Statistik

Rade Kutil, 2013W

Inhaltsverzeichnis

Statistik beschäftigt sich damit, Daten, die aus natürlichen Prozessen (physikalischen, wirtschaftlichen, ...) entstehen bzw. dort gemessen werden, als Produkte von zufälligen Prozessen zu modellieren und daraus Rückschlüsse auf reale Zusammenhänge zu ziehen. Dabei gibt es drei Hauptkapitel. Das erste ist die *deskriptive Statistik*, bei der versucht wird, die Daten in Form von Histogrammen oder Kennwerten zusammenzufassen, die die Daten übersichtlicher beschreiben können. Das zweite ist die *Wahrscheinlichkeitstheorie*, in der der Begriff des zufälligen Ereignisses entwickelt wird, sowie Zufallsvariablen als Ereignissen zugeordnete Werte untersucht werden. Im dritten Kapitel, der *schließenden Statistik*, werden die ersten beiden Kapitel zusammengeführt. Das heißt, Stichproben werden als Realisierungen von Zufallsvariablen betrachtet und aus den Kennwerten der Stichprobe auf Kennwerte der Zufallsvariablen geschlossen.

1 Deskriptive Statistik

Definition 1.1. Eine **Stichprobe** (sample) ist eine Menge von **Merkmalsausprägungen** einer **Population** (auch: Grundgesamtheit). Eine Stichprobe besteht in den meisten Fällen aus n Werten $\{x_1, x_2, \ldots, x_n\}$ aus $\mathbb{R}$ oder $\mathbb{N}$. n heißt **Stichprobengröße**.

Beispiel 1.2. Folgende Menge dient in der Folge als Beispiel für eine Stichprobe: $\{0.14, 0.27, 0.43, 0.68, 0.81, 1.14, 1.45, 1.82, 2.36, 2.53, 2.90, 3.45, 4.51, 5.12, 5.68, 7.84\}$. Die Stichprobengröße ist $n = 16$.

Zur Darstellung wird der Wertebereich einer Stichprobe in Klassen unterteilt.

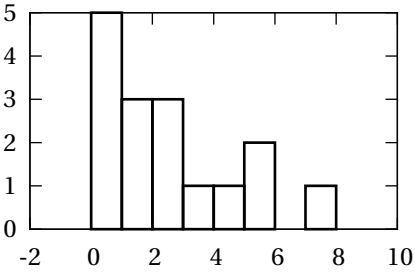
Definition 1.3. Eine Klasseneinteilung ist eine Folge $\{b_0, b_1, \ldots b_n\}$ von aufsteigenden Klassengrenzen. Die Intervalle $[b_{k-1}, b_k)$ werden als Klassen bezeichnet.

Definition 1.4. Die **absolute Häufigkeit** der k-ten Klasse $[b_{k-1}, b_k)$ ist die Anzahl der Stichprobenwerte, die in die Klasse fallen.

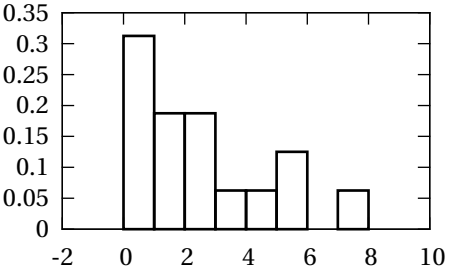
$$H_k := |\{i \mid b_{k-1} \leq x_i < b_k\}|$$

Die **relative Häufigkeit** ist $h_k := \frac{H_k}{n}$.

Definition 1.5. Ein **Histogramm** ist eine graphische Darstellung einer Stichprobe, bei der Balken zwischen den Klassengrenzen b_{k-1}, b_k mit der Höhe der absoluten oder relativen Häufigkeit eingezeichnet werden.

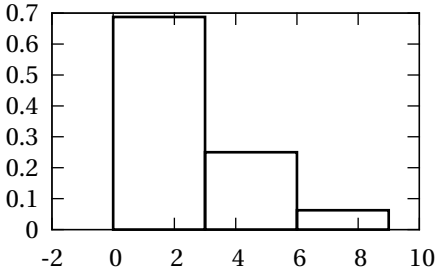


(a) mit absoluten Häufigkeiten

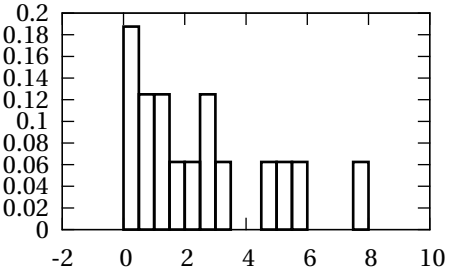


(b) mit relativen Häufigkeiten

Abbildung 1: Histogramme der Beispielstichprobe

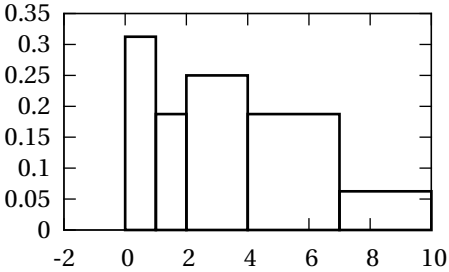


(a) zu große Klassenbreite

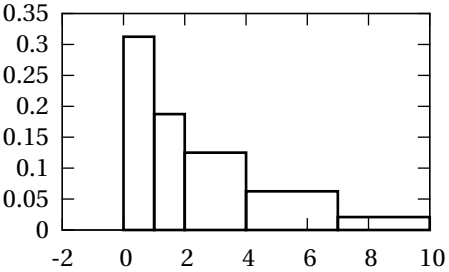


(b) zu kleine Klassenbreite

Abbildung 2: Histogramme mit falsch gewählten Klassengrößen



(a) ohne Skalierung



(b) mit Skalierung

Abbildung 3: Histogramme mit und ohne Skalierung

Beispiel 1.6. Wir wählen die gleichmäßig verteilten Klassengrenzen $\{0,1,2,\ldots\}$ und berechnen zunächst die absoluten Häufigkeiten, z.B. $H_3 = |\{i \mid 2 \le x_i < 3\}| = |\{2.36, 2.53, 2.90\}| = 3$, und die relativen Häufigkeiten, z.B. $h_3 = \frac{H_3}{n} = \frac{3}{16} \approx 0.19$. Dann zeichnen wir die Histogramme wie in Abbildung 1.

Die Wahl der Klassenbreite ist wichtig, wie man in Abbildung 2 sieht. Bei zu großer Klassenbreite zeigen zwar die Balkenhöhen aussagekräftige Werte an, aber die Auflösung des Wertebereichs ist zu gering. Bei zu kleiner Klassenbreite sind wiederum die Balkenhöhen zu gering aufgelöst.

Vor allem wenn die Klassenbreiten nicht gleich sind, ist es sinnvoll, die Balkenhöhe umgekehrt proportional zur Klassenbreite zu skalieren.

Definition 1.7. Die **skalierte Balkenhöhe** der k-ten Klassen ist die relative Häufigkeit dividiert durch die Klassenbreite: $\frac{h_k}{b_k - b_{k-1}}$.

Dadurch entspricht der Flächeninhalt des Balkens der Häufigkeit: $A_k = (b_k - b_{k-1})\frac{h_k}{b_k - b_{k-1}} = h_k$.

Beispiel 1.8. Wir wählen die ungleichmäßig verteilten Klassengrenzen $\{0,1,2,4,7,10\}$ und berechnen die relativen Häufigkeiten, z.B. $h_3 = \frac{4}{16} = 0.25$ und die skalierten Balkenhöhen $\frac{h_3}{b_3 - b_2} = \frac{0.25}{4-2} = 0.125$. Dann zeichnen wir die Histogramme wie in Abbildung 3.

Definition 1.9. Die **empirische Verteilungsfunktion** $F(x)$ gibt für jedes $x \in \mathbb{R}$ die relative Häufigkeit der Werte kleiner oder gleich x an:

$$F(x) := \frac{1}{n}|\{i \mid x_i \le x\}|$$

Beispiel 1.10. Die Werte der Verteilungsfunktion können für jedes beliebige x ausgerechnet werden, z.B. $F(0.75) = \frac{1}{16}|\{0.14, 0.27, 0.43, 0.68\}| = \frac{4}{16} = 0.25$. Am besten rechnet man es aber für jeden Stichprobenwert x_i aus und zeichnet zwischen den Punkten Plateaus ein. Das Ergebnis sieht man in Abbildung 4.

Bei diskreten Merkmalen ($x_i \in \mathbb{N}$) ist es wahrscheinlich, dass gleiche Werte mehrmals auftreten. In diesem Fall ist es geschickter, die Werte in Form einer Häufigkeitstabelle anzugeben.

Definition 1.11. Eine Häufigkeitstabelle gibt eine diskrete Stichprobe als Liste von Paaren (x_i, H_i) an. Das bedeutet, dass der Wert x_i mit der Anzahl H_i auftritt.

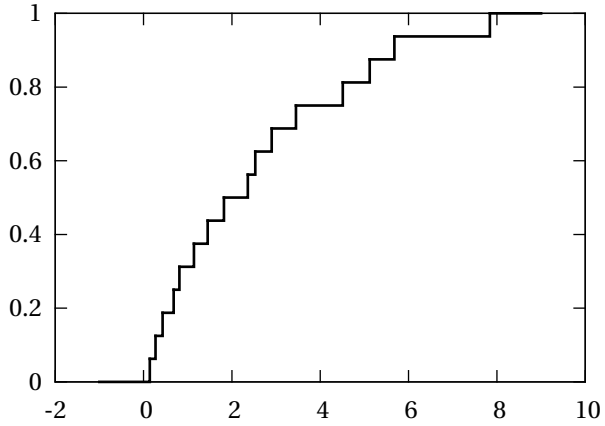


Abbildung 4: Empirische Verteilungsfunktion

Beispiel 1.12.

x_i	3	4	5	6	7
H_i	2	5	7	6	3

$$i = 1 \ldots m, \quad m = 5, \quad n = \sum_{i=1}^{m} H_i = 2+5+7+6+3 = 23$$

Für das Histogramm reicht es meist, *eine* Klasse pro Wert x_i zu verwenden und die absolute (H_i) oder relative Häufigkeit ($h_i = \frac{H_i}{n}$) aufzutragen.

Definition 1.13. Die **diskrete empirische Verteilungsfunktion** ist dann

$$F(x) = \frac{1}{n} \sum_{x_i \leq x} H_i \,.$$

Beispiel 1.14. Z.B.: $x_2 = 4, \quad H_2 = 5, \quad h_2 = \frac{5}{23} \approx 0.22.$

$$F(4.6) = \frac{1}{23} \sum_{x_i \leq 4.6} H_i = \frac{2+5}{23} = \frac{7}{23} \approx 0.3 \,.$$

Eine Möglichkeit, die Verteilung bzw. die Eigenschaften der Verteilung einer Stichprobe zu beschreiben, ist die Analyse mittels Berechnung einer Reihe von Maßzahlen (Kennwerte, Lageparameter).

Definition 1.15. Das **arithmetische Mittel** (der Mittelwert) einer Stichprobe x ist

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

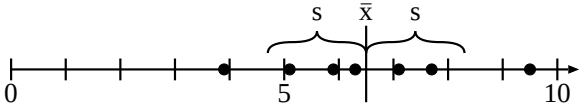


Abbildung 5: Stichprobe, arithmetisches Mittel und Standardabweichung.

Beispiel 1.16. Wir verwenden in diesem und folgenden Beispielen die Stichprobe $x = \{5.1, 7.7, 5.9, 3.9, 9.5, 7.1, 6.3\}$.

$$\bar{x} = \frac{5.1 + \ldots + 6.3}{7} = \frac{45.5}{7} = 6.5\,.$$

Definition 1.17. Die empirische **Standardabweichung** s und die empirische **Varianz** s^2 sind Maße für die mittlere (quadratische) Abweichung der Stichprobenwerte vom Mittelwert.

$$s^2 := \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2, \qquad s := \sqrt{s^2}\,.$$

Beispiel 1.18.

$$s^2 = \frac{(5.1-6.5)^2 + \ldots + (6.3-6.5)^2}{6} = \frac{19.92}{6} = 3.32, \qquad s = \sqrt{2.32} \approx 1.82\,.$$

Siehe Abbildung 5 für eine Veranschaulichung.

Bemerkung: Es gibt auch die analog zu Zufallsvariablen definierte Varianz $\sigma^2 = \frac{1}{n}\sum(x_i - \bar{x})^2$. Diese liefert jedoch im Mittel nicht das zu erwartende Ergebnis (siehe Kapitel Schätzer).

Satz 1.19 (Verschiebungssatz)**.** Auf folgende Weise kann die empirische Varianz oft einfacher berechnet werden.

$$s^2 = \frac{1}{n-1}\left(\left(\sum_{i=1}^{n} x_i^2\right) - n\bar{x}^2\right)$$

Beweis.

$$\begin{aligned}\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(x_i^2 - 2x_i\bar{x} + \bar{x}^2)\\ &= \frac{1}{n-1}\left(\left(\sum_{i=1}^{n} x_i^2\right) - 2\underbrace{\left(\sum_{i=1}^{n} x_i\right)}_{=n\bar{x}}\bar{x} + n\bar{x}^2\right) = \frac{1}{n-1}\left(\left(\sum_{i=1}^{n} x_i^2\right) - n\bar{x}^2\right)\end{aligned}$$ □

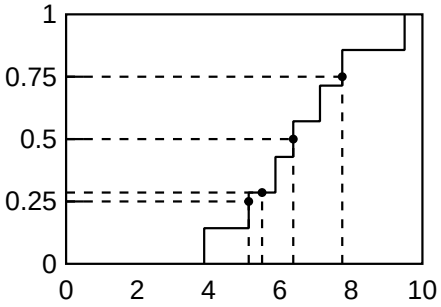


Abbildung 6: Quantile als inverse Verteilungsfunktion.

Beispiel 1.20.

$$s^2 = \frac{5.1^2 + \ldots + 6.3^2 - 7 \cdot 6.5^2}{6} = 3.32\,.$$

Definition 1.21. Das **Quantil** $\tilde{x}_\alpha$ ist der Wert, unter dem ein Anteil von $\alpha \in [0,1]$ der Werte liegt. $\tilde{x}_{\frac{1}{2}}$ heißt **Median**, $\tilde{x}_{\frac{1}{4}}$ heißt **erstes** oder unteres **Quartil**, $\tilde{x}_{\frac{3}{4}}$ heißt **drittes** oder oberes **Quartil**. *Wenn die Werte x_i sortiert sind*, dann ist $\tilde{x}_\alpha$ im Prinzip der Wert an der Stelle $\alpha \cdot n$. Das kann man als die inverse empirische Verteilungsfunktion auffassen. Für $0 < \alpha < 1$ gilt:

$$\tilde{x}_\alpha := \begin{cases} x_{\lceil \alpha n \rceil} & \alpha n \notin \mathbb{N} \\ \frac{x_{\alpha n} + x_{\alpha n + 1}}{2} & \alpha n \in \mathbb{N} \end{cases}$$

Beispiel 1.22.

$$\tilde{x}_{\frac{1}{4}} = x_{\lceil \frac{7}{4} \rceil} = x_2 = 5.1, \quad \tilde{x}_{\frac{1}{2}} = x_4 = 6.3, \quad \tilde{x}_{\frac{3}{4}} = x_6 = 7.7, \quad \tilde{x}_{\frac{2}{7}} = \frac{x_2 + x_3}{2} = 5.5$$

$$\{3.9, \overbrace{5.1}, 5.9, \overbrace{6.3}, 7.1, \overbrace{7.7}, 9.5\}$$

Abbildung 6 zeigt, wie die Quantile an der Verteilungsfunktion abgelesen werden können.

Satz 1.23. Sind die Daten als Häufigkeitstabelle gegeben, so gilt:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{m} H_i x_i\,,$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{m} H_i (x_i - \bar{x})^2 = \frac{1}{n-1} \left(\left(\sum_{i=1}^{m} H_i x_i^2 \right) - n\bar{x}^2 \right)$$

Definition 1.24. Der **Modus** ist der am häufigsten auftretende Wert x_i in der Stichprobe.

Beispiel 1.25. Gegeben sei folgende Häufigkeitstabelle:

$$\begin{array}{c|cccc} x_i & 0 & 1 & 2 & 3 \\ \hline H_i & 2 & 4 & 6 & 4 \end{array} \quad m = 4, \quad n = \sum_{i=1}^{4} H_i = 16$$

$$\bar{x} = \frac{2\cdot 0 + 4\cdot 1 + 6\cdot 2 + 4\cdot 3}{16} = \frac{28}{16} = \frac{7}{4}$$

$$s^2 = \frac{2\cdot 0^2 + 4\cdot 1^2 + 6\cdot 2^2 + 4\cdot 3^2 - 16\cdot\left(\frac{7}{4}\right)^2}{15} = \frac{64 - 16\frac{49}{16}}{15} = \frac{15}{15} = 1$$

Der Modus ist 2. (Und nicht 6!)

2 Korrelation und Regression

Definition 2.1. Bei **mehrdimensionalen Stichproben** werden mehrere Merkmale gleichzeitig gemessen. Eine zweidimensionale Stichproben besteht daher aus n Paaren von Werten $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$, die man natürlich auch in Tabellenform angeben kann.

Definition 2.2. Die **Kovarianz** ist definiert durch

$$s_{x,y} := \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n-1}\left(\left(\sum_{i=1}^{n} x_i y_i\right) - n\bar{x}\bar{y}\right)$$

Beispiel 2.3.

$$\begin{array}{c|cccccc} x_i & 1.2 & 1.8 & 3.2 & 4.6 & 4.9 & 5.8 \\ \hline y_i & 3.1 & 4.2 & 3.8 & 5.8 & 7.2 & 7.9 \end{array} \quad \begin{array}{ll} \bar{x} = 3.583 & s_x = 1.827 \\ \bar{y} = 5.333 & s_y = 1.945 \end{array}$$

$$s_{x,y} = \frac{1.2\cdot 3.1 + \ldots + 5.8\cdot 7.9 - 6\cdot 3.58\cdot 5.33}{5} = 3.31$$

Definition 2.4. Der **Korrelationskoeffizient** $r_{x,y}$ bzw. das **Bestimmtheitsmaß** $r_{x,y}^2$ geben die Stärke des linearen Zusammenhangs zwischen zwei Merkmalen an. Extremfälle sind: maximale (positive) Korrelation ($r_{x,y} = 1$), keine Korrelation ($r_{x,y} = 0$), negative Korrelation ($r_{x,y} = -1$).

$$r_{x,y} := \frac{s_{x,y}}{s_x s_y} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sqrt{\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)\left(\sum_{i=1}^{n} y_i^2 - n\bar{y}^2\right)}}$$

Beispiel 2.5.

$$r_{x,y} = \frac{3.31}{1.83 \cdot 1.95} = 0.928$$

Bei der **Regression** versucht man, eine Funktion f (**Modell**) zu finden, mit der ein Merkmal y durch das andere Merkmal x mittels $f(x)$ möglichst gut geschätzt werden kann (Modellanpassung). Dabei soll die gesamte Abweichung durch die Wahl der Parameter von f minimiert werden.

Definition 2.6. Die **Summe der Fehlerquadrate** (sum of squared errors, SSE) ist

$$S(f,x,y) := \sum_{i=1}^{n} (y_i - f(x_i))^2 .$$

Definition 2.7. Die Regression nach der **Methode der kleinsten Quadrate** ist die Minimierung der Summe der Fehlerquadrate durch Variation der Parameter $a_1, a_2, \ldots$ des Modells $f_{a_1,a_2,\ldots}$.

$$(\hat{a}_1, \hat{a}_2, \ldots) = \operatorname*{argmin}_{a_1,a_2,\ldots} S(f_{a_1,a_2,\ldots}, x, y) = \operatorname*{argmin}_{a_1,a_2,\ldots} \sum_{i=1}^{n} (y_i - f_{a_1,a_2,\ldots}(x_i))^2$$

Definition 2.8. Bei der **linearen Regression** setzt sich f aus einer Linearkombination $f = a_1 f_1 + a_2 f_2 + \ldots + a_m f_m$ zusammen. Die f_k können dabei beliebige Funktionen sein.

Satz 2.9. Die Lösung einer linearen Regression ergibt sich aus der Lösung des linearen Gleichungssystems $Ca = b$, wobei a der Vektor der m Parameter $a_1, a_2, \ldots$ ist, C eine $m \times m$ Matrix und b ein m-Vektor ist mit

$$C_{k,l} = \sum_{i=1}^{n} f_k(x_i) f_l(x_i), \qquad b_k = \sum_{i=1}^{n} y_i f_k(x_i) .$$

Beweis. Das Maximum von $S(f,x,y)$ als Funktion der Parameter a_k können wir durch partielles Ableiten nach den a_k und Nullsetzen ermitteln.

$$\begin{aligned} 0 = \frac{\partial S(f,x,y)}{\partial a_k} &= \frac{\partial}{\partial a_k} \sum_{i=1}^{n} (y_i - a_1 f_1(x_i) - a_2 f_2(x_i) - \ldots - a_m f_m(x_i))^2 \\ &= \sum_{i=1}^{n} 2(y_i - a_1 f_1(x_i) - a_2 f_2(x_i) - \ldots) f_k(x_i) . \end{aligned}$$

Durch Nullsetzen und Hineinziehen und Trennen der Summe bekommen wir

$$a_1 \sum_{i=1}^{n} f_1(x_i) f_k(x_i) + a_2 \sum_{i=1}^{n} f_2(x_i) f_k(x_i) + \ldots a_m \sum_{i=1}^{n} f_m(x_i) f_k(x_i) = \sum_{i=1}^{n} y_i f_k(x_i) ,$$

was der Zeile k des obigen Gleichungssystems entspricht. □

Beispiel 2.10. Das Modell $y = a_1 + a_2 x + a_3 x^2$ soll an die Daten aus Beispiel 2.3 angepasst werden. D.h. $f_1(x) = 1$, $f_2(x) = x$ und $f_3(x) = x^2$. Es ergibt sich das Gleichungssystem

$$\begin{pmatrix} 6 & 21.5 & 93.73 \\ 21.5 & 93.73 & 450.4 \\ 93.73 & 450.4 & 2273 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} 32 \\ 131.2 \\ 618.3 \end{pmatrix},$$

wobei z.B. $C_{1,3} = 1 \cdot 1.2^2 + 1 \cdot 1.8^2 + 1 \cdot 3.2^2 + \ldots = 93.73$ oder $C_{2,2} = 1.2 \cdot 1.2 + 1.8 \cdot 1.8 + \ldots = 93.73$, und $b_3 = 3.1 \cdot 1.2^2 + 4.2 \cdot 1.8^2 \ldots = 618.3$. Die Lösung des Gleichungssystems ergibt

$$a_1 = 3.685, \quad a_2 = -0.4653, \quad a_3 = 0.2123.$$

Definition 2.11. In einem **linearen Regressionsmodell** muss f eine lineare Funktion sein. Bei zweidimensionalen Stichproben heißt das: $y \approx f(x) = ax + b$. f nennt man **Regressionsgerade**.

Satz 2.12. Für das lineare Regressionsmodell ergibt sich

$$a = \frac{s_{x,y}}{s_x^2}, \quad b = \bar{y} - a\bar{x}.$$

Beweis. Mit $f_1 = 1$, $f_2 = x$ und $a_1 = b$, $a_2 = a$ lässt sich Satz 2.9 anwenden und man erhält

$$\begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix} \begin{pmatrix} b \\ a \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix}$$

Multipliziert man die erste Zeile mit $\frac{1}{n}\sum x_i$ und subtrahiert sie von der zweiten Zeile, erhält man

$$b \cdot 0 + a(\sum x_i^2 - \frac{1}{n}(\sum x_i)^2) = \sum x_i y_i - \frac{1}{n} \sum xi \sum y_i$$

$$a = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2}.$$

Erweitert man oben und unten mit $\frac{1}{n-1}$, ergibt sich wie gewünscht $a = \frac{s_{x,y}}{s_x^2}$. Dividiert man die erste Zeile des Gleichungssystems durch n, erhält man $b \cdot 1 + a\bar{x} = \bar{y}$ und daraus $b = \bar{y} - a\bar{x}$. □

Beispiel 2.13. $a = \frac{3.31}{1.83^2} = 0.992$, $b = 5.33 - 0.992 \cdot 3.58 = 1.779$.

Definition 2.14. Die Stichprobe kann in einem **Streudiagramm (Scatterplot)** dargestellt werden, in das man jeden Punkt (x_i, y_i) einzeichnet. In ein Streudiagramm können auch Regressionsgeraden und -kurven eingezeichnet werden.

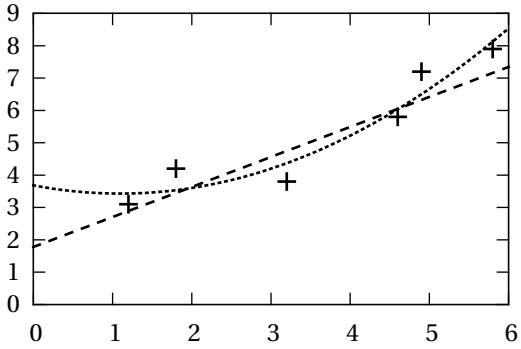


Abbildung 7: Streudiagramm mit Regressions-Kurve und -Gerade.

Beispiel 2.15. Abbildung 7 zeigt das Streudiagramm aus Beispiel 2.3, sowie die Regressionskurve aus Beispiel 2.10 und die Regressionsgerade aus Beispiel 2.13.

Definition 2.16. Wie bei eindimensionalen diskreten Merkmalen gibt es auch hier ein Äquivalent zu Häufigkeitstabellen: **Kontingenztabellen**. Dabei stellt $H_{i,j}$ die absolute Häufigkeit der gemeinsamen Merkmalsausprägung (x_i, y_i) dar. Die **Randhäufigkeiten** $H_{i,\cdot}$ und $H_{\cdot,j}$ sind die Häufigkeiten von x_i bzw. y_j. $h_{i,j} = \frac{H_{i,j}}{n}$, $h_{i,\cdot} = \frac{H_{i,\cdot}}{n}$, $h_{\cdot,j} = \frac{H_{\cdot,j}}{n}$ sind die zugehörigen relativen Häufigkeiten.

	y_1	$\dots$	y_m	$\sum$
x_1	$H_{1,1}$	$\dots$	$H_{1,m}$	$H_{1,\cdot}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_l	$H_{l,1}$	$\dots$	$H_{l,m}$	$H_{l,\cdot}$
$\sum$	$H_{\cdot,1}$	$\dots$	$H_{\cdot,m}$	n

	y_1	$\dots$	y_m	$\sum$
x_1	$h_{1,1}$	$\dots$	$h_{1,m}$	$h_{1,\cdot}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_l	$h_{l,1}$	$\dots$	$h_{l,m}$	$h_{l,\cdot}$
$\sum$	$h_{\cdot,1}$	$\dots$	$h_{\cdot,m}$	1

Satz 2.17. Im Fall von Kontingenztabellen wird die Kovarianz berechnet durch:

$$s_{x,y} = \frac{1}{n-1} \sum_{i=1}^{l} \sum_{j=1}^{m} H_{i,j}(x_i - \bar{x})(y_j - \bar{y}) = \frac{1}{n-1} \left(\left(\sum_{i=1}^{l} \sum_{j=1}^{m} H_{i,j} x_i y_j \right) - n\bar{x}\bar{y} \right).$$

3 Ereignis- und Wahrscheinlichkeitsraum

Als Geburtsstunde der Wahrscheinlichkeitstheorie wird heute ein Briefwechsel zwischen Blaise Pascal und Pierre de Fermat aus dem Jahr 1654 gesehen, in dem es um eine spezielle Fragestellung zum Glücksspiel ging. Solche Fragestellungen waren damals noch schwer zu behandeln, weil die axiomatische Begründung

der Wahrscheinlichkeitstheorie fehlte, die erst Anfang des 20. Jahrhunderts entwickelt wurde.

Ein Ereignis wie z.B. „5 gewürfelt“ ist eine Abstraktion, weil es viele Details wie etwa die Lage des Würfel ignoriert. Es fasst daher viele mögliche Ereignisse zusammen. Weiters gilt auch „Zahl größer 4 gewürfelt“ als Ereignis, welches nochmals zwei Ereignisse zusammenfasst, nämlich „5 gewürfelt“ und „6 gewürfelt“. Daher werden Ereignisse als Mengen abgebildet.

Die Elemente, aus denen diese Mengen gebildet werden, nennt man Ergebnisse, weil sie die möglichen Ergebnisse eines Zufallsprozesses sind. Sie werden in der Ergebnismenge zusammengefasst.

Definition 3.1. Der **Ergebnisraum** (oft auch Grundmenge) Ω ist die Menge aller möglichen Ergebnisse eines Zufallsprozesses.

Definition 3.2. Ein **Ereignisraum** (Ω, Σ) besteht aus einer Menge Σ von Ereignissen e aus dem Ergebnisraum Ω, $e \in \Sigma$, $e \subseteq \Omega$, und erfüllt folgende Axiome (σ-Algebra):

- $\Omega \in \Sigma, \quad \emptyset \in \Sigma$
- $e \in \Sigma \Rightarrow \Omega \setminus e \in \Sigma$
- $e_1, e_2, \ldots \in \Sigma \Rightarrow \bigcup_i e_i \in \Sigma$
- $e_1, e_2, \ldots \in \Sigma \Rightarrow \bigcap_i e_i \in \Sigma$

Das erste Axiom stellt sicher, dass Ω selbst, das sichere Ereignis, sowie die leere Menge $\emptyset$, das unmögliche Ereignis, im Ereignisraum enthalten sind. Das zweite bewirkt, dass zu jedem Ereignis e auch das **Gegenereignis** $\bar{e} = \Omega \setminus e$ enthalten ist. Die letzten zwei Axiome stellen sicher, dass die Vereinigung und der Durchschnitt beliebiger Ereignisse enthalten sind. Man beachte, dass sowohl Ω als auch Σ unendlich groß sein können. Daher reicht es nicht, nur die Vereinigung *zweier* Ereignisse zu inkludieren, sondern es muss auch die Vereinigung abzählbar-unendlich vieler Ereignisse inkludiert sein, was nicht automatisch folgen würde.

Beispiel 3.3. Beim Werfen eines Würfels ist der Ergebnisraum $\Omega = \{1,2,3,4,5,6\}$ und der Ereignisraum $\Sigma = \{\{1\},\{2\},\{1,2\},\ldots,\{1,2,3,4,5,6\}\}$. Das Elementarereignis „Zahl 4 gewürfelt“ entspricht dem Element 4 bzw. der einelementigen Teilmenge $\{4\}$. Das Ereignis „Zahl kleiner 3 gewürfelt“ entspricht der Menge $e_1 = \{1,2\} \subseteq \Omega$, „Zahl größer 4“ entspricht $e_2 = \{5,6\} \subseteq \Omega$.

Die Vereinigung zweier Ereignisse interpretiert man als „oder“. $e_1 \cup e_2$ bedeutet also „Zahl kleiner 3 *oder* größer 4 gewürfelt“. Gleichermaßen interpretiert man

den Durchschnitt als „und“, das Gegenereignis als „nicht“. $\bar{e}_1$ bedeutet also „Zahl größer-gleich 3 gewürfelt“.

Nun soll jedem Ereignis e eine Wahrscheinlichkeit $P(e) \in [0,1]$ zugeordnet werden. P stellt dabei ein *Maß* dar im Sinne der Maßtheorie. Zusammen mit dem Ereignisraum bildet sie einen Maßraum, den Wahrscheinlichkeitsraum.

Definition 3.4. Ein **Wahrscheinlichkeitsraum** (Ω, Σ, P) über einem Ereignisraum (Ω, Σ) enthält das **Wahrscheinlichkeitsmaß** $P : \Sigma \to \mathbb{R}$ und erfüllt folgende Axiome (nach Kolmogorov, 1933):

- $\forall e \in \Sigma : P(e) \geq 0$
- $P(\Omega) = 1$
- $e_1, e_2, \ldots \in \Sigma \wedge \forall i \neq j : e_i \cap e_j = \emptyset \Rightarrow P\big(\bigcup_i e_i\big) = \sum_i P(e_i)$

Die Wahrscheinlichkeit darf also nicht negativ sein, das sichere Ereignis Ω hat Wahrscheinlichkeit 1, und bei der Vereinigung disjunkter Ereignisse summieren sich die Wahrscheinlichkeiten. Daraus ergeben sich nun einige Folgerungen.

Satz 3.5. Das Gegenereignis hat die Gegenwahrscheinlichkeit

$$P(\bar{e}) = 1 - P(e)\,.$$

Beweis. $e \cap \bar{e} = \emptyset \Rightarrow P(e) + P(\bar{e}) = P(e \cup \bar{e}) = P(\Omega) = 1.$ □

Satz 3.6.

$$P(\emptyset) = 0\,.$$

Beweis. $P(\emptyset) = P(\bar{\Omega}) = 1 - P(\Omega) = 1 - 1 = 0.$ □

Satz 3.7.

$$d \subseteq e \Rightarrow P(d) \leq P(e)\,.$$

Beweis. $P(e) = P(d \cup (e \setminus d)) \overset{d \cap (e \setminus d) = \emptyset}{=\!=\!=} P(d) + P(e \setminus d) \geq P(d).$ □

Wenn zwei disjunkte Ereignisse vereinigt werden, addieren sich laut Axiom ihre Wahrscheinlichkeiten. Was passiert, wenn sie nicht disjunkt sind, sehen wir hier:

Satz 3.8 (Additionssatz)**.**

$$P(d \cup e) = P(d) + P(e) - P(d \cap e)\,.$$

Beweis.

$$P(d) = P(d \setminus e \cup d \cap e) = P(d \setminus e) + P(d \cap e) \quad \Rightarrow \quad P(d \setminus e) = P(d) - P(d \cap e),$$

$$P(e) = P(e \setminus d \cup e \cap d) = P(e \setminus d) + P(e \cap d) \quad \Rightarrow \quad P(e \setminus d) = P(e) - P(d \cap e),$$

$$\begin{aligned} P(d \cup e) &= P(d \setminus e \cup e \setminus d \cup d \cap e) = P(d \setminus e) + P(e \setminus d) + P(d \cap e) \\ &= P(d) - P(d \cap e) + P(e) - P(d \cap e) + P(d \cap e). \end{aligned}$$ □

Sehr oft haben Ereignisse gleicher Größe (bezüglich der Ergebnisanzahl) auch die gleiche Wahrscheinlichkeit. Das muss nicht so sein, aber wenn es für den ganzen Wahrscheinlichkeitsraum zutrifft, dann gilt folgendes Modell:

Definition 3.9. Im **Laplace-Modell** (oder nach der **Laplace-Annahme**) für endliche Wahrscheinlichkeitsräume haben alle Elementarereignisse die gleiche Wahrscheinlichkeit p.

$$\forall \omega \in \Omega : P(\{\omega\}) = p.$$

Satz 3.10. Im Laplace-Modell kann die Wahrscheinlichkeit als relative Größe des Ereignisses berechnet werden:

$$\forall e \in \Sigma : P(e) = \frac{|e|}{|\Omega|}$$

Das nennt man oft das Prinzip „günstige durch mögliche Ergebnisse“. Insbesondere gilt für Elementarereignisse

$$\forall \omega \in \Omega : P(\{\omega\}) = \frac{1}{|\Omega|}.$$

Beweis.

$$1 = P(\Omega) = P\Big(\bigcup_i \{\omega_i\}\Big) = \sum_i P(\{\omega_i\}) = \sum_i p = |\Omega| p \quad \Rightarrow \quad p = \frac{1}{|\Omega|}.$$

$$P(e) = P\Big(\bigcup_{\omega \in e} \{\omega\}\Big) = \sum_{\omega \in e} P(\{\omega\}) = |e| p = \frac{|e|}{|\Omega|}.$$ □

Beispiel 3.11. Im Würfel-Beispiel gilt die Laplace-Annahme und daher für die Elementarereignisse $P(\{1\}) = P(\{2\}) = \ldots = P(\{6\}) = \frac{|\{1\}|}{|\{1,2,3,4,5,6\}|} = \frac{1}{6}$. Weiters gilt $P(e_1) = P(\text{„kleiner 3“}) = P(\{1,2\}) = \frac{|\{1,2\}|}{|\{1,2,3,4,5,6\}|} = \frac{2}{6} = \frac{1}{3}$. Und da e_1 und e_2 aus Beispiel 3.3 disjunkt sind, gilt $P(e_1 \cup e_2) = P(\{1,2,5,6\}) = \frac{2}{3} = P(e_1) + P(e_2) = \frac{1}{3} + \frac{1}{3}$. Hingegen ist nach dem Additionssatz $P(\text{„gerade oder kleiner 4“}) = P(\{2,4,6\} \cup \{1,2,3\}) = P(\{2,4,6\}) + P(\{1,2,3\}) - P(\{2\}) = \frac{1}{2} + \frac{1}{2} - \frac{1}{6} = \frac{5}{6}$. Und für das Gegenereignis gilt $P(\text{„nicht kleiner 3“}) = P(\bar{e}_1) = 1 - P(e_1) = 1 - \frac{1}{3} = \frac{2}{3}$.

Für unendliche Wahrscheinlichkeitsräume muss das Laplace-Modell anders formuliert werden, da die Elementarereignisse notgedrungen Wahrscheinlichkeit 0 hätten.

Definition 3.12. Im Laplace-Modell für unendliche Wahrscheinlichkeitsräume soll gelten

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

wobei $|e|$ ein Volumsmaß ist. Das erfüllt die Axiome des Wahrscheinlichkeitsraums, weil jedes Maß die Axiome 1 und 3 erfüllt und Axiom 2 wegen $P(\Omega) = \frac{|\Omega|}{|\Omega|} = 1$ erfüllt ist.

Wie gewünscht haben dann zwei gleich große Ereignisse die gleiche Wahrscheinlichkeit.

Beispiel 3.13. Die Wahrscheinlichkeit, dass der erste Regentropfen, der auf ein A4-Blatt fällt, in einen darauf gezeichneten Kreis mit Durchmesser 4 cm fällt, ist

$$P(\text{„in Kreis“}) = \frac{(2\,\text{cm})^2\pi}{2^{-4}\,\text{m}^2} \approx 2\%.$$

4 Kombinatorik

Zur Berechnung von Wahrscheinlichkeiten in endlichen Wahrscheinlichkeiten sind folgende kombinatorische Formeln essentiell. Als Modell zur Veranschaulichung dient das Urnenmodell, bei dem n nummerierte (also unterscheidbare) Kugeln in einer Urne liegen und auf verschiedene Art daraus gezogen werden. Bei den Variationen ist die Reihenfolge wichtig, mit der die Kugeln gezogen werden; bei den Kombinationen zählt nur, welche Kugeln gezogen werden, nicht ihre Reihenfolge. Weiters gibt es die Möglichkeit, jede Kugel, nachdem sie gezogen wurde, wieder in die Urne zurück zu legen, so dass sich die Kugeln in der Ziehung wiederholen können. Über die Anzahl der Möglichkeiten, so eine Ziehung durchzuführen, kann man die Wahrscheinlichkeit einer bestimmten Ziehung ausrechnen.

Definition 4.1. Als **Permutation** bezeichnet man die Menge der verschiedenen Anordnungen von n Elementen.

$$\{(a_1, a_2, \ldots, a_n) \mid a_i \in \{b_1, b_2, \ldots, b_n\} \wedge \forall i \neq j : a_i \neq a_j\}$$

Das entspricht der Ziehung *aller* n Kugeln aus einer Urne, wobei es auf die Reihenfolge ankommt, mit der die Kugeln gezogen werden.

Satz 4.2. Die Anzahl der Permutationen ist $n! = n \cdot (n-1) \cdots 2 \cdot 1$.

Beweis. Es gibt n Möglichkeiten, das erste Element auszuwählen. Danach gibt es für jede Auswahl $n-1$ Möglichkeiten für die Auswahl des zweiten Elements. Und so weiter, bis nur noch eine Möglichkeit für das letzte Element übrigbleibt. □

Beispiel 4.3.

$$|\{(a,b,c),(a,c,b),(b,a,c),(b,c,a),(c,a,b),(c,b,a)\}| = 3! = 3 \cdot 2 = 6$$

Beispiel 4.4. Wenn sich fünf Kinder nebeneinander stellen, wie groß ist die Wahrscheinlichkeit, dass sie nach Größe geordnet stehen (aufsteigend oder absteigend)? $|\Omega| = 5!$, $G = \{(a_1,a_2,a_3,a_4,a_5),(a_5,a_4,a_3,a_2,a_1)\}$, $P(\text{„geordnet“}) = \frac{|G|}{|\Omega|} = \frac{2}{5!} = \frac{1}{5 \cdot 4 \cdot 3} \approx 1.67\%$.

Definition 4.5. Als **Variation ohne Wiederholung** bezeichnet man die Menge der Anordnungen von je k aus n Elementen.

$$\{(a_1,a_2,\ldots,a_k) \mid a_i \in \{b_1,b_2,\ldots,b_n\} \wedge \forall i \neq j : a_i \neq a_j\}$$

Das entspricht der Ziehung von k Kugeln aus n ohne Zurücklegen, wobei es auf die Reihenfolge ankommt, mit der die Kugeln gezogen werden.

Satz 4.6. Die Anzahl der Variationen ohne Wiederholung ist $\frac{n!}{(n-k)!}$.

Beweis. Wie bei den Permutationen gibt es n Möglichkeiten für das erste Element, $n-1$ für das zweite, und so weiter. Für das letzte, also das k-te Element gibt es dann noch $n-k+1$ Möglichkeiten. Die gesamte Anzahl ist also $n(n-1)\cdots(n-k+1) = \frac{n!}{(n-k)!}$. □

Beispiel 4.7.

$$|\{(a,b),(a,c),(b,a),(b,c),(c,a),(c,b)\}| = \frac{3!}{(3-2)!} = \frac{3!}{1!} = \frac{3 \cdot 2}{1} = 6$$

Definition 4.8. Als **Variation mit Wiederholung** bezeichnet man die Menge der Anordnungen von je k aus n Elementen, wobei Elemente mehrfach vorkommen dürfen.

$$\{(a_1,a_2,\ldots,a_k) \mid a_i \in \{b_1,b_2,\ldots,b_n\}\}$$

Das entspricht der Ziehung von k Kugeln aus n mit Zurücklegen der Kugeln, wobei es auf die Reihenfolge ankommt.

Satz 4.9. Die Anzahl der Variationen mit Zurücklegen ist n^k.

Beweis. Für jedes der k ausgewählten Elemente gibt es n Möglichkeiten. Die Gesamtanzahl ist daher $n \cdot n \cdots n = n^k$. □

Beispiel 4.10.

$$|\{a,b,c\}^2| = |\{(x_1,x_2) \mid x_i \in \{a,b,c\}\}| = |\{(a,a),(a,b),(a,c),(b,a),\ldots,(c,c)\}| = 3^2 = 9$$

Beispiel 4.11. Wie groß ist die Wahrscheinlichkeit, dass in einer PS-Gruppe mit 30 Studenten mindestens zwei am selben Tag Geburtstag haben? Schaltjahre sind der Einfachheit halber zu ignorieren.

Die Menge der möglichen Geburtstagslisten ist $\Omega = \{1.1.,\ldots,31.12.\}^{30}$, also eine Variation mit Wiederholung, und daher ist $|\Omega| = 365^{30}$. Mit $G =$„mindestens zwei gleiche Geburtstage“ ist $\bar{G} =$„keine zwei gleichen Geburtstage“ eine Variation ohne Wiederholung. Daher ist

$$P(G) = 1 - P(\bar{G}) = 1 - \frac{\frac{365!}{(365-30)!}}{365^{30}} \approx 1 - 0.294 = 70.6\%\,.$$

Definition 4.12. Als **Kombination ohne Wiederholung** bezeichnet man die Menge der Auswahlen von je k aus n Elementen.

$$\{A \subseteq \{b_1,b_2,\ldots,b_n\} \mid |A| = k\}$$

Das entspricht der Ziehung von k Kugeln aus n ohne Zurücklegen, wobei die Reihenfolge der Kugeln keine Rolle spielt.

Satz 4.13. Die Anzahl der Kombinationen ohne Wiederholung ist

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

Beweis. Die Variationen ohne Wiederholung sind gerade die Permutationen aller Kombinationen ohne Wiederholung, wobei man auf diese Weise keine Variation doppelt erzeugt. Die Anzahl der Kombinationen ist also die der Variationen dividiert durch die Permutationen der k Elemente, also dividiert durch $k!$, also $\frac{n!}{(n-k)!k!}$. □

Beispiel 4.14.

$$|\{x \subseteq \{a,b,c,d\} \mid |x| = 2\}| = |\{\{a,b\},\{a,c\},\{a,d\},\{b,c\},\{b,d\},\{c,d\}\}|$$
$$= \binom{4}{2} = \frac{4!}{2!(4-2)!} = \frac{4!}{2!2!} = \frac{4\cdot 3\cdot 2}{2\cdot 2} = 6$$

Beispiel 4.15. Wie groß ist die Wahrscheinlichkeit, dass bei einer Lottoziehung (6 aus 45) die Zahl 13 vorkommt?

$$\Omega = \{x \subseteq \{1,\dots,45\} \mid |x| = 6\}, \quad |\Omega| = \binom{45}{6}.$$

$$P(\text{„13 kommt vor“}) = 1 - P(\text{„13 kommt nicht vor“}) = 1 - P(G)$$

mit $G := \{x \subseteq \{1,\dots,12,14,\dots 45\} \mid |x| = 6\}$. $|G| = \binom{44}{6}$. Es gilt die Laplace-Annahme (überlegen Sie warum). Daher ist

$$P(\text{„13 kommt vor“}) = 1 - \frac{|G|}{|\Omega|} = 1 - \frac{\binom{44}{6}}{\binom{45}{6}} = 1 - \frac{44!}{6!38!}\frac{6!39!}{45!} = 1 - \frac{39}{45} = \frac{6}{45} = \frac{2}{15} \approx 13.3\%.$$

Definition 4.16. Als **Kombination mit Wiederholung** bezeichnet man die Menge der Auswahlen von je k aus n Elementen, wobei Elemente mehrfach ausgewählt werden dürfen.

$$\{(a_1, a_2, \dots, a_k) \mid a_i \in \{b_1, b_2, \dots, b_n\} \wedge a_1 \le a_2 \le \dots \le a_k\}$$

Das entspricht der Ziehung von k Kugeln aus n mit Zurücklegen, wobei die Reihenfolge der Kugeln keine Rolle spielt.

Satz 4.17. Die Anzahl der Kombinationen mit Wiederholung ist

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

Beweis. Sowohl $(b_1, b_2, \dots, b_n)$ als auch $(a_1, a_2, \dots, a_k)$ seien aufsteigend sortiert. Jetzt fügt man Trennelemente # in die Kombination ein, und zwar zwischen $a_i = b_k$ und $a_{i+1} = b_l$ genau $l - k$ Stück, wobei vorübergehend $a_0 = b_1$ und $a_{k+1} = b_n$ gesetzt wird. So wird z.B. für $n = 5$ aus $(1,2,4,4)$ nun $(1,\#,2,\#,\#,4,4,\#)$. Die Kombination hat also jetzt $n + k - 1$ Elemente und wird durch die Position der Trennelemente # eineindeutig beschrieben. Alternativ kann man auch die Position der Nicht-Trennelemente festlegen. Diese Positionen sind aber einfach eine Auswahl von k aus den $n + k - 1$ Stellen, also eine Kombination $\binom{n+k-1}{k}$. □

Beispiel 4.18.

$$|\{(a,a),(a,b),(a,c),(a,d),(b,b),(b,c),(b,d),(c,c),(c,d),(d,d)\}|$$
$$= \binom{4+2-1}{2} = \binom{5}{2} = 10$$

Satz 4.19. Werden zufällige Kombinationen mit Wiederholung als aufeinander folgende zufällige Auswahl von Elementen implementiert, z.B. als Ziehung aus der Urne mit Zurücklegen, dann ist die Laplace-Annahme im Allgemeinen *nicht* zu treffen, d.h. die Kombinationen sind nicht gleich wahrscheinlich.

Aus diesem Grund haben Kombinationen mit Wiederholung in der Wahrscheinlichkeitsrechnung eigentlich keine allzu große Bedeutung.

Beweis. Bei den Variationen mit Wiederholung gilt die Laplace-Annahme. Die Kombinationen mit Wiederholung sind nun eine Abstraktion der Variationen, also eine Zusammenfassung von Variationen mit gleicher Elementauswahl aber unterschiedlicher Reihenfolge. Jetzt gibt es zum Beispiel nur eine Variation für $(1,1,1,1,1)$, aber fünf Variationen für $(1,1,1,1,2)$. Letztere Kombination ist also fünf mal so wahrscheinlich wie erstere. Bei anderen Kombinationen ist das Missverhältnis meist noch größer. □

5 Bedingte Wahrscheinlichkeit

Bei der bedingten Wahrscheinlichkeit wird ein zweites Ereignis B (die Bedingung) als neue Grundmenge betrachtet. Man kann das auch als zweistufiges Experiment betrachten: Zuerst tritt Ereignis B ein, dann wird ein Ereignis A unter dieser Voraussetzung untersucht. Jetzt hat A womöglich eine andere Wahrscheinlichkeit, da A von B vielleicht nicht unabhängig ist. Wir beschränken also den Ergebnisraum auf B und die Ereignisse auf jene, die in B liegen. Das Wahrscheinlichkeitsmaß P muss dann so normiert werden, dass $P(B)$ eins wird.

Satz 5.1. Sei (Ω, Σ, P) ein Wahrscheinlichkeitsraum und $B \in \Sigma$. Sei $\Sigma_B := \{e \in \Sigma \mid e \subseteq B\}$ und $P_B := P/P(B)$. Dann ist $(B, \Sigma_B, P_B))$ ein Wahrscheinlichkeitsraum.

Beweis. (B, Σ_B) ist ein Ereignisraum, denn (1) $B \in \Sigma_B$, weil $B \in \Sigma$ und $B \subseteq B$, (2) für $e \in \Sigma_B$ ist $B \setminus e \in \Sigma_B$, weil $B \setminus e = (\Omega \setminus e) \cap B$ und $\Omega \setminus e \in \Sigma$ und daher auch $(\Omega \setminus e) \cap B \in \Sigma$ und $(\Omega \setminus e) \cap B \subseteq B$, und (3,4) für $e_i \in \Sigma_B$ gilt $\bigcup e_i \in \Sigma_B$ und $\bigcap e_i \in \Sigma_B$, weil diese $\in \Sigma$ sind aber auch $\subseteq B$.

P_B ist ein zugehöriges Wahrscheinlichkeitsmaß, weil (1) $P(e) \geq 0 \Rightarrow P_B(e) = P(e)/P(B) \geq 0$, (2) $P_B(B) = P(B)/P(B) = 1$, und (3) für $e_i \in \Sigma_B \wedge e_i \neq e_j$ ist $P_B(\bigcup e_i) = P(\bigcup e_i)/P(B) = \sum P(e_i)/P(B) = \sum P_B(e_i)$. □

Definition 5.2. Die **bedingte Wahrscheinlichkeit** $P(A|B)$ („Wahrscheinlichkeit von A unter der Bedingung B“) ist

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

Abbildung 8: Mengendiagramm zur bedingten Wahrscheinlichkeit.

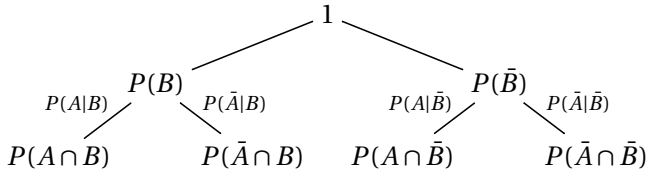


Abbildung 9: Zweistufiger Entscheidungsbaum.

In Abbildung 8 wird der Zusammenhang veranschaulicht. Die Wahrscheinlichkeit von $A \cap B$ ergibt sich also aus der Verkettung (Multiplikation) der Wahrscheinlichkeit von B in Bezug auf Ω ($P(B)$) und der Wahrscheinlichkeit von $A \cap B$ in Bezug auf B ($P(A|B)$).

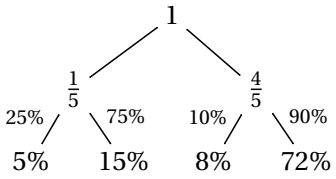
Beispiel 5.3. Gegeben seien folgende Ereignisse: A = „Auto hat nach Verkauf einen Defekt“, B = „Auto wurde an einem Montag produziert“. $P(B) = \frac{1}{5}$. Es sei P(„Auto ist Montagsauto *und* hat Defekt“) $= P(A \cap B) = 5\%$. Dann ist P(„Montagsauto hat Defekt“) $= P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{5\%}{\frac{1}{5}} = 25\%$.

Definition 5.4. Ein **Entscheidungsbaum** ist ein Graph zur Darstellung von zwei- und mehrstufigen Experimenten. Dabei steht an jeder Kante eine bedingte Wahrscheinlichkeit, am Anfang der Kante die Wahrscheinlichkeit der Bedingung und am Ende das Produkt der beiden. Abbildung 9 zeigt den zweistufigen Entscheidungsbaum.

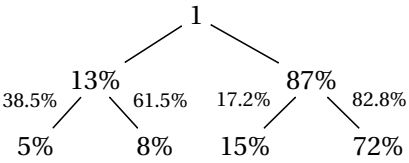
Abbildung 10(a) zeigt das Beispiel 5.3 als Entscheidungsbaum.

Satz 5.5. Der Satz von der **vollständigen Wahrscheinlichkeit** bietet die Möglichkeit, ein Ereignis A in mehrere Bedingungen $B_1, \ldots, B_n$ aufzuteilen. Die Bedingungen müssen vollständig sein ($B_1 \cup \ldots \cup B_n = \Omega$) und sich paarweise ausschließen ($\forall i \neq j : B_i \cap B_j = \emptyset$). Das gilt insbesondere für Bedingung und Gegenbedingung (B und $\bar{B}$). Es gilt

$$P(A) = P(A|B_1)P(B_1) + \ldots + P(A|B_n)P(B_n)\,,$$



(a) B = „Montag“, A = „Defekt“



(b) A und B vertauscht.

Abbildung 10: Beispiel für Entscheidungsbaum.

bzw. für $n = 2$

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

Beweis. $P(A|B_1)P(B_1) + \ldots + P(A|B_n)P(B_n) = P(A \cap B_1) + \ldots + P(A \cap B_n) \overset{B_i \cap B_j = \emptyset}{=\!=\!=} P(A \cap B_1 \cup \ldots \cup A \cap B_n) = P(A \cap (B_1 \cup \ldots \cup B_n)) \overset{\cup B_i = \Omega}{=\!=} P(A)$. □

Beispiel 5.6. P(„Auto nicht am Montag produziert“) $= P(\bar{B}) = \frac{4}{5}$. Es sei P(„Nichtmontagsauto hat Defekt“) $= P(A|\bar{B}) = 10\%$. Dann ist P(„Auto hat Defekt“) $= P(A)$ $= P(A|B)P(B) + P(A|\bar{B})P(\bar{B}) = 25\% \cdot \frac{1}{5} + 10\% \cdot \frac{4}{5} = 5\% + 8\% = 13\%$. Abbildung 10(b) zeigt den Zusammenhang im Entscheidungsbaum mit vertauschten Ereignissen A und B.

Da man den Entscheidungsbaum quasi „umdrehen“ kann, bietet es sich an, auch die bedingte Wahrscheinlichkeit „umzudrehen“, d.h. $P(B_i|A)$ aus $P(A|B_i)$ zu berechnen.

Satz 5.7 (Bayes).

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A)} = \frac{P(A|B_i)P(B_i)}{P(A|B_1)P(B_1) + \ldots + P(A|B_n)P(B_n)},$$

bzw. für $n = 2$

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

Beweis. Aus der bedingten Wahrscheinlichkeit ergibt sich

$$P(A \cap B_i) = P(A|B_i)P(B_i) = P(B_i|A)P(A).$$

Umgeformt wird das zu

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A)}.$$

Setzt man nun für $P(A)$ die vollständige Wahrscheinlichkeit, bekommt man die Aussage. □

Beispiel 5.8.

$$P(\text{„Defektes Auto ist Montagsauto“}) = P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|\bar{B})P(\bar{B})}$$
$$= \frac{25\%\frac{1}{5}}{25\%\frac{1}{5} + 10\%\frac{4}{5}} = \frac{5\%}{13\%} \approx 38.5\%.$$

Abbildung 10(b) zeigt auch diese Wahrscheinlichkeit im Entscheidungsbaum mit vertauschten Ereignissen A und B.

Zwei Ereignisse gelten als unabhängig, wenn man aus dem Wissen über das Eintreten des einen Ereignisses nichts über das Eintreten des anderen sagen kann, d.h. dass $P(A|B) = P(A)$ sein muss. Daraus folgt dann $P(A \cap B) = P(A|B)P(B) = P(A)P(B)$. Deshalb definiert man:

Definition 5.9.

$$A, B \text{ unabhängig } :\Leftrightarrow P(A \cap B) = P(A)P(B)$$

Beispiel 5.10. $P(A \cap B) = 5\%$ (siehe oben). $P(A)P(B) = 13\% \cdot \frac{1}{5} = 2.6\%$. $\Rightarrow A, B$ nicht unabhängig.

6 Zufallsvariablen

Meist ist man nicht unmittelbar am genauen Ausgang eines Experiments (dem Ereignis) interessiert, sondern nur an einem Merkmal, das einem Wert aus $\mathbb{Z}$ oder $\mathbb{R}$ entspricht. Es wird also jedem Elementarereignis $\omega \in \Omega$ ein Wert $X(\omega)$ zugewiesen.

Definition 6.1. Eine Funktion $X : \Omega \longrightarrow B$ heißt **Zufallsvariable**. B ist der Wertebereich der Zufallsvariable, also meist $\mathbb{Z}$ oder $\mathbb{R}$.

Definition 6.2. Das Ereignis, dass die Zufallsvariable X einen bestimmten Wert x aus dem Wertebereich annimmt, wird „$X = x$“ oder einfach $X = x$ geschrieben. Es beinhaltet alle Elementarereignisse, denen der Wert x zugeordnet ist:

$$\text{„}X = x\text{“} := \{\omega \mid X(\omega) = x\} = X^{-1}(x).$$

Analog werden die Ereignisse $X < x$, $X \leq x$, $X > x$ und $X \geq x$ definiert. Zum Beispiel ist

$$\text{„}X \leq x\text{“} := \{\omega \mid X(\omega) \leq x\} = X^{-1}((-\infty, x]).$$

Es reicht nun, die Wahrscheinlichkeiten aller Ereignisse $X = x$ oder besser $X \leq x$ zu kennen, um eine Zufallsvariable vollständig beschreiben zu können.

Definition 6.3. Die **Verteilungsfunktion** $F_X : \mathbb{R} \longrightarrow [0,1]$ ist

$$F_X(x) := P(X \leq x)\,.$$

Ist der Wertebereich der Zufallsvariablen diskret (also z.B. $\mathbb{N}$), dann spricht man von einer **diskreten Verteilung**. Die Verteilungsfunktion, als Funktion reeller x betrachtet, ist dann unstetig in den diskreten Werten und dazwischen konstant, also eine Treppenfunktion, die in $-\infty$ mit dem Wert 0 beginnt, in den diskreten Werten um $P(X = x)$ springt, und in $+\infty$ mit dem Wert 1 endet.

Ist die Verteilungsfunktion dagegen *absolut stetig*, spricht man von einer **stetigen Verteilung**. Es ist dann überall $P(X = x) = 0$. Die Verteilungsfunktion ist fast überall differenzierbar und kann als Integral einer Dichtefunktion dargestellt werden. Wir beschränken uns in der Folge auf diskrete und stetige Verteilungen mit $\mathbb{Z}$ und $\mathbb{R}$ als Wertebereiche.

Definition 6.4. Auf dem Wahrscheinlichkeitsraum soll ein Integral definiert sein, das folgendermaßen geschrieben wird

$$\int_A X(\omega)\,\mathrm{d}P(\omega)\,,$$

und den Inhalt unter der Funktion $X : \Omega \longrightarrow \mathbb{R}$ über der Teilmenge $A \subseteq \Omega$ ($A \in \Sigma$) berechnet, wobei die Teilbereiche von A entsprechend P gewichtet werden sollen. Das Integral soll folgende Kriterien erfüllen. Erstens soll das Integral von 1 dem Wahrscheinlichkeitsmaß entsprechen:

$$\int_A \mathrm{d}P(\omega) = P(A)\,.$$

Zweitens soll das Integral folgende Linearität besitzen:

$$\int_A aX(\omega) + bY(\omega)\,\mathrm{d}P(\omega) = a\int_A X(\omega)\,\mathrm{d}P(\omega) + b\int_A Y(\omega)\,\mathrm{d}P(\omega)\,,$$

wobei a, b konstant sind und $X, Y : \Omega \longrightarrow \mathbb{R}$ integrierbare Funktionen. Der Wertebereich von X, Y kann auch höherdimensional sein, also $\mathbb{R}^n$ oder $\mathbb{C}$. In dem Fall sollen die Koordinatenfunktionen die selben Kriterien erfüllen.

Definition 6.5. Der **Erwartungswert** einer Zufallsvariable X ist definiert durch

$$\mathrm{E}(X) := \int_\Omega X(\omega)\,\mathrm{d}P(\omega)\,.$$

Beachte, dass der Erwartungswert E keine einfache Funktion sondern ein sogenanntes *Funktional* ist, da es eine ganze Funktion, nämlich X, auf eine Zahl abbildet.

Satz 6.6. Aufgrund der Definition leicht ersichtlich ist

$$\mathrm{E}(aX+bY) = a\,\mathrm{E}(X) + b\,\mathrm{E}(Y)\,.$$

Definition 6.7. Varianz $\mathrm{V}(X)$ und **Standardabweichung** $\sigma_X := \sqrt{\mathrm{V}(X)}$.

$$\mathrm{V}(X) := \mathrm{E}((X-\mathrm{E}(X))^2)\,.$$

Satz 6.8. Auch hier gibt es einen Verschiebungssatz.

$$\mathrm{V}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2\,.$$

Beweis. $\mathrm{E}((X-\mathrm{E}(X))^2) = \mathrm{E}(X^2 - 2X\,\mathrm{E}(X) + \mathrm{E}(X)^2) = \mathrm{E}(X^2) - 2\,\mathrm{E}(X)\,\mathrm{E}(X) + \mathrm{E}(X)^2 = \mathrm{E}(X^2) - \mathrm{E}(X)^2$. □

Beispiel 6.9. Für ein Bernoulli-Experiment mit $X \in \{0,1\}$ und $P(X=1) = p$, $P(X=0) = 1-p$ ist

$$\begin{aligned}\mathrm{E}(X) &= \int_\Omega X(\omega)\,\mathrm{d}P(\omega) = \int_{X=0} X(\omega)\,\mathrm{d}P(\omega) + \int_{X=1} X(\omega)\,\mathrm{d}P(\omega)\\ &= \int_{X=0} 0\,\mathrm{d}P(\omega) + \int_{X=1} 1\,\mathrm{d}P(\omega) = 0\cdot\int_{X=0}\mathrm{d}P(\omega) + 1\cdot\int_{X=1}\mathrm{d}P(\omega)\\ &= 0\cdot P(X=0) + 1\cdot P(X=1) = 0\cdot(1-p) + 1\cdot p = p\,.\end{aligned}$$

$$\mathrm{V}(X) = \int_\Omega X^2(\omega)\,\mathrm{d}P(\omega) - \mathrm{E}(X)^2 = 0^2(1-p) + 1^2 p - p^2 = p(1-p)\,.$$

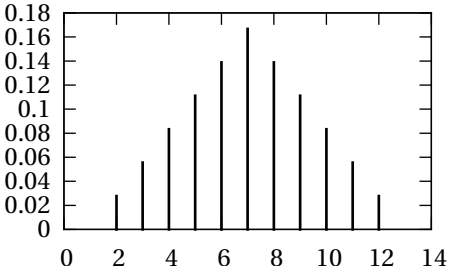
Satz 6.10. Für a, b konstant gilt $\mathrm{V}(aX+b) = a^2\,\mathrm{V}(X)$.

Beweis. $\mathrm{V}(aX+b) = \mathrm{E}((aX+b-\mathrm{E}(aX+b))^2) = \mathrm{E}((aX+b-a\,\mathrm{E}(X)-b)^2) = \mathrm{E}(a^2(X-\mathrm{E}(X))^2) = a^2\,\mathrm{V}(X)$. □

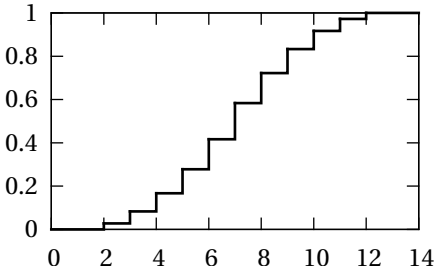
6.1 Diskrete Verteilungen

Definition 6.11. Die **Wahrscheinlichkeitsfunktion** $f_X : \mathbb{Z} \longrightarrow [0,1]$ einer diskreten Zufallsvariablen X gibt die Wahrscheinlichkeit des Werts k an und wird oft auch einfach p_k geschrieben:

$$f_X(k) := p_k := P(X=k)\,.$$



(a) Wahrscheinlichkeitsfunktion f_X



(b) Verteilungsfunktion F_X

Abbildung 11: Augensumme zweier Würfel als Beispiel für eine diskrete Zufallsvariable.

Satz 6.12. Die Verteilungsfunktion ist dann $F_X(x) = \sum_{k \le x} p_k$.

Beispiel 6.13. Wir betrachten die Augensumme bei zweimaligem Würfeln:

$$\Omega = \{1,\dots,6\}^2 = \{(a,b) \mid a,b \in \{1,\dots,6\}\}, \qquad X((a,b)) := a+b.$$

Dann erhalten wir z.B.

$$\begin{aligned}
p_2 &= P(X=2) = P(\{(1,1)\}) = \frac{|\{(1,1)\}|}{|\Omega|} = \frac{1}{6^2} = \frac{1}{36},\\
p_5 &= P(X=5) = P(\{(1,4),(2,3),(3,2),(4,1)\}) = \frac{4}{6^2} = \frac{1}{9},\\
p_{12} &= P(X=12) = P(\{(6,6)\}) = \frac{1}{36},\\
p_{13} &= P(X=13) = P(\emptyset) = 0,\\
F_X(4.3) &= P(X \le 4.3) = P(X \in \{2,3,4\}) = p_2 + p_3 + p_4\\
&= \frac{1}{36} + \frac{2}{36} + \frac{3}{36} = \frac{6}{36} = \frac{1}{6}.
\end{aligned}$$

Abbildung 11 zeigt die Wahrscheinlichkeits- und Verteilungsfunktion.

Der Erwartungswert $\mathrm{E}(X)$ entspricht in etwa dem Mittelwert von Stichproben. Zur Erinnerung (für $x_k = k$): $\bar{x} = \frac{1}{n}\sum_k kH_k = \sum_k kh_k$. Nun wird einfach h_k durch p_k ersetzt.

Satz 6.14.

$$\mathrm{E}(X) = \sum_{k=-\infty}^{\infty} k \cdot P(X=k) = \sum_{k=-\infty}^{\infty} k p_k,$$
$$\mathrm{V}(X) = \sum_{k=-\infty}^{\infty} (k-\mathrm{E}(X))^2 p_k = \sum_{k=-\infty}^{\infty} k^2 p_k - \mathrm{E}(X)^2.$$

Beweis. $\mathrm{E}(X) = \int_\Omega X(\omega)\,\mathrm{d}P(\omega) = \sum_k \int_{X=k} k\,\mathrm{d}P(\omega) = \sum_k k \cdot P(X=k)$. $\mathrm{V}(X)$ analog. □

Beispiel 6.15.

$$\mathrm{E}(X) = \sum_{k=2}^{12} k p_k = 2\frac{1}{36} + 3\frac{2}{36} + \ldots + 12\frac{1}{36} = \frac{252}{36} = 7,$$
$$\mathrm{V}(X) = (2-7)^2\frac{1}{36} + (3-7)^2\frac{2}{36} + \ldots + (12-7)^2\frac{1}{36} = 2^2\frac{1}{36} + \ldots + 12^2\frac{1}{36} - 7^2 \approx 5.83,$$
$$\sigma_X \approx 2.42.$$

Definition 6.16. Die Zufallsvariable X besitzt die diskrete **Gleichverteilung** innerhalb der Grenzen a und b, wenn gilt

$$p_k = \begin{cases} \frac{1}{b-a+1} & a \le k \le b \\ 0 & \text{sonst.} \end{cases}$$

Satz 6.17. Ist X gleichverteilt zwischen a und b, dann gilt

$$\mathrm{E}(X) = \frac{a+b}{2}, \qquad \mathrm{V}(X) = \frac{(b-a+1)^2-1}{12}.$$

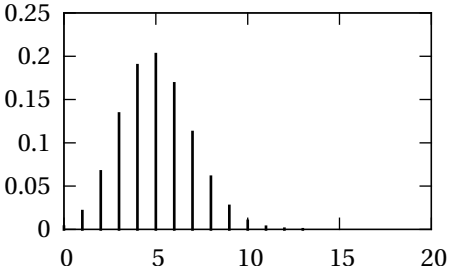
Beweis.

$$\mathrm{E}(X) = \sum_{k=a}^{b} \frac{k}{b-a+1} = \frac{1}{b-a+1} \cdot \frac{1}{2}\left(\sum_a^b k + \sum_a^b a+b-k\right) = \frac{1}{b-a+1} \cdot \frac{1}{2}\sum_a^b a+b$$
$$= \frac{b-a+1}{b-a+1} \cdot \frac{a+b}{2}.$$

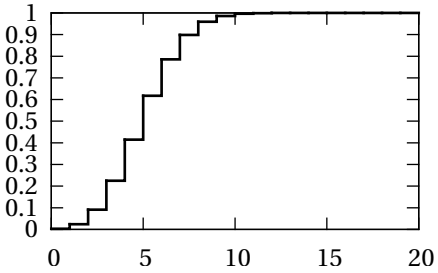
Und mit $\sum_{k=0}^n k^2 = n(n+1)(2n+1)/6$ erhält man

$$\mathrm{V}(X) = \mathrm{V}(X-a) = \frac{1}{b-a+1}\sum_{k=0}^{b-a} k^2 - \mathrm{E}(X-a)^2$$
$$= \frac{(b-a)(b-a+1)(2(b-a)+1)}{6(b-a+1)} - \left(\frac{a+b}{2}-a\right)^2 = \frac{(b-a)(2(b-a)+1)}{6} - \frac{(b-a)^2}{4}$$
$$= \frac{(b-a)(4(b-a)+2-3(b-a))}{12} = \frac{(b-a)^2+2(b-a)+1-1}{12} = \frac{(b-a+1)^2-1}{12}.$$

□



(a) Wahrscheinlichkeitsfunktion f_X



(b) Verteilungsfunktion F_X

Abbildung 12: Binomialverteilung $B(20, 25\%)$.

Definition 6.18. Eine Zufallsvariable X besitzt die **Bernoulliverteilung**, wenn gilt

$$f_X(1) = p\,, \qquad f_X(0) = 1 - p\,, \qquad f_X(k) = 0 \text{ sonst.}$$

Satz 6.19.

$$\mathrm{E}(X) = p\,, \qquad \mathrm{V}(X) = p(1-p)\,.$$

Beweis. $\mathrm{E}(X) = 1 \cdot p + 0 \cdot (1-p) = p$. $\mathrm{V}(X) = 1^2 p + 0^2(1-p) - \mathrm{E}(X)^2 = p - p^2 = p(1-p)$. □

Wird ein Bernoulli-Experiment mit zwei möglichen Ausgängen, erfolgreich bzw. nicht erfolgreich, n-mal wiederholt, wobei das Experiment mit Wahrscheinlichkeit p erfolgreich ist, dann stellt sich die Frage, wie oft das Experiment erfolgreich ist. Die Antwort gibt folgende Verteilung.

Definition 6.20. Eine Zufallsvariable X besitzt die **Binomialverteilung**, wenn gilt

$$f_X(k) = \binom{n}{k} p^k (1-p)^{n-k}\,.$$

Man schreibt $X \sim B(n, p)$.

Satz 6.21. Zur Plausibilität der Definition ist zu zeigen, dass $\sum_k f_X(k) = 1$ ist.

Beweis. $1 = 1^n = (p + (1-p))^n = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k}$. □

Abbildung 12 zeigt Wahrscheinlichkeits- und Verteilungsfunktion einer Binomialverteilung.

Satz 6.22. Es werden n unabhängige Experimente mit Ausgang $e_i \in \{0,1\}$ durchgeführt, wobei das Ereignis A_{i,e_i} =„Experiment i ergibt e_i“ für den erfolgreichen Ausgang $e_i = 1$ die Wahrscheinlichkeit $P(A_{i,1}) = p$ hat. Es sei $e = (e_1, \ldots, e_n)$ ein Ausgang des Gesamtexperiments, $A_e = A_{1,e_1} \cap \ldots \cap A_{n,e_n}$ das Ereignis, dass das Gesamtexperiment diesen Ausgang hat, und $X(\omega) := \sum_i e_i$ für $\omega \in A_e$ die Anzahl der erfolgreichen Ausgänge. Dann ist X binomialverteilt: $X \sim B(n,p)$.

Beweis. Die Wahrscheinlichkeit für „Exp. nicht erfolgreich“ ist $P(A_{i,0}) = P(\bar{A}_{i,1}) = 1-p$. Wegen der Unabhängigkeit der Experimente gilt $P(A_{i,e_i} \cap A_{j,e_j}) = P(A_{i,e_i}) \cdot P(A_{j,e_j})$ für $i \neq j$. Wenn nun $\sum e_i = k$ ist, dann ist daher $P(A_e) = p^k(1-p)^k$. Die Anzahl der möglichen Gesamtexperimente mit k Erfolgen aus n Experimenten ist eine Kombination ohne Wiederholung. Ereignisse für verschiedene Ausgänge schließen sich gegenseitig aus. Daher ist schließlich

$$P(X=k) = P\Big(\bigcup_{\sum e_i = k} A_e\Big) = \sum_{\sum e_i = k} P(A_e) = \binom{n}{k} p^k (1-p)^{n-k}. \qquad \square$$

Satz 6.23. Für $X \sim B(n,p)$ gilt

$$\mathrm{E}(X) = np, \qquad \mathrm{V}(X) = np(1-p).$$

Beweis. X kann als Zusammensetzung $X = X_1 + \ldots + X_n$ von n einzelnen unabhängigen Bernoulli-verteilten Zufallsvariablen X_i dargestellt werden. Dann ist

$$\mathrm{E}(X) = \mathrm{E}(X_1 + \ldots + X_n) = n\,\mathrm{E}(X_1) = np.$$

$$\begin{aligned}\mathrm{V}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2 &= \mathrm{E}((X_1 + \ldots + X_n)^2) - (np)^2 \\ &= \mathrm{E}\Big(\sum_{i,j} X_i X_j\Big) - (np)^2 = \sum_{i,j} \mathrm{E}(X_i X_j) - n^2 p^2.\end{aligned}$$

Nun ist für $i = j$: $\mathrm{E}(X_i^2) = \mathrm{E}(X_i) = p$, weil $0^2 = 0$ und $1^2 = 1$. Für $i \neq j$ ist $\mathrm{E}(X_i X_j) = 0 \cdot 0 \cdot (1-p)^2 + 1 \cdot 0 \cdot p(1-p) + 0 \cdot 1 \cdot (1-p)p + 1 \cdot 1 \cdot p^2 = p^2$. Es gibt n Fälle für $i = j$ und $n^2 - n$ Fälle für $i \neq j$. Daher ist

$$\mathrm{V}(X) = np + (n^2 - n)p^2 - n^2 p^2 = np - np^2 = np(1-p). \qquad \square$$

Beispiel 6.24. Ein Bauteil habe mit einer Wahrscheinlichkeit $p = 0.3$ einen Defekt. In einer Schachtel seien 10 Bauteile. P(„genau 4 Bauteile defekt“) $= P(X = 4) = \binom{10}{4} 0.3^4 0.7^6 \approx 0.2$. P(„maximal 4 Bauteile defekt“) $= p_0 + p_1 + \ldots + p_4 \approx 0.03 + 0.12 + \ldots + 0.2 \approx 0.85$. In einer Schachtel sind durchschnittlich $\mathrm{E}(X) = 10 \cdot 0.3 = 3$ Bauteile defekt mit einer Standardabweichung von $\sigma_X = \sqrt{10 \cdot 0.3 \cdot 0.7} \approx 1.45$.

Wenn man statt der Anzahl der erfolgreichen Versuche misst, wie viele Versuche man benötigt, bis Erfolg eintritt, wobei die Anzahl der gesamten Versuche nicht beschränkt ist, dann erhält man folgende Verteilung.

Definition 6.25. Eine Zufallsvariable besitzt die **geometrische Verteilung**, wenn gilt

$$f_X(k) = p(1-p)^{k-1}.$$

Man schreibt $X \sim G(p)$.

Satz 6.26. Zur Plausibilität zeigen wir wieder $\sum_k f_X(k) = 1$.

Beweis. $S = \sum_{k=1}^{\infty} p(1-p)^{k-1} = p + \sum_{k=2}^{\infty} p(1-p)^{k-1} = p + (1-p)\sum_{k=1}^{\infty} p(1-p)^{k-1} = p + (1-p)S \Rightarrow S(1-(1-p)) = p \Rightarrow Sp = p \Rightarrow S = 1.$ □

Satz 6.27. Seien $e = (e_1, e_2, \ldots)$ die Ausgänge von unendlich vielen Versuchen, A_{i,e_i} das Ereignis „Versuch i ergibt e_i“, $P(A_{i,1}) := p$, A_e das Gesamtereignis für alle Ausgänge, und für $\omega \in A_e$ sei $X(\omega) := k$ so dass $e_k = 1$ und $e_j = 0$ für alle $j < k$. Dann ist $X \sim G(p)$.

Beweis. $P(X = k) = P(A_{1,0} \cap A_{2,0} \cap \ldots \cap A_{k,1}) = (1-p)^{k-1}p.$ □

Satz 6.28. $X \sim G(p) \Rightarrow F_X(m) = 1 - (1-p)^m$.

Beweis. $F_X(m) = \sum_{k=1}^{m} p(1-p)^{k-1} = 1 - \sum_{k=m+1}^{\infty} p(1-p)^{k-1} = 1 - (1-p)^m \sum_{k=1}^{\infty} p(1-p)^{k-1} = 1 - (1-p)^m.$ □

Satz 6.29. $X \sim G(p) \Rightarrow$

$$\mathrm{E}(X) = \frac{1}{p}, \qquad \mathrm{V}(X) = \frac{1-p}{p^2}.$$

Beweis.

$$\mathrm{E}(X) = \sum_{k=1}^{\infty} kp(1-p)^{k-1} = p + \sum_{k=2}^{\infty} kp(1-p)^{k-1} = p + (1-p)\sum_{k=1}^{\infty}(k+1)p(1-p)^{k-1}$$

$$= p + (1-p)\left(\sum_{k=1}^{\infty} kp(1-p)^{k-1} + \sum_{k=1}^{\infty} p(1-p)^{k-1}\right) = p + (1-p)(\mathrm{E}(X)+1)$$

$$\Rightarrow \quad \mathrm{E}(X)(1-(1-p)) = p + (1-p) = 1 \quad \Rightarrow \quad \mathrm{E}(X) = \frac{1}{p}.$$

$$\begin{aligned}\mathrm{E}(X^2) &= \sum_{k=1}^{\infty} k^2 p(1-p)^{k-1} = p + (1-p)\sum_{k=1}^{\infty}(k+1)^2 p(1-p)^{k-1} \\ &= p + (1-p)\left(\sum_{k=1}^{\infty} k^2 p(1-p)^{k-1} + 2\sum_{k=1}^{\infty} kp(1-p)^{k-1} + \sum_{k=1}^{\infty} p(1-p)^{k-1}\right) \\ &= p + (1-p)(\mathrm{E}(X^2) + 2\,\mathrm{E}(X) + 1)\end{aligned}$$

$$\Rightarrow \quad \mathrm{E}(X^2)(1-(1-p)) = p + (1-p)\left(\frac{2}{p}+1\right) = \frac{2}{p} - 1 \quad \Rightarrow \quad \mathrm{E}(X^2) = \frac{2}{p^2} - \frac{1}{p}$$

$$\Rightarrow \quad \mathrm{V}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2 = \frac{2}{p^2} - \frac{1}{p} - \frac{1}{p^2} = \frac{1}{p^2} - \frac{1}{p} = \frac{1-p}{p^2}. \quad \square$$

Wenn man den Parameter n der Binomialverteilung erhöht, dann wandert die ganze Verteilung, die sich ja um den Erwartungswert np konzentriert, nach rechts. Möchte man den Erwartungswert konstant auf $\lambda = np$ halten, dann muss man p entsprechend kleiner machen. Dann kann man n sogar gegen unendlich gehen lassen und erhält folgende Verteilung.

Definition 6.30. Eine Zufallsvariable X besitzt die **Poisson-Verteilung**, wenn gilt

$$f_X(k) = \frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}.$$

Man schreibt $X \sim P(\lambda)$.

Satz 6.31. Plausibilität: $\sum_k f_X(k) = 1$.

Beweis. $\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}\mathrm{e}^{-\lambda} = \mathrm{e}^{\lambda}\mathrm{e}^{-\lambda} = 1.$ $\square$

Satz 6.32. Wenn man in der Wahrscheinlichkeitsfunktion der Binomialverteilung p durch $\frac{\lambda}{n}$ ersetzt, dann konvergiert sie für $n \to \infty$ gegen die der Poisson-Verteilung.

$$\lim_{n\to\infty} \binom{n}{k}\left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k} = \frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}$$

Beweis. Der linke Ausdruck wird zu

$$\lim_{n\to\infty} \underbrace{\frac{n}{n}}_{=1}\underbrace{\frac{n-1}{n}}_{\to 1}\cdots\underbrace{\frac{n-k+1}{n}}_{\to 1}\frac{\lambda^k}{k!}\underbrace{\left(1-\frac{\lambda}{n}\right)^n}_{\to \mathrm{e}^{-\lambda}}\underbrace{\left(1-\frac{\lambda}{n}\right)^{-k}}_{\to 1},$$

was gegen die Poisson-Verteilung konvergiert. $\square$

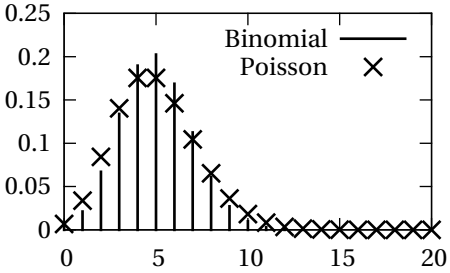


Abbildung 13: Vergleich von $B(20, \frac{1}{4})$ und $P(20 \cdot \frac{1}{4}) = P(5)$.

Abbildung 13 zeigt den Vergleich zwischen Binomial- und Poisson-Verteilung. Die Approximation ist besser, je höher n und je kleiner p ist.

Satz 6.33. $X \sim P(\lambda) \Rightarrow$

$$\mathrm{E}(X) = \lambda\,, \qquad \mathrm{V}(X) = \lambda\,.$$

Beweis.

$$\mathrm{E}(X) = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} \mathrm{e}^{-\lambda} = \mathrm{e}^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} = \mathrm{e}^{-\lambda} \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \mathrm{e}^{-\lambda} \lambda \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \mathrm{e}^{-\lambda} \lambda \mathrm{e}^{\lambda} = \lambda\,.$$

$$\begin{aligned}\mathrm{E}(X^2) &= \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} \mathrm{e}^{-\lambda} = \mathrm{e}^{-\lambda} \lambda \sum_{k=1}^{\infty} k \frac{\lambda^{k-1}}{(k-1)!} = \mathrm{e}^{-\lambda} \lambda \sum_{k=0}^{\infty} (k+1) \frac{\lambda^k}{k!} \\ &= \mathrm{e}^{-\lambda} \lambda \left(\sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} + \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \right) = \mathrm{e}^{-\lambda} \lambda (\mathrm{e}^{\lambda} \lambda + \mathrm{e}^{\lambda}) = \lambda^2 + \lambda\,.\end{aligned}$$

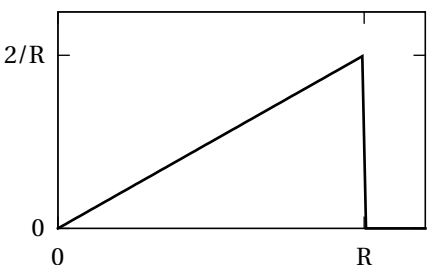
$$\mathrm{V}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda\,. \qquad \square$$

6.2 Stetige Verteilungen

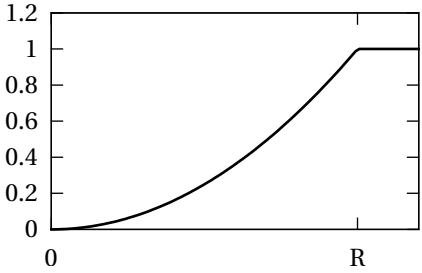
Die Rolle der Wahrscheinlichkeitsfunktion übernimmt bei stetigen Verteilungen die Dichtefunktion.

Definition 6.34. Die **Dichtefunktion** oder **Dichte** einer stetigen Zufallsvariable X ist eine Funktion $f_X : \mathbb{R} \longrightarrow \mathbb{R}$, deren Integral die Verteilungsfunktion ergibt:

$$\int_{-\infty}^{x} f_X(u)\,\mathrm{d}u = F_X(x)\,.$$



(a) Dichtefunktion f_X



(b) Verteilungsfunktion F_X

Abbildung 14: Mittelpunktsabstand als Beispiel für eine stetige Zufallsvariable.

Die Dichtefunktion ist nicht-negativ, weil sonst die Verteilungsfunktion nicht monoton steigend wäre und es negative Wahrscheinlichkeiten gäbe. Wo F differenzierbar ist, ist natürlich $f_X(x) = F'_X(x)$.

Satz 6.35.

$$P(a \le X \le b) = \int_a^b f_X(x)\,\mathrm{d}x = F_X(b) - F_X(a)\,.$$

Satz 6.36.

$$1 = P(-\infty \le X \le +\infty) = \int_{-\infty}^{+\infty} f_X(x)\,\mathrm{d}x = F_X(\infty) - F_X(-\infty) = 1 - 0 = 1\,.$$

Beispiel 6.37. Eine runde Schachtel mit Durchmesser $2R$ und einer kleinen Kugel drin wird geschüttelt. Danach befindet sich die Kugel an der Position (a,b) ($(0,0)$ = Mitte). $\Omega = \{(a,b) \in \mathbb{R}^2 \mid a^2 + b^2 \le R^2\}$. Für eine exakte Position ist natürlich z.B. $P(\{(0.1,0.2)\}) = 0$. *Aber:* Die Wahrscheinlichkeit, dass die Kugel in einem bestimmten Bereich B mit Fläche $I(B)$ liegt, ist $P(B) = \frac{I(B)}{I(\Omega)} = \frac{I(B)}{R^2\pi}$ (Laplace-Annahme). X sei nun der Abstand von der Mitte: $X((a,b)) := \sqrt{a^2+b^2}$. Auch die Wahrscheinlichkeit für einen exakten Abstand r ist $P(X = r) = 0$. Das Ereignis „$X \le r$“ entspricht der Kreisscheibe mit Radius r und Fläche $r^2\pi$. Daraus ergibt sich:

$$F_X(r) = \begin{cases} 0 & r < 0 \\ \frac{r^2\pi}{R^2\pi} = \frac{r^2}{R^2} & 0 \le r \le R \\ 1 & r > R \end{cases}, \qquad f_X(r) = \begin{cases} 0 & r < 0 \\ \frac{\mathrm{d}}{\mathrm{d}r}\frac{r^2}{R^2} = \frac{2r}{R^2} & 0 \le r < R \\ 0 & r > R \end{cases}.$$

Abbildung 14 zeigt die Dichte- und Verteilungsfunktion.

Satz 6.38. Wenn eine Zufallsvariable X eine Dichtefunktion f_X hat und $g : \mathbb{R} \longrightarrow \mathbb{R}$ eine integrierbare Funktion ist, dann gilt

$$\int_{a \le X \le b} g(X(\omega))\,\mathrm{d}P(\omega) = \int_a^b g(x) f_X(x)\,\mathrm{d}x.$$

Das gilt auch mehrdimensional. Wenn $X : \Omega \longrightarrow \mathbb{R}^2$ eine zweidimensionale Zufallsvariable ist und eine Dichtefunktion $f_X : \mathbb{R}^2 \longrightarrow \mathbb{R}$ hat, das heißt $\int_{-\infty}^{a}\int_{-\infty}^{b} f_X(x_1,x_2)\,\mathrm{d}x_2\,\mathrm{d}x_1 = P(X_1 \le a \wedge X_2 \le b)$, weiters $g : \mathbb{R}^2 \longrightarrow \mathbb{R}$ integrierbar, dann gilt

$$\int_{X \in [a,b]\times[c,d]} g(X(\omega))\,\mathrm{d}P(\omega) = \int_a^b \int_c^d g(x_1,x_2) f_X(x_1,x_2)\,\mathrm{d}x_2\,\mathrm{d}x_1.$$

Beweis. Das ist ein wichtiger Satz aus der Maßtheorie. Siehe also dort. □

Satz 6.39. Erwartungswert $\mathrm{E}(X)$, Varianz $\mathrm{V}(X)$ und Standardabweichung $\sigma_X = \sqrt{\mathrm{V}(X)}$ werden hier über Integrale (statt Summen) berechnet.

$$\begin{aligned}
\mathrm{E}(X) &= \int_{-\infty}^{+\infty} x f_X(x)\,\mathrm{d}x \\
\mathrm{V}(X) &= \mathrm{E}((X-\mathrm{E}(X))^2) = \int_{-\infty}^{+\infty} (x-\mathrm{E}(X))^2 f_X(x)\,\mathrm{d}x \\
&= \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = \int_{-\infty}^{+\infty} x^2 f_X(x)\,\mathrm{d}x - (\mathrm{E}(X))^2.
\end{aligned}$$

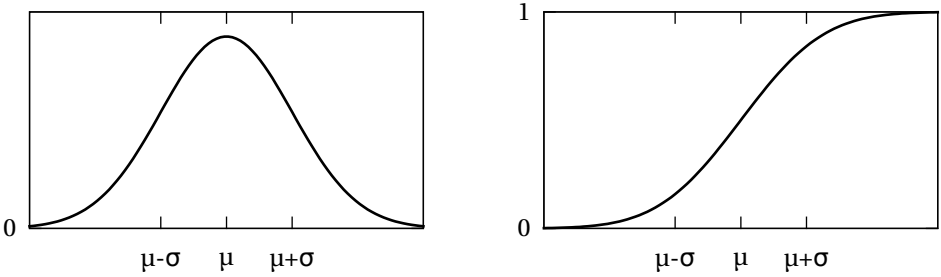
Beweis. Ergibt sich aus Satz 6.38. □

Beispiel 6.40.

$$\begin{aligned}
\mathrm{E}(X) &= \int_0^R r\frac{2r}{R^2}dr = \frac{2}{R^2}\frac{r^3}{3}\bigg|_0^R = \frac{2R}{3}, \\
\mathrm{V}(X) &= \int_0^R r^2\frac{2r}{R^2}dr - \left(\frac{2R}{3}\right)^2 = \frac{2}{R^2}\frac{r^4}{4}\bigg|_0^R - \frac{4R^2}{9} = \frac{R^2}{2} - \frac{4R^2}{9} = \frac{R^2}{18}, \\
\sigma_X &= \frac{R}{\sqrt{18}} \approx 0.236R.
\end{aligned}$$

Definition 6.41. Eine Zufallsvariable X besitzt die stetige **Gleichverteilung** zwischen a und b, wenn gilt

$$f_X(x) = \begin{cases} \frac{1}{b-a} & a \le x \le b \\ 0 & \text{sonst.} \end{cases}$$



(a) Dichtefunktion (b) Verteilungsfunktion

Abbildung 15: Normalverteilung.

Satz 6.42.

$$\mathrm{E}(X) = \frac{a+b}{2}, \qquad \mathrm{V}(X) = \frac{(b-a)^2}{12}.$$

Beweis.

$$\mathrm{E}(X) = \int_a^b \frac{x}{b-a}\,\mathrm{d}x = \frac{x^2}{2(b-a)}\bigg|_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}.$$

Wir setzen $g = \frac{b-a}{2}$, dann ist $a - \mathrm{E}(X) = -g$ und $b - \mathrm{E}(X) = g$ und

$$\mathrm{V}(X) = \mathrm{V}(X - \mathrm{E}(X)) = \int_{-g}^{g} \frac{x^2}{b-a}\,\mathrm{d}x = \frac{x^3}{3(b-a)}\bigg|_{-g}^{g} = \frac{g^3-(-g^3)}{3(b-a)} = \frac{2g^3}{3(b-a)}$$
$$= \frac{2(b-a)^3}{8\cdot 3(b-a)} = \frac{(b-a)^2}{12}. \qquad \square$$

Definition 6.43. Eine Zufallsvariable X besitzt die **Normalverteilung** mit Erwartungswert μ und Standardabweichung σ, wenn gilt

$$f_X(x) := \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Man schreibt $X \sim N(\mu, \sigma^2)$.

Diese Dichtefunktion ist leider nicht in geschlossener Form integrierbar, daher wird die Verteilungsfunktion meist in Tabellenform angegeben wie in Abbildung 25 oder über Näherungsformeln berechnet. Abbildung 15 zeigt beide Funktionen.

Satz 6.44. Das Integral der Dichtefunktion über ganz $\mathbb{R}$ ist 1:

$$\int_{-\infty}^{\infty} f_X(x)\,\mathrm{d}x = 1\,.$$

Beweis. Da das Integral nicht von μ abhängt, setzen wir $\mu = 0$. Die Beweisstrategie geht auf Poisson zurück. Statt des einfachen Integrals wird das Quadrat berechnet:

$$\left(\int_{-\infty}^{\infty} f_X(x)\,\mathrm{d}x\right)^2 = \frac{1}{2\pi\sigma^2}\left(\int_{-\infty}^{\infty} \mathrm{e}^{-\frac{x_1^2}{2\sigma^2}}\,\mathrm{d}x_1 \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{x_2^2}{2\sigma^2}}\,\mathrm{d}x_2\right)$$
$$= \frac{1}{2\pi\sigma^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{e}^{-\frac{x_1^2+x_2^2}{2\sigma^2}}\,\mathrm{d}x_1\,\mathrm{d}x_2\,.$$

Jetzt geht man von kartesischen zu polaren Koordinaten über: $x_1 = r\cos\varphi$, $x_2 = r\sin\varphi$. Dadurch verändert sich der Integrationsbereich auf $(\varphi, r) \in [0, 2\pi] \times [0, \infty)$ und der Integrand muss mit der Determinante der Jacobi-Matrix multipliziert werden.

$$\left|\frac{\partial(x_1, x_2)}{\partial(r, \varphi)}\right| = \begin{vmatrix}\cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi\end{vmatrix} = r(\cos^2\varphi + \sin^2\varphi) = r\,.$$

Außerdem ist $x_1^2 + x_2^2 = r^2$. Wenn wir nun noch sehen, dass $\frac{\mathrm{d}}{\mathrm{d}r} - \sigma^2\mathrm{e}^{-r^2/2\sigma^2} = r\mathrm{e}^{-r^2/2\sigma^2}$ ist, dann ergibt sich

$$= \frac{1}{2\pi\sigma^2}\int_0^{\infty}\int_0^{2\pi} \mathrm{e}^{-\frac{r^2}{2\sigma^2}}\,r\,\mathrm{d}\varphi\,\mathrm{d}r = \frac{1}{\sigma^2}\int_0^{\infty} r\mathrm{e}^{-\frac{r^2}{2\sigma^2}}\,\mathrm{d}r = \frac{-\sigma^2}{\sigma^2}\,\mathrm{e}^{-\frac{r^2}{2\sigma^2}}\Big|_0^{\infty} = -(0-1) = 1\,. \quad \square$$

Satz 6.45. $X \sim N(\mu, \sigma^2) \Rightarrow$

$$\mathrm{E}(X) = \mu\,, \qquad \mathrm{V}(X) = \sigma^2\,.$$

Beweis. Zuerst differenzieren wir $f_X(x)$:

$$f_X'(x) = \frac{\mathrm{d}}{\mathrm{d}x}\frac{1}{\sqrt{2\pi\sigma^2}}\mathrm{e}^{-\frac{x-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi\sigma^2}}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}\left(-\frac{2(x-\mu)}{2\sigma^2}\right) = -\frac{x-\mu}{\sigma^2}f_X(x)\,.$$

Daraus ergibt sich für $x f_X(x)$, welches wir in den Integralen brauchen werden:

$$x f_X(x) = \mu f_X(x) - \sigma^2 f_X'(x)\,.$$

Und damit ist

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} \mu f_X(x)\,\mathrm{d}x - \sigma^2 f_X(x)\big|_{-\infty}^{\infty} = \mu - (0-0) = \mu\,.$$

Für die Varianz können wir wieder $\mu = 0$ setzen. Mit Hilfe der partiellen Integration ($\int uv' = uv - \int u'v$) und $u = x$ und $v = -\sigma^2 f_X(x)$, erhalten wir

$$\mathrm{V}(X) = \int_{-\infty}^{\infty} \underbrace{x}_{u} \underbrace{x f_X(x)}_{v'} \,\mathrm{d}x = -x\sigma^2 f_X(x)\big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} -\sigma^2 f_X(x)\,\mathrm{d}x = 0 + \sigma^2 \cdot 1 = \sigma^2.$$

□

Definition 6.46. Die **Standardnormalverteilung** ist die Normalverteilung mit $\mu = 0$ und $\sigma = 1$, also $N(0,1)$. Die Verteilungsfunktion der Standardnormalverteilung wird als Φ geschrieben.

Satz 6.47. Wenn $X \sim N(\mu, \sigma^2)$, dann ist $\frac{X-\mu}{\sigma} \sim N(0,1)$. Es gilt daher

$$P(a \le X \le b) = \int_a^b f_X(x)\,\mathrm{d}x = F_X(b) - F_X(a) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

Beweis. $P(X \le x) = P\left(\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right)$. □

Aus Symmetriegründen ist $\Phi(-x) = 1 - \Phi(x)$.

Satz 6.48. $X \sim N(\mu, \sigma^2) \Rightarrow P(|X-\mu| \le \sigma) \approx 68.2\%$, $P(|X-\mu| \le 2\sigma) \approx 95.4\%$, $P(|X-\mu| \le 3\sigma) \approx 99.7\%$.

Beispiel 6.49. X sei die Füllmenge in einer Milchpackung in l. $X \sim N(1, 0.05^2)$. $P(X \le 0.98) = \Phi\left(\frac{0.98-1}{0.05}\right) = \Phi(-0.4) \approx 34\%$.

Die Umkehrfunktion Φ^{-1} ist wichtig, wenn Mindestwahrscheinlichkeiten gefragt sind.

Satz 6.50. $P(X \le x) \ge p \Leftrightarrow \Phi\left(\frac{x-\mu}{\sigma}\right) \ge p \Leftrightarrow \frac{x-\mu}{\sigma} \ge \Phi^{-1}(p) \Leftrightarrow x \ge \sigma\Phi^{-1}(p) + \mu$.

Beispiel 6.51. Wie viel Milch ist mit Wahrscheinlichkeit $p \ge 99\%$ in der Packung? $P(X > x) \ge p \Leftrightarrow 1 - \Phi\left(\frac{x-\mu}{\sigma}\right) \ge p \Leftrightarrow \frac{x-\mu}{\sigma} \le \Phi^{-1}(1-p) \Leftrightarrow x \le \sigma\Phi^{-1}(1-p) + \mu = 0.05\Phi^{-1}(0.01) + 1 \approx 0.88$l.

Satz 6.52. Es ist möglich, die Binomialverteilung durch die Normalverteilung zu approximieren (**Normalapproximation**). Sei $X \sim B(n,p)$ und $Y \sim N(\mathrm{E}(X), \mathrm{V}(X)) = N(np, np(1-p))$. Dann ist $P(X \le k) \approx P(Y \le k) = \Phi\left(\frac{k-np}{\sqrt{np(1-p)}}\right)$. Faustregel: es sollte $np(1-p) \ge 9$ sein, sonst ist der Fehler zu groß. Da gerade für ganzzahlige Stellen k die Treppenfunktion der Binomial-Verteilung maximal von der Approximation abweicht, wählt man oft besser $P(X \le k) \approx P(Y \le k + \frac{1}{2})$ (Stetigkeitskorrektur).

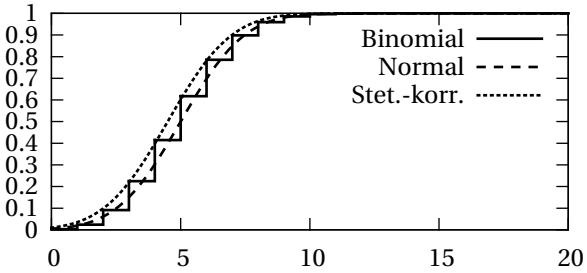


Abbildung 16: Approximation der Binomialverteilung $B(20,25\%)$ durch die Normalverteilung (Normalapproximation).

Der Grund für die Ähnlichkeit der Verteilungen ist der zentrale Grenzwertsatz (siehe Abschnitt 7). Abbildung 16 zeigt die Normalapproximationen.

Beispiel 6.53. Gegeben ist eine Packung mit 1000 Bauteilen. Jedes Bauteil ist mit $p = 0.02$ defekt. $P(\text{„max. 25 kaputt“}) = P(X \le 25) = \sum_{k=0}^{25} \binom{1000}{k} 0.02^k 0.98^{1000-k} \approx \Phi\left(\frac{25-1000\cdot 0.02}{\sqrt{1000\cdot 0.02\cdot 0.98}}\right) = \Phi(1.129) \approx 87\%$.

Definition 6.54. Eine Zufallsvariable besitzt die **Exponentialverteilung** mit Parameter $\lambda > 0$, wenn gilt

$$f_X(x) = \begin{cases} \lambda \mathrm{e}^{-\lambda x} & x \ge 0 \\ 0 & x < 0\,. \end{cases}$$

Satz 6.55.

$$F_X(x) = \int_0^x \lambda \mathrm{e}^{-\lambda u}\,\mathrm{d}u = -\mathrm{e}^{-\lambda u}\Big|_0^x = 1 - \mathrm{e}^{-\lambda x}\,.$$

Satz 6.56.

$$\mathrm{E}(X) = \frac{1}{\lambda}\,, \qquad \mathrm{V}(X) = \frac{1}{\lambda^2}\,.$$

Beweis.

$$\mathrm{E}(X) = \int_0^\infty \underbrace{x}_{u}\,\underbrace{\lambda \mathrm{e}^{-\lambda x}}_{v'}\,\mathrm{d}x = -\underbrace{x}_{u}\,\underbrace{\mathrm{e}^{-\lambda x}}_{v}\Big|_0^\infty + \int_0^\infty \underbrace{1}_{u'}\,\underbrace{\mathrm{e}^{-\lambda x}}_{v}\,\mathrm{d}x = 0 + -\frac{\mathrm{e}^{-\lambda x}}{\lambda}\Big|_0^\infty = \frac{1}{\lambda}\,.$$

$$\mathrm{V}(X) = \int_0^\infty \underbrace{x^2}_{u}\,\underbrace{\lambda \mathrm{e}^{-\lambda x}}_{v'}\,\mathrm{d}x - \mathrm{E}(X)^2 = -x^2\mathrm{e}^{-\lambda x}\Big|_0^\infty + \int_0^\infty 2x\mathrm{e}^{-\lambda x}\,\mathrm{d}x - \frac{1}{\lambda^2}$$

$$= 0 + \frac{2}{\lambda}\,\mathrm{E}(X) - \frac{1}{\lambda^2} = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}\,. \qquad \square$$

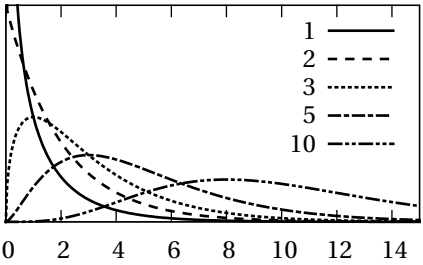


Abbildung 17: Dichtefunktion der χ^2-Verteilung für verschiedene Freiheitsgrade.

Die folgenden Verteilungen sind für die schließende Statistik wichtig. Berechnet man zum Beispiel die Varianz einer Stichprobe und modelliert die einzelnen Stichprobenwerte als normalverteilte Zufallsvariablen, dann ergibt sich eine Summe von quadrierten normalverteilten Zufallsvariablen. Die folgende Verteilung modelliert diese Situation.

Definition 6.57. Seien $Y_1, Y_2, \ldots, Y_n$ unabhängige, standardnormalverteilte Zufallsvariablen. Dann besitzt $X = Y_1^2 + Y_2^2 + \ldots + Y_n^2$ die χ^2-Verteilung („Chi-Quadrat“) mit n Freiheitsgraden. Man schreibt $X \sim \chi^2(n)$.

Satz 6.58.

$$f_X(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} \mathrm{e}^{-\frac{x}{2}} & x \geq 0 \\ 0 & x < 0, \end{cases} \qquad \text{mit} \qquad \Gamma(x) = \int_0^\infty t^{x-1}\mathrm{e}^{-t}\,\mathrm{d}t.$$

Diese Funktionen sind schwer handhabbar, daher werden meistens Näherungsformeln oder Tabellen verwendet wie die in Abbildung 27. Die Verteilungsfunktion ist überhaupt nicht elementar darstellbar. Abbildung 17 zeigt die Dichtefunktion.

Satz 6.59. $X \sim \chi^2(n) \Rightarrow \mathrm{E}(X) = n, \mathrm{V}(X) = 2n.$

Betrachtet man den Mittelwert einer Stichprobe und modelliert die Stichprobenwerte als normalverteilte Zufallsvariablen, dann hängt die Verteilung des Mittelwerts von der Varianz der Zufallsvariablen ab. Ist diese aber nicht bekannt sondern muss aus der Stichprobe konstruiert werden, dann hat der Mittelwert eine leicht andere Verteilung.

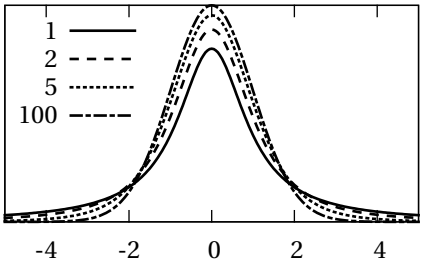


Abbildung 18: Dichtefunktion der Student-t-Verteilung für verschiedene Freiheitsgrade.

Definition 6.60. Sei Y standardnormalverteilt und Z χ^2-verteilt mit n Freiheitsgraden, dann besitzt $X = Y/\sqrt{Z/n}$ die **Student-t-Verteilung**. Man schreibt $X \sim t(n)$.

Abbildung 18 zeigt die Dichtefunktion. Die t-Verteilung konvergiert mit n gegen die Standardnormalverteilung und kann daher für größere n durch diese approximiert werden.

Satz 6.61. Wenn $X \sim t(n)$, dann ist für $n \geq 2$: $\mathrm{E}(X) = 0$, und für $n \geq 3$: $\mathrm{V}(X) = \frac{n}{n-2}$.

Wenn man die Stichprobenvarianzen zweier unabhängiger Stichproben vergleicht, indem man sie durcheinander dividiert, dann ergibt sich folgende Verteilung.

Definition 6.62. Wenn $Y_1 \sim \chi^2(n_1)$ und $Y_2 \sim \chi^2(n_2)$, dann besitzt die Zufallsvariable $X = (Y_1/n_1)/(Y_2/n_2)$ die **F-Verteilung** oder **Fisher-Verteilung**. Man schreibt $X \sim F(n_1, n_2)$.

Abbildung 19 zeigt die Dichtefunktion.

6.3 Transformierte und unabhängige Zufallsvariablen

Sei g eine beliebige Funktion $g : \mathbb{R} \longrightarrow \mathbb{R}$ und X eine Zufallsvariable. Dann ist die Transformation $g(X)$ wieder eine Zufallsvariable und definiert durch: $g(X) : \Omega \longrightarrow \mathbb{R}$, $(g(X))(\omega) := g(X(\omega))$. Für die Verteilungsfunktion sieht es etwas komplizierter aus: $F_{g(X)}(x) = P(g(X) \leq x)$.

Satz 6.63. Für streng monoton steigende Funktionen g gilt $F_{g(X)}(x) = P(X \leq g^{-1}(x)) = F_X(g^{-1}(x))$. Bei diskreten Zufallsvariablen gilt $f_{g(X)}(k) = \sum_{g(j)=k} f_X(j)$.

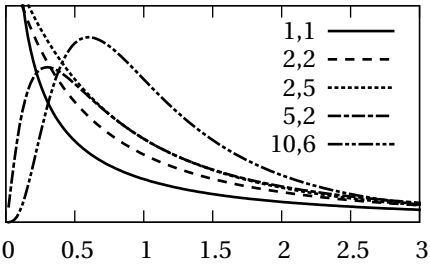
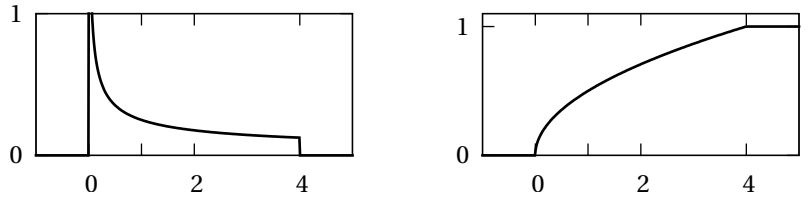


Abbildung 19: Dichtefunktion der F-Verteilung mit verschiedenen Freiheitsgraden n_1, n_2.



(a) Dichtefunktion

(b) Verteilungsfunktion

Abbildung 20: Verteilung von X^2 für X gleichverteilt zwischen 0 und 2.

Beispiel 6.64. Sei X gleichverteilt auf $[0,2]$, also $f_X(x) = \frac{1}{2}$ auf $[0,2]$, 0 außerhalb, und daher $F_X(x) = P(X \le x) = \frac{x}{2}$ auf $[0,2]$, $\mathrm{E}(X) = 1$, $\mathrm{V}(X) = \frac{2}{3}$. Sei nun $Y = X^2$. Die Funktion $g(x) = x^2$ ist streng monoton auf $(0,2]$. Daher ist $F_Y(x) = F_X(g^{-1}(x)) = F_X(\sqrt{x}) = \frac{\sqrt{x}}{2}$ auf $[0,4]$. $f_Y(x) = \frac{\mathrm{d}F_Y(x)}{\mathrm{d}x} = \frac{1}{4\sqrt{x}}$ auf $(0,4]$. Abbildung 20 zeigt die neue Dichte- und Verteilungsfunktion. $\mathrm{E}(Y) = \int_0^4 x f_Y(x)\,\mathrm{d}x = \int_0^4 x\frac{1}{4\sqrt{x}}\,\mathrm{d}x = \frac{1}{6}x^{\frac{3}{2}}\Big|_0^4 = \frac{4}{3}$.

Aufgrund von Satz 6.38 lässt sich das aber auch einfacher rechnen. $\mathrm{E}(Y) = \mathrm{E}(X^2) = \int_0^2 x^2 \frac{1}{2}\,\mathrm{d}x = \frac{1}{2}\cdot\frac{x^3}{3}\Big|_0^2 = \frac{2^3}{2\cdot 3} = \frac{4}{3}$ wie oben.

Beispiel 6.65. Spiel: 7€ setzen, 1× würfeln, 2€ pro Auge bekommen. X = Anzahl der Augen. $\mathrm{E}(X) = 3.5$, $\mathrm{V}(X) = \frac{6^2-1}{12}$, $\sigma_X \approx 1.71$. Der Gewinn ist $Y = 2X - 7$. $\mathrm{E}(Y) = 2\,\mathrm{E}(X) - 7 = 0$€, $\mathrm{V}(Y) = 2^2\,\mathrm{V}(X) = \frac{35}{3}$, $\sigma_Y = 2\sigma_X \approx 3.42$€.

Beispiel 6.66. Es werden zwei Münzen geworfen. X sei die Anzahl der Köpfe (K), Y die Anzahl der Zahlen (Z). $\mathrm{E}(X) = 0\frac{1}{4} + 1\frac{1}{2} + 2\frac{1}{4} = 1$, $\mathrm{E}(Y) = \ldots = 1$, $\mathrm{V}(X) = \mathrm{V}(Y) = \frac{1}{2}$. $(X+Y)(\mathrm{KK}) = X(\mathrm{KK}) + Y(\mathrm{KK}) = 2 + 0 = 2$, $(X+Y)(\mathrm{KZ}) = 1 + 1 = 2$, $(X+Y)(\mathrm{ZK}) = 2$, $(X+Y)(\mathrm{ZZ}) = 2$. $\mathrm{E}(X+Y) = \mathrm{E}(X) + \mathrm{E}(Y) = 1 + 1 = 2$ (stimmt).

Beispiel 6.67. Wir betrachten nun das **Produkt** von Zufallsvariablen. $XY : \Omega \longrightarrow \mathbb{R}$, $(XY)(\omega) := X(\omega)Y(\omega)$. Es ist $(XY)(\mathrm{KK}) = X(\mathrm{KK})Y(\mathrm{KK}) = 2\cdot 0 = 0$, $(XY)(\mathrm{KZ}) = (XY)(\mathrm{ZK}) = 1\cdot 1 = 1$, $(XY)(\mathrm{ZZ}) = 0$.

Definition 6.68. Die **Kovarianz** zweier Zufallsvariablen X und Y ist

$$\sigma_{X,Y} := \mathrm{E}((X - \mathrm{E}(X))(Y - \mathrm{E}(Y))) = \mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y).$$

Der **Korrelationskoeffizient** ist $\rho_{X,Y} = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y}$.

Satz 6.69.

$$\mathrm{V}(X+Y) = \mathrm{V}(X) + \mathrm{V}(Y) + 2\sigma_{X,Y}.$$

Beweis.

$$\begin{aligned}
\mathrm{V}(X+Y) &= \mathrm{E}((X+Y)^2) - \mathrm{E}(X+Y)^2 \\
&= \mathrm{E}(X^2 + 2XY + Y^2) - (\mathrm{E}(X)^2 + 2\,\mathrm{E}(X)\,\mathrm{E}(Y) + \mathrm{E}(Y)^2) \\
&= \mathrm{E}(X^2) - \mathrm{E}(X)^2 + \mathrm{E}(Y^2) - \mathrm{E}(Y)^2 + 2\,\mathrm{E}(XY) - 2\,\mathrm{E}(X)\,\mathrm{E}(Y) \\
&= \mathrm{V}(X) + \mathrm{V}(Y) + 2\sigma_{X,Y}.
\end{aligned}$$

□

Beispiel 6.70. $\mathrm{E}(XY) = 0\frac{1}{2} + 1\frac{1}{2} = \frac{1}{2}$. $\sigma_{X,Y} = \mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y) = \frac{1}{2} - 1\cdot 1 = -\frac{1}{2}$. $\rho_{X,Y} = -\frac{1}{2} \Big/ \left(\frac{1}{\sqrt{2}}\frac{1}{\sqrt{2}}\right) = -1$. $\mathrm{V}(X+Y) = \frac{1}{2} + \frac{1}{2} + 2(-\frac{1}{2}) = 0$ (stimmt).

Definition 6.71. Zwei Zufallsvariablen X, Y sind **unabhängig**, wenn die Ereignisse „$X \le x$“ und „$Y \le y$“ für alle x, y unabhängig sind, d.h. wenn gilt:

$$\forall x, y \in \mathbb{R} : P(X \le x \wedge Y \le y) = P(X \le x)P(Y \le y) = F_X(x)F_Y(y)\,.$$

Beispiel 6.72. $P(X \le 0 \cap Y \le 0) = P(\emptyset) = 0 \neq P(X \le 0)P(Y \le 0) = P(\text{ZZ})P(\text{KK}) = \frac{1}{4}\frac{1}{4} = \frac{1}{16} \Rightarrow$ nicht unabhängig.

Satz 6.73. Die Unabhängigkeit von X und Y ist äquivalent zu $\forall a, b, c, d \in \mathbb{R} : P(a < X \le b \wedge c < Y \le d) = P(a < X \le b)P(c < Y \le d)$. Und daher auch zu $\forall k, l : P(X = k \wedge Y = l) = P(X = k)P(Y = l)$ für diskrete Zufallsvariablen.

Satz 6.74. Bei unabhängigen Zufallsvariablen ist die Kovarianz gleich 0: $\sigma_{X,Y} = 0$, oder anders ausgedrückt: $\mathrm{E}(XY) = \mathrm{E}(X)\,\mathrm{E}(Y)$. *Achtung:* Zufallsvariablen mit $\sigma_{X,Y} = 0$ sind nicht automatisch unabhängig. Weiters folgt $\mathrm{V}(X + Y) = \mathrm{V}(X) + \mathrm{V}(Y)$.

Beweis. Zuerst der diskrete Fall:

$$\begin{aligned}\mathrm{E}(XY) &= \int_\Omega X(\omega)Y(\omega)\,\mathrm{d}P(\omega) = \sum_{x,y} xyP(X = x \wedge Y = y)\\ &= \sum_{x,y} xyP(X = x)P(X = y) = \sum_x xP(X = x)\sum_y yP(Y = y) = \mathrm{E}(X)\,\mathrm{E}(Y)\,.\end{aligned}$$

Für den stetigen Fall stellen wir zuerst fest, dass $f_{X,Y}(x, y) := f_X(x)f_Y(y)$ eine zweidimensionale Dichtefunktion zur zweidimensionalen Zufallsvariable (X, Y) ist, weil

$$\begin{aligned}\int_{-\infty}^{a}\int_{-\infty}^{b} f_X(x)f_Y(y)\,\mathrm{d}y\,\mathrm{d}x &= \int_{-\infty}^{a} f_X(x)\,\mathrm{d}x \int_{-\infty}^{b} f_Y(y)\,\mathrm{d}y\\ &= P(X \le a)P(Y \le b) = P(X \le a \wedge Y \le b)\,.\end{aligned}$$

Dann ist wegen Satz 6.38 mit $g(X, Y) = XY$:

$$\begin{aligned}\mathrm{E}(XY) = \int_\Omega X(\omega)Y(\omega)\,\mathrm{d}P(\omega) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_X(x)f_Y(y)\,\mathrm{d}y\,\mathrm{d}x\\ &= \int_{-\infty}^{\infty} xf_X(x)\,\mathrm{d}x \int_{-\infty}^{\infty} yf_Y(y)\,\mathrm{d}y = \mathrm{E}(X)\,\mathrm{E}(Y)\,. \quad \square\end{aligned}$$

Satz 6.75. Die Dichtefunktion der Summe zweier unabhängiger Zufallsvariablen ergibt sich aus der **Faltung** der zwei Dichtefunktionen.

$$f_{X+Y}(a) = \int_{x=-\infty}^{\infty} f_X(x)f_Y(a - x)\,\mathrm{d}x = (f_X * f_Y)(a)\,.$$

Dabei wird also eine Dichtefunktion umgedreht, über die andere geschoben, und das innere Produkt der beiden Funktionen für jede Verschiebung a gebildet.

Beweis. Die Verteilungsfunktion der Summe $X+Y$ ist

$$F_{X+Y}(a) = \int_{X+Y\le a} \mathrm{d}P(\omega) = \int_{-\infty}^{\infty}\int_{-\infty}^{a-x} f_X(x) f_Y(y)\,\mathrm{d}y\,\mathrm{d}x = \int_{-\infty}^{\infty} f_X(x) F_Y(a-x)\,\mathrm{d}x.$$

Daraus ergibt sich die Dichtefunktion durch

$$f_{X+Y}(a) = \frac{\mathrm{d}}{\mathrm{d}a} F_{X+Y}(a) = \int_{x=-\infty}^{\infty} f_X(x) f_Y(a-x)\,\mathrm{d}x. \qquad \square$$

Satz 6.76. Die Normalverteilung ist **reproduktiv**, d.h. die Summe von unabhängigen normalverteilten Zufallsvariablen ist wieder normalverteilt. Und zwar ist mit $X_1 \sim N(\mu_1,\sigma_1^2)$ und $X_2 \sim N(\mu_2,\sigma_2^2)$: $X_1+X_2 \sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$.

Beweis. Da die Verschiebung um μ_i einer Dichtefunktion nur die Verschiebung des Faltungsprodukts zur Folge hat, setzen wir die $\mu_i = 0$. Wir sind außerdem nur daran interessiert, ob $\mathrm{e}^{-cx^2} * \mathrm{e}^{-dx^2}$ die grobe Form $f\mathrm{e}^{-gx^2}$ hat, die genauen Parameter ergeben sich aus den Sätzen über die Erwartungswerte und Varianzen der Summe von Zufallsvariablen.

$$\mathrm{e}^{-cx^2} * \mathrm{e}^{-dx^2} = \int_{-\infty}^{\infty} \mathrm{e}^{-cx^2-d(a-x)^2}\,\mathrm{d}x = \int_{-\infty}^{\infty} \mathrm{e}^{-(c+d)\left(x-\frac{da}{c+d}\right)^2}\,\mathrm{d}x \cdot \mathrm{e}^{-a^2\left(d-\frac{d^2}{c+d}\right)} = f\mathrm{e}^{-ga^2},$$

wobei das Integral ein gaußsches Integral ist und daher eine Konstante f ergibt, die nicht von a abhängt. $\square$

Beispiel 6.77. Ein Ziegel habe die Höhe X. X sei normalverteilt mit $\mu_X = 20\,\mathrm{cm}$, $\sigma_X = 7\,\mathrm{mm}$. Es werden 20 Ziegel aufeinander gestapelt. Diese haben die Höhen $X_1,\ldots,X_{20}$, die unabhängig sind und alle die selbe Verteilung haben wie X. Die Gesamthöhe ist $Y = X_1+\ldots+X_{20}$. $\mathrm{E}(Y) = \mathrm{E}(X_1)+\ldots+\mathrm{E}(X_{20}) = 20\,\mathrm{E}(X) = 4\,\mathrm{m}$ (klarerweise). $\mathrm{V}(Y) = \mathrm{V}(X_1)+\ldots+\mathrm{V}(X_{20}) = 20\,\mathrm{V}(X)$, $\sigma_Y = \sqrt{20\,\mathrm{V}(X)} = \sqrt{20}\sigma_X \approx 3.13\,\mathrm{cm}$ (und nicht $20{\cdot}7\,\mathrm{mm} = 14\,\mathrm{cm}$!). Y ist also normalverteilt mit $\mu_Y = 4\,\mathrm{m}, \sigma_Y = 3.13\,\mathrm{cm}$.

7 Zentraler Grenzwertsatz

In der Realität misst man meistens Merkmale, die sich aus der Summe von unabhängigen Einzelmerkmalen ergeben. So sind z.B. Abweichungen oder Messfehler bei Messungen meistens die Summe von vielen verschiedenen Ursachen. Die

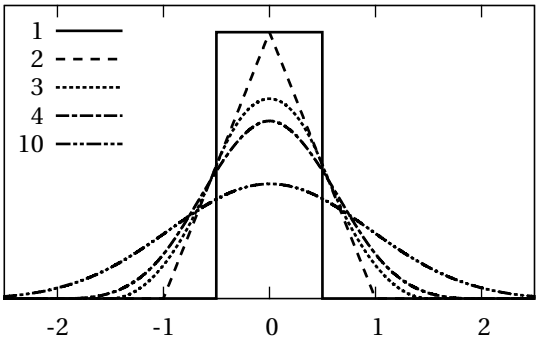


Abbildung 21: Dichtefunktion der Summe von 1, 2, 3, 4 und 10 unabhängigen Zufallsvariablen mit der Gleichverteilung zwischen -0.5 und 0.5.

Abweichungen sind dann meistens normalverteilt, wie sich herausstellt. Dies hat einen theoretischen Grund, und zwar dass die Summe von unabhängigen Zufallsvariablen mit ihrer Anzahl gegen die Normalverteilung konvergiert. Das ist die Aussage des zentralen Grenzwertsatzes.

In Abbildung 21 ist die Verteilung (Dichtefunktion) der Summen von n unabhängigen Zufallsvariablen dargestellt, für $n \in \{1,2,3,4,10\}$. Die n Zufallsvariablen sind dabei alle gleichverteilt zwischen -0.5 und 0.5. Für $n = 2$ hat man also zwei Zufallsvariablen X und Y, die unabhängig sind, aber die gleiche Verteilung besitzen. Durch Faltung bekommt man die Verteilung der Summe von X und Y. Aus den Rechtecksfunktionen der Gleichverteilung wird in Abbildung 21 also eine Dreiecksfunktion. Wenn man diese wiederum mit einer Rechtecksfunktion faltet wird das Ergebnis für $n = 3$ schon rund. Für höhere n nähert sich das Ergebnis immer mehr der Normalverteilung an.

Die Verteilung wird dabei immer breiter. Das ist kein Wunder, wissen wir doch, dass $\mathrm{V}(X+Y) = \mathrm{V}(X) + \mathrm{V}(Y)$ ist. Die Standardabweichung wächst also mit $\sqrt{n}$, d.h. $\sigma(X_1+\ldots+X_n) = \sqrt{n}\sigma(X_1)$. Wir könnten also stattdessen $\frac{1}{\sqrt{n}}(X_1+\ldots+X_n)$ betrachten, dann müssten die Verteilungen in der Breite gleich bleiben. In Abbildung 22 sieht man auch sehr schön, wie dadurch die Konvergenz zur Normalverteilung deutlich wird.

Wir können nun den zentralen Grenzwertsatz in seiner einfachsten Form formulieren.

Satz 7.1 (Zentraler Grenzwertsatz, identische standardisierte Verteilung)**.** Es sei $X_1, X_2, \ldots$ eine Folge von unabhängigen Zufallsvariablen mit gleicher Verteilung

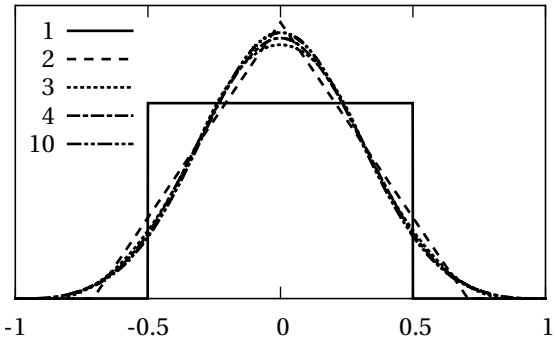


Abbildung 22: Dichtefunktion der Summe von unabhängigen Zufallsvariablen wie in Abbildung 21, nur durch die Wurzel der Anzahl dividiert.

und $\mathrm{E}(X_k) = 0$ und $\mathrm{V}(X_k) = 1$. Dann gilt

$$\lim_{n\to\infty} \frac{1}{\sqrt{n}} \sum_{k=1}^{n} X_k \sim N(0,1)\,.$$

Oder anders formuliert:

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} X_k \xrightarrow{n\to\infty} Y \quad \text{mit} \quad Y \sim N(0,1)\,.$$

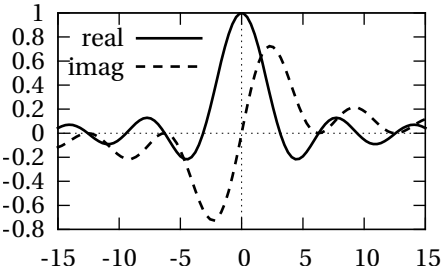
Zum Beweis für diesen Satz benötigen wir ein paar Hilfsmittel. Da die Summe von Zufallsvariablen auf eine Faltung der Dichtefunktionen hinausläuft, bietet sich der Faltungssatz aus der *Fourier*-Analyse an. Dieser besagt, dass $\mathscr{F}(f * g) = \mathscr{F}(f) \cdot \mathscr{F}(g)$, wobei $\mathscr{F}$ die Fourier-Transformation ist und f und g zwei transformierbare Funktionen sind. Das heißt, die Faltung wird durch die Transformation in eine punktweise Multiplikation verwandelt, die viel einfacher handhabbar ist. Aus diesem Grund definieren wir:

Definition 7.2. Die charakteristische Funktion einer Zufallsvariable X bzw. von deren Verteilung ist

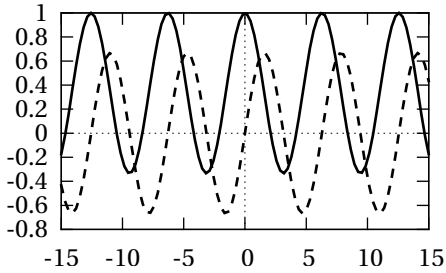
$$\varphi_X(\omega) := \mathrm{E}\left(\mathrm{e}^{\mathrm{i}\omega X}\right).$$

Satz 7.3. Wenn X und Y zwei unabhängige Zufallsvariablen sind, dann gilt

$$\varphi_{X+Y} = \varphi_X \cdot \varphi_Y\,.$$



(a) stetige Gleichverteilung



(b) diskreter Bernoulli-Versuch mit $p_1 = \frac{2}{3}$

Abbildung 23: Charakteristische Funktionen zweier einfacher Verteilungen

Beweis.

$$\varphi_{X+Y}(\omega) = \mathrm{E}\left(\mathrm{e}^{\mathrm{i}\omega(X+Y)}\right) = \mathrm{E}\left(\mathrm{e}^{\mathrm{i}\omega X} \cdot \mathrm{e}^{\mathrm{i}\omega Y}\right) = \mathrm{E}\left(\mathrm{e}^{\mathrm{i}\omega X}\right) \cdot \mathrm{E}\left(\mathrm{e}^{\mathrm{i}\omega Y}\right) = \varphi_X(\omega) \cdot \varphi_Y(\omega)$$

Beachte: die vorletzte Gleichheit gilt, weil mit X und Y auch $\mathrm{e}^{\mathrm{i}\omega X}$ und $\mathrm{e}^{\mathrm{i}\omega Y}$ zwei unabhängige Zufallsvariablen sind. □

Satz 7.4.

$$\varphi_X(0) = 1\,.$$

Beweis.

$$\varphi_X(0) = \mathrm{E}\left(\mathrm{e}^{\mathrm{i}0X}\right) = \mathrm{E}(1) = 1\,.$$ □

Beispiel 7.5. Die charakteristische Funktion der Gleichverteilung auf $[0,1]$ ist

$$\varphi(\omega) = \mathrm{E}\left(\mathrm{e}^{\mathrm{i}\omega X}\right) = \int_0^1 \mathrm{e}^{\mathrm{i}\omega x}\,\mathrm{d}x = \frac{\mathrm{e}^{\mathrm{i}\omega x}}{\mathrm{i}\omega}\Big|_0^1 = \frac{\mathrm{e}^{\mathrm{i}\omega} - 1}{\mathrm{i}\omega} = \frac{\cos\omega + \mathrm{i}\sin\omega - 1}{\mathrm{i}\omega}$$
$$= \frac{\sin\omega}{\omega} - \mathrm{i}\frac{\cos\omega - 1}{\omega}\,.$$

$\varphi(0) = 1$, wie oben bewiesen. Die charakteristische Funktion ist in Abbildung 23(a) dargestellt.

Beispiel 7.6. Auch diskrete Verteilungen haben eine charakteristische Funktion. Betrachten wir einen Bernoulli-Versuch mit $p_0 = \frac{1}{3}$ und $p_1 = \frac{2}{3}$. Dann ist

$$\varphi(\omega) = \sum_{k=-\infty}^{\infty} p_k \mathrm{e}^{\mathrm{i}\omega k} = p_0 + p_1 \mathrm{e}^{\mathrm{i}\omega} = \frac{1}{3} + \frac{2}{3}\cos\omega + \mathrm{i}\frac{2}{3}\sin\omega\,.$$

Das Ergebnis ist in Abbildung 23(b) dargestellt.

Die Idee ist nun, zu zeigen, dass die charakteristische Funktion der Summe von Zufallsvariablen gegen jene der Normalverteilung konvergiert. Letztere schauen wir uns zuerst an.

Satz 7.7. Wenn $X \sim N(0,1)$, dann ist $\varphi_X = \mathrm{e}^{-\frac{\omega^2}{2}}$.

Beweis. Die charakteristische Funktion von X sieht so aus:

$$\varphi_X(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}\omega x} \mathrm{e}^{-\frac{x^2}{2}} \,\mathrm{d}x.$$

Wir schauen uns jetzt die Ableitung davon an, um zu sehen, ob diese mit der charakteristischen Funktion in Beziehung gesetzt werden kann, quasi als Differentialgleichung.

$$\begin{aligned}
\varphi'_X(\omega) = \frac{\mathrm{d}}{\mathrm{d}\omega}\varphi_X(\omega) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{i}x\mathrm{e}^{\mathrm{i}\omega x} \mathrm{e}^{-\frac{x^2}{2}} \,\mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} -\mathrm{i}\mathrm{e}^{\mathrm{i}\omega x} \frac{\mathrm{d}}{\mathrm{d}x} \mathrm{e}^{-\frac{x^2}{2}} \,\mathrm{d}x \\
&= -\frac{\mathrm{i}}{\sqrt{2\pi}} \left(\mathrm{e}^{\mathrm{i}\omega x} \mathrm{e}^{-\frac{x^2}{2}} \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \mathrm{i}\omega \mathrm{e}^{\mathrm{i}\omega x} \mathrm{e}^{-\frac{x^2}{2}} \,\mathrm{d}x \right) \\
&= -\frac{\omega}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}\omega x} \mathrm{e}^{-\frac{x^2}{2}} \,\mathrm{d}x \\
&= -\omega\varphi_X(\omega).
\end{aligned}$$

Das heißt, wir suchen eine Funktion φ_X, für die gilt, dass $\varphi'_X(\omega) = -\omega\varphi_X(\omega)$ ist. Tatsächlich gilt das für $\mathrm{e}^{-\frac{\omega^2}{2}}$:

$$\frac{\mathrm{d}}{\mathrm{d}\omega} \mathrm{e}^{-\frac{\omega^2}{2}} = \mathrm{e}^{-\frac{\omega^2}{2}} \cdot \left(-\frac{2\omega}{2} \right) = -\omega \mathrm{e}^{-\frac{\omega^2}{2}}.$$

Auch muss $\varphi_X(0) = 1$ gelten, was auch erfüllt ist. Damit ist $\varphi_X(\omega)$ mit $\mathrm{e}^{-\frac{\omega^2}{2}}$ eindeutig bestimmt. □

Wir zeigen nun die Konvergenz der Summe nach eben dieser Funktion.

Satz 7.8. Es sei $Y_n := \frac{1}{\sqrt{n}}(X_1 + X_2 + \ldots + X_n)$. Dann gilt $\varphi_{Y_n}(\omega) \xrightarrow{n\to\infty} \mathrm{e}^{-\frac{\omega^2}{2}}$.

Beweis.

$$\varphi_{Y_n} = \varphi_{\frac{X_1}{\sqrt{n}} + \ldots + \frac{X_n}{\sqrt{n}}} = \varphi_{\frac{X_1}{\sqrt{n}}} \cdot \varphi_{\frac{X_2}{\sqrt{n}}} \cdots \varphi_{\frac{X_n}{\sqrt{n}}} = \varphi^n_{\frac{X}{\sqrt{n}}},$$

wobei X eine weitere Zufallsvariable mit der gleichen Verteilung wie X_k ist. Unter Verwendung der Reihenentwicklung der Exponentialfunktion $e^a = 1 + a + \frac{a^2}{2} + \frac{a^3}{3!} + \ldots$ bekommen wir

$$\begin{aligned}
\varphi_{\frac{X}{\sqrt{n}}}(\omega) &= \mathrm{E}\left(e^{i\omega\frac{X}{\sqrt{n}}}\right) \\
&= \mathrm{E}\left(1 + i\omega\frac{X}{\sqrt{n}} - \omega^2\frac{X^2}{2n} + \mathrm{O}\left(n^{-\frac{3}{2}}\right)\right) \\
&= 1 + i\omega\frac{\mathrm{E}(X)}{\sqrt{n}} - \omega^2\frac{\mathrm{E}(X^2)}{2n} + \mathrm{O}\left(n^{-\frac{3}{2}}\right) \\
&= 1 - \frac{\omega^2}{2n} + \mathrm{O}\left(n^{-\frac{3}{2}}\right).
\end{aligned}$$

Und unter Verwendung der Folge $(1 + \frac{a}{n})^n \xrightarrow{n\to\infty} e^a$ bekommen wir

$$\varphi_{Y_n} = \left(1 - \frac{\omega^2}{2n} + \mathrm{O}\left(n^{-\frac{3}{2}}\right)\right)^n \xrightarrow{n\to\infty} e^{-\frac{\omega^2}{2}}.$$

Der Teil $\mathrm{O}\left(n^{-\frac{3}{2}}\right)$ fällt dabei weg, weil er schneller als $\mathrm{O}\left(n^{-1}\right)$ nach 0 geht. □

Damit ist der Beweis aber noch nicht ganz am Ende, wir brauchen noch folgenden Satz.

Satz 7.9 (Lévy's Stetigkeitssatz)**.** Sei $Y_1, Y_2, Y_3, \ldots$ eine Folge von Zufallsvariablen. Wenn $\varphi_{Y_n}(\omega) \xrightarrow{n\to\infty} \varphi(\omega)$ für alle ω, und φ in 0 stetig ist, dann gibt es eine Zufallsvariable Y mit $\varphi_Y = \varphi$ und es gilt $Y_n \xrightarrow{n\to\infty} Y$.

Diesen Satz zu beweisen ist etwas technisch und nicht so erhellend, daher verzichten wir darauf. Jedenfalls schließt er den Kreis und der zentrale Grenzwertsatz in obiger Form ist damit bewiesen, weil für unsere Y_n gilt, dass $\varphi_{Y_n} \longrightarrow e^{-\frac{\omega^2}{2}}$ und die Y_n daher gegen ein standard-normalverteiltes Y konvergieren.

Es gibt nun einige Verallgemeinerungen des zentralen Grenzwertsatzes. Zuerst einmal kann man natürlich die Bedingungen $\mathrm{E}(X_k) = 0$ und $\mathrm{V}(X_k) = 1$ aufheben.

Satz 7.10 (Zentraler Grenzwertsatz nach Lindeberg-Lévy)**.** Es sei $X_1, X_2, \ldots$ eine Folge von unabhängigen Zufallsvariablen mit gleicher Verteilung und $\mathrm{E}(X_k) = \mu$ und $\mathrm{V}(X_k) = \sigma^2$. Dann gilt

$$\lim_{n\to\infty} \sqrt{n}\left(\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\right) \sim N(0, \sigma^2).$$

Der Satz in dieser Form zeigt deutlich, dass die Differenz zwischen Mittelwert und Erwartungswert gegen die Normalverteilung konvergiert und mit $\frac{1}{\sqrt{n}}$ schrumpft.

Beweis. Für $Z_k := \frac{X_k - \mu}{\sigma}$ gilt $\mathrm{E}(Z_k) = 0$ und $\mathrm{V}(Z_k) = 1$ und daher $\lim \frac{1}{\sqrt{n}} \sum_{k=1}^{n} Z_k \sim N(0,1)$. Weiters

$$\lim_{n\to\infty} \sqrt{n}\left(\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\right) = \lim_{n\to\infty} \sqrt{n}\left(\frac{1}{n}\sum_{k=1}^{n} \sigma Z_k\right) = \sigma \lim_{n\to\infty}\left(\frac{1}{\sqrt{n}}\sum_{k=1}^{n} Z_k\right) \sim N(0,\sigma^2)\,.$$

□

Als nächstes verzichten wir auch noch auf die Bedingung, dass alle X_k die gleiche Verteilung haben müssen. Leider muss man diese Bedingung dann z.B. durch die *Lindeberg-Bedingung* ersetzen.

Satz 7.11 (Zentraler Grenzwertsatz nach Lindeberg)**.** Es sei $X_1, X_2, \ldots$ eine Folge von unabhängigen Zufallsvariablen mit verschiedenen Verteilungen und $\mathrm{E}(X_k) = \mu_k$ und $\mathrm{V}(X_k) = \sigma_k^2$. Außerdem sei $s_n^2 := \sum_{k=1}^{n} \sigma_k^2$. Wenn nun die Lindeberg-Bedingung erfüllt ist, also für alle $\varepsilon > 0$ gilt

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{k=1}^{n} \mathrm{E}\left((X_k - \mu_k)^2 \cdot \chi(|X_k - \mu_k| > \varepsilon s_n)\right) = 0\,,$$

wobei $\chi(A) = 1$, wenn A wahr ist, und $\chi(A) = 0$, wenn A falsch ist, dann gilt

$$\lim_{n\to\infty} \frac{1}{s_n} \sum_{k=1}^{n} (X_k - \mu_k) \sim N(0,1)\,.$$

Die Lindeberg-Bedingung muss man sich nicht auswendig merken, es soll damit nur gezeigt werden, dass bei verschiedenen Verteilungen der Grenzwertsatz nicht automatisch gilt.

Des weiteren gibt es noch Verallgemeinerungen für nicht-allzu-abhängige Zufallsvariablen, natürlich mit diversen Zusatzbedingungen.

8 Schätzer

Stichproben werden als Realisierung von Zufallsvariablen betrachtet. Die Ähnlichkeit zwischen Kennwerten von Stichproben und Kennwerten von Zufallsvariablen (z.B. arithmetisches Mittel und Erwartungswert) ist nicht zufällig: die ersteren dienen als Schätzwerte für letztere. Zufallsvariablen sind ein *Modell* der Realität. Deren Kennwerte sind *Modellparameter*, die es gilt durch Beobachtung der Realität (*Empirik*) abzuleiten.

Definition 8.1. Die Realisierung x einer Zufallsvariable X ist ein reeller Wert, der zufällig nach der Verteilung von X erzeugt wurde. Eine Stichprobe $\{x_1, x_2, \ldots, x_n\}$ ist die Realisierung einer Folge von unabhängigen Zufallsvariablen $\{X_1, X_2, \ldots, X_n\}$ mit gleicher Verteilung wie X.

Beispiel 8.2. Ein Würfel wird 10 mal geworfen. Die Stichprobe $x = \{x_1, \ldots, x_{10}\}$, die man dabei erhält, ist die Realisierung der Folge von unabhängigen Zufallsvariablen $\{X_1, \ldots, X_{10}\}$, die alle die selbe Verteilung haben wie der Prototyp X, nämlich die Gleichverteilung auf $\{1, \ldots, 6\}$.

Definition 8.3. Sei $\theta(X)$ ein Kennwert oder Parameter von X. x sei eine Stichprobe von X mit der Größe n. Die Funktion $\hat{\theta} : \mathbb{R}^n \longrightarrow \mathbb{R}, x \mapsto \hat{\theta}(x)$ ist ein **erwartungstreuer (Punkt-)Schätzer** für θ, wenn $\mathrm{E}(\hat{\theta}(X_1, \ldots, X_n)) = \theta(X)$. Der Schätzer ist **konsistent**, wenn er gegen $\theta(X)$ konvergiert (d.h. $\mathrm{E}(\hat{\theta}) \xrightarrow{n\to\infty} \theta$ und $\mathrm{V}(\hat{\theta}) \xrightarrow{n\to\infty} 0$).

Beispiel 8.4. Wir überprüfen, ob die empirische Varianz $\hat{\theta}(x_1, \ldots, x_{10}) := s^2$ tatsächlich ein erwartungstreuer Schätzer für die Varianz $\theta(X) := V(X)$ ist:

$$\mathrm{E}(\hat{\theta}(X_1, \ldots, X_{10})) = \mathrm{E}(\frac{1}{9}(X_1^2 + \ldots + X_{10}^2 - 10\,\bar{x}^2))$$

$$= \frac{1}{9}\mathrm{E}(X_1^2 + \ldots + X_{10}^2 - \frac{1}{10}(X_1 + \ldots + X_{10})^2) = \frac{1}{9}\mathrm{E}(\frac{9}{10}X_1^2 + \ldots + \frac{9}{10}X_{10}^2 - \frac{1}{10}\sum_{i\neq j} X_i X_j)$$

$$= \frac{1}{9}\frac{9}{10}(E(X_1^2) + \ldots + E(X_{10}^2)) - \frac{1}{9}\frac{1}{10}\sum_{i\neq j}\mathrm{E}(X_i)\,\mathrm{E}(X_j) = \frac{1}{10}10\,E(X^2) - \frac{1}{9}\frac{1}{10}10\cdot 9\,(\mathrm{E}(X))^2$$

$$= \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = \mathrm{V}(X)\,.$$

Meistens werden die Parameter einer Verteilung geschätzt wie z.B. μ, σ bei $N(\mu, \sigma^2)$ oder n, p bei $B(n, p)$. D.h. damit wird das Modell (die Art der Verteilung) an die Stichprobe (also an die Realität) angepasst. Es gibt folgende wichtige Methoden, **Parameterschätzer** zu konstruieren: die Momentenmethode und die Maximum-Likelihood-Methode.

Definition 8.5. Bei der **Momentenmethode** werden die Parameter so gewählt, dass der Erwartungswert der Verteilung dem arithmetischen Mittel der Stichprobe entspricht und die Varianz der empirischen Varianz: $\mathrm{E}(X) = \bar{x}$, $\mathrm{V}(X) = s_x^2$. Das ergibt ein Gleichungssystem mit den gesuchten Parametern als Unbekannte. Wird nur ein Parameter gesucht, reicht $\mathrm{E}(X) = \bar{x}$ aus. Werden mehr als zwei Parameter gesucht, muss man weitere Momente (Schiefe, Exzess) heranziehen.

Beispiel 8.6. Gegeben sei die Stichprobe $x = \{4.5, 4.9, 5.1, 5.7, 6.4\}$. Wir nehmen an, dass die Werte gleichverteilt sind auf $[a, b]$ (*Modell*), wobei a und b zu schätzen ist (*Modellanpassung*). Wir setzen

$$\begin{aligned} \mathrm{E}(X) = \bar{x} \quad &\Leftrightarrow \quad \frac{a+b}{2} = 5.32 \quad \Leftrightarrow \quad a+b = 10.64\,, \\ \mathrm{V}(X) = s_x^2 \quad &\Leftrightarrow \quad \frac{(b-a)^2}{12} = 0.552 \quad \Leftrightarrow \quad b-a = 2.574\,. \end{aligned}$$

Die Lösung dieser zwei Gleichungen ergibt: $a = 4.033$, $b = 6.607$.

Definition 8.7. Die Wahrscheinlichkeit, dass die gegebene Stichprobe $\{x_1, \ldots, x_n\}$ auftritt, ist bei diskreten Zufallsvariablen

$$L(x) := P(X_1 = x_1 \cap \ldots \cap X_n = x_n) = f_X(x_1) f_X(x_2) \cdots f_X(x_n)\,.$$

Bei der **Maximum-Likelihood-Methode** wird dieser Ausdruck durch die Wahl der Parameter maximiert. Erleichtert wird das meist dadurch, dass man davon noch den Logarithmus nimmt:

$$\log L(x) = \log f_X(x_1) + \ldots + \log f_X(x_n)\,.$$

Das Maximum findet man nun einfach durch Ableiten nach den Parametern und Nullsetzen der Ableitung. Bei stetigen Verteilungen ist $P(X_1 = x_1 \cap \ldots)$ natürlich gleich null. Man verwendet daher einfach den Term $f_X(x_1) f_X(x_2) \cdots f_X(x_n)$.

Beispiel 8.8. Rote und grüne Äpfel werden in Packungen zu je 12 Äpfel verpackt. Der (unbekannte) Anteil der roten Äpfel sei p. In einer Stichprobe werden 5, 7, 8 und 10 rote Äpfel pro Packung gezählt. X sei die Anzahl der roten Äpfel in einer Packung: $X \sim B(12, p)$.

$$\begin{aligned} &P(X_1 = 5 \cap X_2 = 7 \cap X_3 = 8 \cap X_4 = 10) \\ &\qquad = \binom{12}{5} p^5 (1-p)^7 \binom{12}{7} p^7 (1-p)^5 \binom{12}{8} p^8 (1-p)^4 \binom{12}{10} p^{10} (1-p)^2\,. \end{aligned}$$

$$\begin{aligned} \ln P(\ldots) &= \ln \binom{12}{5} \cdots \binom{12}{10} + 5\ln p + 7\ln(1-p) + 7\ln p + 5\ln(1-p) + 8\ln p \\ &\quad + 4\ln(1-p) + 10\ln p + 2\ln(1-p) = \ln \binom{12}{5} \cdots \binom{12}{10} + 30\ln p + 18\ln(1-p)\,. \end{aligned}$$

$$\frac{\mathrm{d}\ln P(\ldots)}{\mathrm{d}p} = 0 + 30\frac{1}{p} - 18\frac{1}{1-p} \stackrel{!}{=} 0$$

$$\Leftrightarrow \quad 30(1-p) = 18p \quad \Leftrightarrow \quad 30 = 48p \quad \Leftrightarrow \quad p = \frac{30}{48} = \frac{5}{8} = 0.625\,.$$

Beachte: Wegen $np = 12\frac{5}{8} = 7.5 = \bar{x}$ entspricht das auch dem Ergebnis der Momentenmethode.

Es ist zu beachten, dass diese Schätzmethoden *nicht* automatisch in jedem Fall einen erwartungstreuen Schätzer liefern.

9 Konfidenzintervalle

Der wahre Kennwert eines Modells und der zugehörige Schätzwert werden in der Regel von einander abweichen. Es macht daher keinen Sinn, einen Schätzwert bis in die letzte Kommastelle anzugeben, wenn der wahre Kennwert mit großer Wahrscheinlichkeit stark abweicht. Man gibt daher oft besser einen Bereich an, der den Kennwert mit gewisser (großer) Wahrscheinlichkeit enthält: das Konfidenzintervall, auch Intervallschätzer genannt.

Definition 9.1. Ein Intervall $[a,b]$ heißt **Konfidenzintervall** zum Konfidenzniveau (Sicherheit) $1-\alpha$ für den Kennwert θ, wenn gilt:

$$P(a \le \theta \le b) \ge 1-\alpha\,.$$

α heißt auch Irrtumswahrscheinlichkeit.

Beispiel 9.2. Ein 5km langes Datenübertragungskabel ist gerissen. Eine Messmethode kann die Position des Risses mit einer Standardabweichung von 30m feststellen. Die Abweichung ist normalverteilt. Es wird 5 mal gemessen und man erhält die Messwerte $\{2124, 2119, 2080, 2122, 2070\}$. Gesucht ist der Bereich, in dem der Riss mit einer Wahrscheinlichkeit von 90% liegt. Es ist $\bar{x} = 2103$. Wir wissen: $\bar{x} \sim N(\mu, \frac{30^2}{5})$. Daher ist

$$P(a \le \mu \le b) = P(\bar{x}-\Delta \le \mu \le \bar{x}+\Delta) = P(\mu-\Delta \le \bar{x} \le \mu+\Delta) = 2\Phi\left(\frac{\mu+\Delta-\mu}{30/\sqrt{5}}\right) - 1\,.$$

Diese Wahrscheinlichkeit soll nun = 90% sein.

$$\Leftrightarrow \quad \Delta \ge \frac{30}{\sqrt{5}}\Phi^{-1}\left(\frac{1.9}{2}\right) \approx 22.1\,.$$

D.h. der Riss liegt mit einer Wahrscheinlichkeit von 90% im Bereich $2103 \pm 22.1 = [2080.9, 2125.1]$.

Im folgenden sei X normalverteilt: $X \sim N(\mu, \sigma^2)$. $x = \{x_1, \ldots, x_n\}$ sei eine Stichprobe von X mit Mittelwert $\bar{x}$ und empirischer Standardabweichung s.

Satz 9.3. Wenn $X \sim N(\mu, \sigma^2)$ und σ bekannt ist, dann gilt: $\bar{x}$ ist normalverteilt: $\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$ (das ist schon bekannt). Es ist

$$\left[\bar{x} - \frac{K\sigma}{\sqrt{n}}, \bar{x} + \frac{K\sigma}{\sqrt{n}}\right]$$

ein Konfidenzintervall für μ zum Konfidenzniveau $1-\alpha$, wobei $K = \Phi^{-1}(1-\frac{\alpha}{2})$ (siehe Beispiel oben).

Satz 9.4. s^2 verhält sich nach einer χ^2-Verteilung mit $n-1$ Freiheitsgraden:

$$\frac{n-1}{\sigma^2} s^2 \sim \chi^2(n-1). \qquad \text{Es ist} \quad \left[\frac{(n-1)s^2}{F^{-1}(1-\frac{\alpha}{2})}, \frac{(n-1)s^2}{F^{-1}(\frac{\alpha}{2})}\right]$$

ein Konfidenzintervall für σ^2 zum Konfidenzniveau $1-\alpha$, wobei F die Verteilungsfunktion von $\chi^2(n-1)$ ist.

Beweis.

$$\begin{aligned}
\frac{n-1}{\sigma^2} s^2 &= \frac{n-1}{\sigma^2} \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{x})^2 = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \mu + \mu - \bar{x})^2 \\
&= \frac{1}{\sigma^2} \sum_{i=1}^{n} \left((X_i - \mu)^2 + 2(X_i - \mu)(\mu - \bar{x}) + (\mu - \bar{x})^2\right) \\
&= \frac{1}{\sigma^2} \left(\sum_{i=1}^{n} (X_i - \mu)^2 + 2n(\bar{x} - \mu)(\mu - \bar{x}) + n(\mu - \bar{x})^2\right) = \underbrace{\sum_{i=1}^{n} \Big(\underbrace{\frac{X_i - \mu}{\sigma}}_{\sim N(0,1)}\Big)^2}_{\sim \chi^2(n)} - \underbrace{\Big(\underbrace{\frac{\mu - \bar{x}}{\sigma/\sqrt{n}}}_{\sim N(0,1)}\Big)^2}_{\sim \chi^2(1)},
\end{aligned}$$

und da $\chi^2(n-1) + \chi^2(1) = \chi^2(n)$ ist, ist $\chi^2(n) - \chi^2(1) = \chi^2(n-1)$. Für das Konfidenzintervall gilt:

$$\begin{aligned}
P\left(\frac{(n-1)s^2}{F^{-1}(1-\frac{\alpha}{2})} \le \sigma^2 \le \frac{(n-1)s^2}{F^{-1}(\frac{\alpha}{2})}\right) &= P\left(F^{-1}\left(1-\frac{\alpha}{2}\right) \ge \frac{(n-1)s^2}{\sigma^2} \ge F^{-1}\left(\frac{\alpha}{2}\right)\right) \\
&= 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1 - \alpha. \quad \square
\end{aligned}$$

Beispiel 9.5. σ sei nun unbekannt und muss aus der Stichprobe geschätzt werden. Es gilt: σ^2 liegt mit einer Wahrscheinlichkeit von 90% (Konfidenzniveau) in dem Intervall

$$\left[\frac{4\cdot 669}{9.488}, \frac{4\cdot 669}{0.711}\right] \approx [282,3764]\,.$$

σ liegt daher im Intervall $[16.8, 61.3]$. Die Werte $9.488, 0.711$ stammen aus der χ^2-Tabelle für $1-\frac{\alpha}{2} = 0.95$, $\frac{\alpha}{2} = 0.05$ und $n-1 = 4$.

Satz 9.6. Wenn σ *unbekannt* ist, dann verhält sich $\bar{x}$ gemeinsam mit s nach einer Student-t-Verteilung mit $n-1$ Freiheitsgraden:

$$\frac{\bar{x}-\mu}{s/\sqrt{n}} \sim t(n-1)\,. \qquad \text{Es ist} \quad \left[\bar{x} - \frac{Ks}{\sqrt{n}}, \bar{x} + \frac{Ks}{\sqrt{n}}\right]$$

ein Konfidenzintervall für μ zum Konfidenzniveau $1-\alpha$, wobei $K = F^{-1}(1-\frac{\alpha}{2})$ und F die Verteilungsfunktion von t_{n-1} ist.

Beweis.

$$\frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \cdot \frac{\sigma}{s} = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \Bigg/ \sqrt{\frac{n-1}{\sigma^2} s^2 \frac{1}{n-1}}$$

$$\sim N(0,1) \Bigg/ \sqrt{\chi^2(n-1)/(n-1)} = t(n-1)\,.$$

$$P\left(\bar{x} - \frac{Ks}{\sqrt{n}} \le \mu \le \bar{x} + \frac{Ks}{\sqrt{n}}\right) = P\left(-K \le \frac{\mu-\bar{x}}{s/\sqrt{n}} \le K\right) = 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1-\alpha\,. \qquad \square$$

Beispiel 9.7. Im obigen Beispiel sei die Standardabweichung σ (30m) nicht bekannt. Dann gilt: $s^2 = 669$, $s = 25.86$ und μ liegt mit einer Wahrscheinlichkeit von 90% (Konfidenzniveau) im Intervall

$$2103 \pm \frac{2.132 \cdot 25.86}{\sqrt{5}} = [2078, 2127]\,.$$

Der Wert 2.132 stammt aus der t-Tabelle für $1-\frac{\alpha}{2} = 0.95$ und $n-1 = 4$.

Satz 9.8. Für beliebig verteilte X gelten obige Zusammenhänge *asymptotisch*. Das heißt, sie gelten annähernd für große n. Also $\bar{x}$ nähert sich für große n einer Normalverteilung $N(\mathrm{E}(X), \frac{\mathrm{V}(X)}{n})$ an (zentraler Grenzwertsatz), s^2 einer χ^2-Verteilung, u.s.w. ...

Für beide Fälle ist wichtig zu wissen, dass für große Stichproben ($n > 100$) die t-Verteilung gegen eine Standardnormalverteilung konvergiert. Daher kann man auch bei unbekanntem σ bei großen Stichproben den Ansatz wie bei bekanntem σ verwenden und einfach s statt σ einsetzen. Bei beliebig verteilten X muss n sowieso schon groß genug sein, daher wird hier meist nur die Normalverteilung verwendet.

Satz 9.9. Ein Bernoulli-Experiment X mit Parameter p ($p_1 = p$, $p_0 = 1 - p$) wird n-mal wiederholt. p wird nun mittels $\hat{p} = k/n$ geschätzt, wobei k die Anzahl der „Treffer“ ist, also die Anzahl der Versuche, die $X = 1$ ergeben. Für n groß genug ist

$$\left[\hat{p} - \frac{Ks}{\sqrt{n}}, \hat{p} + \frac{Ks}{\sqrt{n}}\right]$$

ein Konfidenzintervall für p zum Konfidenzniveau $1 - \alpha$, wobei $K = \Phi^{-1}(1 - \frac{\alpha}{2})$ und $s = \sqrt{\hat{p}(1-\hat{p})}$.

Beweis. Es gilt $\mathrm{E}(X) = p$. Daher ist $\bar{x} = \frac{k}{n} = \hat{p}$ ein Schätzer für p. Weiters ist

$$s^2 = \frac{1}{n-1}(k \cdot 1^2 - n\bar{x}^2) = \frac{k - n\frac{k^2}{n^2}}{n-1} = \frac{k}{n-1}\left(1 - \frac{k}{n}\right) \approx \hat{p}(1-\hat{p}).$$

Nach Satz 9.8 ist daher $\hat{p} \pm Ks/\sqrt{n}$ ein Konfidenzintervall, wobei für K die Normalverteilung eingesetzt werden kann, weil n ohnehin groß sein muss. □

Es gibt in der Literatur allerdings einige genauere aber kompliziertere Konfidenzintervalle, die auch für kleinere n gelten.

Beispiel 9.10. Eine unfaire Münze wird 1000 mal geworfen und ergibt 423 mal Kopf. Das Konfidenzintervall zum Konfidenzniveau von 95% für die Wahrscheinlichkeit von Kopf ist

$$\frac{423}{1000} \pm \Phi^{-1}(0.975)\frac{\sqrt{0.423(1-0.423)}}{\sqrt{1000}} = 0.423 \pm 1.96 \cdot 0.0156 = [0.392, 0.454].$$

10 Tests

Sehr oft wird über gewisse Phänomene der Natur, der Wirtschaft, etc. eine Behauptung aufgestellt, welche dann mit Hilfe der Statistik untermauert oder angegriffen werden soll. Eine solche Behauptung wird als **Hypothese** H_0 bezeichnet

(**Nullhypothese**). Die zu H_0 komplementäre **Gegenhypothese** oder **Alternativhypothese** wird mit H_1 bezeichnet. Danach wird eine Stichprobe nach ihrer Auftrittswahrscheinlichkeit unter der Hypothese H_0 untersucht. Ist diese zu klein, muss H_0 zugunsten von H_1 verworfen werden.

Definition 10.1. Eine Hypothese über den Kennwert $\theta(X)$ einer Zufallsvariablen kann auf zwei Arten formuliert werden:

- Zweiseitige Fragestellung: $H_0 : \theta(X) = \theta_0, \quad H_1 : \theta(X) \neq \theta_0$.
- Einseitige Fragestellung: $H_0 : \theta(X) \geq \theta_0, \quad H_1 : \theta(X) < \theta_0$.

Beispiel 10.2. Ein Getränkeabfüller behauptet, die befüllten Flaschen beinhalten im Mittel genau einen Liter. D.h. die Hypothese ist $H_0 : \mathrm{E}(X) = 1$, $H_1 : \mathrm{E}(X) \neq 1$, wobei X die Füllmenge in Litern ist.

Wenn der Getränkeabfüller behauptet, die mittlere Füllmenge sei *mindestens* ein Liter, dann muss der Getränkelieferant in Kauf nehmen, in Summe zu viel an die Kunden zu liefern. Die Hypothese lautet: $H_0 : \mathrm{E}(X) \geq 1$, $H_1 : \mathrm{E}(X) < 1$.

Definition 10.3. Ein **statistischer Test** soll nun klären, ob H_0 beibehalten oder verworfen werden soll. Dabei unterscheidet man zwei Arten von Fehlern:

	H_0 wahr	H_0 falsch
H_0 angenommen	korrekt	Fehler 2. Art
H_0 verworfen	Fehler 1. Art	korrekt

Als schwerwiegenderer Fehler wird hier der Fehler 1. Art betrachtet. Die Wahrscheinlichkeit, dass man diesen Fehler begeht, soll nicht größer als α sein (im Zweifel für den Angeklagten). α heißt **Signifikanzniveau**.

Definition 10.4. Für den Schätzwert $\hat{\theta}(x)$ zu einer Stichprobe $x = \{x_1, \ldots, x_n\}$ wird ein **Annahmebereich** A um θ_0 konstruiert. Wenn $\hat{\theta}(x)$ in A liegt, soll H_0 angenommen werden. Um das Signifikanzniveau einzuhalten, sollte A möglichst klein aber so gewählt werden, dass noch gilt:

$$P(\hat{\theta}(X_1, \ldots, X_n) \in A \mid H_0) \geq 1 - \alpha\,.$$

Der komplementäre Wertebereich $\bar{A}$ heißt **Verwerfungsbereich**.

Im Folgenden nehmen wir an, X sei normalverteilt: $X \sim N(\mu, \sigma^2)$. Als $\theta(X)$ wählen wir $\mathrm{E}(X)$ und daher $\hat{\theta}(x) = \bar{x}$. σ sei unbekannt.

Satz 10.5 (Zweiseitiger t-Test)**.** Für die zweiseitige Fragestellung mit der Hypothese $H_0 : \mu = \mu_0$ gilt folgender Annahmebereich: Wenn

$$|\bar{x} - \mu_0| \frac{\sqrt{n}}{s} \leq F^{-1}\left(1 - \frac{\alpha}{2}\right),$$

dann wird die Hypothese $\mu = \mu_0$ beibehalten, ansonsten verworfen, wobei F die Verteilungsfunktion von t_{n-1} ist.

Beweis. Wir konstruieren einen Annahmebereich $\mu_0 \pm \Delta$ für $\bar{x}$. Wir wissen, dass $(\bar{x} - \mu)\frac{\sqrt{n}}{s} \sim t(n-1)$. Daraus folgt

$$\begin{aligned} & P(\mu_0 - \Delta \leq \bar{x} \leq \mu_0 + \Delta \mid \mu = \mu_0) = 1 - \alpha \\ \Leftrightarrow\ & P\left(|\bar{x} - \mu_0| \frac{\sqrt{n}}{s} \leq \Delta \frac{\sqrt{n}}{s} \,\middle|\, \mu = \mu_0\right) = 1 - \alpha \\ \Leftrightarrow\ & \Delta \frac{\sqrt{n}}{s} = F^{-1}\left(1 - \frac{\alpha}{2}\right). \quad \square \end{aligned}$$

Beispiel 10.6. Wir nehmen an, die Füllmengen aus Beispiel 10.2 seien normalverteilt. Eine Stichprobe von 5 Flaschen ergibt

$$x = \{1.0178, 0.9979, 1.0217, 1.0177, 1.0181\}, \qquad \bar{x} = 1.01464, \quad s = 0.0095.$$

Zu einem Signifikanzniveau $\alpha = 5\%$ soll geprüft werden, ob wirklich $\mu = 1$ gilt.

$$|1.01464 - 1| \frac{\sqrt{5}}{0.0095} = 3.444 \not\leq F^{-1}(0.975) = 2.776.$$

Daher ist die Abfüllanlage eher falsch eingestellt.

Satz 10.7 (Einseitiger t-Test)**.** Für die einseitige Fragestellung mit der Hypothese $H_0 : \mu \geq \mu_0$ gilt folgender Annahmebereich: Wenn

$$(\mu_0 - \bar{x}) \frac{\sqrt{n}}{s} \leq F^{-1}(1 - \alpha),$$

dann wird die Hypothese $\mu \geq \mu_0$ beibehalten, ansonsten verworfen.

Beweis. Wir konstruieren den Annahmebereich $\bar{x} \geq \mu_0 - \Delta$.

$$\begin{aligned} & P(\bar{x} \geq \mu_0 - \Delta \mid \mu \geq \mu_0) \geq 1 - \alpha \\ \Leftarrow\ & P(\bar{x} \geq \mu_0 - \Delta \mid \mu = \mu_0) = 1 - \alpha \\ \Leftrightarrow\ & P\left((\mu - \bar{x}) \frac{\sqrt{n}}{s} \leq \Delta \frac{\sqrt{n}}{s}\right) = 1 - \alpha \\ \Leftrightarrow\ & \Delta \frac{\sqrt{n}}{s} = F^{-1}(1 - \alpha). \quad \square \end{aligned}$$

Beispiel 10.8. Zum gleichen Signifikanzniveau $\alpha = 5\%$ soll geprüft werden, ob $\mu \geq 1$ gilt. Es ist $-3.444 \leq F^{-1}(0.95) = 2.132$. Die Kunden werden also nicht betrogen.

Wenn σ bekannt ist oder die Stichprobe groß genug ist, kann wie üblich die Normalverteilung statt der Student-t-Verteilung verwendet werden. In ersterem Fall ist natürlich σ statt s zu verwenden.

Definition 10.9. All diese Tests sind so konzipiert, dass auf der linken Seite der Ungleichung ein Wert steht, der nur aus der Stichprobe berechnet wird. Diesen Wert nennt man **Prüfgröße**, **Testwert** oder **Teststatistik**. Auf der rechten Seite wird ein Vergleichswert aus dem Signifikanzniveau und der Verteilung der Prüfgröße errechnet. Geht die Ungleichung zugunsten der Alternativhypothese aus, wird der Test bzw. die Abweichung der Prüfgröße als **signifikant** bezeichnet.

Oft werden zwei Stichproben miteinander verglichen, um zu sehen, ob sie sich signifikant unterscheiden. Das macht man mit dem sogenannten **doppelten t-Test**. Wenn die echten Varianzen der beiden Stichproben gleich sind, dann kann man die Mittelwerte vergleichen.

Satz 10.10 (Doppelter t-Test)**.** Beim Vergleich zweier Stichproben

$$x_1 = \{x_{1,1}, x_{1,2}, \ldots, x_{1,n_1}\} \quad \text{und} \quad x_2 = \{x_{2,1}, x_{2,2}, \ldots, x_{2,n_2}\}$$

mit den Verteilungen $N(\mu_1, \sigma^2)$ und $N(\mu_2, \sigma^2)$ gilt für die Hypothese $H_0 : \mu_1 = \mu_2$: Wenn

$$\frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \leq F^{-1}\left(1-\frac{\alpha}{2}\right),$$

dann wird H_0 beibehalten, ansonsten verworfen, wobei F die Verteilungsfunktion von $t_{n_1+n_2-2}$ ist.

Beweis. Die Varianz der Mittelwertsdifferenz ist

$$\mathrm{V}(\bar{x}_1 - \bar{x}_2) = \mathrm{V}(\bar{x}_1) + \mathrm{V}(\bar{x}_2) = \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} = \sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right).$$

Daher ist unter der Annahme $H_0 : \mu_1 = \mu_2$:

$$\frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim N(0,1).$$

Weiters ist

$$s_1^2 \frac{n_1-1}{\sigma^2} + s_2^2 \frac{n_2-1}{\sigma^2} \sim \chi^2(n_1-1) + \chi^2(n_2-1) = \chi^2(n_1+n_2-2)\,.$$

Daher ist

$$\frac{\bar{x_1}-\bar{x_2}}{\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{(n_1+n_2-2)\sigma^2}}\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim \frac{N(0,1)}{\sqrt{\frac{\chi^2(n_1+n_2-2)}{n_1+n_2-2}}} = t(n_1+n_2-2)\,. \qquad \square$$

Wenn die Stichproben gleich groß sind, also $n_1 = n_2 = n$, dann reduziert sich der Testwert auf:

$$\frac{|\bar{x}_1-\bar{x}_2|}{\sqrt{\frac{s_1^2(n-1)+s_2^2(n-1)}{2n-2}}\sqrt{\frac{1}{n}+\frac{1}{n}}} = |\bar{x}_1-\bar{x}_2|\sqrt{\frac{n}{s_1^2+s_2^2}}\,.$$

Beispiel 10.11. Die Körpergröße von jeweils 10 Personen aus zwei Ländern A und B wird gemessen. Es ergibt sich:

A	181	162	155	151	168	152	180	190	173	161
B	179	155	189	167	183	186	187	180	190	168

Sind die Personen aus den beiden Ländern im Mittel gleich groß? ($\alpha = 10\%$.)

Die Werte hängen übrigens *nicht* paarweise zusammen, dafür gäbe es wieder einen anderen Test. Es ist $\bar{x}_A = 167.3$, $\bar{x}_B = 178.4$, $s_A = 13.4$, $s_B = 11.5$.

$$|167.3-178.4|\sqrt{\frac{10}{13.4^2+11.5^2}} = 1.988 \not\leq 1.734\,,$$

wobei 1.734 aus der Tabelle der t-Verteilung kommt mit $2\cdot 10-2 = 18$ Freiheitsgraden.

Wenn die Varianzen der zwei Stichproben ungleich ist, dann verwendet man folgende Variante.

Satz 10.12. Beim Vergleich zweier Stichproben $x_1 = \{x_{1,1}, x_{1,2}, \ldots, x_{1,n_1}\}$ und x_2 mit den Verteilungen $N(\mu_1, \sigma_1^2)$ und $N(\mu_2, \sigma_2^2)$ gilt für die Hypothese $H_0 : \mu_1 = \mu_2$: Wenn

$$\frac{|\bar{x}_1-\bar{x}_2|}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_1^2}{n_1}}} \leq F^{-1}\left(1-\frac{\alpha}{2}\right),$$

dann wird H_0 beibehalten, ansonsten verworfen, wobei F die Verteilungsfunktion von t_m ist, mit

$$m \approx \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{(s_1^2/n_1)^2/(n_1-1) + (s_2^2/n_2)^2/(n_2-1)}.$$

Werden mehr als zwei Stichproben verglichen, dann könnte man die Stichproben paarweise mit obigen t-Tests vergleichen. Dabei würde aber die Wahrscheinlichkeit für einen Fehler 1. Art steigen. Man verwendet daher stattdessen eine Verallgemeinerung des doppelten t-Tests, nämlich einen F-Test. Da dieser Test die Stichproben-Varianzen mit der Varianz der Stichproben-Mittelwerte vergleicht, ist das Verfahren als **ANOVA** (ANalysis Of VAriances) bekannt.

Satz 10.13 (ANOVA)**.** Es seien $x_i = \{x_{i,1}, \ldots, x_{i,n_i}\}$, $i = 1, \ldots, m$, m Stichproben mit den Verteilungen $N(\mu_i, \sigma)$. Die Hypothese ist $H_0 : \mu_1 = \mu_2 = \ldots = \mu_m$. Wir definieren nun den Mittelwert der Varianzen $\overline{s^2}$ und die Varianz der Mittelwerte $s_{\bar{x}}^2$ als

$$\overline{s^2} := \frac{1}{n-m}\sum_i (n_i - 1)s_i^2, \qquad s_{\bar{x}}^2 := \frac{1}{m-1}\left(\sum_i n_i \bar{x}_i^2 - n\bar{x}^2\right),$$

wobei $n = n_1 + \ldots + n_m$ die Größe und $\bar{x}$ der Mittelwert der Gesamtstichprobe ist. Gilt H_0 nicht, dann müsste die Varianz der Mittelwerte im Vergleich zu groß ausfallen. Es ergibt sich daher folgender Annahmebereich: Wenn

$$\frac{s_{\bar{x}}^2}{\overline{s^2}} \leq F^{-1}(1-\alpha),$$

dann wird H_0 angenommen, ansonsten verworfen, wobei F die Verteilungsfunktion von $F(m-1, n-m)$ ist.

Beweis. Wir trennen zuerst die Gesamt-Varianz in die beiden obigen Varianzen auf und ermitteln deren Verteilung.

$$\begin{aligned}
\chi^2(n-1) \sim \frac{n-1}{\sigma^2}s^2 &= \frac{1}{\sigma^2}\left(\sum_i\sum_j X_{i,j}^2 - n\bar{x}^2\right) \\
&= \frac{1}{\sigma^2}\left(\sum_i\left(\sum_j X_{i,j}^2 - n_i\bar{x}_i^2 + n_i\bar{x}_i^2\right) - n\bar{x}^2\right) = \underbrace{\sum_i \underbrace{\frac{n_i-1}{\sigma^2}s_i^2}_{\sim\chi^2(n_i-1)}}_{\sim\chi^2(n-m)} + \underbrace{\frac{1}{\sigma^2}\sum_i n_i\bar{x}_i^2 - n\bar{x}^2}_{\Rightarrow\sim\chi^2(m-1)} \\
&= \frac{n-m}{\sigma^2}\overline{s^2} + \frac{m-1}{\sigma^2}s_{\bar{x}}^2.
\end{aligned}$$

Daher ist der Varianz-Quotient F-verteilt:

$$\frac{s_{\bar{x}}^2}{s^2} = \frac{\frac{1}{(m-1)\sigma^2}\left(\sum_i n_i \bar{x}_i^2 - n\bar{x}^2\right)}{\frac{1}{(n-m)\sigma^2}\sum_i (n_i-1)s_i^2} \sim \frac{\chi^2(m-1)/(m-1)}{\chi^2(n-m)/(n-m)} = F(m-1, n-m). \qquad \square$$

Satz 10.14 (Binomialtest)**.** Wird ein Bernoulli-Experiment mit Erfolgswahrscheinlichkeit p n-mal wiederholt, und ist das Experiment k-mal erfolgreich, dann wird die Nullhypothese $H_0 : p \geq p_0$ (bzw. $H_0 : p = p_0$) beibehalten, wenn für $X \sim B(n, p_0)$ gilt

$$P(X > k) < 1 - \alpha \qquad (\text{bzw. } P([2np - k, k]) < 1 - \alpha).$$

Die Wahrscheinlichkeiten können anhand der Binomialverteilung gerechnet oder normalapproximiert werden.

Beweis. Für den Annahmebereich A gilt gerade noch $P(X \in A) \geq 1-\alpha$. Es ist $k \in A \Leftrightarrow [k, n] \subseteq A \Leftrightarrow P(X \geq k) \leq P(X \in A) \Leftrightarrow P(X > k) < 1 - \alpha$. Zweiseitiger Fall analog. $\square$

Beispiel 10.15. Eine Münze wird 100-mal geworfen und zeigt 56-mal Kopf. Ist die Münze fair ($p = 1/2$)? $P(44 < X < 56) \approx 2\Phi((55-50)/\sqrt{25}) - 1 = 68\% < 95\%$. Ja, die Münze ist wohl fair.

Die bisherigen Tests nennt man parametrische Tests, weil sie sich auf einen Parameter der jeweiligen Verteilung beziehen. Im folgenden betrachten wir einige *nicht-parametrische Tests*. Der folgende Test überprüft die Hypothese, dass X eine bestimmte Verteilung V besitzt: $H_0 : X \sim V$.

Satz 10.16 (χ^2-Anpassungstest)**.** Der Wertebereich wird in m Klassen $[a_i, b_i)$ eingeteilt. Es fallen n_i Stichprobenwerte in die i-te Klasse. n_i sollte für $X \sim V$ annähernd gleich $np_i = nP(a_i \leq X < b_i)$ sein. Es gilt annähernd: $\sum_{i=1}^m \frac{(n_i - np_i)^2}{np_i} \sim \chi^2(m-1)$. Deshalb heißt dieser Test χ^2-Test. Es sollte $n \geq 30$ und für jede Klasse sowohl $n_i \geq 5$ als auch $np_i \geq 5$ erfüllt sein. Also gilt: Wenn

$$\sum_{i=1}^m \frac{(n_i - np_i)^2}{np_i} \leq F^{-1}(1-\alpha),$$

dann wird $H_0 : X \sim V$ beibehalten, ansonsten verworfen, wobei F die Verteilungsfunktion von $\chi^2(m-1)$ ist.

Es ist zu bemerken, dass *ungefähr* jede Einschränkung der Stichprobenwerte durch eine Gleichung die Anzahl der Freiheitsgrade um eins verringert. So führt hier die Einschränkung $\sum n_i = n$ zu den Freiheitsgraden $m-1$ statt m. Werden Verteilungs-Parameter aus der Stichprobe geschätzt, bevor der Anpassungstest durchgeführt wird, dann reduziert sich die Anzahl der Freiheitsgrade für jeden geschätzten Parameter weiter um eins. Außerdem müssen dabei noch weitere Bedingungen erfüllt sein, Vorsicht ist also geboten.

Beispiel 10.17. In einem Betrieb ereigneten sich in einem Jahr 147 Arbeitsunfälle, die folgendermaßen auf die Wochentage verteilt waren:

Mo: 32, Di: 43, Mi: 18, Do: 23, Fr: 31.

Sind die Arbeitsunfälle auf die Wochentage gleichverteilt? Unter dieser Hypothese wären an einem Wochentag 29.4 Unfälle zu erwarten. Wir wählen $\alpha = 0.01$.

$$\frac{(32-29.4)^2}{29.4} + \ldots + \frac{(31-29.4)^2}{29.4} = 12.422 \leq F^{-1}(0.99) = 13.277\,.$$

Daher kann die Gleichverteilungs-Hypothese nicht verworfen werden.

Der folgende Test überprüft die Unabhängigkeit zweier Merkmale.

Satz 10.18 (χ^2-Unabhängigkeitstest)**.** Gegeben ist eine zweidimensionale diskrete Stichprobe mit den absoluten Häufigkeiten $H_{i,j}$ und den Randhäufigkeiten $H_{i,\cdot}$ und $H_{\cdot,j}$. $i = 1,\ldots,l$, $j = 1,\ldots,m$. Wir definieren $\tilde{H}_{i,j} = H_{i,\cdot}H_{\cdot,j}/n$. Bei Unabhängigkeit der beiden Merkmale sollte $\tilde{H}_{i,j} \approx H_{i,j}$ sein. Wenn

$$\sum_{i=1}^{l}\sum_{j=1}^{m} \frac{(H_{i,j}-\tilde{H}_{i,j})^2}{\tilde{H}_{i,j}} \leq F^{-1}(1-\alpha)\,,$$

dann wird die Nullhypothese H_0 :„Merkmale sind unabhängig“ beibehalten, ansonsten verworfen, wobei F die Verteilungsfunktion von $\chi^2((l-1)(m-1))$ ist. Es sollte für alle i,j gelten: $H_{i,j} \geq 10$, $\tilde{H}_{i,j} > 1$ und $\tilde{H}_{i,j} > 5$ für mindestens 80% der Bins i,j. Ist das nicht erfüllt, müssen Klassen zusammengefasst werden. Bei nicht-diskreten Merkmalen können ebenfalls Klassen gebildet werden, damit der Test anwendbar ist.

Beispiel 10.19. Beim Exit-Poll einer Wahl ergibt sich folgendes Ergebnis:

	ABC	DEF	GHI	JKL	MNO	PQR
Männer	198	175	112	65	12	5
Frauen	164	192	89	39	20	3

Wählen Frauen anders als Männer? Anders gefragt: Sind die Merkmale Partei und Geschlecht unabhängig? Das Signifikanzniveau sei 5%.

Wir müssen die Parteien MNO und PQR zusammenfassen, damit die Bedingung $H_{i,j} \geq 10$ erfüllt ist. Dann berechnen wir die Randhäufigkeiten und die Sollwerte $\tilde{H}$:

	ABC	DEF	GHI	JKL	MNO/PQR	
Männer	198	175	112	65	17	567
$\tilde{H}_{\text{Männer}}$	191.1	193.8	106.1	54.9	21.1	
Frauen	164	192	89	39	23	507
$\tilde{H}_{\text{Frauen}}$	170.9	173.2	94.9	49.1	18.9	
	362	367	201	104	40	1074

Nun können wir die Prüfgröße überprüfen:

$$\frac{(198-191.1)^2}{191.1} + \cdots + \frac{(23-18.9)^2}{18.9} = 0.248 + \cdots + 0.898 = 10.69 \not\leq 9.49,$$

wobei 9.49 von der Tabelle der χ^2-Verteilung mit $(5-1)(2-1) = 4$ Freiheitsgraden stammt. Frauen wählen also tatsächlich anders als Männer.

Dieser Unabhängigkeitstest kann auch als Homogenitätstest verwendet werden, d.h. als Test, ob mehrere Stichproben aus der gleichen Verteilung (Grundgesamtheit) stammen. Dabei wird als zweites Merkmal die Stichproben-Nummer eingesetzt und die daraus entstehende zweidimensionale Stichprobe auf Unabhängigkeit der Merkmale getestet.

Der folgende Test ist eine Alternative für die χ^2-Anpassungs- und Homogenitätstests.

Satz 10.20 (Kolmogorow-Smirnow-Anpassungstest)**.** Die Nullhypothese ist, dass die Stichprobe x einer Verteilung V folgt, also $H_0 : X \sim V$. Es sollte dann die empirische Verteilungsfunktion F_x ungefähr gleich der Verteilungsfunktion von V sein. Wenn

$$\sup_u |F_x(u) - F_V(u)| = \max_k \max(|F_x(x_k) - F_V(x_k)|, |F_x(x_{k-1}) - F_V(x_k)|) \leq \Delta_n\left(1 - \frac{\alpha}{2}\right),$$

dann wird H_0 beibehalten, ansonsten verworfen. Die Werte für Δ werden für $n \leq 40$ einer Tabelle entnommen, für $n > 40$ gilt näherungsweise

$$\Delta_n\left(1 - \frac{\alpha}{2}\right) = \sqrt{\frac{\ln(2/\alpha)}{2n}}.$$

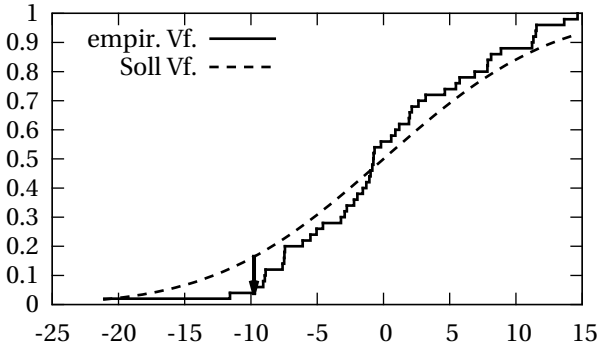


Abbildung 24: Empirische Verteilungsfunktion einer Stichprobe mit $n = 50$ sowie die Verteilungsfunktion von $N(0, 10^2)$. Die größte Abweichung ist mit einem Pfeil gekennzeichnet.

Beispiel 10.21. Abbildung 24 zeigt die Verteilungsfunktion einer Stichprobe, die nach $N(0, 8^2)$ erzeugt wurde. Können wir mit $\alpha = 5\%$ zeigen, dass diese Stichprobe nicht nach $N(0, 10^2)$ verteilt ist?

Es ist also $H_0 : X \sim N(0, 10^2)$. Die größte Abweichung tritt bei $x_3 = -9.773$ auf. Nun ist

$$\sup_u |F_x(u) - F_V(u)| = |F_x(x_2) - \Phi(-9.733/10)| = |0.04 - 0.1642| = 0.1242$$
$$\leq \Delta_{50}\left(1 - \frac{\alpha}{2}\right) = \sqrt{\frac{\ln(2/0.05)}{2 \cdot 50}} = 0.192\,.$$

Wir können die Verteilung $N(0, 10^2)$ also *nicht* widerlegen.

Satz 10.22 (Kolmogorow-Smirnow-Homogenitätstest)**.** Es gibt hier zwei Stichproben, x und y, mit den Größen n_x und n_y. Die Nullhypothese ist, dass beide Stichproben aus der gleichen Verteilung stammen. Wenn

$$\sup_u |F_x(u) - F_y(u)| = \max_k |F_x(x_k) - F_y(y_k)| \leq \Delta_{n_x,n_y}\left(1 - \frac{\alpha}{2}\right)$$

ist, dann wird H_0 beibehalten, ansonsten verworfen. Δ_{n_x,n_y} werden wieder einer Tabelle entnommen. Für große Stichproben gilt wiederum

$$\Delta_{n_x,n_y} = \Delta_m \qquad \text{mit} \qquad m = \frac{n_x n_y}{n_x + n_y}\,.$$

11 Simulation

Sehr oft sind gewisse Fragestellungen in der Statistik und der Wahrscheinlichkeitstheorie nicht oder nur schwer analytisch zu lösen. In diesem Fall können die zugrunde liegenden Zufallsprozesse simuliert und daraus Erkenntnisse abgeleitet werden. Dazu müssen zuerst Zufallszahlen erzeugt werden.

Definition 11.1. Gleichverteilte **Pseudozufallszahlen** können mit Hilfe von sogenannten Kongruenz-Generatoren erzeugt werden. Dabei berechnet man eine Folge x_i von natürlichen Zahlen mittels $x_{i+1} = (ax_i + c) \mod m$ (**additiver Kongruenz-Generator**) oder $x_{i+1} = ax_i \mod m$ (**multiplikativer Kongruenz-Generator**). Dabei müssen die Parameter a, c und m gewisse Bedingungen erfüllen. Kriterien für „gute“ Generatoren findet man in der Literatur. Bei geeigneten Parametern sind die $\frac{x_i}{m}$ gleichverteilt auf $[0,1]$.

Beispiel 11.2. $m = 714025$, $a = 1366$, $c = 150889$. x_0 sei 400000 (random seed).

$$x_1 = 321764, x_2 = 555138, x_3 = 174847, x_4 = 507541, \ldots .$$

$$\frac{x}{m} = (0.560204, 0.450634, 0.777477, 0.244875, 0.710817, \ldots) .$$

Satz 11.3. Nicht gleichverteilte Zufallszahlen können mittels der **Methode der inversen Transformation** erzeugt werden. Da Verteilungsfunktionen (F) bis auf Teilintervalle mit konstantem Funktionswert bijektiv sind, können Umkehrfunktionen (Quantilfunktionen) F^{-1} gebildet werden. Die konstanten Teilintervalle werden bei der Umkehrung einfach zu Unstetigkeitsstellen. Ist nun $(x_1, x_2, \ldots)$ eine Folge von gleichverteilten Zufallszahlen, dann ist $(F^{-1}(x_1), F^{-1}(x_2), \ldots)$ eine Folge von Zufallszahlen mit der Verteilungsfunktion F.

Beweis. Wenn X gleichverteilt ist auf $[0,1]$, dann ist $P(X \leq x) = x$ und daher

$$P(F^{-1}(X) \leq x) = P(X \leq F(x)) = F(x) . \qquad \square$$

Beispiel 11.4. Exponentialverteilung: $f(x) = \beta e^{-\beta x}, F(x) = 1 - e^{-\beta x}$ für $x > 0$. Es folgt $F^{-1}(y) = -\frac{1}{\beta} \ln(1-y)$. Mit obigen Zufallszahlen erhalten wir für $\beta = 1$:

$$(0.8214, 0.599, 1.5027, 0.281, 1.2407) .$$

Da $\Phi^{-1}(p)$ schwer berechenbar ist, gibt es für die Normalverteilung eine bessere Lösung:

Satz 11.5. Wenn $(x_1, x_2, \ldots)$ auf $(0,1)$ gleichverteilt sind, dann sind mit

$$y_{2i} = \sqrt{-2\ln x_{2i}}\cos(2\pi x_{2i+1}) \quad \text{und} \quad y_{2i+1} = \sqrt{-2\ln x_{2i}}\sin(2\pi x_{2i+1})$$

die $(y_1, y_2, \ldots)$ standardnormalverteilt. Mittels $z_i = \sigma y_i + \mu$ erhält man eine mit $N(\mu, \sigma^2)$ verteilte Folge.

Beweis. Es seien X und Y zwei unabhängige standardnormalverteilte Zufallsvariablen. Wir ersetzen diese durch die Polarkoordinaten R und S ($X = R\cos S$, $Y = R\sin S$) und ermitteln deren gemeinsame Verteilungs- und Dichtefunktion. Das führt zu einer Variablensubstitution wie im Beweis zu Satz 6.44. Sei A der Bereich von X, Y, wo $R \le r$ und $S \le s$.

$$F_{R,S}(r,s) = \iint_A \frac{1}{2\pi} \mathrm{e}^{-\frac{x^2+y^2}{2}}\,\mathrm{d}x\,\mathrm{d}y = \int_0^r \int_0^s \frac{1}{2\pi} r\mathrm{e}^{-\frac{r^2}{2}}\,\mathrm{d}s\,\mathrm{d}r = \frac{s}{2\pi}\int_0^r r\mathrm{e}^{-\frac{r^2}{2}}\,\mathrm{d}r,$$
$$f_{R,S}(r,s) = \frac{1}{2\pi}\cdot r\mathrm{e}^{-\frac{r^2}{2}} = f_S(s)\cdot f_R(r).$$

Die Dichtefunktion ist also separierbar in $f_S(s)$ und $f_R(r)$. S und R sind also unabhängig, und S ist gleichverteilt. Wir zeigen nun noch, dass $G := \mathrm{e}^{-\frac{R^2}{2}}$ gleichverteilt ist.

$$F_G(g) = P\left(\mathrm{e}^{-\frac{R^2}{2}} \le g\right) = P\left(-\frac{R^2}{2} \le \ln g\right) = P\left(R \ge \sqrt{-2\ln g}\right) = 1 - F_R\left(\sqrt{-2\ln g}\right),$$
$$f_G(g) = \frac{\mathrm{d}}{\mathrm{d}g}\left(1 - F_R\left(\sqrt{-2\ln g}\right)\right) = -f_R\left(\sqrt{-2\ln g}\right)\frac{-2}{2\sqrt{-2\ln g}\,g}$$
$$= \sqrt{-2\ln g}\cdot \mathrm{e}^{-\frac{\left(\sqrt{-2\ln g}\right)^2}{2}}\frac{1}{\sqrt{-2\ln g}\cdot g} = \frac{g}{g} = 1.$$

Durch Rückeinsetzen bekommen wir X und Y als Funktion zweier gleichverteilter Zufallsvariablen G und S:

$$X = \sqrt{-2\ln G}\cos S, \qquad Y = \sqrt{-2\ln G}\sin S.$$

Durch Skalierung von S erhalten wird die Aussage des Satzes. □

Definition 11.6. Eine **Simulation** bildet einen zufälligen Prozess nach, indem Ereignisse mit bekannter Verteilung durch Pseudozufallszahlen mit der selben Verteilung ersetzt werden und daraus die Messgrößen errechnet (statt gemessen) werden. Die resultierende Folge von „Ersatzgrößen“ wird dann mit Hilfe der deskriptiven Statistik untersucht.

Definition 11.7. Mit **Monte-Carlo-Methoden** berechnet man näherungsweise numerische Integrale. Wenn $g : [0,1] \longrightarrow \mathbb{R}$ und X auf $[0,1]$ gleichverteilt ist, dann ist $\mathrm{E}(g(X)) = \int_0^1 g(x)\,\mathrm{d}x$. Die Vorgangsweise ist nun, n Zufallszahlen $x_1, \ldots, x_n$ mit einer Gleichverteilung auf $[0,1]$ zu erzeugen. Mit $y_i := g(x_i)$ ist $\bar{y}$ ein Schätzwert für $\int_0^1 g(x)\,\mathrm{d}x$.

Diese Methode funktioniert vor allem in höheren Dimensionen sehr gut.

Beispiel 11.8. Es ist das Integral

$$\int_0^1 \int_0^1 \exp(\sqrt{1 + x_1 + \cos(\pi x_2)})\,\mathrm{d}x_1\,\mathrm{d}x_2$$

näherungsweise zu lösen. Folgendes Programm berechnet die Approximation:

$c \leftarrow 0$, $q \leftarrow 0$	Summen initialisieren
Wiederhole n-mal:	
$x_1 \leftarrow \mathrm{rand}(0,1)$, $x_2 \leftarrow \mathrm{rand}(0,1)$	Zufallszahlen generieren
$y \leftarrow \exp(\sqrt{1 + x_1 + \cos(\pi x_2)})$	Funktion berechnen
$c \leftarrow c + y$, $q \leftarrow q + y^2$	Aufsummieren von $c = \sum y_i$, $q = \sum y_i^2$
$\bar{y} \leftarrow \frac{c}{n}$, $V(y) \leftarrow \frac{q - c\cdot\bar{y}}{n-1}$, $s \leftarrow \sqrt{V(y)}$	Schätzer berechnen

Für $n = 100000$ ergibt sich: $\bar{y} \approx 3.426$, $s \approx 1.09$. Die Abweichung der Schätzung kann wiederum mit $s_{\bar{y}} \approx \frac{s}{\sqrt{n}} \approx 0.00345$ geschätzt werden.

Definition 11.9. Bei der Simulation zeitlicher Prozesse werden Ereignisse in chronologischer Abfolge nachgebildet. Das System hat einen aktuellen Zustand und einen aktuellen Zeitpunkt. Iterativ wird der aktuelle Zeitpunkt auf das nächste Ereignis verschoben und der neue Systemzustand berechnet.

Beispiel 11.10. Ein Router benötigt T ms, um ein Paket zu routen. Die Zeit Δ zwischen der Ankunft zweier Pakete sei exponentialverteilt mit Parameter β. Pakete müssen in einer **Warteschlange** warten. Der Systemzustand ist w, der nächstmögliche Zeitpunkt zur Bearbeitung eines neuen Pakets. Gesucht ist die durchschnittliche Verweilzeit.

$c \leftarrow 0$, $q \leftarrow 0$, $w \leftarrow 0$	Initialisierung
Wiederhole n-mal:	
$\Delta \leftarrow -\frac{1}{\beta}\ln(\mathrm{rand}(0,1))$	nächstes Paket in Δ ms
$w \leftarrow \max(w - \Delta, 0) + T$	neuen Systemzustand berechnen
$c \leftarrow c + w$, $q \leftarrow q + w^2$	Paket verweilt also w ms
$\bar{w} \leftarrow \frac{c}{n}$, $V(w) \leftarrow \frac{q - c\cdot\bar{w}}{n-1}$, $s \leftarrow \sqrt{V(w)}$	Schätzer berechnen

Für $\frac{1}{\beta} = 10$, $T = 7$, $n = 1000000$ ergibt sich: $\bar{w} \approx 15.21$, $s \approx 10.28$.

	+0.00	+0.01	+0.02	+0.03	+0.04	+0.05	+0.06	+0.07	+0.08	+0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
	+0.00	+0.02	+0.04	+0.06	+0.08	+0.10	+0.12	+0.14	+0.16	+0.18
1.0	0.8413	0.8461	0.8508	0.8554	0.8599	0.8643	0.8686	0.8729	0.8770	0.8810
1.2	0.8849	0.8888	0.8925	0.8962	0.8997	0.9032	0.9066	0.9099	0.9131	0.9162
1.4	0.9192	0.9222	0.9251	0.9279	0.9306	0.9332	0.9357	0.9382	0.9406	0.9429
1.6	0.9452	0.9474	0.9495	0.9515	0.9535	0.9554	0.9573	0.9591	0.9608	0.9625
1.8	0.9641	0.9656	0.9671	0.9686	0.9699	0.9713	0.9726	0.9738	0.9750	0.9761
	+0.00	+0.05	+0.10	+0.15	+0.20	+0.25	+0.30	+0.35	+0.40	+0.45
2.0	0.9772	0.9798	0.9821	0.9842	0.9861	0.9878	0.9893	0.9906	0.9918	0.9929
2.5	0.9938	0.9946	0.9953	0.9960	0.9965	0.9970	0.9974	0.9978	0.9981	0.9984
3.0	0.9987	0.9989	0.9990	0.9992	0.9993	0.9994	0.9995	0.9996	0.9997	0.9997
3.5	0.9998	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	1.0000	1.0000

Abbildung 25: Tabelle Standardnormalverteilung $N(0,1)$: $\Phi(x)$

n	0.9	0.95	0.975	0.99	0.995	0.999
1	3.078	6.314	12.706	31.821	63.657	318.309
2	1.886	2.920	4.303	6.965	9.925	22.327
3	1.638	2.353	3.182	4.541	5.841	10.215
4	1.533	2.132	2.776	3.747	4.604	7.173
5	1.476	2.015	2.571	3.365	4.032	5.893
6	1.440	1.943	2.447	3.143	3.707	5.208
7	1.415	1.895	2.365	2.998	3.499	4.785
8	1.397	1.860	2.306	2.896	3.355	4.501
9	1.383	1.833	2.262	2.821	3.250	4.297
10	1.372	1.812	2.228	2.764	3.169	4.144
11	1.363	1.796	2.201	2.718	3.106	4.025
12	1.356	1.782	2.179	2.681	3.055	3.930
13	1.350	1.771	2.160	2.650	3.012	3.852
14	1.345	1.761	2.145	2.624	2.977	3.787
15	1.341	1.753	2.131	2.602	2.947	3.733
16	1.337	1.746	2.120	2.583	2.921	3.686
17	1.333	1.740	2.110	2.567	2.898	3.646
18	1.330	1.734	2.101	2.552	2.878	3.610
19	1.328	1.729	2.093	2.539	2.861	3.579
20	1.325	1.725	2.086	2.528	2.845	3.552
22	1.321	1.717	2.074	2.508	2.819	3.505
24	1.318	1.711	2.064	2.492	2.797	3.467
25	1.316	1.708	2.060	2.485	2.787	3.450
29	1.311	1.699	2.045	2.462	2.756	3.396
30	1.310	1.697	2.042	2.457	2.750	3.385
40	1.303	1.684	2.021	2.423	2.704	3.307
50	1.299	1.676	2.009	2.403	2.678	3.261
100	1.290	1.660	1.984	2.364	2.626	3.174
200	1.286	1.653	1.972	2.345	2.601	3.131
∞	1.282	1.645	1.960	2.326	2.576	3.090

Abbildung 26: Tabelle Student-t-Verteilung t_n: $F^{-1}(p)$

n	.005	0.01	.025	0.05	0.95	0.975	0.99	0.995
1	0.00	0.00	0.00	0.00	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	7.81	9.35	11.34	12.84
4	0.21	0.30	0.48	0.71	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	15.51	17.53	20.09	21.95
9	1.73	2.09	2.70	3.33	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	26.30	28.85	32.00	34.27
19	6.84	7.63	8.91	10.12	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	31.41	34.17	37.57	40.00
24	9.89	10.86	12.40	13.85	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	37.65	40.65	44.31	46.93
29	13.12	14.26	16.05	17.71	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	67.50	71.42	76.15	79.49
100	67.33	70.06	74.22	77.93	124.3	129.6	135.8	140.2
200	152.2	156.4	162.7	168.3	234.0	241.0	249.5	255.3

Abbildung 27: Tabelle Chiquadrat-Verteilung $\chi^2(n)$: $F^{-1}(p)$

$p = 0.9$

$n_2 \backslash n_1$	1	2	3	4	5	10	20	50	100
1	39.86	49.5	53.59	55.83	57.24	60.19	61.74	62.69	63.01
2	8.526	9	9.162	9.243	9.293	9.392	9.441	9.471	9.481
3	5.538	5.462	5.391	5.343	5.309	5.23	5.184	5.155	5.144
4	4.545	4.325	4.191	4.107	4.051	3.92	3.844	3.795	3.778
5	4.06	3.78	3.619	3.52	3.453	3.297	3.207	3.147	3.126
10	3.285	2.924	2.728	2.605	2.522	2.323	2.201	2.117	2.087
20	2.975	2.589	2.38	2.249	2.158	1.937	1.794	1.69	1.65
50	2.809	2.412	2.197	2.061	1.966	1.729	1.568	1.441	1.388
100	2.756	2.356	2.139	2.002	1.906	1.663	1.494	1.355	1.293

$p = 0.95$

$n_2 \backslash n_1$	1	2	3	4	5	10	20	50	100
1	161.4	199.5	215.7	224.6	230.2	241.9	248	251.8	253
2	18.51	19	19.16	19.25	19.3	19.4	19.45	19.48	19.49
3	10.13	9.552	9.277	9.117	9.013	8.786	8.66	8.581	8.554
4	7.709	6.944	6.591	6.388	6.256	5.964	5.803	5.699	5.664
5	6.608	5.786	5.409	5.192	5.05	4.735	4.558	4.444	4.405
10	4.965	4.103	3.708	3.478	3.326	2.978	2.774	2.637	2.588
20	4.351	3.493	3.098	2.866	2.711	2.348	2.124	1.966	1.907
50	4.034	3.183	2.79	2.557	2.4	2.026	1.784	1.599	1.525
100	3.936	3.087	2.696	2.463	2.305	1.927	1.676	1.477	1.392

$p = 0.99$

$n_2 \backslash n_1$	1	2	3	4	5	10	20	50	100
1	4052	5000	5403	5625	5764	6056	6209	6303	6334
2	98.5	99	99.17	99.25	99.3	99.4	99.45	99.48	99.49
3	34.12	30.82	29.46	28.71	28.24	27.23	26.69	26.35	26.24
4	21.2	18	16.69	15.98	15.52	14.55	14.02	13.69	13.58
5	16.26	13.27	12.06	11.39	10.97	10.05	9.553	9.238	9.13
10	10.04	7.559	6.552	5.994	5.636	4.849	4.405	4.115	4.014
20	8.096	5.849	4.938	4.431	4.103	3.368	2.938	2.643	2.535
50	7.171	5.057	4.199	3.72	3.408	2.698	2.265	1.949	1.825
100	6.895	4.824	3.984	3.513	3.206	2.503	2.067	1.735	1.598

Abbildung 28: Tabelle F-Verteilung $F(n_1, n_2)$: $F^{-1}(p)$