

Kapitel 3

Allgemeine Wahrscheinlichkeitsräume

3.1 Einleitung

Wir hatten schon bemerkt, dass der Begriff des diskreten Wahrscheinlichkeitsraums nicht ausreicht, um das unendliche Wiederholen eines Zufallsexperiments zu modellieren. Der Grund dafür ist, dass die Menge aller möglichen Ausgänge nicht mehr abzählbar ist. Ein weiterer Unterschied zu diskreten Wahrscheinlichkeitsräumen ist, dass es nicht mehr notwendigerweise Elementarereignisse positiver Wahrscheinlichkeit geben muss. Wir diskutieren dies an einem Beispiel.

Beispiel 3.1.1. Werfen wir unendlich oft eine faire Münze, so bietet sich als Grundraum der Raum

$$\Omega := \{\omega = (\omega_1, \omega_2, \omega_3, \dots) : \omega_n \in \{K, Z\} \forall n \in \mathbb{N}\}$$

aller Folgen in $\{K, Z\}$ an. Das Elementarereignis $\omega = (\omega_1, \omega_2, \dots)$ bedeutet hierbei gerade, dass im ersten Wurf ω_1 , im zweiten Wurf ω_2 usw. geworfen wurde. Beachte, dass Ω nicht abzählbar ist.

Es sei

$$A_1 = \{\omega \in \Omega : \omega_1 = K\} \subseteq \mathcal{P}(\Omega).$$

Dann bezeichnet A das Ereignis “Im ersten Wurf Kopf” und sollte gerade Wahrscheinlichkeit $\frac{1}{2}$ haben. Ist allgemeiner

$$A_n = \{\omega \in \Omega : \omega_1 = \omega_2 = \dots = \omega_n = K\}$$

das Ereignis, dass in den ersten n Würfeln Kopf gefallen ist, so sollte A_n Wahrscheinlichkeit 2^{-n} haben.

Betrachten wir nun das Elementarereignis $A_\infty := \{(K, K, K, \dots)\}$, dass in allen Würfeln Kopf fällt, so ist $A_\infty \subseteq A_n$ für alle $n \in \mathbb{N}$. Aufgrund der Monotonie der Wahrscheinlichkeit sollte $\mathbb{P}(A_\infty) \leq \mathbb{P}(A_n) = 2^{-n}$ für alle $n \in \mathbb{N}$ gelten. Es folgt also $\mathbb{P}(A_\infty) = 0$.

Analog folgt, dass alle Elementarereignisse Wahrscheinlichkeit 0 haben müssen.

Ein weiteres Beispiel ergibt sich bei Zufallsvariablen die stetige Merkmale, wie Temperatur oder Zeit, beschreiben. Hier möchte man auf, dass die Zufallsvariablen keinen Wert mit positiver Wahrscheinlichkeit annimmt. In einem Wahrscheinlichkeitsraum, auf dem eine solche Zufallsvariable definiert ist, müssen alle Elementarereignisse Wahrscheinlichkeit 0 haben.

Ein weiterer Unterschied zwischen allgemeinen und diskreten Wahrscheinlichkeitsräumen betrifft die Additivität: Haben alle Elementarereignisse Wahrscheinlichkeit 0, so muss diese eingeschränkt werden, damit

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\}) = 0$$

nicht für alle Ereignisse $A \subseteq \Omega$ gilt.

Es bleibt die Frage, ob es in dieser Situation stets ein Wahrscheinlichkeitsmaß (im Sinne einer σ -additiven Abbildung von $\mathcal{P}(\Omega)$ nach $[0, 1]$) gibt. Leider ist dies nicht immer der Fall. Es zeigt sich, dass die Potenzmenge $\mathcal{P}(\Omega)$ in der Regel zu groß ist.

Deshalb werden wir nicht mehr jede Teilmenge von Ω als Ereignis betrachten. Wir nehmen aber an, dass die Menge aller Ereignisse eine sogenannte σ -Algebra bildet.

Definition 3.1.2. Es sei Ω eine Menge. Eine σ -Algebra auf Ω ist ein Mengensystem $\Sigma \subseteq \mathcal{P}(\Omega)$, sodass

- (i) $\emptyset \in \Sigma$.
- (ii) Ist $A \in \Sigma$, so ist auch $A^c = \Omega \setminus A \in \Sigma$.
- (iii) Ist $A_n \in \Sigma$ für $n \in \mathbb{N}$, so ist auch $\bigcup_{n \in \mathbb{N}} A_n \in \Sigma$.

Beispiel 3.1.3. a) Offensichtlich ist $\mathcal{P}(\Omega)$ eine σ -Algebra auf Ω .

b) Eine weitere σ -Algebra auf Ω ist $\{\emptyset, \Omega\}$.

c) Ist Ω eine Menge mit mindestens zwei Elementen und $\emptyset \neq A \subseteq \Omega$ mit $A \neq \Omega$, so ist $\{\emptyset, A, A^c, \Omega\}$ eine σ -Algebra.

Bemerkung 3.1.4. Weiterhin erfüllt eine σ -Algebra Σ auf Ω auch folgende Eigenschaften:

- (1) Sind $A_1, \dots, A_k \in \Sigma$, so auch $A_1 \cup \dots \cup A_k$. Das folgt aus (iii), indem man $A_n = \emptyset$ für $n > k$ wählt.
- (2) Sind $A_1, \dots, A_k \in \Sigma$, so auch $A_1 \cap \dots \cap A_k$. Das folgt aus (1), (ii) und deMorgan's Gesetz (siehe Proposition 3.1.5) unten.
- (3) Genau so sieht man, dass Σ mit der Folge $(A_n)_{n \in \mathbb{N}}$ auch deren Durchschnitt $\bigcap_{n \in \mathbb{N}} A_n$ enthält.

Proposition 3.1.5. Seien $A_i, i \in I$, endlich oder unendlich viele Teilmengen einer Grundmenge Ω . Dann ist

$$\left(\bigcup_{i \in I} A_i \right)^c = \bigcap_{i \in I} A_i^c \quad \text{und} \quad \left(\bigcap_{i \in I} A_i \right)^c = \bigcup_{i \in I} A_i^c.$$

In der Regel ist es schwer, eine σ -Algebra konkret anzugeben. Daher ist folgendes Resultat wichtig:

Lemma 3.1.6. Ist $S \subseteq \mathcal{P}(\Omega)$, so gibt es eine kleinste σ -Algebra auf Ω , die S enthält. Präziser: Es gibt eine σ -Algebra Σ_0 auf Ω , so dass $S \subseteq \Sigma_0$ gilt und für jede σ -Algebra Σ auf Ω mit $S \subseteq \Sigma$ auch $\Sigma_0 \subseteq \Sigma$ gilt.

Diese σ -Algebra Σ_0 bezeichnet man mit $\sigma(S)$ und nennt sie die von S erzeugte σ -Algebra.

Definition 3.1.7. Die von den offenen Intervallen in \mathbb{R} erzeugte σ -Algebra heißt *Borel σ -Algebra* und wird mit $\mathcal{B}(\mathbb{R})$ bezeichnet. Es ist also $\mathcal{B}(\mathbb{R}) = \sigma(\{(a, b) : a, b \in \mathbb{R}, a < b\})$.

Es ist nicht möglich $\mathcal{B}(\mathbb{R})$ genauer zu beschreiben, aber alle praktisch relevanten Teilmengen von \mathbb{R} liegen in $\mathcal{B}(\mathbb{R})$. Es gilt $\mathcal{B}(\mathbb{R}) \neq \mathcal{P}(\mathbb{R})$, aber dies ist schwer zu zeigen.

Wir diskutieren dies nicht näher und geben stattdessen Beispiele für Mengen in $\mathcal{B}(\mathbb{R})$.

Beispiel 3.1.8. Für $a \in \mathbb{R}$ ist $(-\infty, a) \in \mathcal{B}(\mathbb{R})$ und $(a, \infty) \in \mathcal{B}(\mathbb{R})$. Es ist nämlich $(-\infty, a) = \bigcup_{n \in \mathbb{N}} (a - n, a)$ und $(a, \infty) = \bigcup_{n \in \mathbb{N}} (a, a + n)$. Somit liegen für $a \in \mathbb{N}$ auch die Komplemente $[a, \infty) = (-\infty, a)^c$ und $(-\infty, a] = (a, \infty)^c$ in \mathcal{B} . Schließlich sind auch kompakte Intervalle in $\mathcal{B}(\mathbb{R})$ enthalten, denn $[a, b] = (-\infty, b] \cap [a, \infty)$.

Vereinbarung: In Übungs- und Klausuraufgaben zur Vorlesung für “Stochastik für Wirtschaftswissenschaftler” braucht für Teilmengen von \mathbb{R} nicht nachgewiesen zu werden, dass diese in der Borel- σ -Algebra liegen.

Wir kommen nun zur zentralen Definition:

Definition 3.1.9. Ein *Messraum* ist ein Paar (Ω, Σ) , bestehend aus einer Menge Ω und einer σ -Algebra Σ auf Ω . Ist (Ω, Σ) ein Messraum, so heißt eine Abbildung $\mathbb{P} : \Sigma \rightarrow [0, 1]$ *Wahrscheinlichkeitsmaß* auf (Ω, Σ) , falls

1. $\mathbb{P}(\Omega) = 1$ (Normiertheit)
2. Ist $(A_n)_{n \in \mathbb{N}}$ eine Folge paarweise disjunkter Mengen in Σ , so ist

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

Ist (Ω, Σ) ein Messraum und \mathbb{P} ein Wahrscheinlichkeitsmaß darauf, so nennt man das Tripel $(\Omega, \Sigma, \mathbb{P})$ einen *Wahrscheinlichkeitsraum*.

Bemerkung 3.1.10. (a) Beachte, dass Σ mit der Folge A_n auch deren Vereinigung $\bigcup A_n$ enthält. Daher ist $\mathbb{P}(\bigcup A_n)$ in (ii) wohldefiniert.

- (b) Jeder diskrete Wahrscheinlichkeitsraum ist ein Wahrscheinlichkeitsraum mit der Potenzmenge als σ -Algebra.
- (c) Die Eigenschaften von Wahrscheinlichkeitsmaßen in Proposition 1.1.5 gelten auch in beliebigen Wahrscheinlichkeitsräumen. Der Beweis bleibt unverändert.

Auch andere Konzepte (wie beispielsweise die Begriffe “Unabhängigkeit” und “bedingte Wahrscheinlichkeit”) übertragen sich ohne Änderungen auf allgemeine Wahrscheinlichkeitsräume.

Lemma 3.1.11. *Es sei $(\Omega, \Sigma, \mathbb{P})$ ein Wahrscheinlichkeitsraum und A_n eine Folge in Σ mit $A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots$ und $A := \bigcup_{n \in \mathbb{N}} A_n$. (Wir sagen: A_n wächst gegen A und schreiben $A_n \uparrow A$. Dann ist*

$$\mathbb{P}(A) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n).$$

Ist C_n eine Folge in Σ mit $C_1 \supseteq C_2 \supseteq \dots$ und $C := \bigcap_{n \in \mathbb{N}} C_n$ (Wir sagen C_n fällt gegen C und schreiben $C_n \downarrow C$), so gilt ebenfalls

$$\mathbb{P}(C) = \lim_{n \rightarrow \infty} \mathbb{P}(C_n).$$

Beweis. Übung!

□

3.2 Zufallsvariablen und ihre Verteilungen

Gegeben einen Wahrscheinlichkeitsraum $(\Omega, \Sigma, \mathbb{P})$ sind wir versucht, wiederum jede Abbildung X von Ω nach \mathbb{R} Zufallsvariable zu nennen. Dabei gibt es jedoch folgendes Problem:

Wenn wir die Verteilung \mathbb{P}_X von X durch $\mathbb{P}_X(A) := \mathbb{P}(X \in A)$ definieren, so muss $\{X \in A\}$ in der σ -Algebra Σ liegen.

Wir definieren daher:

Definition 3.2.1. Sei $(\Omega, \Sigma, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Eine Abbildung $X : \Omega \rightarrow \mathbb{R}$ heißt *Zufallsvariable*, falls für $A \in \mathcal{B}(\mathbb{R})$ stets $\{\omega \in \Omega : X(\omega) \in A\} \in \Sigma$ liegt. In diesem Fall heißt das Wahrscheinlichkeitsmaß \mathbb{P}_X auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, gegeben durch

$$\mathbb{P}_X(A) := \mathbb{P}(X \in A),$$

die *Verteilung von X* .

Wir hatten bereits bemerkt, dass die Borel σ -Algebra $\mathcal{B}(\mathbb{R})$ sehr groß und unübersichtlich ist. Daher ist es schwierig, die Verteilung konkret anzugeben. Es ist auch nicht mehr richtig, dass Verteilungen durch ihre Werte auf einelementigen Mengen bestimmt sind. Es stellt sich also die Frage, ob die Verteilung einer Zufallsvariablen bereits durch die Werte auf gewissen Mengen eindeutig bestimmt ist.

Dies ist in der Tat der Fall. Wir definieren:

Definition 3.2.2. Sei $(\Omega, \Sigma, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable. Dann heißt $F_X : \mathbb{R} \rightarrow [0, 1]$, definiert durch

$$F_X(x) := \mathbb{P}(X \leq x)$$

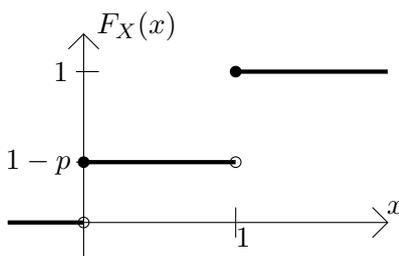
Verteilungsfunktion von X .

Proposition 3.2.3. Seien X und Y zwei Zufallsvariablen mit der selben Verteilungsfunktion, $F_X(t) = F_Y(t)$ für alle $t \in \mathbb{R}$. Dann haben X und Y auch die selbe Verteilung, $\mathbb{P}_X(A) = \mathbb{P}_Y(A)$ für beliebige $A \in \mathcal{B}(\mathbb{R})$.

Beispiel 3.2.4. Ist $X \sim \text{Bin}(1, p)$ Bernoulli-verteilt, so ist

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 1 - p, & 0 \leq x < 1 \\ 1, & x \geq 1. \end{cases}$$

Denn für $x < 0$ ist $\{X \leq x\} = \emptyset$, also $\mathbb{P}(X \leq x) = 0$. Für $0 \leq x < 1$ ist $\{X \leq x\} = \{X = 0\}$ und daher $\mathbb{P}(X \leq x) = \mathbb{P}(X = 0) = 1 - p$. Schließlich ist für $x \geq 1$ gerade $\{X \leq x\} = \Omega$ und daher $F_X(x) = \mathbb{P}(\Omega) = 1$.



Ähnliche Überlegungen zeigen folgendes:

Beispiel 3.2.5. Ist X eine Zufallsvariable die lediglich die Werte $x_1 < x_2 < \dots < x_n$ annimmt mit $\mathbb{P}(X = x_k) = p_k$, so ist

$$F_X(x) = \begin{cases} 0 & x < x_1 \\ \sum_{j=1}^k p_j & x_k \leq x < x_{k+1} \\ 1 & x \geq x_n \end{cases}$$

Das Schaubild der Verteilungsfunktion sieht ähnlich wie oben aus, hat aber nun n Sprungstellen.

Nimmt allgemeiner die Zufallsvariable X abzählbar viele Werte $x_k, k \in \mathbb{N}$, an mit $\mathbb{P}(X = x_k) = p_k$, so ist

$$F_X(x) = \sum_{k: x_k \leq x} p_k$$

In diesem Fall sagt man X habe *diskrete Verteilung* oder manchmal X sei diskret.

Die Verteilungsfunktion F_X hat in diesem Fall unendlich viele Sprungstellen – es handelt sich jedoch weiterhin um eine Funktion, die nicht stetig wächst, sondern Sprungstellen hat und dazwischen konstant ist.

Ein Sprung in einer Verteilungsfunktion an der Stelle x in Höhe p bedeutet, dass die Zufallsvariable den Wert x mit Wahrscheinlichkeit p annimmt.

Wir stellen einige Eigenschaften von Verteilungsfunktionen zusammen:

Eine Funktion $g : \mathbb{R} \rightarrow \mathbb{R}$ heißt *rechtsseitig stetig*, falls für jede monoton fallende, konvergente Folge $(x_n)_{n \in \mathbb{N}}$ gilt $\lim_{n \rightarrow \infty} g(x_n) = g(\lim_{n \rightarrow \infty} x_n)$.

Lemma 3.2.6. *Es sei X eine Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \Sigma, \mathbb{P})$. Dann ist die Verteilungsfunktion F_X monoton wachsend, rechtsseitig stetig und erfüllt*

$$\lim_{x \rightarrow -\infty} F_X(x) = 0 \quad \text{und} \quad \lim_{x \rightarrow \infty} F_X(x) = 1.$$

Beweis. Zunächst ist F offensichtlich monoton wachsend (jedoch nicht notwendigerweise strikt). Ist nämlich $x \leq y$ so ist $\{X \leq x\} \subseteq \{X \leq y\}$ und daher, wegen der Monotonie des Wahrscheinlichkeitsmaßes, $F_X(x) = \mathbb{P}(X \leq x) \leq \mathbb{P}(X \leq y) = F_X(y)$.

Ist nun x_n eine fallende Folge, die gegen x konvergiert, so ist $\{X \leq x_n\} \downarrow \{X \leq x\}$. Es folgt aus Lemma 3.1.11 dass $F_X(x_n) \rightarrow F_X(x)$. Also ist F rechtsseitig stetig.

Es gilt $\lim_{x \rightarrow -\infty} F_X(x) = 0$ und $\lim_{x \rightarrow \infty} F_X(x) = 1$. Das folgt aus Lemma 3.1.11 und den Beziehungen $\{X \leq -x_n\} \downarrow \emptyset$ und $\{X \leq x_n\} \uparrow \Omega$ für jede Folge $(x_n)_{n \in \mathbb{N}}$, die gegen ∞ konvergiert. \square

Es gilt auch die Umkehrung dieses Lemmas. Da der Beweis deutlich schwieriger ist, übergehen wir ihn.

Satz 3.2.7. *Es sei $F : \mathbb{R} \rightarrow \mathbb{R}$ eine monoton wachsende, rechtsseitig stetige Funktion mit $\lim_{x \rightarrow -\infty} F(x) = 0$ und $\lim_{x \rightarrow \infty} F(x) = 1$. Dann gibt es genau ein Wahrscheinlichkeitsmaß \mathbb{P}_F auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit*

$$\mathbb{P}_F((-\infty, x]) = F(x) \tag{3.1}$$

Eine wichtige Klasse von Verteilungen sind *absolutstetige Verteilungen*.

Definition 3.2.8. Es sei X eine Zufallsvariable mit Verteilungsfunktion F_X . Wir sagen, X hat *absolutstetige Verteilung* (gelegentlich auch X sei *absolutstetig*), falls es eine nichtnegative Funktion f auf \mathbb{R} mit

$$\int_{-\infty}^{\infty} f(t) dt = 1$$

gibt (hierbei soll das Integral wohldefiniert sein), sodass

$$F_X(x) = \int_{-\infty}^x f(t) dt.$$

In diesem Fall heißt f *Dichte* der Verteilung von X .

Bemerkung 3.2.9. Ist die Verteilungsfunktion stetig (insbesondere, bei absolutstetigen Verteilungsfunktionen), so nimmt die Zufallsvariable keinen Wert mit positiver Wahrscheinlichkeit an. In diesem Fall gilt also

$$\mathbb{P}(X \in (a, b]) = \mathbb{P}(X \in (a, b)) = \mathbb{P}(X \in [a, b)) = \mathbb{P}(X \in [a, b]) = F(b) - F(a) = \int_a^b f(t) dt.$$

Bemerkung 3.2.10. Ist f eine nichtnegative Funktion auf \mathbb{R} mit $\int_{-\infty}^{\infty} f(t) dt = 1$, so kann man zeigen, dass $F(x) = \int_{-\infty}^x f(t) dt$ eine monoton wachsende, rechtsseitig stetige Funktion mit $\lim_{x \rightarrow -\infty} F(x) = 0$ und $\lim_{x \rightarrow \infty} F(x) = 1$ ist. Somit ist F eine Verteilungsfunktion und f die Dichte dieser Verteilungsfunktion.

In gewisser Weise sind die Dichten von absolutstetigen Zufallsvariablen ähnlich zu den Zähldichten von diskreten Zufallsvariablen in Beispiel 3.2.5. Ist f_X die Zähldichte der diskreten Zufallsvariablen X , so ist die Verteilungsfunktion F_X gegeben durch

$$F_X(x) = \sum_{t \leq x} f_X(t),$$

wobei zu beachten ist, dass f_X ja nur an höchstens abzählbar vielen Stellen verschieden von 0 ist. Bei absolutstetigen Zufallsvariablen hat man stattdessen eine Dichte f , die man bis x "aufintegriert" um die Verteilungsfunktion zu erhalten.

Diese Analogie verwenden wir auch bei der Definition von Erwartungswert, Varianz, etc. von absolutstetigen Zufallsvariablen.

Definition 3.2.11. Es sei X eine absolutstetige Zufallsvariable und f die Dichte der Verteilung von X . Wir sagen, X hat *endlichen Erwartungswert*, falls

$$\int_{-\infty}^{\infty} |t|f(t) dt < \infty.$$

In diesem Fall heißt

$$\mathbb{E}X := \int_{-\infty}^{\infty} tf(t) dt$$

der Erwartungswert von X .

Wie im diskreten Fall gilt für eine Zufallsvariable X mit Dichte f und eine Funktion $g: \mathbb{R} \rightarrow \mathbb{R}$, dass

$$\mathbb{E}g(X) = \int_{-\infty}^{\infty} f(t)g(t) dt,$$

falls $g(X)$ endlichen Erwartungswert hat.

Wir sagen, X habe endliche Varianz, falls

$$\text{Var}X := \mathbb{E}[(X - \mathbb{E}X)^2] = \int_{-\infty}^{\infty} (t - \mathbb{E}X)^2 f(t) dt$$

endlich ist.

Bemerkung 3.2.12. Ähnlich wie im diskreten Fall kann man zeigen, dass X genau dann endliche Varianz hat, wenn $\mathbb{E}X^2 = \int_{-\infty}^{\infty} t^2 f(t) dt$ endlich ist. In diesem Fall ist

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

Beispiel 3.2.13. Setzen wir

$$f(t) = \begin{cases} \frac{1}{t^2} & \text{falls } t \geq 1 \\ 0 & \text{falls } t < 1 \end{cases} = \frac{1}{t^2} \mathbb{1}_{[1, \infty)},$$

so ist f Dichte einer Verteilungsfunktion. Es ist nämlich

$$\int_{-\infty}^{\infty} f(t) dt = \int_1^{\infty} \frac{1}{t^2} dt = \left[\frac{-1}{t} \right]_1^{\infty} = 0 - (-1) = 1.$$

Ist f die Dichte der Zufallsvariablen X , so ist

$$\mathbb{P}(1 \leq X \leq 2) = \int_1^2 \frac{1}{t^2} dt = \left[\frac{-1}{t} \right]_1^2 = -\frac{1}{2} + 1 = \frac{1}{2},$$

also nimmt X mit Wahrscheinlichkeit $1/2$ einen Wert zwischen 1 und 2 an.

X hat *keinen* endlichen Erwartungswert, es ist nämlich

$$\int_{-\infty}^{\infty} |t|f(t) dt = \int_1^{\infty} \frac{1}{t} dt = \left[\log t \right]_1^{\infty} = \infty.$$

Gelegentlich ist es wichtig (beispielsweise beim Berechnen von Konfidenzintervallen), Zahlen x zu finden, sodass $\mathbb{P}(X \leq x) = F_X(x) = p$ für ein vorgegebenes $p \in (0, 1)$. Wir behandeln diese Frage nur in dem Spezialfall, dass es ein Intervall I gibt, so dass die Verteilungsfunktion F_X aufgefasst als Funktion $I \rightarrow (0, 1)$ bijektiv ist. Dies ist z.B. dann erfüllt, wenn die Verteilungsfunktion auf \mathbb{R} stetig und streng monoton wachsend ist; dann ist $I = \mathbb{R}$. Dann hat $x := (F_X)^{-1}(p)$ die gewünschte Eigenschaft und heißt *p-Quantil*.

3.3 Wichtige absolutstetige Verteilungen

Stetige Gleichverteilung

Wir sagen X ist *gleichverteilt auf dem Intervall* (a, b) und schreiben $X \sim U(a, b)$, falls X die Dichte

$$f(t) = \frac{1}{b-a} \mathbb{1}_{(a,b)}(t) = \begin{cases} \frac{1}{b-a}, & a < t < b \\ 0, & \text{sonst} \end{cases}$$

besitzt.

Ein Beispiel, bei dem diese Verteilung auftritt ist beim Drehen eines ‘‘Glücksrades’’. Der Winkel, in dem das Rad im Vergleich zur Ausgangslage zum stehen kommt, ist gleichverteilt in $(0, 2\pi)$.

Wir berechnen Erwartungswert und Varianz einer gleichverteilten Zufallsvariable.

Lemma 3.3.1. *Ist $X \sim U(a, b)$, so ist*

$$\mathbb{E}X = \frac{a+b}{2} \quad \text{und} \quad \text{Var}X = \frac{(b-a)^2}{12}.$$

Beweis. Es ist

$$\mathbb{E}X = \frac{1}{b-a} \int_a^b t \, dt = \frac{1}{b-a} \left[\frac{t^2}{2} \right]_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}.$$

Weiter ist für $\mu = (a+b)/2$ gerade

$$\begin{aligned} \text{Var}X &= \frac{1}{b-a} \int_a^b (t-\mu)^2 \, dt = \frac{1}{b-a} \left[\frac{(t-\mu)^3}{3} \right]_a^b \\ &= \frac{1}{3(b-a)} \frac{1}{8} ((b-a)^3 - (a-b)^3) = \frac{(b-a)^2}{12}. \end{aligned}$$

□

Exponentialverteilung

Eine Zufallsvariable X heißt exponentialverteilt mit Parameter $\lambda > 0$, wenn X die Dichte

$$f_\lambda(t) = \lambda e^{-\lambda t} \mathbb{1}_{(0, \infty)}(t)$$

besitzt. Wir schreiben in diesem Fall $X \sim \text{Exp}(\lambda)$. Beachte, dass

$$\int_{-\infty}^{\infty} f_\lambda(t) \, dt = \int_0^{\infty} \lambda e^{-\lambda t} \, dt = \left[-e^{-\lambda t} \right]_0^{\infty} = 0 - (-1) = 1,$$

sodass f_λ in der Tat eine Dichte ist.

Die Exponentialverteilung tritt bei sogenannten Wartezeitproblemen auf. Wichtige Beispiele sind: Die Lebensdauer von Glühbirnen oder die Wartezeit auf den nächsten Anruf.

Wir berechnen wiederum Erwartungswert und Varianz.

Lemma 3.3.2. *Es sei $X \sim \text{Exp}(\lambda)$. Dann ist*

$$\mathbb{E}X = \frac{1}{\lambda} \quad \text{und} \quad \text{Var}X = \frac{1}{\lambda^2}.$$

Beweis. Mit partieller Integration erhalten wir

$$\mathbb{E}X = \int_0^{\infty} t \lambda e^{-\lambda t} \, dt = \left[-te^{-\lambda t} \right]_0^{\infty} - \int_0^{\infty} -e^{-\lambda t} \, dt = 0 - \left[\frac{e^{-\lambda t}}{\lambda} \right]_0^{\infty} = \frac{1}{\lambda}.$$

Mit zweifacher partieller Integration folgt

$$\mathbb{E}X^2 = \frac{2}{\lambda^2},$$

wobei die Details Übung sind. Daher ist

$$\text{Var}X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

□

Die Exponentialverteilung hat eine wichtige Eigenschaft, die man *Gedächtnislosigkeit* nennt. Ist nämlich $X \sim \text{Exp}(\lambda)$, so ist

$$\mathbb{P}(X \geq a + b | X \geq a) = \mathbb{P}(X \geq b). \quad (3.2)$$

Bei Wartezeitproblemen interpretiert man diese Gleichheit wie folgt: Die Wahrscheinlichkeit, noch mindestens die Zeitspanne b warten zu müssen, wenn man bereits a gewartet hat, ist genau so groß, wie von Anfang an mindestens b warten zu müssen. Um Gleichung (3.2) zu zeigen, rufen wir uns die Definition der bedingten Wahrscheinlichkeit ins Gedächtnis: $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$. Hier haben wir $A = \{X \geq a + b\}$ und $B = \{X \geq a\}$. Beachte, dass $A \subseteq B$ und daher $A \cap B = A$. Für $x \in (0, \infty)$ gilt

$$\mathbb{P}(X \geq x) = \int_x^\infty \lambda e^{-\lambda t} dt = \left[-e^{-\lambda t} \right]_x^\infty = e^{-\lambda x}.$$

Somit ergibt sich

$$\frac{\mathbb{P}(X \geq a + b, X \geq a)}{\mathbb{P}(X \geq a)} = \frac{e^{-\lambda(a+b)}}{e^{-\lambda a}} = e^{-\lambda b} = \mathbb{P}(X \geq b),$$

wie behauptet.

Normalverteilung

Es seien $\mu \in \mathbb{R}$ und $\sigma^2 > 0$. Eine Zufallsvariable X heißt *normalverteilt* mit Parametern μ und σ^2 , falls X die Dichte

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

besitzt. Wir schreiben $X \sim \mathcal{N}(\mu, \sigma^2)$. Ist $\mu = 0$ und $\sigma^2 = 1$, so sagen wir, X ist *standard-normalverteilt*.

Da der Nachweis von $\int_{-\infty}^{\infty} f(t) dt = 1$ sehr schwierig ist, übergehen wir ihn.

Die Bedeutung der Normalverteilung entsteht vor allem durch den Zentralen Grenzwertsatz (den wir später behandeln), demzufolge viele Zufallsvariablen zumindest "annähernd" normalverteilt sind.

Proposition 3.3.3. *Ist $X \sim \mathcal{N}(\mu, \sigma^2)$ und sind $a, b \in \mathbb{R}$, so ist $aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.*

Beweis. Es gilt

$$\begin{aligned} \mathbb{P}(aX + b \leq y) &= \mathbb{P}\left(X \leq \frac{y-b}{a}\right) \\ &= \int_{-\infty}^{(y-b)/a} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \\ &= \int_{-\infty}^y \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{((u-b)/a-\mu)^2}{2\sigma^2}} \frac{1}{a} du \\ &= \int_{-\infty}^y \frac{1}{\sqrt{2\pi a^2 \sigma^2}} e^{-\frac{(u-b-a\mu)^2}{2a^2\sigma^2}} du. \end{aligned}$$

Also hat $aX + b$ die Verteilungsfunktion einer $\mathcal{N}(a\mu + b, a^2\sigma^2)$ -Verteilung. \square

Insbesondere folgt, dass $\frac{X-\mu}{\sigma}$ standardnormalverteilt ist, wenn $X \sim \mathcal{N}(\mu, \sigma^2)$. Somit kann man normalverteilte Zufallsvariablen durch Transformation immer in standardnormalverteilte Zufallsvariablen überführen. Daher erhalten Dichte und Verteilungsfunktion der Standardnormalverteilung besondere Bezeichner. Es sei

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \quad \text{und} \quad \Phi(x) = \int_{-\infty}^x \varphi(t) dt.$$

Man kann Φ nicht elementar ausdrücken. In Anwendungen verwendet man daher oft Tabellen mit Werten von Φ . Häufig sind dort nur Werte $\Phi(x)$ für $x \geq 0$ aufgeführt. Aus der Symmetrie von φ , d.h. $\varphi(-t) = \varphi(t)$, folgt dass

$$\Phi(-x) = \int_{-\infty}^{-x} \varphi(t) dt = \int_x^{\infty} \varphi(-t) dt = \int_{-\infty}^{\infty} \varphi(t) dt - \int_{-\infty}^x \varphi(t) dt = 1 - \Phi(x),$$

sodass diese Information ausreicht.

Wir können nun Erwartungswert und Varianz einer normalverteilten Zufallsvariablen bestimmen.

Lemma 3.3.4. *Ist $X \sim \mathcal{N}(\mu, \sigma^2)$, so ist $\mathbb{E}X = \mu$ und $\text{Var}X = \sigma^2$. Man sagt daher auch, X sei normalverteilt mit Erwartungswert μ und Varianz σ^2 .*

Beweis. Es sei zunächst $X \sim \mathcal{N}(0, 1)$. Aus der Symmetrie von φ folgt mit der Substitution $t = -s$, dass

$$\mathbb{E}X = \int_{-\infty}^{\infty} t\varphi(t) dt = \int_{-\infty}^0 t\varphi(t) dt + \int_0^{\infty} t\varphi(t) dt = - \int_0^{\infty} s\varphi(s) ds + \int_0^{\infty} t\varphi(t) dt = 0.$$

Mit partieller Integration folgt nun

$$\text{Var}X = \mathbb{E}X^2 = \int_{-\infty}^{\infty} t \cdot \frac{t \cdot e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} dt = \left[\frac{-te^{-\frac{t^2}{2}}}{\sqrt{2\pi}} \right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \frac{-e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} dt = 1.$$

Sei nun $X \sim \mathcal{N}(\mu, \sigma^2)$. Dann ist $Y := \frac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$. Daher ist

$$0 = \mathbb{E}Y = \sigma^{-1} \mathbb{E}(X - \mu) = \sigma^{-1} ((\mathbb{E}X) - \mu)$$

und daher $\mathbb{E}X = \mu$. Weiter ist $\text{Var}X = \text{Var}(X - \mu) = \mathbb{E}((\sigma Y)^2) = \sigma^2 \mathbb{E}Y^2 = \sigma^2$. \square

3.4 Zufallsvektoren

Definition 3.4.1. Es seien X_1, \dots, X_n Zufallsvariablen, die auf einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \Sigma, \mathbb{P})$ definiert sind. Dann heißt die Abbildung $X : \Omega \rightarrow \mathbb{R}^n$, gegeben durch

$$X(\omega) = (X_1(\omega), \dots, X_n(\omega))$$

Zufallsvektor. Die Funktion $F : \mathbb{R}^n \rightarrow [0, 1]$ gegeben durch

$$F(x_1, \dots, x_n) := \mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n)$$

heißt *Verteilungsfunktion* des Zufallsvektors X . Die Verteilungsfunktion (oder auch der Zufallsvektor X selbst) heißt *absolutstetig*, falls es eine Funktion $f : \mathbb{R}^n \rightarrow [0, \infty)$ gibt mit

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(t_1, \dots, t_n) dt_1 \dots dt_n = 1$$

sodass

$$F(x_1, \dots, x_n) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f(t_1, \dots, t_n) dt_1 \dots dt_n.$$

In diesem Fall heißt f *Dichte* von F (oder von X).

Bemerkung 3.4.2. (a) Wir nennen die von den Rechtecken $(a_1, b_1) \times \dots \times (a_n, b_n)$ mit $a_i, b_i \in \mathbb{R}$, $a_i < b_i$, erzeugte σ -Algebra auf \mathbb{R}^n Borel- σ -Algebra $\mathcal{B}(\mathbb{R}^n)$. Wiederum enthält die Borel- σ -Algebra aller praktisch relevanten Teilmengen, weshalb wir in Übungs- und Klausuraufgaben darauf verzichten, bei Teilmengen des \mathbb{R}^n zu überprüfen, ob diese in der Borel- σ -Algebra liegen. Ähnlich wie im Fall von Zufallsvariablen kann man zeigen, dass die Verteilungsfunktion ein Wahrscheinlichkeitsmaß auf $\mathcal{B}(\mathbb{R}^n)$ bestimmt. Dieses Maß ist gerade die Verteilung des Vektors X .

(b) Aus der Verteilungsfunktion des Vektors können auch die Verteilungsfunktionen der einzelnen Komponenten bestimmt werden. Mit Lemma 3.1.11 folgt nämlich

$$\begin{aligned} F_{X_j}(x) &= \mathbb{P}(X_j \leq x) = \lim_{N \rightarrow \infty} \mathbb{P}(X_1 \leq N, \dots, X_{j-1} \leq N, X_j \leq x, X_{j+1} \leq N, \dots, X_n \leq N) \\ &= \lim_{N \rightarrow \infty} F_X(N, \dots, N, x, N, \dots, N). \end{aligned}$$

In dieser Situation nennt man manchmal den letzten Grenzwert *Randverteilungsfunktion*.

(c) Auf ähnliche Weise lassen sich auch die Dichten der einzelnen Komponenten aus der Dichte des Vektors berechnen. Es ist nämlich

$$f_j(s) = \int_{\mathbb{R}^{n-1}} f(t_1, \dots, t_{j-1}, s, t_{j+1}, \dots, t_n) dt_1 \dots dt_n,$$

wobei f_j die Dichte von X_j bezeichnet. Diese wird also aus der "gemeinsamen Dichte" durch ausintegrieren der anderen Variablen berechnet. Man sagt die f_j seien die *Randdichten*.

(d) Mittels der Dichte lassen sich auch Wahrscheinlichkeiten berechnen. Sei nämlich X ein Zufallsvektor mit Dichte f . Dann gilt für $A \in \mathcal{B}(\mathbb{R}^n)$, dass

$$\mathbb{P}(X \in A) = \int_A f(t_1, \dots, t_n) dt_1 \dots dt_n.$$

(e) Für eine Funktion $g : \mathbb{R}^n \rightarrow \mathbb{R}$ und einen Zufallsvektor X mit Dichte f ist

$$\mathbb{E}g(X_1, \dots, X_n) = \int_{\mathbb{R}^n} g(t_1, \dots, t_n) f(t_1, \dots, t_n) dt_1 \dots dt_n$$

sofern der Erwartungswert endlich ist. Insbesondere kann man auf diese Art die Kovarianz zweier Zufallsvariablen berechnen, die wie im diskreten Fall definiert ist.

Beispiel 3.4.3. Es sei (X, Y) ein Vektor mit Dichte $f(x, y) = (x + 2xy) \mathbb{1}_{(0,1)}(x) \mathbb{1}_{(0,1)}(y)$. Beachte, dass dies in der Tat eine Dichte ist, es ist nämlich $f(x, y) \geq 0$ für alle $x, y \in \mathbb{R}$ und

$$\begin{aligned} \int_{\mathbb{R}^2} f(x, y) dx dy &= \int_0^1 \int_0^1 (x + 2xy) dx dy = \int_0^1 \left[\frac{1}{2}x^2 + x^2y \right]_0^1 dy \\ &= \int_0^1 \frac{1}{2} + y dy = \left[\frac{1}{2}y + \frac{1}{2}y^2 \right]_0^1 = 1. \end{aligned}$$

Die Dichten der Zufallsvariablen X und Y erhält man wie folgt:

$$f_X(t) = \int_{\mathbb{R}} f(t, y) dy = \int_0^1 t + 2ty dy \mathbb{1}_{(0,1)}(t) = 2t \mathbb{1}_{(0,1)}(t)$$

und

$$f_Y(t) = \int_{\mathbb{R}} f(x, t) dx = \int_0^t x + 2xt dx \mathbb{1}_{(0,1)}(t) = \left(\frac{1}{2} + t\right) \mathbb{1}_{(0,1)}(t)$$

Mittels der gemeinsamen Dichte kann man die Wahrscheinlichkeit berechnen, dass $X \leq Y$ ist. Ist nämlich $A = \{(x, y) \in \mathbb{R}^2 : x \leq y\}$, so ist

$$\begin{aligned} \mathbb{P}(X \leq Y) &= \mathbb{P}((X, Y) \in A) = \int_A f(x, y) dx dy = \int_0^1 \int_x^1 x + 2xy dy dx \\ &= \int_0^1 \left[xy - xy^2 \Big|_{y=x}^{y=1} \right] dx = \int_0^1 2x - x^2 - x^3 dx \\ &= 1 - \frac{1}{3} - \frac{1}{4} = \frac{5}{12}. \end{aligned}$$

Schließlich berechnen wir noch die Kovarianz von X und Y . Hierzu benötigen wir zunächst die Erwartungswerte. Es ist

$$\mathbb{E}X = \int_{\mathbb{R}} t f_X(t) dt = \int_0^1 t \cdot 2t dt = \left[\frac{2}{3} t^3 \Big|_0^1 \right] = \frac{2}{3}$$

und

$$\mathbb{E}Y = \int_{\mathbb{R}} t f_Y(t) dt = \int_0^1 t \left(\frac{1}{2} + t\right) dt = \left[\frac{t^2}{4} + \frac{t^3}{3} \Big|_0^1 \right] = \frac{7}{12}.$$

Somit ist

$$\begin{aligned} \text{Cov}(X, Y) &= \mathbb{E}\left(X - \frac{2}{3}\right)\left(Y - \frac{7}{12}\right) \\ &= \int_0^1 \int_0^1 \left(x - \frac{2}{3}\right)\left(y - \frac{7}{12}\right)(x + 2xy) dx dy \\ &= \int_0^1 \left(y - \frac{7}{12}\right) \int_0^1 x^2 + 2x^2y - \frac{2}{3}x - \frac{4}{3}xy dx dy \\ &= \int_0^1 \left(y - \frac{7}{12}\right) \left[\frac{x^3}{3} + \frac{2}{3}x^3y - \frac{x^2}{3} - \frac{2}{3}x^2y \Big|_0^1 \right] dy \\ &= 0. \end{aligned}$$

Folglich sind X und Y unkorreliert.

Definition 3.4.4. Zufallsvariablen X_1, \dots, X_n , die auf einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \Sigma, \mathbb{P})$ definiert sind, heißen *unabhängig*, falls für alle $A_1, \dots, A_n \in \mathcal{B}(\mathbb{R})$ stets

$$\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \mathbb{P}(X_1 \in A_1) \cdot \dots \cdot \mathbb{P}(X_n \in A_n)$$

gilt.

Man kann nun folgenden Satz beweisen:

Satz 3.4.5. Seien X_1, \dots, X_n Zufallsvariablen auf einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \Sigma, \mathbb{P})$. Weiter sei F_X die Verteilungsfunktion des Vektors (X_1, \dots, X_n) und F_{X_j} die Verteilungsfunktion der Zufallsvariable X_j für $j = 1, \dots, n$. Dann sind folgende Aussagen äquivalent:

(i) X_1, \dots, X_n sind unabhängig.

(ii) $F_X(x_1, \dots, x_n) = F_{X_1}(x_1) \cdot \dots \cdot F_{X_n}(x_n)$ für alle $x_1, \dots, x_n \in \mathbb{R}$.

Besitzt X eine Dichte f und X_j die Dichte f_j , so sind obige Aussagen äquivalent zu

(iii) $f(t_1, \dots, t_n) = f_1(t_1) \cdot \dots \cdot f_n(t_n)$, wobei f die Dichte von X und f_j die Dichte von X_j bezeichnet.

Beispiel 3.4.6. Die Zufallsvariablen X und Y aus Beispiel 3.4.3 sind unabhängig. Es ist nämlich

$$f_X(x)f_Y(y) = 2x\mathbb{1}_{(0,1)}(x)\left(\frac{1}{2} + y\right)\mathbb{1}_{(0,1)}(y) = (x + 2xy)\mathbb{1}_{(0,1)}(x)\mathbb{1}_{(0,1)}(y) = f_{(X,Y)}(x, y).$$

Wir betrachten ein weiteres Beispiel.

Beispiel 3.4.7. Es sei $D = \{(x, y) : 0 \leq y \leq 2x, 0 \leq x \leq 1\}$ und $f = \mathbb{1}_D$. Dann ist f eine Dichte. Es ist nämlich

$$\int_{\mathbb{R}^2} \mathbb{1}_D dx dy = \int_0^1 \int_0^{2x} dy dx = \int_0^1 2x dx = 1.$$

Ist (X, Y) ein Zufallsvektor mit Dichte f , so sagt man (X, Y) sei *gleichverteilt auf dem Dreieck D* .

Die Randdichten von f sind wie folgt gegeben:

$$f_X(t) = \int_{\mathbb{R}} \mathbb{1}_D(t, y) dy = \mathbb{1}_{(0,1)}(t) \int_0^{2t} dy = 2t\mathbb{1}_{(0,1)}(t)$$

und

$$f_Y(t) = \int_{\mathbb{R}} \mathbb{1}_D(x, t) dx = \mathbb{1}_{(0,2)}(t) \int_{t/2}^1 dx = \left(1 - \frac{t}{2}\right)\mathbb{1}_{(0,1)}(t).$$

Hat der Vektor (X, Y) Dichte f , so sind X und Y nicht unabhängig, es ist nämlich

$$\begin{aligned} f_X(x)f_Y(y) &= 2x\mathbb{1}_{(0,1)}(x)\left(1 - \frac{y}{2}\right)\mathbb{1}_{(0,2)}(y) \\ &= (2x - xy)\mathbb{1}_{(0,1)}(x)\mathbb{1}_{(0,2)}(y) \\ &\neq \mathbb{1}_D(x, y) = f_{X,Y}(x, y). \end{aligned}$$

3.5 Der Zentrale Grenzwertsatz

Nun kommen wir zu einem zentralen Resultat, welches die Bedeutung der Normalverteilung erklärt.

Satz 3.5.1. (Zentraler Grenzwertsatz) Es sei X_1, X_2, \dots eine Folge von unabhängigen und identisch verteilten Zufallsvariablen mit endlichen Momenten zweiter Ordnung und positiver Varianz. Wir setzen $\sigma^2 := \text{Var}X_1$ und $\mu := \mathbb{E}X_1$. Weiter sei $S_n := \sum_{k=1}^n X_k$. Dann ist

$$\mathbb{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right) \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt = \Phi(x), \quad n \rightarrow \infty.$$

Wir bemerken: Es ist $\mathbb{E}S_n = n\mu$ und, wegen der Unabhängigkeit, $\text{Var}(S_n) = n\sigma^2$. Daher sind

$$\mathbb{E}\left[\frac{S_n - n\mu}{\sigma\sqrt{n}}\right] = 0 \quad \text{und} \quad \text{Var}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}}\right) = 1$$

unabhängig von n .

Alternative Interpretation: Ist $M_n := \frac{1}{n}S_n$ das Mittel der ersten n Zufallsvariablen, so ist $\frac{\sqrt{n}}{\sigma}(M_n - \mu) = \frac{S_n - n\mu}{\sigma\sqrt{n}}$.

Der Zentrale Grenzwertsatz besagt gerade, dass die Verteilungsfunktion dieser standardisierten Mittel gegen die Verteilungsfunktion der Standardnormalverteilung konvergiert.

Bemerkung 3.5.2. Beachte, dass $\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a)$ ist. Es folgt aus dem Zentralen Grenzwertsatz, dass

$$\mathbb{P}\left(a < \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq b\right) \rightarrow \int_a^b e^{-\frac{t^2}{2}} \frac{dt}{\sqrt{2\pi}}, \quad n \rightarrow \infty.$$

Man darf hier $<$ - und \leq - Zeichen austauschen.

Dank des zentralen Grenzwertsatzes können wir (unter Kenntnis der Werte von Φ , die tabelliert sind) Wahrscheinlichkeiten approximieren. Wir geben hierzu ein Beispiel:

Beispiel 3.5.3. Ein Würfel werde 600 mal geworfen. Was ist die Wahrscheinlichkeit zwischen 90 und 100 Sechsen zu werfen?

Hierbei handelt es sich um das 600-fache (unabhängige) Wiederholen eines Bernoulli-Experiments mit Erfolgswahrscheinlichkeit $p = \frac{1}{6}$. Damit ist die Zahl der Erfolge $\text{Bin}(600, \frac{1}{6})$ -verteilt und die gesuchte Wahrscheinlichkeit ist

$$\sum_{k=90}^{100} \binom{600}{k} \frac{1}{6}^k \frac{5}{6}^{600-k}.$$

Allerdings ist diese Summe relativ schwierig zu berechnen. Wir verwenden den zentralen Grenzwertsatz, um die Wahrscheinlichkeit zu approximieren.

Seien hierzu X_1, \dots, X_{600} unabhängige, Bernoulli verteilte Zufallsvariablen mit Erfolgswahrscheinlichkeit $\frac{1}{6}$. Es ist also $\mu := \mathbb{E}X_1 = p = \frac{1}{6}$ und $\sigma^2 = p(1-p) = \frac{5}{36}$. Es ist also $n\mu = 100$ und $\sqrt{n}\sigma = \sqrt{500/6} \approx 9,13$.

Die Anzahl der Erfolge bei 600 Versuchen ist $S_{600} = \sum_{k=1}^{600} X_k$. Nach dem zentralen Grenzwertsatz gilt

$$\begin{aligned} \mathbb{P}(90 \leq S_{600} \leq 100) &= \mathbb{P}\left(\frac{90 - 100}{9,13} \leq \frac{S_{600} - 600\mu}{\sqrt{600}\sigma} \leq \frac{100 - 100}{9,13}\right) \\ &\approx \Phi(0) - \Phi(-1,095) = 0,5 - (1 - \Phi(1,095)) \approx 0,363, \end{aligned}$$

also ungefähr 36 Prozent.

Bemerkung 3.5.4. Bei der Approximation der Binomialverteilung durch die Normalverteilung ist folgendes zu beachten:

Die Binomialverteilung ist eine diskrete Verteilung, genauer nimmt sie nur natürliche Zahlen als Werte an. In Beispiel 3.5.3 ist somit

$$\mathbb{P}(90 \leq S_{600} \leq 100) = \mathbb{P}(89 < S_{600} < 101) = \mathbb{P}(89,5 \leq S_{600} \leq 100,5).$$

Berechnet man diese Wahrscheinlichkeiten näherungsweise mit dem zentralen Grenzwertsatz, so ergeben sich 0,363, 0,430 bzw. 0,397. Vergleichen wir diesen Wert mit dem exakten Wert 0,4025 (die Berechnung ist zwar aufwendig, aber nicht unmöglich), so sehen wir, dass der letztgenannte Wert am besten ist. Es entspricht auch der allgemeinen Erfahrung, dass, wenn die Wahrscheinlichkeit, dass eine diskrete Zufallsvariable in einem bestimmten Intervall liegt, mit Hilfe des zentralen Grenzwertsatzes approximiert wird, man am besten die Mittelwerte zwischen den beiden in Frage kommenden Zahlen als Intervallgrenzen einsetzt.

Wir geben nun noch ein etwas anderes Anwendungsbeispiel.

Beispiel 3.5.5. Ein Airbus A380 hat gewöhnlich 526 Sitzplätze. Aus Erfahrungen ist bekannt, dass ein verkaufte Ticket mit einer Wahrscheinlichkeit von 0,1 storniert wird. Wir nehmen an, dass die Fluggäste, die Entscheidung, ob sie stornieren oder nicht, unabhängig voneinander treffen. Wie viele Tickets kann man für einen Flug verkaufen, wenn dieser mit einer Wahrscheinlichkeit von höchstens 2% überbucht sein soll?

Es sei X_1, X_2, \dots eine Folge unabhängiger $\text{Bin}(1, 0,9)$ -verteilter Zufallsvariablen und $S_n = \sum_{k=1}^n X_k$. Wir suchen die größte Zahl n , sodass $\mathbb{P}(S_n \geq 526) \leq 0,02$. Es ist

$$0,02 \stackrel{!}{=} \mathbb{P}(S_n \geq 526) = \mathbb{P}\left(\frac{S_n - n \cdot 0,9}{\sqrt{n \cdot 0,1 \cdot 0,9}} \geq \frac{526 - n \cdot 0,9}{\sqrt{n \cdot 0,09}}\right) \approx 1 - \Phi\left(\frac{526 - n \cdot 0,9}{\sqrt{n \cdot 0,09}}\right)$$

Wir benötigen das 0,98-Quantil der Normalverteilung. kann man in Tabellen nachschlagen. Es gilt $\Phi^{-1}(0,98) \approx 2,05$. Wir erhalten also

$$2,05 \approx \Phi^{-1}(0,98) \approx \frac{526 - n \cdot 0,9}{\sqrt{n \cdot 0,09}}.$$

Wir schreiben $m = \sqrt{n}$ und erhalten die Gleichung

$$\frac{526 - m^2 \cdot 0,9}{m \sqrt{0,09}} = 2,05 \Leftrightarrow 526 - m^2 \cdot 0,9 = m \cdot 0,3 \cdot 2,05$$

Löst man diese quadratische Gleichung, so erhält man $m = 23,8$, also $n = 568,1$. Die Fluggesellschaft darf also höchstens 568 Tickets verkaufen.

3.6 Schätzung der Parameter in der Normalverteilung

In Kapitel 2 hatten wir bereits die Schätzung von Parametern gewisser Verteilungen diskutiert, dabei aber (weil wir absolutstetige Verteilungen noch nicht eingeführt hatten) die Normalverteilung außen vor gelassen. Allerdings ist diese Verteilung in Anwendungen von größtem Interesse. Daher wollen wir das Schätzen der Parameter der Normalverteilung hier nachholen.

Es gilt also aus einer Zufallsstichprobe X_1, \dots, X_n zur Normalverteilung $\mathcal{N}(\mu, \sigma^2)$ die Parameter μ und σ^2 zu schätzen.

Wir nutzen zunächst die *Momentenmethode*. Für $Y \sim \mathcal{N}(\mu, \sigma^2)$ gilt $(\mu, \sigma^2) = (\mathbb{E}Y, \mathbb{E}Y^2 - (\mathbb{E}Y)^2)$. Somit ergibt sich als Momentenschätzer, wenn Realisierungen x_1, \dots, x_n beobachtet werden,

$$(\hat{\mu}, \widehat{\sigma^2}) = \left(\bar{x}, \frac{1}{n} \sum_{k=1}^n (x_k)^2 - (\bar{x})^2\right) = \left(\bar{x}, \frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2\right).$$

Nu wenden wir uns der *Maximum-Likelihood-Methode* zu. Im Falle diskreter Wahrscheinlichkeiten hatten wir als Likelihood-Funktion L gerade das Produkt der Zähldichten verwendet: $L(x_1, \dots, x_n; \theta) = f(x_1, \theta) \cdot \dots \cdot f(x_n, \theta)$. Im Falle von absolutstetigen Verteilungen mit Dichte $f(t, \theta)$ verwenden wir das Produkt der Dichten als Likelihood-Funktion.

Definition 3.6.1. Es sei X_1, \dots, X_n eine Zufallsstichprobe zur Verteilungsfunktion $P \in \{P_\theta : \theta \in \Theta\}$. Weiter sei P_θ absolutstetig mit Dichte $f(\cdot, \theta)$. Dann heißt $L : \mathbb{R}^n \times \Theta \rightarrow [0, \infty)$, definiert durch

$$L(x_1, \dots, x_n; \theta) := f(x_1, \theta) \cdot \dots \cdot f(x_n, \theta)$$

Likelihood-Funktion. Ist $\hat{\theta} : \mathbb{R}^n \rightarrow \Theta$ eine Funktion mit

$$L(x_1, \dots, x_n; a) \leq L(x_1, \dots, x_n; \hat{\theta}(x_1, \dots, x_n))$$

für alle $a \in \Theta$, so heißt $\hat{\theta}(X_1, \dots, X_n)$ *Maximum-Likelihood-Schätzer* für θ .

Wir berechnen nun Maximum-Likelihood-Schätzer für die Parameter μ und σ^2 einer Normalverteilung. Es ist

$$L(x_1, \dots, x_n; \mu, \sigma^2) = \prod_{k=1}^n \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x_k - \mu)^2}{2\sigma^2}} = \left(\frac{1}{\sqrt{2\pi\sigma}}\right)^n \exp\left(-\frac{\sum_{k=1}^n (x_k - \mu)^2}{2\sigma^2}\right).$$

Es ist einfacher, die log-Likelihood Funktion

$$\ell(\mu, \sigma^2) := \log L(x_1, \dots, x_n; \mu, \sigma^2) = -\frac{1}{2\sigma^2} \sum_{k=1}^n (x_k - \mu)^2 - \frac{n}{2} \log(2\pi\sigma^2)$$

zu betrachten. Um einen Kandidaten für die Maximumstelle zu finden, berechnen wir die kritischen Punkte der Log-Likelihood-Funktion, also die Lösungen des Gleichungssystems $\nabla \ell = 0$. Für die partielle Ableitung nach μ finden wir

$$\frac{\partial \ell}{\partial \mu} = \left(-\frac{1}{2\sigma^2}\right) \sum_{k=1}^n 2(x_k - \mu) \stackrel{!}{=} 0 \quad \Leftrightarrow \quad 0 = \sum_{k=1}^n (x_k - \mu) = n\bar{x} - n\mu \quad \Leftrightarrow \quad \mu = \bar{x}$$

Für die partielle Ableitung nach σ^2 ist

$$\frac{\partial \ell}{\partial \sigma^2} = \frac{1}{2\sigma^4} \sum_{k=1}^n (x_k - \mu)^2 - \frac{n}{2} \frac{2\pi}{2\pi\sigma^2} = \frac{1}{2\sigma^2} \left(\frac{1}{\sigma^2} \sum_{k=1}^n (x_k - \mu)^2 - n\right) \stackrel{!}{=} 0$$

Wir wissen bereits, dass in einem kritischen Punkt $\mu = \bar{x}$ sein muss. Wir setzen das in obige Gleichung ein und erhalten

$$\frac{\partial \ell}{\partial \sigma^2} = 0 \quad \Leftrightarrow \quad \frac{1}{\sigma^2} \sum_{k=1}^n (x_k - \bar{x})^2 = n \quad \Leftrightarrow \quad \sigma^2 = \frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2.$$

Die einzige kritische Stelle von ℓ ist demnach der Punkt

$$(\mu, \sigma^2) = \left(\bar{x}, \frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2\right).$$

Eine genauere Untersuchung der log-Likelihood Funktion zeigt, dass dies in der Tat eine (globale!) Maximumstelle ist, d.h. dies ist der (in diesem Falle eindeutige) Maximum-Likelihood-Schätzer. Für die Normalverteilung stimmen also Momenten- und Maximum-Likelihood-Schätzer überein.

Obwohl sich sowohl aus der Momenten- als auch aus der Maximum-Likelihood-Methode $\frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2$ als Schätzer für σ^2 ergibt, zieht man in der Praxis meist $\frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^2$ vor, weil dieser Schätzer erwartungstreu ist.

Konfidenzintervall für μ bei bekannter Varianz

Als nächstes konstruieren wir Konfidenzintervalle für den Mittelwert μ . Dabei muss unterschieden werden, ob die Varianz σ^2 bekannt ist, oder nicht

Es sei also X_1, \dots, X_n eine Zufallsstichprobe zur Verteilung $\mathcal{N}(\mu, \sigma_0^2)$, wobei σ_0^2 eine bekannte, feste, positive Zahl ist und μ ein unbekannter, reeller Parameter ist. Um μ zu schätzen verwenden wir den Schätzer $\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k$. Wir könnten nun mittels Tschebyscheff-Ungleichung ein Konfidenzintervall herleiten. Allerdings ergibt sich ein kürzeres Konfidenzintervall, wenn wir folgende Proposition verwenden.

Proposition 3.6.2. *Seien $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, $i = 1, \dots, n$, unabhängige Zufallsvariablen. Dann gilt $X_1 + \dots + X_n \sim \mathcal{N}(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2)$.*

Insbesondere ist $X_1 + \dots + X_n \sim \mathcal{N}(n\mu, n\sigma_0^2)$ und daher $\bar{X} \sim \mathcal{N}(\mu, \sigma_0^2/n)$. Es folgt, dass $V := \sqrt{n} \frac{\bar{X} - \mu}{\sigma_0} \sim \mathcal{N}(0, 1)$. Beachte, dass diese Verteilung *nicht* mehr von dem unbekanntem Parameter μ abhängt. Um ein Konfidenzintervall zum Konfidenzniveau α zu konstruieren, geht man nun wie folgt vor:

Zunächst bestimmt man ein Intervall, in dem V mit Wahrscheinlichkeit α liegt. Weil V symmetrisch verteilt ist, wählen wir auch ein symmetrisches Intervall, also ein $c > 0$ sodass

$$1 - \alpha = \mathbb{P}_\mu(|V| > c) = 2\mathbb{P}_\mu(V > c) = 2(1 - \Phi(c))$$

also $c = \Phi^{-1}(\frac{1+\alpha}{2})$. Somit ergibt sich

$$\alpha = \mathbb{P}_\mu(|V| \leq c) = \mathbb{P}_\mu\left(-c \leq \sqrt{n} \frac{\bar{X} - \mu}{\sigma_0} \leq c\right) \quad (3.3)$$

$$= \mathbb{P}_\mu\left(\mu \in \left[\bar{X} - \frac{c\sigma_0}{\sqrt{n}}, \bar{X} + \frac{c\sigma_0}{\sqrt{n}}\right]\right). \quad (3.4)$$

Somit ist $[\bar{X} - c\sigma_0/\sqrt{n}, \bar{X} + c\sigma_0/\sqrt{n}]$ mit $c = \Phi^{-1}(1 + \alpha/2)$ ein α -Konfidenzintervall für μ .

An diesem Beispiel kann man sehen, wie man allgemein vorgeht, um (unter Ausnutzung spezieller Eigenschaften der zu Grunde liegenden Verteilung) Konfidenzintervalle für einen Parameter $\vartheta \in \Theta$ zu konstruieren:

- 1) Wähle einen (Punkt-) Schätzer $\hat{\theta}$ für θ .
- 2) Bestimme die Verteilung von $\hat{\theta}$.
- 3) Wähle eine Transformation g_θ , so dass die Verteilung von $g_\theta(\hat{\theta})$ nicht von θ abhängt.
- 4) Wähle nun ein Intervall J mit $\mathbb{P}(g_\theta(\hat{\theta}) \in J) = \alpha$.
- 5) Löse die Bedingung $g_\theta(\hat{\theta}) \in J$ nach θ auf. Falls sich ein Intervall ergibt, ist dies das gesuchte Konfidenzintervall.

Insbesondere sind Konfidenzintervalle nicht eindeutig bestimmt.

Beispiel 3.6.3. Bei einer Stichprobe von 5 Brötchen wurden folgende Gewichte (in Gramm) gemessen (nach aufsteigendem Gewicht):

43 44 44 47 50

Wir nehmen an, dass das Gewicht normalverteilt ist mit unbekanntem Mittelwert und Standardabweichung 3. Wir bestimmen ein 95%-Konfidenzintervall für μ wie folgt:

Zunächst schlagen wir $c = \Phi^{-1}(\frac{1,95}{2}) = \Phi^{-1}(0,975) = 1,96$ in einer Tabelle der Verteilung der Standardnormalverteilung nach. Es ist nun

$$\frac{c\sigma_0}{\sqrt{n}} = \frac{1,96 \cdot 3}{\sqrt{5}} = 2,63 \quad \text{und} \quad \bar{X} = 45,6.$$

Daher ist $[42,97, 48,23]$ ein 95%-Konfidenzintervall für den Mittelwert μ .

Mit der Tschebyscheff-Ungleichung hätte sich

$$\left[\bar{X} - \sqrt{\frac{\sigma^2}{n \cdot (1-\alpha)}}, \bar{X} + \sqrt{\frac{\sigma^2}{n \cdot (1-\alpha)}} \right] = [39,6, 51,6]$$

ergeben.

Konfidenzintervall für μ bei unbekannter Varianz

Im allgemeinen kann man in Anwendungen *nicht* annehmen, dass die Varianz bekannt ist. Sie muss also ebenfalls geschätzt werden. Es ist naheliegend, in obigem Zugang die (nun nicht mehr bekannte) Standardabweichung σ_0 durch S , die Wurzel aus dem Schätzer S^2 für die Varianz, zu ersetzen. Um ein Konfidenzintervall zu bestimmen, muss man nun die Verteilung der resultierenden Zufallsvariable kennen. Wir diskutieren dies in allgemeiner Situation.

Definition 3.6.4. 1. Es seien X_1, \dots, X_r unabhängig und $\mathcal{N}(0,1)$ -verteilt. Dann heißt die Verteilung von $Y := X_1^2 + \dots + X_r^2$ χ^2 -Verteilung mit r Freiheitsgraden (sprich: chi-quadrat). Wir schreiben $Y \sim \chi_r^2$.

2. Es seien X, Y unabhängig mit $X \sim \mathcal{N}(0,1)$ und $Y \sim \chi_r^2$. Dann heißt die Verteilung von $Z := \frac{X}{\sqrt{Y/r}}$ t -Verteilung mit r Freiheitsgraden. Wir schreiben $Z \sim t_r$.

Bemerkung 3.6.5. (a) Sowohl die χ^2 -Verteilungen als auch die t -Verteilungen sind absolutstetig. Die Dichten dieser Verteilungen können explizit angegeben werden. Für uns sind jedoch insbesondere die Verteilungsfunktion und deren Umkehrfunktion (d.h. die Quantilfunktion) von Bedeutung. Diese liegen in Tabellen vor.

(b) Die t -Verteilung ist symmetrisch.

Zur Bestimmung von Konfidenzintervallen wichtig ist folgendes Resultat, welches wir hier ohne Beweis angeben.

Satz 3.6.6. *Es seien X_1, \dots, X_n unabhängig und $\mathcal{N}(\mu, \sigma^2)$ -verteilt. Dann sind \bar{X} und S^2 unabhängig und es gilt*

$$\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right), \quad \frac{n-1}{\sigma^2} S^2 \sim \chi_{n-1}^2, \quad \sqrt{n} \frac{\bar{X} - \mu}{S} \sim t_{n-1}.$$

Wir können also statt der Größe $V = \sqrt{n} \frac{\bar{X} - \mu}{\sigma}$, die von der Standardabweichung σ abhängt, die Größe $T = \sqrt{n} \frac{\bar{X} - \mu}{S}$, die nicht von der Standardabweichung abhängt, betrachten und haben wiederum eine Größe, deren Verteilung nicht von den Parametern μ und σ^2 abhängt. Somit können wir nun ein α -Konfidenzintervall für μ wie folgt bestimmen:

Es sei c ein $(1 + \alpha)/2$ -Quantil der t_{n-1} -Verteilung, sodass

$$\mathbb{P}_{\mu, \sigma^2}(|T| > c) = 2\mathbb{P}_{\mu, \sigma^2}(T > c) = 2\left(1 - \frac{1 + \alpha}{2}\right) = 1 - \alpha.$$

Somit ist

$$\mathbb{P}_{\mu, \sigma^2}\left(\bar{X} - \frac{cS}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{cS}{\sqrt{n}}\right) = \mathbb{P}_{\mu, \sigma^2}(|T| \leq c) = \alpha.$$

Wir haben also ein α -Konfidenzintervall für μ gefunden.

Beispiel 3.6.7. Wir betrachten wiederum Beispiel 3.6.3. Man rechnet leicht nach, dass $S^2 = 8,3$, also $S = 2,88$ ist. Um ein 95%-Konfidenzintervall zu berechnen, bestimmen wir zunächst das 0,975-Quantil der t -Verteilung mit 4 Freiheitsgraden. Aus einer Tabelle entnehmen wir den Wert 2,78. Somit ergibt sich

$$\frac{cS}{\sqrt{n}} = \frac{2,78 \cdot 2,88}{\sqrt{5}} \approx 3,58$$

Daher ist $[42,02, 49,18]$ ein 95%-Konfidenzintervall für μ . Beachte, dass dieses Intervall länger ist als das in Beispiel 3.6.3 berechnete, obwohl die dort angenommene Standardabweichung 3 größer als die nun geschätzte Standardabweichung 2,88 ist. Der Preis dafür, dass wir σ^2 als unbekannt annehmen (dürfen!) ist ein längeres Konfidenzintervall.

Konfidenzintervall für σ^2

Wir wollen nun noch ein Konfidenzintervall für die Varianz einer Stichprobe X_1, \dots, X_n zur Normalverteilung $\mathcal{N}(\mu, \sigma^2)$ bestimmen. Wir nehmen an, dass der Erwartungswert μ bekannt sei. Wir wissen bereits, dass $R := \frac{n-1}{\sigma^2} S^2 \sim \chi_{n-1}^2$. Insbesondere hängt die Verteilung dieser Größe nicht mehr von den Parametern μ und σ^2 ab. Wir können daher Konfidenzintervalle für σ^2 mittels Quantilen der χ_{n-1}^2 -Verteilung bestimmen.

Um ein α -Konfidenzintervall zu bestimmen, gehen wie folgt vor. Wir bestimmen a und b , sodass $\mathbb{P}(R \in [a, b]) = \alpha$. Dann ist

$$\alpha = \mathbb{P}_{\mu, \sigma^2}\left(\frac{n-1}{\sigma^2} S^2 \in [a, b]\right) = \mathbb{P}_{\mu, \sigma^2}\left(\frac{(n-1)S^2}{b} \leq \sigma^2 \leq \frac{(n-1)S^2}{a}\right).$$

Wir haben also ein α -Konfidenzintervall gefunden. Es bleibt noch a und b zu wählen. Hier ist zu beachten, dass (anders als bei der Standardnormalverteilung und der t -Verteilung), die χ^2 -Verteilung *nicht* symmetrisch ist, genauer gesagt nimmt eine χ^2 -verteilte Zufallsvariable nur positive Werte an. Daher wählt man a und b in der Regel wie folgt:

Wir wählen a so, dass $R \in [0, a]$ Wahrscheinlichkeit $(1 - \alpha)/2$ hat, also ist a das $(1 - \alpha)/2$ -Quantil der χ_{n-1}^2 -Verteilung. Dann wählen wir b so, dass $R \in [b, \infty)$ ebenfalls Wahrscheinlichkeit $(1 - \alpha)/2$ hat, also ist b das $1 - (1 - \alpha)/2 = (1 + \alpha)/2$ -Quantil der χ_{n-1}^2 -Verteilung.

Beispiel 3.6.8. Wir betrachten wieder Beispiel 3.6.3, wo wir $S^2 = 8,3$ bei einem Stichprobenumfang von $n = 5$ beobachtet hatten. Um ein Konfidenzintervall zum Niveau 0,95 zu bestimmen benötigen wir noch Quantile der χ_4^2 -Verteilung. Das 2,5%-Quantil von χ_4^2 ist

gegeben durch $a = 8,91$. Andererseits ist das 97,5%-Quantil gegeben durch $b = 32,85$. Somit ergibt sich als 95%-Konfidenzintervall für σ^2

$$\left[\frac{4 \cdot 8,3}{32,85}, \frac{4 \cdot 8,3}{8,91} \right] = [2,98, 68,5].$$

Auf Grund des kleinen Stichprobenumfangs ist dies recht ungenau.