlichkeit auch über $P(A \cap B)$ und $P(B)$ bestimmen können.

Das Beispiel legt folgende Definition nahe.

Definition 2.2.2 Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum.

Seien $A, B \in \mathcal{A}$.

Die bedingte Wahrscheinlichkeit von A gegeben B ist für $P(B) > 0$ definiert durch

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \quad (2.33)$$

Ist $P(B) = 0$, so ist $P(A|B)$ nicht definiert.

Der folgende Satz zeigt, daß $P(A|B)$ ein Wahrscheinlichkeitsmaß über (Ω, \mathcal{A}, P) ist.

Satz 2.2.16 $P(A|B)$ ist ein Wahrscheinlichkeitsmaß über (Ω, \mathcal{A}, P) .
Es gilt also

1.

$$P(\Omega|B) = 1$$

2. Für jedes $A \in \mathcal{A}$.

$$0 \leq P(A|B)$$

3. Für $A_i, i = 1, 2, \dots$ mit $A_i \cap A_j = \emptyset$ für $i \neq j$ gilt

$$P\left(\bigcup_{i=1}^{\infty} A_i \mid B\right) = \sum_{i=1}^{\infty} P(A_i|B)$$

Beweis:

1. Wegen $\Omega \cap B = B$ gilt

$$P(\Omega|B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$$

2. Wegen

$$P(A \cap B) \geq 0$$

und

$$P(B) \geq 0$$

folgt

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \geq 0$$

3. Da gilt $A_i \cap A_j = \emptyset$ für $i \neq j$, gilt auch $(A_i \cap B) \cap (A_j \cap B) = \emptyset$ für $i \neq j$.

Also folgt

$$\begin{aligned} P\left(\bigcup_{i=1}^{\infty} A_i \mid B\right) &= \frac{P\left(\left(\bigcup_{i=1}^{\infty} A_i\right) \cap B\right)}{P(B)} \\ &= \frac{P\left(\bigcup_{i=1}^{\infty} (A_i \cap B)\right)}{P(B)} \\ &= \frac{\sum_{i=1}^{\infty} P(A_i \cap B)}{P(B)} \\ &= \sum_{i=1}^{\infty} \frac{P(A_i \cap B)}{P(B)} \\ &= \sum_{i=1}^{\infty} P(A_i \mid B) \end{aligned}$$

2.2.7 Multiplikationssätze

Bei vielen Anwendungen ist die Wahrscheinlichkeit $P(A \cap B)$ gesucht, daß die Ereignisse A und B gleichzeitig eintreten.

Sind $P(A|B)$ und $P(B)$ bekannt, so erhält man aus

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

durch Multiplikation mit $P(B)$ die Wahrscheinlichkeit:

$$P(A \cap B) = P(A|B) P(B) \quad (2.34)$$

Offensichtlich gilt auch

$$P(A \cap B) = P(B|A) P(A) \quad (2.35)$$

Beispiel 2.2.21 *In einer Kiste sind 10 Glühbirnen, von denen 4 defekt sind. Zwei Glühbirnen werden ohne Zurücklegen gezogen.*

Wie groß ist die Wahrscheinlichkeit, daß beide defekt sind?

Seien

- A_1 : *die erste gezogene Glühbirne ist defekt*
- A_2 : *die zweite gezogene Glühbirne ist defekt*

Dann ist

$$\begin{aligned} P(A_1 \cap A_2) &= P(A_2|A_1) P(A_1) \\ &= \frac{3}{9} \cdot \frac{4}{10} \\ &= \frac{2}{15} \end{aligned}$$

Dieses Ergebnis hätten wir auch mit Hilfe der Kombinatorik bestimmen können.

Die gesuchte Wahrscheinlichkeit ist

$$\frac{\binom{4}{2} \binom{6}{0}}{\binom{10}{2}} = \frac{4 \cdot 3 \cdot 2 \cdot 1}{2 \cdot 1 \cdot 10 \cdot 9} = \frac{2}{15}$$

Die Verallgemeinerung auf mehr als zwei Ereignisse liefert der folgende Satz.

Satz 2.2.17 *Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum.*

Seien A_1, \dots, A_n Ereignisse mit $A_i \in \mathcal{A}$ für $i = 1, \dots, n$ und

$$P(A_1 \cap \dots \cap A_{n-1}) > 0.$$

Dann gilt

$$P\left(\bigcap_{i=1}^n A_i\right) = P\left(A_n \mid \bigcap_{i=1}^{n-1} A_i\right) \cdot P\left(A_{n-1} \mid \bigcap_{i=1}^{n-2} A_i\right) \cdot \dots \cdot P(A_1) \quad (2.36)$$

Beweis:

Es gilt

$$(A_1 \cap \dots \cap A_{n-1}) \subset (A_1 \cap \dots \cap A_{n-2}) \subset \dots \subset (A_1 \cap A_2) \subset A_1$$

Wegen

$$P(A_1 \cap \dots \cap A_{n-1}) > 0.$$

und Satz folgt

$$0 < P(A_1 \cap \dots \cap A_{n-1}) \leq P(A_1 \cap \dots \cap A_{n-2}) \leq \dots \leq P(A_1)$$

Also gilt

$$\begin{aligned} P\left(\bigcap_{i=1}^n A_i\right) &= \frac{P\left(\bigcap_{i=1}^n A_i\right)}{P\left(\bigcap_{i=1}^{n-1} A_i\right)} \frac{P\left(\bigcap_{i=1}^{n-1} A_i\right)}{P\left(\bigcap_{i=1}^{n-2} A_i\right)} \dots \frac{P(A_1 \cap A_2)}{P(A_1)} \cdot P(A_1) \\ &= P\left(A_n \mid \bigcap_{i=1}^{n-1} A_i\right) \cdot P\left(A_{n-1} \mid \bigcap_{i=1}^{n-2} A_i\right) \cdot \dots \cdot P(A_2 \mid A_1) \cdot P(A_1) \end{aligned}$$

Beispiel 2.2.22 *In einer Kiste sind 10 Glühbirnen, von denen 4 defekt sind. Vier Glühbirnen werden ohne Zurücklegen gezogen.*

Wie groß ist die Wahrscheinlichkeit, daß alle vier defekt sind?

Seien

- A_1 : *die erste gezogene Glühbirne ist defekt*
- A_2 : *die zweite gezogene Glühbirne ist defekt*
- A_3 : *die dritte gezogene Glühbirne ist defekt*
- A_4 : *die vierte gezogene Glühbirne ist defekt*

Dann ist

$$\begin{aligned}
 P(A_1 \cap A_2 \cap A_3 \cap A_4) &= P(A_4|A_1 \cap A_2 \cap A_3) \cdot P(A_3|A_1 \cap A_2) \\
 &\quad \cdot P(A_2|A_1) \cdot P(A_1) \\
 &= \frac{1}{7} \cdot \frac{2}{8} \cdot \frac{3}{9} \cdot \frac{4}{10} \\
 &= \frac{1}{210}
 \end{aligned}$$

Dieses Ergebnis hätten wir auch mit Hilfe der Kombinatorik bestimmen können.

Die gesuchte Wahrscheinlichkeit ist

$$\frac{\binom{4}{4} \binom{6}{0}}{\binom{10}{4}} = \frac{4 \cdot 3 \cdot 2 \cdot 1}{10 \cdot 9 \cdot 8 \cdot 7} = \frac{1}{210}$$

2.2.8 Satz von der totalen Wahrscheinlichkeit

Definition 2.2.3 Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum.

Eine Folge von Ereignissen $A_1, A_2, \dots \in \mathcal{A}$ heißt vollständiges System von Ereignissen, wenn gilt

$$A_i \cap A_j = \emptyset \quad \text{für } i \neq j$$

und

$$\bigcup_{i=1}^{\infty} A_i = \Omega.$$

Satz 2.2.18 Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $A_1, A_2, \dots \in \mathcal{A}$ ein vollständiges System von Ereignissen mit $P(A_i) > 0$ für $i = 1, 2, \dots$

Dann gilt für jedes $B \in \mathcal{A}$

$$P(B) = \sum_{I=1}^{\infty} P(B|A_i)P(A_i).$$

Beweis

Es gilt

$$\Omega = \bigcup_{i=1}^{\infty} A_i$$

und damit wegen (2.10)

$$B = B \cap \Omega = B \cap \left(\bigcup_{i=1}^{\infty} A_i \right) = \bigcup_{i=1}^{\infty} (B \cap A_i)$$

Wegen

$$A_i \cap A_j = \emptyset \quad \text{für } i \neq j$$

gilt auch

$$(B \cap A_i) \cap (B \cap A_j) = \emptyset \quad \text{für } i \neq j.$$

Also folgt

$$\begin{aligned} P(B) &= P\left(\bigcup_{i=1}^{\infty} (B \cap A_i)\right) \\ &= \sum_{i=1}^{\infty} P(B \cap A_i) \\ &= \sum_{i=1}^{\infty} P(B|A_i) P(A_i) \end{aligned}$$

Man spricht vom Satz von der totalen Wahrscheinlichkeit.

Beispiel 2.2.23 5 Urnen sind von 1 bis 5 durchnummeriert.

Jede Urne enthält 10 Kugeln, wobei in der i -ten Urne i Kugeln weiß sind.

Wir betrachten folgendes Zufallsexperiment.

Wir wählen zuerst eine Urne zufällig aus und ziehen dann eine Kugel aus dieser Urne.

Wie groß ist die Wahrscheinlichkeit, daß die gezogene Kugel weiß ist?

Seien

- B : gezogene Kugel ist weiß
- A_i : gezogene Kugel stammt aus Urne i , $i = 1, 2, 3, 4, 5$.

Es gilt

$$\begin{aligned}P(B|A_i) &= \frac{i}{10} \\P(A_i) &= \frac{1}{5}\end{aligned}$$

Aufgrund des Satzes von der totalen Wahrscheinlichkeit folgt:

$$\begin{aligned}P(B) &= \sum_{i=1}^5 P(B|A_i) P(A_i) \\&= \sum_{i=1}^5 \frac{i}{10} \cdot \frac{1}{5} \\&= \frac{1}{50} \sum_{i=1}^5 i \\&= \frac{1}{50} \cdot \frac{5 \cdot 6}{2} \\&= 0.3\end{aligned}$$

Beispiel 2.2.24 Eine Urne enthält N Kugeln, von denen K weiß und die restlichen $N - K$ schwarz sind.

Es werden zwei Kugeln nacheinander gezogen.

Wie groß ist die Wahrscheinlichkeit, daß die zweite gezogene Kugel weiß ist?
seien

- W_1 : die erste gezogene Kugel ist weiß
- W_2 : die zweite gezogene Kugel ist weiß

Wir betrachten zunächst das Ziehen **mit** Zurücklegen.

Es gilt

$$\begin{aligned}
 P(W_2) &= P(W_2|W_1)P(W_1) + P(W_2|\overline{W_1})P(\overline{W_1}) \\
 &= \frac{K}{N} \cdot \frac{K}{N} + \frac{K}{N} \cdot \frac{N-K}{N} \\
 &= \frac{K}{N} \left(\frac{K}{N} + \frac{N-K}{N} \right) \\
 &= \frac{K}{N}
 \end{aligned}$$

Und nun zum Ziehen **ohne** Zurücklegen.

Es gilt

$$\begin{aligned}
 P(W_2) &= P(W_2|W_1)P(W_1) + P(W_2|\overline{W_1})P(\overline{W_1}) \\
 &= \frac{K-1}{N} \cdot \frac{K}{N} + \frac{K}{N-1} \cdot \frac{N-K}{N} \\
 &= \frac{1}{N(N-1)} ((K-1)K + K(N-K)) \\
 &= \frac{1}{N(N-1)} (K^2 - K + KN - K^2) \\
 &= \frac{1}{N(N-1)} K(N-1) \\
 &= \frac{K}{N}
 \end{aligned}$$

Wir sehen, daß beim Ziehen mit Zurücklegen und beim Ziehen ohne Zurücklegen die unbedingte Wahrscheinlichkeit für eine weiße Kugel identisch ist, während die bedingten Wahrscheinlichkeiten sich unterscheiden.

Beispiel 2.2.25 Bei vielen Fragen kann man sich nicht sicher sein, daß sie wahrheitsgemäß beantwortet werden. So wird nicht jeder zugeben, daß er Drogen genommen hat oder regelmäßig Alkohol trinkt.

Von Warner wurde ein Verfahren vorgeschlagen, daß es erlaubt die Wahrscheinlichkeit einer positiven Antwort zu bestimmen.

Dieses Verfahren ist zweistufig.

Wir schauen es uns in der von Hutchinson vorgeschlagenen Form an.

Auf der ersten Stufe führt der Befragte ein Zufallsexperiment durch, dessen Ergebnis er nur selber kennt.

Wir lassen den Befragten zweimal eine Münze werfen.

Fällt beim ersten Wurf Kopf, so soll er die **Frage 1**, ansonsten die **Frage 2** beantworten:

Frage 1: *Trinken Sie regelmäßig Alkohol?*

Frage 2: *Erschien beim zweiten Münzwurf Kopf?*

Wir definieren die folgenden Ereignisse:

$F1$ die Frage 1 wird beantwortet

J die Antwort ist 'ja'

Wir wissen

$$\begin{aligned} P(F1) &= 0.5 \\ P(\overline{F1}) &= 0.5 \\ P(J|\overline{F1}) &= 0.5 \end{aligned}$$

Es gilt

$$\begin{aligned} P(J) &= P(J|F1) \cdot P(F1) + P(J|\overline{F1}) \cdot P(\overline{F1}) \\ &= P(J|F1) \cdot 0.5 + 0.5 \cdot 0.5 \\ &= P(J|F1) \cdot 0.5 + 0.25 \end{aligned}$$

Ist $P(J)$ bekannt, so können wir $P(J|F1)$ bestimmen durch

$$P(J|F1) = \frac{P(J) - 0.25}{0.5} = 2 \cdot P(J) - 0.5$$

2.2.9 Satz von Bayes

Es sei bekannt, daß ein Promille der Bevölkerung an einer bestimmten Krankheit leide.

Bei der Diagnose wird ein Test verwendet, der in 90 Prozent der Fälle einen Kranken als krank einstuft. Man sagt, daß der Test positiv ist. Weiterhin stuft der Test in 99 Prozen der Fälle einen Gesunden als gesund ein.

Ein Patient wird getestet. Das Ergebnis ist positiv. Wie groß ist die Wahrscheinlichkeit, daß der Patient an der Krankheit leidet?

Wir kennen $P(A|B)$ und suchen $P(B|A)$.

Der folgende Satz gibt die Lösung an.

Satz 2.2.19 Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $A_1, A_2, \dots \in \mathcal{A}$ ein vollständiges System von Ereignissen mit $P(A_i) > 0$ für $i = 1, 2, \dots$

Dann gilt für jedes $B \in \mathcal{A}$

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{I=1}^{\infty} P(B|A_i)P(A_i)}.$$

Beweis

Es gilt

$$P(A_i|B)P(B) = P(B|A_i)P(A_i)$$

und damit

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)}$$

Da A_1, A_2, \dots ein vollständiges System von Ereignissen bildet, gilt aufgrund von Satz

$$P(B) = \sum_{i=1}^{\infty} P(B|A_i)P(A_i)$$

und damit

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{I=1}^{\infty} P(B|A_i)P(A_i)}.$$

Beispiel 2.2.26 *Wir setzen das Eingangsbeispiel fort.*

Seien

- K eine Person leidet an der Krankheit
- PO das Testergebnis ist positiv

Wir wissen

$$\begin{aligned}P(K) &= 0.001 & P(\bar{K}) &= 0.999 \\P(PO|K) &= 0.9 & P(PO|\bar{K}) &= 0.01\end{aligned}$$

Gesucht ist $P(K|PO)$.

Es gilt

$$\begin{aligned}P(K|PO) &= \frac{P(PO|K)P(K)}{P(PO|K)P(K) + P(PO|\bar{K})P(\bar{K})} \\&= \frac{0.9 \cdot 0.001}{0.9 \cdot 0.001 + 0.01 \cdot 0.999} \\&= 0.08264463\end{aligned}$$

Diese Wahrscheinlichkeit ist sehr gering.

Beispiel 2.2.27 5 Urnen sind von 1 bis 5 durchnummeriert.

Jede Urne enthält 10 Kugeln, wobei in der i -ten Urne i Kugeln weiß sind.

Wir betrachten folgendes Zufallsexperiment.

Wir wählen zuerst eine Urne zufällig aus und ziehen dann eine Kugel aus dieser Urne.

Die gezogene Kugel ist weiß .

Wie groß ist die Wahrscheinlichkeit, daß sie aus der 5. Urne kommt?

Seien

- B : gezogene Kugel ist weiß
- A_i : gezogene Kugel stammt aus Urne i , $i = 1, 2, 3, 4, 5$.

Es gilt

$$\begin{aligned}P(B|A_i) &= \frac{i}{10} \\P(A_i) &= \frac{1}{5}\end{aligned}$$

Wir haben gesehen:

$$P(B) = 0.3$$

Also gilt

$$\begin{aligned}P(A_5|B) &= \frac{P(B|A_5)P(A_5)}{P(B)} \\&= \frac{0.5 \cdot 0.2}{0.3} \\&= \frac{1}{3}\end{aligned}$$

2.2.10 Unabhängigkeit

Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum.

Wir betrachten zwei Ereignisse $A, B \in \mathcal{A}$.

Welcher Zusammenhang besteht zwischen $P(A|B)$ und $P(A)$?

Schauen wir uns ein Beispiel an.

Beispiel 2.2.28 *Wir werfen einen fairen Würfel und betrachten das Ereignis*

$$B = \{2, 4, 6\}$$

Nun betrachten wir 3 Fälle.

Fall 1

$$A = \{1, 2, 3\}$$

Es gilt

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1}{3}$$

Wegen $P(A) = 0.5$ gilt also

$$P(A) > P(A|B)$$

Fall 2

$$A = \{2, 3, 4\}$$

Es gilt

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{2}{3}$$

Wegen $P(A) = 0.5$ gilt also

$$P(A) < P(A|B)$$

Fall 3

$$A = \{1, 2\}$$

Es gilt

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1}{3}$$

Wegen $P(A) = \frac{1}{3}$ gilt also

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

Zwischen $P(A)$ und $P(A|B)$ sind also alle möglichen Beziehungen möglich.

Was bedeutet $P(A) = P(A|B)$?

Schauen wir uns ein Beispiel an.

Beispiel 2.2.29 *Eine Population besteht aus 1000 Personen.*

Von diesen sind 600 männlich.

Unter den Männern sind 150 Raucher und unter den Frauen sind 100 Raucher.

Eine Person wird zufällig ausgewählt.

Sei

- *R: die Person ist Raucher*
- *M: die Person ist ein Mann*

Es gilt

$$P(R) = 0.25$$

und

$$P(R|M) = \frac{P(R \cap M)}{P(M)} = \frac{0.15}{0.6} = 0.25$$

Das Wissen, daß eine Person männlich ist, ändert nichts an der Wahrscheinlichkeit, Raucher zu sein.

Wir sagen, daß im Beispiel Rauchen und Geschlecht unabhängig sind.

Definition 2.2.4 *Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $A, B \in \mathcal{A}$ mit $P(B) > 0$.*

Dann heißen die Ereignisse A und B unabhängig, wenn gilt

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

Satz 2.2.20 *Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $A, B \in \mathcal{A}$ mit $P(B) > 0$.*

Sind die Ereignisse A und B unabhängig, so gilt

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

Beweis:

Aus

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

folgt

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

Sind die Ereignisse A und B unabhängig, so benötigt man nur die Wahrscheinlichkeiten $P(A)$ und $P(B)$, um die Wahrscheinlichkeit zu bestimmen, daß A und B gleichzeitig eintreten.

Satz 2.2.21 Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $A, B \in \mathcal{A}$ mit $P(B) > 0$.

Sind die Ereignisse A und B unabhängig, so sind auch die folgenden Paare von Ereignissen unabhängig:

1. A und \bar{B}
2. \bar{A} und B
3. \bar{A} und \bar{B}

Beweis:

Wir zeigen nur 1.. Die anderen Beweise verlaufen analog.

Es ist zu zeigen

$$P(A \cap \bar{B}) = P(A)P(\bar{B})$$

Es gilt

$$\begin{aligned} P(A \cap \bar{B}) &= P(A) - P(A \cap B) \\ &= P(A) - P(A)P(B) \\ &= P(A)(1 - P(B)) \\ &= P(A)P(\bar{B}) \end{aligned}$$

Der folgende Satz zeigt, wie die Begriffe Disjunktheit und Unabhängigkeit zusammenhängen.

Satz 2.2.22 Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $A, B \in \mathcal{A}$.

Sind A und B disjunkt und gilt $P(A) > 0$ und $P(B) > 0$, so sind A und B nicht unabhängig.

Beweis:

Aus

$$A \cap B = \emptyset$$

folgt

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = 0$$

Da gilt $P(A) > 0$ folgt

$$P(A|B) \neq P(A)$$

Die Aussage des Satzes ist auch intuitiv klar:

Sind die Ereignisse disjunkt, so können sie nicht gleichzeitig eintreten, und somit im höchsten Maße abhängig. Tritt nämlich A ein, so tritt B nicht ein und umgekehrt.

Wir schauen uns noch ein Beispiel an, das unabhängige Ereignisse verwendet.

Beispiel 2.2.30 *Oft werden Entscheidungen durch einen Münzwurf getroffen. Hierbei wird unterstellt, daß die verwendete Münze fair ist. Dies muß aber nicht der Fall sein.*

John von Neumann hat ein Verfahren vorgeschlagen, bei dem man sich mit gleicher Wahrscheinlichkeit für eine der beiden Alternativen entscheidet.

Die Münze wird zweimal hintereinander geworfen. Fällt beim ersten Mal Kopf und beim zweiten Mal Zahl, so entscheidet man sich für Alternative 1. Fällt beim ersten Mal Zahl und beim zweiten Mal Kopf, so entscheidet man sich für Alternative 2. Fällt bei beiden Würfeln das gleiche Symbol, so wird die Münze wiederum zweimal geworfen, und es wird genauso verfahren wie bei der ersten Runde. Die ganze Prozedur wird solange durchgeführt, bis zum ersten Mal zwei unterschiedliche Symbole auftreten.

Inwiefern ist diese Prozedur fair?

Sei K_i das Ereignis, daß beim i -ten Wurf Kopf fällt, $i=1,2$.

Wir unterstellen, daß die Wahrscheinlichkeit für Kopf bei beiden Würfeln gleich ist, und daß die Münze kein Gedächtnis hat.

Sei $P(K_i) = p$ die Wahrscheinlichkeit für Kopf beim i -ten Wurf.

Dann gilt wegen der Unabhängigkeit

$$P(K_1 \cap \overline{K_2}) = P(K_1) \cdot P(\overline{K_2}) = p \cdot (1 - p)$$

und

$$P(\overline{K_1} \cap K_2) = P(\overline{K_1}) \cdot P(K_2) = (1 - p) \cdot p.$$

Wir sehen, daß die Wahrscheinlichkeiten der beiden Alternativen bei jeder Runde gleich sind.

Ein Problem hat die Prozedur jedoch:

Es kann sehr lange dauern, bis eine Entscheidung fällt.

Kapitel 3

Univariate Verteilungen

3.1 Zufallsvariablen und ihre Verteilungen

Bei einem Zufallsvorgang interessiert man sich in der Regel nicht für die Ergebnisse, sondern für Zahlenwerte, die diesen Ergebnissen zugeordnet sind. Schauen wir uns ein Beispiel an:

Ein Teilchen bewegt sich auf den ganzen Zahlen, wobei es im Nullpunkt startet. Bei jedem Schritt geht es zufällig nach rechts oder links. Das Teilchen möge zwei Schritte machen. Uns interessiert die Anzahl der Schritte nach links. Offensichtlich kann es keinen, einen oder zwei Schritte nach links machen.

Wie groß ist die Wahrscheinlichkeit jedes dieser Werte?

Um diese Frage zu beantworten, schauen wir uns das Beispiel zunächst unter dem Aspekt der Wahrscheinlichkeitsrechnung an. Wir konstruieren also einen geeigneten Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) .

Die Ergebnismenge ist

$$\Omega = \{LL, LR, RL, RR\}$$

Als σ -Algebra \mathcal{A} von Ereignissen wählen wir die Potenzmenge $\mathcal{P}(\Omega)$.

Es gilt

$$\begin{aligned} \mathcal{P}(\Omega) = & \{\emptyset, \{LL\}, \{LR\}, \{RL\}, \{RR\}, \\ & \{LL, LR\}, \{LL, RL\}, \{LL, RR\}, \\ & \{LR, RL\}, \{LR, RR\}, \{RL, RR\} \\ & \{LL, LR, RL\}, \{LL, LR, RR\}, \{LL, RL, RR\}, \{LR, RL, RR\}, \\ & \{LL, LR, RL, RR\}\} \end{aligned}$$

Der letzte Schritt besteht in der Festlegung eines Wahrscheinlichkeitsmaßes

$$P : \mathcal{A} \rightarrow [0, 1].$$

Da das Teilchen sich zufällig bewegt, ist jedes der Elementarereignisse

$$\{LL\} \quad \{LR\} \quad \{RL\} \quad \{RR\}$$

gleichwahrscheinlich.

Aus dem dritten Axiom der Wahrscheinlichkeitsrechnung folgt dann für jedes $A \in \mathcal{P}$:

$$P(A) = \frac{\text{Anzahl der Elemente in } A}{\text{Anzahl der Elemente in } \Omega}$$

So gilt zum Beispiel für das Teilchen

$$\begin{aligned} P(\{RR\}) &= 0.25 \\ P(\{RR, RL, LR\}) &= 0.75 \\ P(\{RR, RL, LR, LL\}) &= 1 \end{aligned}$$

Nun haben wir einen geeigneten Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) für unser Beispiel gefunden und können uns nun der ursprünglichen Fragestellung zuwenden.

Uns interessiert die Anzahl der Schritte nach links.

Die Abbildung $X : \Omega \rightarrow \mathfrak{R}$ ordne nun jedem $\omega \in \Omega$ die Anzahl der L zu.

Es gilt

$$\begin{aligned} X(RR) &= 0 \\ X(RL) &= 1 \\ X(LR) &= 1 \\ X(LL) &= 2 \end{aligned}$$

Das Teilchen kann also keinen, einen oder zwei Schritte nach links machen. Mit welcher Wahrscheinlichkeit macht es keinen, mit welcher einen und mit welcher zwei Schritte nach links?

Neben diesen Fragen könnte auch die Wahrscheinlichkeit von Interesse sein, höchstens einen Schritt nach links zu machen oder auch die Wahrscheinlichkeit, mindestens einen Schritt nach links zu machen.

Die folgende Definition erlaubt die Bestimmung all dieser Wahrscheinlichkeiten.

Definition 3.1.1 Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum.

Eine Abbildung $X : \Omega \rightarrow \mathfrak{R}$ heißt Zufallsvariable X , falls für jedes $x \in \mathfrak{R}$ gilt

$$\{\omega | X(\omega) \leq x\} \in \mathcal{A}$$

Beispiel 3.1.1 Bei dem obigen Beispiel handelt es sich um eine Zufallsvariable, denn es gilt

$$\{\omega | X(\omega) \leq x\} = \begin{cases} \emptyset & \text{für } x < 0 \\ \{RR\} & \text{für } 0 \leq x < 1 \\ \{RR, RL, LR\} & \text{für } 1 \leq x < 2 \\ \{RR, RL, LR, LL\} & \text{für } 2 \leq x \end{cases}$$

Die Definition der Zufallsvariablen stellt sicher, daß allen Ereignissen der Form $\{\omega | X(\omega) \leq x\}$ eine Wahrscheinlichkeit $P(\{\omega | X(\omega) \leq x\})$ zugeordnet wird, da wir jedem $A \in \mathcal{A}$ eine Wahrscheinlichkeit zuordnen.

Für

$$P(\{\omega | X(\omega) \leq x\})$$

schreiben wir im folgenden kurz

$$P(X \leq x).$$

Definition 3.1.2 Sei X eine Zufallsvariable.

Dann heißt

$$F_X(x) = P(X \leq x) \tag{3.1}$$

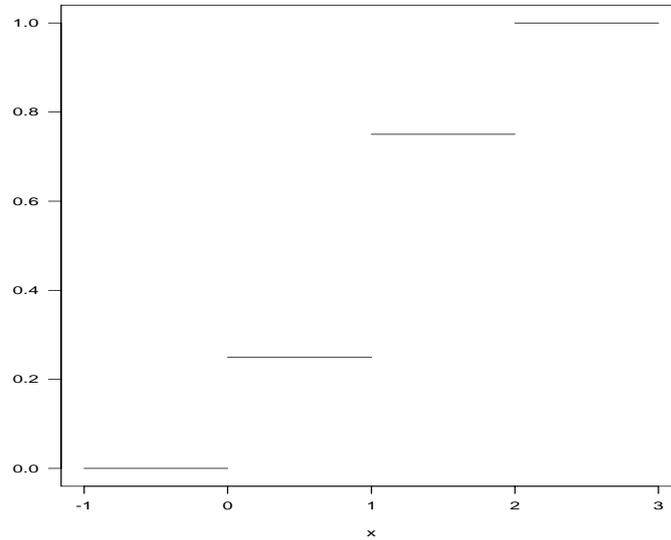
die Verteilungsfunktion von X .

Beispiel 3.1.2 Wir haben gesehen, daß die Anzahl der Schritte nach links, die das Teilchen bei zwei Schritten macht, eine Zufallsvariable X ist.

Die Verteilungsfunktion $F_X(x)$ ist gegeben durch

$$F_X(x) = \begin{cases} 0 & \text{für } x < 0 \\ 0.25 & \text{für } 0 \leq x < 1 \\ 0.75 & \text{für } 1 \leq x < 2 \\ 1 & \text{für } 2 \leq x \end{cases}$$

Die folgende Graphik zeigt die empirische Verteilungsfunktion.



Wir sehen, daß sie eine Treppenfunktion ist, die monoton wächst.

Die Verteilungsfunktion $F_X(x)$ einer Zufallsvariablen X besitzt folgende Eigenschaften, die wir ohne Beweis angeben.

Satz 3.1.1 Die Verteilungsfunktion $F_X(x)$ einer Zufallsvariablen X besitzt folgende Eigenschaften.

1. $F_X(x)$ ist monoton wachsend.
2. $F_X(x)$ ist rechtsseitig stetig.
3. $\lim_{x \rightarrow -\infty} F_X(x) = 0$
4. $\lim_{x \rightarrow \infty} F_X(x) = 1$

Mit Hilfe der Verteilungsfunktion kann man unter anderem folgende Wahrscheinlichkeiten bestimmen.

1.

$$P(X = a) = F_X(a) - \lim_{x \uparrow a} F_X(x) \quad (3.2)$$

2.

$$P(X \leq a) = F_X(a) \quad (3.3)$$

3.

$$P(X < a) = F_X(a) - P(X = a) \quad (3.4)$$

4.

$$P(X > a) = 1 - F_X(a) \quad (3.5)$$

5.

$$P(a < X \leq b) = F_X(b) - F_X(a) \quad (3.6)$$

6.

$$P(a \leq X \leq b) = F_X(b) - F_X(a) + P(X = a) \quad (3.7)$$

7.

$$P(a \leq X < b) = F_X(b) - F_X(a) - P(X = b) + P(X = a) \quad (3.8)$$

8.

$$P(a < X < b) = F_X(b) - F_X(a) - P(X = b) \quad (3.9)$$

Nach der Anzahl der Realisierungsmöglichkeiten unterscheiden wir diskrete und stetige Zufallsvariablen.

Definition 3.1.3 Eine Zufallsvariable X , die höchstens abzählbar viele Werte annehmen kann, heißt diskret.

Dabei heißt $P(X = x)$ die Wahrscheinlichkeitsfunktion von X .

$\mathcal{T}_X = \{x | P(X = x) > 0\}$ heißt der Träger von x .

Es muß gelten

$$\sum_x P(X = x) = 1 \quad (3.10)$$

und

$$P(X = x) \geq 0 \quad (3.11)$$

für alle $x \in \mathfrak{R}$.

Die Anzahl der Schritte des Teilchens nach links ist offensichtlich eine diskrete Zufallsvariable.

Der Träger ist $\mathcal{T}_X = \{0, 1, 2\}$.

Schauen wir uns jetzt stetige Zufallsvariablen an.

Definition 3.1.4 Eine Zufallsvariable X heißt stetig, wenn eine Funktion $f_X : \mathfrak{R} \rightarrow \mathfrak{R}$ existiert, so daß für die Verteilungsfunktion $F_X(x)$ von X gilt:

$$F_X(x) = \int_{-\infty}^x f_X(u) du \quad (3.12)$$

Die Funktion $f_X(x)$ heißt Dichtefunktion der Zufallsvariablen X .

Die Dichtefunktion $f_X(x)$ erfüllt folgende Bedingungen:

1.

$$f_X(x) \geq 0 \quad (3.13)$$

für alle $x \in \mathfrak{R}$

2.

$$\int_{-\infty}^{\infty} f_X(u) du = 1 \quad (3.14)$$

Man kann zeigen, daß jede Funktion, die diese Bedingungen erfüllt, als Dichtefunktion einer stetigen Zufallsvariablen aufgefaßt werden kann.

Beispiel 3.1.3 Gegeben sei folgende Funktion $f : \mathfrak{R} \rightarrow \mathfrak{R}$ mit

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

Offensichtlich gilt

$$f(x) \geq 0 \quad \text{für alle } x \in \mathfrak{R}$$

Außerdem gilt

$$\int_{-\infty}^{\infty} f(u) du = \int_0^1 1 du = [u]_0^1 = 1$$

Es handelt sich also um die Dichtefunktion einer Zufallsvariablen. Dies ist die Dichtefunktion einer auf $(0, 1)$ gleichverteilten Zufallsvariablen.

Die Verteilungsfunktion $F_X(x)$ einer stetigen Zufallsvariablen X ist eine stetige Funktion.

Hieraus folgt, daß für eine stetige Zufallsvariable X für alle $x \in \mathfrak{R}$ gilt

$$P(X = x) = 0,$$

denn

$$\begin{aligned} P(X = x) &= F_X(x) - \lim_{u \uparrow x} F_X(u) \\ &= F_X(x) - F_X(x) \\ &= 0 \end{aligned}$$

3.2 Diskrete Verteilungsmodelle

Eine Zufallsvariable X , die höchstens abzählbar viele Werte annehmen kann, heißt diskret.

Dabei heißt $P(X = x)$ die Wahrscheinlichkeitsfunktion von X .

Man kann die Wahrscheinlichkeitsfunktion einer diskreten Zufallsvariablen dadurch gewinnen, daß man den Träger \mathcal{T}_X vorgibt und für jedes $x \in \mathcal{T}_X$ die Wahrscheinlichkeit $P(X = x)$ festlegt. Dabei müssen natürlich (3.10) und (3.11) erfüllt sein.

Ein Beispiel hierfür ist die diskrete **Gleichverteilung**.

Definition 3.2.1 Die Zufallsvariable X heißt gleichverteilt, wenn gilt ,

$$P(X = x) = \frac{1}{N} \quad (3.15)$$

für $x = 1, \dots, N$.

Die Gleichverteilung ist ein sinnvolles Modell für die Augenzahl beim einmaligen Wurf eines fairen Würfels.

Hier gilt

$$P(X = x) = \frac{1}{6}$$

für $x = 1, 2, \dots, 6$.

Es gibt eine Reihe von Möglichkeiten, diskrete Verteilungsmodelle zu gewinnen.

Eine Möglichkeit besteht darin, auf einem geeigneten Wahrscheinlichkeitsraum eine Zufallsvariable zu definieren.

Hier spielt sogenannte **Bernoulli-prozeß** eine zentrale Rolle.

Definition 3.2.2 Wir gehen aus von einem Zufallsvorgang, bei dem ein Ereignis A mit $P(A) = p$ von Interesse ist.

Man spricht auch von einem **Bernoullivorgang** und nennt das Ereignis A oft auch **Erfolg**.

Einen **Bernoulli-prozeß** erhält man nun dadurch, daß man einen Bernoullivorgang mehrmals beobachtet, wobei folgende Annahmen getroffen werden:

- Die einzelnen Bernoullivorgänge sind voneinander unabhängig.
- Die Erfolgswahrscheinlichkeit p bleibt konstant.

Beispiel 3.2.1 *Im letzten Abschnitt haben wir ein Teilchen betrachtet, daß sich zufällig auf den ganzen Zahlen bewegt, wobei es im Nullpunkt startet.*

Jeder Schritt des Teilchens ist ein Bernoullixperiment.

Es gibt nur 2 Möglichkeiten. Das Teilchen geht entweder nach links, was wir mit A , oder es geht nach rechts, was wir mit \bar{A} bezeichnen.

Da das Teilchen sich zufällig entscheidet, gilt $P(A) = p = 0.5$.

Wenn wir unterstellen, daß die Entscheidung des Teilchens bei einem Schritt unabhängig von den anderen Schritten fällt, und die Erfolgswahrscheinlichkeit p konstant bleibt, beobachten wir bei n Schritten einen Bernoulliprozeß der Länge n .

Beim Bernoulliprozeß sind 3 Zufallsvariablen von Interesse:

- Anzahl der Erfolge bei n Durchführungen.
- Anzahl der Mißerfolge vor dem ersten Erfolg
- Anzahl der Mißerfolge vor dem r .ten Erfolg

Schauen wir uns die Verteilungen dieser Zufallsvariablen im Detail an.

Definition 3.2.3 *Die Zufallsvariable X heißt binomialverteilt mit den Parametern n und p , wenn ihre Wahrscheinlichkeitsfunktion gegeben ist durch:*

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x} \quad (3.16)$$

für $x = 0, 1, \dots, n$.

Wir schreiben auch $X \sim \text{BINOMIAL}(n, p)$.

Die Binomialverteilung erhält man, wenn man bei einem Bernoulliprozeß der Länge n die Erfolge zählt.

Dies kann man sich folgendermaßen klarmachen:

Sei X die Anzahl der Erfolge bei einem Bernoulliprozeß der Länge n .

Die Zufallsvariable X kann die Werte $0, 1, 2, \dots, n$ annehmen.

Nimmt X den Wert x an, so wurden x Erfolge A und $n - x$ Mißerfolge \bar{A} beobachtet.

Aufgrund der Unabhängigkeit der einzelnen Durchführungen beträgt die Wahrscheinlichkeit einer Folge aus x Erfolgen A und $n - x$ Mißerfolgen \bar{A} gleich

$$p^x (1 - p)^{n-x}.$$

Eine Folge, die x -mal ein A und $n - x$ -mal ein \bar{A} enthält, ist eindeutig durch die Positionen der A festgelegt.

Wir haben gesehen, daß es

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

Möglichkeiten gibt, die Positionen der A zu wählen.

Also gilt

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

Die Binomialverteilung mit $n = 1$ heißt Bernoulliverteilung.

Man schreibt $X \sim \text{BERNOULLI}(p)$.

Die Wahrscheinlichkeitsfunktion der Bernoulliverteilung ist gegeben durch:

$$P(X = x) = p^x (1-p)^{1-x} \quad \text{für } x = 0, 1.$$

Es gilt also

$$P(X = 0) = 1 - p$$

und

$$P(X = 1) = p.$$

Eine bernoulliverteilte Zufallsvariable X zählt, wie oft A bei einem Bernoullivorgang eingetreten ist.

Beispiel 3.2.2 *Das Teilchen aus Beispiel 2.1 macht 3 Schritte.*

Sei X die Anzahl der Schritte nach links.

Die Zufallsvariable X ist binomialverteilt mit den Parametern $n = 3$ und $p = 0.5$.

Es gilt

$$P(X = x) = \binom{3}{x} 0.5^x (1 - 0.5)^{3-x} = \binom{3}{x} 0.5^3 \quad \text{für } x = 0, 1, 2, 3$$

Es gilt speziell

$$P(X = 0) = 0.125$$

$$P(X = 1) = 0.375$$

$$P(X = 2) = 0.375$$

$$P(X = 3) = 0.125$$

Beispiel 3.2.3 *Bei einer Vielzahl statistischer Tests ist die Teststatistik binomialverteilt. Mit Hilfe eines statistischen Tests soll eine Hypothese überprüft werden.*

Schauen wir uns ein Beispiel an, das ursprünglich von Fisher stammt und von Neyman modifiziert wurde.

Eine englische Dame behauptet, daß sie erkennt, ob bei einer Tasse Tee mit Milch zuerst die Milch oder zuerst der Tee eingegossen wurde.

Wir wollen überprüfen, ob die Behauptung der Dame zutrifft. Hierzu formulieren wir das Gegenteil der Behauptung der Dame als Hypothese und die Behauptung der Dame als Gegenhypothese.

Wir erhalten also:

*Hypothese: Die Dame rät
Gegenhypothese: Die Dame rät nicht*

Man bezeichnet die Hypothese mit H_0 und die Gegenhypothese mit H_1 .

Versuchen wir nun auch die Hypothese und die Gegenhypothese in die Sprache der Statistik zu übersetzen. In der Statistik formuliert man Hypothesen in der Regel über Parameter.

Sei R das Ereignis, daß die Dame die richtige Wahl trifft, und \bar{R} das Ereignis, daß sie die falsche Wahl trifft.

Außerdem sei $p = P(R)$. Rät die Dame, so ist $p = 0.5$, rät sie nicht, so gilt $p > 0.5$.

Die Hypothese und Gegenhypothese lauten also

$$\begin{aligned} H_0 &: p = 0.5 \\ H_1 &: p > 0.5 \end{aligned}$$

Um die Hypothesen zu überprüfen, beobachtet man den Sachverhalt. Es liegt nahe, der Dame eine Tasse zu reichen, in die ohne ihr Wissen zuerst die Milch gefüllt wurde. Sie soll dann entscheiden, welche Situation vorliegt. Es ist aber sicherlich fairer, ihr von vornherein die Möglichkeit zu geben, zu vergleichen. Wir reichen ihr also zwei Tassen, wobei in die eine zuerst die Milch und in die andere zuerst der Tee gefüllt wurde. Auf der Basis der Beobachtung fällen wir nun die Entscheidung.

Man nennt die Beobachtung eine Teststatistik.

Folgende Entscheidungsregel liegt nahe:

*Entscheidung für H_0 , wenn die Dame die Tassen falsch zugeordnet hat.
Entscheidung für H_1 , wenn die Dame die Tassen richtig zugeordnet hat.*

Die Entscheidung ist fehlerbehaftet.

Wir können zwei Fehler begehen.

Entscheiden wir uns für H_1 , obwohl H_0 zutrifft, so begehen wir einen Fehler 1. Art.

In unserem Beispiel heißt dies, daß wir zu der Entscheidung kommen, daß die Dame differenzieren kann, obwohl sie in Wirklichkeit geraten hat.

Entscheiden wir uns für H_0 , obwohl H_1 zutrifft, so begehen wir einen Fehler 2. Art.

In unserem Beispiel heißt dies, daß wir zu der Entscheidung kommen, daß die Dame geraten hat, obwohl sie in Wirklichkeit differenzieren kann.

Die Wahrscheinlichkeit des Fehlers 1. Art ist

$$\alpha = P(\text{Entscheidung für } H_1 | H_0 \text{ trifft zu})$$

In unserem Beispiel gilt

$$\alpha = P_{H_0}(R) = 0.5$$

Die Wahrscheinlichkeit des Fehlers 2. Art ist

$$\beta = P(\text{Entscheidung für } H_0 | H_1 \text{ trifft zu})$$

In unserem Beispiel gilt

$$\beta = P_{H_1}(\bar{R})$$

Dies Wahrscheinlichkeit können wir nicht angeben, da sie davon abhängt, wie gut die Dame differenzieren kann.

Wir haben also nur den Fehler 1. Art unter Kontrolle. Dies ist die übliche Situation beim statistischen Test. Da wir die Wahrscheinlichkeit des Fehlers 1. Art unter Kontrolle haben, geben wir diese vor. In der Regel wählt man $\alpha = 0.05$. Hierdurch ist sichergestellt, daß man sich ziemlich sicher sein kann, wenn man sich für H_1 entscheidet, da die Wahrscheinlichkeit für einen Fehler klein ist. Nun ist auch klar, warum wir die Behauptung der Dame als Gegenhypothese formuliert haben.

In unserem Beispiel beträgt die Wahrscheinlichkeit für den Fehler 1. Art 0.5. Sie ist viel zu groß. Wir können sie verkleinern, indem wir das Experiment mehrmals wiederholen. Nehmen wir an, wir führen 6 Versuche durch und erhalten das Ergebnis RRRRR \bar{R} .

Spricht dies für oder gegen die Fähigkeit der Dame zu differenzieren?

Wir brauchen eine geeignete Teststatistik. Es liegt nahe, die Anzahl S der Fälle zu wählen, bei denen die Dame sich richtig entschieden hat.

In unserem Fall beträgt sie 5.

Ist diese sehr groß, so würden wir sagen, daß die Dame differenzieren kann.

Was heißt groß?

Dies hängt von dem vorgegebenen Wert von α ab. Wir müssen den Wert, ab dem wir uns für die Gegenhypothese entscheiden so wählen, daß die Wahrscheinlichkeit für den Fehler 1. Art gleich α ist. Hierzu benötigen wir die Verteilung der Teststatistik, wenn H_0 zutrifft. Man spricht auch von der Verteilung der Teststatistik unter H_0 . Und hier kommt in unserem Beispiel die Binomialverteilung ins Spiel.

Wir haben sechsmal einen Bernoullivorgang beobachtet. Wenn wir der Dame nicht nach jedem Versuch sagen, ob ihre Entscheidung richtig oder falsch ist, lernt sie bei den Versuchen nichts dazu. Also sind die einzelnen Versuche unabhängig. Außerdem bleibt $P(R) = 0.5$ konstant.

Wir beobachten also einen Bernoulliprozeß der Länge 6 mit Erfolgswahrscheinlichkeit $p = 0.5$.

Dann ist die Anzahl der Erfolge mit den Parametern $n = 6$ und $p = 0.5$ binomialverteilt.

Es gilt also

$$P(S = s) = \binom{6}{s} 0.5^6 \quad \text{für } s = 0, 1, \dots, 6.$$

Dies ist die Verteilung von S unter H_0 . Wir haben gesagt, daß wir H_0 ablehnen, wenn S zu groß ist.

Wir bestimmen also für das Beispiel

$$P_{H_0}(S \geq 5) = \left[\binom{6}{5} + \binom{6}{6} \right] 0.5^6 = 0.109375$$

Wir nennen diese Wahrscheinlichkeit Überschreitungswahrscheinlichkeit.

Wir können die Testentscheidung mit Hilfe der Überschreitungswahrscheinlichkeit durchführen.

Da sie größer als $\alpha = 0.05$ ist, lehnen wir H_0 nicht ab.

Wie gut ist die Entscheidungsregel?

Wir messen die Güte der Entscheidungsregel durch die Wahrscheinlichkeit, uns für H_1 zu entscheiden, wenn H_1 zutrifft.

Wir sprechen auch von der Gütefunktion $G(p)$:

$$G(p) = P(\text{Entscheidung für } H_1 | p)$$

Diese hängt vom Wert von p ab, der nicht bekannt ist.

Sinnvollerweise sollte die Güte mit wachsendem p zunehmen, da wir uns ja immer weiter von $p = 0.5$ entfernen.

In unserem Beispiel ist es einfach die Gütefunktion zu bestimmen.

Wir wollen diese für den Fall bestimmen, daß wir zum Niveau $\alpha = 0.015625$ testen.

Wir lehnen H_0 ab, wenn $S = 6$ gilt.

Offensichtlich gilt

$$P_{H_0}(S = 6) = 0.5^6 = 0.015625$$

Die Teststatistik S ist auch unter H_1 binomialverteilt. Nur der Wert von p ändert sich.

Es gilt also

$$P(S = s) = \binom{6}{s} p^s (1-p)^{6-s} \quad \text{für } s = 0, 1, \dots, n.$$

Die Gütefunktion lautet

$$\begin{aligned} G(p) &= P_{H_1}(S = 6) \\ &= p^6 \end{aligned}$$

Die folgende Tabelle zeigt die Gütefunktion für ausgewählte Werte von p .

| p | $G(p)$ |
|------|--------|
| 0.60 | 0.047 |
| 0.70 | 0.118 |
| 0.80 | 0.262 |
| 0.90 | 0.531 |
| 0.95 | 0.735 |

Wir sehen, daß mit wachsendem P die Güte immer größer wird.

Die Entscheidungsregel im Beispiel beruht auf der Anzahl der Fehlversuche unter 6 Versuchen.

Wir hätten auch eine andere Teststatistik wählen können.

Lindley schlägt vor, die Dame so lange probieren zu lassen, bis sie zum ersten Mal falsch klassifiziert.

Wenn wir mit A den Fall bezeichnen, bei dem die Dame falsch klassifiziert, dann beobachten wir also einen Bernoulli-prozeß bis zum ersten Eintreten von A . Man erhält dann die geometrische Verteilung.

Definition 3.2.4 Die Zufallsvariable X heißt geometrisch verteilt mit Parameter p , wenn ihre Wahrscheinlichkeitsfunktion gegeben ist durch

$$P(X = x) = p(1 - p)^x \quad \text{für } x = 0, 1, \dots \quad (3.17)$$

Wir schreiben auch $X \sim \text{GEOM}(p)$

Die geometrische Verteilung erhält man, wenn man bei einem Bernoulli-prozeß die Anzahl der Mißerfolge vor dem ersten Erfolg zählt.

Dies kann man sich folgendermaßen klarmachen:

Sei X die Anzahl der Mißerfolge vor dem ersten Erfolg bei einem Bernoulli-prozeß. Die Zufallsvariable X kann die Werte $0, 1, 2, \dots$ annehmen.

Nimmt X den Wert x an, so wurden x Mißerfolge \bar{A} und genau ein Erfolg A beobachtet, wobei das A das letzte Symbol in der Folge ist.

Wir erhalten also

$$\underbrace{\bar{A}\bar{A}\dots\bar{A}}_{x\text{-mal}}A$$

Aufgrund der Unabhängigkeit der einzelnen Durchführungen beträgt die Wahrscheinlichkeit dieser Folge

$$\underbrace{(1 - p)(1 - p)\dots(1 - p)}_{x\text{-mal}}p = (1 - p)^x p.$$

Satz 3.2.1 *Sei X mit Parameter p geometrisch verteilt. Dann gilt*

$$P(X \geq x) = (1 - p)^x \quad (3.18)$$

Beweis:

$$\begin{aligned} P(X \geq x) &= \sum_{i=x}^{\infty} p(1-p)^i \\ &= p \sum_{i=x}^{\infty} (1-p)^i \\ &= p \sum_{j=0}^{\infty} (1-p)^{j+x} \\ &= p(1-p)^x \sum_{j=0}^{\infty} (1-p)^j \\ &= p(1-p)^x \frac{1}{1-(1-p)} \\ &= (1-p)^x \end{aligned}$$

Beispiel 3.2.4 *Schauen wir uns noch einmal die teetrinkende Dame an.*

Anstatt die Anzahl der Versuchen vorab festzulegen, lassen wir die Dame so lange probieren, bis sie zum ersten Mal falsch klassifiziert.

Dies ist beim sechsten Versuch.

Wir zählen die Anzahl T der Versuche vor der ersten Mißklassifikation bei der Dame.

Unter H_0 gilt

$$P(T = t) = 0.5^{t+1} \quad \text{für } t = 0, 1, \dots$$

Je mehr Erfolge die Dame vor dem ersten Mißerfolg hat, um so mehr spricht dies für H_1 .

Wir lehnen H_0 also ab, wenn T zu groß wird.

Wir können für das Beispiel die Überschreitungswahrscheinlichkeit bestimmen.

Es gilt

$$P(T \geq 5) = 0.5^5 = 0.03125$$

Wir sehen, daß die Überschreitungswahrscheinlichkeit kleiner als 0.05 ist. Wir würden uns also für die Gegenhypothese entscheiden, die besagt, daß die Dame nicht rät.

Wir gelangen also beim gleichen Datensatz zu unterschiedlichen Entscheidungen.

Eine detaillierte Analyse findet man in Lindley.

Beispiel 3.2.5 *Wir warten so lange, bis das Teilchen zum ersten Mal nach links geht, und zählen die Anzahl X der Schritte nach rechts vor dem ersten Schritt nach links.*

Die Zufallsvariable X ist geometrisch verteilt mit $p = 0.5$.

Es gilt

$$P(X = x) = 0.5^{x+1} \quad \text{für } x = 0, 1, 2, \dots$$

Beobachtet man einen Bernoulli-prozeß bis zum r .ten Erfolg, so erhält man die Pascal-Verteilung.

Definition 3.2.5 *Die Zufallsvariable X heißt pascalverteilt mit den Parametern r und p , wenn ihre Wahrscheinlichkeitsfunktion gegeben ist durch:*

$$P(X = x) = \binom{x+r-1}{x} p^r (1-p)^x \quad (3.19)$$

für $x = 0, 1, \dots$

Wir schreiben auch $X \sim \text{PASCAL}(r, p)$.

Die Pascalverteilung erhält man, wenn man bei einem Bernoulli-prozeß die Anzahl der Mißerfolge vor dem r .ten Erfolg zählt.

Dies kann man sich folgendermaßen klarmachen:

Sei X die Anzahl der Mißerfolge vor dem r .ten Erfolg bei einem Bernoulli-prozeß.

Die Zufallsvariable X kann die Werte $0, 1, 2, \dots$ annehmen.

Nimmt X den Wert x an, so wurden x Mißerfolge \bar{A} und r Erfolge A beobachtet. Aufgrund der Unabhängigkeit der einzelnen Durchführungen beträgt die Wahrscheinlichkeit einer Folge aus r Erfolgen A und x Mißerfolgen \bar{A} $p^r (1-p)^x$.

Wieviele derartige Folgen gibt es?

Von den $x + r$ Symbolen ist das letzte auf jeden Fall ein A , da wir den Prozeß solange beobachten, bis das r .te A eintritt.

Für die restlichen x Mißerfolge \bar{A} und $r - 1$ Erfolge A gibt es

$$\binom{x + r - 1}{x}$$

unterschiedliche Anordnungen.

Somit gilt

$$P(X = x) = \binom{x + r - 1}{x} p^r (1 - p)^x \quad \text{für } x = 0, 1, \dots$$

Ist bei der Pascal-Verteilung $r = 1$, so erhält man die geometrische Verteilung, denn es gilt

$$P(X = x) = \binom{x + 1 - 1}{x} p^1 (1 - p)^x = p (1 - p)^x$$

Die Realisation eines Bernoulliprozesses der Länge n kann man folgendermaßen erhalten.

Man betrachtet eine Urne, die N Kugeln enthält, von denen M Kugeln weiß und die restlichen $N - M$ Kugeln schwarz sind.

Man zieht n Kugeln mit Zurücklegen aus dieser Urne

Sei X die Anzahl der weißen Kugeln unter den n gezogenen Kugeln.

Dann gilt

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x} \tag{3.20}$$

mit

$$p = \frac{M}{N}$$

Zieht man ohne Zurücklegen aus der Urne, so besitzt die Anzahl X der weißen Kugeln eine andere Verteilung.

Die Wahrscheinlichkeit $P(X = x)$ erhalten wir in diesem Fall folgendermaßen:

Insgesamt gibt es

$$\binom{N}{n}$$

Möglichkeiten, von den N Kugeln n ohne Zurücklegen auszuwählen.

Es gibt

$$\binom{M}{x}$$

Möglichkeiten, von den M weißen Kugeln x Kugeln auszuwählen, und es gibt

$$\binom{N-M}{n-x}$$

Möglichkeiten, von den $N - M$ schwarzen Kugeln $n - x$ Kugeln auszuwählen.

Somit gilt

$$P(X = x) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}}$$

für $\max\{0, n - M\} \leq x \leq \min\{n, N\}$.

Definition 3.2.6 Die Zufallsvariable X heißt *hypergeometrisch verteilt* mit den Parametern N , M und n , wenn ihre Wahrscheinlichkeitsfunktion gegeben ist durch

$$P(X = x) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} \quad (3.21)$$

für $\max\{0, n - M\} \leq x \leq \min\{n, N\}$.

Wir schreiben auch $X \sim \text{HYPERGEOM}(N, M, n)$

Im Beispiel 3.2.3 haben wir einen statistischen Test kennengelernt.

Statistische Tests gehören zum Bereich der schließenden Statistik. Hier soll auf der Basis einer Stichprobe eine Aussage über eine Grundgesamtheit gemacht werden.

Neben Tests betrachtet man in der schließenden Statistik noch Schätzer.

Schauen wir uns ein Beispiel an.

Beispiel 3.2.6 *Es soll die Größe N einer Population, zum Beispiel die Anzahl Fische in einem Teich, geschätzt werden. Hierzu entnimmt man der Population M Objekte und markiert diese. Danach entläßt man die Objekte in die Population. Nach einiger Zeit entnimmt man der Population n Objekte und zählt die markierten.*

Wie kann man auf der Basis dieser Information die Größe der Population schätzen?

Wir gehen zunächst davon aus, daß die Größe N der Population bekannt ist, und bestimmen die Wahrscheinlichkeit, unter den n Objekten x markierte zu finden. Offensichtlich ist die Anzahl X der markierten Objekte unter den gefangenen Objekten mit den Parametern N , M und n hypergeometrisch verteilt.

Es gilt also

$$P(X = x) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}}.$$

Wir kennen nun aber nicht N , x hingegen ist bekannt.

Wir fassen die Wahrscheinlichkeit $P(X = x)$ als Funktion $L(N)$ von N auf. Für einen Wert N ist $L(N)$ die Wahrscheinlichkeit, x markierte Objekte in der Stichprobe zu finden, wenn sich in der Population N Objekte befinden. Je größer diese Wahrscheinlichkeit ist, um so plausibler ist es, daß der zugehörige Wert n von N der wahre Wert ist.

*Man nennt $L(N)$ auch die **Likelihoodfunktion**.*

*Den Wert, für den die Likelihood am größten ist, nennt man den **Maximum-Likelihood-Schätzer**.*

Wie sieht der Maximum-Likelihood-Schätzer für das Beispiel aus?

Hierzu benutzen wir folgende Eigenschaft der Binomialkoeffizienten, die wir in Satz 2.2.15 bewiesen haben.

$$\binom{n}{k} = \frac{n}{n-k} \binom{n-1}{k}$$

Somit gilt

$$\begin{aligned}
 L(N) &= \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} \\
 &= \frac{\frac{N-M}{N-M-n+x} \binom{M}{x} \binom{N-M-1}{n-x}}{\frac{N}{N-n} \binom{N-1}{n}} \\
 &= \frac{(N-M)(N-n)}{N(N-M-n+x)} L(N-1)
 \end{aligned}$$

Somit gilt

$$\frac{L(N)}{L(N-1)} = \frac{(N-M)(N-n)}{N(N-M-n+x)}$$

Dieser Bruch ist größer als 1 und somit $L(N)$ monoton wachsend für:

$$\frac{(N-M)(N-n)}{N(N-M-n+x)} > 1$$

$$(N-M)(N-n) > N(N-M-n+x)$$

$$N^2 - Nn - MN + Mn > N^2 - MN - Nn - Nx$$

$$Mn > Nx$$

$$\frac{Mn}{x} > N$$

$L(N)$ wächst also für

$$N < \frac{Mn}{x}$$

und fällt für

$$N > \frac{Mn}{x}.$$

Also nimmt die Likelihood ihr Maximum an für

$$N = \lfloor \frac{M}{x} n \rfloor$$

Was passiert mit der Wahrscheinlichkeitsfunktion der hypergeometrischen Verteilung, wenn N beliebig groß wird, wobei aber $p = \frac{M}{N}$ konstant bleibt? Es gilt:

$$\begin{aligned}
 P(X = x) &= \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} \\
 &= \frac{(M)_x (N-M)_{n-x} n!}{x! (n-x)! (N)_n} \\
 &= \binom{n}{x} \frac{(M)_x (N-M)_{n-x}}{(N)_n} \\
 &= \binom{n}{x} \frac{M}{N} \cdots \frac{M-x+1}{N-x+1} \frac{N-M}{N-x} \cdots \frac{N-M-n+x+1}{N-n+1}
 \end{aligned}$$

Also gilt

$$\begin{aligned}
 &\lim_{N \rightarrow \infty} P(X = x) \\
 &= \lim_{N \rightarrow \infty} \binom{n}{x} \frac{M}{N} \frac{M-1}{N-1} \cdots \frac{M-x+1}{N-x+1} \frac{N-M}{N-x} \cdots \frac{N-M-n+x+1}{N-n+1} \\
 &= \binom{n}{x} p^x (1-p)^{n-x}
 \end{aligned}$$

Für große Werte von N können wir also die hypergeometrische Verteilung durch die Binomialverteilung approximieren.

Außerdem bestätigt die Aussage die intuitive Annahme, daß es keinen Unterschied macht, ob man aus einer sehr großen Grundgesamtheit mit oder ohne Zurücklegen zieht.

Schauen wir uns noch an, was mit der Wahrscheinlichkeitsfunktion der Binomialverteilung geschieht, wenn wir n beliebig groß werden lassen. Auch hier müssen wir beim Grenzübergang eine Restriktion berücksichtigen, da sonst die Wahrscheinlichkeit gegen Null geht. Wir fordern, daß $\lambda = np$ konstant bleibt.

Somit gilt

$$p = \frac{\lambda}{n}.$$

$$\begin{aligned} \lim_{n \rightarrow \infty} P(X = x) &= \lim_{n \rightarrow \infty} \binom{n}{x} p^x (1-p)^{n-x} \\ &= \lim_{n \rightarrow \infty} \frac{n!}{x!(n-x)!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \\ &= \frac{\lambda^x}{x!} \lim_{n \rightarrow \infty} \frac{(n)_x}{n^x} \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{-x} \\ &= \frac{\lambda^x}{x!} e^{-\lambda} \end{aligned}$$

da gilt

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{(n)_x}{n^x} &= \lim_{n \rightarrow \infty} \frac{n}{n} \frac{n-1}{n} \dots \frac{n-x+1}{n} = 1 \\ \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n &= e^{-\lambda} \\ \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{-x} &= 1 \end{aligned}$$

Definition 3.2.7 Die Zufallsvariable X heißt *poissonverteilt* mit dem Parameter λ , wenn ihre Wahrscheinlichkeitsfunktion gegeben ist durch:

$$P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda} \quad (3.22)$$

für $x = 0, 1, \dots$

Wir schreiben auch $X \sim \text{POISSON}(\lambda)$.

Die Poissonverteilung wird auch die Verteilung der seltenen Ereignisse genannt, da $p = \frac{\lambda}{n}$ mit wachsendem n immer kleiner wird.

Die Poissonverteilung spielt eine zentrale Rolle bei Ankunftsprozessen. Hier soll die Anzahl $X(a, t)$ der Ankünfte im Intervall $(a, a + t]$ modelliert werden. Dabei geht man von folgenden Annahmen aus:

1. $X(a, t)$ und $X(a^*, t^*)$ sind unabhängig, wenn $(a, a + t]$ und $(a^*, a^* + t^*]$ disjunkt sind.
2. $X(a, h)$ hängt von h aber nicht von a ab.
3. $P(X(a, h) = 1) = \lambda h + o(h)$.
4. $P(X(a, h) > 1) = o(h)$.

Für die Funktion $o(h)$ gilt

$$\lim_{h \rightarrow 0} \frac{o(h)}{h} = 0.$$

Annahme 1 besagt, daß die Häufigkeit des Auftretens von A in einem Intervall ist von der Häufigkeit des Auftretens von A in einem dazu disjunkten Intervall.

Annahme 2 besagt, daß die Häufigkeit des Auftretens von A in einem Intervall der Länge h nur von der Länge, aber nicht von der Lage des Intervalls abhängt.

Annahme 3 besagt, daß die Wahrscheinlichkeit, daß A in einem sehr kleinen Intervall der Länge h genau einmal eintritt, proportional zur Länge des Intervalls ist.

Annahme 4 besagt, daß die Wahrscheinlichkeit, daß A in einem sehr kleinen Intervall der Länge h mehr als einmal eintritt, vernachlässigt werden kann.

Schauen wir uns dafür ein Beispiel an:

Beispiel 3.2.7 *Der Verkehrsstrom einer Fernstraße wird an einem festen Beobachtungspunkt protokolliert, wobei jeweils festgehalten wird, wann ein Fahrzeug den Beobachtungspunkt passiert.*

Annahme 1 besagt, daß der Verkehrsstrom völlig regellos ist. Ereignisse in nichtüberlappenden Intervallen sind unabhängig.

Annahme 2 besagt, daß der Verkehrsstrom von konstanter Intensität ist. Die Wahrscheinlichkeit, daß in einem Zeitintervall der Länge h k Fahrzeuge eintreffen, hängt von der Länge des Intervalls, nicht jedoch von seiner Lage ab.

Annahme 4 besagt, daß das Auftreten geschlossener Fahrzeuggruppen vernachlässigt werden kann.

Unter den Annahmen 1 bis 4 ist $X(a, t)$ poissonverteilt mit dem Parameter λt .

Es gilt also

$$P(X = x) = \frac{(\lambda t)^x}{x!} e^{-\lambda t} \quad \text{für } x = 0, 1, \dots$$

Dies sieht man folgendermaßen:

Wir teilen das Intervall $(a, a + t]$ in n gleichgroße Intervalle der Länge $h = \frac{t}{n}$. Aufgrund von Annahme 4 kann A in jedem dieser Intervalle nur einmal oder keinmal eintreten.

In jedem dieser Intervalle beträgt die Wahrscheinlichkeit, A genau einmal zu beobachten,

$$p = \lambda h = \frac{\lambda t}{n}$$

Da die Intervalle disjunkt sind, ist das Eintreten von A in den einzelnen Intervallen unabhängig. Außerdem ist jedes Intervall gleich lang, so daß p konstant ist.

Es liegt also ein Bernoulli-Prozeß der Länge n mit Erfolgswahrscheinlichkeit $p = \frac{\lambda t}{n}$ vor.

Somit gilt

$$P(X(a, t) = x) = \binom{n}{x} \left(\frac{\lambda t}{n}\right)^x \left(1 - \frac{\lambda t}{n}\right)^{n-x}$$

Lassen wir n über alle Grenzen wachsen, so erhalten wir

$$P(X = x) = \frac{(\lambda t)^x}{x!} e^{-\lambda t} \quad \text{für } x = 0, 1, \dots$$

3.3 Stetige Verteilungsmodelle

Eine Zufallsvariable heißt stetig, wenn eine Funktion $f_X : \mathfrak{R} \rightarrow \mathfrak{R}$ existiert, so daß für die Verteilungsfunktion $F_X(x)$ von X gilt:

$$F_X(x) = \int_{-\infty}^x f_X(u) du \quad (3.23)$$

Die Funktion $f_X(x)$ heißt Dichtefunktion der Zufallsvariablen X .

Die Dichtefunktion $f_X(x)$ erfüllt folgende Annahmen:

1.

$$f_X(x) \geq 0 \quad (3.24)$$

für alle $x \in \mathfrak{R}$.

2.

$$\int_{-\infty}^{\infty} f_X(u) du = 1 \quad (3.25)$$

Man kann nun für ein stetiges Merkmal dadurch ein Modell gewinnen, daß man eine Funktion $f_X(x)$ vorgibt, die (3.24) und (3.25) erfüllt.

So werden wir im folgenden vorgehen.

Wir suchen zunächst geeignete Modelle für Meßfehler.

Bei der wiederholten Messung einer Größe Y stellt man fest, daß die einzelnen Messungen unterschiedliche Werte ergeben.

Wie kann diesen Tatbestand modellieren?

Ein einfaches Modell ist

$$Y = \mu + X.$$

Dabei ist μ der wahre Wert der zu messenden Größe und X der Meßfehler.

Aus den Annahmen über die Verteilung des Meßfehlers ergeben sich eine Reihe von stetigen Verteilungsmodellen.

Annahme 1

Die Meßfehler sollten nicht systematisch sein.

Aus Annahme 1 folgt, daß die Verteilung der Meßfehler das Zentrum 0 haben sollte.

Annahme 2

Positive Meßfehler sollten die gleiche Verteilung besitzen wie negative Meßfehler.

Aus Annahme 2 folgt, daß die Dichtefunktion der Meßfehler symmetrisch bezüglich ihres Zentrums sein sollte.

Aus den beiden Annahmen folgt, daß die Dichtefunktion einer sinnvollen Meßfehlerverteilung also symmetrisch bezüglich 0 sein sollte.

Wir schauen uns nun eine Reihe von Dichtefunktionen an, die diese Bedingung erfüllen.

Die einfachste Verteilung, die diese Annahme erfüllt, ist die Gleichverteilung auf dem Intervall $(-c, c)$.

Diese ist ein Spezialfall der Gleichverteilung auf (a, b) .

Definition 3.3.1 Die Zufallsvariable X heißt gleichverteilt auf dem Intervall (a, b) , wenn ihre Dichtefunktion gegeben ist durch:

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{für } a < x < b \\ 0 & \text{sonst} \end{cases} \quad (3.26)$$

Wir schreiben auch $X \sim \text{GLEICH}(a, b)$.

Die Verteilungsfunktion der Gleichverteilung auf (a, b) ist gegeben durch

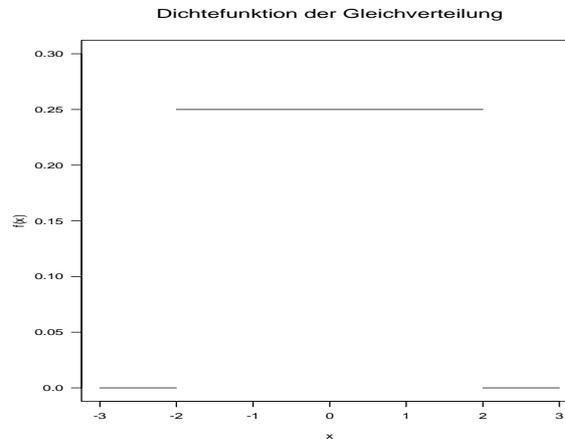
$$F_X(x) = \begin{cases} 0 & \text{für } x \leq a \\ \frac{x-a}{b-a} & \text{für } a < x < b \\ 1 & \text{für } x \geq b \end{cases} \quad (3.27)$$

Für $a = -c$ und $b = c$ erhalten wir die Gleichverteilung auf $(-c, c)$.

Die Dichtefunktion der Gleichverteilung auf $(-c, c)$ ist gegeben durch

$$f_X(x) = \begin{cases} \frac{1}{2c} & \text{für } -c < x < c \\ 0 & \text{sonst} \end{cases} \quad (3.28)$$

Das folgende Bild zeigt die Dichtefunktion für $c = 2$.



Bei der Gleichverteilung wird unterstellt, daß betragsmäßig große Fehler genauso häufig sind wie kleine. Sinnvollerweise sollte die Wahrscheinlichkeit betragsmäßig großer Meßfehler geringer sein als die betragsmäßig kleiner Meßfehler.

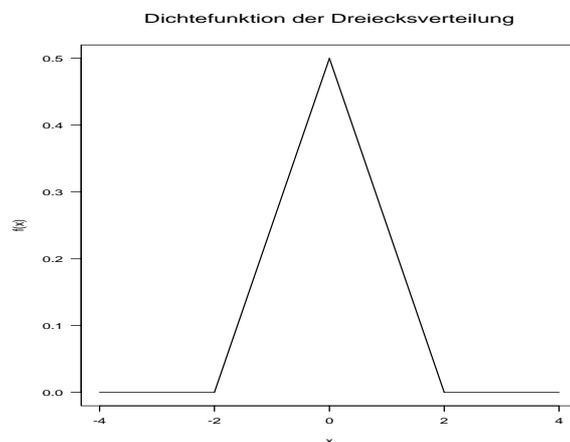
Dies ist bei der Dreiecksverteilung der Fall.

Definition 3.3.2 Die Zufallsvariable X heißt dreiecksverteilt auf dem Intervall $(-c, c)$, wenn ihre Dichtefunktion gegeben ist durch:

$$f_X(x) = \begin{cases} \frac{1}{c} + \frac{1}{c^2}x & \text{für } -c \leq x \leq 0 \\ \frac{1}{c} - \frac{1}{c^2}x & \text{für } 0 \leq x \leq c \\ 0 & \text{sonst} \end{cases} \quad (3.29)$$

Wir schreiben auch $X \sim \text{DREIECK}(-c, c)$.

Die nachstehende Graphik zeigt die Dichtefunktion der Dreiecksverteilung für $c = 2$.



Sowohl die Gleichverteilung als auch die Dreiecksverteilung haben als Modell für Meßfehler den Nachteil, daß die Größe der Meßfehler nach oben und unten begrenzt ist.

Von Laplace wurde folgende Verteilung als Modell vorgeschlagen, die diesen Nachteil nicht hat:

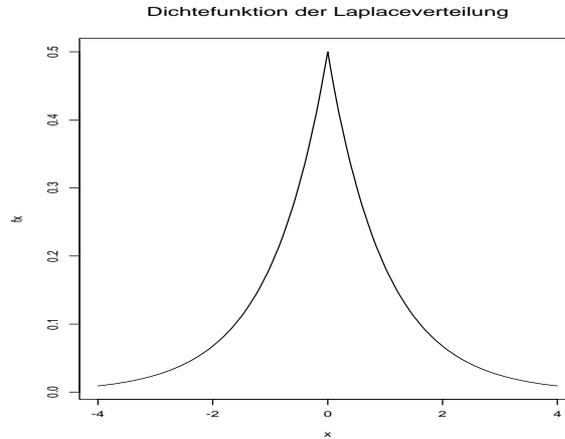
Definition 3.3.3 Die Zufallsvariable X heißt laplaceverteilt, wenn ihre Dichtefunktion gegeben ist durch:

$$f_X(x) = \frac{1}{2} e^{-|x|} \quad \text{für } x \in \mathfrak{R} \quad (3.30)$$

Wir schreiben auch $X \sim \text{LAPLACE}$.

Man nennt die Laplaceverteilung auch Doppelsexponentialverteilung.

Das folgende Bild zeigt die Dichtefunktion der Laplace-Verteilung:



Die Laplaceverteilung hat den Nachteil, daß die Dichtefunktion an der Stelle 0 nicht differenzierbar ist.

Diesen Nachteil hat die von Gauss vorgeschlagene Normalverteilung nicht.

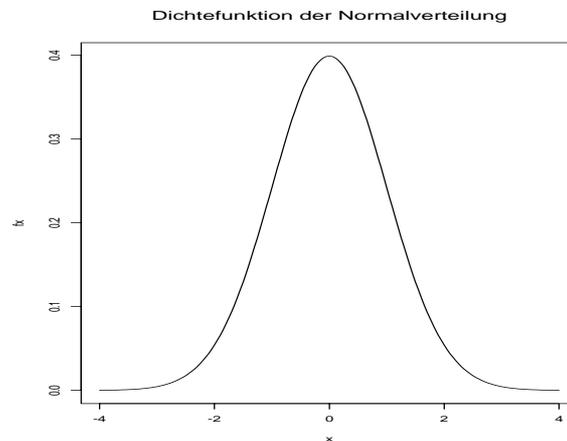
Definition 3.3.4 Die Zufallsvariable X heißt normalverteilt mit den Parametern μ und σ^2 , wenn ihre Dichtefunktion gegeben ist durch:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{für } x \in \mathfrak{R} \quad (3.31)$$

Wir schreiben $X \sim N(\mu, \sigma^2)$.

Die Normalverteilung ist die wichtigste Verteilung der Statistik. Sie wird in diesem Skript also immer wieder auftauchen.

Das folgende Bild zeigt die Dichtefunktion der Normalverteilung mit $\mu = 0$ und $\sigma = 1$, die **Standardnormalverteilung** heißt:



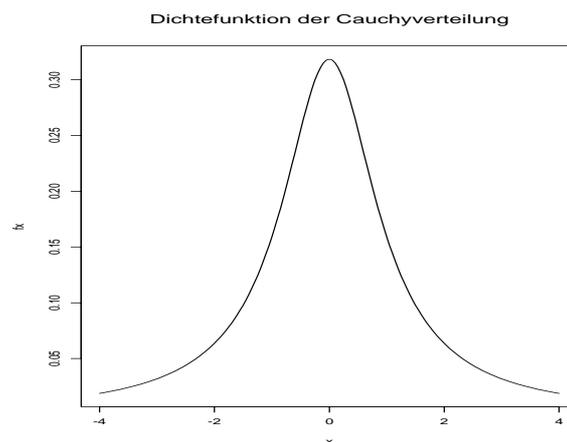
Neben der Normalverteilung könnte man auch die Cauchyverteilung verwenden.

Definition 3.3.5 Die Zufallsvariable X heißt *cauchyverteilt*, wenn ihre Dichtefunktion gegeben ist durch:

$$f_X(x) = \frac{1}{\pi(1+x^2)} \quad (3.32)$$

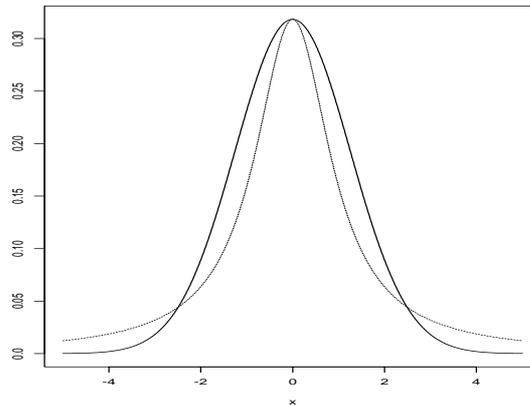
Wir schreiben auch $X \sim \text{CAUCHY}$.

Das folgende Bild zeigt die Dichtefunktion der Cauchyverteilung:



Auf den ersten Blick sieht die Cauchyverteilung der Normalverteilung sehr ähnlich.

In der folgenden Graphik sind die Dichtefunktionen der Normalverteilung (durchgezogene Linie) und Cauchyverteilung (gestrichelte Linie) eingezeichnet.



Um die beiden Dichtefunktionen vergleichbar zu machen, sind sie so skaliert, daß sie den gleichen Funktionswert am Nullpunkt annehmen.

Wir sehen, daß die Cauchyverteilung mehr Wahrscheinlichkeitsmasse an den Rändern besitzt.

Kommen wir noch einmal zum Poisson-Prozeß zurück. Bei einem Poisson-Prozeß im Intervall $(0, t]$ ist die absolute Häufigkeit des Ereignisses A poissonverteilt mit Parameter λt .

Es gilt also

$$P(X = x) = \frac{(\lambda t)^x}{x!} e^{-\lambda t} \quad \text{für } x = 0, 1, \dots$$

Wir wollen nun einen Poisson-Prozeß so lange beobachten, bis A zum ersten Mal eintritt. Gesucht ist Dichtefunktion $f_T(t)$ und die Verteilungsfunktion $F_T(t)$ der Wartezeit T bis zum ersten Eintreten von A .

Für $t < 0$ gilt

$$F_T(t) = 0.$$

Für $t \geq 0$ gilt:

$$\begin{aligned} F_T(t) &= P(T \leq t) \\ &= 1 - P(T > t) \\ &= 1 - P(X = 0) \\ &= 1 - e^{-\lambda t} \end{aligned}$$

Somit gilt:

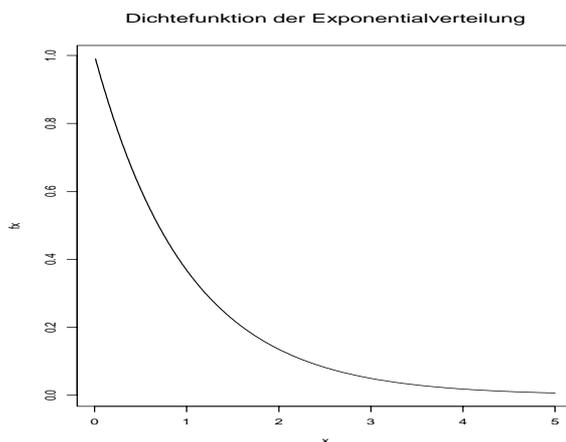
$$F_T(t) = \begin{cases} 1 - e^{-\lambda t} & \text{für } t > 0 \\ 0 & \text{sonst} \end{cases}$$

Definition 3.3.6 Die Zufallsvariable X heißt *exponentialverteilt* mit Parameter λ , wenn ihre Dichtefunktion gegeben ist durch:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{für } x > 0 \\ 0 & \text{sonst} \end{cases} \quad (3.33)$$

Wir schreiben auch $X \sim EXPO(\lambda)$.

Das folgende Bild zeigt die Dichtefunktion der Exponentialverteilung mit $\lambda = 1$.



Die Exponentialverteilung ist eine schiefe Verteilung. Sie ist ein Spezialfall der Gammaverteilung.

Definition 3.3.7 Die Zufallsvariable X heißt *gammaverteilt* mit den Parametern r und λ , falls ihre Dichtefunktion lautet:

$$f_X(x) = \begin{cases} \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} & \text{für } x > 0 \\ 0 & \text{sonst} \end{cases} \quad (3.34)$$

Dabei heißt

$$\Gamma(r) = \int_0^{\infty} x^{r-1} e^{-x} dx$$

die *Gammafunktion*.

Satz 3.3.1 Für die Gammafunktion $\Gamma(r)$ gilt:

1. $\Gamma(1) = 1$
2. $\Gamma(n) = (n - 1)\Gamma(n - 1)$
3. $\Gamma(0.5) = \sqrt{\pi}$

Beweis:

1.

$$\Gamma(1) = \int_0^{\infty} x^{1-1} e^{-x} dx = \int_0^{\infty} e^{-x} dx = \left[-e^{-x} \right]_0^{\infty} = 1$$

2.

$$\begin{aligned} \Gamma(n) &= \int_0^{\infty} x^{n-1} e^{-x} dx = \\ &= \left[-x^{n-1} e^{-x} \right]_0^{\infty} + \int_0^{\infty} (n-1)x^{n-2} e^{-x} dx \\ &= (n-1)\Gamma(n-1) \end{aligned}$$

3. Auf diesen Beweis verzichten wir.

Die Gamma-Verteilung mit den Parametern $r = \frac{k}{2}$ und $\lambda = \frac{1}{2}$ heißt Chi-Quadrat-Verteilung mit k Freiheitsgraden. In der schließenden Statistik spielt die Chi-Quadrat-Verteilung eine zentrale Rolle.

Definition 3.3.8 Die Zufallsvariable X heißt chi-Quadrat-verteilt mit k Freiheitsgraden, falls ihre Dichtefunktion lautet:

$$f_X(x) = \begin{cases} \frac{1}{\Gamma(k/2)} \left(\frac{1}{2}\right)^{k/2} x^{k/2-1} e^{-0.5x} & \text{für } x > 0 \\ 0 & \text{sonst} \end{cases} \quad (3.35)$$

Betrachten wir noch einmal einen Poissonprozeß im Intervall $(0, t]$.

Sei T die Wartezeit bis zum r -ten Erfolg.

Wir suchen die Verteilungsfunktion $F_T(t)$ und die Dichtefunktion $f_T(t)$.

Wir wissen, daß die Anzahl X der Erfolge in $(0, t]$ poissonverteilt ist mit Parameter λt .

Wenn die Wartezeit bis zum r -ten Erfolg mehr als t Zeiteinheiten beträgt, so sind höchstens $r - 1$ Erfolge in $(0, t]$ beobachtet worden.

Es gilt also

$$\begin{aligned} P(T > t) &= P(X \leq r - 1) \\ &= \sum_{k=0}^{r-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t} \end{aligned}$$

Also ist die Verteilungsfunktion von T gegeben durch

$$\begin{aligned} F_T(t) &= 1 - P(T > t) \\ &= 1 - \sum_{k=0}^{r-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t} \end{aligned}$$

Man spricht von der Erlang-Verteilung mit den Parametern r und λ . Für die Dichtefunktion gilt

$$\begin{aligned} f_T(t) &= \frac{d}{dt} F_T(t) \\ &= \frac{d}{dt} \left(1 - \sum_{k=0}^{r-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t} \right) \\ &= - \sum_{k=0}^{r-1} \left(\frac{k (\lambda t)^{k-1}}{k!} e^{-\lambda t} - \frac{(\lambda t)^k}{k!} \lambda e^{-\lambda t} \right) \\ &= \sum_{k=0}^{r-1} \frac{(\lambda t)^k}{k!} \lambda e^{-\lambda t} - \sum_{k=1}^{r-1} \frac{(\lambda t)^{k-1}}{(k-1)!} e^{-\lambda t} \\ &= \sum_{k=0}^{r-1} \frac{(\lambda t)^k}{k!} \lambda e^{-\lambda t} - \sum_{k=0}^{r-2} \frac{(\lambda t)^k}{k!} e^{-\lambda t} \\ &= \frac{(\lambda t)^{r-1}}{(r-1)!} \lambda e^{-\lambda t} + \sum_{k=0}^{r-2} \frac{(\lambda t)^k}{k!} \lambda e^{-\lambda t} - \sum_{k=0}^{r-2} \frac{(\lambda t)^k}{k!} e^{-\lambda t} \\ &= \frac{t^{r-1}}{(r-1)!} \lambda^r e^{-\lambda t} \end{aligned}$$

Da für natürliche Zahlen r gilt

$$\Gamma(r) = (r - 1)!$$

gilt

$$f_T(t) = \frac{t^{r-1}}{\Gamma(r)} \lambda^r e^{-\lambda t}$$

Dies ist aber die Dichtefunktion einer Gammaverteilung mit den Parametern r und λ .

Die Erlangverteilung ist also eine spezielle Gammaverteilung.

Wir sehen, daß die Wartezeit bis zum r -ten Erfolg bei einem Poissonprozeß gammaverteilt ist mit den Parametern r und λ .

Schauen wir uns noch ein weiteres stetiges Verteilungsmodell an.

Definition 3.3.9 Die Zufallsvariable X heißt betaverteilt mit den Parametern a und b , wenn ihre Dichtefunktion lautet:

$$f_X(x) = \begin{cases} \frac{1}{B(a, b)} x^{a-1} (1-x)^{b-1} & \text{für } 0 < x < 1, a, b > 0 \\ 0 & \text{sonst} \end{cases}$$

Dabei ist

$$B(a, b) = \int_0^1 w^{a-1} (1-w)^{b-1} dw$$

die Betafunktion.

Wir schreiben $X \sim \text{BETA}(a, b)$.

Für die Betafunktion gilt

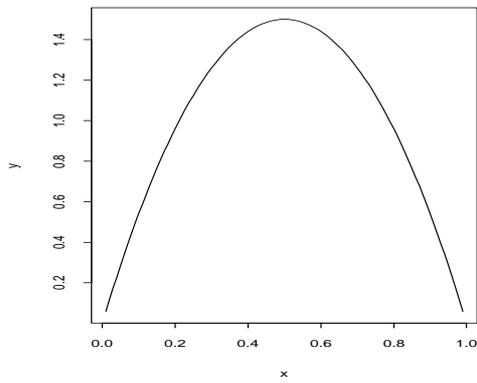
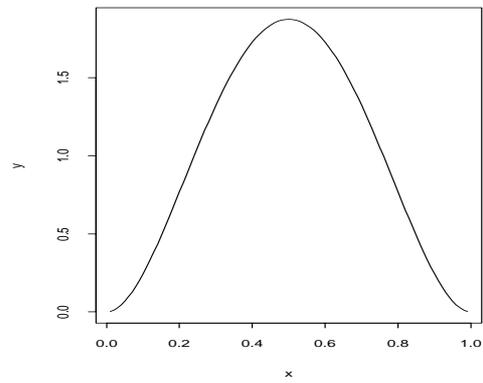
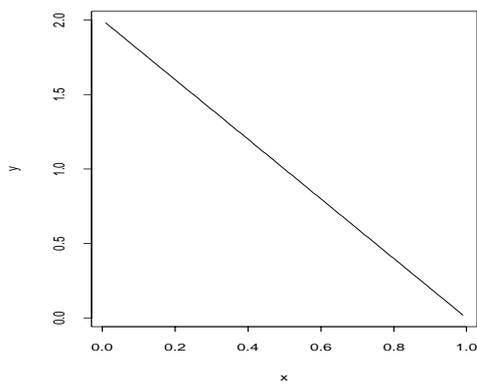
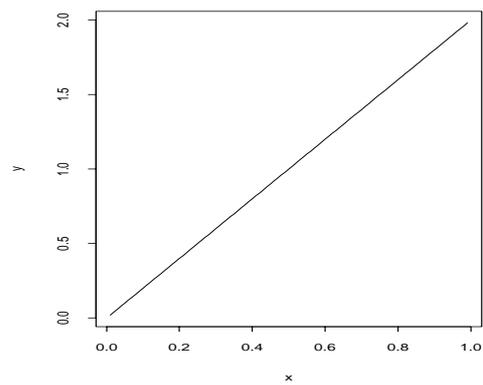
$$B(a, b) = \frac{\Gamma(a) \Gamma(b)}{\Gamma(a+b)}$$

mit

$$\Gamma(r) = \int_0^\infty x^{r-1} e^{-x} dx$$

Für $a = 1$ und $b = 1$ erhalten wir die Gleichverteilung auf $(0, 1)$.

Die folgenden Graphiken zeigen die Dichtefunktion der Beta-Verteilung für ausgewählte Werte der Parameter.

Beta-Verteilung mit $a=2$ und $b=2$ Beta-Verteilung mit $a=3$ und $b=3$ Beta-Verteilung mit $a=1$ und $b=2$ Beta-Verteilung mit $a=2$ und $b=1$ 

3.4 Die Exponentialfamilie

Viele der Verteilungen, die wir in den letzten beiden Abschnitten kennengelernt haben, besitzen eine gemeinsame Struktur.

Die Dichtefunktion bzw. Wahrscheinlichkeitsfunktion $f_X(x)$ kann in der folgenden Form dargestellt werden:

$$f_X(x) = e^{\theta x - b(\theta) + d(x)} \quad (3.36)$$

wobei der Träger \mathcal{T}_X von X nicht von θ abhängt.

Man spricht von der **Exponentialfamilie** in kanonischer Form.

Der Parameter θ heißt auch **natürlicher Parameter**.

Die Exponentialfamilie spielt eine zentrale bei sogenannten verallgemeinerten linearen Modellen (**GLIM: Generalized Linear Models**). Hier werden beim Schätzen und Testen Eigenschaften ausgenutzt, auf die wir an späterer Stelle detailliert eingehen werden.

Beispiel 3.4.1 Für die Bernoulli-Verteilung gilt

$$P(X = x) = e^{\theta x - b(\theta) + d(x)}$$

mit

$$b(\theta) = \ln(1 + e^\theta)$$

und

$$d(x) = 0.$$

Zwischen θ und p besteht folgende Beziehung:

$$\theta = \ln \frac{p}{1-p}. \quad (3.37)$$

Dies sieht man folgendermaßen:

$$p^x(1-p)^{1-x} = \left(\frac{p}{1-p}\right)^x (1-p) = e^{x \ln \frac{p}{1-p} + \ln(1-p)}$$

Wir setzen

$$\theta = \ln \frac{p}{1-p}.$$

Lösen wir diese Gleichung nach p auf, so erhalten wir

$$p = \frac{e^\theta}{1 + e^\theta}.$$

Somit gilt

$$e^{x \ln \frac{p}{1-p} + \ln(1-p)} = e^{x\theta + \ln(1 - \frac{e^\theta}{1+e^\theta})} = e^{x\theta - \ln(1+e^\theta)}$$

Beispiel 3.4.2 Für die Binomialverteilung gilt

$$P(X = x) = e^{\theta x - b(\theta) + d(x)}$$

mit

$$b(\theta) = n \ln(1 + e^\theta)$$

und

$$d(x) = \ln \binom{n}{x}.$$

Zwischen θ und p besteht folgende Beziehung:

$$\theta = \ln \frac{p}{1-p}. \quad (3.38)$$

Dies sieht man folgendermaßen:

$$\binom{n}{x} p^x (1-p)^{n-x} = e^{x \ln \frac{p}{1-p} + n \ln(1-p) + \ln \binom{n}{x}}$$

Wir setzen

$$\theta = \ln \frac{p}{1-p}.$$

Lösen wir diese Gleichung nach p auf, so erhalten wir

$$p = \frac{e^\theta}{1 + e^\theta}$$

Somit gilt

$$\begin{aligned} e^{x \ln \frac{p}{1-p} + n \ln(1-p) + \ln \binom{n}{x}} &= e^{x\theta + n \ln(1 - \frac{e^\theta}{1+e^\theta}) + \ln \binom{n}{x}} \\ &= e^{x\theta - n \ln(1+e^\theta) + \ln \binom{n}{x}} \end{aligned}$$

Beispiel 3.4.3 Für die Poissonverteilung gilt

$$P(X = x) = e^{\theta x - b(\theta) + d(x)}$$

mit

$$b(\theta) = e^{\theta}$$

und

$$d(x) = -\ln x!.$$

Zwischen θ und λ besteht folgende Beziehung:

$$\theta = \ln \lambda. \tag{3.39}$$

Dies sieht man folgendermaßen:

$$\frac{\lambda^x}{x!} e^{-\lambda} = e^{x \ln \lambda - \lambda - \ln x!}$$

Wir setzen

$$\theta = \ln \lambda.$$

Lösen wir diese Gleichung nach λ auf, so erhalten wir

$$\lambda = e^{\theta}$$

Somit gilt

$$e^{x \ln \lambda - \lambda - \ln x!} = e^{x\theta - e^{\theta} - \ln x!}$$

Beispiel 3.4.4 Für die geometrische Verteilung gilt

$$P(X = x) = e^{\theta x - b(\theta) + d(x)}$$

mit

$$b(\theta) = -\ln(1 - e^\theta)$$

und

$$d(x) = 0.$$

Zwischen θ und p besteht folgende Beziehung:

$$\theta = \ln(1 - p). \tag{3.40}$$

Dies sieht man folgendermaßen:

$$p(1 - p)^x = e^{x \ln(1-p) + \ln p}$$

Wir setzen

$$\theta = \ln(1 - p).$$

Lösen wir diese Gleichung nach p auf, so erhalten wir

$$p = 1 - e^\theta$$

Somit gilt

$$e^{x \ln(1-p) + \ln p} = e^{x\theta + \ln(1-e^\theta)}$$

Beispiel 3.4.5 Für die Pascalverteilung mit festem r gilt

$$P(X = x) = e^{\theta x - b(\theta) + d(x)}$$

mit

$$b(\theta) = -r \ln(1 - e^\theta)$$

und

$$d(x) = \ln \binom{x+r-1}{x}.$$

Zwischen θ und p besteht folgende Beziehung:

$$\theta = \ln(1 - p). \quad (3.41)$$

Dies sieht man folgendermaßen:

$$\binom{x+r-1}{x} p^r (1-p)^x = e^{x \ln(1-p) + r \ln p + \ln \binom{x+r-1}{x}}$$

Wir setzen

$$\theta = \ln(1 - p).$$

Lösen wir diese Gleichung nach p auf, so erhalten wir

$$p = 1 - e^\theta$$

Somit gilt

$$e^{x \ln(1-p) + r \ln p + \ln \binom{x+r-1}{x}} = e^{x\theta + r \ln(1-e^\theta) + \ln \binom{x+r-1}{x}}$$

Beispiel 3.4.6 Für die Exponentialverteilung gilt

$$f_X(x) = e^{\theta x - b(\theta) + d(x)}$$

mit

$$b(\theta) = -\ln \lambda$$

und

$$d(x) = 0.$$

Zwischen θ und λ besteht folgende Beziehung:

$$\theta = -\lambda. \tag{3.42}$$

Dies sieht man folgendermaßen:

$$\lambda e^{-\lambda x} = e^{-\lambda x + \ln \lambda}$$

Wir setzen

$$\theta = -\lambda.$$

Lösen wir diese Gleichung nach λ auf, so erhalten wir

$$\lambda = -\theta$$

Somit gilt

$$e^{-\lambda x + \ln \lambda} = e^{\theta x + \ln(-\theta)}$$

Beispiel 3.4.7 Wir betrachten die Normalverteilung mit festem σ^2 und setzen $\sigma = 1$

$$f_X(x) = e^{\theta x - b(\theta) + d(x)}$$

mit

$$b(\theta) = 0.5 \theta^2$$

und

$$d(x) = -0.5 (x^2 + \ln(2\pi)).$$

Zwischen θ und μ besteht folgende Beziehung:

$$\theta = \mu. \tag{3.43}$$

Dies sieht man folgendermaßen:

$$\begin{aligned} \frac{1}{\sqrt{(2\pi)}} e^{-0.5(x-\mu)^2} &= e^{-0.5x^2 + x\mu - 0.5\mu^2 - 0.5 \ln(2\pi)} = \\ &= e^{x\mu - 0.5\mu^2 - 0.5(\ln(2\pi) + x^2)} \end{aligned}$$

Wir setzen

$$\theta = \mu.$$

Lösen wir diese Gleichung nach μ auf, so erhalten wir

$$\mu = \theta$$

Somit gilt

$$e^{x\mu - 0.5\mu^2 - 0.5(\ln(2\pi) + x^2)} = e^{x\theta - 0.5\theta^2 - 0.5(\ln(2\pi) + x^2)}$$

Kapitel 4

Funktionen von Zufallsvariablen

Bei vielen statistischen Fragestellungen ist nicht eine Zufallsvariable X , sondern eine Funktion $g(X)$ dieser Zufallsvariablen von Interesse.

So betrachtet man häufig die logarithmierte Zufallsvariable.

Wir wollen uns in diesem Abschnitt damit beschäftigen, wie man die Verteilung der Funktion $g(X)$ einer Zufallsvariablen X aus der Verteilung von X herleiten kann.

Wir schauen uns zunächst diskrete Zufallsvariablen an und beginnen mit einem Beispiel:

Beispiel 4.0.8 *Wir schauen uns wieder das Teilchen an, das sich zufällig auf den ganzen Zahlen bewegt.*

Wir betrachten nun die Position X des Teilchens nach 3 Schritten.

Die folgende Tabelle gibt für jedes Ergebnis $\omega \in \Omega$ den Wert der Zufallsvariablen X an.

| ω | x |
|------------|-----------|
| <i>RRR</i> | <i>3</i> |
| <i>RRL</i> | <i>1</i> |
| <i>RLR</i> | <i>1</i> |
| <i>LRR</i> | <i>1</i> |
| <i>RLL</i> | <i>-1</i> |
| <i>LRL</i> | <i>-1</i> |
| <i>LLR</i> | <i>-1</i> |
| <i>LLL</i> | <i>-3</i> |

Somit erhalten wir folgende Wahrscheinlichkeitsfunktion von X :

| x | $P(X = x)$ |
|-----|------------|
| -3 | 0.125 |
| -1 | 0.375 |
| 1 | 0.375 |
| 3 | 0.125 |

Uns interessiert nun die Verteilung von $Y = |X|$, dem Abstand des Teilchens vom Nullpunkt.

Die Zufallsvariable Y kann die Werte 1 und 3 annehmen.

Es gilt

$$P(Y = 1) = P(X = -1) + P(X = 1) = 0.75$$

$$P(Y = 3) = P(X = -3) + P(X = 3) = 0.25$$

Wie das Beispiel zeigt, kann man die Verteilung einer Funktion $Y = g(X)$ einer diskreten Zufallsvariablen X folgendermaßen bestimmen:

$$P(Y = y) = \sum_{\{x|g(x)=y\}} P(X = x) \quad (4.1)$$

Schauen wir uns zwei weitere Beispiele an:

Beispiel 4.0.9 X sei binomialverteilt mit den Parametern n und p .

Es gilt also

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

für $x = 0, 1, \dots, n$.

Wie ist

$$Y = n - X$$

verteilt?

Y kann auch die Werte $0, 1, \dots, n$ annehmen.

Die Wahrscheinlichkeitsfunktion von Y ist dann:

$$\begin{aligned}
 P(Y = y) &= P(n - X = y) \\
 &= P(X = n - y) \\
 &= \binom{n}{n - y} p^{n-y} (1 - p)^y \\
 &= \binom{n}{y} (1 - p)^y p^{n-y}
 \end{aligned}$$

Also ist Y binomialverteilt mit den Parametern n und $1-p$.

Beispiel 4.0.10 Sei X geometrisch verteilt mit Parameter p .

Also gilt

$$P(X = x) = p(1 - p)^x$$

für $x = 0, 1, \dots$

Gesucht ist die Verteilung von

$$Y = X + 1.$$

Ist X gleich 0, so ist Y gleich 1.

Für X gegen ∞ geht auch Y gegen ∞ .

Somit kann Y die Werte $1, 2, \dots$ annehmen.

Es gilt

$$\begin{aligned}
 P(Y = y) &= P(X + 1 = y) \\
 &= P(X = y - 1) \\
 &= p(1 - p)^{y-1}
 \end{aligned}$$

für $y = 1, 2, \dots$

Die Zufallsvariable X zählt bei einem Bernoulliprozess die Anzahl der Mißerfolge vor dem ersten Erfolg, während die Zufallsvariable Y die Anzahl der Versuche bis zum ersten Erfolg zählt.

Schauen wir uns nun stetige Zufallsvariablen an.
Wir beginnen mit einem Beispiel.

Beispiel 4.0.11 Sei X exponentialverteilt mit Parameter λ . Also gilt

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{für } x > 0 \\ 0 & \text{sonst} \end{cases}$$

und

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & \text{für } x > 0 \\ 0 & \text{sonst} \end{cases}$$

Gesucht ist die Verteilung von

$$Y = \sqrt{X}.$$

Der Träger $\mathcal{T}_X = \{x | f_X(x) > 0\}$ von X ist $\{x | x \in \mathbb{R}, x > 0\}$.

Aus $x \rightarrow 0$ folgt $y \rightarrow 0$.

Aus $x \rightarrow \infty$ folgt $y \rightarrow \infty$.

Der Träger \mathcal{T}_Y von Y ist also $\{y | y \in \mathbb{R}, y > 0\}$.

Wir bestimmen zunächst die Verteilungsfunktion von Y .

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(\sqrt{X} \leq y) \\ &= P(X \leq y^2) \\ &= F_X(y^2) \end{aligned}$$

Somit ist die Verteilungsfunktion von Y gegeben durch

$$F_Y(y) = \begin{cases} 1 - e^{-\lambda y^2} & \text{für } y > 0 \\ 0 & \text{sonst} \end{cases}$$

Die Dichtefunktion von Y erhält man durch Differenzieren der Verteilungsfunktion.

$$f_Y(y) = \begin{cases} 2 \lambda y e^{-\lambda y^2} & \text{für } y > 0 \\ 0 & \text{sonst} \end{cases}$$

Man spricht auch von der **Rayleigh**-Verteilung.

Schauen wir uns das Beispiel genauer an.

Ausgangspunkt ist eine stetige Zufallsvariable X mit Träger \mathcal{T}_X , Dichtefunktion $f_X(x)$, Verteilungsfunktion $F_X(x)$ und eine Abbildung $g : \mathcal{T}_X \rightarrow \mathfrak{R}$, wobei $\frac{d}{dx} g(x)$ auf \mathcal{T}_X existiert und grösser als 0 ist. Also ist die Abbildung g auf \mathcal{T}_X streng monoton wachsend.

Gesucht ist die Dichtefunktion $f_Y(y)$ von $Y = g(X)$.

Wir bestimmen zunächst die Verteilungsfunktion $F_Y(y)$ von Y .

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(g(X) \leq y) \\ &= P(X \leq g^{-1}(y)) \\ &= F_X(g^{-1}(y)) \end{aligned}$$

Da $\frac{d}{dx} g(x)$ auf \mathcal{T}_X existiert, existiert auch $\frac{d}{dy} g^{-1}(y)$ auf $\mathcal{T}_Y = g(\mathcal{T}_X)$.

Also gilt

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} F_Y(y) \\ &= \frac{d}{dy} F_X(g^{-1}(y)) \\ &= \frac{d}{dy} g^{-1}(y) f_X(g^{-1}(y)) \end{aligned}$$

Ist die Funktion g streng monoton fallend auf \mathcal{T}_X , so erhalten wir:

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(g(X) \leq y) \\ &= P(X \geq g^{-1}(y)) \\ &= 1 - P(X \leq g^{-1}(y)) \\ &= 1 - F_X(g^{-1}(y)) \end{aligned}$$

Die Dichtefunktion erhalten wir entsprechend:

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} F_Y(y) \\ &= \frac{d}{dy} (1 - F_X(g^{-1}(y))) \\ &= -\frac{d}{dy} g^{-1}(y) f_X(g^{-1}(y)) \end{aligned}$$

Wir fassen die Ergebnisse in folgendem Satz zusammen.

Satz 4.0.1 *Sei X eine stetige Zufallsvariable mit Träger \mathcal{T}_X und Dichtefunktion $f_X(x)$.*

Sei weiterhin $g : \mathcal{T}_X \rightarrow \mathfrak{R}$ streng monoton und differenzierbar auf \mathcal{T}_X .

Dann gilt

$$f_Y(y) = \left| \frac{d}{dy} g^{-1}(y) \right| f_X(g^{-1}(y)) \quad (4.2)$$

Gehen wir also von einer stetigen Zufallsvariablen X mit Dichtefunktion $f_X(x)$ und suchen die Dichtefunktion der stetigen Zufallsvariablen $Y = g(X)$, wobei die Funktion $g : \mathcal{T}_X \rightarrow \mathfrak{R}$ streng monoton und differenzierbar auf \mathcal{T}_X ist, dann müssen wir folgendermaßen vorgehen:

1. Bestimme den Träger $\mathcal{T}_Y = g(\mathcal{T}_X)$ von Y .

2. Bestimme

$$x = g^{-1}(y)$$

3. Bestimme

$$\frac{d}{dy} g^{-1}(y)$$

4. Berechne

$$f_Y(y) = \left| \frac{d}{dy} g^{-1}(y) \right| f_X(g^{-1}(y))$$

Schauen wir uns ein Beispiel an:

Beispiel 4.0.12 Sei X exponentialverteilt mit Parameter λ .

Es gilt also

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{für } x > 0 \\ 0 & \text{sonst} \end{cases}$$

Gesucht ist die Verteilung von

$$Y = 1 - e^{-\lambda X}.$$

Wir führen die 4 Schritte durch.

1. Wir bestimmen den Träger von Y .

Der Träger \mathcal{T}_X von X ist $\{x | x \in \mathbb{R}, x > 0\}$.

Aus $x \rightarrow 0$ folgt $y \rightarrow 0$.

Aus $x \rightarrow \infty$ folgt $y \rightarrow 1$.

Der Träger \mathcal{T}_Y von Y ist also $(0, 1)$.

2. Wir bestimmen

$$x = g^{-1}(y)$$

Aus

$$y = g(x) = 1 - e^{-\lambda x}$$

folgt

$$x = g^{-1}(y) = -\frac{1}{\lambda} \ln(1 - y)$$

3. Wir bestimmen

$$\frac{d}{dy} g^{-1}(y)$$

Es gilt

$$\frac{d}{dy} g^{-1}(y) = \frac{1}{\lambda(1 - y)}$$

4. Wir berechnen

$$\begin{aligned} f_Y(y) &= \left| \frac{d}{dy} g^{-1}(y) \right| f_X(g^{-1}(y)) \\ &= \frac{1}{\lambda(1 - y)} f_X\left(\frac{\ln(1 - y)}{-\lambda}\right) \end{aligned}$$

Also gilt

$$\begin{aligned} f_Y(y) &= \frac{1}{\lambda(1-y)} \lambda e^{-\lambda(-\frac{1}{\lambda} \ln(1-y))} \\ &= 1 \end{aligned}$$

Die Dichtefunktion von Y ist also gegeben durch

$$f_Y(y) = \begin{cases} 1 & \text{für } 0 < y < 1 \\ 0 & \text{sonst} \end{cases}$$

Y ist also gleichverteilt auf $(0,1)$.

Daß Y auf $(0,1)$ gleichverteilt ist, ist kein Zufall. Wir haben nämlich die Zufallsvariable X mit ihrer eigenen Verteilungsfunktion $F_X(x)$ transformiert. Allgemein gilt

Satz 4.0.2 Gegeben sei eine stetige Zufallsvariable X mit Verteilungsfunktion $F_X(x)$.

Sei

$$U = F_X(X)$$

Dann ist die Zufallsvariable U gleichverteilt auf $(0,1)$.

Beweis:

Wir beweisen den Satz nur für den Fall einer streng monoton wachsenden Verteilungsfunktion $F_X(x)$.

Aus $x \rightarrow -\infty$ folgt $u \rightarrow 0$.

Aus $x \rightarrow \infty$ folgt $u \rightarrow 1$.

Also ist der Träger von U gleich $(0,1)$.

Für die Verteilungsfunktion von U gilt:

für $u \leq 0$:

$$F_U(u) = 0$$

für $0 < u < 1$:

$$\begin{aligned} F_U(u) &= P(U \leq u) \\ &= P(F_X(X) \leq u) \\ &= P(X \leq F_X^{-1}(u)) \\ &= F_X(F_X^{-1}(u)) \\ &= u \end{aligned}$$

für $u \geq 1$:

$$F_U(u) = 1$$

Dies ist aber die Verteilungsfunktion einer auf $(0, 1)$ gleichverteilten Zufallsvariablen U .

Es gilt auch die Umkehrung.

Satz 4.0.3 Sei U eine auf $(0,1)$ -gleichverteilte Zufallsvariable und G eine auf $(0,1)$ monoton wachsende stetige Funktion.

Dann besitzt die Zufallsvariable $Y = G(U)$ die Verteilungsfunktion G^{-1} .

Der Satz zeigt, wie man sich eine Zufallszahl aus einer stetigen Verteilung mit Verteilungsfunktion $F_X(x)$ erzeugen kann:

Man erzeugt eine auf $(0,1)$ gleichverteilte Zufallszahl u und bildet dann $x = F_X^{-1}(u)$.

Beispiel 4.0.13 Bei der Modellierung eines Bedienungsprozesses sei die Bedienungszeit X exponentialverteilt mit Parameter λ .

Wir wollen eine Bedienungszeit simulieren.

Hierzu erzeugen wir eine gleichverteilte Zufallszahl u .

Die Verteilungsfunktion von X ist gegeben durch

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & \text{für } x > 0 \\ 0 & \text{sonst} \end{cases}$$

Wir setzen

$$1 - e^{-\lambda x} = u,$$

lösen diese Gleichung nach x auf und erhalten

$$x = -\frac{1}{\lambda} \ln(1 - u)$$

Das nächste Beispiel zeigt eine wichtige Eigenschaft der Normalverteilung.

Beispiel 4.0.14 Sei X normalverteilt mit den Parametern μ und σ^2 .

Es gilt also:

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

Der Träger von X ist die Menge der reellen Zahlen.

Gesucht ist die Dichtefunktion von $Y = a + bX$.

Der Träger von Y ist offensichtlich auch die Menge der reellen Zahlen.

Wir bilden

$$x = g^{-1}(y) = \frac{y - a}{b}.$$

Es gilt

$$\frac{d}{dy} g^{-1}(y) = \frac{1}{b}.$$

Somit lautet die Dichtefunktion von Y :

$$\begin{aligned} f_Y(y) &= \left| \frac{d}{dy} g^{-1}(y) \right| f_X(g^{-1}(y)) \\ &= \frac{1}{|b|\sqrt{2\pi\sigma^2}} e^{-\frac{(y-a-b\mu)^2}{2b^2\sigma^2}} \end{aligned}$$

Also ist Y normalverteilt mit den Parametern $a + b\mu$ und $(b\sigma)^2$.

Setzen wir

$$a = -\frac{\mu}{\sigma}$$

und

$$b = \frac{1}{\sigma},$$

so ist

$$Y = \frac{X - \mu}{\sigma}$$

normalverteilt mit den Parametern 0 und 1.

Man spricht in diesem Fall von Standardisierung. Die zugehörige Normalverteilung heißt Standardnormalverteilung. Die Dichtefunktion der Standardnormalverteilung wird mit $\phi(z)$ und die Verteilungsfunktion mit $\Phi(z)$ bezeichnet.

Es gilt also

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad (4.3)$$

und

$$\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \quad (4.4)$$

Man kann jede Wahrscheinlichkeitsaussage bei der Normalverteilung auf eine Wahrscheinlichkeitsaussage bei der Standardnormalverteilung zurückführen. Ist X normalverteilt mit den Parametern μ und σ^2 , so gilt:

$$F_X(x) = \Phi\left(\frac{x - \mu}{\sigma}\right) \quad (4.5)$$

Dies sieht man folgendermaßen:

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) \\ &= P\left(Z \leq \frac{x - \mu}{\sigma}\right) \\ &= \Phi\left(\frac{x - \mu}{\sigma}\right) \end{aligned}$$

Man muss also nur die Standardnormalverteilung tabellieren.

Beispiel 4.0.15 Sei X normalverteilt mit den Parametern μ und σ^2 .

Es gilt also:

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

Der Träger von X ist die Menge der reellen Zahlen.

Gesucht ist die Dichtefunktion von $Y = e^X$.

Aus $x \rightarrow -\infty$ folgt $y \rightarrow 0$.

Aus $x \rightarrow \infty$ folgt $y \rightarrow \infty$.

Der Träger von Y ist die Menge der positiven reellen Zahlen.

Wir bilden

$$x = g^{-1}(y) = \ln y.$$

Es gilt

$$\frac{d}{dy} g^{-1}(y) = \frac{1}{y}.$$

Somit lautet die Dichtefunktion von Y :

$$\begin{aligned} f_Y(y) &= \left| \frac{d}{dy} g^{-1}(y) \right| f_X(g^{-1}(y)) \\ &= \frac{1}{y\sqrt{2\pi\sigma^2}} e^{-\frac{(\ln y - \mu)^2}{2\sigma^2}} \end{aligned}$$

Man sagt, daß Y lognormalverteilt ist.

Die Lognormalverteilung spielt eine wichtige Rolle in der Zuverlässigkeitstheorie.

Wir betrachten nun noch ein Beispiel:

Beispiel 4.0.16 Sei Z standardnormalverteilt.

Gesucht ist die Dichtefunktion von $Y = Z^2$.

Offensichtlich ist die Funktion $y = z^2$ auf dem Träger von Z nicht monoton.

Wir können die Dichtefunktion von Y aber über die Verteilungsfunktion erhalten:

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(Z^2 \leq y) \\ &= P(\sqrt{Z^2} \leq \sqrt{y}) \\ &= P(|Z| \leq \sqrt{y}) \\ &= P(-\sqrt{y} \leq Z \leq \sqrt{y}) \\ &= \Phi(\sqrt{y}) - \Phi(-\sqrt{y}) \end{aligned}$$

Also gilt

$$\begin{aligned}
 f_Y(y) &= \frac{d}{dy} F_Y(y) \\
 &= \frac{d}{dy} (\Phi(\sqrt{y}) - \Phi(-\sqrt{y})) \\
 &= \frac{1}{2\sqrt{y}} \phi(\sqrt{y}) + \frac{1}{2\sqrt{y}} \phi(-\sqrt{y}) \\
 &= \frac{2}{2\sqrt{y}} \phi(\sqrt{y}) \\
 &= \frac{1}{\sqrt{2\pi y}} e^{-y/2} \\
 &= \sqrt{\frac{1}{2}} \frac{1}{\Gamma 0.5} y^{-0.5} e^{-0.5 y}
 \end{aligned}$$

Das ist die Dichtefunktion einer Chiquadratverteilung mit einem Freiheitsgrad.

Das Quadrat einer standardnormalverteilten Zufallsvariablen ist also chiquadratverteilt mit einem Freiheitsgrad.

Kapitel 5

Parameter univariater Verteilungen

5.1 Erwartungswerte

Beispiel 5.1.1 *Tversky und Kahnemann fragten Personen, welche der beiden folgenden Alternativen sie vorzögen.*

Alternative A

Man muß 750 DM zahlen.

Alternative B

Eine faire Münze wird zweimal hintereinander geworfen.

Fällt beide Male Kopf, so muß man nichts zahlen, ansonsten muß man 1000 DM zahlen.

Wir wollen zunächst untersuchen, ob eine der beiden Situationen günstiger ist.

Die Konsequenz von Alternative A ist klar. Wir müssen auf jeden Fall 750 DM bezahlen.

Bei Alternative B müssen wir mit Wahrscheinlichkeit 0.25 nicht und mit Wahrscheinlichkeit 0.75 1000 DM bezahlen.

Welche Auszahlung würden wir in diesem Fall erwarten.?

Wir gewichten die Werte mit ihren Wahrscheinlichkeiten und summieren auf:

$$0 \cdot 0.25 + 1000 \cdot 0.75 = 750$$

Wir erwarten also eine Auszahlung von 750 DM.

Bei Alternative A haben wir eine sichere Auszahlung von 750 DM, bei Alternative B erwarten wir eine Auszahlung von 750 DM.

Das Beispiel legt folgende Definition nahe:

Definition 5.1.1 Sei X eine Zufallsvariable mit Wahrscheinlichkeitsfunktion $P(X = x)$ bzw. Dichtefunktion $f_X(x)$.

Der Erwartungswert $E(X)$ von X ist definiert durch

$$E(X) = \sum_i x_i P(X = x_i) \quad (5.1)$$

wenn X diskret ist, und durch

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx \quad (5.2)$$

wenn X stetig ist.

Beispiel 5.1.2 Im letzten Kapitel haben wir ein Teilchen betrachtet, das sich zufällig auf den ganzen Zahlen bewegt, wobei es im Nullpunkt startet.

Sei X die Position des Teilchens nach 3 Schritten. Die Wahrscheinlichkeitsverteilung von X ist in der folgenden Tabelle zu finden:

| x | $P(X=x)$ |
|-----|----------|
| -3 | 0.125 |
| -1 | 0.375 |
| 1 | 0.375 |
| 3 | 0.125 |

Welche Position erwarten wir nach 3 Schritten?

Es gilt:

$$E(X) = (-3) \cdot 0.125 + (-1) \cdot 0.375 + 1 \cdot 0.375 + 3 \cdot 0.125 = 0$$

Wir erwarten das Teilchen nach 3 Schritten im Nullpunkt.

Bemerkenswert ist, daß sich das Teilchen nach 3 Schritten niemals im Nullpunkt befindet. Der Erwartungswert ist also ein Wert, der gar nicht angenommen werden muß.

Schauen wir uns den Erwartungswert einiger spezieller Wahrscheinlichkeitsverteilungen an:

Beispiel 5.1.3 Die Zufallsvariable X sei gleichverteilt.

Also gilt

$$P(X = x) = \frac{1}{N} \quad \text{für } x = 1, 2, \dots, N$$

Dann gilt

$$E(X) = \frac{N + 1}{2} \quad (5.3)$$

Dies sieht man folgendermaßen:

$$\begin{aligned} E(X) &= \sum_{x=1}^N x \frac{1}{N} \\ &= \frac{1}{N} \sum_{x=1}^N x \\ &= \frac{1}{N} \frac{N(N+1)}{2} \\ &= \frac{N+1}{2} \end{aligned}$$

Beispiel 5.1.4 X sei bernoulliverteilt mit Parameter p .

Es gilt also

$$P(X = x) = p^x (1-p)^{1-x} \quad \text{für } x = 0, 1.$$

Dann gilt

$$E(X) = p \quad (5.4)$$

Dies sieht man folgendermaßen:

$$\begin{aligned} E(X) &= 0 \cdot (1-p) + 1 \cdot p \\ &= p \end{aligned}$$

Beispiel 5.1.5 X sei binomialverteilt mit den Parametern n und p .

Es gilt also

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

für $x = 0, 1, \dots, n$.

Dann gilt

$$E(X) = np \tag{5.5}$$

Dies sieht man folgendermaßen:

$$\begin{aligned} E(X) &= \sum_{x=0}^n x \binom{n}{x} p^x (1-p)^{n-x} \\ &= \sum_{x=1}^n \frac{x \cdot n!}{x!(n-x)!} p^x (1-p)^{n-x} \\ &= np \sum_{x=1}^n \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} (1-p)^{n-x} \\ &= np \sum_{x=1}^n \binom{n-1}{x-1} p^{x-1} (1-p)^{n-x} \\ &= np \sum_{y=0}^{n-1} \binom{n-1}{y} p^y (1-p)^{n-1-y} \\ &= np \end{aligned}$$

Beispiel 5.1.6 Sei X poissonverteilt mit Parameter λ .

Es gilt also

$$P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

für $x = 0, 1, \dots$

Dann gilt

$$E(X) = \lambda \tag{5.6}$$

Dies sieht man folgendermaßen:

$$\begin{aligned} E(X) &= \sum_{x=0}^{\infty} x P(X = x) \\ &= \sum_{x=0}^{\infty} x \frac{\lambda^x}{x!} e^{-\lambda} \\ &= \sum_{x=1}^{\infty} x \frac{\lambda^x}{x!} e^{-\lambda} \\ &= \sum_{x=1}^{\infty} \frac{\lambda^x}{(x-1)!} e^{-\lambda} \\ &= \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} \\ &= \lambda e^{-\lambda} e^{\lambda} \\ &= \lambda \end{aligned}$$

da gilt

$$\begin{aligned} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} &= \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} \\ &= e^{\lambda} \end{aligned}$$

Der folgende Satz erleichtert die Bestimmung des Erwartungswertes in bestimmten Situationen.

Satz 5.1.1 *X sei eine diskrete Zufallsvariable, die die Werte $0, 1, 2, \dots$ annehmen kann.*

Dann gilt

$$E(X) = \sum_{y=1}^{\infty} P(X \geq y) \quad (5.7)$$

Beweis:

Sei $p_i = P(X = i)$.

Dann gilt

$$\begin{aligned} E(X) &= 1 \cdot p_1 + 2 \cdot p_2 + \cdots + n \cdot p_n + \cdots = \\ &= p_1 \\ &+ p_2 + p_2 \\ &+ p_3 + p_3 + p_3 \\ &\cdots \quad \cdots \quad \cdots \\ &+ p_n + p_n + p_n + \cdots + p_n + \cdots \\ &\quad \cdots \quad \cdots \quad \cdots \quad \cdots \end{aligned}$$

In der 1. Spalte steht $P(X \geq 1)$, in der 2. Spalte $P(X \geq 2)$,...

Also gilt

$$E(X) = \sum_{y=1}^{\infty} P(X \geq y)$$

Betrachten wir also den Erwartungswert der geometrischen Verteilung.

Satz 5.1.2 *Sei X geometrisch verteilt mit Parameter p .*

Also gilt

$$P(X = x) = p(1 - p)^x$$

für $x = 0, 1, \dots$

Dann gilt

$$E(X) = \frac{1 - p}{p}.$$

Beweis:

Für die geometrische Verteilung gilt:

$$P(X \geq x) = (1 - p)^x$$

Somit lautet der Erwartungswert

$$\begin{aligned} E(X) &= \sum_{x=1}^{\infty} (1 - p)^x \\ &= \sum_{x=0}^{\infty} (1 - p)^x - 1 \\ &= \frac{1}{1 - (1 - p)} - 1 \\ &= \frac{1 - p}{p} \end{aligned}$$

Schauen wir uns einige stetige Verteilungen an.

Beispiel 5.1.7 X sei gleichverteilt auf (a, b) .

Es gilt also

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{für } a < x < b \\ 0 & \text{sonst} \end{cases}$$

Dann gilt

$$E(X) = \frac{a+b}{2} \tag{5.8}$$

Dies sieht man folgendermaßen:

$$\begin{aligned} E(X) &= \int_a^b x \frac{1}{b-a} dx \\ &= \frac{1}{b-a} \int_a^b x dx \\ &= \frac{1}{b-a} \left[\frac{x^2}{2} \right]_a^b \\ &= \frac{1}{b-a} \frac{b^2 - a^2}{2} \\ &= \frac{1}{b-a} \frac{(b+a)(b-a)}{2} \\ &= \frac{a+b}{2} \end{aligned}$$

Den Erwartungswert der Standardnormalverteilung erhalten wir folgendermaßen:

Beispiel 5.1.8 Z sei standardnormalverteilt.

Es gilt also

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

Dann gilt

$$E(Z) = 0 \tag{5.9}$$

Beweis:

$$\begin{aligned} E(Z) &= \int_{-\infty}^{\infty} z\phi(z)dz \\ &= \int_{-\infty}^0 z\phi(z)dz + \int_0^{\infty} z\phi(z)dz \\ &= \int_{-\infty}^0 z\phi(z)dz - \int_{-\infty}^0 z\phi(-z)dz \\ &= \int_{-\infty}^0 z\phi(z)dz - \int_{-\infty}^0 z\phi(z)dz \\ &= 0 \end{aligned}$$

Beispiel 5.1.9 Sei X betaverteilt mit den Parametern a und b .

Es gilt also

$$f_X(x) = \begin{cases} \frac{1}{B(a,b)} x^{a-1} (1-x)^{b-1} & \text{für } 0 < x < 1, a, b > 0 \\ 0 & \text{sonst} \end{cases}$$

Dann gilt

$$E(X) = \frac{a}{a+b} \quad (5.10)$$

Dies sieht man folgendermaßen:

$$\begin{aligned} E(X) &= \int_0^1 x \frac{1}{B(a,b)} x^{a-1} (1-x)^{b-1} dx \\ &= \int_0^1 \frac{1}{B(a,b)} x^a (1-x)^{b-1} dx \\ &= \frac{B(a+1,b)}{B(a,b)} \int_0^1 \frac{1}{B(a+1,b)} x^a (1-x)^{b-1} dx \\ &= \frac{B(a+1,b)}{B(a,b)} \\ &= \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \\ &= \frac{a\Gamma(a)\Gamma(a+b)}{(a+b)\Gamma(a+b)\Gamma(a)} \\ &= \frac{a}{a+b} \end{aligned}$$

wegen

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

und

$$\Gamma(a+1) = a\Gamma(a)$$

Beispiel 5.1.10 Sei X gammaverteilt mit den Parametern r und λ .

Es gilt also

$$f_X(x) = \begin{cases} \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} & \text{für } x > 0 \\ 0 & \text{sonst} \end{cases}$$

Dann gilt

$$E(X) = \frac{r}{\lambda} \quad (5.11)$$

Dies sieht man folgendermaßen:

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f_X(x) dx \\ &= \int_0^{\infty} x \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} dx \\ &= \frac{1}{\lambda} \int_0^{\infty} \frac{\lambda^{r+1}}{\Gamma(r)} x^r e^{-\lambda x} dx \\ &= \frac{\Gamma(r+1)}{\lambda \Gamma(r)} \int_0^{\infty} \frac{\lambda^{r+1}}{\Gamma(r+1)} x^r e^{-\lambda x} dx \\ &= \frac{r \Gamma(r)}{\lambda \Gamma(r)} \\ &= \frac{r}{\lambda} \end{aligned}$$

Da die Exponentialverteilung mit Parameter λ eine Gammaverteilung mit Parametern $r = 1$ und λ ist, ist der Erwartungswert der Exponentialverteilung gleich $\frac{1}{\lambda}$.

Der Erwartungswert einer Zufallsvariablen muß nicht existieren.

Satz 5.1.3 *Die Zufallsvariable X besitze die Dichtefunktion*

$$f_X(x) = \begin{cases} \frac{1}{x^2} & \text{für } x > 1 \\ 0 & \text{sonst} \end{cases}$$

Dann existiert der Erwartungswert von X nicht.

Beweis:

$$\int_1^{\infty} x \frac{1}{x^2} dx = [\ln x]_1^{\infty} \rightarrow \infty$$

Auch bei der Cauchyverteilung existiert der Erwartungswert nicht.

Angenommen, wir kennen den Erwartungswert $E(X)$ einer nichtnegativen Zufallsvariablen X . Wie groß kann dann die Wahrscheinlichkeit $P(X \geq a)$ höchstens sein, wobei $a > 0$ gilt.

Die Antwort liefert der folgende Satz. Bevor wir diesen aber angeben, schauen wir uns eine heuristische Herleitung an, die bei Pitman zu finden ist.

Wir betrachten den Fall $E(X) = 3$ und $a = 100$. Wie sollen wir die Wahrscheinlichkeitsmasse verteilen, wenn $P(X \geq 100)$ möglichst gross sein soll?

Auf jeden Fall sollte man Wahrscheinlichkeitsmasse auf die Null legen, damit man noch für den Rand Wahrscheinlichkeitsmasse übrig hat.

Die restliche Wahrscheinlichkeitsmasse sollte man auf 100 legen.

Legt man nämlich Wahrscheinlichkeitsmasse auf einen Punkt grösser als 100, so muss man mit einer geringeren Masse als in 100 dafür bezahlen, wenn der Erwartungswert gleich 3 sein soll.

Damit der Erwartungswert 3 ist, muss 100 die Wahrscheinlichkeitswert $3/100$ erhalten.

Dies ist die Aussage des folgenden Satzes von Markow.

Satz 5.1.4 *Sei Y eine nichtnegative Zufallsvariable mit Erwartungswert $E(Y)$ existiert.*

Dann gilt

$$P(Y \geq a) \leq \frac{E(Y)}{a} \quad (5.12)$$

Beweis:

Wir zeigen den stetigen Fall

$$\begin{aligned} E(Y) &= \int_0^{\infty} y f_Y(y) dy \\ &= \int_0^a y f_Y(y) dy + \int_a^{\infty} y f_Y(y) dy \\ &\geq \int_a^{\infty} y f_Y(y) dy \\ &\geq a \int_a^{\infty} f_Y(y) dy \\ &\geq a P(Y \geq a) \end{aligned}$$

5.2 Erwartungswerte von Funktionen von Zufallsvariablen

Wir betrachten nun noch einmal das Teilchen, das sich auf den ganzen Zahlen bewegt.

Im letzten Kapitel haben wir die Verteilung von $Y = |X|$ hergeleitet, wobei X die Position des Teilchens ist.

Die Zufallsvariable Y ist der Abstand des Teilchens vom Nullpunkt.

Wir haben gesehen, daß gilt

| y | P(Y=y) |
|---|--------|
| 1 | 0.75 |
| 3 | 0.25 |

Somit gilt

$$E(Y) = 1 \cdot 0.75 + 3 \cdot 0.25 = 1.5$$

Wir können $E(|X|)$ auch bestimmen, ohne die Verteilung von $Y = |X|$ herzuleiten, da gilt

$$\begin{aligned} E(Y) &= 1 \cdot 0.75 + 3 \cdot 0.25 \\ &= 1 \cdot (0.375 + 0.375) + 3 \cdot (0.125 + 0.125) \\ &= 1 \cdot 0.375 + 1 \cdot 0.375 + 3 \cdot 0.125 + 3 \cdot 0.125 \\ &= |-3| \cdot 0.125 + |-1| \cdot 0.375 + |1| \cdot 0.375 + |3| \cdot 0.125 \end{aligned}$$

Es gilt der folgende Satz:

Satz 5.2.1 *Sei X eine diskrete Zufallsvariable mit Wahrscheinlichkeitsfunktion $P(X = x)$ und $Y = g(X)$ eine Funktion von X .*

Dann gilt

$$E(Y) = \sum_x g(x) P(X = x) \quad (5.13)$$

Beweis:

Es gilt:

$$\begin{aligned}
 E(Y) &= \sum_y y P(Y = y) \\
 &= \sum_y y P(g(X) = y) \\
 &= \sum_y y \sum_{x|g(x)=y} P(X = x) \\
 &= \sum_y \sum_{\{x|g(x)=y\}} y P(X = x) \\
 &= \sum_x g(x) P(X = x)
 \end{aligned}$$

Für eine stetige Zufallsvariable X mit Dichtefunktion $f_X(x)$ gilt:

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x) dx \quad (5.14)$$

Schauen wir uns die Gleichverteilung an:

Beispiel 5.2.1 X sei gleichverteilt auf (a, b) .

Es gilt also

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{für } a < x < b \\ 0 & \text{sonst} \end{cases}$$

Dann gilt

$$\begin{aligned}
 E(X^2) &= \int_a^b x^2 \frac{1}{b-a} dx \\
 &= \frac{1}{b-a} \int_a^b x^2 dx \\
 &= \frac{1}{b-a} \left[\frac{x^3}{3} \right]_a^b
 \end{aligned}$$

$$\begin{aligned}
&= \frac{b^3 - a^3}{3(b - a)} \\
&= \frac{a^2 + ab + b^2}{3}
\end{aligned}$$

Beispiel 5.2.2 Sei X gammaverteilt mit den Parametern r und λ .

Es gilt also

$$f_X(x) = \begin{cases} \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} & \text{für } x > 0 \\ 0 & \text{sonst} \end{cases}$$

Dann gilt

$$\begin{aligned}
E(X^2) &= \int_{-\infty}^{\infty} x^2 f_X(x) dx \\
&= \int_0^{\infty} x^2 \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} dx \\
&= \int_0^{\infty} \frac{\lambda^r}{\Gamma(r)} x^{r+1} e^{-\lambda x} dx = \\
&= \frac{\Gamma(r+2)}{\lambda^2 \Gamma(r)} \int_0^{\infty} \frac{\lambda^{r+2}}{\Gamma(r+2)} x^{r+1} e^{-\lambda x} dx \\
&= \frac{(r+1)r \Gamma(r)}{\lambda^2 \Gamma(r)} \\
&= \frac{r^2 + r}{\lambda^2}
\end{aligned}$$

Beispiel 5.2.3 Sei X betaverteilt mit den Parametern a und b .

Es gilt also

$$f_X(x) = \begin{cases} \frac{1}{B(a,b)} x^{a-1} (1-x)^{b-1} & \text{für } 0 < x < 1, a, b > 0 \\ 0 & \text{sonst} \end{cases}$$

Dann ist

$$\begin{aligned} E(X^2) &= \int_0^1 x^2 \frac{1}{B(a,b)} x^{a-1} (1-x)^{b-1} dx \\ &= \int_0^1 \frac{1}{B(a,b)} x^{a+1} (1-x)^{b-1} dx \\ &= \frac{B(a+2,b)}{B(a,b)} \int_0^1 \frac{1}{B(a+2,b)} x^{a+1} (1-x)^{b-1} dx \\ &= \frac{B(a+2,b)}{B(a,b)} \\ &= \frac{\Gamma(a+2)\Gamma(b)}{\Gamma(a+b+2)} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \\ &= \frac{(a+1)a\Gamma(a)\Gamma(a+b)}{(a+b+1)(a+b)\Gamma(a+b)\Gamma(a)} \\ &= \frac{(a+1)a}{(a+b+1)(a+b)} \end{aligned}$$

wegen

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

und

$$\Gamma(a+1) = a\Gamma(a)$$

Der folgende Satz gibt einige Eigenschaften des Erwartungswertes an.

Satz 5.2.2 *X sei eine Zufallsvariable mit Erwartungswert $E(X)$ und Träger \mathcal{X} . Außerdem seien $g : \mathcal{X} \rightarrow \mathfrak{R}$ und $h : \mathcal{X} \rightarrow \mathfrak{R}$ Funktionen.*

Dann gilt

1. $E(c) = c$ für jedes reelle c
2. $E(cg(X)) = cE(g(X))$
3. $E(g(X) + h(X)) = E(g(X)) + E(h(X))$

Beweis:

Wir zeigen den diskreten Fall:

1.

$$\begin{aligned} E(c) &= \sum_x c P(X = x) = \\ &= c \sum_x P(X = x) \\ &= c \end{aligned}$$

2.

$$\begin{aligned} E(cg(X)) &= \sum_x cg(x) P(X = x) \\ &= c \sum_x g(x) P(X = x) \\ &= cE(g(X)) \end{aligned}$$

3.

$$\begin{aligned} E(g(X) + E(h(X))) &= \sum_x (g(x) + h(x)) P(X = x) \\ &= \sum_x (g(x)P(X = x) + h(x)P(X = x)) \\ &= \sum_x g(x)P(X = x) + \sum_x h(x)P(X = x) \\ &= E(g(X)) + E(h(X)) \end{aligned}$$

5.2. ERWARTUNGSWERTE VON FUNKTIONEN VON ZUFALLSVARIABLEN 141

Aus dem Satz folgt folgende wichtige Eigenschaft des Erwartungswerts:

$$E(a + bX) = a + bE(X) \quad (5.15)$$

Mit

$$g(x) = a + bX$$

gilt also

$$E(g(X)) = g(E(X)).$$

Dies gilt aber nicht immer, wie das folgende Beispiel zeigt:

Bei der Gleichverteilung auf (a, b) gilt

$$E(X^2) = \frac{a^2 + ab + b^2}{3}$$

und

$$E(X)^2 = \frac{a^2 + 2ab + b^2}{4}$$

Nun gilt

$$\begin{aligned} E(X^2) - E(X)^2 &= \frac{a^2 + ab + b^2}{3} - \frac{a^2 + 2ab + b^2}{4} \\ &= \frac{(b - a)^2}{12} \end{aligned}$$

Also ist bei der Gleichverteilung

$$E(X^2) > E(X)^2.$$

Der folgende Satz gibt den Erwartungswert der Normalverteilung an.

Satz 5.2.3 *Sei X normalverteilt mit den Parametern μ und σ^2 .
Dann gilt*

$$E(X) = \mu \tag{5.16}$$

Beweis:

Ist Z standardnormalverteilt, so ist

$$X = \mu + Z\sigma$$

normalverteilt mit den Parametern μ und σ^2 .

Aus

$$E(Z) = 0$$

folgt

$$\begin{aligned} E(X) &= E(\mu + Z\sigma) \\ &= \mu + \sigma E(Z) \\ &= \mu \end{aligned}$$

Der Parameter μ ist also der Erwartungswert der Normalverteilung. Wir betrachten noch folgende Funktion einer Zufallsvariablen X :

$$g(X) = X - E(X)$$

Man spricht in diesem Fall von Zentrierung der Zufallsvariablen X . Der Erwartungswert einer zentrierten Zufallsvariablen ist 0:

$$E(X - E(X)) = E(X) - E(E(X)) = E(X) - E(X) = 0$$

5.3 Die Varianz

Der Erwartungswert einer Zufallsvariablen X ist ein Maß für die Lage von X .

Neben der Lage der Verteilung einer Zufallsvariablen X ist auch die Streuung von Interesse.

Die folgende Definition gibt ein Maß für die Streuung an.

Definition 5.3.1 Sei X eine Zufallsvariable.

Die Varianz $Var(X)$ von X ist definiert durch

$$Var(X) = E((X - E(X))^2)$$

Wir berechnen die Varianz im diskreten Fall also durch

$$Var(X) = \sum_x (x - E(X))^2 P(X = x) \quad (5.17)$$

und im stetigen Fall durch

$$Var(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f_X(x) dx \quad (5.18)$$

Wir wählen für die Varianz oft die Abkürzung σ_X^2 .

Die Varianz besitzt nicht die gleiche Maßeinheit wie X , die Standardabweichung $\sqrt{Var(X)}$ hingegen doch.

Wir kürzen im folgenden die Standardabweichung mit σ_X ab.

Beispiel 5.3.1 Im Beispiel 5.1.1 haben wir zwei Alternativen betrachtet.

Bei Alternative A müssen wir in jedem Fall 750 DM zahlen.

Wir bezeichnen die zugehörige Zufallsvariable mit X .

Es gilt

$$P(X = 750) = 1$$

Bei Alternative B wird eine faire Münze zweimal hintereinander geworfen. Fällt beide Male Kopf, so muß man nichts zahlen, ansonsten muß man 1000 DM zahlen.

Wir bezeichnen die zugehörige Zufallsvariable mit Y .

Es gilt

$$P(Y = 0) = 0.25$$

$$P(Y = 1000) = 0.75$$

Der Erwartungswert beträgt in beiden Fällen 750 DM.

Die beiden Zufallsvariablen unterscheiden sich aber hinsichtlich ihrer Streuung.

Offensichtlich gilt

$$\text{Var}(X) = 0$$

Für Y gilt

$$\begin{aligned} \text{Var}(Y) &= \sum_x (x - E(X))^2 P(X = x) \\ &= (0 - 750)^2 \cdot 0.25 + (1000 - 750)^2 \cdot 0.75 \\ &= 562500 \cdot 0.25 + 62500 \cdot 0.75 \\ &= 187500 \end{aligned}$$

Die zweite Alternative hat die größere Varianz.

Der folgende Satz zeigt, wie man die Varianz einfach berechnen kann.

Satz 5.3.1 Sei X eine Zufallsvariable mit Varianz $\text{Var}(X)$. Dann gilt

$$\text{Var}(X) = E(X^2) - E(X)^2 \quad (5.19)$$

Beweis:

$$\begin{aligned} \text{Var}(X) &= E((X - E(X))^2) \\ &= E(X^2 - 2XE(X) + E(X)^2) \\ &= E(X^2) - E(2XE(X)) + E(E(X)^2) \\ &= E(X^2) - 2E(X)E(X) + E(X)^2 \\ &= E(X^2) - E(X)^2 \end{aligned}$$

Beispiel 5.3.2 Wir berechnen die Varianz der Zufallsvariablen Y aus Beispiel 5.3.1 mit Hilfe von (5.19).

Es gilt

$$\begin{aligned} E(Y^2) &= \sum_y y^2 P(Y = y) \\ &= 0^2 \cdot 0.25 + 1000^2 \cdot 0.75 \\ &= 750000 \end{aligned}$$

Also gilt

$$\begin{aligned} \text{Var}(Y) &= E(Y^2) - E(Y)^2 \\ &= 750000 - 750^2 \\ &= 187500 \end{aligned}$$

Schauen wir nun die Varianzen einiger Verteilungsmodelle an.

Beispiel 5.3.3 Für die Gleichverteilung auf (a, b) gilt:

$$\text{Var}(X) = \frac{(b-a)^2}{12} \quad (5.20)$$

Wegen

$$\begin{aligned} E(X) &= \frac{a+b}{2} \\ E(X^2) &= \frac{a^2 + ab + b^2}{3} \end{aligned}$$

gilt

$$\begin{aligned} \text{Var}(X) &= E(X^2) - E(X)^2 \\ &= \frac{a^2 + ab + b^2}{3} - \frac{a^2 + 2ab + b^2}{4} \\ &= \frac{(b-a)^2}{12} \end{aligned}$$

Daß die Varianz eine Maßzahl für die Streuung ist, kann man am Beispiel der Gleichverteilung sehr schön sehen.

Je weiter a und b entfernt sind, um so weiter entfernt vom Mittelwert $\frac{a+b}{2}$ sind die Werte der Zufallsvariablen und um so größer ist die Varianz.

Beispiel 5.3.4 Für die Gammaverteilung mit den Parametern r und λ gilt:

$$\operatorname{Var}(X) = \frac{r}{\lambda^2} \quad (5.21)$$

Wegen

$$E(X) = \frac{r}{\lambda}$$

$$E(X^2) = \frac{r^2 + r}{\lambda^2}$$

gilt

$$\begin{aligned} \operatorname{Var}(X) &= E(X^2) - E(X)^2 \\ &= \frac{r^2 + r}{\lambda^2} - \frac{r^2}{\lambda^2} \\ &= \frac{r}{\lambda^2} \end{aligned}$$

Beispiel 5.3.5 Für die Betaverteilung mit den Parametern a und b gilt:

$$\operatorname{Var}(X) = \frac{ab}{(a+b+1)(a+b)^2} \quad (5.22)$$

Wegen

$$E(X) = \frac{a}{a+b}$$

$$E(X^2) = \frac{(a+1)a}{(a+b+1)(a+b)}$$

gilt

$$\begin{aligned} \operatorname{Var}(X) &= E(X^2) - E(X)^2 \\ &= \frac{(a+1)a}{(a+b+1)(a+b)} - \frac{a^2}{(a+b)^2} \\ &= \frac{ab}{(a+b+1)(a+b)^2} \end{aligned}$$

Der nächste Satz zeigt, wie sich die Varianz unter Lineartransformationen verhält.

Satz 5.3.2 Sei $Var(X)$ die Varianz der Zufallsvariablen X .
Dann gilt:

$$Var(a + bX) = b^2 Var(X) \quad (5.23)$$

Beweis:

Es gilt

$$\begin{aligned} Var(a + bX) &= E((a + bX - E(a + bX))^2) \\ &= E((a + bX - a - bE(X))^2) \\ &= E((bX - bE(X))^2) \\ &= E(b^2(X - E(X))^2) \\ &= b^2 E((X - E(X))^2) \\ &= b^2 Var(X) \end{aligned}$$

Der folgende Satz gibt die Varianz der Standardnormalverteilung an.

Satz 5.3.3 Sei Z standardnormalverteilt. Dann gilt

$$Var(Z) = 1. \quad (5.24)$$

Beweis:

Es gilt

$$Var(Z) = E(Z^2) - E(Z)^2$$

Nun gilt

$$\begin{aligned} E(Z^2) &= \int_{-\infty}^{\infty} z^2 \frac{1}{\sqrt{2\pi}} e^{-0.5z^2} dz \\ &= \left[z \frac{1}{\sqrt{2\pi}} e^{-0.5z^2} \right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-0.5z^2} dz \\ &= 0 + 1 \\ &= 1 \end{aligned}$$

Also gilt

$$\begin{aligned} \text{Var}(Z) &= E(Z^2) - E(Z)^2 \\ &= 1 \end{aligned}$$

Ist Z standardnormalverteilt, so ist $X = \mu + Z\sigma$ normalverteilt mit den Parametern μ und σ^2 .

Wegen $\text{Var}(Z) = 1$ folgt

$$\begin{aligned} \text{Var}(X) &= \text{Var}(\mu + Z\sigma) \\ &= \sigma^2 \text{Var}(Z) \\ &= \sigma^2 \end{aligned}$$

Der Parameter σ^2 ist also die Varianz der Normalverteilung.

Wann ist eine Varianz groß?

Die Tschebyscheff-Ungleichung gibt eine Antwort auf diese Frage.

Satz 5.3.4 Sei X eine Zufallsvariable mit Erwartungswert μ_X und Varianz σ_X^2 .

Dann gilt

$$P(|X - \mu_X| \geq t) \leq \frac{\sigma_X^2}{t^2} \quad (5.25)$$

Beweis:

Wir setzen in der Markow-Ungleichung (5.12)

$$Y = (X - \mu_X)^2.$$

Dann gilt

$$P\left((X - \mu_X)^2 \geq a\right) \leq \frac{E\left((X - \mu_X)^2\right)}{a}$$

Also gilt

$$P\left(|X - \mu_X| \geq \sqrt{a}\right) \leq \frac{\sigma_X^2}{a}$$

und die Behauptung ist erfüllt mit $t = \sqrt{a}$.

Die Tschebyscheff-Ungleichung unterstützt die Interpretation der Varianz als Streuungsmass.

Ist nämlich $\text{Var}(X)$ klein, so ist die Wahrscheinlichkeit gross, daß X nicht stark von $E(X)$ abweichen wird.

Setzen wir in der Tschebyscheff-Ungleichung $t = k\sigma$, so erhalten wir folgende Ungleichung

$$P(|X - \mu_X| \geq k \sigma_X) \leq \frac{1}{k^2} \quad (5.26)$$

Wir erhalten eine Abschätzung für die Wahrscheinlichkeit, daß die Zufallsvariable X um mindestens k Standardabweichungen von ihrem Erwartungswert abweicht.

In der folgenden Tabelle ist diese Abschätzung für einige Werte von k zusammengestellt:

| k | $P(X - \mu_X \geq k\sigma_X) \leq$ |
|---|--------------------------------------|
| 1 | 1.000 |
| 2 | 0.250 |
| 3 | 0.110 |
| 4 | 0.063 |

5.4 Momente und momenterzeugende Funktion

Die Varianz einer Zufallsvariablen X berechnet man durch

$$\text{Var}(X) = E(X^2) - E(X)^2. \quad (5.27)$$

Die dabei auftretenden Größen $E(X)$ und $E(X^2)$ sind spezielle Momente der Zufallsvariablen X .

Definition 5.4.1 Sei X eine Zufallsvariable.

Dann heißt $E(X^r)$ r .tes Moment von X und $E((X - \mu)^r)$ heißt r .tes zentrales Moment

Offensichtlich ist das erste zentrale Moment gleich Null und das zweite zentrale Moment gleich der Varianz.

Das dritte und vierte zentrale Moment beschreiben bestimmte Charakteristika einer Verteilung.

Definition 5.4.2 Die Schiefe einer Zufallsvariablen X ist definiert durch

$$\gamma_1 = \frac{E((X - \mu)^3)}{\sigma^3}. \quad (5.28)$$

Die Kurtosis einer Zufallsvariablen X ist definiert durch

$$\gamma_2 = \frac{E((X - \mu)^4)}{\sigma^4}. \quad (5.29)$$

Ist eine Verteilung symmetrisch, so ist die Schiefe gleich 0.

Die Kurtosis nimmt bei Normalverteilung den Wert 3 an.

Beispiel 5.4.1 Sei X bernoulliverteilt mit Parameter p .

Dann gilt

$$E(X^k) = p \quad (5.30)$$

Dies sieht man folgendermaßen:

$$\begin{aligned} E(X^k) &= 0^k \cdot (1 - p) + 1^k \cdot p \\ &= p \end{aligned}$$

Alle k .ten Momente der Bernoulliverteilung sind also gleich p .

Beispiel 5.4.2 Wie wir gesehen haben, ist bei einer standardnormalverteilten Zufallsvariablen Z :

$$\begin{aligned} E(Z) &= 0 \\ E(Z^2) &= 1 \end{aligned}$$

Die Berechnung der Momente erleichtert die momenterzeugende Funktion.

Definition 5.4.3 Sei X eine Zufallsvariable. Dann besitzt X eine momenterzeugende Funktion, wenn es ein $h > 0$ gibt, so daß für alle $t \in (-h, h)$ gilt:

$$E(e^{tX}) < \infty$$

Die momenterzeugende Funktion von X ist dann für $t \in (-h, h)$ definiert durch:

$$m_X(t) = E(e^{tX}) \quad (5.31)$$

Schauen wir uns einige Beispiele an.

Beispiel 5.4.3 Sei X bernoulliverteilt mit Parameter p .

Dann gilt

$$m_X(t) = 1 - p + pe^t \quad (5.32)$$

Dies sieht man folgendermaßen:

$$\begin{aligned} m_X(t) &= E(e^{tX}) \\ &= e^{t \cdot 0}(1 - p) + e^{t \cdot 1} p \\ &= 1 - p + pe^t \end{aligned}$$

Beispiel 5.4.4 Sei X binomialverteilt mit den Parametern n und p .

Dann gilt

$$m_X(t) = (e^t p + 1 - p)^n$$

Dies sieht man folgendermaßen:

$$\begin{aligned} m_X(t) &= \sum_{x=0}^n e^{tx} \binom{n}{x} p^x (1 - p)^{n-x} \\ &= \sum_{x=0}^n \binom{n}{x} (e^t p)^x (1 - p)^{n-x} \\ &= (e^t p + 1 - p)^n \end{aligned}$$

Beispiel 5.4.5 Sei Z standardnormalverteilt.

Dann gilt

$$m_X(t) = e^{0.5t^2}$$

Dies sieht man folgendermaßen:

$$\begin{aligned} m_X(t) &= E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}+tx} dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-0.5(x^2-2tx+t^2)+0.5t^2} dx \\ &= e^{0.5t^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-t)^2}{2}} dx \\ &= e^{0.5t^2} \end{aligned}$$

da

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

die Dichtefunktion einer mit den Parametern $\mu = t$ und $\sigma^2 = 1$ normalverteilten Zufallsvariablen ist.

Woher kommt die Bezeichnung momenterzeugende Funktion?

Schauen wir uns die erste Ableitung der momenterzeugenden Funktion nach t an:

$$\begin{aligned} m_X^{(1)}(t) &= \frac{d}{dt} m_X(t) \\ &= \frac{d}{dt} E(e^{tX}) \\ &= E\left(\frac{d}{dt} e^{tX}\right) \\ &= E(X e^{tX}) \end{aligned}$$

Somit gilt

$$\begin{aligned} m_X^{(1)}(0) &= E(Xe^{0 \cdot X}) \\ &= E(X) \end{aligned}$$

Die erste Ableitung der momenterzeugenden Funktion an der Stelle 0 liefert also das erste Moment.

Für die zweite Ableitung gilt

$$\begin{aligned} m_X^{(2)}(t) &= \frac{d}{dt} m_X^{(1)}(t) \\ &= \frac{d}{dt} E(X e^{tX}) \\ &= E\left(\frac{d}{dt} X e^{tX}\right) \\ &= E(X^2 e^{tX}) \end{aligned}$$

Somit gilt

$$\begin{aligned} m_X^{(2)}(0) &= E(X^2 e^{0 \cdot X}) \\ &= E(X^2) \end{aligned}$$

Entsprechend erhält man das k.te Moment als k.te Ableitung der momenterzeugenden Funktion an der Stelle 0.

$$m_X^{(k)}(0) = E(X^k) \quad (5.33)$$

Beispiel 5.4.6 *Wie wir gesehen haben, lautet die momenterzeugende Funktion der Bernoulliverteilung*

$$m_X(t) = 1 - p + pe^t.$$

Es gilt

$$m_X^{(1)}(t) = pe^t$$

Offensichtlich gilt

$$m_X^{(k)}(t) = pe^t$$

Also gilt

$$m_X^{(k)}(0) = p$$

und alle Momente der Bernoulliverteilung sind gleich p, wie wir bereits gesehen haben.

Beispiel 5.4.7 *Wie wir gesehen haben, lautet die momenterzeugende Funktion der Standardnormalverteilung*

$$m_Z(t) = e^{0.5t^2}$$

Es gilt

$$m_Z^{(1)}(t) = te^{0.5t^2}$$

und

$$m_Z^{(2)}(t) = t^2e^{0.5t^2} + e^{0.5t^2}$$

Also gilt

$$\begin{aligned} m_Z^{(1)}(0) &= 0 \cdot e^{0.5 \cdot 0^2} \\ &= 0 \end{aligned}$$

und

$$\begin{aligned} m_Z^{(2)}(0) &= 0^2e^{0.5 \cdot 0^2} + e^{0.5 \cdot 0^2} \\ &= 1 \end{aligned}$$

Also ist bei der Standardnormalverteilung

$$E(Z) = 0$$

und

$$E(Z^2) = 1$$

Der folgende Satz zeigt, daß die momenterzeugende Funktion eindeutig ist.

Satz 5.4.1 *Seien X und Y Zufallsvariablen mit momenterzeugenden Funktionen $m_X(t)$ bzw. $m_Y(t)$.*

Gilt

$$m_X(t) = m_Y(t)$$

für alle t aus einem Intervall um 0, so gilt

$$F_X(x) = F_Y(x)$$

für alle $x \in \mathfrak{R}$.

Aufgrund des Satzes kann man also von der momenterzeugenden Funktion eindeutig auf die Verteilungsfunktion zurückschließen.

Der folgende Satz zeigt, wie die momenterzeugende Funktion sich unter Lineartransformationen verhält.

Satz 5.4.2 Sei $m_X(t)$ die momenterzeugende Funktion der Zufallsvariablen X .

Dann ist die momenterzeugende Funktion der Zufallsvariablen

$$Y = a + bX$$

gegeben durch

$$m_Y(t) = e^{at} m_X(bt). \quad (5.34)$$

Beweis:

$$\begin{aligned} m_Y(t) &= E(e^{tY}) \\ &= E(e^{t(a+bX)}) \\ &= E(e^{ta+tbX}) \\ &= e^{ta} E(e^{btX}) \\ &= e^{at} m_X(bt) \end{aligned}$$

Beispiel 5.4.8 Im Beispiel 4.0.1 haben wir gezeigt, daß

$$X = \mu + Z\sigma$$

normalverteilt ist mit den Parametern μ und σ^2 , wenn Z standardnormalverteilt ist.

Die momenterzeugende Funktion der Standardnormalverteilung ist gegeben durch:

$$m_Z(t) = e^{0.5 t^2}$$

Aufgrund von Satz 5.4.2 ist die momenterzeugende Funktion einer mit Parametern μ und σ^2 normalverteilten Zufallsvariablen also gegeben durch

$$m_X(t) = e^{\mu t + 0.5 \sigma^2 t^2} \quad (5.35)$$

Dies sieht man folgendermaßen:

$$\begin{aligned} m_X(t) &= e^{\mu t} e^{0.5 \sigma^2 t^2} \\ &= e^{\mu t + 0.5 \sigma^2 t^2} \end{aligned}$$

Wir können auch hier wieder Erwartungswert und Varianz bestimmen.
Es gilt

$$\begin{aligned} m_X^{(1)}(t) &= \frac{d}{dt} m_X(t) \\ &= \frac{d}{dt} e^{\mu t + 0.5\sigma^2 t^2} \\ &= e^{\mu t + 0.5\sigma^2 t^2} (\mu + \sigma^2 t) \end{aligned}$$

Außerdem gilt

$$\begin{aligned} m_X^{(2)}(t) &= \frac{d^2}{dt^2} m_X(t) \\ &= \frac{d}{dt} e^{\mu t + 0.5\sigma^2 t^2} (\mu + \sigma^2 t) \\ &= e^{\mu t + 0.5\sigma^2 t^2} (\mu + \sigma^2 t)^2 + e^{\mu t + 0.5\sigma^2 t^2} \sigma^2 \end{aligned}$$

Es gilt

$$m_X^{(1)}(0) = \mu$$

Also ist der Erwartungswert gleich μ .

Weiterhin gilt

$$m_X^{(2)}(0) = \mu^2 + \sigma^2$$

Also gilt

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

5.5 Momente spezieller Verteilungen

Im folgenden sind Erwartungswert und Varianz für einige wichtige Verteilungen zusammengefaßt.

- Diskrete Gleichverteilung

Wahrscheinlichkeitsfunktion

$$P(X = x) = \frac{1}{N}$$

für $x = 1, 2, \dots, N$

Erwartungswert $E(X) = \frac{N + 1}{2}$

Varianz $Var(X) = \frac{N^2 - 1}{12}$

- Bernoulliverteilung

Wahrscheinlichkeitsfunktion

$$P(X = x) = p^x (1 - p)^{(1-x)}$$

für $x = 0, 1$

Erwartungswert $E(X) = p$

Varianz $Var(X) = p(1 - p)$

- Binomialverteilung

Wahrscheinlichkeitsfunktion

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

Erwartungswert $E(X) = n p$

Varianz $Var(X) = n p (1 - p)$

- geometrische Verteilung

Wahrscheinlichkeitsfunktion

$$P(X = x) = p(1 - p)^x$$

für $x = 0, 1, \dots$

Erwartungswert $E(X) = \frac{1 - p}{p}$

Varianz $Var(X) = \frac{1 - p}{p^2}$

- Pascalverteilung

Wahrscheinlichkeitsfunktion

$$P(X = x) = \binom{x + r - 1}{x} p^r (1 - p)^x$$

für $x = 0, 1, \dots$

Erwartungswert $E(X) = r \frac{1 - p}{p}$

Varianz $Var(X) = r \frac{1 - p}{p^2}$

- Poissonverteilung

Wahrscheinlichkeitsfunktion

$$P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

für $x = 0, 1, \dots$

Erwartungswert $E(X) = \lambda$

Varianz $Var(X) = \lambda$

- Stetige Gleichverteilung auf (a, b)

Dichtefunktion

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{für } a < x < b \\ 0 & \text{sonst} \end{cases}$$

Erwartungswert $E(X) = \frac{a+b}{2}$

Varianz $Var(X) = \frac{(b-a)^2}{12}$

- Normalverteilung

Dichtefunktion

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{für } x \in \mathfrak{R}$$

Erwartungswert $E(X) = \mu$

Varianz $Var(X) = \sigma^2$

- Betaverteilung

Dichtefunktion

$$f_X(x) = \begin{cases} \frac{1}{B(a,b)} x^{a-1} (1-x)^{b-1} & \text{für } 0 < x < 1, a, b > 0 \\ 0 & \text{sonst} \end{cases}$$

Erwartungswert $E(X) = \frac{a}{a+b}$

Varianz $Var(X) = \frac{ab}{(a+b+1)(a+b)^2}$

- Gammaverteilung

Dichtefunktion

$$f_T(t) = \begin{cases} \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda t} & \text{für } x > 0 \\ 0 & \text{sonst} \end{cases}$$

Erwartungswert $E(X) = \frac{r}{\lambda}$

Varianz $Var(X) = \frac{r}{\lambda^2}$

- Exponentialverteilung

Dichtefunktion

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{für } x > 0 \\ 0 & \text{sonst} \end{cases}$$

Erwartungswert $E(X) = \frac{1}{\lambda}$

Varianz $Var(X) = \frac{1}{\lambda^2}$

- Cauchyverteilung

Dichtefunktion

$$f_X(x) = \frac{1}{\pi(1+x^2)}$$

Erwartungswert existiert nicht

Varianz existiert nicht

5.6 Score und Fisher-Information

Wir bezeichnen in diesem Abschnitt auch die Wahrscheinlichkeitsfunktion mit $f_X(x)$. Wir betrachten eine Zufallsvariable X mit Wahrscheinlichkeits- bzw. Dichtefunktion $f_X(x)$, die von einem Parameter θ abhängt. Wir schreiben $f_X(x, \theta)$.

Die Größe

$$U(\theta) = \frac{\partial}{\partial \theta} \ln f_X(X, \theta) \quad (5.36)$$

heißt **Score**.

Wegen

$$\frac{\partial}{\partial \theta} \ln f_X(X, \theta) = \frac{\frac{\partial}{\partial \theta} f_X(X, \theta)}{f_X(X, \theta)}$$

gibt sie die relative Veränderung der Wahrscheinlichkeits- bzw. Dichtefunktion in Abhängigkeit von θ an.

Der folgende Satz gibt den Erwartungswert und die Varianz des Score unter bestimmte Regularitätsbedingungen an.

Satz 5.6.1 *Die Wahrscheinlichkeits- bzw. Dichtefunktion $f_X(x, \theta)$ erfülle die folgenden Bedingungen:*

$$\int_{-\infty}^{\infty} \frac{\partial}{\partial \theta} f_X(x, \theta) dx = \frac{\partial}{\partial \theta} \int_{-\infty}^{\infty} f_X(x, \theta) dx$$

$$\int_{-\infty}^{\infty} \frac{\partial^2}{\partial \theta^2} f_X(x, \theta) dx = \frac{\partial^2}{\partial \theta^2} \int_{-\infty}^{\infty} f_X(x, \theta) dx$$

$$\sum_x \frac{\partial}{\partial \theta} f_X(x, \theta) = \frac{\partial}{\partial \theta} \sum_x f_X(x, \theta)$$

$$\sum_x \frac{\partial^2}{\partial \theta^2} f_X(x, \theta) = \frac{\partial^2}{\partial \theta^2} \sum_x f_X(x, \theta)$$

Sei

$$U(\theta) = \frac{\partial}{\partial \theta} \ln f_X(X, \theta).$$

Dann gilt

$$E(U(\theta)) = 0 \quad (5.37)$$

und

$$\text{Var}(U(\theta)) = -E\left(\frac{\partial}{\partial\theta}U(\theta)\right) \quad (5.38)$$

Beweis:

Wir beweisen den diskreten Fall

$$\begin{aligned} E(U(\theta)) &= E\left(\frac{\partial}{\partial\theta} \ln f_X(X, \theta)\right) \\ &= E\left(\frac{\frac{\partial}{\partial\theta} f_X(X, \theta)}{f_X(X, \theta)}\right) \\ &= \sum_x \frac{\frac{\partial}{\partial\theta} f_X(x, \theta)}{f_X(x, \theta)} f_X(x, \theta) \\ &= \sum_x \frac{\partial}{\partial\theta} f_X(x, \theta) \\ &= \frac{\partial}{\partial\theta} \sum f_X(x, \theta) \\ &= \frac{\partial}{\partial\theta} 1 \\ &= 0 \end{aligned}$$

Weiterhin gilt

$$\begin{aligned} \frac{\partial^2}{\partial\theta^2} \ln f_X(X, \theta) &= \frac{\partial}{\partial\theta} \frac{\frac{\partial}{\partial\theta} f_X(X, \theta)}{f_X(X, \theta)} = \\ &= \frac{f_X(X, \theta) \frac{\partial^2}{\partial\theta^2} f_X(X, \theta) - \left(\frac{\partial}{\partial\theta} f_X(X, \theta)\right)^2}{f_X(X, \theta)^2} = \\ &= \frac{\frac{\partial^2}{\partial\theta^2} f_X(X, \theta)}{f_X(X, \theta)} - \left(\frac{\frac{\partial}{\partial\theta} f_X(X, \theta)}{f_X(X, \theta)}\right)^2 \end{aligned}$$

Da gilt

$$\begin{aligned}
 E\left(\frac{\frac{\partial^2}{\partial\theta^2} f_X(X, \theta)}{f_X(X, \theta)}\right) &= \sum_x \frac{\frac{\partial^2}{\partial\theta^2} f_X(x, \theta)}{f_X(x, \theta)} f_X(x, \theta) \\
 &= \sum_x \frac{\partial^2}{\partial\theta^2} f_X(x, \theta) \\
 &= \frac{\partial^2}{\partial\theta^2} \sum_x f_X(x, \theta) \\
 &= \frac{\partial^2}{\partial\theta^2} 1 \\
 &= 0
 \end{aligned}$$

folgt

$$\begin{aligned}
 E\left(\frac{\partial}{\partial\theta} U(\theta)\right) &= E\left(\frac{\partial^2}{\partial\theta^2} \ln f_X(X, \theta)\right) \\
 &= E\left(\frac{\frac{\partial^2}{\partial\theta^2} f_X(X, \theta)}{f_X(X, \theta)}\right) - E\left(\left(\frac{\frac{\partial}{\partial\theta} f_X(X, \theta)}{f_X(X, \theta)}\right)^2\right) \\
 &= -E(U(\theta)^2) \\
 &= -E[(U(\theta) - E(U(\theta)))^2] \\
 &= -\text{Var}(U(\theta))
 \end{aligned}$$

Definition 5.6.1 Die Größe

$$I(\theta) = \text{Var}(U(\theta))$$

heißt **Fisher-Information**.

Beispiel 5.6.1 Sei X bernoulliverteilt mit Parameter p .

Es gilt also

$$f_X(x, p) = p^x(1-p)^{1-x} \quad \text{für } x = 0, 1.$$

Dann gilt für die Fisher-Information

$$I(p) = \frac{1}{p(1-p)} \tag{5.39}$$

Dies sieht man folgendermaßen:

Es gilt

$$\ln f_X(x, p) = x \ln p + (1-x) \ln(1-p)$$

Somit ist der Score:

$$\begin{aligned} U(p) &= \frac{\partial}{\partial p} \ln f_X(X, p) \\ &= \frac{X}{p} - \frac{1-X}{1-p}. \end{aligned}$$

Hieraus folgt

$$\begin{aligned} I(p) &= E(U(p)^2) = \\ &= E \left[\left(\frac{X-p}{p(1-p)} \right)^2 \right] \\ &= \frac{1}{p^2(1-p)^2} E[(X-p)^2] \\ &= \frac{1}{p^2(1-p)^2} \text{Var}(X) \\ &= \frac{p(1-p)}{p^2(1-p)^2} \\ &= \frac{1}{p(1-p)} \end{aligned}$$

Beispiel 5.6.2 Sei X poissonverteilt mit Parameter λ .

Es gilt also

$$f_X(x, \lambda) = \frac{\lambda^x}{x!} e^{-\lambda} \quad \text{für } x = 0, 1, \dots$$

Dann gilt für die Fisher-Information

$$I(\lambda) = \frac{1}{\lambda} \tag{5.40}$$

Dies sieht man folgendermaßen:

Es gilt

$$\ln f_X(x, \lambda) = x \ln \lambda - \ln x! - \lambda$$

Somit ist der Score:

$$\begin{aligned} U(\lambda) &= \frac{\partial}{\partial \lambda} \ln f_X(X, \lambda) \\ &= \frac{X}{\lambda} - 1 \end{aligned}$$

Es folgt

$$\begin{aligned} I(\lambda) &= -E \left(\frac{\partial}{\partial \lambda} U(\lambda) \right) \\ &= -E \left(-\frac{X}{\lambda^2} \right) \\ &= \frac{1}{\lambda^2} E(X) \\ &= \frac{1}{\lambda} \end{aligned}$$

Beispiel 5.6.3 Sei X exponentialverteilt mit Parameter λ .

Es gilt also

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{für } x > 0 \\ 0 & \text{sonst} \end{cases}$$

Dann gilt für die Fisher-Information

$$I(\lambda) = \frac{1}{\lambda^2} \quad (5.41)$$

Dies sieht man folgendermaßen:

Es gilt

$$\ln f_X(x, \lambda) = \ln \lambda - \lambda x$$

Somit ist der Score:

$$\begin{aligned} U(\lambda) &= \frac{\partial}{\partial \lambda} \ln f_X(X, \lambda) \\ &= \frac{1}{\lambda} - X \end{aligned}$$

Es folgt

$$\begin{aligned} I(\lambda) &= -E \left(\frac{\partial}{\partial \lambda} U(\lambda) \right) \\ &= -E \left(-\frac{1}{\lambda^2} \right) \\ &= \frac{1}{\lambda^2} \end{aligned}$$

5.7 Momente der Exponentialfamilie

Wie wir in Kapitel gesehen haben, ist die Wahrscheinlichkeits- bzw. Dichtefunktion der Exponentialfamilie von der folgenden Form

$$f_X(x, \theta) = e^{\theta x - b(\theta) + d(x)}$$

Der Erwartungswert $E(X)$ und die Varianz $Var(X)$ einer Zufallsvariablen, die eine Verteilung aus der Exponentialfamilie besitzt, hängen in einfacher Form von der Funktion $b(\theta)$ ab.

Mit den Ergebnissen aus dem letzten Abschnitt können wir diese Abhängigkeit einfach bestimmen.

Die Exponentialfamilie erfüllt nämlich die Bedingungen von Satz 3.12 (siehe dazu Lehmann[1986]: Testing Statistical Hypotheses, S.59).

Satz 5.7.1 *Die Wahrscheinlichkeits- bzw. Dichtefunktion der Zufallsvariablen X gehöre zur Exponentialfamilie.*

Dann gilt

$$E(X) = b'(\theta) \tag{5.42}$$

und

$$Var(X) = b''(\theta) \tag{5.43}$$

Dabei gilt

$$b'(\theta) = \frac{d}{d\theta} b(\theta)$$

und

$$b''(\theta) = \frac{d^2}{d\theta^2} b(\theta)$$

Beweis:

Wir bestimmen den Score der Exponentialfamilie.

Es gilt

$$\ln f_X(X, \theta) = \theta X - b(\theta) + d(X)$$

und

$$\begin{aligned} U(\theta) &= \frac{\partial}{\partial \theta} \ln f_X(X, \theta) \\ &= X - b'(\theta) \end{aligned}$$

Wegen

$$E(U(\theta)) = 0$$

folgt

$$E(X - b'(\theta)) = 0$$

und somit

$$E(X) = b'(\theta)$$

Außerdem gilt

$$\begin{aligned} \text{Var}(X) &= \text{Var}(U(\theta)) \\ &= -E\left(\frac{\partial}{\partial\theta} U(\theta)\right) \\ &= b''(\theta) \end{aligned}$$

Wir können diese Beziehung auch über die momenterzeugende Funktion bestimmen.

Wir berechnen also zunächst die momenterzeugende Funktion der Exponentialfamilie.

Satz 5.7.2 *Die Zufallsvariable X besitze eine Verteilung aus der Exponentialfamilie.*

Dann gilt

$$m_X(t) = e^{b(\theta+t)-b(\theta)}$$

Beweis:

Wir zeigen den diskreten Fall.

$$\begin{aligned} E(e^{tX}) &= \sum_x e^{tx} e^{\theta x - b(\theta) + d(x)} \\ &= \sum_x e^{tx + \theta x - b(\theta) + d(x)} \\ &= \sum_x e^{(\theta+t)x - b(\theta) + d(x)} \\ &= \sum_x e^{(\theta+t)x + b(\theta+t) - b(\theta+t) - b(\theta) + d(x)} \\ &= e^{b(\theta+t) - b(\theta)} \sum_x e^{(\theta+t)x - b(\theta+t) + d(x)} \\ &= e^{b(\theta+t) - b(\theta)} \end{aligned}$$

Der folgende Satz gibt den Erwartungswert und die Varianz einer Zufallsvariable an, die eine Verteilung aus der Exponentialfamilie besitzt:

Satz 5.7.3 *Die Zufallsvariable X besitze eine Verteilung aus der Exponentialfamilie.*

Dann gilt

$$E(X) = b'(\theta)$$

und

$$\text{Var}(X) = b''(\theta)$$

Beweis:

Es gilt

$$m_X^{(1)}(t) = e^{b(\theta+t)-b(\theta)} b'(\theta+t)$$

und

$$m_X^{(2)}(t) = e^{b(\theta+t)-b(\theta)} b''(\theta+t) + e^{b(\theta+t)-b(\theta)} b'(\theta+t)^2$$

Also gilt

$$m_X^{(1)}(0) = b'(\theta)$$

und

$$m_X^{(2)}(0) = b''(\theta) + b'(\theta)^2$$

Somit gilt

$$\begin{aligned} E(X) &= m_X^{(1)}(0) \\ &= b'(\theta) \end{aligned}$$

und

$$\begin{aligned} \text{Var}(X) &= m_X^{(2)}(0) - (m_X^{(1)}(0))^2 \\ &= b''(\theta) + b'(\theta)^2 - b'(\theta)^2 \\ &= b''(\theta) \end{aligned}$$

Schauen wir uns einige Beispiele an:

Beispiel 5.7.1 Für die Binomialverteilung gilt

$$b(\theta) = n \ln(1 + e^\theta)$$

Somit gilt

$$b'(\theta) = n \frac{e^\theta}{1 + e^\theta}$$

und

$$b''(\theta) = n \frac{e^\theta}{(1 + e^\theta)^2}$$

Wegen

$$\theta = \ln \frac{p}{1 - p}$$

und

$$p = \frac{e^\theta}{1 + e^\theta}$$

gilt also

$$E(X) = n p$$

und

$$\text{Var}(X) = n p (1 - p)$$

Beispiel 5.7.2 Für die Poissonverteilung gilt

$$b(\theta) = e^\theta$$

Also gilt

$$b'(\theta) = e^\theta$$

und

$$b''(\theta) = e^\theta.$$

Wegen

$$\theta = \ln \lambda$$

und

$$\lambda = e^\theta$$

gilt

$$E(X) = \lambda$$

und

$$\text{Var}(X) = \lambda$$

Beispiel 5.7.3 Für die geometrische Verteilung gilt

$$b(\theta) = -\ln(1 - e^\theta)$$

Somit gilt

$$b'(\theta) = \frac{e^\theta}{1 - e^\theta}$$

und

$$b''(\theta) = \frac{e^\theta}{(1 - e^\theta)^2}$$

Wegen

$$\theta = \ln(1 - p)$$

und

$$p = 1 - e^\theta$$

gilt

$$E(X) = \frac{1 - p}{p}$$

und

$$\text{Var}(X) = \frac{1 - p}{p^2}$$

Beispiel 5.7.4 Für die Pascalverteilung gilt

$$b(\theta) = -r \ln(1 - e^\theta)$$

Somit gilt

$$b'(\theta) = r \frac{e^\theta}{1 - e^\theta}$$

und

$$b''(\theta) = r \frac{e^\theta}{(1 - e^\theta)^2}$$

Wegen

$$\theta = \ln(1 - p)$$

und

$$p = 1 - e^\theta$$

gilt

$$E(X) = r \frac{1 - p}{p}$$

und

$$\text{Var}(X) = r \frac{1 - p}{p^2}$$

Beispiel 5.7.5 Für die Exponentialverteilung gilt

$$b(\theta) = -\ln -\theta$$

Also gilt

$$b'(\theta) = -\frac{1}{\theta}$$

und

$$b''(\theta) = \frac{1}{\theta^2}$$

Wegen

$$\theta = -\lambda$$

und

$$\lambda = -\theta$$

gilt

$$E(X) = \frac{1}{\lambda}$$

und

$$\text{Var}(X) = \frac{1}{\lambda^2}$$

Beispiel 5.7.6 Wir betrachten die Normalverteilung mit festem σ^2 und setzen $\sigma = 1$.

Es gilt

$$b(\theta) = 0.5\theta^2.$$

Also gilt

$$b'(\theta) = \theta$$

und

$$b''(\theta) = 1$$

Wegen

$$\theta = \mu$$

und

$$\mu = \theta$$

gilt

$$E(X) = \mu$$

und

$$\text{Var}(X) = 1$$

Kapitel 6

Mehrdimensionale Verteilungen

6.1 Mehrdimensionale Zufallsvariablen

Bisher haben wir immer nur eine Zufallsvariable betrachtet. Bei vielen Anwendungen sind aber mehrere Zufallsvariablen von Interesse. So besteht ein Fragebogen in der Regel aus mehreren Fragen, mit denen die Ausprägungen von Merkmalen erfragt werden. Es sollen Zusammenhänge zwischen den Merkmalen aufgedeckt werden.

Die Definition einer univariaten Zufallsvariablen legt folgende Definition nahe:

Definition 6.1.1 Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum.

Eine Abbildung $X : \Omega \rightarrow \mathfrak{R}^k$ heißt *k-dimensionale Zufallsvariable* $X = (X_1, \dots, X_k)$, falls gilt

$$\{\omega | X_1(\omega) \leq x_1, X_2(\omega) \leq x_2, \dots, X_k(\omega) \leq x_k\} \in \mathcal{A} \quad \text{für jedes } x_i \in \mathfrak{R}$$

Wir werden im folgenden in der Regel zweidimensionale Zufallsvariablen betrachten. In diesem Fall schreiben wir für die zweidimensionale Zufallsvariable (X, Y) .

Beispiel 6.1.1 Wir betrachten wieder das Teilchen, das sich zufällig auf den ganzen Zahlen bewegt.

Uns interessieren folgende Zufallsvariablen:

X Anzahl der Schritte nach rechts beim ersten Schritt

Y Position des Teilchens

In der folgenden Tabelle sind die Ergebnisse mit den zugehörigen Werten von X und Y zu finden.

| ω | x | y |
|----------|-----|-----|
| RR | 1 | 2 |
| RL | 1 | 0 |
| LR | 0 | 0 |
| LL | 0 | -2 |

Es gilt

$$\{\omega | X(\omega) \leq x, Y(\omega) \leq y\} = \begin{cases} \emptyset & \text{für } x < 0 \text{ oder } y < -2 \\ \{LL\} & \text{für } x \geq 0, -2 \leq y < 0 \\ \{LL, LR\} & \text{für } 0 \leq x < 1, y \geq 0 \\ \{RL, LR, LL\} & \text{für } x \geq 1, 0 \leq y < 2 \\ \Omega & \text{für } x \geq 1, y \geq 2 \end{cases}$$

Wählen wir als σ -Algebra die Potenzmenge von Ω , so handelt es sich um eine zweidimensionale Zufallsvariable.

Es hätte gereicht, zu zeigen, daß X und Y eindimensionale Zufallsvariablen sind.

Sind nämlich X_1, X_2, \dots, X_k eindimensionale Zufallsvariablen über dem gleichen Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) , so ist auch (X_1, X_2, \dots, X_k) eine k -dimensionale Zufallsvariable über (Ω, \mathcal{A}, P) .

Dies sieht man folgendermaßen:

Da X_1, X_2, \dots, X_k eindimensionale Zufallsvariablen sind, gilt:

$$\{\omega | X_i(\omega) \leq x_i\} \in \mathcal{A} \quad \text{für jedes } x_i \in \mathfrak{R}$$

Da \mathcal{A} eine σ -Algebra ist, gilt auch

$$\begin{aligned} \{\omega | X_1(\omega) \leq x_1, X_2(\omega) \leq x_2, \dots, X_k(\omega) \leq x_k\} &= \\ = \bigcap_{i=1}^k \{\omega | X_i(\omega) \leq x_i\} &\in \mathcal{A} \quad \text{für jedes } x_i \in \mathfrak{R} \end{aligned}$$

Die Definition der mehrdimensionalen Zufallsvariablen stellt sicher, daß allen Ereignissen der Form

$$\{\omega | X_1(\omega) \leq x_1, X_2(\omega) \leq x_2, \dots, X_k(\omega) \leq x_k\}$$

eine Wahrscheinlichkeit

$$P(\{\omega | X_1(\omega) \leq x_1, X_2(\omega) \leq x_2, \dots, X_k(\omega) \leq x_k\})$$

zugeordnet wird.

Für

$$P(\{\omega | X_1(\omega) \leq x_1, X_2(\omega) \leq x_2, \dots, X_k(\omega) \leq x_k\})$$

schreiben wir im folgenden kurz

$$P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_k \leq x_k).$$

Wir sprechen wie auch bei eindimensionalen Zufallsvariablen von der Verteilungsfunktion.

Definition 6.1.2 Sei (X_1, \dots, X_k) eine k -dimensionale Zufallsvariable.

Dann heißt

$$F_{X_1, \dots, X_k}(x_1, \dots, x_k) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_k \leq x_k)$$

die Verteilungsfunktion von (X_1, \dots, X_k) .

Beispiel 6.1.2 Für das Beispiel gilt:

$$F_{X,Y}(x, y) = \begin{cases} 0.00 & \text{für } x < 0 \quad \text{oder } y < -2 \\ 0.25 & \text{für } x \geq 0, -2 \leq y < 0 \\ 0.50 & \text{für } 0 \leq x < 1, y \geq 0 \\ 0.75 & \text{für } x \geq 1, 0 \leq y < 2 \\ 1.00 & \text{für } x \geq 1, y \geq 2 \end{cases}$$

Wie auch bei eindimensionalen Zufallsvariablen unterscheiden wir diskrete und stetige mehrdimensionale Zufallsvariablen.

Wir definieren diese nur für den zweidimensionalen Fall. Die Übertragung auf den höherdimensionalen Fall ist offensichtlich.

Definition 6.1.3 Eine zweidimensionale Zufallsvariable heißt diskret, wenn sie nur abzählbar viele Werte annehmen kann.

dabei heißt

$$P(X = x, Y = y)$$

die Wahrscheinlichkeitsfunktion von (X, Y) .

Beispiel 6.1.3 Die Wahrscheinlichkeitsfunktion der zweidimensionalen Zufallsvariablen aus Beispiel 4.1.1 lautet:

$$P(X = 0, Y = -2) = 0.25$$

$$P(X = 0, Y = 0) = 0.25$$

$$P(X = 1, Y = 0) = 0.25$$

$$P(X = 1, Y = 2) = 0.25$$

Wir stellen die Wahrscheinlichkeitsfunktion einer zweidimensionalen Zufallsvariablen (X, Y) in der Regel in der Form einer Tabelle dar.

Für das Beispiel erhalten wir dann:

| y | -2 | 0 | 2 |
|---|------|------|------|
| x | | | |
| 0 | 0.25 | 0.25 | 0 |
| 1 | 0 | 0.25 | 0.25 |

Wie im univariaten Fall betrachten wir auch im zweidimensionalen Fall stetige Zufallsvariablen.

Definition 6.1.4 Eine Zufallsvariable (X, Y) heißt stetig, wenn eine Funktion $f_{X,Y} : \mathbb{R}^2 \rightarrow \mathbb{R}$ existiert, so daß für die Verteilungsfunktion $F_{X,Y}(x, y)$ von (X, Y) gilt:

$$F_{X,Y}(x, y) = \int_{-\infty}^y \int_{-\infty}^x f_{X,Y}(u, v) du dv$$

Die Funktion $f_{X,Y}(x, y)$ heißt Dichtefunktion der Zufallsvariablen (X, Y) . Die Dichtefunktion $f_{X,Y}(x, y)$ erfüllt folgende Bedingungen:

1. $f_{X,Y}(x, y) \geq 0$ für alle $(x, y) \in \mathbb{R}^2$
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx dy = 1$

Wie auch im univariaten Fall kann man zeigen, daß jede Funktion, die die beiden Bedingungen erfüllt als Dichtefunktion einer zweidimensionalen stetigen Zufallsvariablen aufgefaßt werden kann.

Wie auch im eindimensionalen Fall gewinnen wir die Dichtefunktion durch Ableiten der Verteilungsfunktion:

$$f_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y)$$

Beispiel 6.1.4 *Wir betrachten folgende zweidimensionale stetige Zufallsvariable (X, Y) mit gemeinsamer Dichtefunktion*

$$f_{X,Y}(x, y) = \begin{cases} 1 & \text{für } 0 < x < 1, 0 < y < 1 \\ 0 & \text{sonst} \end{cases}$$

*Offensichtlich ist die Nichtnegativitätsbedingung erfüllt.
Außerdem gilt*

$$\begin{aligned} \int_0^1 \int_0^1 f_{X,Y}(u, v) du dv &= \int_0^1 \int_0^1 1 du dv \\ &= \int_0^1 [u]_0^1 dv \\ &= \int_0^1 1 dv \\ &= [v]_0^1 \\ &= 1 \end{aligned}$$

Also handelt es sich um eine Dichtefunktion.

Die gemeinsame Verteilungsfunktion $F_{X,Y}(x, y)$ lautet dann:
für $0 \leq x \leq 1, 0 \leq y \leq 1$

$$\begin{aligned}
 F_{X,Y}(x, y) &= \int_0^y \int_0^x f_{X,Y}(u, v) \, du \, dv \\
 &= \int_0^y \int_0^x 1 \, du \, dv \\
 &= \int_0^y [u]_0^x \, dv \\
 &= \int_0^y x \, dv \\
 &= [xv]_0^y \\
 &= xy
 \end{aligned}$$

für $0 \leq x \leq 1, 1 < y$

$$\begin{aligned}
 F_{X,Y}(x, y) &= \int_0^x \int_0^1 1 \, dv \, du \\
 &= \int_0^x 1 \, du \\
 &= x
 \end{aligned}$$

für $0 \leq y \leq 1, 1 < x$

$$\begin{aligned}
 F_{X,Y}(x, y) &= \int_0^y \int_0^1 1 \, du \, dv \\
 &= \int_0^y 1 \, dv \\
 &= y
 \end{aligned}$$

Also gilt

$$F_{X,Y}(x, y) = \begin{cases} xy & \text{für } 0 \leq x \leq 1, 0 \leq y \leq 1 \\ x & \text{für } 0 \leq x \leq 1, 1 < y \\ y & \text{für } 0 \leq y \leq 1, 1 < x \\ 1 & \text{für } x \geq 1, y \geq 1 \\ 0 & \text{sonst} \end{cases}$$

Schauen wir uns noch ein Beispiel an.

Beispiel 6.1.5 Sei (X, Y) eine zweidimensionale Zufallsvariable mit gemeinsamer Dichtefunktion

$$f_{X,Y}(x, y) = \begin{cases} 2 & \text{für } 0 \leq x \leq y \leq 1 \\ 0 & \text{sonst} \end{cases}$$

Offensichtlich ist die Nichtnegativitätsbedingung erfüllt.

Außerdem gilt

$$\begin{aligned} \int_0^1 \int_u^1 2 \, dv \, du &= \int_0^1 [2v]_u^1 \, du \\ &= \int_0^1 (2 - 2u) \, du \\ &= [2u - u^2]_0^1 \\ &= 1 \end{aligned}$$

Also handelt es sich um eine Dichtefunktion.

Die gemeinsame Verteilungsfunktion $F_{X,Y}(x, y)$ lautet dann:
für $0 \leq x \leq 1, 1 < y$

$$\begin{aligned} F_{X,Y}(x, y) &= \int_0^x \int_u^1 2 \, dv \, du \\ &= \int_0^x (2 - 2u) \, du \\ &= 2x - x^2 \end{aligned}$$

für $0 \leq x \leq y \leq 1$

$$\begin{aligned}
 F_{X,Y}(x,y) &= \int_0^x \int_u^y 2 \, dv \, du \\
 &= \int_0^x [2v]_u^y \, du \\
 &= \int_0^x (2y - 2u) \, du \\
 &= [2yu - u^2]_0^x \\
 &= 2xy - x^2
 \end{aligned}$$

für $0 \leq y \leq 1, y \leq x$

$$\begin{aligned}
 F_{X,Y}(x,y) &= \int_0^y \int_0^v 2 \, du \, dv \\
 &= \int_0^y 2v \, dv \\
 &= y^2
 \end{aligned}$$

Also gilt

$$F_{X,Y}(x,y) = \begin{cases} 2xy - x^2 & \text{für } 0 \leq x \leq y \leq 1 \\ 2x - x^2 & \text{für } 0 \leq x \leq 1, 1 < y \\ y^2 & \text{für } 0 \leq y \leq 1, y \leq x_1 \\ 1 & \text{für } x \geq 1, y \geq 1 \\ 0 & \text{sonst} \end{cases}$$

6.2 Randverteilungen

Bei einer zweidimensionalen Zufallsvariablen (X, Y) interessiert man sich neben der gemeinsamen Verteilung auch für die Verteilungen von X und Y .

Man spricht auch von Randverteilungen.

Diese kann man leicht aus der gemeinsamen Verteilung bestimmen.

Es gilt

$$\{\omega | X(\omega) \leq x\} = \{\omega | X(\omega) \leq x, Y(\omega) < \infty\}.$$

Somit gilt

$$P(X \leq x) = P(X \leq x, Y < \infty)$$

Somit gilt für die Verteilungsfunktion von X :

$$F_X(x) = F_{X,Y}(x, \infty)$$

Entsprechend läßt sich die Verteilungsfunktion $F_Y(y)$ von Y aus der gemeinsamen Verteilungsfunktion $F_{X,Y}(x, y)$ bestimmen.

Man nennt $F_X(x)$ die Randverteilungsfunktion von X und $F_Y(y)$ die Randverteilungsfunktion von Y .

Wegen

$$\begin{aligned} F_X(x) &= F_{X,Y}(x, \infty) \\ &= \int_{-\infty}^x \int_{-\infty}^{\infty} f_{X,Y}(u, y) dy du \end{aligned}$$

erhält man bei stetigen zweidimensionalen Zufallsvariablen die Randdichtefunktionen $f_X(x)$ folgendermaßen aus der gemeinsamen Dichtefunktion

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy$$

Eine analoge Beziehung gilt für $f_Y(y)$.

Will man also die Randdichtefunktion von X an der Stelle x bestimmen, muß man für festes x die gemeinsame Dichtefunktion $f_{X,Y}(x, y)$ bezüglich y integrieren. In der Regel wird dabei der Bereich, auf dem die gemeinsame Dichtefunktion $f_{X,Y}(x, y)$ für festes x positiv ist, vom Wert x abhängen. Wir werden dies in Beispiel 6.2.2 sehen.

Wir bestimmen die Randdichtefunktionen der Verteilungen der Beispiele:

Beispiel 6.2.1 *Wir betrachten folgende zweidimensionale stetige Zufallsvariable (X, Y) mit gemeinsamer Dichtefunktion*

$$f_{X,Y}(x, y) = \begin{cases} 1 & \text{für } 0 < x < 1, 0 < y < 1 \\ 0 & \text{sonst} \end{cases}$$

Dann gilt für $0 < x < 1$:

$$f_X(x) = \int_0^1 1 \, dy = 1$$

Also lautet die Randdichtefunktion von X :

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

Analog erhalten wir

$$f_Y(y) = \begin{cases} 1 & \text{für } 0 < y < 1 \\ 0 & \text{sonst} \end{cases}$$

Die Randdichtefunktion von X ist also die Dichtefunktion einer auf $(0, 1)$ gleichverteilten Zufallsvariablen.

Das gleich gilt für die Randdichtefunktion von Y .

Beispiel 6.2.2 Sei (X, Y) eine zweidimensionale Zufallsvariable mit gemeinsamer Dichtefunktion

$$f_{X,Y}(x, y) = \begin{cases} 2 & \text{für } 0 \leq x \leq y \leq 1 \\ 0 & \text{sonst} \end{cases}$$

Für festes x ist $f_{X,Y}(x, y)$ positiv, wenn für y gilt $x \leq y \leq 1$.

Also gilt für $0 \leq x \leq 1$:

$$\begin{aligned} f_X(x) &= \int_x^1 2 \, dy \\ &= [2y]_x^1 \\ &= 2 - 2x \end{aligned}$$

Also lautet die Randdichtefunktion von X :

$$f_X(x) = \begin{cases} 2(1-x) & \text{für } 0 \leq x \leq 1 \\ 0 & \text{sonst} \end{cases}$$

Dies ist die Dichtefunktion einer Beta-Verteilung mit den Parametern $a = 1$ und $b = 2$.

Für festes y ist $f_{X,Y}(x, y)$ positiv, wenn für x gilt $0 \leq x \leq y$.

Also gilt für $0 \leq y \leq 1$:

$$\begin{aligned} f_Y(y) &= \int_0^y 2 \, dx \\ &= [2x]_0^y \\ &= 2y \end{aligned}$$

Also lautet die Randdichtefunktion von Y :

$$f_Y(y) = \begin{cases} 2y & \text{für } 0 \leq y \leq 1 \\ 0 & \text{sonst} \end{cases}$$

Dies ist die Dichtefunktion einer Beta-Verteilung mit den Parametern $a = 2$ und $b = 1$.

Die Randwahrscheinlichkeitsfunktionen einer diskreten zweidimensionalen Zufallsvariablen (X, Y) erhält man durch

$$P(X = x) = \sum_y P(X = x, Y = y)$$

und

$$P(Y = y) = \sum_x P(X = x, Y = y)$$

Beispiel 6.2.3 Für das Beispiel 6.1.1 gilt:

$$P(X = 0, Y = -2) = 0.25$$

$$P(X = 0, Y = 0) = 0.25$$

$$P(X = 1, Y = 0) = 0.25$$

$$P(X = 1, Y = 2) = 0.25$$

Somit gilt

$$\begin{aligned} P(X = 0) &= P(X = 0, Y = -2) + P(X = 0, Y = 0) \\ &= 0.25 + 0.25 \\ &= 0.5 \end{aligned}$$

$$\begin{aligned} P(X = 1) &= P(X = 1, Y = 0) + P(X = 1, Y = 2) \\ &= 0.25 + 0.25 \\ &= 0.5 \end{aligned}$$

$$\begin{aligned} P(Y = -2) &= P(X = 0, Y = -2) \\ &= 0.25 \end{aligned}$$

$$\begin{aligned} P(Y = 0) &= P(X = 0, Y = 0) + P(X = 1, Y = 0) \\ &= 0.25 + 0.25 \\ &= 0.5 \end{aligned}$$

$$\begin{aligned} P(Y = 2) &= P(X = 1, Y = 2) \\ &= 0.25 \end{aligned}$$

Wir können die Randverteilung aber auch direkt durch Rändern der Tabelle erreichen.

| | | | | |
|-----|--------|--------|--------|-------|
| y | -2 | 0 | 2 | |
| x | | | | |
| 0 | 0.25 | 0.25 | 0 | 0.5 |
| 1 | 0 | 0.25 | 0.25 | 0.5 |
| | 0.25 | 0.5 | 0.25 | 1 |

Dies erklärt dann auch den Namen Randverteilung.

6.3 Bedingte Verteilungen

Ein wichtiges Konzept der Wahrscheinlichkeitsrechnung, das nun auf Verteilungen übertragen werden soll, ist die bedingte Wahrscheinlichkeit.

Sind A und B Ereignisse, so ist die bedingte Wahrscheinlichkeit von B unter der Bedingung, daß A eingetreten ist, definiert durch

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

Dieses Konzept kann problemlos auf diskrete Zufallsvariablen X und Y übertragen werden.

Mit

$$A = \{\omega | X(\omega) = x\}$$

und

$$B = \{\omega | Y(\omega) = y\}$$

erhalten wir

$$P(Y = y | X = x) = \frac{P(Y = y, X = x)}{P(X = x)} \quad \text{für } P(X = x) > 0$$

Ist $P(X = x) = 0$, so setzen wir $P(Y = y | X = x) = 0$.

Beispiel 6.3.1 *Im Beispiel 6.1.3 gilt:*

$$P(X = 0, Y = -2) = 0.25$$

$$P(X = 0, Y = 0) = 0.25$$

$$P(X = 1, Y = 0) = 0.25$$

$$P(X = 1, Y = 2) = 0.25$$

Die Randwahrscheinlichkeitsfunktionen lauten:

$$P(X = 0) = 0.5$$

$$P(X = 1) = 0.5$$

$$P(Y = -2) = 0.25$$

$$P(Y = 0) = 0.5$$

$$P(Y = 2) = 0.25$$

Wir bestimmen zunächst die bedingte Verteilung von Y gegeben $X = 0$:
Es gilt

$$\begin{aligned} P(Y = -2|X = 0) &= \frac{P(X = 0, Y = -2)}{P(X = 0)} \\ &= \frac{0.25}{0.5} \\ &= 0.5 \end{aligned}$$

und

$$\begin{aligned} P(Y = 0|X = 0) &= \frac{P(X = 0, Y = 0)}{P(X = 0)} \\ &= \frac{0.25}{0.5} \\ &= 0.5 \end{aligned}$$

Wir erhalten also als bedingte Verteilung von Y gegeben $X = 0$

$$P(Y = y|X = 0) = \begin{cases} 0.5 & \text{für } y = -2 \\ 0.5 & \text{für } y = 0 \\ 0 & \text{für } y = 2 \end{cases}$$

Analog erhalten wir für $X = 1$:

$$P(Y = y|X = 1) = \begin{cases} 0 & \text{für } y = -2 \\ 0.5 & \text{für } y = 0 \\ 0.5 & \text{für } y = 2 \end{cases}$$

Analog erhalten wir die bedingte Verteilung von X gegeben $Y = -2$

$$P(X = x|Y = -2) = \begin{cases} 1 & \text{für } x = 0 \\ 0 & \text{für } x = 1 \end{cases}$$

und für $Y = 0$:

$$P(X = x|Y = 0) = \begin{cases} 0.5 & \text{für } x = 0 \\ 0.5 & \text{für } x = 1 \end{cases}$$

und für $Y = 2$:

$$P(X = x|Y = 2) = \begin{cases} 0 & \text{für } x = 0 \\ 1 & \text{für } x = 1 \end{cases}$$

Im stetigen Fall gilt $P(X = x) = 0$ für alle $x \in \mathfrak{R}$, so daß die obige Definition nicht sinnvoll ist.

Folgende Definition liegt nahe

Definition 6.3.1 Sei (X, Y) eine zweidimensionale Zufallsvariable mit Dichtefunktion $f_{X,Y}(x, y)$.

Dann ist die bedingte Dichtefunktion von Y gegeben $X = x$ definiert durch:

$$f_{Y|X}(y|x) = \begin{cases} \frac{f_{X,Y}(x, y)}{f_X(x)} & \text{für } f_X(x) > 0 \\ 0 & \text{sonst} \end{cases}$$

Entsprechend ist $f_{X|Y}(x|y)$ definiert.

Beispiel 6.3.2 Für die zweidimensionale stetige Zufallsvariable (X, Y) mit gemeinsamer Dichtefunktion

$$f_{X,Y}(x, y) = \begin{cases} 1 & \text{für } 0 < x < 1, 0 < y < 1 \\ 0 & \text{sonst} \end{cases}$$

gilt

$$f_{Y|X}(y|x) = \begin{cases} 1 & \text{für } 0 < y < 1 \\ 0 & \text{sonst} \end{cases}$$

und

$$f_{X|Y}(x|y) = \begin{cases} 1 & \text{für } 0 < x < 1 \\ 0 & \text{sonst} \end{cases}$$

Beispiel 6.3.3 Für die zweidimensionale stetige Zufallsvariable (X, Y) mit gemeinsamer Dichtefunktion

$$f_{X,Y}(x, y) = \begin{cases} 2 & \text{für } 0 \leq x \leq y \leq 1 \\ 0 & \text{sonst} \end{cases}$$

gilt

$$f_{Y|X}(y|x) = \begin{cases} \frac{1}{1-x} & \text{für } x \leq y \leq 1 \\ 0 & \text{sonst} \end{cases}$$

und

$$f_{X|Y}(x|y) = \begin{cases} \frac{1}{y} & \text{für } 0 \leq x \leq y \\ 0 & \text{sonst} \end{cases}$$

Wir sehen, daß Y für festes x gleichverteilt auf $[x, 1]$ ist, während X für festes y gleichverteilt auf $[0, y]$ ist.

Ein wichtiges Anwendungsgebiet von bedingten Verteilungen sind sogenannte Mischverteilungen.

Beispiel 6.3.4 Für die Modellierung von Zählvariablen bietet sich die Poissonverteilung an. In der Praxis stellt man oft fest, daß die Poissonverteilung bis auf die Null gut paßt. Die Null tritt häufiger auf als man bei Poissonverteilung erwarten würde.

Ein Beispiel hierfür sind Produktionsprozesse. In einem Zeitintervall ist der Produktionsprozeß entweder völlig störungsfrei, oder es können Störungen auftreten.

Der erste Fall möge mit Wahrscheinlichkeit p eintreten. Alle produzierten Stücke sind hier fehlerfrei.

Der zweite Fall ist durch das Auftreten von Fehlern gekennzeichnet. Ein Modell hierfür ist die Poissonverteilung mit Parameter λ .

Sei Y die Anzahl der Fehler.

Wir wollen die Wahrscheinlichkeitsverteilung von Y herleiten.

Dazu führen wir eine mit dem Parameter p bernoulliverteilte Zufallsvariable U ein mit

$$U = \begin{cases} 1 & \text{wenn der Prozeß fehlerfrei ist} \\ 0 & \text{sonst} \end{cases}$$

Nun gilt

$$\begin{aligned}
 P(Y = y) &= P(Y = y, U = 1) + P(Y = y, U = 0) \\
 &= P(Y = y|U = 1)P(U = 1) + P(Y = y|U = 0)P(U = 0) \\
 &= \begin{cases} p + (1 - p)e^{-\lambda} & \text{für } y = 0 \\ (1 - p)\frac{\lambda^y}{y!}e^{-\lambda} & \text{sonst} \end{cases}
 \end{aligned}$$

Man spricht von der nullinflationierten Poissonverteilung.

Beispiel 6.3.5 In vielen Datensätzen tauchen sogenannte Ausreißer auf. Dies sind Beobachtungen, die stark von den anderen Beobachtungen abweichen. Es gibt unterschiedliche Ansätze zur Modellierung von Ausreißern, von denen wir hier einen näher betrachten wollen.

Wie schon im letzten Beispiel ist die Grundidee, daß eine Beobachtung aus einer von zwei möglichen Grundgesamtheiten kommen kann.

Die erste Grundgesamtheit ist eine Normalverteilung mit den Parametern μ und σ^2 , die andere Grundgesamtheit ist eine Normalverteilung mit den Parametern μ und $k \cdot \sigma^2$.

Die Auswahl wird durch ein Zufallsexperiment getroffen, wobei die Beobachtung mit Wahrscheinlichkeit $1 - \epsilon$ aus der ersten Grundgesamtheit und mit Wahrscheinlichkeit ϵ aus der zweiten Grundgesamtheit kommt. Dabei sollte epsilon in der Nähe von 0 liegen. Typische Werte sind $\epsilon = 0.05$ oder $\epsilon = 0.10$. Man spricht auch von der ϵ -kontaminierten Normalverteilung.

Wie sieht die Verteilungsfunktion und die Dichtefunktion der ϵ -kontaminierten Normalverteilung aus?

Wie auch schon im letzten Beispiel definieren wir eine Binärvariable U mit

$$U = \begin{cases} 1 & \text{wenn die erste Grundgesamtheit gewählt wird} \\ 0 & \text{wenn die zweite Grundgesamtheit gewählt wird} \end{cases}$$

Es gilt

$$P(U = 1) = 1 - \epsilon$$

und

$$P(U = 0) = \epsilon$$

Dann gilt für die Verteilungsfunktion $F_X(x)$ von X .

$$\begin{aligned} P(X \leq x) &= P(X \leq x, U = 1) + P(X \leq x, U = 0) \\ &= P(X \leq x|U = 1)P(U = 1) + P(X \leq x|U = 0)P(U = 0) \\ &= (1 - \epsilon)\Phi\left(\frac{x - \mu}{\sigma}\right) + \epsilon\Phi\left(\frac{x - \mu}{k\sigma}\right) \end{aligned}$$

Dabei ist $\Phi(z)$ die Verteilungsfunktion der Standardnormalverteilung an der Stelle z .

In den Beispielen ist die mischende Verteilung die Bernoulliverteilung. Die mischende Verteilung kann aber auch stetig sein.

Beispiel 6.3.6 *Wir betrachten die Anzahl der Schadensfälle einer versicherten Person. Diese wird bei einer Person einer Poissonverteilung mit Parameter λ folgen. Der Wert von λ wird aber nicht bei allen Personen gleich sein. Man spricht von Heterogenität. Wir fassen λ als Realisation einer Zufallsvariablen auf. Ein geeignetes Modell ist die Gammaverteilung mit den Parametern r und β . Die Anzahl der Schadensfälle einer versicherten Person kommt also dadurch zustande, daß zuerst eine Realisation λ von Λ und eine Realisation einer mit diesem λ poissonverteilten Zufallsvariablen beobachtet wird.*

Es gilt also

$$f_{Y|\Lambda}(y|\lambda) = \frac{\lambda^y}{y!} e^{-\lambda}$$

Die Dichtefunktion von Λ ist gegeben durch

$$f_{\Lambda}(\lambda) = \frac{\beta^r}{\Gamma(r)} \lambda^{r-1} e^{-\beta\lambda}$$

Somit gilt

$$\begin{aligned} f_{Y,\Lambda}(y, \lambda) &= f_{Y|\Lambda}(y|\lambda) f_{\Lambda}(\lambda) \\ &= \frac{\beta^r}{\Gamma(r)} \lambda^{r-1} e^{-\beta\lambda} \frac{\lambda^y}{y!} e^{-\lambda} \\ &= \frac{1}{y!} \frac{\beta^r}{\Gamma(r)} \lambda^{y+r-1} e^{-\lambda(\beta+1)} \end{aligned}$$

Die Randverteilung von Y erhalten wir dann durch

$$\begin{aligned}
 f_Y(y) &= \int_0^{\infty} f_{Y,\Lambda}(y, \lambda) d\lambda \\
 &= \int_0^{\infty} \frac{1}{y!} \frac{\beta^r}{\Gamma(r)} \lambda^{y+r-1} e^{-\lambda(\beta+1)} d\lambda \\
 &= \frac{1}{y!} \frac{\beta^r}{\Gamma(r)} \int_0^{\infty} \lambda^{y+r-1} e^{-\lambda(\beta+1)} d\lambda \\
 &= \frac{1}{y!} \frac{\beta^r}{\Gamma(r)} \frac{\Gamma(y+r)}{(\beta+1)^{y+r}} \\
 &= \left(\frac{\beta}{\beta+1}\right)^r \left(\frac{1}{\beta+1}\right)^y \frac{(y+r-1)!}{y!(r-1)!} \\
 &= \binom{y+r-1}{y} \left(\frac{\beta}{\beta+1}\right)^r \left(\frac{1}{\beta+1}\right)^y
 \end{aligned}$$

Y ist also pascalverteilt mit den Parametern r und $\frac{\beta}{\beta+1}$.

Das Beispiel zeigt, warum die Pascalverteilung oft als Modell für Zählvariablen verwendet wird.

Die "klassischen Modelle für Zählvariablen sind

- Binomialverteilung mit den Parametern n und p
- Poissonverteilung mit dem Parameter λ
- Pascalverteilung mit den Parametern r und p

Die Erwartungswert und Varianzen der Verteilungen sind
Binomialverteilung

$$\begin{aligned}
 E(X) &= np \\
 Var(X) &= np(1-p)
 \end{aligned}$$

Poissonverteilung

$$\begin{aligned}E(X) &= \lambda \\ \text{Var}(X) &= \lambda\end{aligned}$$

Pascalverteilung

$$\begin{aligned}E(X) &= r \frac{1-p}{p} \\ \text{Var}(X) &= r \frac{1-p}{p^2}\end{aligned}$$

Für die Poissonverteilung gilt

$$E(X) = \text{Var}(X)$$

Für die Binomialverteilung gilt

$$E(X) > \text{Var}(X)$$

da gilt

$$\text{Var}(X) = n p (1-p) = E(X) (1-p)$$

und $p < 1$ gilt.

Für die Pascalverteilung gilt

$$E(X) < \text{Var}(X)$$

da gilt

$$\text{Var}(X) = r \frac{1-p}{p^2} = \frac{E(X)}{p}$$

und $p < 1$ gilt.

6.4 Unabhängigkeit

Wie das Konzept der bedingten Wahrscheinlichkeit läßt sich auch das Konzept der Unabhängigkeit auf Zufallsvariablen übertragen.

Die Ereignisse A und B heißen unabhängig, wenn gilt

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

Sind X und Y diskrete Zufallsvariablen, erhalten wir mit

$$A = \{\omega | X(\omega) = x\}$$

und

$$B = \{\omega | Y(\omega) = y\} :$$

$$P(X = x, Y = y) = P(X = x) P(Y = y)$$

Definition 6.4.1 Die diskreten Zufallsvariablen X und Y heißen unabhängig, wenn für alle $(x, y) \in \mathfrak{R}^2$ gilt:

$$P(X = x, Y = y) = P(X = x) P(Y = y)$$

Im Fall stetiger Zufallsvariablen müssen wir anders vorgehen.

Wir setzen

$$A = \{\omega | X(\omega) \leq x\}$$

und

$$B = \{\omega | Y(\omega) \leq y\}$$

und erhalten:

$$P(X \leq x, Y \leq y) = P(X \leq x) P(Y \leq y)$$

Definition 6.4.2 Die Zufallsvariablen X und Y heißen unabhängig, wenn für alle $(x, y) \in \mathfrak{R}^2$ gilt:

$$F_{X,Y}(x, y) = F_X(x) F_Y(y)$$

Wegen

$$f_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y)$$

sind die stetigen Zufallsvariablen X und Y unabhängig, wenn gilt

$$f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y) \quad \text{für alle } (x, y) \in \mathfrak{R}^2$$

Beispiel 6.4.1 Für die zweidimensionale stetige Zufallsvariable (X, Y) mit gemeinsamer Dichtefunktion

$$f_{X,Y}(x, y) = \begin{cases} 1 & \text{für } 0 < x < 1, 0 < y < 1 \\ 0 & \text{sonst} \end{cases}$$

gilt

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

und

$$f_Y(y) = \begin{cases} 1 & \text{für } 0 < y < 1 \\ 0 & \text{sonst} \end{cases}$$

Somit gilt für alle $(x, y) \in \mathfrak{R}^2$

$$f_{X,Y}(x, y) = f_X(x) f_Y(y)$$

und die Zufallsvariablen X und Y sind unabhängig.

Da X und Y im letzten Beispiel nicht nur unabhängig sind, sondern auch die gleiche Verteilung haben, spricht man von **unabhängigen, identisch verteilten Zufallsvariablen**.

Im Rahmen der schließenden Statistik geht man in der Regel davon aus, daß die Beobachtungen x_1, \dots, x_n Realisationen von unabhängigen, identisch verteilten Zufallsvariablen X_1, \dots, X_n sind.

Im Falle der Unabhängigkeit benötigt man dann zur Bestimmung der gemeinsamen Verteilung nur noch die Randverteilungen.

Beispiel 6.4.2 *Aus zwei Urnen soll jeweils eine Kugel gezogen werden.*

Die Urnen weisen folgende Zusammensetzungen auf:

In der ersten Urne befinden sich 2 Kugeln, die 10 g wiegen und 3 Kugeln, die 20 g wiegen.

In der zweiten Urne befinden sich 4 Kugeln, die 10 g wiegen und 1 Kugel, die 20 g wiegen.

Sei X das Gewicht der aus Urne I gezogenen Kugel und Y das Gewicht der aus Urne II gezogenen Kugel.

Es gilt:

$$P(X = 10) = 0.4$$

$$P(X = 20) = 0.6$$

$$P(Y = 10) = 0.8$$

$$P(Y = 20) = 0.2$$

Aufgrund der Unabhängigkeit erhalten wir die gemeinsame Verteilung von X und Y ohne Probleme:

$$\begin{aligned} P(X = 10, Y = 10) &= P(X = 10) \cdot P(Y = 10) \\ &= 0.4 \cdot 0.8 \\ &= 0.32 \end{aligned}$$

$$\begin{aligned} P(X = 20, Y = 10) &= P(X = 20) \cdot P(Y = 10) \\ &= 0.6 \cdot 0.8 \\ &= 0.48 \end{aligned}$$

$$\begin{aligned} P(X = 10, Y = 20) &= P(X = 10) \cdot P(Y = 20) \\ &= 0.4 \cdot 0.2 \\ &= 0.08 \end{aligned}$$

$$\begin{aligned} P(X = 20, Y = 20) &= P(X = 20) \cdot P(Y = 20) \\ &= 0.6 \cdot 0.2 \\ &= 0.12 \end{aligned}$$

In der folgenden Tabelle sehen wir die gemeinsame Wahrscheinlichkeitsfunktion:

| y | 10 | 20 | |
|-----|------|------|------|
| x | | | |
| 10 | 0.32 | 0.08 | 0.40 |
| 20 | 0.48 | 0.12 | 0.60 |
| | 0.80 | 0.20 | 1.00 |

Die Zufallsvariablen X und Y des Beispiels sind unabhängig, sie besitzen aber nicht die gleiche Verteilung. Dies erreichen wir, indem wir aus der gleichen Urne zweimal mit Zurücklegen ziehen.

Beispiel 6.4.3 In einer Urne befinden sich 2 Kugeln, die 10 g wiegen und 3 Kugeln, die 20 g wiegen.

Es werden zwei Kugeln mit Zurücklegen gezogen.

Sei X das Gewicht der ersten gezogenen Kugel und Y das Gewicht der zweiten gezogenen Kugel.

Es gilt:

$$P(X = 10) = 0.4$$

$$P(X = 20) = 0.6$$

$$P(Y = 10) = 0.4$$

$$P(Y = 20) = 0.6$$

Aufgrund der Unabhängigkeit erhalten wir die gemeinsame Verteilung von X und Y ohne Probleme:

$$\begin{aligned} P(X = 10, Y = 10) &= P(X = 10) \cdot P(Y = 10) \\ &= 0.4 \cdot 0.4 \\ &= 0.16 \end{aligned}$$

$$\begin{aligned} P(X = 20, Y = 10) &= P(X = 20) \cdot P(Y = 10) \\ &= 0.6 \cdot 0.4 \\ &= 0.24 \end{aligned}$$

$$\begin{aligned}
 P(X = 10, Y = 20) &= P(X = 10) \cdot P(Y = 20) \\
 &= 0.4 \cdot 0.6 \\
 &= 0.24
 \end{aligned}$$

$$\begin{aligned}
 P(X = 20, Y = 20) &= P(X = 20) \cdot P(Y = 20) \\
 &= 0.6 \cdot 0.6 \\
 &= 0.36
 \end{aligned}$$

In der folgenden Tabelle sehen wir die gemeinsame Wahrscheinlichkeitsfunktion:

| y | 10 | 20 | |
|-----|------|------|------|
| x | | | |
| 10 | 0.16 | 0.24 | 0.40 |
| 20 | 0.24 | 0.36 | 0.60 |
| | 0.40 | 0.60 | 1.00 |

Und was passiert, wenn man ohne Zurücklegen zieht?

Beispiel 6.4.4 In einer Urne befinden sich 2 Kugeln, die 10 g wiegen und 3 Kugeln, die 20 g wiegen.

Es werden zwei Kugeln ohne Zurücklegen gezogen.

Sei X das Gewicht der ersten gezogenen Kugel und Y das Gewicht der zweiten gezogenen Kugel.

Für den ersten Zug gilt:

$$P(X = 10) = 0.4$$

$$P(X = 20) = 0.6$$

Wir müssen beim zweiten Zug das Ergebnis des ersten Zuges berücksichtigen.

$$P(Y = 10|X = 10) = 0.25$$

$$P(Y = 20|X = 10) = 0.75$$

$$P(Y = 10|X = 20) = 0.50$$

$$P(Y = 20|X = 20) = 0.50$$

Somit gilt für die gemeinsame Wahrscheinlichkeitsfunktion:

$$\begin{aligned} P(X = 10, Y = 10) &= P(Y = 10|X = 10) \cdot P(X = 10) \\ &= 0.25 \cdot 0.4 \\ &= 0.10 \end{aligned}$$

$$\begin{aligned} P(X = 20, Y = 10) &= P(Y = 20|X = 10) \cdot P(X = 10) \\ &= 0.75 \cdot 0.4 \\ &= 0.30 \end{aligned}$$

$$\begin{aligned} P(X = 10, Y = 20) &= P(Y = 10|X = 20) \cdot P(X = 10) \\ &= 0.50 \cdot 0.6 \\ &= 0.30 \end{aligned}$$

$$\begin{aligned} P(X = 20, Y = 20) &= P(Y = 20|X = 20) \cdot P(X = 20) \\ &= 0.50 \cdot 0.6 \\ &= 0.30 \end{aligned}$$

In der folgenden Tabelle sehen wir die gemeinsame Wahrscheinlichkeitsfunktion:

| | | | |
|-----|------|------|------|
| y | 10 | 20 | |
| x | | | |
| 10 | 0.10 | 0.30 | 0.40 |
| 20 | 0.30 | 0.30 | 0.60 |
| | 0.40 | 0.60 | 1.00 |

Wenn wir also mit unabhängigen, identisch verteilten Zufallsvariablen arbeiten wollen, müssen wir mit Zurücklegen ziehen.

Beim genauen Betrachten der Randverteilungen fällt auf, daß beim Ziehen aus der gleichen Urne die Randverteilungen beim Ziehen mit Zurücklegen und beim Ziehen ohne Zurücklegen identisch sind.

6.5 Die bivariate Normalverteilung

Das wichtigste Verteilungsmodell für eine stetige zweidimensionale Zufallsvariable ist die bivariate Normalverteilung.

Definition 6.5.1 Die Zufallsvariable (X, Y) heißt *bivariat normalverteilt* mit den Parametern $\mu_X, \mu_Y, \sigma_X, \sigma_Y$ und ρ , wenn ihre Dichtefunktion lautet

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_X)^2}{\sigma_X^2} - 2\rho\frac{x-\mu_X}{\sigma_X}\frac{y-\mu_Y}{\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right)}$$

Die Höhenlinien sind bei der bivariaten Normalverteilung Ellipsen.

Ist

$$\rho = 0$$

so sind die Ellipsen achsenparallel.

Ist

$$\sigma_X = \sigma_Y$$

und

$$\rho = 0$$

so sind die Höhenlinien Kreise.

Ist

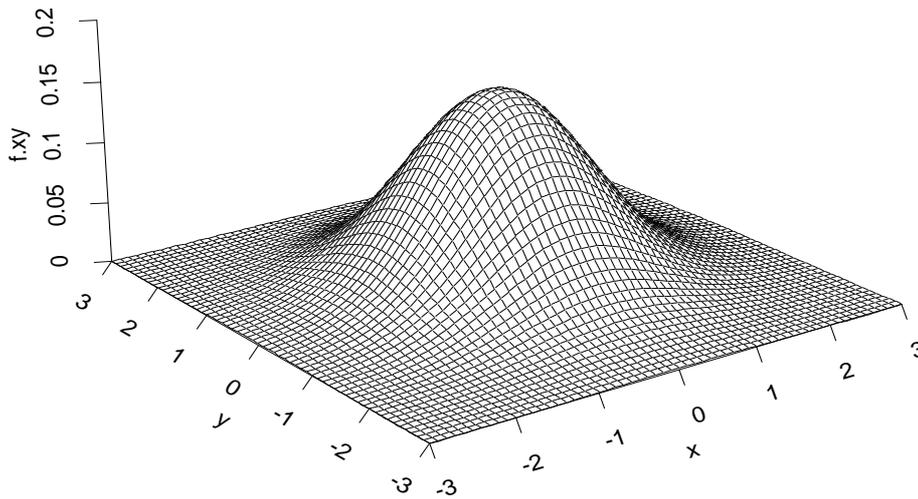
$$\sigma_X > \sigma_Y$$

und

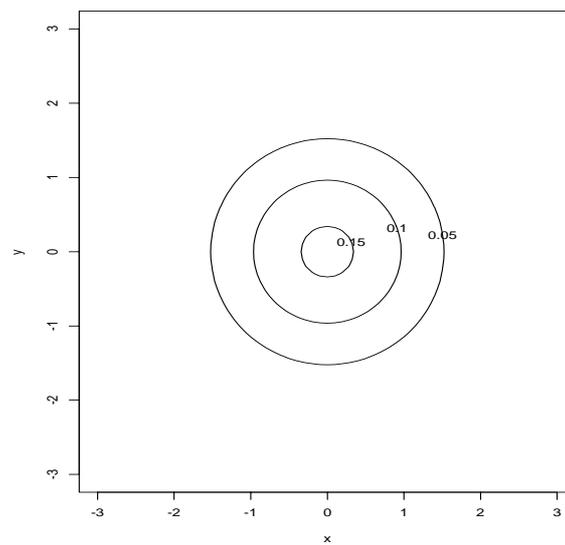
$$\rho = 0$$

so ist die Ausrichtung der Ellipse in X-Richtung größer als in Y-Richtung.

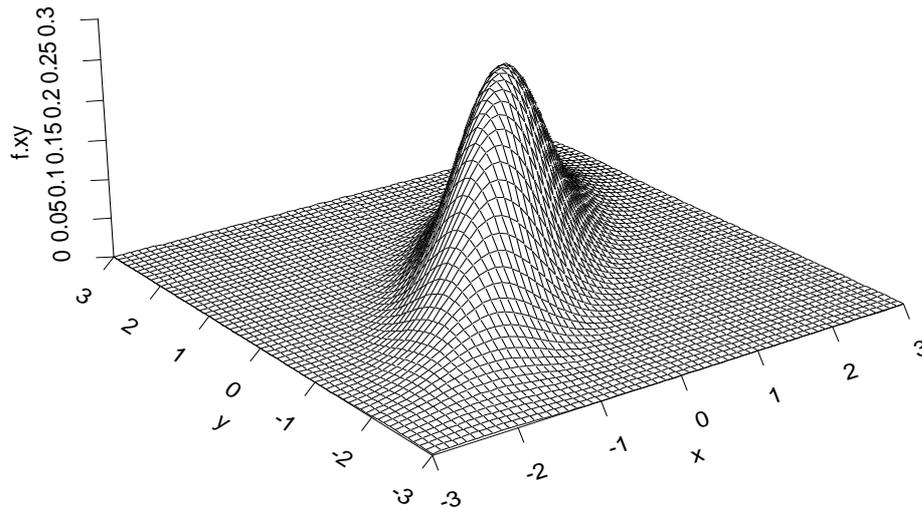
Das folgende Bild zeigt die Dichtefunktion der bivariaten Normalverteilung mit den Parametern $\mu_X = 0$, $\mu_Y = 0$, $\sigma_X = 1$, $\sigma_Y = 1$ und $\rho = 0$:



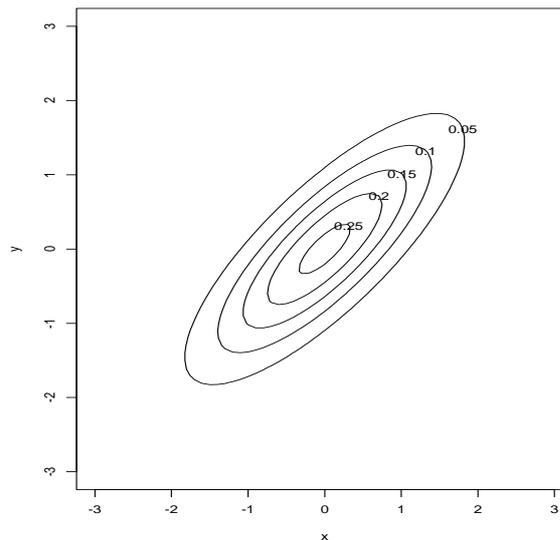
Das folgende Bild zeigt die zugehörigen Höhenlinien.



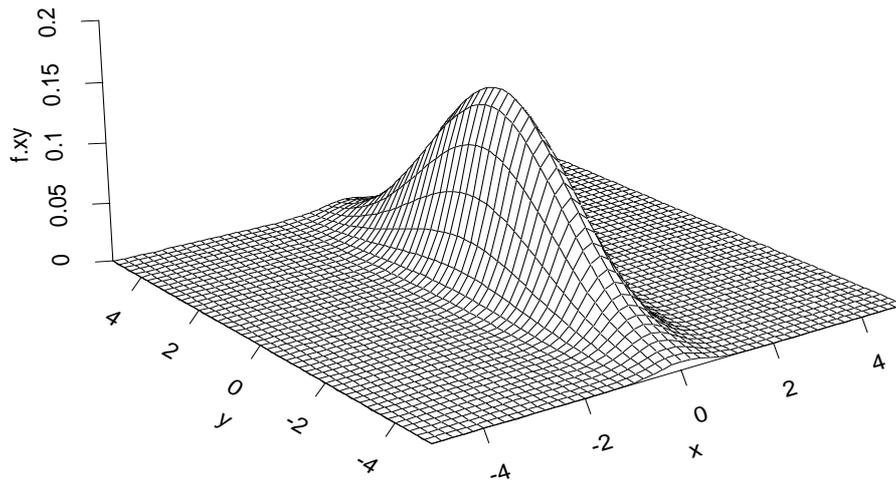
Das folgende Bild zeigt die Dichtefunktion der bivariaten Normalverteilung mit den Parametern $\mu_X = 0$, $\mu_Y = 0$, $\sigma_X = 1$, $\sigma_Y = 1$ und $\rho = 0.8$:



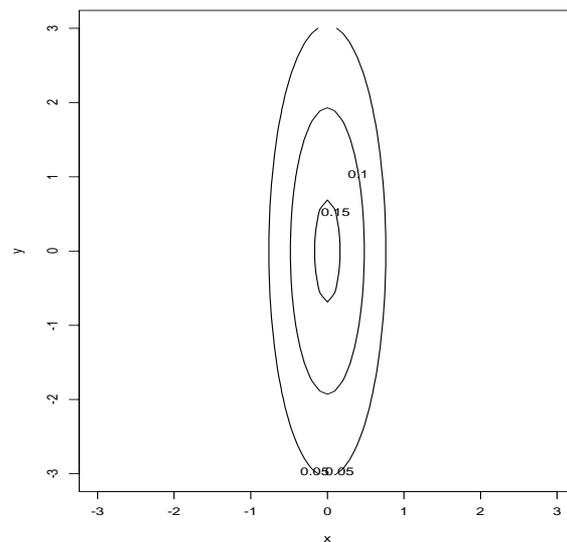
Das folgende Bild zeigt die zugehörigen Höhenlinien.



Das folgende Bild zeigt die Dichtefunktion der bivariaten Normalverteilung mit den Parametern $\mu_X = 0$, $\mu_Y = 0$, $\sigma_X = 0.5$, $\sigma_Y = 2$ und $\rho = 0$:



Das folgende Bild zeigt die zugehörigen Höhenlinien.



Satz 6.5.1 Die Zufallsvariable (X, Y) sei bivariat normalverteilt mit den Parametern $\mu_X, \mu_Y, \sigma_X, \sigma_Y$ und ρ .

Dann gilt

1. X ist normalverteilt mit den Parametern μ_X und σ_X^2 .
2. Y ist normalverteilt mit den Parametern μ_Y und σ_Y^2 .
3. $f_{Y|X}(y|x)$ ist die Dichtefunktion einer Normalverteilung mit den Parametern

$$\mu_Y - \rho \frac{\sigma_Y}{\sigma_X} (x - \mu_X)$$

und

$$\sigma_Y^2 (1 - \rho^2).$$

4. $f_{X|Y}(x|y)$ ist die Dichtefunktion einer Normalverteilung mit den Parametern

$$\mu_X - \rho \frac{\sigma_X}{\sigma_Y} (y - \mu_Y)$$

und

$$\sigma_X^2 (1 - \rho^2).$$

Beweis:

Es gilt

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)} \left(\frac{(x-\mu_X)^2}{\sigma_X^2} - 2\rho \frac{x-\mu_X}{\sigma_X} \frac{y-\mu_Y}{\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} \right)}$$

Wir betrachten zunächst den Ausdruck:

$$e^{-\frac{1}{2(1-\rho^2)} \left(\frac{(x-\mu_X)^2}{\sigma_X^2} - 2\rho \frac{x-\mu_X}{\sigma_X} \frac{y-\mu_Y}{\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} \right)}$$

Wir setzen

$$u = \frac{x - \mu_X}{\sigma_X}$$

und

$$v = \frac{y - \mu_Y}{\sigma_Y}$$

und erhalten:

$$\begin{aligned} e^{-\frac{1}{2(1-\rho^2)}(u^2 - 2\rho uv + v^2)} &= e^{-\frac{1}{2(1-\rho^2)}(u^2 - (\rho u)^2 + (\rho u)^2 - 2\rho uv + v^2)} \\ &= e^{-\frac{1}{2(1-\rho^2)}((1-\rho^2)u^2 + (v-\rho u)^2)} \\ &= e^{-\frac{u^2}{2}} e^{-\frac{(v-\rho u)^2}{2(1-\rho^2)}} \end{aligned}$$

Also gilt

$$f_{X,Y}(x, y) = \frac{1}{\sqrt{2\pi} \sigma_X} e^{-\frac{(x-\mu_X)^2}{2\sigma_X^2}} \frac{1}{\sqrt{2\pi} \sigma_Y \sqrt{1-\rho^2}} e^{-\frac{1}{2\sigma_Y^2(1-\rho^2)} \left(y - \mu_Y - \rho \frac{\sigma_Y}{\sigma_X} (x - \mu_X) \right)^2}$$

Es gilt also

$$f_{X,Y}(x, y) = g(x) h(x, y)$$

mit

$$g(x) = \frac{1}{\sqrt{2\pi} \sigma_X} e^{-\frac{(x-\mu_X)^2}{2\sigma_X^2}}$$

und

$$h(x, y) = \frac{1}{\sqrt{2\pi} \sigma_Y \sqrt{1-\rho^2}} e^{-\frac{1}{2\sigma_Y^2(1-\rho^2)} \left(y - \mu_Y - \rho \frac{\sigma_Y}{\sigma_X} (x - \mu_X) \right)^2}$$

Die Funktion $g(x)$ ist die Dichtefunktion einer univariaten Normalverteilung mit den Parametern μ_X und σ_X^2 , während $h(x, y)$ für festes x die Dichtefunktion einer univariaten Normalverteilung mit den Parametern

$$\mu_Y - \rho \frac{\sigma_Y}{\sigma_X} (x - \mu_X)$$

und

$$\sigma_Y^2 (1 - \rho^2)$$

ist.

Für festes x ist das Integral über $h(x, y)$ mit y als Integrationsvariablen gleich 1. Also ist die Randdichtefunktion von X eine Normalverteilung mit den Parametern μ_X und σ_X^2 .

Wegen

$$h(x, y) = \frac{f_{X,Y}(x, y)}{f_X(x)}$$

gilt

$$h(x, y) = f_{Y|X}(y|x)$$

Vertauschen wir X und Y , so sehen wir, daß die Randdichtefunktion von Y eine Normalverteilung mit den Parametern μ_Y und σ_Y^2 ist.

Satz 6.5.2 Die Zufallsvariable (X, Y) sei bivariat normalverteilt mit den Parametern $\mu_X, \mu_Y, \sigma_X, \sigma_Y$ und ρ .

Wenn $\rho = 0$ gilt, dann sind X und Y unabhängig.

Beweis:

Es gilt

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_X)^2}{\sigma_X^2} - 2\rho\frac{x-\mu_X}{\sigma_X}\frac{y-\mu_Y}{\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right)}$$

Ist $\rho = 0$, so folgt

$$\begin{aligned} f_{X,Y}(x, y) &= \frac{1}{2\pi\sigma_X\sigma_Y} e^{-\frac{1}{2}\left(\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right)} \\ &= \frac{1}{\sigma_X\sqrt{2\pi}} e^{-\frac{(x-\mu_X)^2}{2\sigma_X^2}} \frac{1}{\sigma_Y\sqrt{2\pi}} e^{-\frac{(y-\mu_Y)^2}{2\sigma_Y^2}} \end{aligned}$$

Die gemeinsame Dichtefunktion von X und Y ist also das Produkt der Dichtefunktion einer mit den Parametern μ_X und σ_X^2 normalverteilten Zufallsvariablen und der Dichtefunktion einer mit den Parametern μ_Y und σ_Y^2 normalverteilten Zufallsvariablen.

Dies sind aber gerade die Randdichten von X und Y .

Ist also bei der bivariaten Normalverteilung der Parameter ρ gleich 0, so sind die beiden Komponenten unabhängig. Wir werden später sehen, wie man den Parameter ρ interpretieren kann.

Kapitel 7

Funktionen mehrdimensionaler Zufallsvariablen

7.1 Funktionen von diskreten Zufallsvariablen

Bei einer eindimensionalen Zufallsvariablen X haben wir die Verteilung einer Funktion $g(X)$ betrachtet.

In diesem Kapitel werden wir uns damit beschäftigen, wie man die Verteilung einer Funktion $g(X, Y)$ einer zweidimensionalen Zufallsvariablen (X, Y) bestimmt.

Sei (X, Y) also eine zweidimensionale Zufallsvariable mit Wahrscheinlichkeitsfunktion $P(X = x, Y = y)$.

Wir suchen die Wahrscheinlichkeitsfunktion $P(Z = z)$ der Funktion $Z = g(X, Y)$.

Beginnen wir mit einem Beispiel.

Beispiel 7.1.1 *Wir betrachten wiederum das Teilchen, das sich zufällig auf den ganzen Zahlen bewegt und 3 Schritte macht.*

Wir betrachten zunächst die Zufallsvariablen

X Anzahl der Schritte nach rechts

Y Anzahl der Schritte nach links.

Die folgende Tabelle zeigt, welche Werte den einzelnen Ergebnissen durch X und Y zugeordnet werden.

| ω | x | y |
|----------|-----|-----|
| RRR | 3 | 0 |
| RRL | 2 | 1 |
| RLR | 2 | 1 |
| LRR | 2 | 1 |
| RLL | 1 | 2 |
| LRL | 1 | 2 |
| LLR | 1 | 2 |
| LLL | 0 | 3 |

Wir erhalten somit folgende gemeinsame Wahrscheinlichkeitsfunktion von X und Y .

| y | 0 | 1 | 2 | 3 |
|-----|---------|---------|---------|---------|
| x | | | | |
| 0 | 0 | 0 | 0 | 0.125 |
| 1 | 0 | 0 | 0.375 | 0 |
| 2 | 0 | 0.375 | 0 | 0 |
| 3 | 0.125 | 0 | 0 | 0 |

Die Position Z des Teilchens erhalten wir, indem wir von der Anzahl X der Schritte nach rechts die Anzahl Y der Schritte nach links subtrahieren.

Es gilt also $Z = X - Y$.

Offensichtlich nimmt Z den Wert 3 an, wenn X den Wert 3 und Y den Wert 0 annimmt.

Somit gilt

$$\begin{aligned} P(Z = 3) &= P(X = 3, Y = 0) \\ &= 0.125 \end{aligned}$$

Analog erhalten wir

$$\begin{aligned} P(Z = 1) &= P(X = 2, Y = 1) = 0.375 \\ P(Z = -1) &= P(X = 1, Y = 2) = 0.375 \\ P(Z = -3) &= P(X = 0, Y = 3) = 0.125 \end{aligned}$$

Das Beispiel zeigt die allgemeine Vorgehensweise:

Suchen wir Wahrscheinlichkeit eines Wertes z von Z , so suchen wir alle Urbilder (x, y) von z unter g , bestimmen deren Wahrscheinlichkeiten und summieren diese auf.

Es ist

$$P(Z = z) = P(g(X, Y) = z) = \sum_{\{(x, y) | g(x, y) = z\}} P(X = x, Y = y)$$

Ein wichtiger Spezialfall ist die Summe Z von X und Y .

Es gilt

$$\begin{aligned} P(Z = z) &= P(X + Y = z) \\ &= \sum_{\{(x, y) | x + y = z\}} P(X = x, Y = y) \\ &= \sum_y P(X = z - y, Y = y) \\ &= \sum_x P(X = x, Y = z - x) \end{aligned}$$

Dabei werden die Summen über alle zulässigen Realisationsmöglichkeiten gebildet.

Wir schauen uns zwei wichtige Anwendungsbeispiele an:

Satz 7.1.1 *Die Zufallsvariablen X und Y seien unabhängig, wobei X poissonverteilt mit Parameter λ und Y poissonverteilt mit Parameter μ ist.*

Dann ist

$$Z = X + Y$$

poissonverteilt mit Parameter

$$\lambda + \mu.$$

Beweis:

$$\begin{aligned} P(Z = z) &= \sum_{y=0}^z P(X = z - y, Y = y) \\ &= \sum_{y=0}^z P(X = z - y) P(Y = y) \\ &= \sum_{y=0}^z \frac{\lambda^{z-y}}{(z-y)!} e^{-\lambda} \frac{\mu^y}{y!} e^{-\mu} \\ &= \frac{1}{z!} e^{-(\lambda+\mu)} \sum_{y=0}^z \frac{z!}{y!(z-y)!} \mu^y \lambda^{z-y} \\ &= \frac{1}{z!} e^{-(\lambda+\mu)} \sum_{y=0}^z \binom{z}{y} \mu^y \lambda^{z-y} \\ &= \frac{(\mu + \lambda)^z}{z!} e^{-(\lambda+\mu)} \end{aligned}$$

da gilt

$$\sum_{i=0}^n \binom{n}{i} a^i b^{n-i} = (a + b)^n$$

Die Summe von unabhängigen poissonverteilten Zufallsvariablen ist also poissonverteilt, wobei der Parameter der Summe gleich der Summe der Parameter der Summanden ist.

Speziell folgt, daß die Summe von n unabhängigen, identisch mit Parameter λ poissonverteilten Zufallsvariablen mit dem Parameter $n\lambda$ poissonverteilt ist.

Satz 7.1.2 Die Zufallsvariablen X und Y seien unabhängig, wobei X mit den Parametern m und p und Y mit den Parametern n und p binomialverteilt ist.

Dann ist

$$Z = X + Y$$

binomialverteilt mit Parametern

$$m + n$$

und

$$p.$$

Beweis:

$$\begin{aligned} P(Z = z) &= \sum_{y=0}^z P(X = z - y, Y = y) \\ &= \sum_{y=0}^z P(X = z - y)P(Y = y) \\ &= \sum_{y=0}^z \binom{m}{z - y} p^{z-y} (1-p)^{m-z+y} \binom{n}{y} p^y (1-p)^{n-y} \\ &= p^z (1-p)^{m+n-z} \sum_{y=0}^z \binom{m}{z - y} \binom{n}{y} \\ &= \binom{m+n}{z} p^z (1-p)^{m+n-z} \end{aligned}$$

da gilt

$$\sum_{y=0}^z \binom{m}{z - y} \binom{n}{y} = \binom{m+n}{z}$$

Daß die Aussage des Satzes richtig ist, kann man sich auch so klarmachen: Beobachtet man zwei unabhängige Bernoulli-Prozesse der Länge m bzw. n mit gleicher Erfolgswahrscheinlichkeit p , so ist die Anzahl der Erfolge bei beiden Prozessen zusammen binomialverteilt mit den Parametern $m + n$ und p .

Speziell folgt aus dem Satz noch, daß die Summe von n unabhängigen, identisch mit dem Parameter p bernoulliverteilten Zufallsvariablen mit den Parametern n und p binomialverteilt ist.

Für die schließende Statistik sind zwei Konsequenzen der beiden Sätze wichtig

- Sind die Zufallsvariablen X_1, X_2, \dots, X_n unabhängig und identisch mit dem Parameter λ poissonverteilt, so ist

$$Y = X_1 + X_2 + \dots + X_n$$

poissonverteilt mit dem Parameter $n \lambda$.

- Sind die Zufallsvariablen X_1, X_2, \dots, X_n unabhängig und identisch mit dem Parameter p bernoulliverteilt, so ist

$$Y = X_1 + X_2 + \dots + X_n$$

binomialverteilt mit den Parametern n und p .

Schauen wir uns noch einmal die Beispiele 6.4.3 und 6.4.4 an.

Beispiel 7.1.2 *In einer Urne befinden sich 2 Kugeln, die 10 g wiegen und 3 Kugeln, die 20 g wiegen.*

Es werden zwei Kugeln mit Zurücklegen gezogen.

Sei X das Gewicht der ersten gezogenen Kugel und Y das Gewicht der zweiten gezogenen Kugel.

In der folgenden Tabelle sehen wir die gemeinsame Wahrscheinlichkeitsfunktion:

| | | | |
|-----|------|------|------|
| y | 10 | 20 | |
| x | | | |
| 10 | 0.16 | 0.24 | 0.40 |
| 20 | 0.24 | 0.36 | 0.60 |
| | 0.40 | 0.60 | 1.00 |

Wir bestimmen die Wahrscheinlichkeitsfunktion von $Z = X + Y$.

Es gilt

$$\begin{aligned} P(Z = 20) &= P(X = 10, Y = 10) \\ &= 0.16 \end{aligned}$$

$$\begin{aligned} P(Z = 30) &= P(X = 10, Y = 20) + P(X = 20, Y = 10) \\ &= 0.24 + 0.24 \\ &= 0.48 \end{aligned}$$

$$\begin{aligned} P(Z = 40) &= P(X = 20, Y = 20) \\ &= 0.36 \end{aligned}$$

Und was passiert, wenn man ohne Zurücklegen zieht?

Beispiel 7.1.3 *In einer Urne befinden sich 2 Kugeln, die 10 g wiegen und 3 Kugeln, die 20 g wiegen.*

Es werden zwei Kugeln ohne Zurücklegen gezogen.

Sei X das Gewicht der ersten gezogenen Kugel und Y das Gewicht der zweiten gezogenen Kugel.

In der folgenden Tabelle sehen wir die gemeinsame Wahrscheinlichkeitsfunktion:

| | | | | |
|-----|-----|------|------|------|
| | y | 10 | 20 | |
| x | | | | |
| 10 | | 0.10 | 0.30 | 0.40 |
| 20 | | 0.30 | 0.30 | 0.60 |
| | | 0.40 | 0.60 | 1.00 |

Wir bestimmen die Wahrscheinlichkeitsfunktion von $Z = X + Y$.

Es gilt

$$\begin{aligned} P(Z = 20) &= P(X = 10, Y = 10) \\ &= 0.10 \end{aligned}$$

$$\begin{aligned} P(Z = 30) &= P(X = 10, Y = 20) + P(X = 20, Y = 10) = \\ &= 0.30 + 0.30 \\ &= 0.60 \end{aligned}$$

$$\begin{aligned} P(Z = 40) &= P(X = 20, Y = 20) \\ &= 0.30 \end{aligned}$$

Wodurch unterscheidet sich die Verteilung von $X + Y$ beim Ziehen mit Zurücklegen von der Verteilung von $X + Y$ beim Ziehen ohne Zurücklegen?

In der folgenden Tabelle sind die Verteilungen gegenübergestellt:

| z | $P(Z=z)$ m.Z. | $P(Z=z)$ o.Z. |
|-----|------------------|------------------|
| 20 | 0.16 | 0.10 |
| 30 | 0.48 | 0.60 |
| 40 | 0.36 | 0.30 |

Es fällt auf, daß beim Ziehen ohne Zurücklegen extreme Werte seltener auftreten als beim Ziehen mit Zurücklegen. Dies liegt daran, daß die extremen Stichproben $(10, 10)$ und $(20, 20)$ beim Ziehen ohne Zurücklegen seltener auftreten.

Schauen wir uns auch den Erwartungswert und die Varianz von $X + Y$ in den beiden Situationen für das Beispiel an.

Ziehen mit Zurücklegen

$$\begin{aligned} E(X + Y) &= 20 \cdot 0.16 + 30 \cdot 0.48 + 40 \cdot 0.36 \\ &= 32 \end{aligned}$$

Ziehen ohne Zurücklegen

$$\begin{aligned} E(X + Y) &= 20 \cdot 0.10 + 30 \cdot 0.60 + 40 \cdot 0.30 \\ &= 32 \end{aligned}$$

Die Erwartungswerte unterscheiden sich nicht, hingegen die Varianzen.

Schauen wir uns die beiden Fälle an:

Ziehen mit Zurücklegen

$$\begin{aligned} E((X + Y)^2) &= 400 \cdot 0.16 + 900 \cdot 0.48 + 1600 \cdot 0.36 \\ &= 1072 \end{aligned}$$

Also gilt

$$\begin{aligned} \text{Var}(X + Y) &= 1072 - 32^2 \\ &= 48 \end{aligned}$$

Ziehen ohne Zurücklegen

$$\begin{aligned} E((X + Y)^2) &= 400 \cdot 0.10 + 900 \cdot 0.60 + 1600 \cdot 0.30 \\ &= 1060 \end{aligned}$$

Also gilt

$$\begin{aligned} \text{Var}(X + Y) &= 1060 - 32^2 \\ &= 36 \end{aligned}$$

Wie zu erwarten war, ist die Varianz beim Ziehen ohne Zurücklegen kleiner. Schauen wir uns an, wie der Erwartungswert bzw. die Varianz der Summe von den Erwartungswerten bzw. Varianzen der Summanden abhängen.

Beim Ziehen mit Zurücklegen gilt:

$$\begin{aligned} E(X) &= E(Y) \\ &= 10 \cdot 0.4 + 20 \cdot 0.6 \\ &= 16 \end{aligned}$$

Wegen

$$E(X + Y) = 32$$

gilt

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

Weiterhin gilt

$$\begin{aligned} E(X^2) &= E(Y^2) \\ &= 100 \cdot 0.4 + 400 \cdot 0.6 \\ &= 280 \end{aligned}$$

Also gilt

$$\text{Var}(X) = \text{Var}(Y) = 24$$

Wegen

$$\text{Var}(X + Y) = 48$$

gilt

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

Beim Ziehen ohne Zurücklegen stimmen die Randverteilungen von X und Y mit denen von X und Y beim Ziehen mit Zurücklegen überein.

Offensichtlich gilt auch hier

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

Wegen

$$\text{Var}(X + Y) = 36$$

gilt hier nicht

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

Wodurch unterscheiden sich die beiden Situationen?

Beim Ziehen mit Zurücklegen sind X und Y unabhängig.

Im Beispiel gilt

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

und

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

Beim Ziehen ohne Zurücklegen sind X und Y nicht unabhängig.

Im Beispiel gilt

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

und

$$\text{Var}(X + Y) \neq \text{Var}(X) + \text{Var}(Y)$$

Wir werden im folgenden Kapitel sehen, daß der Erwartungswert einer Summe gleich der Summe der Erwartungswerte ist, wenn alle Erwartungswerte existieren.

Außerdem wird dann bewiesen, daß die Varianz einer Summe von unabhängigen Zufallsvariablen gleich der Summe der Varianzen ist.

7.2 Funktionen von stetigen Zufallsvariablen

Die Dichtefunktion der Funktion $Y = g(X)$ einer univariaten Zufallsvariablen X erhält man durch:

$$f_Y(y) = \left| \frac{d}{dy} g^{-1}(y) \right| f_X(g^{-1}(y))$$

Dabei muß die Funktion $g(x)$ monoton und differenzierbar sein.

Den bivariaten Fall kann man dem folgenden Satz entnehmen, den wir ohne Beweis angeben. Eine sehr schöne Herleitung ist in dem Buch von Blake zu finden.

Satz 7.2.1 *Sei (X, Y) eine stetige zweidimensionale Zufallsvariable mit Dichtefunktion $f_{X,Y}(x, y)$.*

Sei $T_{X,Y}$ der Träger von (X, Y) .

Es gilt also

$$T_{X,Y} = \{(x, y) | f_{X,Y}(x, y) > 0\}$$

Seien

$$g_1 : T_{X,Y} \rightarrow \mathfrak{R}$$

mit

$$(x, y) \mapsto u = g_1(x, y)$$

und

$$g_2 : T_{X,Y} \rightarrow \mathfrak{R}$$

mit

$$(x, y) \mapsto v = g_2(x, y)$$

Abbildungen.

Sei

$$T_{U,V} = \{(u, v) | u = g_1(x, y), v = g_2(x, y), (x, y) \in T_{X,Y}\}$$

Die Abbildung

$$g : T_{X,Y} \rightarrow T_{U,V}$$

mit

$$(x, y) \mapsto u = g_1(x, y)$$

und

$$(x, y) \mapsto v = g_2(x, y)$$

sei bijektiv, so daß die inverse Abbildung mit

$$x = h_1(u, v)$$

und

$$y = h_2(u, v)$$

existiert.

Die Abbildungen g_1, g_2, h_1 und h_2 seien stetig und die partiellen Ableitungen

$$\frac{\partial}{\partial u} h_1(u, v) \quad \frac{\partial}{\partial v} h_1(u, v)$$

$$\frac{\partial}{\partial u} h_2(u, v) \quad \frac{\partial}{\partial v} h_2(u, v)$$

mögen existieren und stetig sein.

Sei J die Jacobi-Determinante der inversen Abbildung :

$$J = \det \begin{pmatrix} \frac{\partial}{\partial u} h_1(u, v) & \frac{\partial}{\partial v} h_1(u, v) \\ \frac{\partial}{\partial u} h_2(u, v) & \frac{\partial}{\partial v} h_2(u, v) \end{pmatrix}$$

Ist J von Null verschieden, so gilt

$$f_{U,V}(u, v) = |J| f_{X,Y}(h_1(u, v), h_2(u, v))$$

Ist die zweidimensionale stetige Zufallsvariable (X, Y) gegeben, so legt der Satz folgende Vorgehensweise zur Bestimmung der gemeinsamen Verteilung von

$$U = g_1(X, Y)$$

und

$$V = g_2(X, Y)$$

nahe.

1. Bilde die inverse Abbildung mit

$$x = h_1(u, v)$$

und

$$y = h_2(u, v)$$

2. Bestimme den Träger $T_{U,V}$ von (U, V) .

3. Bilde die partiellen Ableitungen

$$\frac{\partial}{\partial u} h_1(u, v) \quad \frac{\partial}{\partial v} h_1(u, v)$$

$$\frac{\partial}{\partial u} h_2(u, v) \quad \frac{\partial}{\partial v} h_2(u, v)$$

4. Berechne die Jacobi-Determinante

$$J = \det \begin{pmatrix} \frac{\partial}{\partial u} h_1(u, v) & \frac{\partial}{\partial v} h_1(u, v) \\ \frac{\partial}{\partial u} h_2(u, v) & \frac{\partial}{\partial v} h_2(u, v) \end{pmatrix}$$

5. Bilde

$$f_{U,V}(u, v) = |J| f_{X,Y}(h_1(u, v), h_2(u, v))$$

Schauen wir uns einige Beispiele an.

Beispiel 7.2.1 Die Zufallsvariablen X und Y seien unabhängig und identisch auf $(0, 1)$ gleichverteilt.

Die gemeinsame Dichtefunktion lautet also:

$$\begin{aligned} f_{X,Y}(x, y) &= f_X(x) f_Y(y) \\ &= \begin{cases} 1 & \text{für } 0 < x < 1, 0 < y < 1 \\ 0 & \text{sonst} \end{cases} \end{aligned}$$

Gesucht ist die gemeinsame Dichtefunktion von

$$\begin{aligned} U &= g_1(X, Y) = X + Y \\ V &= g_2(X, Y) = X - Y. \end{aligned}$$

Wir bestimmen zunächst die inverse Transformation.

Dazu lösen wir das Gleichungssystem

$$\begin{aligned} u &= x + y \\ v &= x - y \end{aligned}$$

nach x und y auf.

Wir erhalten

$$\begin{aligned}x &= h_1(u, v) = 0.5u + 0.5v \\y &= h_2(u, v) = 0.5u - 0.5v.\end{aligned}$$

Nun bestimmen wir den Träger von U und V .

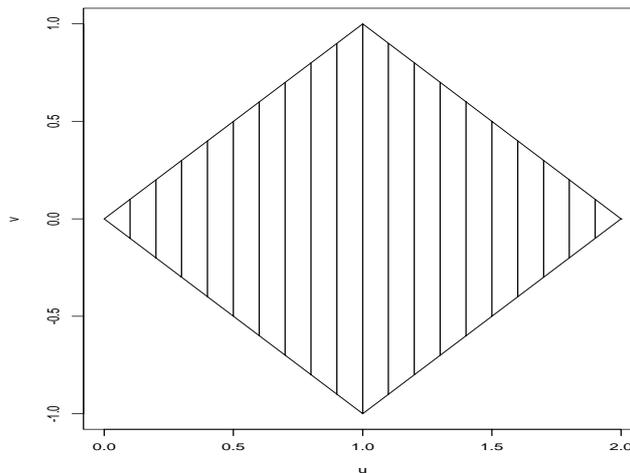
Aus

$$\begin{aligned}0 &< x < 1 \\0 &< y < 1\end{aligned}$$

folgt

$$\begin{aligned}0 &< u + v < 2 \\0 &< u - v < 2\end{aligned}$$

Das folgende Bild zeigt den Träger von (U, V) :



Wir bestimmen die partiellen Ableitungen.

Es gilt

$$\begin{aligned}\frac{\partial}{\partial u} h_1(u, v) &= \frac{\partial}{\partial u} 0.5u + 0.5v \\&= 0.5 \\ \frac{\partial}{\partial v} h_1(u, v) &= \frac{\partial}{\partial v} 0.5u + 0.5v \\&= 0.5\end{aligned}$$

$$\begin{aligned}\frac{\partial}{\partial u} h_2(u, v) &= \frac{\partial}{\partial u} 0.5 u - 0.5 v \\ &= 0.5\end{aligned}$$

$$\begin{aligned}\frac{\partial}{\partial v} h_2(u, v) &= \frac{\partial}{\partial v} 0.5 u - 0.5 v \\ &= -0.5\end{aligned}$$

Somit ist die Jacobi-Matrix

$$\begin{aligned}J &= \det \begin{bmatrix} \frac{\partial}{\partial u} h_1(u, v) & \frac{\partial}{\partial v} h_1(u, v) \\ \frac{\partial}{\partial u} h_2(u, v) & \frac{\partial}{\partial v} h_2(u, v) \end{bmatrix} \\ &= \det \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & -0.5 \end{bmatrix} \\ &= -0.5\end{aligned}$$

Somit gilt

$$|J| = 0.5.$$

Also gilt

$$f_{U,V}(u, v) = \begin{cases} 0.5 & \text{für } 0 < u + v < 2, 0 < u - v < 2 \\ 0 & \text{sonst} \end{cases}$$

Wir schauen uns nun noch die Randverteilungen von U und V an.

Für die Randdichte von U gilt für $0 < u < 1$:

$$\begin{aligned}f_U(u) &= \int_{-u}^u 0.5 \, dv \\ &= u\end{aligned}$$

und für $1 \leq u < 2$:

$$\begin{aligned}f_U(u) &= \int_{u-2}^{2-u} 0.5 \, dv \\ &= 2 - u\end{aligned}$$

Also gilt

$$f_U(u) = \begin{cases} u & \text{für } 0 < u < 1 \\ 2 - u & \text{für } 1 \leq u < 2 \\ 0 & \text{sonst} \end{cases}$$

U ist also dreiecksverteilt auf $(0, 2)$.

Für die Randdichte von V gilt für $-1 < v < 0$:

$$\begin{aligned} f_V(v) &= \int_{-v}^{2+v} 0.5 \, du \\ &= 1 + v \end{aligned}$$

und für $0 \leq v < 1$:

$$\begin{aligned} f_V(v) &= \int_v^{2-v} 0.5 \, du \\ &= 1 - v \end{aligned}$$

Also gilt

$$f_V(v) = \begin{cases} 1 + v & \text{für } -1 < v < 0 \\ 1 - v & \text{für } 0 \leq v < 1 \\ 0 & \text{sonst} \end{cases}$$

V ist also dreiecksverteilt auf $(-1, 1)$.

Beispiel 7.2.2 Die Zufallsvariablen X und Y seien unabhängig und identisch mit Parameter λ exponentialverteilt.

Die gemeinsame Dichtefunktion lautet also:

$$\begin{aligned} f_{X,Y}(x, y) &= f_X(x) f_Y(y) \\ &= \begin{cases} \lambda^2 e^{-\lambda(x+y)} & \text{für } x, y > 0 \\ 0 & \text{sonst} \end{cases} \end{aligned}$$

Gesucht ist die gemeinsame Dichtefunktion von

$$\begin{aligned} U &= g_1(X, Y) = X + Y \\ V &= g_2(X, Y) = X. \end{aligned}$$

Die inverse Transformation lautet

$$\begin{aligned} x &= h_1(u, v) = v \\ y &= h_2(u, v) = u - v. \end{aligned}$$

Somit gilt

$$\begin{aligned} J &= \det \begin{bmatrix} \frac{\partial}{\partial u} h_1(u, v) & \frac{\partial}{\partial v} h_1(u, v) \\ \frac{\partial}{\partial u} h_2(u, v) & \frac{\partial}{\partial v} h_2(u, v) \end{bmatrix} \\ &= \det \begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix} \\ &= -1 \end{aligned}$$

Somit gilt

$$|J| = 1.$$

Aus

$$\begin{aligned} x &> 0 \\ y &> 0 \end{aligned}$$

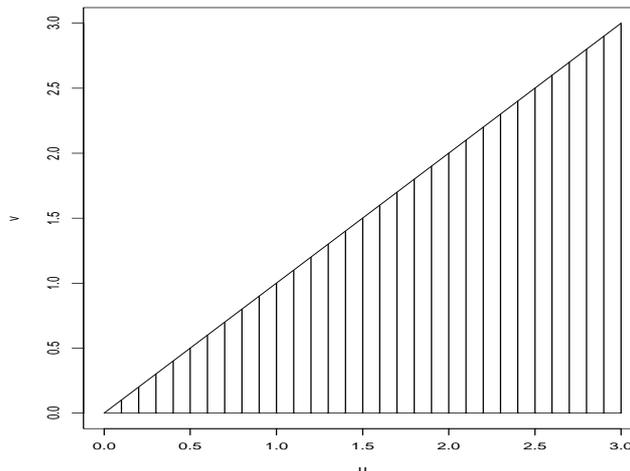
folgt mit

$$\begin{aligned} x &= v \\ y &= u - v. \end{aligned}$$

also

$$u > v > 0.$$

Das folgende Bild zeigt einen Teil des Trägers von (U, V) :



Also gilt für $u > v > 0$:

$$\begin{aligned} f_{U,V}(u, v) &= |J| f_{X,Y}(v, u - v) \\ &= \lambda^2 e^{-\lambda(v+u-v)} \\ &= \lambda^2 e^{-\lambda u} \end{aligned}$$

Ansonsten ist die gemeinsame Dichtefunktion von U und V gleich Null.

Für die Randdichte von U gilt:

$$\begin{aligned} f_U(u) &= \int_0^u \lambda^2 e^{-\lambda u} dv \\ &= \lambda^2 u e^{-\lambda u} \end{aligned}$$

Dies ist die Dichtefunktion einer Gammaverteilung mit Parametern $r = 2$ und λ .

Das Beispiel zeigt, wie man mit Hilfe der zweidimensionalen Funktion einfach die Verteilung einer Funktion $g_1(X, Y)$ bestimmen kann.

Man setzt

$$g_2(X, Y) = X$$

oder

$$g_2(X, Y) = Y.$$

Beispiel 7.2.3 Die Zufallsvariablen X und Y seien unabhängig und identisch standardnormalverteilt.

Die gemeinsame Dichtefunktion lautet also:

$$f_{X,Y}(x, y) = \frac{1}{2\pi} e^{-0.5(x^2+y^2)}$$

Gesucht ist die Dichtefunktion von

$$U = g_1(X, Y) = \frac{X}{Y}.$$

Wir setzen

$$V = g_2(X, Y) = Y.$$

Die inverse Transformation lautet

$$\begin{aligned} x &= h_1(u, v) = uv \\ y &= h_2(u, v) = v. \end{aligned}$$

Somit gilt

$$\begin{aligned} J &= \det \begin{bmatrix} \frac{\partial}{\partial u} h_1(u, v) & \frac{\partial}{\partial v} h_1(u, v) \\ \frac{\partial}{\partial u} h_2(u, v) & \frac{\partial}{\partial v} h_2(u, v) \end{bmatrix} \\ &= \det \begin{bmatrix} v & u \\ 0 & 1 \end{bmatrix} \\ &= v. \end{aligned}$$

Somit gilt

$$|J| = |v|.$$

Aus

$$x \in \mathfrak{R}$$

und

$$y \in \mathfrak{R}$$

folgt

$$u \in \mathfrak{R}$$

und

$$v \in \mathfrak{R}.$$

Also gilt:

$$\begin{aligned} f_{U,V}(u, v) &= |J| f_{X,Y}(u v, v) \\ &= \frac{1}{2\pi} |v| e^{-0.5 u^2 v^2 - 0.5 v^2} \end{aligned}$$

Für die Randdichte von U gilt:

$$\begin{aligned} f_U(u) &= \int_{-\infty}^{\infty} \frac{1}{2\pi} |v| e^{-0.5 u^2 v^2 - 0.5 v^2} dv \\ &= \int_{-\infty}^0 \frac{1}{2\pi} (-v) e^{-0.5 u^2 v^2 - 0.5 v^2} dv + \int_0^{\infty} \frac{1}{2\pi} v e^{-0.5 u^2 v^2 - 0.5 v^2} dv \\ &= 2 \int_0^{\infty} \frac{1}{2\pi} v e^{-0.5 (u^2+1) v^2} dv \\ &= \frac{1}{\pi} \int_0^{\infty} v e^{-0.5 (u^2+1) v^2} dv \\ &= \frac{1}{\pi} \left[-\frac{1}{1+u^2} e^{-0.5 (u^2+1) v^2} \right]_0^{\infty} \\ &= \frac{1}{\pi (1+u^2)}. \end{aligned}$$

Dies ist die Dichtefunktion der Cauchyverteilung.

Der Quotient von zwei unabhängigen standardnormalverteilten Zufallsvariablen ist also cauchyverteilt.

Beispiel 7.2.4 Die Zufallsvariablen X und Y seien unabhängig, wobei X mit Parameter λ exponentialverteilt ist und Y mit Parameter μ exponentialverteilt ist.

Die gemeinsame Dichtefunktion lautet also:

$$f_{X,Y}(x,y) = \begin{cases} \lambda \mu e^{-\lambda x - \mu y} & \text{für } x, y > 0 \\ 0 & \text{sonst} \end{cases}$$

Gesucht ist die gemeinsame Dichtefunktion von

$$U = g_1(X, Y) = Y - X$$

$$V = g_2(X, Y) = X.$$

Die inverse Transformation lautet

$$x = h_1(u, v) = v$$

$$y = h_2(u, v) = u + v.$$

Somit gilt

$$\begin{aligned} J &= \det \begin{bmatrix} \frac{\partial}{\partial u} h_1(u, v) & \frac{\partial}{\partial v} h_1(u, v) \\ \frac{\partial}{\partial u} h_2(u, v) & \frac{\partial}{\partial v} h_2(u, v) \end{bmatrix} \\ &= \det \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \\ &= -1 \end{aligned}$$

Somit gilt

$$|J| = 1.$$

Aus

$$x > 0$$

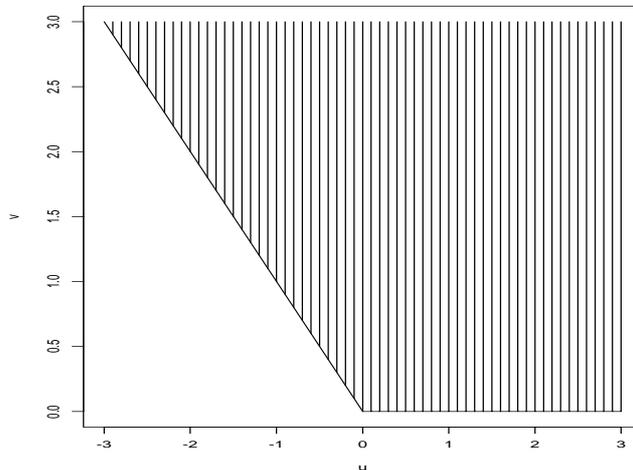
$$y > 0$$

folgt

$$u + v > 0$$

$$v > 0$$

Das folgende Bild zeigt einen Teil des Trägers von (U, V) :



Also gilt für $u + v > 0$ und $v > 0$:

$$\begin{aligned} f_{U,V}(u, v) &= |J| f_{X,Y}(v, u + v) \\ &= \lambda \mu e^{-\lambda v - \mu(u+v)} \\ &= \lambda \mu e^{-(\lambda+\mu)v - \mu u} \end{aligned}$$

Ansonsten ist die gemeinsame Dichtefunktion von U und V gleich Null.
Für die Randdichte von U gilt für $u \geq 0$:

$$\begin{aligned} f_U(u) &= \int_0^{\infty} \lambda \mu e^{-(\lambda+\mu)v - \mu u} dv \\ &= \lambda \mu e^{-\mu u} \left[-\frac{1}{\lambda + \mu} e^{-(\lambda+\mu)v} \right]_0^{\infty} \\ &= \frac{\lambda \mu}{\lambda + \mu} e^{-\mu u} \end{aligned}$$

und für $u < 0$

$$\begin{aligned}
 f_U(u) &= \int_{-u}^{\infty} \lambda \mu e^{-(\lambda+\mu)v - \mu u} dv \\
 &= \lambda \mu e^{-\mu u} \left[-\frac{1}{\lambda + \mu} e^{-(\lambda+\mu)v} \right]_{-u}^{\infty} \\
 &= \lambda \mu e^{-\mu u} \frac{1}{\lambda + \mu} e^{(\lambda+\mu)u} \\
 &= \frac{\lambda \mu}{\lambda + \mu} e^{\lambda u}
 \end{aligned}$$

Also gilt

$$f_U(u) = \begin{cases} \frac{\lambda \mu}{\lambda + \mu} e^{\lambda u} & \text{für } u < 0 \\ \frac{\lambda \mu}{\lambda + \mu} e^{-\mu u} & \text{für } u \geq 0 \end{cases}$$

Für

$$\lambda = \mu = 1$$

gilt

$$f_U(u) = \begin{cases} \frac{1}{2} e^u & \text{für } u < 0 \\ \frac{1}{2} e^{-u} & \text{für } u \geq 0 \end{cases}$$

Dafür können wir auch schreiben

$$f_U(u) = 0.5 e^{-|u|}.$$

Dies ist die Dichtefunktion einer Doppelsexponentialverteilung.

7.3 Verteilung des Maximums und des Minimums

Eine Kette ist nur so stark wie ihr schwächstes Glied. Deshalb spielen in der Zuverlässigkeitstheorie die Verteilung des Maximums und des Minimums eine zentrale Rolle.

Satz 7.3.1 *Seien X_1, \dots, X_n unabhängige, identisch mit stetiger Verteilungsfunktion $F_X(x)$ verteilte Zufallsvariablen.*

Dann ist die Verteilungsfunktion von

$$V = \min\{X_1, \dots, X_n\} :$$

gegeben durch

$$F_V(v) = 1 - (1 - F_X(v))^n$$

und die Verteilungsfunktion von

$$W = \max\{X_1, \dots, X_n\} :$$

gegeben durch

$$F_W(w) = F_X(w)^n$$

Beweis:

Für V gilt

$$\begin{aligned} F_V(v) &= P(V \leq v) \\ &= 1 - P(V \geq v) \\ &= 1 - P(\min\{X_1, \dots, X_n\} \geq v) \\ &= 1 - P(\text{alle } X_i \geq v) \\ &= 1 - P(X_1 \geq v, \dots, X_n \geq v) \\ &= 1 - P(X_1 \geq v) \cdots P(X_n \geq v) \\ &= 1 - (1 - P(X_1 \leq v)) \cdots (1 - P(X_n \leq v)) \\ &= 1 - (1 - F_{X_1}(v)) \cdots (1 - F_{X_n}(v)) \\ &= 1 - (1 - F_X(v))^n \end{aligned}$$

Für W gilt

$$\begin{aligned}
 F_W(w) &= P(W \leq w) \\
 &= P(\max\{X_1, \dots, X_n\} \leq w) \\
 &= P(\text{alle } X_i \leq w) \\
 &= P(X_1 \leq w, \dots, X_n \leq w) \\
 &= P(X_1 \leq w) \cdots P(X_n \leq w) \\
 &= F_X(w)^n
 \end{aligned}$$

Eine große Rolle im Rahmen der Zuverlässigkeitstheorie spielt die Exponentialverteilung.

Satz 7.3.2 Die Zufallsvariablen X_1, \dots, X_n seien unabhängig, wobei X_i mit Parameter λ_i exponentialverteilt ist.

Dann ist

$$V = \min\{X_1, \dots, X_n\}$$

exponentialverteilt mit Parameter

$$\sum_{i=1}^n \lambda_i.$$

Beweis:

Für $v > 0$ gilt:

$$\begin{aligned}
 F_V(v) &= P(V \leq v) \\
 &= 1 - P(V \geq v) \\
 &= 1 - P(\min\{X_1, \dots, X_n\} \geq v) \\
 &= 1 - P(\text{alle } X_i \geq v) \\
 &= 1 - P(X_1 \geq v, \dots, X_n \geq v) \\
 &= 1 - P(X_1 \geq v) \cdots P(X_n \geq v) \\
 &= 1 - (1 - P(X_1 \leq v)) \cdots (1 - P(X_n \leq v)) \\
 &= 1 - (1 - F_{X_1}(v)) \cdots (1 - F_{X_n}(v)) \\
 &= 1 - (1 - (1 - e^{-\lambda_1 v})) \cdots (1 - (1 - e^{-\lambda_n v})) \\
 &= 1 - e^{-\lambda_1 v} \cdots e^{-\lambda_n v} \\
 &= 1 - e^{-(\lambda_1 + \dots + \lambda_n) v}
 \end{aligned}$$

Ansonsten gilt

$$F_V(v) = 0.$$

Dies ist aber die Verteilungsfunktion der Exponentialverteilung mit Parameter

$$\sum_{i=1}^n \lambda_i.$$

Im Rahmen von Warteschlangenmodellen betrachtet man oft zwei unabhängige Poissonprozesse mit Parametern λ beziehungsweise μ .

Bei einem Schalter werden die Wartezeiten zwischen den Ankünften von Kunden als mit Parameter λ unabhängige exponentialverteilte Zufallsvariablen modelliert.

Die Bedienungszeiten werden als mit Parameter μ unabhängige exponentialverteilte Zufallsvariablen modelliert.

Es wird unterstellt, daß Ankunftsprozeß und Bedienungsprozeß unabhängig sind.

Die Wartezeit bis zum ersten Kunden ist also exponentialverteilt mit Parameter λ .

Nachdem der erste Kunde angekommen ist, warten wir auf das nächste Ereignis.

Für dieses gibt es zwei Möglichkeiten:

1. Die Bedienung des ersten Kunden endet.
2. Ein neuer Kunde kommt an.

Die Wartezeit X auf das erste Ereignis ist exponentialverteilt mit Parameter μ , während die Wartezeit Y auf das zweite Ereignis exponentialverteilt ist mit Parameter λ .

Die minimale Wartezeit $\min X, Y$ bis zum Eintreten eines von beiden ist exponentialverteilt mit Parameter $\lambda + \mu$.

Wie groß ist nun die Wahrscheinlichkeit, daß das erste Ereignis zuerst eintritt, daß also der erste Kunde bedient ist ehe der zweite Kunde ankommt?

Gesucht ist also

$$P(X = \min\{X, Y\})$$

Es gilt

$$P(X = \min\{X, Y\}) = \frac{\lambda}{\lambda + \mu}$$

Dies sieht man folgendermaßen:

$$\begin{aligned} P(X = \min\{X, Y\}) &= P(X < Y) \\ &= P(Y - X > 0) \end{aligned}$$

In Satz 7.2.4 wurde die Verteilung von $Y - X$ hergeleitet.
Die Dichtefunktion lautet

$$f_U(u) = \begin{cases} \frac{\lambda\mu}{\lambda+\mu} e^{\lambda u} & \text{für } u < 0 \\ \frac{\lambda\mu}{\lambda+\mu} e^{-\mu u} & \text{für } u \geq 0 \end{cases}$$

Die gesuchte Wahrscheinlichkeit erhalten wir durch

$$\begin{aligned} P(Y - X > 0) &= \int_0^{\infty} \frac{\lambda\mu}{\lambda+\mu} e^{-\mu u} du \\ &= \left[\frac{\lambda\mu}{\lambda+\mu} \frac{1}{\mu} e^{-\mu u} \right]_0^{\infty} \\ &= \frac{\lambda}{\lambda+\mu} \end{aligned}$$

Man kann nun die ersten zwei Ereignisse eines solches System folgendermaßen simulieren:

Das erste Ereignis muß eine Ankunft sein. Also erzeugt man eine mit Parameter λ exponentialverteilte Zufallszahl.

Das zweite Ereignis ist entweder die Ende der Bedienung des ersten Kunden oder die Ankunft eines neuen Kunden.

Da die Wartezeit seit dem ersten Ereignis bis zum Eintreten dieses Ereignisses eine mit dem Parameter $\lambda + \mu$ exponentialverteilte Zufallszahl ist, erzeugen wir eine derartige.

Hierdurch ist der Zeitpunkt des Eintretens des zweiten Ereignisses gegeben. Wir wissen nur noch, ob es sich um das Bedienungsende oder eine neue Ankunft handelt.

Die Wahrscheinlichkeit, daß es sich um eine Ankunft handelt, beträgt

$$\frac{\lambda}{\lambda + \mu}$$

Wir erzeugen nun eine auf $(0, 1)$ gleichverteilte Zufallszahl.

Ist diese kleiner als

$$\frac{\lambda}{\lambda + \mu}$$

so handelt es sich um eine Ankunft, ansonsten um das Bedienungsende.

Wir schauen uns noch die Verteilung des Minimums und Maximums einer Zufallsstichprobe aus einer Gleichverteilung an.

Satz 7.3.3 *Die Zufallsvariablen X_1, \dots, X_n seien unabhängig und identisch auf $(0, 1)$ gleichverteilt.*

Dann ist die Verteilungsfunktion von

$$V = \min\{X_1, \dots, X_n\} :$$

gegeben durch

$$F_V(v) = \begin{cases} 1 - (1 - v)^n & \text{für } v > 0 \\ 0 & \text{sonst} \end{cases}$$

und die Verteilungsfunktion von

$$W = \max\{X_1, \dots, X_n\} :$$

gegeben durch

$$F_W(w) = \begin{cases} w^n & \text{für } w > 0 \\ 0 & \text{sonst} \end{cases}$$

Beweis:

Für X_i gilt

$$F_{X_i}(x_i) = \begin{cases} x_i & \text{für } w > 0 \\ 0 & \text{sonst} \end{cases}$$

Also folgt

für V :

$$\begin{aligned} F_V(v) &= 1 - (1 - F_X(v))^n \\ &= 1 - (1 - v)^n \end{aligned}$$

und für W :

$$\begin{aligned} F_W(w) &= F_X(w)^n \\ &= w^n \end{aligned}$$

Somit gilt für die Dichtefunktion von V :

$$f_V(v) = \begin{cases} n(1-v)^{n-1} & \text{für } 0 < v < 1 \\ 0 & \text{sonst} \end{cases}$$

und für die Dichtefunktion von W :

$$f_W(w) = \begin{cases} n w^{n-1} & \text{für } 0 < w < 1 \\ 0 & \text{sonst} \end{cases}$$

Das Minimum von n auf $(0, 1)$ gleichverteilten Zufallsvariablen ist also mit den Parametern $a = 1$ und $b = n$ betaverteilt, während das Maximum mit den Parametern $a = n$ und $b = 1$ betaverteilt ist.

7.4 Funktionen von unabhängigen Zufallsvariablen

Bei vielen Anwendungen geht man aus von unabhängigen Zufallsvariablen X_1, \dots, X_n und fragt sich, welcher Zusammenhang zwischen $g_1(X_1), \dots, g_n(X_n)$ besteht.

Der folgende Satz zeigt, daß Funktionen von unabhängigen Zufallsvariablen ebenfalls unabhängig sind.

Satz 7.4.1 *Die Zufallsvariablen X und Y seien unabhängig.*

Sei T_X der Träger von X und T_Y der Träger von Y .

Seien $g : T_X \rightarrow \mathfrak{R}$ und $h : T_Y \rightarrow \mathfrak{R}$ Abbildungen.

Dann sind $V = g(X)$ und $W = h(Y)$ unabhängig.

Beweis:

Wir beweisen den diskreten Fall.

$$\begin{aligned}
 P(g(X) = v, h(Y) = w) &= P(V = v, W = w) \\
 &= P(g(X) = v, h(Y) = w) \\
 &= \sum_{\{x|g(x)=v\}} \sum_{\{y|h(y)=w\}} P(X = x, Y = y) \\
 &= \sum_{\{x|g(x)=v\}} \sum_{\{y|h(y)=w\}} P(X = x)P(Y = y) \\
 &= \sum_{\{x|g(x)=v\}} P(X = x) \sum_{\{y|h(y)=w\}} P(Y = y) \\
 &= P(g(X) = v) P(h(Y) = w)
 \end{aligned}$$

Beispiel 7.4.1 *Es soll überprüft werden, ob der Median M einer Grundgesamtheit mit stetiger Verteilungsfunktion $F_X(x)$ den Wert 0 annimmt.*

Hierzu wird eine Zufallsstichprobe vom Umfang n aus der Grundgesamtheit gezogen. Wir beobachten also die Realisationen x_1, \dots, x_n der unabhängigen, identisch verteilten Zufallsvariablen X_1, \dots, X_n .

Zu testen ist

$$H_0 : M = 0$$

$$H_1 : M \neq 0$$

Ein geeigneter Test ist der Vorzeichentest, dessen Teststatistik S lautet:

$$S = \sum_{i=1}^n s(X_i)$$

mit

$$s(X_i) = \begin{cases} 1 & \text{für } X_i > 0 \\ 0 & \text{sonst} \end{cases}$$

S ist also die Anzahl der Beobachtungen, die größer als 0 ist.

Ist nun 0 der Wert des Medians M , so erwartet man, daß die Hälfte der Beobachtungen größer als 0 ist.

Sind sehr viele oder sehr wenige Beobachtungen größer als 0, so spricht dies dagegen, daß 0 der Median der Grundgesamtheit ist.

Ab welcher Anzahl soll man sich für die Gegenhypothese entscheiden?

Dies hängt vom vorgegebenen Signifikanzniveau α ab und der Verteilung von S , wenn H_0 zutrifft.

Ist 0 der Median der Grundgesamtheit, so gilt

$$P(X_i > 0) = 0.5.$$

Da die X_i unabhängig sind, sind auch die $s(X_i)$ unabhängig.

Also sind $s(X_1), \dots, s(X_n)$ unabhängige, identisch mit Parameter $p = 0.5$ bernoulliverteilte Zufallsvariablen.

Also ist S binomialverteilt mit den Parametern n und $p = 0.5$.

Kapitel 8

Parameter multivariater Verteilungen

8.1 Erwartungswerte

Wir können auch bei mehrdimensionalen Zufallsvariablen den Erwartungswert betrachten.

Definition 8.1.1 Sei $X = (X_1, X_2, \dots, X_k)$ eine mehrdimensionalen Zufallsvariable.

Dann ist der Erwartungswert $E(X)$ definiert durch

$$E(X) = \begin{pmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_k) \end{pmatrix}$$

Wir müssen also nur den Erwartungswert jeder Komponente bestimmen.

Beispiel 8.1.1 Sei (X, Y) bivariat normalverteilt mit den Parametern μ_X , μ_Y , σ_X , σ_Y und ρ .

Dann gilt

$$E(X) = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}$$

In der schließenden Statistik sind die Erwartungswerte von Funktionenn $g(X_1, X_2, \dots, X_k)$ von mehrdimensionalen Zufallsvariablen (X_1, X_2, \dots, X_k) von zentraler Bedeutung.

Im eindimensionalen Fall haben wir gesehen, daß man den Erwartungswert einer Funktion $g(X)$ der Zufallsvariablen X mit der Verteilung von X bestimmen kann.

Im diskreten Fall gilt

$$E(g(X)) = \sum_x g(x) P(X = x)$$

und im stetigen Fall

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x) dx$$

Dies legt folgende Definition nahe:

Definition 8.1.2 Sei (X, Y) eine zweidimensionale Zufallsvariable mit Wahrscheinlichkeitsfunktion $P(X = x, Y = y)$ bzw. Dichtefunktion $f_{X,Y}(x, y)$ und $g : \mathfrak{R}^2 \rightarrow \mathfrak{R}$ eine Funktion.

Dann ist im diskreten Fall

$$E(g(X, Y)) = \sum_x \sum_y g(x, y) P(X = x, Y = y) \quad (8.1)$$

und im stetigen Fall

$$E(g(X, Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{X,Y}(x, y) dx dy$$

In letzten Kapitel haben wir an zwei Beispielen gesehen, daß gilt

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

Wie der folgende Satz zeigt, gilt dies allgemein.

Satz 8.1.1 Sei (X, Y) eine zweidimensionale Zufallsvariable mit Wahrscheinlichkeitsfunktion $P(X = x, Y = y)$ bzw. Dichtefunktion $f_{X,Y}(x, y)$.

Falls $E(X)$ und $E(Y)$ existieren, gilt

$$E(aX + bY) = aE(X) + bE(Y)$$

Beweis:

Wir beweisen den diskreten Fall:

$$\begin{aligned} E(aX + bY) &= \sum_x \sum_y (ax + by) P(X = x, Y = y) = \\ &= \sum_x \sum_y (ax P(X = x, Y = y) + by P(X = x, Y = y)) \\ &= \sum_x \sum_y ax P(X = x, Y = y) + \sum_x \sum_y by P(X = x, Y = y) \\ &= a \sum_x \sum_y x P(X = x, Y = y) + b \sum_x \sum_y y P(X = x, Y = y) \\ &= a \sum_x x \sum_y P(X = x, Y = y) + b \sum_y y \sum_x P(X = x, Y = y) \\ &= a \sum_x x P(X = x) + b \sum_y y P(Y = y) = \\ &= aE(X) + bE(Y) \end{aligned}$$

Aus dem Satz folgt, daß der Erwartungswert einer Summe von Zufallsvariablen gleich der Summe der Erwartungswerte ist, wenn diese existieren.

In der schliessenden Statistik betrachten wir in der Regel unabhängige und identisch verteilte Zufallsvariablen X_1, X_2, \dots, X_n .

Es gelte

$$E(X_i) = \mu$$

für $i = 1, \dots, n$.

Dann gilt

$$E\left(\sum_{i=1}^n X_i\right) = n\mu$$

Dies sieht man folgendermaßen:

$$\begin{aligned} E\left(\sum_{i=1}^n X_i\right) &= \sum_{i=1}^n E(X_i) \\ &= \sum_{i=1}^n \mu \\ &= n\mu \end{aligned}$$

In der schließenden Statistik geht man aus von einem Verteilungsmodell mit unbekanntem Parameter θ .

Aus den Realisationen x_1, x_2, \dots, x_n der Zufallsvariablen X_1, X_2, \dots, X_n sucht man einen geeigneten Wert für den Parameter θ .

Sinnvollerweise bildet man eine Funktion $g(x_1, x_2, \dots, x_n)$ der Beobachtungen x_1, x_2, \dots, x_n .

Vor der Realisation der Stichprobe hängt diese Funktion von den Zufallsvariablen X_1, X_2, \dots, X_n ab und ist somit eine Zufallsvariable $g(X_1, X_2, \dots, X_n)$.

Man nennt sie auch Schätzfunktion

$$\hat{\theta} = g(X_1, X_2, \dots, X_n)$$

Schätzfunktionen sollten gute Eigenschaften besitzen.

Eine dieser Eigenschaften ist die Erwartungstreue.

Definition 8.1.3 *Eine Schätzfunktion $\hat{\theta}$ heißt erwartungstreue Schätzfunktion von θ , wenn für alle Werte von θ gilt*

$$E(\hat{\theta}) = \theta$$

Im Falle einer Zufallsstichprobe ist das arithmetische Mittel

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

eine erwartungstreue Schätzfunktion für den Erwartungswert μ .

Dies sieht man folgendermaßen:

$$\begin{aligned} E(\bar{X}) &= E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \\ &= \frac{1}{n} E\left(\sum_{i=1}^n X_i\right) \\ &= \frac{1}{n} n \mu \\ &= \mu \end{aligned}$$

Wir betrachten noch einmal das Beispiel

Beispiel 8.1.2 *Wir betrachten folgende zweidimensionale stetige Zufallsvariable (X, Y) mit gemeinsamer Dichtefunktion*

$$f_{X,Y}(x, y) = \begin{cases} 1 & \text{für } 0 < x < 1, 0 < y < 1 \\ 0 & \text{sonst} \end{cases}$$

Gesucht ist $E(XY)$.

Es gilt

$$\begin{aligned} E(XY) &= \int_0^1 \int_0^1 xy \, dx dy \\ &= \int_0^1 0.5 y \, dy \\ &= 0.25 \end{aligned}$$

Wir haben gesehen, daß X und Y unabhängige, auf $(0, 1)$ gleichverteilte Zufallsvariablen sind.

Es gilt somit

$$E(X) = E(Y) = 0.5.$$

Somit gilt für das Beispiel

$$E(XY) = E(X) E(Y).$$

Dies ist kein Zufall:

Satz 8.1.2 *Seien X und Y unabhängige Zufallsvariablen.
Dann gilt*

$$E(XY) = E(X) E(Y)$$

Beweis:

Wir beweisen den diskreten Fall:

$$\begin{aligned} E(XY) &= \sum_x \sum_y xy P(X = x, Y = y) \\ &= \sum_x \sum_y xy P(X = x) P(Y = y) \\ &= \sum_x x P(X = x) \sum_y y P(Y = y) \\ &= E(X) E(Y) \end{aligned}$$

8.2 Kovarianz und Korrelationskoeffizient

Wie wir im letzten Abschnitt gesehen haben, gilt immer

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

Gilt auch

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

Schauen wir uns ein Beispiel an.

Beispiel 8.2.1 *Ein Teilchen bewegt sich auf den ganzen Zahlen. Bei jedem Schritt geht es entweder nach links oder nach rechts.*

Seien

$$X = \begin{cases} 1 & \text{wenn es beim ersten Schritt nach links geht} \\ -1 & \text{wenn es beim ersten Schritt nach rechts geht} \end{cases}$$

und

$$Y = \begin{cases} 1 & \text{wenn es beim zweiten Schritt nach links geht} \\ -1 & \text{wenn es beim zweiten Schritt nach rechts geht} \end{cases}$$

Von Interesse ist die Position Z des Teilchens nach zwei Schritten.

Es gilt

$$Z = X + Y$$

Beim ersten Schritt geht das Teilchen zufällig nach links oder rechts.

Es gilt also

$$P(X = -1) = 0.5$$

und

$$P(X = 1) = 0.5$$

Somit gilt

$$E(X) = 0$$

Beim zweiten Schritt unterscheiden wir drei Fälle:

1.Fall

Beim zweiten Schritt geht das Teilchen zufällig nach links oder rechts.

Die gemeinsame Verteilung von X und Y ist in diesem Fall:

| | | | | |
|------|-----|--------|--------|-------|
| | y | -1 | 1 | |
| x | | | | |
| -1 | | 0.25 | 0.25 | 0.5 |
| 1 | | 0.25 | 0.25 | 0.5 |
| | | 0.5 | 0.5 | |

Somit erhalten wir folgende Verteilung für Z :

| | |
|------|----------|
| z | $P(Z=z)$ |
| -2 | 0.25 |
| 0 | 0.5 |
| 2 | 0.25 |

Somit gilt

$$\begin{aligned} E(Z) &= (-2) \cdot 0.25 + 0 \cdot 0.5 + 2 \cdot 0.25 \\ &= 0 \end{aligned}$$

und

$$\begin{aligned} E(Z^2) &= (-2)^2 \cdot 0.25 + 0 \cdot 0.5 + 2^2 \cdot 0.25 \\ &= 2 \end{aligned}$$

Die Varianz von Z ist also

$$\begin{aligned} \text{Var}(Z) &= E(Z^2) - E(Z)^2 \\ &= 2 \end{aligned}$$

2.Fall

Beim zweiten Schritt geht das Teilchen in die entgegengesetzte Richtung.

Die gemeinsame Verteilung von X und Y ist in diesem Fall:

| | | | |
|-----|-----|-----|-----|
| y | -1 | 1 | |
| x | | | |
| -1 | 0 | 0.5 | 0.5 |
| 1 | 0.5 | 0 | 0.5 |
| | 0.5 | 0.5 | |

Somit erhalten wir folgende Verteilung für Z :

| | |
|-----|----------|
| z | $P(Z=z)$ |
| 0 | 1 |

Somit gilt

$$\text{Var}(Z) = 0$$

3.Fall

Beim zweiten Schritt geht das Teilchen in die gleiche Richtung wie beim ersten.

Die gemeinsame Verteilung von X und Y ist in diesem Fall:

| | | | |
|-----|-----|-----|-----|
| y | -1 | 1 | |
| x | | | |
| -1 | 0.5 | 0 | 0.5 |
| 1 | 0 | 0.5 | 0.5 |
| | 0.5 | 0.5 | |

Somit erhalten wir folgende Verteilung für Z :

| | |
|-----|----------|
| z | $P(Z=z)$ |
| -2 | 0.5 |
| 2 | 0.5 |

Somit gilt

$$\begin{aligned} E(Z) &= (-2) \cdot 0.5 + 2 \cdot 0.5 \\ &= 0 \end{aligned}$$

und

$$\begin{aligned} E(Z^2) &= (-2)^2 \cdot 0.5 + 2^2 \cdot 0.5 \\ &= 4 \end{aligned}$$

Die Varianz von Z ist

$$\begin{aligned} \text{Var}(Z) &= E(Z^2) - E(Z)^2 \\ &= 4 \end{aligned}$$

Die Randverteilungen von X und Y sind in allen drei Fällen identisch.

Also gilt auch

$$\text{Var}(Y) = 1$$

Beim ersten Fall gilt

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

Beim zweiten Fall gilt

$$\text{Var}(X + Y) < \text{Var}(X) + \text{Var}(Y)$$

Beim dritten Fall gilt

$$\text{Var}(X + Y) > \text{Var}(X) + \text{Var}(Y)$$

Das Beispiel zeigt, daß die Größe $\text{Var}(X + Y)$ neben $\text{Var}(X)$ und $\text{Var}(Y)$ offensichtlich noch von einer dritten Größe abhängt.

Es gilt

$$\begin{aligned} \text{Var}(X + Y) &= E \left[(X + Y - E(X + Y))^2 \right] \\ &= E \left[(X + Y - E(X) - E(Y))^2 \right] \\ &= E \left[(X - E(X) + Y - E(Y))^2 \right] \\ &= E \left[(X - E(X))^2 \right] + E \left[(Y - E(Y))^2 \right] \\ &\quad + 2 E \left[(X - E(X))(Y - E(Y)) \right] \\ &= \text{Var}(X) + \text{Var}(Y) + 2 E \left[(X - E(X))(Y - E(Y)) \right] \end{aligned}$$

Definition 8.2.1 Seien X und Y Zufallsvariablen.

Dann heißt

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

die Kovarianz zwischen X und Y .

Es gilt also

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y)$$

Satz 8.2.1 Für die Zufallsvariablen X_1, \dots, X_n gilt:

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{i < j} \text{Cov}(X_i, X_j)$$

Beweis

$$\begin{aligned} \text{Var}\left(\sum_{i=1}^n X_i\right) &= E\left[\left(\sum_{i=1}^n X_i - E\left(\sum_{i=1}^n X_i\right)\right)^2\right] \\ &= E\left[\left(\sum_{i=1}^n (X_i - E(X_i))\right)^2\right] \\ &= E\left[\sum_{i=1}^n \sum_{j=1}^n (X_i - E(X_i))(X_j - E(X_j))\right] \\ &= \sum_{i=1}^n \sum_{j=1}^n E[(X_i - E(X_i))(X_j - E(X_j))] \\ &= \sum_{i=1}^n \text{Var}(X_i) + \sum_{i \neq j} \text{Cov}(X_i, X_j) \\ &= \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{i < j} \text{Cov}(X_i, X_j) \end{aligned}$$

da gilt

$$\begin{aligned}
 \left(\sum_{i=1}^n a_i \right)^2 &= \left(\sum_{i=1}^n a_i \right) \left(\sum_{i=1}^n a_i \right) = \\
 &= (a_1 + \dots + a_n) (a_1 + \dots + a_n) = \\
 &= a_1 a_1 + \dots + a_1 a_n + \dots + a_n a_1 + \dots + a_n a_n = \\
 &= \sum_{i=1}^n \sum_{j=1}^n (a_i a_j)
 \end{aligned}$$

Im folgenden Satz sind die Eigenschaften der Kovarianz zu finden.

Satz 8.2.2 Für die Kovarianz zwischen den Zufallsvariablen X und Y gilt:

1. $Cov(X, Y) = E(XY) - E(X)E(Y)$
2. $Cov(X, X) = Var(X)$
3. $Cov(X, Y) = Cov(Y, X)$
4. $Cov(a + X, Y) = Cov(X, Y)$
5. $Cov(aX, bY) = ab Cov(X, Y)$
6. $Cov(X, Y + Z) = Cov(X, Y) + Cov(X, Z)$
7. $Cov(aV + bW, cX + dY) =$
 $= ac Cov(V, X) + ad Cov(V, Y) + bc Cov(W, X) + bd Cov(W, Y)$

Beweis:

1.

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

2.

$$\begin{aligned}
 \text{Cov}(X, X) &= E[(X - E(X))(X - E(X))] \\
 &= E[(X - E(X))^2] \\
 &= \text{Var}(X)
 \end{aligned}$$

3.

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

4.

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

5.

$$\begin{aligned}
 \text{Cov}(aX, bY) &= E[(aX - E(aX))(bY - E(bY))] \\
 &= E[(aX - aE(X))(bY - bE(Y))] \\
 &= abE[(X - E(X))(Y - E(Y))] \\
 &= ab\text{Cov}(X, Y)
 \end{aligned}$$

6.

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

7.

$$\begin{aligned} \text{Cov}(aV + bW, cX + dY) &= \text{Cov}(aV, cX + dY) \\ &\quad + \text{Cov}(bW, cX + dY) \\ &= \text{Cov}(aV, cX) + \text{Cov}(aV, dY) \\ &\quad + \text{Cov}(bW, cX) + \text{Cov}(bW, dY) \\ &= ac \text{Cov}(V, X) + ad \text{Cov}(V, Y) \\ &\quad + bc \text{Cov}(W, X) + bd \text{Cov}(W, Y) \end{aligned}$$

Beispiel 8.2.2 Wir betrachten folgende zweidimensionale Zufallsvariable mit gemeinsamer Dichtefunktion

$$f_{X,Y}(x, y) = \begin{cases} 2 & \text{für } 0 \leq x \leq y \leq 1 \\ 0 & \text{sonst} \end{cases}$$

Die Randdichtefunktion von X lautet:

$$f_X(x) = \begin{cases} 2 - 2x & \text{für } 0 \leq x \leq 1 \\ 0 & \text{sonst} \end{cases}$$

Die Randdichtefunktion von Y lautet:

$$f_Y(y) = \begin{cases} 2y & \text{für } 0 \leq y \leq 1 \\ 0 & \text{sonst} \end{cases}$$

Also gilt:

$$\begin{aligned} E(X) &= \int_0^1 x(2 - 2x) dx \\ &= \frac{1}{3} \end{aligned}$$

$$\begin{aligned} E(Y) &= \int_0^1 y \cdot 2y dx \\ &= \frac{2}{3} \end{aligned}$$

$$\begin{aligned} E(XY) &= \int_0^1 \int_0^y xy \cdot 2 dx dy \\ &= \int_0^1 y^3 dy \\ &= \frac{1}{4} \end{aligned}$$

Also gilt

$$\begin{aligned} \text{Cov}(X, Y) &= E(XY) - E(X)E(Y) \\ &= \frac{1}{4} - \frac{2}{9} \\ &= \frac{1}{36} \end{aligned}$$

Der folgende Satz zeigt, daß die Kovarianz bei Unabhängigkeit gleich 0 ist.

Satz 8.2.3 *Sind die Zufallsvariablen X und Y unabhängig, so gilt*

$$\text{Cov}(X, Y) = 0.$$

Beweis:

Sind X und Y unabhängig, so gilt

$$E(XY) = E(X)E(Y).$$

Also gilt

$$\begin{aligned} \text{Cov}(X, Y) &= E(XY) - E(X)E(Y) \\ &= E(X)E(Y) - E(X)E(Y) \\ &= 0 \end{aligned}$$

Bei Unabhängigkeit kann man die Varianz der Summe von Zufallsvariablen einfach bestimmen.

Satz 8.2.4 *Sind die Zufallsvariablen X und Y unabhängig, so gilt*

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

Beweis:

Sind X und Y unabhängig, so gilt $\text{Cov}(X, Y) = 0$.

Also gilt

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

Satz 8.2.5 Die Zufallsvariablen X_1, \dots, X_n seien unabhängig und identisch verteilt mit

$$E(X_i) = \mu$$

und

$$\text{Var}(X_i) = \sigma^2.$$

Dann gilt

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = n\sigma^2$$

Beweis:

$$\begin{aligned} \text{Var}\left(\sum_{i=1}^n X_i\right) &= \sum_{i=1}^n \text{Var}(X_i) \\ &= n\sigma^2 \end{aligned}$$

Für

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

gilt

$$\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

denn

$$\begin{aligned} \text{Var}(\bar{X}) &= \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \\ &= \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^n X_i\right) \\ &= \frac{1}{n^2} n\sigma^2 \\ &= \frac{\sigma^2}{n} \end{aligned}$$

Wir sehen, daß die Varianz von \bar{X} mit wachsendem Stichprobenumfang immer kleiner wird.

Da gilt

$$E(\bar{X}) = \mu,$$

konzentriert sich die Verteilung von \bar{X} mit wachsendem Stichprobenumfang immer mehr um μ .

Das folgende Beispiel zeigt, daß aus $Cov(X, Y) = 0$ nicht notwendigerweise die Unabhängigkeit von X und Y folgt.

Beispiel 8.2.3 Die Zufallsvariablen X und Y besitzen folgende gemeinsame Wahrscheinlichkeitsfunktion:

| y | 0 | 1 |
|-----|-----|-----|
| x | | |
| -1 | 0 | 0.2 |
| 0 | 0.6 | 0 |
| 1 | 0 | 0.2 |

Es gilt

$$\begin{aligned} E(X) &= (-1) \cdot 0.2 + 0 \cdot 0.6 + 1 \cdot 0.2 \\ &= 0 \end{aligned}$$

und

$$\begin{aligned} E(XY) &= (-1) \cdot 0 \cdot 0 + (-1) \cdot 1 \cdot 0.2 \\ &\quad + 0 \cdot 0 \cdot 0.6 + 0 \cdot 1 \cdot 0 \\ &\quad + 1 \cdot 0 \cdot 0 + 1 \cdot 1 \cdot 0.2 \\ &= 0 \end{aligned}$$

Also gilt

$$Cov(X, Y) = 0.$$

Die Zufallsvariablen X und Y sind aber nicht unabhängig, denn

$$P(X = -1, Y = 0) = 0 \neq 0.12 = P(X = -1)P(Y = 0)$$

Vielmehr besteht zwischen X und Y mit Wahrscheinlichkeit 1 ein Zusammenhang.

Es gilt

$$\begin{aligned} P(Y = 1 | X = -1) &= 1 \\ P(Y = 0 | X = 0) &= 1 \\ P(Y = 1 | X = 1) &= 1 \end{aligned}$$

Also gilt

$$P(Y = X^2) = 1$$

Wie der folgende Satz zeigt, sind bei bivariater Normalverteilung die Zufallsvariablen X und Y genau dann unabhängig, wenn die Kovarianz zwischen X und Y Null ist.

Satz 8.2.6 Die Zufallsvariable (X, Y) sei bivariat normalverteilt mit den Parametern $\mu_X, \mu_Y, \sigma_X, \sigma_Y$ und ρ .

Dann gilt

$$\text{Cov}(X, Y) = \rho \sigma_X \sigma_Y$$

Beweis:

Es gilt

$$\text{Cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y) f_{X,Y}(x, y) dx dy$$

mit

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_X)^2}{\sigma_X^2} - 2\rho\frac{x-\mu_X}{\sigma_X}\frac{y-\mu_Y}{\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right)}$$

Wir substituieren

$$u = \frac{x - \mu_X}{\sigma_X}$$

$$v = \frac{y - \mu_Y}{\sigma_Y}$$

Es gilt

$$dx = \sigma_X du$$

und

$$dy = \sigma_Y dv.$$

Dann gilt

$$\begin{aligned} \text{Cov}(X, Y) &= \frac{\sigma_X \sigma_Y}{2\pi \sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u v e^{-\frac{1}{2(1-\rho^2)}(u^2 - 2\rho uv + v^2)} du dv \\ &= \frac{\sigma_X \sigma_Y}{2\pi \sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u v e^{-\frac{1}{2(1-\rho^2)}(u^2 - 2\rho uv + \rho^2 v^2 + v^2 - \rho^2 v^2)} du dv \\ &= \frac{\sigma_X \sigma_Y}{\sqrt{2\pi}} \int_{-\infty}^{\infty} v e^{-0.5 v^2} \left[\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi} \sqrt{1-\rho^2}} u e^{-\frac{1}{2(1-\rho^2)}(u-\rho v)^2} du \right] dv \end{aligned}$$

Da

$$\frac{1}{\sqrt{2\pi} \sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(u-\rho v)^2}$$

die Dichtefunktion einer mit den Parametern ρv und $1-\rho^2$ normalverteilten Zufallsvariablen ist, ist

$$\int_{-\infty}^{\infty} u \frac{1}{\sqrt{2\pi} \sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(u-\rho v)^2} du$$

gleich dem Erwartungswert

$$\rho v.$$

Also gilt

$$\begin{aligned} \text{Cov}(X, Y) &= \frac{\sigma_X \sigma_Y}{\sqrt{2\pi}} \int_{-\infty}^{\infty} v e^{-0.5 v^2} \rho v dv \\ &= \sigma_X \sigma_Y \rho \int_{-\infty}^{\infty} v^2 \frac{1}{\sqrt{2\pi}} e^{-0.5 v^2} dv \\ &= \sigma_X \sigma_Y \rho \end{aligned}$$

denn

$$\int_{-\infty}^{\infty} v^2 \frac{1}{\sqrt{2\pi}} e^{-0.5 v^2} dv$$

ist gleich der Varianz einer standardnormalverteilten Zufallsvariablen und somit gleich 1.

Die Kovarianz hat den Nachteil, daß sie von den Maßeinheiten der Variablen abhängt.

Dieses Problem kann man beheben, indem man zu den standardisierten Variablen übergeht.

Satz 8.2.7 *Die Zufallsvariable X besitze den Erwartungswert μ und die Varianz σ .*

Dann gilt für die Zufallsvariable

$$\tilde{X} = \frac{X - \mu_X}{\sigma_X}$$

$$E(\tilde{X}) = 0$$

und

$$\text{Var}(\tilde{X}) = 1$$

Beweis:

$$\begin{aligned} E(\tilde{X}) &= E\left(\frac{X - \mu}{\sigma}\right) \\ &= \frac{1}{\sigma} E(X - \mu) \\ &= \frac{1}{\sigma} (E(X) - \mu) \\ &= 0 \end{aligned}$$

$$\begin{aligned} \text{Var}(\tilde{X}) &= \text{Var}\left(\frac{X - \mu}{\sigma}\right) \\ &= \frac{1}{\sigma^2} \text{Var}(X - \mu) \\ &= 1 \end{aligned}$$

Nun können wir den Korrelationskoeffizienten definieren.

Definition 8.2.2 Seien X und Y Zufallsvariablen mit Varianzen σ_X^2 und σ_Y^2 und Kovarianz $Cov(X, Y)$.

Dann heißt

$$\rho_{X,Y} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

Korrelationskoeffizient zwischen X und Y .

Der Korrelationskoeffizient ist die Kovarianz der standardisierten Variablen, denn

$$\begin{aligned} Cov(\tilde{X}, \tilde{Y}) &= Cov\left(\frac{X - \mu_X}{\sigma_X}, \frac{Y - \mu_Y}{\sigma_Y}\right) \\ &= \frac{1}{\sigma_X \sigma_Y} Cov(X - \mu_X, Y - \mu_Y) \\ &= \frac{1}{\sigma_X \sigma_Y} Cov(X, Y) \\ &= \rho_{X,Y} \end{aligned}$$

Im folgende Satz sind Eigenschaften des Korrelationskoeffizienten zu finden.

Satz 8.2.8 Für den Korrelationskoeffizienten $\rho_{X,Y}$ zwischen den Zufallsvariablen X und Y gilt:

$$-1 \leq \rho_{X,Y} \leq 1$$

Dabei ist $\rho_{X,Y} = \pm 1$ genau dann, wenn Konstanten a und $b \neq 0$ existieren, so daß gilt

$$P(Y = a \pm b X) = 1.$$

Beweis:

Seien \tilde{X} und \tilde{Y} die standardisierten Variablen.

Dann gilt

$$\begin{aligned} Var(\tilde{X} + \tilde{Y}) &= Var(\tilde{X}) + Var(\tilde{Y}) + 2Cov(\tilde{X}, \tilde{Y}) \\ &= 2 + 2\rho_{X,Y} \end{aligned}$$

Da die Varianz nichtnegativ ist, gilt

$$2 + 2\rho_{X,Y} \geq 0$$

und somit

$$\rho_{X,Y} \geq -1$$

Außerdem gilt

$$\begin{aligned} \text{Var}(\tilde{X} - \tilde{Y}) &= \text{Var}(\tilde{X}) + \text{Var}(\tilde{Y}) - 2 \text{Cov}(\tilde{X}, \tilde{Y}) \\ &= 2 - 2\rho_{X,Y} \end{aligned}$$

Hieraus folgt

$$\rho_{X,Y} \leq 1$$

Also gilt

$$-1 \leq \rho_{X,Y} \leq 1$$

Ist

$$\rho_{X,Y} = 1,$$

so gilt

$$\text{Var}(\tilde{X} - \tilde{Y}) = 0.$$

Somit gilt

$$P(\tilde{X} - \tilde{Y} = 0) = 1.$$

Also gilt

$$P(Y = a + bX) = 1$$

mit

$$a = \mu_Y - \frac{\sigma_X}{\sigma_Y} \mu_X$$

und

$$b = \frac{\sigma_X}{\sigma_Y}.$$

Eine analoge Beziehung erhält man für $\rho_{X,Y} = -1$.

Der Korrelationskoeffizient charakterisiert also den linearen Zusammenhang zwischen X und Y .

8.3 Kovarianzen in der Zeitreihenanalyse

Eine wichtige Rolle spielen Kovarianzen immer dann, wenn Abhängigkeitsstrukturen zwischen Zufallsvariablen modelliert werden sollen.

Dies ist zum Beispiel in der Zeitreihenanalyse der Fall.

Ausgangspunkt ist hier eine Folge X_1, \dots, X_n, \dots von Zufallsvariablen.

Die Abhängigkeitsstruktur zwischen den Zufallsvariablen soll durch ein einfaches Modell beschrieben werden.

Das einfachste Modell ist der White-Noise-Prozeß.

Dies ist eine Folge von unabhängigen, identisch verteilten Zufallsvariablen $\epsilon_1, \dots, \epsilon_n, \dots$, wobei wir in der Regel unterstellen, daß alle ϵ_i den Erwartungswert 0 besitzen.

Es gilt also

$$E(\epsilon_t) = 0$$

und

$$\text{Cov}(\epsilon_s, \epsilon_t) = E(\epsilon_s \cdot \epsilon_t) = \begin{cases} \sigma^2 & \text{für } s = t \\ 0 & \text{für } \textit{sonst} \end{cases}$$

Man nennt ϵ_t auch einen Zufallsschock.

White-Noise-Prozesse sind die Bausteine komplexerer Prozesse.

Ist X_t eine Linearkombination von gegenwärtigem und vergangenen Zufallsschocks, so spricht man von einem Moving-Average-Prozeß.

Definition 8.3.1 *Eine Folge X_1, \dots, X_n, \dots von Zufallsvariablen heißt Moving-Average-Prozeß der Ordnung q , kurz MA[q]-Prozeß, wenn gilt*

$$X_t = \epsilon_t + \beta_1 \epsilon_{t-1} + \dots + \beta_q \epsilon_{t-q}$$

wobei $\epsilon_1, \dots, \epsilon_t, \dots$ ein White-Noise-Prozeß ist.

Schauen wir uns einen MA[1]-Prozeß an.

Es gilt also

$$X_t = \epsilon_t + \beta \epsilon_{t-1}$$

Wir gehen davon aus, daß für alle t gilt

$$E(\epsilon_t) = 0.$$

Für den Erwartungswert von X_t gilt

$$E(X_t) = 0$$

Dies sieht man folgendermaßen:

$$\begin{aligned} E(X_t) &= E(\epsilon_t + \beta \epsilon_{t-1}) \\ &= E(\epsilon_t) + \beta E(\epsilon_{t-1}) \\ &= 0 \end{aligned}$$

Die Varianz von X_t ist gegeben durch

$$\text{Var}(X_t) = \sigma^2 (1 + \beta^2)$$

Dies sieht man folgendermaßen:

$$\begin{aligned} \text{Var}(X_t) &= \text{Var}(\epsilon_t + \beta \epsilon_{t-1}) \\ &= \text{Var}(\epsilon_t) + \beta^2 \text{Var}(\epsilon_{t-1}) \\ &= \sigma^2 (1 + \beta^2) \end{aligned}$$

Wir schauen uns die Kovarianz zwischen benachbarten Beobachtungen an. Es gilt

$$\text{Cov}(X_t, X_{t-1}) = \beta \sigma^2$$

Dies sieht man folgendermaßen:

$$\begin{aligned} \text{Cov}(X_t, X_{t-1}) &= \text{Cov}(\epsilon_t + \beta \epsilon_{t-1}, \epsilon_{t-1} + \beta \epsilon_{t-2}) = \\ &= \text{Cov}(\epsilon_t, \epsilon_{t-1}) + \text{Cov}(\epsilon_t, \beta \epsilon_{t-2}) \\ &+ \text{Cov}(\beta \epsilon_{t-1}, \epsilon_{t-1}) + \text{Cov}(\beta \epsilon_{t-1}, \beta \epsilon_{t-2}) \\ &= \text{Cov}(\epsilon_t, \epsilon_{t-1}) + \beta \text{Cov}(\epsilon_t, \epsilon_{t-2}) \\ &+ \beta \text{Cov}(\epsilon_{t-1}, \epsilon_{t-1}) + \beta^2 \text{Cov}(\epsilon_{t-1}, \epsilon_{t-2}) \\ &= \beta \text{Var}(\epsilon_{t-1}) = \\ &= \beta \sigma^2 \end{aligned}$$

da für $s > 0$ gilt

$$\text{Cov}(\epsilon_t, \epsilon_{t-s}) = 0$$

Somit gilt für den Korrelationskoeffizienten zwischen X_t und X_{t-1} :

$$\rho_{X_t, X_{t-1}} = \frac{\beta}{1 + \beta^2}$$

Dies sieht man folgendermaßen.

$$\begin{aligned} \rho_{X_t, X_{t-1}} &= \frac{\text{Cov}(X_t, X_{t-1})}{\sqrt{\text{Var}(X_t) \text{Var}(X_{t-1})}} \\ &= \frac{\beta \sigma^2}{\sigma^2 (1 + \beta^2)} \\ &= \frac{\beta}{1 + \beta^2} \end{aligned}$$

Man nennt $\rho_{X_t, X_{t-1}}$ auch den Autokorrelationskoeffizienten zum Lag 1 und schreibt ρ_1 .

Für $s > 1$ gilt

$$\begin{aligned} \text{Cov}(X_t, X_{t-s}) &= \text{Cov}(\epsilon_t + \beta \epsilon_{t-1}, \epsilon_{t-s} + \beta \epsilon_{t-s-1}) \\ &= \text{Cov}(\epsilon_t, \epsilon_{t-s}) + \text{Cov}(\epsilon_t, \beta \epsilon_{t-s-1}) \\ &\quad + \text{Cov}(\beta \epsilon_{t-1}, \epsilon_{t-s}) + \text{Cov}(\beta \epsilon_{t-1}, \beta \epsilon_{t-s-1}) \\ &= 0 \end{aligned}$$

Somit gilt für die Autokorrelationsfunktion (ACF) eines MA[1]-Prozesses:

$$\rho_s = \begin{cases} \frac{\beta}{1 + \beta^2} & \text{für } s = 1 \\ 0 & \text{sonst} \end{cases}$$

Die ACF eines MA[1]-Prozesses wird also nach dem Lag 1 gleich 0. Man sagt auch, daß die nach dem Lag 0 abbricht.

Auf Moving-Average-Prozesse höherer Ordnung wollen wir hier nicht eingehen.

Erwartungswerte und Varianzen eines MA[1]-Prozesses sind für jeden Zeitpunkt gleich. Außerdem hängt die Kovarianz zwischen X_t und X_{t-s} nur von s und nicht von t ab.

Man spricht in diesem Fall von einem schwach stationären Prozeß.

Eine weitere Klasse von linearen stochastischen Prozessen bilden die autoregressiven Prozesse.

Definition 8.3.2 Eine Folge X_1, \dots, X_n, \dots von Zufallsvariablen heißt autoregressiver Prozeß der Ordnung p , kurz AR[p]-Prozeß, wenn gilt

$$X_t = \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + \epsilon_t$$

wobei $\epsilon_1, \dots, \epsilon_t, \dots$ ein White-Noise-Prozeß ist.

Schauen wir uns einen AR[1]-Prozeß an.

Es gilt also

$$X_t = \alpha X_{t-1} + \epsilon_t$$

Wir leiten nun den Erwartungswert, die Varianz und die Autokorrelationsfunktion unter der Nebenbedingung her, daß der Prozeß schwach stationär ist.

Es muß also gelten

$$E(X_t) = E(X_{t-s})$$

für alle $s > 0$,

$$Var(X_t) = Var(X_{t-s})$$

für alle $s > 0$ und

$$Cov(X_t, X_{t-s}) = Cov(X_{t-k}, X_{t-s-k})$$

für alle $s, k > 0$.

Für den Erwartungswert eines AR[1]-Prozesses gilt

$$E(X_t) = 0$$

Dies sieht man folgendermaßen:

Es gilt

$$\begin{aligned} E(X_t) &= E(\alpha X_{t-1} + \epsilon_t) \\ &= \alpha E(X_{t-1}) + E(\epsilon_t) \\ &= \alpha E(X_{t-1}) \end{aligned}$$

Aufgrund der schwachen Stationarität muß gelten

$$E(X_t) = E(X_{t-1})$$

Also gilt

$$E(X_t) = \alpha E(X_t)$$

Hieraus folgt

$$E(X_t) = 0$$

Für die Varianz gilt

$$\text{Var}(X_t) = \alpha^2 \text{Var}(X_{t-1}) + \sigma^2$$

Dies sieht man folgendermaßen:

$$\begin{aligned} \text{Var}(X_t) &= \text{Var}(\alpha X_{t-1} + \epsilon_t) \\ &= \alpha^2 \text{Var}(X_{t-1}) + \text{Var}(\epsilon_t) + 2\alpha \text{Cov}(X_{t-1}, \epsilon_t) \\ &= \alpha^2 \text{Var}(X_{t-1}) + \sigma^2 \end{aligned}$$

da der Zufallsschock ϵ_t zeitlich auf X_{t-1} folgt und somit mit X_{t-1} unkorreliert ist.

Aufgrund der schwachen Stationarität muß gelten

$$\text{Var}(X_t) = \text{Var}(X_{t-1})$$

Also gilt

$$\text{Var}(X_t) = \alpha^2 \text{Var}(X_t) + \sigma^2$$

Hieraus folgt

$$\text{Var}(X_t) = \frac{\sigma^2}{1 - \alpha^2}$$

Offensichtlich muß gelten

$$1 - \alpha^2 > 0$$

und somit

$$|\alpha| < 1.$$

Für die Kovarianz zwischen X_t und X_{t-s} gilt:

$$\text{Cov}(X_t, X_{t-s}) = \alpha \text{Cov}(X_{t-1}, X_{t-s})$$

Dies sieht man folgendermaßen:

$$\begin{aligned} \text{Cov}(X_t, X_{t-s}) &= \text{Cov}((\alpha X_{t-1} + \epsilon_t), X_{t-s}) \\ &= \text{Cov}(\alpha X_{t-1}, X_{t-s}) + \text{Cov}(\epsilon_t, X_{t-s}) \\ &= \alpha \text{Cov}(X_{t-1}, X_{t-s}) \end{aligned}$$

Da der Prozeß stationär ist, gilt auch

$$\rho_s = \alpha \rho_{s-1},$$

Man nennt diese Gleichungen auch die Yule-Walker-Gleichungen.

Wegen

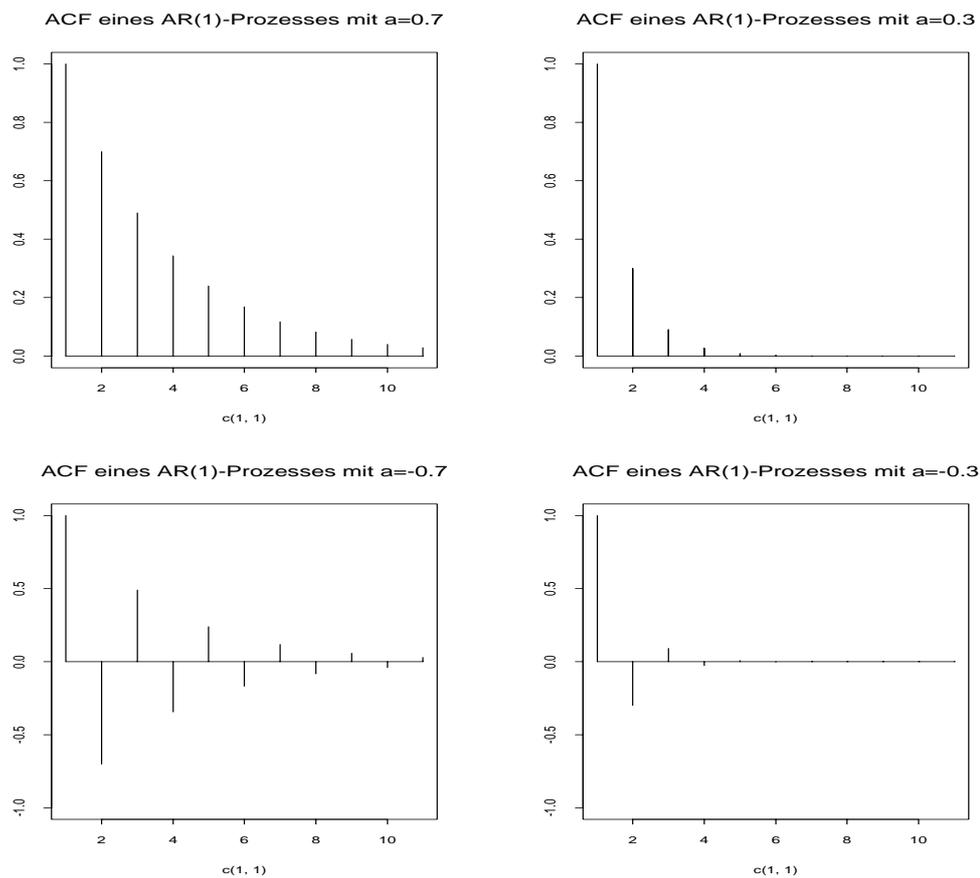
$$\rho_0 = 1$$

gilt

$$\rho_s = \alpha^s$$

Die Autokorrelationsfunktion eines AR[1]-Prozesses nimmt also mit wachsendem Lag geometrisch ab.

Die folgende Graphik zeigt die Autokorrelationsfunktionen einiger AR[1]-Prozesse.



Wir betrachten nun noch folgende Größe:

$$DW = \frac{\text{Var}(X_t - X_{t-1})}{\text{Var}(X_t)}$$

Dies ist die normierte Varianz der Differenz zweier benachbarter Beobachtungen. Bei einem stationären Prozeß nimmt diese eine einfache Gestalt an.

Es gilt

$$DW = 2 - 2\rho_1$$

Dies sieht man folgendermaßen:

$$\begin{aligned} DW &= \frac{\text{Var}(X_t - X_{t-1})}{\text{Var}(X_t)} \\ &= \frac{\text{Var}(X_t) + \text{Var}(X_{t-1}) - 2\text{Cov}(X_t, X_{t-1})}{\text{Var}(X_t)} \\ &= 1 + 1 - 2\rho_1 \\ &= 2 - 2\rho_1 \end{aligned}$$

Gilt

$$\rho_1 = 1,$$

so ist

$$DW = 0,$$

gilt

$$\rho_1 = -1,$$

so ist

$$DW = 4$$

und gilt

$$\rho_1 = 0,$$

so ist

$$DW = 2.$$

Eine notwendige Bedingung für die schwache Stationarität eines AR[1]-Prozesses ist

$$|\alpha| < 1.$$

Ist nun

$$\alpha = 1,$$

so gilt

$$X_t = X_{t-1} + \epsilon_t.$$

Man nennt diesen Prozeß Random-Walk.

Mit $X_0 = 0$ erhalten wir

$$X_t = \sum_{i=1}^t \epsilon_i$$

Somit gilt für einen Random-Walk:

$$\begin{aligned} E(X_t) &= E\left(\sum_{i=1}^t \epsilon_i\right) \\ &= \sum_{i=1}^t E(\epsilon_i) \\ &= 0 \end{aligned}$$

und

$$\begin{aligned} Var(X_t) &= Var\left(\sum_{i=1}^t \epsilon_i\right) \\ &= \sum_{i=1}^t Var(\epsilon_i) \\ &= t\sigma^2 \end{aligned}$$

Für die Kovarianz erhalten wir mit $s \leq t$:

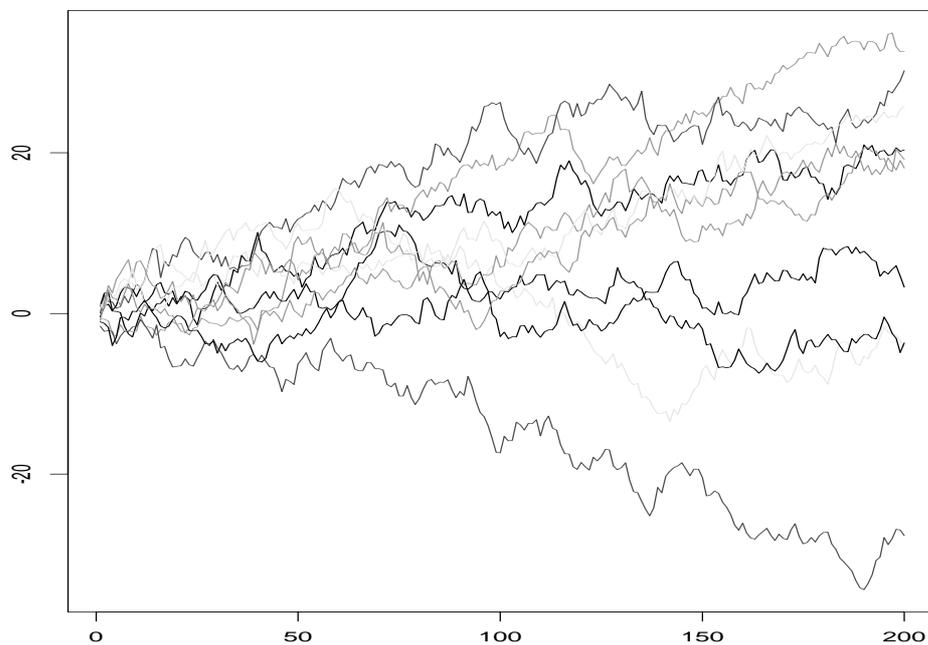
$$\begin{aligned} Cov(X_s, X_t) &= E(X_s X_t) = E\left(\sum_{i=1}^s \epsilon_i \sum_{j=1}^t \epsilon_j\right) \\ &= E\left(\sum_{i=1}^s \sum_{j=1}^t \epsilon_i \epsilon_j\right) \end{aligned}$$

$$\begin{aligned} &= \sum_{i=1}^s \sum_{j=1}^t E(\epsilon_i \epsilon_j) \\ &= \sum_{j=i}^s \text{Var}(\epsilon_i) \\ &= s \sigma^2 \end{aligned}$$

Also gilt

$$\begin{aligned} \rho_{X_t, X_s} &= \frac{s}{\sqrt{st}} \\ &= \sqrt{\frac{s}{t}} \end{aligned}$$

Das folgende Bild zeigt 10 Realisationen eines Random-Walks der Länge 200.



Hier zeigen sich die Eigenschaften des Random-Walk sehr schön.

Die Varianz nimmt mit der Zeit immer mehr zu.

Außerdem sind Beobachtungen, die nahe beieinander liegen, hoch korreliert.

Der Random-Walk ist nicht stationär.

Es gilt aber

$$\begin{aligned}\Delta X_t &= X_t - X_{t-1} \\ &= \epsilon_t\end{aligned}$$

Also sind die ersten Differenzen eines Random-Walk stationär.

Man sagt, daß der Random-Walk integriert vom Grade 1, kurz I(1), ist.

8.4 Kovarianzen in der Multivariaten Analyse

Auch in der Multivariaten Analyse spielen Kovarianzen eine zentrale Rolle. Hier werden an jedem von n Objekten p Variablen erhoben.

Es ist sehr schwierig, bei so vielen Variablen Strukturen zu erkennen. Man versucht deshalb, die Anzahl der Variablen zu reduzieren.

Eine Möglichkeit besteht darin, die Variablen auszuwählen, die die meiste Information enthalten.

Dies sind aber die Variablen mit der größten Varianz.

Werden nur diese bei der weiteren Analyse verwendet, verliert man natürlich die in den anderen Variablen enthaltene Information.

Aus diesem Grunde wählt man eine andere Vorgehensweise.

Man betrachtet alle Linearkombinationen der Variablen und wählt die mit der größten Streuung aus.

Man nennt diese Vorgehensweise Hauptkomponentenanalyse.

Schauen wir uns dies für zwei Zufallsvariablen an, die wir mit X_1 und X_2 bezeichnen wollen.

Wir suchen reelle Zahlen a_1 und a_2 , so daß die Varianz der Linearkombination

$$a_1 X_1 + a_2 X_2$$

maximal wird.

Dieses Problem hat aber ohne Nebenbedingungen keine endliche Lösung.

Wir betrachten die Nebenbedingung

$$a_1^2 + a_2^2 = 1.$$

Betrachten wir zunächst die Varianz der Linearkombination:

$$\begin{aligned} \text{Var}(a_1 X_1 + a_2 X_2) &= \text{Var}(a_1 X_1) + \text{Var}(a_2 X_2) + 2 \text{Cov}(a_1 X_1, a_2 X_2) \\ &= a_1^2 \text{Var}(X_1) + a_2^2 \text{Var}(X_2) + 2 a_1 a_2 \text{Cov}(X_1, X_2) \\ &= a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 + 2 a_1 a_2 \sigma_{12} \end{aligned}$$

Dieser Ausdruck ist also zu maximieren unter der Nebenbedingung

$$a_1^2 + a_2^2 = 1$$

Wir stellen die Lagrange-Funktion auf:

$$L(a_1, a_2, \lambda) = a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 + 2 a_1 a_2 \sigma_{12} - \lambda (a_1^2 + a_2^2 - 1)$$

Die notwendigen Bedingungen für einen Extremwert lauten

$$\frac{\partial}{\partial a_1} L(a_1, a_2, \lambda) = 0,$$

$$\frac{\partial}{\partial a_2} L(a_1, a_2, \lambda) = 0$$

und

$$\frac{\partial}{\partial \lambda} L(a_1, a_2, \lambda) = 0$$

Wir bilden also die partiellen Ableitungen:

$$\frac{\partial}{\partial a_1} L(a_1, a_2, \lambda) = 2 a_1 \sigma_1^2 + 2 a_2 \sigma_{12} - 2 \lambda a_1,$$

$$\frac{\partial}{\partial a_2} L(a_1, a_2, \lambda) = 2 a_1 \sigma_{12} + 2 a_2 \sigma_2^2 - 2 \lambda a_2$$

und

$$\frac{\partial}{\partial \lambda} L(a_1, a_2, \lambda) = a_1^2 + a_2^2 - 1$$

Es muß also gelten

$$a_1 \sigma_1^2 + a_2 \sigma_{12} = \lambda a_1$$

und

$$a_1 \sigma_{12} + a_2 \sigma_2^2 = \lambda a_2$$

Wenn wir dies in Matrixschreibweise schreiben, können wir die Struktur besser erkennen.

Seien

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$$

und

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$$

Also muß gelten

$$\Sigma \mathbf{a} = \lambda \mathbf{a}$$

Es handelt sich also um ein Eigenwertproblem.

Wir formen um

$$(\boldsymbol{\Sigma} - \lambda \mathbf{I}) \mathbf{a} = \mathbf{0}$$

mit

$$\mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

und

$$\mathbf{0} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Dieses Gleichungssystem wird trivialerweise von $\mathbf{a} = \mathbf{0}$ erfüllt.

Für das Vorliegen nichttrivialer Lösungen muß

$$\boldsymbol{\Sigma} - \lambda \mathbf{I}$$

singulär sein.

Es muß also gelten

$$\det(\boldsymbol{\Sigma} - \lambda \mathbf{I}) = 0$$

Es gilt

$$\begin{aligned} \det(\boldsymbol{\Sigma} - \lambda \mathbf{I}) &= \det \begin{pmatrix} \sigma_1^2 - \lambda & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 - \lambda \end{pmatrix} \\ &= (\sigma_1^2 - \lambda)(\sigma_2^2 - \lambda) - \sigma_{12}^2 \\ &= \sigma_1^2 \sigma_2^2 - \sigma_1^2 \lambda - \sigma_2^2 \lambda + \lambda^2 - \sigma_{12}^2 \end{aligned}$$

Man kann zeigen, daß die beiden Lösungen λ_1 und λ_2 dieser quadratischen Gleichung reell sind.

Man nennt sie die Eigenwerte von $\boldsymbol{\Sigma}$.

Die zugehörigen auf Länge 1 normierten Vektoren \mathbf{a}_1 und \mathbf{a}_2 heißen Eigenvektoren von $\boldsymbol{\Sigma}$.

Welcher der beiden Eigenvektoren liefert nun die Linearkombination mit der größten Varianz?

Wir haben gesehen, daß a_1 und a_2 die folgenden Gleichungen erfüllen müssen:

$$a_1 \sigma_1^2 + a_2 \sigma_{12} = \lambda a_1$$

und

$$a_1 \sigma_{12} + a_2 \sigma_2^2 = \lambda a_2$$

Multiplizieren wir die erste Gleichung mit a_1 und die zweite Gleichung mit a_2 , so erhalten wir:

$$a_1^2 \sigma_1^2 + a_1 a_2 \sigma_{12} = \lambda a_1^2$$

und

$$a_1 a_2 \sigma_{12} + a_2^2 \sigma_2^2 = \lambda a_2^2$$

Addieren wir die beiden Gleichungen, so steht auf der linken Seite

$$\text{Var}(a_1 X_1 + a_2 X_2) = a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 + 2 a_1 a_2 \sigma_{12}$$

und auf der rechten Seite steht wegen $a_1^2 + a_2^2 = 1$:

$$\lambda (a_1^2 + a_2^2) = \lambda$$

Also wählen wir den Eigenvektor, der zum größeren Eigenwert gehört. Schauen wir uns ein Beispiel an.

Beispiel 8.4.1 Sei

$$\Sigma = \begin{pmatrix} 6.8 & 2.4 \\ 2.4 & 8.2 \end{pmatrix}$$

Es gilt

$$\det(\Sigma - \lambda \mathbf{I}_a) = (6.8 - \lambda)(8.2 - \lambda) - 5.76 = 0$$

Wir lösen diese quadratische Gleichung

$$\lambda^2 - 15 \lambda = -50$$

$$\lambda^2 - 15 \lambda + 7.5^2 = -50 + 7.5^2$$

$$(\lambda - 7.5)^2 = 6.25$$

$$\lambda - 7.5 = \pm 2.5$$

Die beiden Eigenwerte sind also

$$\lambda_1 = 10$$

und

$$\lambda_2 = 5.$$

Der Eigenvektor zum Eigenwert $\lambda_1 = 10$ muß folgendes Gleichungssystem erfüllen:

$$-3.2 a_1 + 2.4 a_2 = 0$$

$$2.4 a_1 - 1.8 a_2 = 0$$

Hieraus folgt

$$a_1 = \frac{3}{4} a_2.$$

Also lautet der Eigenvektor

$$\mathbf{a} = \begin{pmatrix} 0.6 \\ 0.8 \end{pmatrix}$$

Die Linearkombination mit der größten Varianz ist also

$$0.6 X_1 + 0.8 X_2$$

8.5 Bedingte Erwartungswerte

In Kapitel 6.3 haben wir uns mit bedingten Verteilungen beschäftigt. In diesem Kapitel wollen wir nun die Erwartungswerte von solchen Verteilungen betrachten.

Definition 8.5.1 Sei (X, Y) eine zweidimensionale Zufallsvariable mit Wahrscheinlichkeitsfunktion

$$P(X = x, Y = y)$$

bzw. Dichtefunktion

$$f_{X,Y}(x, y).$$

Der bedingte Erwartungswert von Y gegeben $X = x$ ist definiert durch

$$E(Y|X = x) = \sum_y y P(Y = y|X = x) \quad (8.2)$$

bzw.

$$E(Y|X = x) = \int_{-\infty}^{\infty} y f_{Y|X}(y|x) dy \quad (8.3)$$

Analog ist $E(X|Y = y)$ definiert.

Anmerkung:

Der bedingte Erwartungswert einer Funktion $g(Y)$ der Zufallsvariablen Y bei gegebenem $X = x$ ist definiert durch

$$E(g(Y)|X = x) = \sum_y g(y) P(Y = y|X = x) \quad (8.4)$$

bzw.

$$E(g(Y)|X = x) = \int_{-\infty}^{\infty} g(y) f_{Y|X}(y|x) dy \quad (8.5)$$

Schauen wir uns einige Beispiele an.

Beispiel 8.5.1 *Im Beispiel 6.4.3 haben wir eine Urne betrachtet, die 5 Kugeln enthält, von denen zwei 10 g und drei 20 g wiegen.*

Es werden zwei Kugeln ohne Zurücklegen gezogen.

Sei X das Gewicht der ersten gezogenen Kugel und Y das Gewicht der zweiten gezogenen Kugel.

Die gemeinsame Wahrscheinlichkeitsfunktion von X und Y ist gegeben durch

| | | | |
|-----|------|------|------|
| y | 10 | 20 | |
| x | | | |
| 10 | 0.10 | 0.30 | 0.40 |
| 20 | 0.30 | 0.30 | 0.60 |
| | 0.40 | 0.60 | 1.00 |

Es gilt

$$P(Y = y|X = 10) = \begin{cases} 0.25 & \text{für } y = 10 \\ 0.75 & \text{für } y = 20 \end{cases}$$

Also gilt

$$\begin{aligned} E(Y|X = 10) &= 10 \cdot 0.25 + 20 \cdot 0.75 \\ &= 17.5 \end{aligned}$$

Außerdem gilt

$$P(Y = y|X = 20) = \begin{cases} 0.5 & \text{für } y = 10 \\ 0.5 & \text{für } y = 20 \end{cases}$$

Also gilt

$$\begin{aligned} E(Y|X = 20) &= 10 \cdot 0.5 + 20 \cdot 0.5 \\ &= 15 \end{aligned}$$

Wiegt die erste gezogene Kugel 10 g, so erwarten wir, daß die zweite gezogene Kugel 17.5 g wiegt.

Wiegt die erste gezogene Kugel hingegen 20 g, so erwarten wir, daß die zweite gezogene Kugel 15 g wiegt.

Beispiel 8.5.2 Für die zweidimensionale stetige Zufallsvariable (X, Y) mit gemeinsamer Dichtefunktion

$$f_{X,Y}(x, y) = \begin{cases} 2 & \text{für } 0 \leq x \leq y \leq 1 \\ 0 & \text{sonst} \end{cases}$$

gilt

$$f_{Y|X}(y|x) = \begin{cases} \frac{1}{1-x} & \text{für } x \leq y \leq 1 \\ 0 & \text{sonst} \end{cases}$$

Also gilt

$$\begin{aligned} E(Y|X = x) &= \int_x^1 y \frac{1}{1-x} dy \\ &= \frac{1}{1-x} \left[\frac{y^2}{2} \right]_x^1 \\ &= \frac{1-x^2}{2(1-x)} \\ &= \frac{x+1}{2} \end{aligned}$$

Dies überrascht uns nicht, denn Y ist bei gegebenem Wert x von X gleichverteilt auf $[x, 1]$.

Beispiel 8.5.3 Die Zufallsvariable (X, Y) sei bivariat normalverteilt mit den Parametern $\mu_X, \mu_Y, \sigma_X, \sigma_Y$ und ρ .

Dann Y für $X = x$ normalverteilt mit den Parametern

$$\mu_Y - \rho \frac{\sigma_Y}{\sigma_X} (x - \mu_X)$$

und

$$\sigma_Y^2 (1 - \rho^2).$$

Also gilt

$$E(Y|X = x) = \mu_Y - \rho \frac{\sigma_Y}{\sigma_X} (x - \mu_X)$$

Schauen wir uns noch einmal das Beispiel 8.5.1 an.

Es gilt

$$\begin{aligned} E(Y|X = 10) &= 10 \cdot 0.25 + 20 \cdot 0.75 \\ &= 17.5 \end{aligned}$$

und

$$\begin{aligned} E(Y|X = 20) &= 10 \cdot 0.5 + 20 \cdot 0.5 \\ &= 15 \end{aligned}$$

Der bedingte Erwartungswert hängt vom Wert von X ab.

Er kann also als Zufallsvariable aufgefaßt werden, genauer als Funktion der Zufallsvariablen X .

Wir sprechen von der bedingten Erwartung $E(Y|X)$.

Die Zufallsvariable $E(Y|X)$ kann die Werte 17.5 und 15 annehmen, wobei gilt

$$\begin{aligned} P(E(Y|X) = 17.5) &= P(X = 10) \\ &= 0.4 \end{aligned}$$

und

$$\begin{aligned} P(E(Y|X) = 15) &= P(X = 20) \\ &= 0.6 \end{aligned}$$

Somit gilt

$$\begin{aligned} E(E(Y|X)) &= 17.5 \cdot 0.4 + 15 \cdot 0.6 \\ &= 16 \end{aligned}$$

Dies ist aber gerade $E(Y)$, denn es ist

$$\begin{aligned} E(Y) &= 10 \cdot 0.4 + 20 \cdot 0.6 \\ &= 16 \end{aligned}$$

Dies ist kein Zufall, wie der folgende Satz zeigt.

Satz 8.5.1 *Es gilt*

$$E(E(Y|X)) = E(Y)$$

Beweis:

Wir beweisen den diskreten Fall.

$$\begin{aligned} E(E(Y|X)) &= \sum_x E(Y|X = x) P(X = x) \\ &= \sum_x \sum_y y P(Y = y|X = x) P(X = x) \\ &= \sum_x \sum_y y P(Y = y, X = x) \\ &= \sum_y y \sum_x P(Y = y, X = x) \\ &= \sum_y y P(Y = y) \\ &= E(Y) \end{aligned}$$

Neben dem bedingten Erwartungswert können wir auch die bedingte Varianz betrachten

Definition 8.5.2 *Die bedingte Varianz von Y gegeben $X = x$ ist definiert durch*

$$\text{Var}(Y|X = x) = E(Y^2|X = x) - E(Y|X = x)^2$$

Beispiel 8.5.4 Die gemeinsame Wahrscheinlichkeitsfunktion der Zufallsvariablen X und Y ist gegeben durch

| | | | | |
|-----|-----|------|------|------|
| | y | 10 | 20 | |
| x | | | | |
| 10 | | 0.10 | 0.30 | 0.40 |
| 20 | | 0.30 | 0.30 | 0.60 |
| | | 0.40 | 0.60 | 1.00 |

Es gilt

$$P(Y = y|X = 10) = \begin{cases} 0.25 & \text{für } y = 10 \\ 0.75 & \text{für } y = 20 \end{cases}$$

Also gilt

$$\begin{aligned} E(Y^2|X = 10) &= 100 \cdot 0.25 + 400 \cdot 0.75 \\ &= 325 \end{aligned}$$

und somit

$$\begin{aligned} \text{Var}(Y|X = 10) &= 325 - 17.5^2 \\ &= 18.75 \end{aligned}$$

Außerdem gilt

$$P(Y = y|X = 20) = \begin{cases} 0.5 & \text{für } y = 10 \\ 0.5 & \text{für } y = 20 \end{cases}$$

Also gilt

$$\begin{aligned} E(Y^2|X = 20) &= 100 \cdot 0.5 + 400 \cdot 0.5 \\ &= 250 \end{aligned}$$

Also gilt

$$\begin{aligned} \text{Var}(Y|X = 20) &= 250 - 15^2 \\ &= 25 \end{aligned}$$

Beispiel 8.5.5 Für die zweidimensionale stetige Zufallsvariable (X, Y) mit gemeinsamer Dichtefunktion

$$f_{X,Y}(x, y) = \begin{cases} 2 & \text{für } 0 \leq x \leq y \leq 1 \\ 0 & \text{sonst} \end{cases}$$

gilt

$$f_{Y|X}(y|x) = \begin{cases} \frac{1}{1-x} & \text{für } x \leq y \leq 1 \\ 0 & \text{sonst} \end{cases}$$

Also gilt

$$\begin{aligned} E(Y^2|X = x) &= \int_x^1 y^2 \frac{1}{1-x} dy \\ &= \frac{1}{1-x} \left[\frac{y^3}{3} \right]_x^1 \\ &= \frac{1-x^3}{3(1-x)} \\ &= \frac{1+x+x^2}{3} \end{aligned}$$

Also gilt

$$\begin{aligned} \text{Var}(Y|X = x) &= \frac{1+x+x^2}{3} - \frac{(1+x)^2}{4} \\ &= \frac{(1-x)^2}{12} \end{aligned}$$

Beispiel 8.5.6 Die Zufallsvariable (X, Y) sei bivariat normalverteilt mit den Parametern $\mu_X, \mu_Y, \sigma_X, \sigma_Y$ und ρ .

Dann Y für $X = x$ normalverteilt mit den Parametern

$$\mu_Y - \rho \frac{\sigma_Y}{\sigma_X} (x - \mu_X)$$

und

$$\sigma_Y^2 (1 - \rho^2).$$

Also gilt

$$\text{Var}(Y|X = x) = \sigma_Y^2 (1 - \rho^2)$$

Die bedingte Varianz $Var(Y|X = x)$ hängt wie der bedingte Erwartungswert vom Wert von X ab. Wir können auch hier $Var(Y|X)$ betrachten.

Der folgende Satz zeigt den Zusammenhang zwischen

$$Var(Y|X)$$

und

$$Var(Y).$$

Satz 8.5.2 *Es gilt*

$$Var(Y) = E(Var(Y|X)) + Var(E(Y|X))$$

Außerdem gilt

$$Var(Y) = Var(E(Y|X))$$

genau dann, wenn gilt

$$P(Y = g(X)) = 1$$

Beweis:

$$\begin{aligned} Var(Y) &= E(Y^2) - E(Y)^2 \\ &= E(E(Y^2|X)) - E(E(Y|X))^2 \\ &= E(E(Y^2|X)) - E(E(Y|X))^2 - E(E(Y|X)^2) + E(E(Y|X)^2) \\ &= E(E(Y^2|X)) - E(E(Y|X)^2) + E(E(Y|X)^2) - E(E(Y|X))^2 \\ &= E\left(E(Y^2|X) - E(Y|X)^2\right) + Var(E(Y|X)) \\ &= E(Var(Y|X)) + Var(E(Y|X)) \end{aligned}$$

$$Var(Y) = Var(E(Y|X)) \iff E(Var(Y|X)) = 0$$

$$\iff Var(Y|X = x) = 0 \quad \text{für alle } x \in \mathfrak{R}$$

$$\iff P(Y = E(Y|X = x)) = 1 \quad \text{für alle } x \in \mathfrak{R}$$

Beispiel 8.5.7 Die Zufallsvariablen X und Y besitzen folgende gemeinsame Wahrscheinlichkeitsfunktion:

| | y | 0 | 1 |
|------|-----|-------|-------|
| x | | | |
| -1 | | 0 | 0.2 |
| 0 | | 0.6 | 0 |
| 1 | | 0 | 0.2 |

Es gilt

$$\begin{aligned} E(Y) &= 0 \cdot 0.6 + 1 \cdot 0.4 \\ &= 0.4 \end{aligned}$$

und

$$\begin{aligned} E(Y^2) &= 0 \cdot 0.6 + 1 \cdot 0.4 \\ &= 0.4 \end{aligned}$$

Also gilt

$$\begin{aligned} \text{Var}(Y) &= 0.4 - 0.4^2 \\ &= 0.24 \end{aligned}$$

Wegen

$$P(Y = X^2) = 1$$

gilt auch

$$\text{Var}(E(Y|X)) = 0.24.$$

Die nachfolgenden Berechnungen bestätigen dies.

Es gilt

$$E(Y|X = -1) = 1,$$

$$E(Y|X = 0) = 0$$

und

$$E(Y|X = 1) = 1$$

Somit gilt

$$\begin{aligned} E(E(Y|X)) &= 1 \cdot 0.2 + 0 \cdot 0.6 + 1 \cdot 0.2 \\ &= 0.4 \end{aligned}$$

und

$$\begin{aligned} E(E(Y|X)^2) &= 1 \cdot 0.2 + 0 \cdot 0.6 + 1 \cdot 0.2 \\ &= 0.4 \end{aligned}$$

Also gilt

$$\begin{aligned} \text{Var}(E(Y|X)) &= E(E(Y|X)^2) - E(E(Y|X))^2 \\ &= 0.4 - 0.4^2 \\ &= 0.24 \end{aligned}$$

Der Übergang zu bedingten Erwartungswerten erleichtert oft die Berechnung von unbedingten Erwartungswerten.

Dies ist im folgenden Beispiel der Fall.

Beispiel 8.5.8 Ein Modell für eine Zählvariable Y bei vorgegebener Anzahl von Durchführungen ist die Binomialverteilung. Neben n besitzt diese noch den Parameter w , die Erfolgswahrscheinlichkeit bei einmaliger Durchführung. Bei einer Reihe von Anwendungen ist w nicht fest, sondern wird als Realisation einer Zufallsvariablen W aufgefaßt.

In der Regel unterstellt man, daß W mit den Parametern a und b betaverteilt ist.

Die Dichtefunktion von W lautet:

$$f_W(w) = \begin{cases} \frac{1}{B(a,b)} w^{a-1} (1-w)^{b-1} & \text{für } 0 < w < 1, a, b > 0 \\ 0 & \text{sonst} \end{cases}$$

mit

$$B(a,b) = \int_0^1 w^{a-1} (1-w)^{b-1} dw$$

Es gilt

$$E(W) = \frac{a}{a+b},$$

$$E(W^2) = \frac{a(a+1)}{(a+b)(a+b+1)}$$

und

$$\text{Var}(W) = \frac{ab}{(a+b)^2(a+b+1)}$$

Gegeben $W = w$ ist Y binomialverteilt mit den Parametern n und w .

Es gilt also

$$f_{Y|W}(y|w) = \binom{n}{y} w^y (1-w)^{n-y}$$

Somit folgt

$$\begin{aligned} f_{Y,W}(y, w) &= f_{Y|W}(y|w) f_W(w) \\ &= \binom{n}{y} w^y (1-w)^{n-y} \frac{1}{B(a, b)} w^{a-1} (1-w)^{b-1} \\ &= \binom{n}{y} \frac{1}{B(a, b)} w^{y+a-1} (1-w)^{n+b-1-y} \end{aligned}$$

Die Randverteilung von Y erhält man dann durch

$$\begin{aligned} f_Y(y) &= \int_0^1 f_{Y,W}(y, w) dw \\ &= \int_0^1 \binom{n}{y} \frac{1}{B(a, b)} w^{y+a-1} (1-w)^{n+b-1-y} dw \\ &= \binom{n}{y} \frac{B(a+y, n+b-y)}{B(a, b)} \end{aligned}$$

Man spricht von der Beta-Binomial-Verteilung.

Den Erwartungswert von Y erhalten wir dann folgendermaßen:

$$\begin{aligned} E(Y) &= E(E(Y|W)) \\ &= E(nW) \\ &= n \frac{a}{a+b} \end{aligned}$$

Und die Varianz können wir folgendermaßen bestimmen:

$$\begin{aligned}
 \text{Var}(Y) &= E(\text{Var}(Y|W)) + \text{Var}(E(Y|W)) \\
 &= E(nW(1-W)) + \text{Var}(nW) \\
 &= nE(W) - nE(W^2) + n^2\text{Var}(W) \\
 &= n\frac{a}{a+b} - n\frac{a(a+1)}{(a+b)(a+b+1)} + n^2\frac{ab}{(a+b+1)(a+b)^2} \\
 &= n\frac{ab(n+a+b)}{(a+b+1)(a+b)^2}
 \end{aligned}$$

Mit

$$w = \frac{a}{a+b}$$

gilt also:

$$E(Y) = nw$$

und

$$\text{Var}(Y) = nw(1-w)\frac{n+a+b}{1+a+b}$$

Die Varianz der Beta-Binomial-Verteilung ist somit größer als die Varianz der Binomialverteilung. Mit Hilfe der Beta-Binomial-Verteilung wird Überdispersion modelliert.

Beispiel 8.5.9 In Beispiel 4.13 haben wir eine Zählvariable Y dadurch modelliert, daß wir angenommen haben, daß Y gegeben $\Lambda = \lambda$ poissonverteilt ist mit Parameter λ .

Es gilt also

$$E(Y|\Lambda) = \Lambda$$

und

$$\text{Var}(Y|\Lambda) = \Lambda$$

Weiterhin wurde unterstellt, daß Λ gammaverteilt ist mit den Parametern r und β .

Es gilt also

$$E(\Lambda) = \frac{r}{\beta}$$

und

$$\text{Var}(\Lambda) = \frac{r}{\beta^2}$$

Somit können wir den Erwartungswert und die Varianz von Y einfach bestimmen.

Es gilt

$$\begin{aligned} E(Y) &= E(E(Y|\Lambda)) \\ &= E(\Lambda) \\ &= \frac{r}{\beta} \end{aligned}$$

und

$$\begin{aligned} \text{Var}(Y) &= E(\text{Var}(Y|\Lambda)) + \text{Var}(E(Y|\Lambda)) \\ &= E(\Lambda) + \text{Var}(\Lambda) \\ &= \frac{r}{\beta} + \frac{r}{\beta^2} \\ &= \frac{r(1 + \beta)}{\beta^2} \end{aligned}$$

In Beispiel 4.13 haben wir gesehen, daß Y pascalverteilt ist mit den Parametern r und $p = \frac{\beta}{1+\beta}$.

Wir können $E(Y)$ und $Var(Y)$ auch direkt bestimmen.

Es gilt

$$\begin{aligned} E(Y) &= r \frac{1-p}{p} \\ &= r \frac{1 - \frac{\beta}{1+\beta}}{\frac{\beta}{1+\beta}} \\ &= \frac{r}{\beta} \end{aligned}$$

+ und

$$\begin{aligned} Var(Y) &= r \frac{1-p}{p^2} \\ &= r \frac{1 - \frac{\beta}{1+\beta}}{\left(\frac{\beta}{1+\beta}\right)^2} \\ &= r \frac{(1+\beta)}{\beta^2} \end{aligned}$$

8.6 Prognose

Oft ist der Wert x einer Zufallsvariablen X bekannt. Von Interesse ist nun der Wert y der Zufallsvariablen Y .

Man will den Wert von Y mit Hilfe von X prognostizieren.

Wir beginnen mit dem einfachen Fall, daß die Zufallsvariablen X und Y unabhängig sind. In diesem Fall wird die Prognose nicht von X abhängen.

Wir prognostizieren Y also durch eine Konstante c .

Welchen Wert sollen wir für c wählen?

Schauen wir uns eine diskrete Zufallsvariable Y mit den Realisierungsmöglichkeiten y_1, \dots, y_k und zugehörigen Wahrscheinlichkeiten p_1, \dots, p_k an.

Es liegt nahe, c so zu wählen, daß c möglichst nahe an allen Realisationsmöglichkeiten von Y liegt.

Als Maß für die Nähe bietet sich der Abstand zwischen y_i und c für $i = 1, \dots, k$ an.

Wir betrachten also $|y_1 - c|, \dots, |y_k - c|$.

Die Konstante c sollte möglichst nah an allen Realisierungsmöglichkeiten von Y liegen. Wir gewichten also die $|y_i - c|$ mit ihren Wahrscheinlichkeiten p_i und summieren auf.

Die Konstante c sollte also so gewählt werden, daß

$$E(|Y - c|) = \sum_{i=1}^k |y_i - c| p_i$$

minimal wird.

Dieses Kriterium hat sich aber in der Statistik nicht durchgesetzt.

Es werden stattdessen die quadrierten Abstände betrachtet.

$$E((Y - c)^2) = \sum_{i=1}^k (y_i - c)^2 p_i$$

Man nennt $E((Y - c)^2)$ auch den mittleren quadratischen Fehler.

Wir suchen nun den Wert von c , für den $E((Y - c)^2)$ minimal wird.

Es gilt

$$\begin{aligned} E((Y - c)^2) &= \text{Var}(Y - c) + (E(Y - c))^2 \\ &= \text{Var}(Y) + (E(Y) - c)^2 \end{aligned}$$

Der erste Summand hängt nicht von c ab. Der zweite Summand wird minimal, wenn gilt

$$c = E(Y).$$

Die im Sinne des mittleren quadratischen Fehlers beste Prognose der Zufallsvariablen Y ist also der Erwartungswert von Y .

Nun wollen wir annehmen, daß X und Y nicht unabhängig sind. Wir wollen also Y durch eine geeignete Funktion $g(X)$ von X prognostizieren.

Wie sollte $g(X)$ gewählt werden?

Wir betrachten den mittleren quadratischen Fehler

$$E((Y - g(X))^2)$$

als Entscheidungskriterium.

Es gilt

$$E((Y - g(X))^2) = E[E((Y - g(X))^2)|X]$$

Wie wir oben gesehen haben, wird für jeden Wert von x

$$E((Y - g(X))^2)$$

durch

$$E(Y|X = x)$$

minimiert.

Also ist die optimale Wahl von $g(X)$ durch $g(X) = E(Y|X)$ gegeben.

Beispiel 8.6.1 *Bei der bivariaten Normalverteilung mit den Parametern μ_X , μ_Y , σ_X , σ_Y und ρ gilt:*

$$E(Y|X = x) = \mu_Y - \rho \frac{\sigma_Y}{\sigma_X} (x - \mu_X)$$

Die beste Prognose von Y durch X ist somit durch eine lineare Funktion in X gegeben.

Beispiel 8.6.2 *Für die zweidimensionale stetige Zufallsvariable (X, Y) mit gemeinsamer Dichtefunktion*

$$f_{X,Y}(x, y) = \begin{cases} 3 & \text{für } 0 \leq y \leq x^2 \leq 1 \\ 0 & \text{sonst} \end{cases}$$

gilt

$$\begin{aligned} f_X(x) &= \int_0^{x^2} 3 dy \\ &= 3x^2 \end{aligned}$$

für $0 \leq x \leq 1$. Also gilt

$$f_{Y|X}(y|x) = \begin{cases} \frac{1}{x^2} & \text{für } 0 \leq y \leq x^2 \\ 0 & \text{sonst} \end{cases}$$

Somit gilt

$$\begin{aligned} E(Y|X = x) &= \int_0^{x^2} y \frac{1}{x^2} dy \\ &= \frac{1}{x^2} \left[\frac{y^2}{2} \right]_0^{x^2} \\ &= \frac{x^2}{2} \end{aligned}$$

Für den mittleren quadratischen Fehler gilt

$$\begin{aligned} E((Y - E(Y|X))^2) &= E((Y - 0.5 X^2)^2) \\ &= \int_0^1 \int_0^{x^2} (y - 0.5 x^2)^2 3 dy dx \\ &= \int_0^1 \int_0^{x^2} (3 y^2 - 3 y x^2 + 0.75 x^4) dy dx \\ &= \int_0^1 [y^3 - 1.5 y^2 x^2 + 0.75 y x^4]_0^{x^2} dx \\ &= \int_0^1 (x^6 - 1.5 x^6 + 0.75 x^6) dx = \\ &= \int_0^1 0.25 x^6 dx = \left[\frac{x^7}{28} \right]_0^1 \\ &= \frac{1}{28} \\ &= 0.0357 \end{aligned}$$

Es ist nicht immer einfach, $E(Y|X)$ zu bestimmen.

In diesen Fällen sucht man nach einer einfachen Approximation von $h(X)$.

Der einfachste Fall ist

$$h(X) = a + bX.$$

Wir suchen also Werte für a und b , so daß

$$g(a, b) = E((Y - a - bX)^2)$$

minimal ist.

Es gilt

$$\begin{aligned} E((Y - a - bX)^2) &= \text{Var}(Y - a - bX) + (E(Y - a - bX))^2 \\ &= \text{Var}(Y - bX) + (E(Y) - a - bE(X))^2 \end{aligned}$$

Der erste Summand hängt nicht von a ab.

Deshalb wählen wir a so, daß es den zweiten Summanden minimiert.

Dies ist für

$$a = E(Y) - bE(X)$$

der Fall.

In diesem Fall ist der zweite Summand 0.

Somit müssen wir noch den Wert von b finden, für den der erste Summand minimal wird.

Es gilt

$$\text{Var}(Y - bX) = \text{Var}(Y) - 2b\text{Cov}(X, Y) + b^2\text{Var}(X)$$

Die notwendige Bedingung für einen Extremwert dieser in b quadratischen Funktion ist, daß die erste Ableitung nach b gleich 0 ist:

$$\frac{d}{db} \text{Var}(Y - bX) = 0$$

Es muß also gelten:

$$-2\text{Cov}(X, Y) + 2b\text{Var}(X) = 0$$

Diese Gleichung wird erfüllt durch:

$$b = \frac{\text{Cov}(X, Y)}{\text{Var}(X)}$$

Wegen

$$\frac{d^2}{db^2} \text{Var}(Y - bX) = 2 \text{Var}(X)$$

handelt es sich um ein Minimum, wenn X keine Einpunktverteilung besitzt. Der minimale Wert von $E((Y - a - bX)^2)$ ist wegen $a = E(Y) - bE(X)$:

$$E((Y - a - bX)^2) = \text{Var}(Y) (1 - \rho_{X,Y}^2)$$

Dies sieht man folgendermaßen:

$$\begin{aligned} E((Y - a - bX)^2) &= \text{Var}(Y - bX) \\ &= \text{Var}(Y) - 2b \text{Cov}(X, Y) + b^2 \text{Var}(X) \\ &= \text{Var}(Y) - 2 \frac{\text{Cov}(X, Y)}{\text{Var}(X)} \text{Cov}(X, Y) + \\ &\quad + \frac{\text{Cov}(X, Y)^2}{\text{Var}(X)^2} \text{Var}(X) \\ &= \text{Var}(Y) - 2 \frac{\text{Cov}(X, Y)^2}{\text{Var}(X)} + \frac{\text{Cov}(X, Y)^2}{\text{Var}(X)} \\ &= \text{Var}(Y) - \frac{\text{Cov}(X, Y)^2}{\text{Var}(X)} \\ &= \text{Var}(Y) - \frac{\text{Cov}(X, Y)^2 \text{Var}(Y)}{\text{Var}(X) \text{Var}(Y)} \\ &= \text{Var}(Y) - \text{Var}(Y) \rho_{X,Y}^2 \\ &= \text{Var}(Y) (1 - \rho_{X,Y}^2) \end{aligned}$$

Schauen wir uns ein Beispiel an:

Beispiel 8.6.3 *Wir bestimmen die beste lineare Prognose von Y durch X für die zweidimensionale stetige Zufallsvariable (X, Y) mit gemeinsamer Dichtefunktion*

$$f_{X,Y}(x, y) = \begin{cases} 3 & \text{für } 0 \leq y \leq x^2 \leq 1 \\ 0 & \text{sonst} \end{cases}$$

Wie wir gesehen haben, gilt

$$f_X(x) = \begin{cases} 3x^2 & \text{für } 0 \leq x \leq 1 \\ 0 & \text{sonst} \end{cases}$$

Also gilt

$$\begin{aligned} E(X) &= \int_0^1 x \cdot 3x^2 dx \\ &= \frac{3}{4} \end{aligned}$$

und

$$\begin{aligned} E(X^2) &= \int_0^1 x^2 \cdot 3x^2 dx \\ &= \frac{3}{5} \end{aligned}$$

und somit gilt

$$\text{Var}(X) = \frac{3}{80}$$

Außerdem gilt für $0 \leq y \leq 1$:

$$\begin{aligned} f_Y(y) &= \int_{\sqrt{y}}^1 3 dx \\ &= 3 - 3\sqrt{y} \end{aligned}$$

Also gilt

$$\begin{aligned} E(Y) &= \int_0^1 y(3 - 3\sqrt{y}) dy \\ &= \frac{3}{10} \end{aligned}$$

und

$$\begin{aligned} E(Y^2) &= \int_0^1 y^2 (3 - 3\sqrt{y}) dy \\ &= \frac{1}{7} \end{aligned}$$

und somit gilt

$$\text{Var}(Y) = \frac{37}{700}$$

Um $\text{Cov}(X, Y)$ zu berechnen, bestimmen wir

$$\begin{aligned} E(XY) &= \int_0^1 \int_0^{x^2} x y 3 dy dx \\ &= \int_0^1 3x \left[\frac{y^2}{2} \right]_0^{x^2} dx \\ &= \int_0^1 3x \frac{x^4}{2} dx \\ &= \left[\frac{x^6}{4} \right]_0^1 \\ &= \frac{1}{4} \end{aligned}$$

Also gilt

$$\text{Cov}(X, Y) = \frac{1}{40}$$

Somit gilt

$$\begin{aligned} b &= \frac{\text{Cov}(X, Y)}{\text{Var}(X)} \\ &= \frac{2}{3} \end{aligned}$$

und

$$\begin{aligned} a &= E(Y) - b E(X) \\ &= \frac{3}{10} - \frac{2}{3} \frac{3}{4} \\ &= -\frac{1}{5} \end{aligned}$$

Der minimale Wert von $E((Y - a - bX)^2)$ ist somit

$$\begin{aligned} \text{Var}(Y - bX) &= \text{Var}(Y) (1 - \rho_{X,Y}^2) \\ &= \frac{37}{700} \left(1 - \frac{35}{111}\right) \\ &= 0.0363 \end{aligned}$$

Beispiel 8.6.4 Sei (X, Y) eine zweidimensionale diskrete Zufallsvariable mit gemeinsamer Wahrscheinlichkeitsfunktion

$$P(X = x, Y = y) = \begin{cases} 0.2 & \text{für } x = 1, y = 3 \\ 0.2 & \text{für } x = 2, y = 2 \\ 0.2 & \text{für } x = 3, y = 5 \\ 0.2 & \text{für } x = 4, y = 4 \\ 0.2 & \text{für } x = 5, y = 6 \end{cases}$$

Es gilt

$$\begin{aligned} E(X) &= 1 \cdot 0.2 + 2 \cdot 0.2 + 3 \cdot 0.2 + 4 \cdot 0.2 + 5 \cdot 0.2 \\ &= 3, \end{aligned}$$

$$\begin{aligned} E(X^2) &= 1 \cdot 0.2 + 4 \cdot 0.2 + 9 \cdot 0.2 + 16 \cdot 0.2 + 25 \cdot 0.2 \\ &= 11, \end{aligned}$$

$$\begin{aligned} \text{Var}(X) &= E(X^2) - E(X)^2 \\ &= 11 - 9 \\ &= 2, \end{aligned}$$

$$\begin{aligned} E(Y) &= 3 \cdot 0.2 + 2 \cdot 0.2 + 5 \cdot 0.2 + 4 \cdot 0.2 + 6 \cdot 0.2 \\ &= 4, \end{aligned}$$

$$\begin{aligned} E(Y^2) &= 9 \cdot 0.2 + 4 \cdot 0.2 + 25 \cdot 0.2 + 16 \cdot 0.2 + 36 \cdot 0.2 \\ &= 18 \end{aligned}$$

und

$$\begin{aligned} \text{Var}(Y) &= E(Y^2) - E(Y)^2 \\ &= 18 - 16 \\ &= 2 \end{aligned}$$

Außerdem gilt

$$\begin{aligned} E(XY) &= 1 \cdot 3 \cdot 0.2 + 2 \cdot 2 \cdot 0.2 + 3 \cdot 5 \cdot 0.2 + 4 \cdot 4 \cdot 0.2 + 5 \cdot 6 \cdot 0.2 \\ &= 13.6 \end{aligned}$$

Also gilt

$$\begin{aligned} \text{Cov}(X, Y) &= 13.6 - 3 \cdot 4 \\ &= 1.6 \end{aligned}$$

Somit gilt

$$\begin{aligned} b &= \frac{\text{Cov}(X, Y)}{\text{Var}(X)} \\ &= \frac{1.6}{2} \\ &= 0.8 \end{aligned}$$

und

$$\begin{aligned} a &= E(Y) - b E(X) \\ &= 4 - 0.8 \cdot 3 \\ &= 1.6 \end{aligned}$$