

Statistik und Wahrscheinlichkeit

für Ingenieurstudiengänge

Matthias Künzer

Universität Stuttgart

6. August 2026

Inhalt

1	Deskriptive Statistik	4
1.1	Arithmetisches Mittel und Median	4
1.2	Standardabweichung	5
1.3	Ausgleichsgerade	7
1.4	Empirischer Korrelationskoeffizient	12
2	Wahrscheinlichkeitstheorie	16
2.1	Begriff der Wahrscheinlichkeit	16
2.2	Laplace-Experimente	19
2.3	Bedingte Wahrscheinlichkeit und Pfadregel	22
2.4	Zufallsgrößen	26
2.5	Erwartungswert	28
2.6	Varianz	31
2.7	Verteilungen	33
2.7.1	Binomialverteilung	33
2.7.2	Poisson-Verteilung	35
2.7.3	Stetige Verteilungen	38
2.7.4	Exponentialverteilung	39
2.7.5	Normalverteilung	41
2.7.6	Normalverteilung als Grenzfall der Binomialverteilung	44
3	Induktive Statistik	48
3.1	Schätzung von Erwartungswert und Varianz	48
3.1.1	Schätzung des Erwartungswerts	48
3.1.2	Schätzung der Varianz	49
3.2	Parameterschätzung bei Verteilungen bekannter Art	50
3.2.1	Parameterschätzung bei Binomialverteilung	50
3.2.2	Parameterschätzung bei Poissonverteilung	51
3.2.3	Parameterschätzung bei Normalverteilung	52
3.2.4	Parameterschätzung bei Exponentialverteilung	53
3.3	Hypothesentests	54
3.3.1	Hypothesentest bei binomialverteilten Größen	54
3.3.2	Levene-Test	58
3.3.3	t-Test für unabhängige Stichproben (Student)	61
3.3.4	U-Test für unabhängige Stichproben (Wilcoxon-Mann-Whitney)	65

Vorwort

Wir geben eine Einführung in die Statistik und die Wahrscheinlichkeitsrechnung im Hinblick auf Anwendungen im Ingenieurbereich.

Inhaltlich geht es um folgendes.

In der deskriptiven Statistik werden Median, Mittelwert, Varianz und Korrelation aus vorliegenden Daten berechnet.

In der Wahrscheinlichkeitsrechnung werden Binomial-, Poisson-, Exponential- und Normalverteilung besprochen.

In der induktiven Statistik liegen Meßreihen vor, die die Rolle von Stichproben spielen. Wir behandeln, wie man daraus durch Parameterschätzungen Aussagen über Wahrscheinlichkeiten erhalten kann.

Ferner besprechen wir Hypothesentests: Kann man anhand einer Meßreihe mit zureichender Sicherheit eine eingangs getroffene Annahme, die Nullhypothese, ausschließen?

Insbesondere behandeln wir den Levene-Test für übereinstimmende Varianzen, den t-Test von Student für übereinstimmene Erwartungswerte und den U-Test von Wilcoxon, Mann und Whitney für übereinstimmende Mediane.

Im Verlauf habe ich oft das Skript von J. DIPPON konsultiert [3].

Stuttgart, im Sommersemester 2025

Matthias Künzer

Kapitel 1

Deskriptive Statistik

1.1 Arithmetisches Mittel und Median

Seien $x_1, \dots, x_n \in \mathbb{R}$ gegeben.

Definition. Das *arithmetische Mittel* von $x_1, \dots, x_n$, auch *Mittelwert* genannt, ist

$$\bar{x} := \frac{1}{n} \sum_{j=1}^n x_j = \frac{1}{n} (x_1 + \dots + x_n) .$$

Bemerkung. Es gilt

$$\sum_{j=1}^n (x_j - \bar{x}) = 0 .$$

Diese Gleichung nennt man auch *Schwerpunkteigenschaft*.

Denn es ist

$$\sum_{j=1}^n (x_j - \bar{x}) = \left(\sum_{j=1}^n x_j \right) - n \cdot \bar{x} = \left(\sum_{j=1}^n x_j \right) - n \cdot \frac{1}{n} \sum_{j=1}^n x_j = 0 .$$

Falls man Ausreißer nicht einrechnen möchte, kann man auch wie folgt den “Wert an der mittleren Stelle” betrachten, genannt Median.

Definition. Seien $x_1 \leq x_2 \leq x_3 \leq \dots \leq x_n$ angeordnet.

Der *Median* von $x_1, \dots, x_n$ ist

$$x_{\text{med}} := \begin{cases} x_{\frac{n+1}{2}} & \text{falls } n \text{ ungerade} \\ \frac{1}{2}(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}) & \text{falls } n \text{ gerade} . \end{cases}$$

Beispiel.

(1). Sei $x_1 = 1$, $x_2 = 3$, $x_3 = 11$.

Das arithmetische Mittel ist

$$\bar{x} = \frac{1}{3}(x_1 + x_2 + x_3) = \frac{1}{3}(1 + 3 + 11) = 5.$$

Es ist $n = 3$ ungerade. Der Median ist also

$$x_{\text{med}} = x_{\frac{3+1}{2}} = x_2 = 3.$$

(2) Sei $x_1 = 1$, $x_2 = 3$, $x_3 = 10$, $x_4 = 11$.

Es ist $n = 4$ gerade. Der Median ist also

$$x_{\text{med}} = \frac{1}{2}(x_{\frac{4}{2}} + x_{\frac{4}{2}+1}) = \frac{1}{2}(x_2 + x_3) = 6,5.$$

1.2 Standardabweichung

Seien $x_1, \dots, x_n \in \mathbb{R}$ gegeben.

Wir erinnern an den Mittelwert $\bar{x}$ von $x_1, \dots, x_n$ aus §1.1.

Definition. Die *Varianz* ist

$$\frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^2.$$

Die *Standardabweichung* ist die Wurzel der Varianz, also

$$\sigma := \left(\frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^2 \right)^{1/2} = \sqrt{\frac{1}{n} ((x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2)}.$$

Falls $x_j > 0$ für $1 \leq j \leq n$, dann definiert man auch den *Variationskoeffizienten* $v := \frac{\sigma}{\bar{x}}$. Dieser ist dimensionslos und maßstabsunabhängig.

Die Standardabweichung ist ein Maß für die Streuung der Werte $x_1, \dots, x_n$ um ihren Mittelwert $\bar{x}$. Je größer die Standardabweichung, desto breiter gestreut sind die Werte. Umgekehrt ist die Standardabweichung gleich 0 genau dann, wenn alle Werte $x_1, \dots, x_n$ gleich dem Mittelwert $\bar{x}$ sind, wenn also keine Streuung vorliegt.

Bemerkung. Alternativ kann die Standardabweichung auch berechnet werden als

$$\sigma = \left(\left(\frac{1}{n} \sum_{j=1}^n x_j^2 \right) - \bar{x}^2 \right)^{1/2}.$$

Denn unter Verwendung der Schwerpunkteigenschaft folgt

$$\begin{aligned}
 \sigma^2 &= \frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^2 \\
 &= \frac{1}{n} \sum_{j=1}^n (x_j^2 - 2x_j\bar{x} + \bar{x}^2) \\
 &= \frac{1}{n} \sum_{j=1}^n (x_j^2 - 2(x_j - \bar{x})\bar{x} - 2\bar{x}^2 + \bar{x}^2) \\
 &= \frac{1}{n} \sum_{j=1}^n (x_j^2 - 2(x_j - \bar{x})\bar{x} - \bar{x}^2) \\
 &= \frac{1}{n} \sum_{j=1}^n (x_j^2 - \bar{x}^2) \\
 &= \left(\frac{1}{n} \sum_{j=1}^n x_j^2 \right) - \bar{x}^2 .
 \end{aligned}$$

Beispiel. Sei $x_1 = 1$, $x_2 = 3$, $x_3 = 11$; vgl. Beispiel (1) in §1.1.

Es ist $\bar{x} = 5$.

Die Variation ist

$$\sigma^2 = \frac{1}{3} ((1-5)^2 + (3-5)^2 + (11-5)^2) = \frac{56}{3} .$$

Die Standardabweichung ist also

$$\sigma = \sqrt{\frac{56}{3}} = 2\sqrt{\frac{14}{3}} \approx 4,32 .$$

Alternativ kann man rechnen:

$$\sigma = \sqrt{\frac{1}{3}(1^2 + 3^2 + 11^2) - 5^2} = \sqrt{\frac{56}{3}} \approx 4,32 .$$

Der Variationskoeffizient ist

$$v = \frac{\sigma}{\bar{x}} = \frac{1}{5}\sqrt{\frac{56}{3}} \approx 0,86 .$$

Beispiel. Eine Messung der Länge von 100 Werkstücken ergab folgendes.

1	Werkstück hat Länge	35 cm
4	Werkstücke haben Länge	38 cm
23	Werkstücke haben Länge	39 cm
60	Werkstücke haben Länge	40 cm (die Soll-Länge)
12	Werkstücke haben Länge	41 cm

Der Mittelwert ist

$$\bar{x} = \frac{1}{100}(1 \cdot 35 + 4 \cdot 38 + 23 \cdot 39 + 60 \cdot 40 + 12 \cdot 41) = 39,76 \text{ (cm)} .$$

Der Median ist der Mittelwert aus der 50ten und der 51ten gemessenen Länge, wenn man sie der Größe nach anordnet. Diese Werte sind beide gleich 40 cm. Also wird

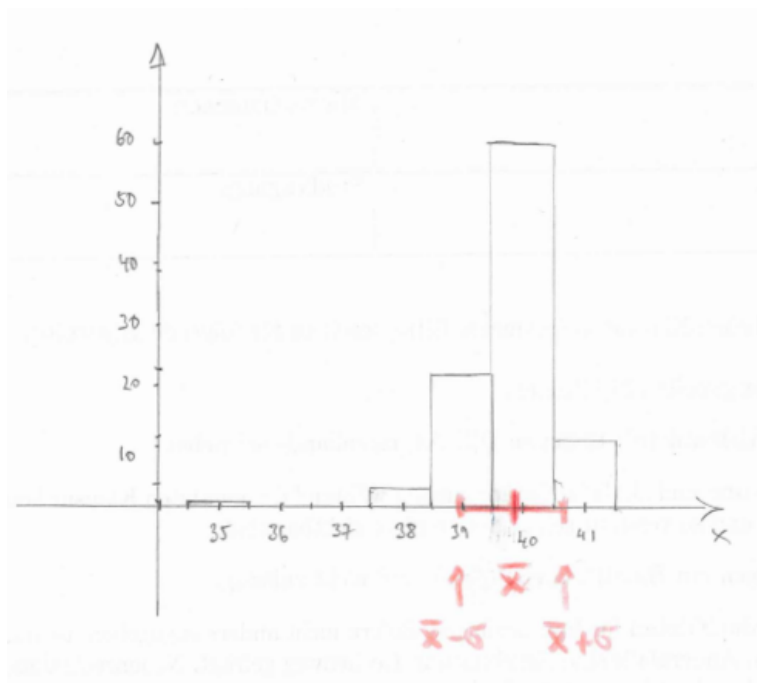
$$x_{\text{med}} = \frac{1}{2}(40 + 40) = 40 \text{ (cm)} .$$

Die Standardabweichung wird

$$\sigma = \left(\frac{1}{100} \left(1 \cdot \underbrace{(35 - \bar{x})^2}_{\approx 22,66} + 4 \cdot \underbrace{(38 - \bar{x})^2}_{\approx 3,10} + 23 \cdot \underbrace{(39 - \bar{x})^2}_{\approx 0,58} + 60 \cdot \underbrace{(40 - \bar{x})^2}_{\approx 0,06} + 12 \cdot \underbrace{(41 - \bar{x})^2}_{\approx 1,54} \right) \right)^{1/2}$$

$$\approx 0,8381 \text{ (cm)} .$$

Der Variationskoeffizient ist $v = \frac{\sigma}{\bar{x}} \approx \frac{0,8381}{39,76} \approx 0,0211$. Derselbe maßstabslose Wert für v hätte sich ergeben, hätte man alles in mm statt in cm gerechnet.

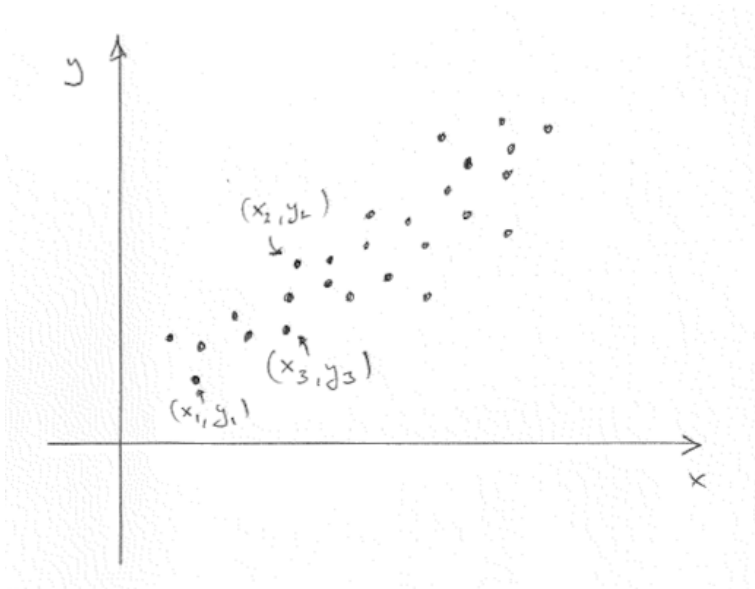


1.3 Ausgleichsgerade

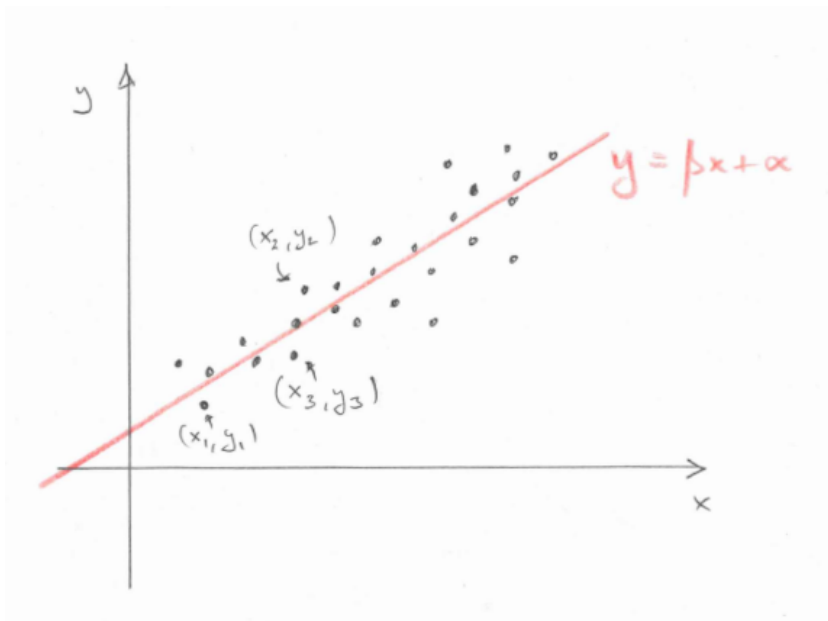
Seien $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ in $\mathbb{R}^2$ gegeben. Es soll dabei $(x_1, x_2, \dots, x_n)$ nicht aus lauter gleichen Einträgen bestehen.

Man stelle sich unter (x_j, y_j) ein Paar von zusammengehörenden Meßdaten vor. Wie z.B. die Temperatur eines Metallstabs und die bei dieser Temperatur gemessene Länge.

Zeichnet man diese Paare in der Ebene ein, so ergibt sich ein Bild wie folgt.



Gesucht ist nun eine Gerade der Form $y = f(x) = \beta x + \alpha$, die diese Sammlung von gemessenen Daten möglichst gut wiedergibt, genannt *Ausgleichsgerade*.



Folgende Bedingung ist dazu zu erfüllen:

$$e(\alpha, \beta) := \sum_{j=1}^n (f(x_j) - y_j)^2 = \text{Min!}$$

Mit anderen Worten, die Quadrate der vertikalen Abstände der einzelnen Punkte von der Ausgleichsgeraden sollen in der Summe minimal werden.

Um die kritische Stelle für $e(\alpha, \beta)$ zu bestimmen, setzen wir die partiellen Ableitungen

$e_\alpha(\alpha, \beta) = \frac{\partial}{\partial \alpha} e(\alpha, \beta)$ und $e_\beta(\alpha, \beta) = \frac{\partial}{\partial \beta} e(\alpha, \beta)$ gleich 0.

Wir rechnen. Es ist

$$\begin{aligned} e(\alpha, \beta) &= \sum_{j=1}^n (f(x_j) - y_j)^2 \\ &= \sum_{j=1}^n (\beta x_j + \alpha - y_j)^2 \\ &= \sum_{j=1}^n (\beta^2 x_j^2 + \alpha^2 + y_j^2 + 2\beta x_j \alpha - 2\beta x_j y_j - 2\alpha y_j) . \end{aligned}$$

Somit soll sein

$$e_\alpha(\alpha, \beta) = \sum_{j=1}^n (2\alpha + 2\beta x_j - 2y_j) = 2\alpha n + 2\beta \left(\sum_{j=1}^n x_j \right) - 2 \left(\sum_{j=1}^n y_j \right) \stackrel{!}{=} 0$$

und

$$e_\beta(\alpha, \beta) = \sum_{j=1}^n (2\beta x_j^2 + 2x_j \alpha - 2x_j y_j) = 2\beta \left(\sum_{j=1}^n x_j^2 \right) + 2\alpha \left(\sum_{j=1}^n x_j \right) - 2 \left(\sum_{j=1}^n x_j y_j \right) \stackrel{!}{=} 0$$

Wir kürzen ab: $\sum x_j := \sum_{j=1}^n x_j$, $\sum y_j := \sum_{j=1}^n y_j$, $\sum x_j^2 := \sum_{j=1}^n x_j^2$, $\sum x_j y_j := \sum_{j=1}^n x_j y_j$.

Zu erfüllen ist also das lineare Gleichungssystem

$$\begin{aligned} \alpha n + \beta \sum x_j &= \sum y_j \\ \alpha \sum x_j + \beta \sum x_j^2 &= \sum x_j y_j . \end{aligned}$$

Aus der Forderung

$$\begin{pmatrix} n & \sum x_j \\ \sum x_j & \sum x_j^2 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \sum y_j \\ \sum x_j y_j \end{pmatrix}$$

folgt durch Multiplikation mit der inversen Matrix

$$\begin{aligned} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} &= \begin{pmatrix} n & \sum x_j \\ \sum x_j & \sum x_j^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum y_j \\ \sum x_j y_j \end{pmatrix} \\ &= \frac{1}{n(\sum x_j^2) - (\sum x_j)^2} \begin{pmatrix} \sum x_j^2 & -\sum x_j \\ -\sum x_j & n \end{pmatrix} \begin{pmatrix} \sum y_j \\ \sum x_j y_j \end{pmatrix} \\ &= \frac{1}{n(\sum x_j^2) - (\sum x_j)^2} \begin{pmatrix} (\sum x_j^2)(\sum y_j) - (\sum x_j)(\sum x_j y_j) \\ -(\sum x_j)(\sum y_j) + n(\sum x_j y_j) \end{pmatrix} . \end{aligned}$$

Nebenrechnung 1. Es wird

$$\begin{aligned} n(\sum x_j^2) - (\sum x_j)^2 &= n^2 \left(\frac{1}{n} (\sum x_j^2) - \bar{x}^2 \right) \\ &\stackrel{\S 1.2}{=} n^2 \sigma^2 \\ &= n^2 \cdot \frac{1}{n} \sum (x_j - \bar{x})^2 \\ &= n \sum (x_j - \bar{x})^2 . \end{aligned}$$

Nebenrechnung 2. Dank Schwerpunkteigenschaft für $\bar{y}$ ist $\sum (y_j - \bar{y}) = 0$; vgl. §1.1. Damit

wird

$$\begin{aligned}
 n(\sum x_j y_j) - (\sum x_j)(\sum y_j) &= n(\sum x_j y_j) - n(\sum x_j) \bar{y} \\
 &= n(\sum x_j y_j - x_j \bar{y}) \\
 &= n(\sum x_j (y_j - \bar{y})) \\
 &= n(\sum x_j (y_j - \bar{y})) - n(\sum \bar{x} (y_j - \bar{y})) \\
 &= n(\sum (x_j - \bar{x})(y_j - \bar{y})) .
 \end{aligned}$$

Mit diesen beiden Nebenrechnungen erhalten wir aus dem unteren Vektoreintrag

$$\beta = \frac{n \sum (x_j - \bar{x})(y_j - \bar{y})}{n \sum (x_j - \bar{x})^2} = \frac{\sum (x_j - \bar{x})(y_j - \bar{y})}{\sum (x_j - \bar{x})^2} .$$

Aus der ersten Gleichung des linearen Gleichungssystems folgt $\alpha n + \beta n \bar{x} = n \bar{y}$, also

$$\beta \bar{x} + \alpha = \bar{y}$$

und somit $\alpha = \bar{y} - \beta \bar{x}$.

Nun haben wir die einzige kritische Stelle (α, β) von $e(\alpha, \beta)$ bestimmt. Man kann nun noch nachrechnen, daß an dieser tatsächlich ein globales Minimum vorliegt.

Wir fassen zusammen:

Satz. Mit

$$\beta = \frac{\sum_{j=1}^n (x_j - \bar{x})(y_j - \bar{y})}{\sum_{j=1}^n (x_j - \bar{x})^2}$$

und

$$\alpha = \bar{y} - \beta \bar{x}$$

ist die Ausgleichsgerade gegeben durch

$$y = \beta x + \alpha .$$

Bemerkung. Die Gleichung $\beta \bar{x} + \alpha = \bar{y}$ besagt: Es liegt der Punkt $(\bar{x}, \bar{y})$ auf der Ausgleichsgeraden.

Hierbei sind $\bar{x} = \frac{1}{n} \sum_{j=1}^n x_j$ und $\bar{y} = \frac{1}{n} \sum_{j=1}^n y_j$ die jeweiligen Mittelwerte.

Beispiel. Sei $(x_1, y_1) = (2, 2)$, $(x_2, y_2) = (3, 4)$ und $(x_3, y_3) = (7, 3)$.

Durch diese drei Punkte soll eine Ausgleichsgerade gelegt werden.

Wir erhalten $\bar{x} = \frac{1}{3}(x_1 + x_2 + x_3) = \frac{1}{3}(2 + 3 + 7) = 4$ und $\bar{y} = \frac{1}{3}(y_1 + y_2 + y_3) = \frac{1}{3}(2 + 4 + 3) = 3$.

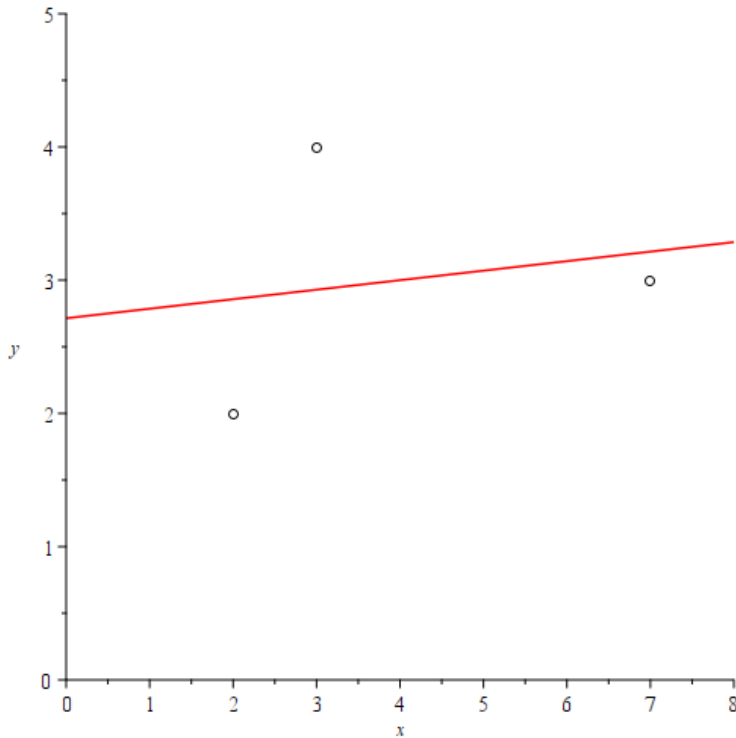
Es wird

$$\begin{aligned}
 \beta &= \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + (x_3 - \bar{x})(y_3 - \bar{y})}{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2} \\
 &= \frac{(2-4)(2-3) + (3-4)(4-3) + (7-4)(3-3)}{(2-4)^2 + (3-4)^2 + (7-4)^2} \\
 &= \frac{1}{14} \approx 0,0714 .
 \end{aligned}$$

Damit wird

$$\alpha = \bar{y} - \beta \bar{x} = 3 - \frac{1}{14} \cdot 4 = \frac{19}{7} \approx 2,7143 .$$

Die Ausgleichsgerade ist also durch $y = \frac{1}{14}x + \frac{19}{7} \approx 0,0714 \cdot x + 2,7143$ gegeben.



Beispiel. Ein Experiment liefert folgende Daten. Zu einer Temperatur x_j (in °C) eines Metallstabs wurde immer seine Länge y_j (in cm) gemessen.

$$\begin{aligned} (x_1, y_1) &= (0, 120) \\ (x_2, y_2) &= (11, 123) \\ (x_3, y_3) &= (24, 127) \\ (x_4, y_4) &= (50, 132) \\ (x_5, y_5) &= (68, 141) \end{aligned}$$

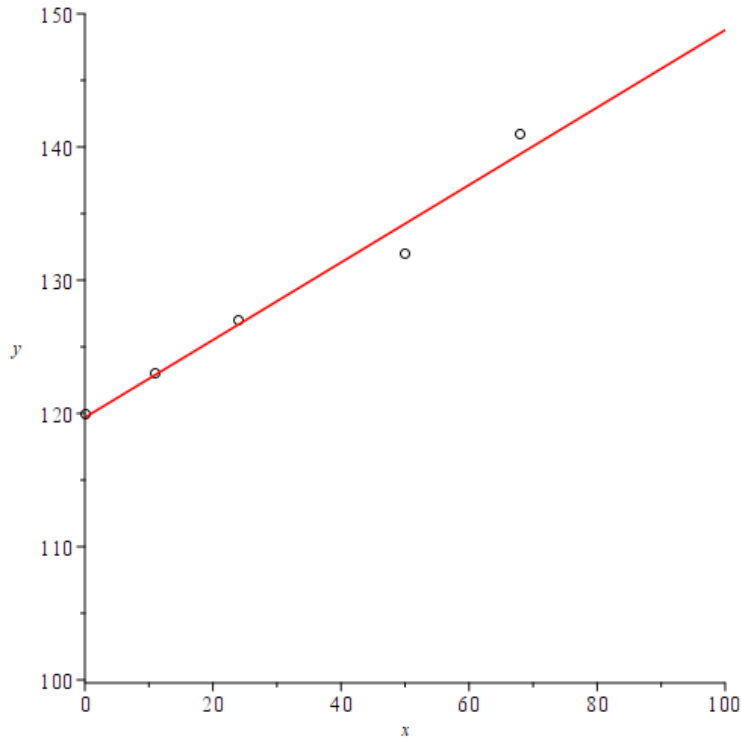
Es ergibt sich $\bar{x} = \frac{1}{5} \sum_{j=1}^5 x_j = 30,6$ und $\bar{y} = \frac{1}{5} \sum_{j=1}^5 y_j = 128,6$.

Damit wird

$$\begin{aligned} \beta &= \frac{(0-30,6)(120-128,6)+(11-30,6)(123-128,6)+(24-30,6)(127-128,6)+(50-30,6)(132-128,6)+(68-30,6)(141-128,6)}{(0-30,6)^2+(11-30,6)^2+(24-30,6)^2+(50-30,6)^2+(68-30,6)^2} \\ &\approx 0,2909 . \end{aligned}$$

und

$$\alpha = \bar{y} - \beta \bar{x} \approx 128,6 - 0,2909 \cdot 30,6 \approx 119,69 .$$



1.4 Empirischer Korrelationskoeffizient

Seien $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n) \in \mathbb{R}^2$.

Es sollen dabei weder $\vec{x} := (x_1, x_2, \dots, x_n)$ noch $\vec{y} := (y_1, y_2, \dots, y_n)$ aus lauter gleichen Einträgen bestehen.

Gesucht ist ein Maß dafür, wie ausgeprägt ein linearer Zusammenhang zwischen den Größen x_j und y_j besteht.

Definition. Der *empirische Korrelationskoeffizient* ist gegeben durch

$$r_{\vec{x}, \vec{y}} := \frac{\sum_{j=1}^n (x_j - \bar{x})(y_j - \bar{y})}{\sqrt{\left(\sum_{j=1}^n (x_j - \bar{x})^2\right) \left(\sum_{j=1}^n (y_j - \bar{y})^2\right)}}$$

Kurz heißt $r_{\vec{x}, \vec{y}}$ auch die *Korrelation*.

Bemerkung. Es kann $r_{\vec{x}, \vec{y}}$ gelesen werden als der Cosinus des Winkels zwischen den Vektoren $(x_1 - \bar{x}, x_2 - \bar{x}, \dots, x_n - \bar{x})$ und $(y_1 - \bar{y}, y_2 - \bar{y}, \dots, y_n - \bar{y})$. Also ist

$$-1 \leq r_{\vec{x}, \vec{y}} \leq +1.$$

Ansonsten braucht man die geometrische Interpretation dieser Vektoren aber nicht.

Bemerkung.

(1) Es ist $r_{\vec{x}, \vec{y}} = r_{\vec{y}, \vec{x}}$.

(2) Für $c \in \mathbb{R}$ sei $\vec{x} + c := (x_1 + c, x_2 + c, \dots, x_n + c)$. Es ist

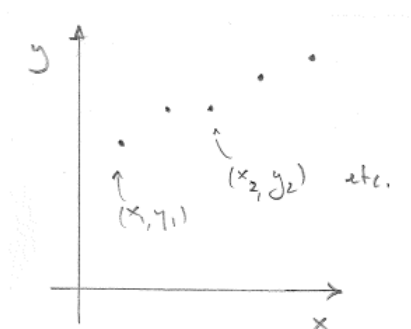
$$r_{\vec{x}+c, \vec{y}} = r_{\vec{x}, \vec{y}} = r_{\vec{x}, \vec{y}+c}.$$

(2) Für $a \in \mathbb{R} \setminus \{0\}$ sei $a\vec{x} := (ax_1, ax_2, \dots, ax_n)$. Es ist

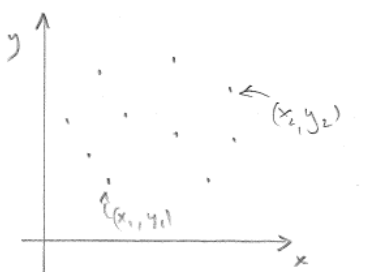
$$r_{\vec{x}, a\vec{y}} = \begin{cases} r_{\vec{x}, \vec{y}} & \text{falls } a > 0 \\ -r_{\vec{x}, \vec{y}} & \text{falls } a < 0. \end{cases}$$

Beispiel zur Vorstellung.

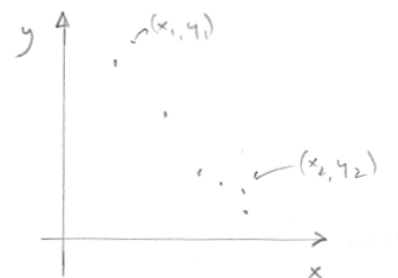
Korrelation $r_{\vec{x}, \vec{y}}$ knapp unter 1:



Korrelation $r_{\vec{x}, \vec{y}}$ nahe bei 0:



Korrelation $r_{\vec{x}, \vec{y}}$ knapp über -1:

**Bemerkung.**

Ist die Korrelation knapp unter 1, so sagt man, die x -Werte und die y -Werte sind stark korreliert. Die Punkte liegen dann nahe an der Ausgleichsgeraden.

Ist die Korrelation nahe bei 0, so sagt man, die x -Werte und die y -Werte sind schwach korreliert. Die Punkte liegen dann weit entfernt von der Ausgleichsgeraden.

Ist die Korrelation knapp über -1, so sagt man, die x -Werte und die y -Werte sind stark negativ korreliert. Die Punkte liegen dann nahe an der Ausgleichsgeraden.

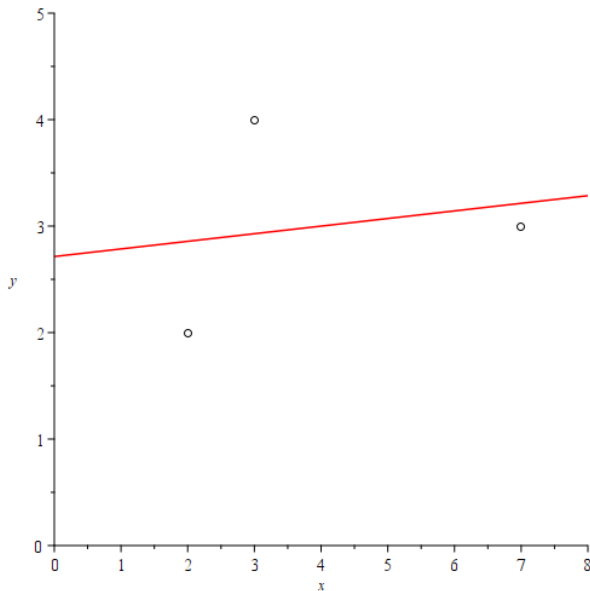
Beispiel. Sei $(x_1, y_1) = (2, 2)$, $(x_2, y_2) = (3, 4)$ und $(x_3, y_3) = (7, 3)$.

Vgl. erstes Beispiel in §1.3. Es ist $\bar{x} = 4$ und $\bar{y} = 3$.

Die Korrelation ergibt sich zu

$$\begin{aligned} r_{\vec{x}, \vec{y}} &= \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + (x_3 - \bar{x})(y_3 - \bar{y})}{((x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2)^{1/2} \cdot ((y_1 - \bar{y})^2 + (y_2 - \bar{y})^2 + (y_3 - \bar{y})^2)^{1/2}} \\ &= \frac{(2-4)(2-3) + (3-4)(4-3) + (7-4)(3-3)}{((2-4)^2 + (3-4)^2 + (7-4)^2)^{1/2} \cdot ((2-3)^2 + (4-3)^2 + (3-3)^2)^{1/2}} \\ &= \frac{1}{\sqrt{14} \cdot 2} = \frac{1}{2\sqrt{7}} \approx 0,1890. \end{aligned}$$

Die x -Werte und die y -Werte sind nur schwach korreliert. Die Ausgleichsgerade hat großen Abstand von den Punkten.



Beispiel. Wir betrachten wieder das zweite Beispiel in §1.3. Dort lieferte ein Experiment folgende Daten. Zu einer Temperatur x_j (in °C) eines Metallstabs wurde immer seine Länge y_j (in cm) gemessen.

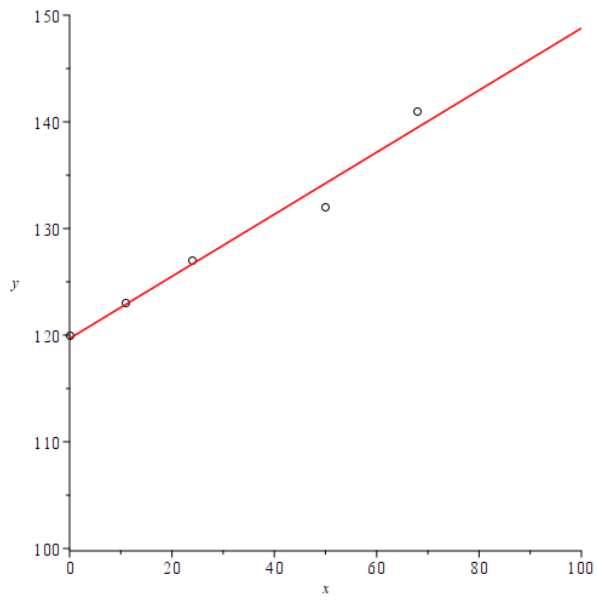
$$\begin{aligned}
 (x_1, y_1) &= (0, 120) \\
 (x_2, y_2) &= (11, 123) \\
 (x_3, y_3) &= (24, 127) \\
 (x_4, y_4) &= (50, 132) \\
 (x_5, y_5) &= (68, 141)
 \end{aligned}$$

Es ist $\bar{x} = 30,6$ und $\bar{y} = 128,6$.

Die Korrelation ergibt sich zu

$$\begin{aligned}
 r_{\vec{x}, \vec{y}} &= \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + (x_3 - \bar{x})(y_3 - \bar{y}) + (x_4 - \bar{x})(y_4 - \bar{y}) + (x_5 - \bar{x})(y_5 - \bar{y})}{((x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + (x_4 - \bar{x})^2 + (x_5 - \bar{x})^2)^{1/2} \cdot ((y_1 - \bar{y})^2 + (y_2 - \bar{y})^2 + (y_3 - \bar{y})^2 + (y_4 - \bar{y})^2 + (y_5 - \bar{y})^2)^{1/2}} \\
 &\approx 0,9861.
 \end{aligned}$$

Die x -Werte (jeweilige Temperatur des Metallstabs) und die y -Werte (jeweilige Länge des Metallstabs) sind stark korreliert. Die Ausgleichsgerade hat kleinen Abstand von den Punkten.



Kapitel 2

Wahrscheinlichkeitstheorie

2.1 Begriff der Wahrscheinlichkeit

Definition. Gegeben sei als *Ergebnisraum* eine Menge Ω .

Es ist Ω die Menge der möglichen Ergebnisse eines Zufallsexperiments.

Beispiel. Beim Würfeln mit einem Würfel ist $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Beispiel. Beim Messen der Temperatur (in °C) um 8:00 morgens ist $\Omega = \mathbb{R}_{\geq -273,15}$.

Bemerkung. Aus mathematik-technischen Gründen muß man manchmal noch zwischen meßbaren und nicht meßbaren Teilmengen $A \subseteq \Omega$ unterscheiden. Nicht meßbare Teilmengen von Ω sind nicht ganz einfach zu finden. Die in der Praxis üblicherweise auftretenden Teilmengen sind in aller Regel meßbar. Wir werden dieses Problem daher weitgehend ignorieren.

Ist z.B. Ω endlich, so stellt sich dieses Problem nicht: Jede Teilmenge von Ω ist meßbar.

Definition. Eine meßbare Teilmenge $A \subseteq \Omega$ heißt *Ereignis*.

Diese Teilmenge stellt das Ereignis dar, daß das Ergebnis des Zufallsexperiment in A liegt.

Beispiel. Sei wieder beim Würfeln $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Das Ereignis, eine ungerade Zahl zu würfeln, ist die Teilmenge $A = \{1, 3, 5\} \subseteq \Omega$.

Beispiel. Sei wieder beim Temperaturmessen $\Omega = \mathbb{R}_{\geq -273,15}$.

Das Ereignis, Frost zu haben, also eine Temperatur ≤ 0 , ist die Teilmenge

$$\{x \in \Omega : x \leq 0\} = [-273,15, 0] \subseteq \Omega.$$

Definition. Gegeben sei ein *Wahrscheinlichkeitsmaß* P , das jedem Ereignis $A \subseteq \Omega$ eine *Wahrscheinlichkeit* $P(A) \in \mathbb{R}$ zuordnet.

Dabei sollen die folgenden Bedingungen von Kolmogorov gelten.

(K 1) Es ist $P(A) \geq 0$ für jedes Ereignis $A \subseteq \Omega$.

(K 2) Es ist $P(\Omega) = 1$.

(K 3) Es ist $P(A \cup B) = P(A) + P(B)$ für Ereignisse $A, B \subseteq \Omega$ mit $A \cap B = \emptyset$.

Allgemeiner ist auch noch $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$ für Ereignisse $A_1, A_2, \dots \subseteq \Omega$ mit $A_i \cap A_j = \emptyset$ für $i, j \geq 1$ mit $i \neq j$.

Interpretation:

- Es ist $P(A)$ die Wahrscheinlichkeit, daß das Zufallsexperiment ein Ergebnis in A liefert. (“ P ” für “probability”.)
- $P(A) = 0$ heißt: Es ist fast unmöglich, ein Ergebnis in A zu haben.
- $P(A) = 1$ heißt: Es ist fast sicher, ein Ergebnis in A zu haben.
Zum Beispiel ist $P(\Omega) = 1$, da es sicher ist, ein Ergebnis in Ω zu haben.
- Sei $A \subseteq \Omega$ ein Ereignis. Es heißt $\bar{A} := \Omega \setminus A \subseteq \Omega$ das zugehörige *Gegenereignis*.

Dank (K 3) ist

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

Insbesondere ist

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

Da $P(\bar{A}) \geq 0$ ist, folgt

$$0 \leq P(A) \leq 1.$$

Ferner folgt

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

- (K 3) heißt: Wahrscheinlichkeiten disjunkter Ereignisse werden addiert.

Beispiel. Sei wieder beim Würfeln $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Es ist

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

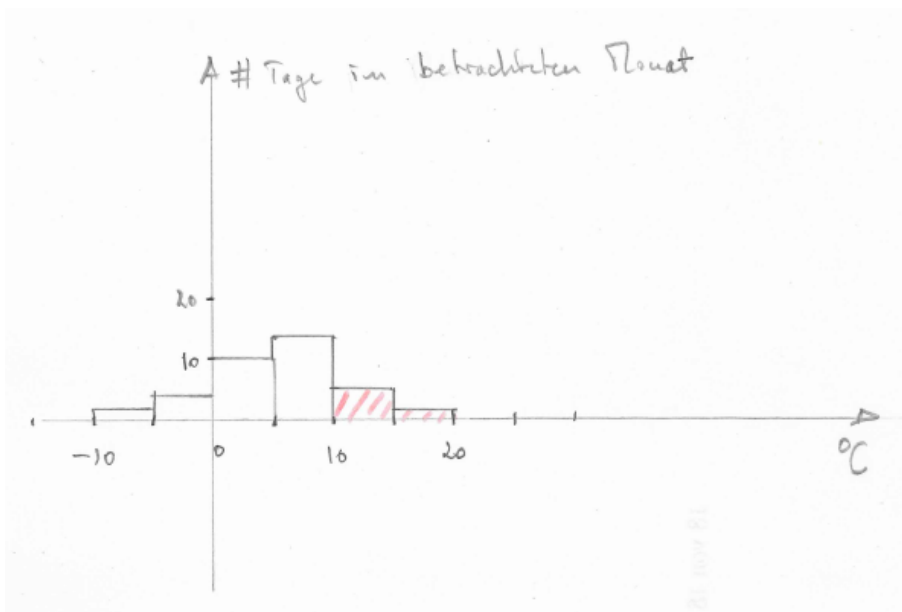
Die Wahrscheinlichkeit, eine ungerade Zahl zu würfeln, ist

$$P(\{1, 3, 5\}) = P(\{1\}) + P(\{3\}) + P(\{5\}) = 3 \cdot \frac{1}{6} = \frac{1}{2}.$$

Das Gegenereignis, eine gerade Zahl zu würfeln, hat also Wahrscheinlichkeit $1 - \frac{1}{2} = \frac{1}{2}$.

Beispiel. (Temperaturmessung)

Im Monat März sei folgende langjährige Temperaturverteilung bekannt, die wir unseren Wahrscheinlichkeitsbetrachtungen zugrunde legen wollen.



Es gebe also z.B. 10 Tage, an denen die Durchschnittstemperatur zwischen 0 und 5 Grad liegt.

Hier ist $\Omega = \mathbb{R}_{\geq -273,15}$.

Es ist $P(\{10\}) = 0$, da es fast unmöglich ist, eine Temperatur von exakt 10 Grad zu erhalten.

Es ist $P([10, 20])$ die Wahrscheinlichkeit, eine Temperatur zwischen 10 und 20 Grad zu erhalten.

Es ist

$$P([10, 20]) = \frac{\text{Anzahl der Tage mit Temperatur zwischen 10 und 20 Grad}}{\text{Anzahl der Tage im Monat}}$$

anschaulich

$$= \frac{\text{Flächeninhalt in rot}}{\text{Flächeninhalt unter Kurve insgesamt}}.$$

An diesem Beispiel kann man erkennen, warum allgemein P immer Ereignissen, also Teilmengen von Ω , eine Wahrscheinlichkeit zuweist, und nicht einem einzelnen Ergebnis.

2.2 Laplace-Experimente

Definition. Ist M eine endliche Menge, so schreiben wir $\#M$ für die Anzahl ihrer Elemente.

Definition. Ist Ω endlich und nichtleer und ist

$$P(\{x\}) = \frac{1}{\#\Omega}$$

für jedes $x \in \Omega$, so sprechen wir von einem *Laplace-Experiment*.

Diesemfalls gilt dank (K 3) für $A \subseteq \Omega$

$$P(A) = \frac{\#A}{\#\Omega}.$$

In Worten, die Wahrscheinlichkeit für ein Ereignis A ist gleich der Anzahl der für A günstigen Ergebnisse (also $\#A$), geteilt durch die Anzahl aller möglichen Ergebnisse (also $\#\Omega$).

Beispiel. Es werde eine Münze dreimal geworfen. Es stehe W für Wappen, Z für Zahl.

Es ist

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

Es liegt ein Laplace-Experiment vor, da jedes Ergebnis Wahrscheinlichkeit $\frac{1}{\#\Omega} = \frac{1}{8}$ hat.

Wir betrachten das Ereignis, daß genau zweimal Wappen geworfen wird, also

$$A = \{(W, W, Z), (W, Z, W), (Z, W, W)\}.$$

Dafür ist die Wahrscheinlichkeit

$$P(A) = \frac{\#A}{\#\Omega} = \frac{3}{8}.$$

Beispiel. Wir betrachten das Lottospiel “3 aus 9”. Es werden aus den Zahlen $\{1, 2, \dots, 9\}$ drei Zahlen gezogen.

Der Ergebnisraum ist also

$$\begin{aligned} \Omega = \{ & (1, 2, 3), (1, 2, 4), \dots, (1, 2, 9), \\ & (1, 3, 2), (1, 3, 4), \dots, (1, 3, 9), \\ & (1, 4, 2), (1, 4, 3), \dots, (1, 4, 9), \\ & \vdots \\ & (9, 8, 1), (9, 8, 2), \dots, (9, 8, 7) \} \end{aligned}$$

Es ist $\#\Omega = 9 \cdot 8 \cdot 7$.

Jemand hat die Zahlen 1, 4 und 5 getippt.

Er will wissen, wie hoch die Wahrscheinlichkeit für drei Richtige ist. Das zugehörige Ereignis ist

$$A_3 = \{(1, 4, 5), (1, 5, 4), (4, 1, 5), (4, 5, 1), (5, 1, 4), (5, 4, 1)\}.$$

Es ist $\#A_3 = 3 \cdot 2 \cdot 1 = 6$. Also ist

$$P(A_3) = \frac{\#A_3}{\#\Omega} = \frac{3 \cdot 2 \cdot 1}{9 \cdot 8 \cdot 7} = \frac{1}{\binom{9}{3}} = \frac{1}{84}.$$

Er will auch noch wissen, wie hoch die Wahrscheinlichkeit für zwei Richtige ist. Das zugehörige Ereignis ist

$$\begin{aligned} A_2 = \{ & (1, 4, x), (4, 1, x), (1, 5, x), (5, 1, x), (4, 5, x), (5, 4, x), \\ & (1, x, 4), (4, x, 1), (1, x, 5), (5, x, 1), (4, x, 5), (5, x, 4), \\ & (x, 1, 4), (x, 4, 1), (x, 1, 5), (x, 5, 1), (x, 4, 5), (x, 5, 4) \\ & : x \in \{2, 3, 6, 7, 8, 9\} \}. \end{aligned}$$

Es ist $\#A_2 = 3 \cdot 6 \cdot (9 - 3)$. Also ist

$$P(A_2) = \frac{\#A_2}{\#\Omega} = \frac{6 \cdot 6 \cdot 3}{9 \cdot 8 \cdot 7} = \frac{3}{14} \approx 0,2143.$$

Beispiel. Beim Skatspiel gibt es 32 Karten, vier davon sind Buben. Jeder Spieler erhält 10 Karten.

Sei A_4 das Ereignis, daß ein Spieler 4 Buben unter seinen 10 Karten findet.

Wir ziehen nacheinander 10 Karten aus den 32 Karten. Sei Ω die Menge der Ziehungsprotokolle. Also z.B.

(Herz-7, Karo-9, Kreuz-7, Pik-9, Pik-10, Kreuz-Dame, Pik-Dame, Pik-König, Pik-Bube, Karo-Dame) $\in \Omega$

Die Anzahlen der beim jeweiligen Ziehungsschritt noch verfügbaren Karten geben

$$\#\Omega = 32 \cdot 31 \cdot 30 \cdot 29 \cdot 28 \cdot 27 \cdot 26 \cdot 25 \cdot 24 \cdot 23.$$

Sei A_4 die Menge der Ziehungsprotokolle, bei denen 4 Buben gefunden werden. Wir haben $\#A_4$ zu bestimmen.

Bezeichnen wir mit B im Protokoll, daß ein Bube gezogen wurde. Bezeichnen wir mit N im Protokoll, daß ein Nicht-Bube gezogen wurde. Für die Stellen im Protokoll mit Buben gibt es dann die Möglichkeiten

(B, B, B, B, N, N, N, N, N, N), (B, B, B, N, B, N, N, N, N, N), ..., (N, N, N, N, N, N, N, B, B, B, B)

Das sind $\binom{10}{4}$ Möglichkeiten. Für jede dieser Möglichkeiten gibt es dann wiederum

$$4 \cdot 3 \cdot 2 \cdot 1 \cdot 28 \cdot 27 \cdot 26 \cdot 25 \cdot 24 \cdot 23$$

mögliche Protokolle. Also ist

$$P(A_4) = \frac{\binom{10}{4} \cdot 4 \cdot 3 \cdot 2 \cdot 1 \cdot 28 \cdot 27 \cdot 26 \cdot 25 \cdot 24 \cdot 23}{32 \cdot 31 \cdot 30 \cdot 29 \cdot 28 \cdot 27 \cdot 26 \cdot 25 \cdot 24 \cdot 23} = \frac{\binom{10}{4}}{\binom{32}{4}} = \frac{21}{3594} \approx 0,005840$$

die Wahrscheinlichkeit, vier Buben zu finden.

Darf man übrigens als Alleinspieler noch zwei weitere Karten aufnehmen, dann ist dabei entsprechend die Wahrscheinlichkeit

$$\frac{\binom{12}{4}}{\binom{32}{4}} = \frac{99}{7192} \approx 0,01377 ,$$

vier Buben zu finden.

Beispiel. Wir würfeln.

- (1) Sei A_1 das Ereignis, nach einem Wurf eine 6 zu würfeln. Es ist also

$$A_1 = \{1\} \subseteq \{1, 2, 3, 4, 5, 6\} = \Omega .$$

Die Wahrscheinlichkeit, nach einem Wurf eine 6 zu würfeln, ist also

$$P(A_1) = \frac{\#A_1}{\#\Omega} = \frac{1}{6} \approx 0,1667 .$$

- (2) Sei A_2 das Ereignis, nach zwei Würfeln mindestens eine 6 zu würfeln.

Hier gehen wir besser über das Gegenereignis $\bar{A}_2$, welches beinhaltet, nach zwei Würfeln keine 6 zu würfeln. Dafür gibt es $\#\bar{A}_2 = 5 \cdot 5 = 25$ Möglichkeiten, aus einer Ergebnismenge mit $\#\Omega = 6 \cdot 6 = 36$. Also ist

$$P(\bar{A}_2) = \frac{\#\bar{A}_2}{\#\Omega} = \frac{25}{36} .$$

Folglich ergibt sich die Wahrscheinlichkeit, bei zwei Würfeln mindestens eine 6 zu würfeln, zu

$$P(A_2) = 1 - P(\bar{A}_2) = 1 - \frac{25}{36} = \frac{11}{36} \approx 0,3056 .$$

- (3) Sei A_3 das Ereignis, nach drei Würfeln mindestens eine 6 zu würfeln. Analog zu (2) wird

$$P(A_3) = 1 - \frac{5^3}{6^3} = \frac{91}{216} \approx 0,4213 .$$

2.3 Bedingte Wahrscheinlichkeit und Pfadregel

Definition (Bedingte Wahrscheinlichkeit).

Seien $A, B \subseteq \Omega$ Ereignisse, mit $P(B) \neq 0$. Es heißt

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

die *bedingte Wahrscheinlichkeit* von A unter B .

Es gibt $P(A|B)$ die Wahrscheinlichkeit, daß A eintritt, unter Voraussetzung, daß B eintritt.

Beispiel (Alarmanlage). Wir wollen die Funktionalität einer Alarmanlage evaluieren.

Sei A das Ereignis “Alarm wurde ausgelöst”. Sei B das Ereignis “Ein Einbruch hat stattgefunden”.

Die Gegenereignisse sind $\bar{A}$, “Alarm wurde nicht ausgelöst”, und $\bar{B}$, “Ein Einbruch hat nicht stattgefunden”.

Wir erstellen eine *Vierfeldertafel* wie folgt.

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

Auf der Grundlage einer Versuchsreihe mit 1000 simulierten Verläufen einer Nacht ergab sich folgende Tafel.

P	A	$\bar{A}$	
B	0,08	0,02	0,1
$\bar{B}$	0,1	0,8	0,9
	0,18	0,82	1

Hierbei ist die Summe in den Zeilen immer die rechts stehende Zahl. Z.B. ist ja

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

Ferner ist die Summe in den Spalten immer die unten stehende Zahl. Z.B. ist ja

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

Wir entnehmen: Die Wahrscheinlichkeit für Fehlalarm beträgt $P(A \cap \bar{B}) = 0,1$. Die Wahrscheinlichkeit für ein Versagen der Alarmanlage beträgt $P(\bar{A} \cap B) = 0,02$.

Im Fall, daß ein Einbruch stattgefunden hat, war die bedingte Wahrscheinlichkeit, daß die Alarmanlage funktioniert,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0,08}{0,1} = 0,8 .$$

Das ist zufriedenstellend.

Im Fall, daß die Alarmanlage losging, war die Wahrscheinlichkeit, daß ein Fehlalarm vorlag, bei

$$P(\bar{B}|A) = \frac{P(A \cap \bar{B})}{P(A)} = \frac{0,1}{0,18} = \frac{5}{9} \approx 0,5556 .$$

Das sollte dem Hersteller zu denken geben.

Definition (Unabhängigkeit).

Seien $A, B \subseteq \Omega$ Ereignisse. Es heißen A und B *unabhängig*, falls

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

Falls A und B unabhängig sind und falls $P(B) \neq 0$ ist, folgt

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

Mit anderen Worten, ob das Ereignis B zur Bedingung gemacht wird oder auch nicht, ändert diesfalls nichts an der Wahrscheinlichkeit für das Ereignis A .

Falls A und B unabhängig sind und falls $P(A) \neq 0$ ist, folgt analog

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

Beispiel (Alarmanlage). Es sind im vorigen Beispiel die Ereignisse “Alarm wurde ausgelöst” und “Ein Einbruch hat stattgefunden” nicht unabhängig. Denn

$$P(A \cap B) = 0,08 \neq 0,18 \cdot 0,1 = P(A) \cdot P(B) .$$

Beispiel (Kugeln). Es befinden sich drei Kugeln in einem Gefäß, mit den Nummern 1, 2, 3.

Es sollen nacheinander zwei Kugeln gezogen werden. Für eine Ziehung einer Kugel sollen alle Kugeln dieselbe Wahrscheinlichkeit haben.

Es gibt zwei Möglichkeiten des Vorgehens.

- (1) Die Kugel wird nach dem ersten Zug wieder in das Gefäß zurückgelegt.

Es ist

$$\Omega = \{(1, 1), (1, 2), (1, 3), \\ (2, 1), (2, 2), (2, 3), \\ (3, 1), (3, 2), (3, 3)\}$$

Sei A das Ereignis, im ersten Zug eine 1 zu ziehen. Es ist $A = \{(1, 1), (1, 2), (1, 3)\}$ und also $P(A) = \frac{3}{9} = \frac{1}{3}$.

Sei B das Ereignis, im zweiten Zug eine 3 zu ziehen. Es ist $B = \{(1, 3), (2, 3), (3, 3)\}$ und also $P(B) = \frac{3}{9} = \frac{1}{3}$.

Dann ist $A \cap B$ das Ereignis, im ersten Zug eine 1 und im zweiten Zug eine 3 zu ziehen. Es ist $A \cap B = \{(1, 3)\}$ und also $P(A \cap B) = \frac{1}{9}$.

Die Ereignisse A und B sind unabhängig, da

$$P(A) \cdot P(B) = \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9} = P(A \cap B)$$

ist.

- (2) Die Kugel wird nach dem ersten Zug nicht wieder in das Gefäß zurückgelegt.

Es ist

$$\Omega = \left\{ \begin{array}{ll} (1, 2), (1, 3), \\ (2, 1), (2, 3), \\ (3, 1), (3, 2) \end{array} \right\}$$

Sei A das Ereignis, im ersten Zug eine 1 zu ziehen. Es ist $A = \{(1, 2), (1, 3)\}$ und also $P(A) = \frac{2}{6} = \frac{1}{3}$.

Sei B das Ereignis, im zweiten Zug eine 3 zu ziehen. Es ist $B = \{(1, 3), (2, 3)\}$ und also $P(B) = \frac{2}{6} = \frac{1}{3}$.

Dann ist $A \cap B$ das Ereignis, im ersten Zug eine 1 und im zweiten Zug eine 3 zu ziehen. Es ist $A \cap B = \{(1, 3)\}$ und also $P(A \cap B) = \frac{1}{6}$.

Die Ereignisse A und B sind nicht unabhängig, da

$$P(A) \cdot P(B) = \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9} \neq \frac{1}{6} = P(A \cap B)$$

ist.

Hier ist z.B. $P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{1/6}{1/3} = \frac{1}{2}$ die Wahrscheinlichkeit, beim zweiten Zug die Kugel 3 zu ziehen, unter der Bedingung, im ersten Zug die Kugel 1 gezogen zu haben. Insbesondere ist $P(B|A) = \frac{1}{2} \neq \frac{1}{3} = P(B)$.

Bemerkung (Pfadregel).

Sei $n \geq 1$. Seien $A_1, \dots, A_n \subseteq \Omega$. Dann ist

$$\begin{aligned} & P(A_1 \cap A_2 \cap \dots \cap A_n) \\ &= P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdot \dots \cdot P(A_n|A_1 \cap A_2 \cap \dots \cap A_{n-1}). \end{aligned}$$

Denn tatsächlich können wir durch Kürzen erreichen:

$$\begin{aligned} & P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdot \dots \cdot P(A_n|A_1 \cap A_2 \cap \dots \cap A_{n-1}) \\ &= P(A_1) \cdot \frac{P(A_1 \cap A_2)}{P(A_1)} \cdot \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)} \cdot \dots \cdot \frac{P(A_1 \cap A_2 \cap A_3 \cap \dots \cap A_n)}{P(A_1 \cap A_2 \cap \dots \cap A_{n-1})} \\ &= P(A_1 \cap A_2 \cap \dots \cap A_n). \end{aligned}$$

Beispiel (Streichhölzer ziehen).

Seien 4 Streichhölzer vorliegen. Sei unter diesen ein kurzes.

Es ziehen 4 Personen je ein Streichholz, ohne zu sehen, welches das kurze ist. Jeder behält nach dem Zug das Streichholz.

Wer das kurze Streichholz zieht, verliert.

Welche Position im Ablauf ist vorteilhaft? Die als erste Person, als zweite Person, als dritte Person oder als vierte Person?

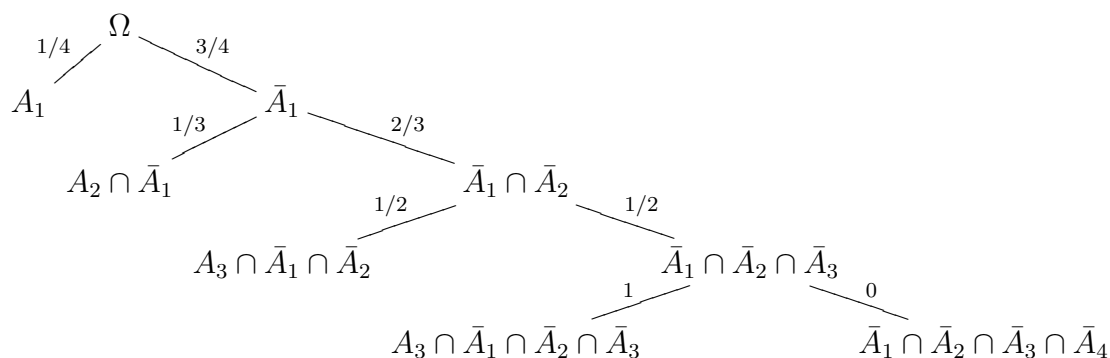
Sei A_1 das Ereignis, daß die erste Person ein kurzes Streichholz zieht. Dann ist $P(\bar{A}_1) = \frac{3}{4}$ und $P(A_1) = \frac{1}{4}$.

Sei A_2 das Ereignis, daß die zweite Person ein kurzes Streichholz zieht. Da dies nur eintritt, wenn die erste Person ein langes Streichholz gezogen hat und da danach noch 2 lange und ein kurzes Streichholz vorhanden sind, ist dann $P(\bar{A}_2|\bar{A}_1) = \frac{2}{3}$ und $P(A_2|\bar{A}_1) = \frac{1}{3}$.

Sei A_3 das Ereignis, daß die dritte Person ein kurzes Streichholz zieht. Da dies nur eintritt, wenn die erste und die zweite Person je ein langes Streichholz gezogen haben und da danach noch ein langes und ein kurzes Streichholz vorhanden sind, ist dann $P(\bar{A}_3|A_1 \cap A_2) = \frac{1}{2}$ und $P(A_3|A_1 \cap A_2) = \frac{1}{2}$.

Sei A_4 das Ereignis, daß die dritte Person ein kurzes Streichholz zieht. Da dies nur eintritt, wenn die erste und die zweite und die dritte Person je ein langes Streichholz gezogen haben und da danach noch ein kurzes Streichholz vorhanden ist, ist dann $P(\bar{A}_4|A_1 \cap A_2 \cap A_3) = 0$ und $P(A_4|A_1 \cap A_2 \cap A_3) = 1$.

Wir erstellen folgenden Baum.



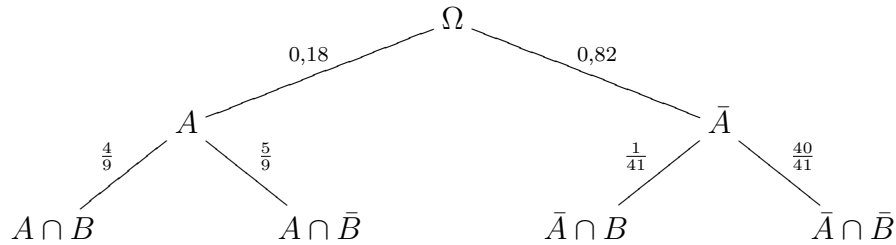
An den Ästen notieren wir die bedingten Wahrscheinlichkeiten. Um zur Wahrscheinlichkeit eines Ereignisses weiter unten im Baum zu kommen, multiplizieren wir gemäß Pfadregel die bedingten Wahrscheinlichkeiten entlang dem Pfad, der dorthin führt.

Es wird

$$\begin{aligned}
 P(A_1) &= P(A_1) &= \frac{1}{4} &= \frac{1}{4} \\
 P(A_2) &= P(A_2 \cap \bar{A}_1) &= \frac{3}{4} \cdot \frac{1}{3} &= \frac{1}{4} \\
 P(A_3) &= P(A_3 \cap \bar{A}_1 \cap \bar{A}_2) &= \frac{3}{4} \cdot \frac{2}{3} \cdot \frac{1}{2} &= \frac{1}{4} \\
 P(A_4) &= P(A_4 \cap \bar{A}_1 \cap \bar{A}_2 \cap \bar{A}_3) &= \frac{3}{4} \cdot \frac{2}{3} \cdot \frac{1}{2} \cdot 1 &= \frac{1}{4}
 \end{aligned}$$

Es ist also egal, ob man als erster, zweiter, dritter oder vierter zieht.

Beispiel (Alarmanlage). Zum obigen Beispiel mit Alarm und Einbruch kann man auch einen Baum erstellen, die bedingten Wahrscheinlichkeiten eintragen und dann die Pfadregel anwenden.



Auch hiermit ergibt sich z.B. $P(A \cap \bar{B}) = 0,18 \cdot \frac{5}{9} = 0,1$.

2.4 Zufallsgrößen

Definition. Eine Funktion

$$X : \Omega \rightarrow \mathbb{R} : t \mapsto X(t)$$

heißt *Zufallsgröße* oder *Zufallsvariable*.

Wir schreiben kurz

$$\begin{aligned} P(X = a) &:= P(\{t \in \Omega : X(t) = a\}) \\ P(X \leq a) &:= P(\{t \in \Omega : X(t) \leq a\}) \end{aligned}$$

usf., wobei $a \in \mathbb{R}$. Allgemein, ist $U \subseteq \mathbb{R}$, so schreiben wir kurz

$$P(X \in U) := P(\{t \in \Omega : X(t) \in U\}).$$

Beispiel. Wir würfeln mit 2 Würfeln. Die Ergebnismenge ist $\Omega = \{(a, b) : 1 \leq a, b \leq 6\}$.

Wir betrachten die Zufallsgröße X mit $X((a, b)) := a + b$ für $(a, b) \in \Omega$.

Es wird

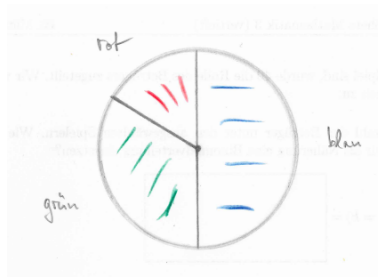
$$P(X = 7) = P(\{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}) = \frac{6}{36} = \frac{1}{6}.$$

Es wird

$$P(X \in \{7, 10\}) = P(X = 7) + P(X = 10) = \frac{6}{36} + \frac{3}{36} = \frac{1}{4}.$$

Beispiel. Wir betrachten ein Glücksrad, bei welchem Sektoren rot (r), grün (g) oder blau (b) eingefärbt sind. Es ist also $\Omega = \{r, g, b\}$.

Sei das Glücksrad so eingerichtet, daß $P(\{r\}) = \frac{1}{6}$, $P(\{g\}) = \frac{1}{3}$, $P(\{b\}) = \frac{1}{2}$.



Der Veranstalter setzt nun Gewinne bei den einzelnen Ergebnissen fest:

$$X(r) = 2, \quad X(g) = 0, \quad X(b) = -1.$$

Es ist

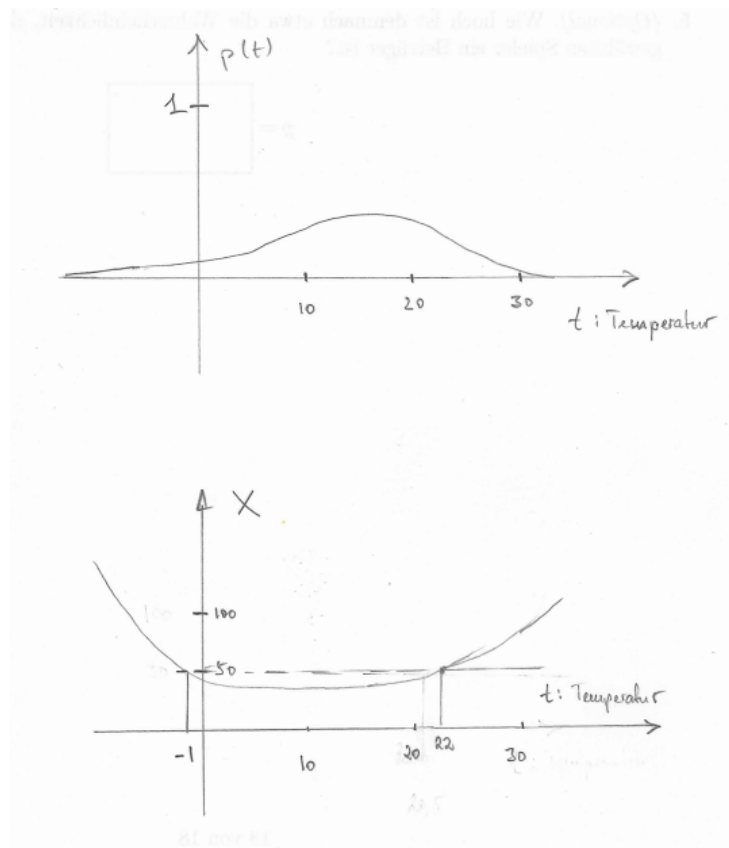
$$P(X \geq 0) = P(\{r, g\}) = \frac{1}{6} + \frac{1}{3} = \frac{1}{2}$$

die Wahrscheinlichkeit dafür, als Teilnehmer keinen Verlust zu erleiden.

Definition. Sei $\Omega \subseteq \mathbb{R}$. Diesenfalls heißt eine Funktion $p : \Omega \rightarrow \mathbb{R}$ eine *Wahrscheinlichkeitsdichte* für das Wahrscheinlichkeitsmaß P , falls für alle Ereignisse $A \subseteq \Omega$ gilt:

$$P(A) = \int_A p(t) \, dt.$$

Beispiel (graphisch). Sei die Wahrscheinlichkeitsdichte $p(t)$ für die Temperatur im März bekannt. Und seien die Betriebskosten X einer Maschine temperaturabhängig:



Laut Graphik wird

$$P(X \leq 50) = P([-1, 22]) = \int_{-1}^{22} p(t) \, dt .$$

Beispiel. Sei $\Omega = \mathbb{R}_{\geq 0}$. Sei die Wahrscheinlichkeitsdichte

$$p : \Omega \rightarrow \mathbb{R} : t \mapsto p(t) := e^{-t} .$$

Dann wird für $0 \leq a \leq b$

$$P([a, b]) = \int_a^b p(t) \, dt = \int_a^b e^{-t} \, dt = [-e^{-t}]_a^b = e^{-a} - e^{-b} .$$

Wir betrachten die Zufallsgröße Y mit $Y(t) = t^2$ auf Ω . Es wird z.B.

$$P(Y \leq \tfrac{1}{4}) = P(\{t \in \mathbb{R}_{\geq 0} : t^2 \leq \tfrac{1}{4}\}) = P([0, \tfrac{1}{2}]) = 1 - e^{-1/2} = 1 - \tfrac{1}{\sqrt{e}} .$$

2.5 Erwartungswert

Sei $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsgröße.

Definition (Erwartungswert).

Fall: Ω endlich. Der *Erwartungswert* der Zufallsgröße X ist

$$E(X) := \sum_{t \in \Omega} X(t) \cdot P(\{t\}) .$$

Indem wir die Summanden mit einem festen Wert a von $X(t)$ zusammenfassen, erhalten wir alternativ auch

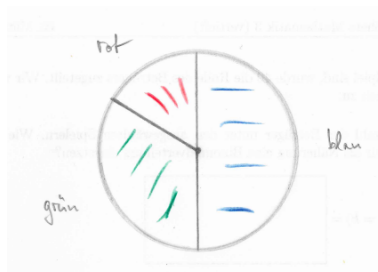
$$E(X) = \sum_{a \in X(\Omega)} a \cdot P(X = a) .$$

Fall: $\Omega \subseteq \mathbb{R}$ unendlich. Sei p eine Wahrscheinlichkeitsdichte für P .

Der *Erwartungswert* der Zufallsgröße X ist

$$E(X) := \int_{t \in \Omega} X(t) \cdot p(t) \, dt .$$

Beispiel (Glücksrad). Sei wieder $\Omega = \{r, g, b\}$ mit $P(r) = \frac{1}{6}$, $P(g) = \frac{1}{3}$, $P(b) = \frac{1}{2}$.



Sei der Gewinn wieder jeweils gesetzt auf:

$$X(r) = 2, \quad X(g) = 0, \quad X(b) = -1.$$

Der Erwartungswert für den Gewinn wird

$$E(X) = X(r) \cdot P(r) + X(g) \cdot P(g) + X(b) \cdot P(b) = 2 \cdot \frac{1}{6} + 0 \cdot \frac{1}{3} + (-1) \cdot \frac{1}{2} = -\frac{1}{6}.$$

Im langfristigen Mittel hat ein Spielteilnehmer also pro Spiel ein Verlust von $\frac{1}{6}$ zu erwarten. Der Veranstalter kann das Spiel also so anbieten, er wird langfristig Gewinn machen.

Beispiel. Sei $\Omega = \mathbb{R}_{\geq 0}$. Sei darauf die Wahrscheinlichkeitsdichte $p(t) = e^{-t}$ bekannt.

Sei darauf ferner die Zufallsgröße X durch $X(t) = 3t$ gegeben.

Der Erwartungswert wird

$$\begin{aligned} E(X) &= \int_{t \in \Omega} X(t) \cdot p(t) \, dt \\ &= \int_0^{+\infty} 3t \cdot e^{-t} \, dt. \end{aligned}$$

Wir rechnen mit β statt $+\infty$ und lassen im Nachgang β gegen $+\infty$ gehen.

$$\begin{aligned} \int_0^\beta 3t \cdot e^{-t} \, dt &= [3t \cdot (-e^{-t})]_0^\beta - \int_0^\beta 3 \cdot (-e^{-t}) \, dt \\ &= -3\beta \cdot e^{-\beta} + 3 \int_0^\beta e^{-t} \, dt \\ &= -3\beta \cdot e^{-\beta} + 3[-e^{-t}]_0^\beta \\ &= -3\beta \cdot e^{-\beta} - 3e^{-\beta} + 3 \end{aligned}$$

Für $\beta \rightarrow +\infty$ geht dies gegen

$$E(X) = \int_0^{+\infty} 3t \cdot e^{-t} \, dt = 3.$$

Bemerkung. Seien X und Y Zufallsgrößen. Seien $a, b \in \mathbb{R}$ konstant.

Sei $(aX + bY)(t) := aX(t) + bY(t)$ für $t \in \Omega$.

Es wird

$$E(aX + bY) = a E(X) + b E(Y).$$

Definition. Die Zufallsgrößen $X : \Omega \rightarrow \mathbb{R}$ und $Y : \Omega \rightarrow \mathbb{R}$ heißen *unabhängig*, wenn für $B, C \subseteq \Omega$ die Ereignisse

$$\{t \in \Omega : X(t) \in B\} \quad \text{und} \quad \{t \in \Omega : Y(t) \in C\}$$

unabhängig sind, d.h. wenn stets

$$P(X \in B \text{ und } Y \in C) = P(X \in B) \cdot P(Y \in C).$$

Ist Ω endlich, so genügt es, hierfür einelementige Teilmengen $B = \{b\}$ und $C = \{c\}$ heranzuziehen:

$$P(X = b \text{ und } Y = c) = P(X = b) \cdot P(Y = c)$$

soll für alle $b, c \in X(\Omega)$ gelten.

Bemerkung. Seien unabhängige Zufallsgrößen X und Y gegeben.

Sei $(X \cdot Y)(t) = X(t) \cdot Y(t)$ für $t \in \Omega$.

Es wird

$$E(X \cdot Y) = E(X) \cdot E(Y) .$$

Dies wollen wir im Fall Ω endlich nachrechnen. Es wird

$$\begin{aligned} E(X) \cdot E(Y) &= \left(\sum_{b \in X(\Omega)} b \cdot P(X = b) \right) \cdot \left(\sum_{c \in Y(\Omega)} c \cdot P(Y = c) \right) \\ &= \sum_{b \in X(\Omega), c \in Y(\Omega)} bc \cdot P(X = b) \cdot P(Y = c) \\ &\stackrel{X, Y \text{ unabhängig}}{=} \sum_{b \in X(\Omega), c \in Y(\Omega)} bc \cdot P(X = b \text{ und } Y = c) . \\ &= \sum_{b \in X(\Omega), c \in Y(\Omega)} bc \cdot P(\{t \in \Omega : X(t) = b \text{ und } Y(t) = c\}) \\ &= \sum_{b \in X(\Omega), c \in Y(\Omega)} bc \cdot \sum_{\substack{t \in \Omega \text{ mit} \\ X(t) = b \text{ und } Y(t) = c}} P(\{t\}) \\ &= \sum_{t \in \Omega} X(t) \cdot Y(t) \cdot P(\{t\}) \\ &= E(X \cdot Y) . \end{aligned}$$

Beispiel. Wir würfeln mit 2 Würfeln.

Die Ergebnismenge ist $\Omega = \{(\alpha, \beta) : 1 \leq \alpha, \beta \leq 6\}$.

Wir betrachten die Zufallsgröße X mit

$$X((\alpha, \beta)) := \begin{cases} 2 & \text{falls } \alpha = 6 \\ -1 & \text{sonst} \end{cases}$$

für $(\alpha, \beta) \in \Omega$. Wir betrachten die Zufallsgröße Y mit

$$Y((\alpha, \beta)) := \begin{cases} 3 & \text{falls } \beta = 6 \\ -1 & \text{sonst} \end{cases}$$

für $(\alpha, \beta) \in \Omega$.

Der Wert von X hängt nur vom ersten Würfel ab. Der Wert von Y hängt nur vom zweiten Würfel ab. Das legt nahe, daß X und Y unabhängig sind.

Tatsächlich ist

$$\begin{aligned} P(X = 2 \text{ und } Y = 3) &= \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6} = P(X = 2) \cdot P(Y = 3) \\ P(X = 2 \text{ und } Y = -1) &= \frac{5}{36} = \frac{1}{6} \cdot \frac{5}{6} = P(X = 2) \cdot P(Y = -1) \\ P(X = -1 \text{ und } Y = 3) &= \frac{5}{36} = \frac{5}{6} \cdot \frac{1}{6} = P(X = -1) \cdot P(Y = 3) \\ P(X = -1 \text{ und } Y = -1) &= \frac{25}{36} = \frac{5}{6} \cdot \frac{5}{6} = P(X = -1) \cdot P(Y = -1) . \end{aligned}$$

Und weitere Werte können X und Y nicht annehmen. Also sind die Zufallsgrößen X und Y unabhängig.

Für die Erwartungswerte ergibt sich

$$\begin{aligned} E(X) &= 2 \cdot \frac{1}{6} + (-1) \cdot \frac{5}{6} = -\frac{1}{2} \\ E(Y) &= 3 \cdot \frac{1}{6} + (-1) \cdot \frac{5}{6} = -\frac{1}{3} . \end{aligned}$$

Damit wird in der Tat

$$E(X \cdot Y) = 2 \cdot 3 \cdot \frac{1}{36} + (-1) \cdot 3 \cdot \frac{5}{36} + 2 \cdot (-1) \cdot \frac{5}{36} + (-1) \cdot (-1) \cdot \frac{25}{36} = \frac{1}{6} = E(X) \cdot E(Y) .$$

2.6 Varianz

Sei $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsgröße. Sei $\mu := E(X)$ ihr Erwartungswert.

Definition (Varianz).

Fall: Ω endlich. Die *Varianz* der Zufallsgröße X ist

$$\text{Var}(X) := E((X - \mu)^2) = \sum_{t \in \Omega} (X(t) - \mu)^2 \cdot P(\{t\}) .$$

Indem wir die Summanden mit einem festen Wert a von $X(t)$ zusammenfassen, erhalten wir alternativ auch

$$\text{Var}(X) = \sum_{a \in X(\Omega)} (a - \mu)^2 \cdot P(X = a) .$$

Fall: $\Omega \subseteq \mathbb{R}$ unendlich. Sei p eine Wahrscheinlichkeitsdichte für P .

Der *Erwartungswert* der Zufallsgröße X ist

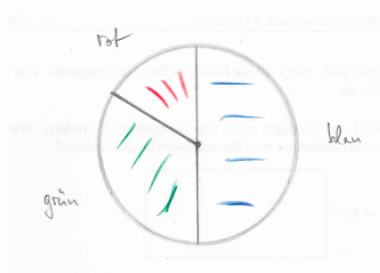
$$\text{Var}(X) := E((X - \mu)^2) = \int_{t \in \Omega} (X(t) - \mu)^2 \cdot p(t) \, dt .$$

In beiden Fällen schreibt man auch noch

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

für die *Standardabweichung*. Also $\sigma^2 = \text{Var}(X)$.

Beispiel (Glücksrad). Sei wieder $\Omega = \{r, g, b\}$ mit $P(r) = \frac{1}{6}$, $P(g) = \frac{1}{3}$, $P(b) = \frac{1}{2}$.



Sei der Gewinn wieder jeweils gesetzt auf:

$$X(r) = 2, \quad X(g) = 0, \quad X(b) = -1.$$

Der Erwartungswert für den Gewinn ergab sich zu $\mu = -\frac{1}{6}$. Die Varianz wird

$$\begin{aligned} \text{Var}(X) &= (X(r) - (-\frac{1}{6}))^2 \cdot P(r) + (X(g) - (-\frac{1}{6}))^2 \cdot P(g) + (X(b) - (-\frac{1}{6}))^2 \cdot P(b) \\ &= (2 + \frac{1}{6})^2 \cdot \frac{1}{6} + (0 + \frac{1}{6})^2 \cdot \frac{1}{3} + (-1 + \frac{1}{6})^2 \cdot \frac{1}{2} \\ &= \frac{41}{36}. \end{aligned}$$

Beispiel. Sei $\Omega = \mathbb{R}_{\geq 0}$.

Sei $p(t) := e^{-t}$ auf Ω die Wahrscheinlichkeitsdichte für das Wahrscheinlichkeitsmaß P .

Sei die Zufallsgröße X auf Ω gegeben durch $X(t) = 3t$.

Der Erwartungswert von X ist $\mu = E(X) = 3$; vgl. Beispiel in §2.5.

Die Varianz von X wird

$$\begin{aligned} \text{Var}(X) &= \int_0^{+\infty} (X(t) - \mu)^2 \cdot p(t) \, dt \\ &= \int_0^{+\infty} (3t - 3)^2 e^{-t} \, dt \\ &= \lim_{\beta \rightarrow +\infty} \int_0^{\beta} (3t - 3)^2 e^{-t} \, dt \\ &= \lim_{\beta \rightarrow +\infty} 9 \left([(t - 1)^2 (-e^{-t})]_0^{\beta} - \int_0^{\beta} 2(t - 1)(-e^{-t}) \, dt \right) \\ &= 9 \left(1 + \lim_{\beta \rightarrow +\infty} \int_0^{\beta} 2(t - 1) e^{-t} \, dt \right) \\ &= 9 \left(1 + \lim_{\beta \rightarrow +\infty} ([2(t - 1)(-e^{-t})]_0^{\beta} - \int_0^{\beta} 2(-e^{-t}) \, dt) \right) \\ &= 9 \left(1 - 2 + \lim_{\beta \rightarrow +\infty} \int_0^{\beta} 2 e^{-t} \, dt \right) \\ &= 9 \left(1 - 2 + \lim_{\beta \rightarrow +\infty} [2(-e^{-t})]_0^{\beta} \right) \\ &= 9(1 - 2 + 2) = 9. \end{aligned}$$

Bemerkung. Seien X und Y Zufallsgrößen. Sei $c \in \mathbb{R}$ eine Konstante.

Dann ist

$$\begin{aligned}\text{Var}(cX) &= c^2 \text{Var}(X) \\ \text{Var}(X + Y) &= \text{Var}(X) + \text{Var}(Y), \text{ falls } X \text{ und } Y \text{ unabhängig sind.}\end{aligned}$$

Verifizieren wir dies im endlichen Fall.

Es ist $E(cX) = cE(X)$. Also wird

$$\begin{aligned}\text{Var}(cX) &= \sum_{t \in \Omega} (cX(t) - E(cX))^2 \cdot P(\{t\}) \\ &= \sum_{t \in \Omega} (cX(t) - cE(X))^2 \cdot P(\{t\}) \\ &= c^2 \sum_{t \in \Omega} (X(t) - E(X))^2 \cdot P(\{t\}) \\ &= c^2 \text{Var}(X).\end{aligned}$$

Ferner wird

$$\begin{aligned}& -(\text{Var}(X) + \text{Var}(Y)) + \text{Var}(X + Y) \\ &= -(\text{Var}(X) + \text{Var}(Y)) + \sum_{t \in \Omega} ((X + Y)(t) - E(X + Y))^2 \cdot P(\{t\}) \\ &= -(\text{Var}(X) + \text{Var}(Y)) + \sum_{t \in \Omega} (X(t) + Y(t) - E(X) - E(Y))^2 \cdot P(\{t\}) \\ &= -(\text{Var}(X) + \text{Var}(Y)) + \sum_{t \in \Omega} ((X(t) - E(X)) + (Y(t) - E(Y)))^2 \cdot P(\{t\}) \\ &= -(\text{Var}(X) + \text{Var}(Y)) + \sum_{t \in \Omega} ((X(t) - E(X))^2 + 2(X(t) - E(X))(Y(t) - E(Y)) + (Y(t) - E(Y))^2) \cdot P(\{t\}) \\ &= 2 \sum_{t \in \Omega} (X(t) - E(X))(Y(t) - E(Y)) \cdot P(\{t\}) \\ &= 2 \sum_{t \in \Omega} (X(t)Y(t) - X(t)E(Y) - E(X)Y(t) + E(X)E(Y))P(\{t\}) \\ &= 2 \left(\sum_{t \in \Omega} X(t)Y(t)P(\{t\}) - (\sum_{t \in \Omega} X(t)P(\{t\}))E(Y) - E(X)(\sum_{t \in \Omega} Y(t)P(\{t\})) + E(X)E(Y)(\sum_{t \in \Omega} P(\{t\})) \right) \\ &= 2(E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y) \cdot 1) \\ &= 2(E(XY) - E(X)E(Y)) \\ &= 0,\end{aligned}$$

letzteres, da X und Y unabhängig sind.

2.7 Verteilungen

2.7.1 Binomialverteilung

Wir führen einen Versuch durch, mit Ergebnismenge Ω .

Dabei sei $A \subseteq \Omega$ ein Ereignis, *Erfolg* genannt. Das Gegenereignis $\bar{A}$ sei dementsprechend *kein Erfolg*.

Sei

$$\begin{aligned}P(A) &=: p \\ P(\bar{A}) &= 1 - p.\end{aligned}$$

Bei unserem Versuch ist also die Wahrscheinlichkeit für einen Erfolg gleich p , die Wahrscheinlichkeit für keinen Erfolg gleich $1 - p$.

Wir führen diesen Versuch nun n -mal durch, wobei $n \geq 1$. Die Ergebnismenge dafür ist also

$$\Omega^n = \{ (t_1, t_2, \dots, t_n) : t_1, t_2, \dots, t_n \in \Omega \}.$$

Sei auf Ω^n die Zufallsgröße X erklärt durch

$$\begin{aligned} X : \quad \Omega^n &\rightarrow \mathbb{R} \\ (t_1, t_2, \dots, t_n) &\mapsto X((t_1, t_2, \dots, t_n)) := \#\{i \in [1, n] : t_i \in A\}. \end{aligned}$$

Es ist also $X((t_1, t_2, \dots, t_n))$ die Anzahl der Erfolge unter den Einzelergebnissen $t_1, t_2, \dots, t_n$.

Bemerkung. Für $0 \leq k \leq n$ ist die Wahrscheinlichkeit für k Erfolge unter n Einzelergebnissen gegeben durch

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

Bemerkung. Es ist der Erwartungswert

$$E(X) = np.$$

Es ist die Varianz

$$\text{Var}(X) = np(1 - p).$$

Beispiel (5-mal würfeln).

Sei $A = \{6\} \subseteq \{1, 2, 3, 4, 5, 6\} = \Omega$ als Erfolg festgelegt.

Die Wahrscheinlichkeit für einen Erfolg bei einem solchen Einzelereignis ist $p = \frac{1}{6}$.

Wir wollen nun 5-mal würfeln. Sei also $n = 5$.

Wir erinnern an das Pascalsche Dreieck, aus dem wir die Binomialkoeffizienten entnehmen können.

$$\begin{array}{ccccccc} & & & & 1 & & \\ & & & 1 & & 1 & \\ & & 1 & & 2 & & 1 \\ & 1 & & 3 & & 3 & & 1 \\ 1 & & 4 & & 6 & & 4 & & 1 \\ & 1 & 5 & 10 & 10 & 5 & 1 \end{array}$$

So wird $((\binom{5}{0}), (\binom{5}{1}), (\binom{5}{2}), (\binom{5}{3}), (\binom{5}{4}), (\binom{5}{5})) = (1, 5, 10, 10, 5, 1)$.

Das ist manchmal schneller, als die Formel $\binom{n}{k} = \frac{n \cdot (n-1) \cdot \dots \cdot (n-k+1)}{k!}$ zu verwenden.

Die Wahrscheinlichkeiten verteilen sich nun wie folgt:

$$\begin{aligned}
 P(X = 5) &= \binom{5}{5} \cdot \left(\frac{1}{6}\right)^5 \cdot \left(\frac{5}{6}\right)^0 = \frac{1}{7776} \approx 0,000129 \quad (5 \text{ Sechsen}) \\
 P(X = 4) &= \binom{5}{4} \cdot \left(\frac{1}{6}\right)^4 \cdot \left(\frac{5}{6}\right)^1 = \frac{25}{7776} \approx 0,003215 \quad (4 \text{ Sechsen}) \\
 P(X = 3) &= \binom{5}{3} \cdot \left(\frac{1}{6}\right)^3 \cdot \left(\frac{5}{6}\right)^2 = \frac{250}{7776} \approx 0,032150 \quad (3 \text{ Sechsen}) \\
 P(X = 2) &= \binom{5}{2} \cdot \left(\frac{1}{6}\right)^2 \cdot \left(\frac{5}{6}\right)^3 = \frac{1250}{7776} \approx 0,160751 \quad (2 \text{ Sechsen}) \\
 P(X = 1) &= \binom{5}{1} \cdot \left(\frac{1}{6}\right)^1 \cdot \left(\frac{5}{6}\right)^4 = \frac{3125}{7776} \approx 0,401878 \quad (1 \text{ Sechs}) \\
 P(X = 0) &= \binom{5}{0} \cdot \left(\frac{1}{6}\right)^0 \cdot \left(\frac{5}{6}\right)^5 = \frac{3125}{7776} \approx 0,401878 \quad (0 \text{ Sechsen})
 \end{aligned}$$

Der Erwartungswert ist $E(X) = n \cdot p = 5 \cdot \frac{1}{6} = \frac{5}{6} \approx 0,8333$.

Die Varianz ist $\text{Var}(X) = np(1-p) = 5 \cdot \frac{1}{6} \cdot \frac{5}{6} = \frac{25}{36} \approx 0,6944$.

Beispiel (Werkstück-Herstellung wird kontrolliert).

Bei der Produktion eines Werkstücks sei es mit Wahrscheinlichkeit $p = 0,9$ ohne Fehler, also mit Wahrscheinlichkeit $1 - p = 0,1$ mit Fehler und damit Ausschuß.

Es werden $n = 20$ Werkstücke geprüft. Gebe die Zufallsgröße X die Anzahl der Werkstücke ohne Fehler an, die unter diesen 20 festgestellt wurden.

Der Erwartungswert für die Anzahl der Werkstücke ohne Fehler ist $E(X) = n \cdot p = 20 \cdot 0,9 = 18$.

Die Wahrscheinlichkeit für ≥ 18 Werkstücke ohne Fehler ist

$$P(X \geq 18) = \underbrace{\binom{20}{18} \cdot 0,9^{18} \cdot 0,1^2}_{\approx 0,2852} + \underbrace{\binom{20}{19} \cdot 0,9^{19} \cdot 0,1^1}_{\approx 0,2702} + \underbrace{\binom{20}{20} \cdot 0,9^{20} \cdot 0,1^0}_{\approx 0,1216} \approx 0,6769.$$

Die Wahrscheinlichkeit für ≤ 17 Werkstücke ohne Fehler ist

$$P(X \leq 17) = 1 - P(X \geq 18) \approx 0,3231.$$

Die Varianz ist

$$\text{Var}(X) = np(1-p) = 20 \cdot 0,9 \cdot (1 - 0,9) = 1,8.$$

2.7.2 Poisson-Verteilung

Die Poisson-Verteilung wird sich als ein Grenzfall der Binomialverteilung ergeben.

Hierzu machen wir ein Experiment.

Es werden $n = 5000$ Werkstücke geprüft.

Sei die Fehlerwahrscheinlichkeit bei jedem Werkstück gleich $p = \frac{1}{1000}$. Sei also p klein.

Wir wollen die Wahrscheinlichkeit für k fehlerhafte Werkstücke feststellen, wobei k klein sei im Vergleich zu n .

Es wird

$$\begin{aligned}
 P(X = k) &= \binom{n}{k} p^k (1-p)^{n-k} \\
 &= \frac{n \cdot (n-1) \cdot \dots \cdot (n-k+1)}{k!} \cdot p^k \cdot (1-p)^{n-k} \\
 &\approx \frac{n^k}{k!} \cdot p^k \cdot (1-p)^{n-k} \\
 &= \frac{n^k}{k!} \cdot p^k \cdot (1-p)^n \cdot (1-p)^{-k} \\
 &\approx \frac{n^k}{k!} \cdot p^k \cdot (1-p)^n \\
 &= \frac{(np)^k}{k!} \cdot ((1-p)^{\frac{1}{p}})^{np} \\
 &\approx \frac{(np)^k}{k!} \cdot e^{-np} .
 \end{aligned}$$

Für den letzten Schritt haben wir $(1-x)^{\frac{1}{x}} = \exp\left(\frac{\ln(1-x)}{x}\right)$ sowie

$$\lim_{x \rightarrow 0} \frac{\ln(1-x)}{x} \stackrel{\text{l'Hôpital}}{=} \lim_{x \rightarrow 0} \frac{\frac{1}{1-x} \cdot (-1)}{1} = -1$$

verwendet.

Also, mit $np = 5000 \cdot \frac{1}{1000} = 5$,

$$\begin{aligned}
 P(X = 0) &\approx \frac{5^0}{0!} e^{-5} \approx 0,006738 \\
 P(X = 1) &\approx \frac{5^1}{1!} e^{-5} \approx 0,033690 \\
 P(X = 2) &\approx \frac{5^2}{2!} e^{-5} \approx 0,084224 \\
 P(X = 3) &\approx \frac{5^3}{3!} e^{-5} \approx 0,140373896 \\
 P(X = 4) &\approx \frac{5^4}{4!} e^{-5} \approx 0,175468 \\
 P(X = 5) &\approx \frac{5^5}{5!} e^{-5} \approx 0,175468 \\
 P(X = 6) &\approx \frac{5^6}{6!} e^{-5} \approx 0,146223 .
 \end{aligned}$$

Zum Vergleich: Z.B. ist $P(X = 3) = \binom{5000}{3} \cdot \left(\frac{1}{1000}\right)^3 \cdot \left(1 - \frac{1}{1000}\right)^{5000-3} \approx 0,140359822$.

Die verwandte Näherung erklären wir nun zur Poisson-Verteilung:

Definition. Sei Ω ein Ergebnisraum. Sei $\lambda \in \mathbb{R}_{>0}$.

Eine Zufallsgröße $X : \Omega \rightarrow \mathbb{R}$ heißt *poissonverteilt* mit *Rate* λ , falls gilt:

$$P(X = k) = \begin{cases} \frac{\lambda^k}{k!} e^{-\lambda} & \text{für } k \in \mathbb{Z}_{\geq 0} \\ 0 & \text{sonst} \end{cases}$$

Wir erhalten dann den Erwartungswert

$$E(X) = \lambda$$

und die Varianz

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

Bemerkung. Es ist dann in der Tat

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

Beispiel. Die Wahrscheinlichkeit, sich innerhalb eines Jahres in einer Stadt mit einem seltenen Virus zu infizieren, sei nicht bekannt.

Es gebe die Zufallsgröße X an, wieviele Personen sich dort in diesem Zeitraum infizieren.

Sei aus langjährigen Statistiken die plausible Annahme $E(X) = 10$ gewonnen. Sei X als poissonverteilt angenommen.

Dann ist die Rate $\lambda = E(X) = 10$. Wir erhalten

$$\begin{aligned} P(X = 10) &= \frac{10^{10}}{10!} e^{-10} \approx 0,12511 \\ P(X = 0) &= \frac{10^0}{0!} e^{-10} \approx 4,54 \cdot 10^{-5} = 0,0000454 \\ P(X = 15) &= \frac{10^{15}}{15!} e^{-10} \approx 0,03472 . \end{aligned}$$

Die Wahrscheinlichkeit, daß sich dort in diesem Zeitraum zwischen 10 und 15 Personen infizieren, ergibt sich zu

$$P(10 \leq X \leq 15) = \sum_{k=10}^{15} \frac{10^k}{k!} e^{-10} \approx 0,493330 ,$$

liegt also nahe bei $0,5 = 50\%$.

Die Varianz ist übrigens $\text{Var}(X) = \lambda = 10$.

Beispiel. Die Werkstück-Produktion wird kontrolliert. Die Fehlerwahrscheinlichkeit bei jedem Werkstück sei nicht bekannt.

Sei aber für die Anzahl Y der fehlerhaften Werkstücke aus einer Tranche der Erwartungswert bekannt: $E(Y) = 2$.

Bei angenommener Poisson-Verteilung muß man sich die Mühe nicht machen, zu zählen, wieviele Werkstücke eine Tranche umfaßt.

Es ist dann die Rate $\lambda = E(Y) = 2$. Es wird

$$\begin{aligned} P(Y = 2) &= \frac{2^2}{2!} e^{-2} \approx 0,2707 \\ P(Y = 1) &= \frac{2^1}{1!} e^{-2} \approx 0,2707 \\ P(Y = 0) &= \frac{2^0}{0!} e^{-2} \approx 0,13523 \end{aligned}$$

Somit wird die Wahrscheinlichkeit für höchstens zwei fehlerhafte Werkstücke in der Tranche

$$P(Y \leq 2) = \sum_{k=0}^2 \frac{2^k}{k!} \approx 0,6767 .$$

Der Kollege, der im Beispiel von §2.7.1 für die Produktion dieser Maschine Binomialverteilung verwendet hat, hat gezählt, daß eine Tranche 20 Werkstücke umfaßt und bekam dort als Wahrscheinlichkeit für höchstens 2 fehlerhafte Werkstücke $P(X \geq 18) \approx 0,6769$. Was sich nicht stark unterscheidet vom Resultat $P(Y \leq 2) \approx 0,6767$, das hier mit Poisson-Verteilung berechnet wurde.

2.7.3 Stetige Verteilungen

Sei Ω ein Ergebnisraum. Sei $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsgröße.

Definition. Es heißt X eine *stetige* Zufallsgröße, falls es eine *Wahrscheinlichkeitsdichte*

$$p_X : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto p_X(x)$$

gibt, für welche für $B \subseteq \mathbb{R}$ sich für das Ereignis $\{t \in \Omega : X(t) \in B\}$ die Wahrscheinlichkeit

$$P(X \in B) = P(\{t \in \Omega : X(t) \in B\}) = \int_B p_X(x) \, dx$$

ergibt.

So ergibt sich speziell für $a, b \in \mathbb{R}$ mit $a \leq b$ die Wahrscheinlichkeit

$$P(a \leq X \leq b) = \int_a^b p_X(x) \, dx$$

dafür, daß die Zufallsgröße X einen Wert zwischen a und b annimmt.

Bemerkung. Es ist der Erwartungswert

$$\mu = \mu_X := E(X) = \int_{-\infty}^{+\infty} x \cdot p_X(x) \, dx .$$

Es ist die Varianz

$$\text{Var}(X) = E((X - \mu)^2) = \int_{-\infty}^{+\infty} (x - \mu)^2 \cdot p_X(x) \, dx .$$

Bemerkung. Ist $\Omega \subseteq \mathbb{R}$ und schreiben wir wieder

$$p : \Omega \rightarrow \mathbb{R}$$

für die Wahrscheinlichkeitsdichte für das Wahrscheinlichkeitsmaß P wie in §2.4, und ist $X : \Omega \rightarrow \mathbb{R}$ differenzierbar, dann wird

$$p(t) = p_X(X(t)) \cdot X'(t) .$$

Dann ist

$$\begin{aligned}
 \mu = E(X) &= \int_{-\infty}^{+\infty} x \cdot p_X(x) \, dx && \stackrel{x=X(t)}{=} \int_{\frac{dx}{dt}=X'(t)} X(t) \cdot p_X(X(t)) \cdot X'(t) \, dt \\
 & && = \int_{\Omega} X(t) \cdot p(t) \, dt \\
 \text{Var}(X) &= \int_{-\infty}^{+\infty} (x - \mu)^2 \cdot p_X(x) \, dx && \stackrel{x=X(t)}{=} \int_{\frac{dx}{dt}=X'(t)} (X(t) - \mu)^2 \cdot p_X(X(t)) \cdot X'(t) \, dt \\
 & && = \int_{\Omega} (X(t) - \mu)^2 \cdot p(t) \, dt ,
 \end{aligned}$$

wie in §2.5 und §2.6.

2.7.4 Exponentialverteilung

Sei Ω ein Ergebnisraum.

Sei $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsgröße.

Definition. Sei $\lambda \in \mathbb{R}_{>0}$.

Es heißt X *exponentialverteilt* mit *Rate* $\lambda > 0$, falls die zugehörige Wahrscheinlichkeitsdichte sich zu

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

ergibt.

Bemerkung. Sei X exponentialverteilt mit Rate $\lambda > 0$.

Wir berechnen den Erwartungswert zu

$$\begin{aligned}
 \mu_X &= E(X) \\
 &= \int_{-\infty}^{+\infty} x \cdot p_X(x) \, dx \\
 &= \int_0^{+\infty} x \cdot \lambda e^{-\lambda x} \, dx \\
 &= \lim_{\beta \rightarrow +\infty} \underbrace{\int_0^{\beta} x}_{u} \cdot \underbrace{\lambda e^{-\lambda x}}_{v'} \, dx \\
 &= \lim_{\beta \rightarrow +\infty} \left(\underbrace{[x \cdot (-e^{-\lambda x})]_0^{\beta}}_{\rightarrow 0} - \int_0^{\beta} 1 \cdot (-e^{-\lambda x}) \, dx \right) \\
 &= \lim_{\beta \rightarrow +\infty} \int_0^{\beta} e^{-\lambda x} \, dx \\
 &= \lim_{\beta \rightarrow +\infty} \left[\frac{1}{(-\lambda)} e^{-\lambda x} \right]_0^{\beta} \\
 &= \frac{1}{\lambda} .
 \end{aligned}$$

Wir berechnen die Varianz zu

$$\begin{aligned}
 \text{Var}(X) &= \int_{-\infty}^{+\infty} (x - \mu)^2 \cdot p_X(x) \, dx \\
 &= \int_0^{+\infty} (x - \frac{1}{\lambda})^2 \cdot \lambda e^{-\lambda x} \, dx \\
 &= \lim_{\beta \rightarrow +\infty} \int_0^\beta \underbrace{(x - \frac{1}{\lambda})^2}_u \cdot \underbrace{\lambda e^{-\lambda x}}_{v'} \, dx \\
 &= \lim_{\beta \rightarrow +\infty} \left(\underbrace{[(x - \frac{1}{\lambda})^2 \cdot (-e^{-\lambda x})]_0^\beta}_{\rightarrow \frac{1}{\lambda^2}} - \int_0^\beta 2(x - \frac{1}{\lambda}) \cdot (-e^{-\lambda x}) \, dx \right) \\
 &= \frac{1}{\lambda^2} + 2 \lim_{\beta \rightarrow +\infty} \int_0^\beta (x - \frac{1}{\lambda}) \cdot e^{-\lambda x} \, dx \\
 &= \frac{1}{\lambda^2} + 2 \lim_{\beta \rightarrow +\infty} \left(\underbrace{[(x - \frac{1}{\lambda}) \frac{1}{(-\lambda)} e^{-\lambda x}]_0^\beta}_{\rightarrow -\frac{1}{\lambda^2}} - \int_0^\beta 1 \cdot \frac{1}{(-\lambda)} e^{-\lambda x} \, dx \right) \\
 &= \frac{1}{\lambda^2} - 2 \frac{1}{\lambda^2} + 2 \lim_{\beta \rightarrow +\infty} \int_0^\beta \frac{1}{\lambda} e^{-\lambda x} \, dx \\
 &= \frac{1}{\lambda^2} - 2 \frac{1}{\lambda^2} + 2 \lim_{\beta \rightarrow +\infty} \left[\frac{1}{\lambda} \frac{1}{(-\lambda)} e^{-\lambda x} \right]_0^\beta \\
 &= \frac{1}{\lambda^2} - 2 \frac{1}{\lambda^2} + 2 \frac{1}{\lambda^2} \\
 &= \frac{1}{\lambda^2} .
 \end{aligned}$$

Beispiel. Von Maschinen einer bestimmten Bauart sei bekannt, daß pro Jahr 3% ausfallen und entsorgt werden müssen.

Sei $\Omega = \mathbb{R}_{\geq 0}$ die Menge der möglichen Ausfallzeitpunkte der betrachteten Maschine, gemessen in Jahren.

Sei $X(t) = t$ die Lebensdauer der betrachteten Maschine.

Es ist X exponentialverteilt mit Rate $\lambda = 0,03$. Es hat dann λ die Einheit $\frac{1}{\text{Jahr}}$.

(1) Der Erwartungswert für die Lebensdauer einer Maschine ist

$$E(X) = \frac{1}{\lambda} = \frac{1}{0,03} \approx 33,3 \text{ (Jahre)} .$$

(2) Die Wahrscheinlichkeit dafür, daß die Maschine zwischen ihrem 10ten und 20ten Betriebsjahr ausfällt, ist

$$\begin{aligned}
 P(10 \leq X \leq 20) &= \int_{10}^{20} p_X(x) \, dx \\
 &= \int_{10}^{20} \lambda e^{-\lambda x} \, dx \\
 &= \int_{10}^{20} 0,03 e^{-0,03x} \, dx \\
 &= [-e^{-0,03x}]_{10}^{20} \\
 &= -e^{-0,03 \cdot 20} + e^{-0,03 \cdot 10} \\
 &\approx 0,192 .
 \end{aligned}$$

(3) Die Wahrscheinlichkeit dafür, daß die Maschine mindestens während ihrer erwarteten Lebensdauer von $\approx 33,3$ Jahren läuft, ist

$$\begin{aligned}
 P(X \geq 33,3) &= \int_{100/3}^{+\infty} p_X(x) \, dx \\
 &= \int_{100/3}^{+\infty} 0,03 e^{-0,03x} \, dx \\
 &= \lim_{\beta \rightarrow +\infty} [-e^{-0,03x}]_{100/3}^{\beta} \\
 &= e^{-0,03 \cdot 100/3} \\
 &= e^{-1} \\
 &\approx 0,368 .
 \end{aligned}$$

Also weniger als $\frac{1}{2}$, was einen überraschen kann.

Das kann damit erklärt werden, daß eine Maschine mit einer gewissen Wahrscheinlichkeit sehr lange läuft und dieser Effekt dahingehend in den Erwartungswert für den Ausfallzeitpunkt einzurechnen ist, daß er nach dem Zeitpunkt liegt, zu dem die Maschine mit Wahrscheinlichkeit $\frac{1}{2}$ ausgefallen ist.

(4) Zu welchem Zeitpunkt kann man davon ausgehen, daß die Hälfte der Maschinen in der Fabrik ausgefallen ist, während die andere Hälfte noch läuft?

Genauer gesagt, für welchen Zeitpunkt t_0 ist die Wahrscheinlichkeit dafür, daß eine betrachtete Maschine erst zum Zeitpunkt t_0 oder später ausfällt, gleich $\frac{1}{2}$?

Zunächst ist

$$\begin{aligned}
 P(X \geq t_0) &= \int_{t_0}^{+\infty} p_X(x) \, dx \\
 &= \int_{t_0}^{+\infty} 0,03 e^{-0,03x} \, dx \\
 &= \lim_{\beta \rightarrow +\infty} [-e^{-0,03x}]_{t_0}^{\beta} \\
 &= e^{-0,03 \cdot t_0} .
 \end{aligned}$$

Aus $e^{-0,03 \cdot t_0} \stackrel{!}{=} \frac{1}{2}$ folgt $-0,03 \cdot t_0 \stackrel{!}{=} \ln(\frac{1}{2})$ und damit

$$t_0 = \frac{1}{(-0,03)} \cdot \ln(\frac{1}{2}) \approx 23,1 \text{ (Jahre)} .$$

2.7.5 Normalverteilung

Sei Ω ein Ergebnisraum.

Sei $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsgröße.

Definition. Sei $\sigma \in \mathbb{R}_{>0}$. Sei $\mu \in \mathbb{R}$. Es heißt X *normalverteilt* mit den Parametern μ und σ , falls für die zugehörige Wahrscheinlichkeitsdichte gilt:

$$p_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) .$$

Bemerkung. Die Maximalstelle von $p_X(x)$ befindet sich dort, wo der Exponentialfaktor gleich 1 ist, also bei $x = \mu$. Dort ist

$$p_X(\mu) = \frac{1}{\sigma\sqrt{2\pi}}.$$

Es ist $p'_X(x) = -\frac{1}{\sigma^3\sqrt{2\pi}}(x - \mu) \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$. Sodann hat

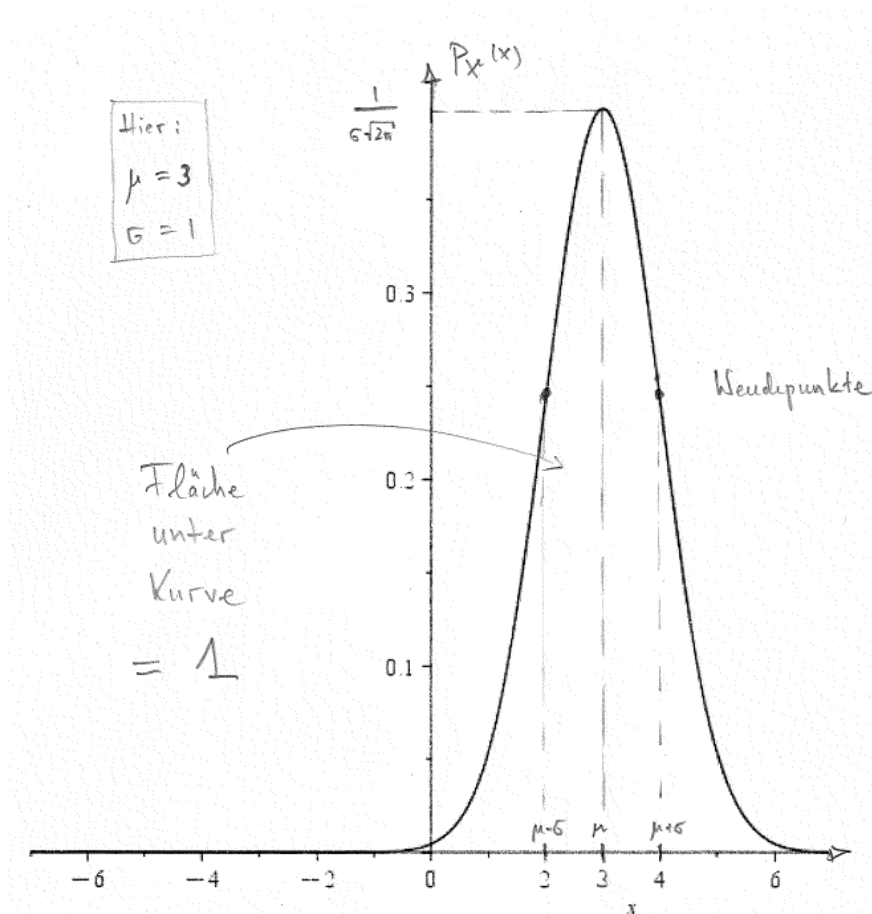
$$p''_X(x) = \frac{1}{\sigma^5\sqrt{2\pi}} \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \cdot (x - (\mu - \sigma)) \cdot (x - (\mu + \sigma))$$

die Nullstellen $\mu - \sigma$ und $\mu + \sigma$, an welchen sich die Wendepunkte der Funktion $p_X(x)$ befinden.

Der Vorfaktor $\frac{1}{\sigma\sqrt{2\pi}}$ resultiert aus der Bedingung, daß

$$\int_{-\infty}^{+\infty} p_X(x) \, dx = P(-\infty \leq X \leq +\infty) = 1$$

zu gelten hat, als Wahrscheinlichkeit dafür, daß X irgendeinen Wert zwischen $-\infty$ und $+\infty$ annimmt – was sicher ist. Wie dieser Vorfaktor aus dieser Bedingung resultiert, ist eine Rechnung, die ein Integral in zwei Variablen involviert, und die wir hier nicht durchführen.



Bemerkung. Sei X normalverteilt mit den Parametern μ und σ .

Der Erwartungswert von X berechnet sich zu

$$\begin{aligned}
 E(X) &= \int_{-\infty}^{+\infty} x \cdot p_X(x) \, dx \\
 &= \int_{-\infty}^{+\infty} x \cdot \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \, dx \\
 &\quad = 0 \text{ da Integrand ungerade} \\
 &= \overbrace{\int_{-\infty}^{+\infty} (x-\mu) \cdot \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \, dx} \\
 &\quad + \int_{-\infty}^{+\infty} \mu \cdot \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \, dx \\
 &= \mu \cdot 1 = \mu,
 \end{aligned}$$

letzteres, da $\int_{-\infty}^{+\infty} p_X(x) \, dx = 1$ gilt. Der Parameter μ trägt seinen Namen also zu Recht.

Sobald $E(X) = \mu$ bekannt ist, kann auch noch die Varianz zu

$$\text{Var}(X) = \int_{-\infty}^{+\infty} (x-\mu)^2 \cdot \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \, dx = \sigma^2$$

berechnet werden. Auch diese Rechnung führen wir nicht durch.

Die Standardabweichung ist also in der Tat gleich σ . Auch der Parameter σ trägt seinen Namen zu Recht.

Bemerkung. Seien $-\infty \leq z \leq w \leq +\infty$.

Die lineare Substitution $u = \frac{1}{\sigma}(x-\mu)$ und $\sigma \, du = dx$ gibt

$$\begin{aligned}
 P(\mu + z\sigma \leq X \leq \mu + w\sigma) &= \int_{\mu+z\sigma}^{\mu+w\sigma} p_X(x) \, dx \\
 &= \frac{1}{\sigma\sqrt{2\pi}} \int_{x=\mu+z\sigma}^{x=\mu+w\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \, dx \\
 &= \frac{1}{\sigma\sqrt{2\pi}} \int_{u=z}^{u=w} \exp\left(-\frac{u^2}{2}\right) \sigma \, du \\
 &= \frac{1}{\sqrt{2\pi}} \int_z^w \exp\left(-\frac{u^2}{2}\right) \, du.
 \end{aligned}$$

Das ist unabhängig von den Parametern μ und σ .

Für $z = -1$ und $w = 1$ wird

$$P(\mu - \sigma \leq X \leq \mu + \sigma) = \frac{1}{\sqrt{2\pi}} \int_{-1}^{+1} \exp\left(-\frac{u^2}{2}\right) \, du \approx 0,6827.$$

Für $z = -2$ und $w = 2$ wird

$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = \frac{1}{\sqrt{2\pi}} \int_{-2}^{+2} \exp\left(-\frac{u^2}{2}\right) \, du \approx 0,9545.$$

Für $z = -3$ und $w = 3$ wird

$$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = \frac{1}{\sqrt{2\pi}} \int_{-3}^{+3} \exp\left(-\frac{u^2}{2}\right) du \approx 0,9973 .$$

Beispiel. Eine Maschine soll 50mm lange Schrauben produzieren.

Normalverteilung mit $\mu = 50\text{mm}$ und Standardabweichung $\sigma = 2\text{mm}$ sei vorausgesetzt.

- (a) Wie groß ist die Wahrscheinlichkeit für eine Schraube von Länge $\geq 50\text{mm}$?
- (b) Wie groß ist die Wahrscheinlichkeit für eine Schraube von Länge $\geq 51\text{mm}$?

Es ist also $\mu = 50$ und $\sigma = 2$.

Zu (a). In vorstehender Bemerkung können wir $z = 0$ und $w = +\infty$ setzen und erhalten

$$P(X \geq 50) = P(50 \leq X \leq +\infty) = \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} \exp\left(-\frac{u^2}{2}\right) du = \frac{1}{2} ,$$

da der Graph des Integranden achsensymmetrisch zu $u = 0$ ist und da die Wahrscheinlichkeit des sicheren Ereignisses gleich $1 = P(-\infty \leq X \leq +\infty) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left(-\frac{u^2}{2}\right) du$ ist.

Zu (b). Es ist $51 = 50 + \frac{1}{2} \cdot 2 = \mu + \frac{1}{2} \cdot \sigma$. Also können wir in vorstehender Bemerkung $z = \frac{1}{2}$ und $w = +\infty$ setzen und erhalten

$$P(X \geq 51) = P(51 \leq X \leq +\infty) = \frac{1}{\sqrt{2\pi}} \int_{1/2}^{+\infty} \exp\left(-\frac{u^2}{2}\right) du \stackrel{\text{Tabelle}}{\approx} 0,3085 .$$

2.7.6 Normalverteilung als Grenzfall der Binomialverteilung

Satz (de Moivre-Laplace). Sei $X = X_n$ eine Binomialverteilung mit Erfolgswahrscheinlichkeit p des Einzelversuchs und Versuchsanzahl n . Wie in §2.7.1 ist dann für $0 \leq k \leq n$ die Wahrscheinlichkeit für k erfolgreiche Versuche

$$P(X_n = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} .$$

Ferner ist der Erwartungswert

$$\mu_n = E(X_n) = np$$

und die Standardabweichung

$$\sigma_n = \sqrt{\text{Var}(X_n)} = \sqrt{np(1-p)} ,$$

in Abhängigkeit von n . Nun wollen wir $n \rightarrow \infty$ laufen lassen.

Sei $-\infty \leq a \leq b \leq +\infty$. Der Satz von de Moivre-Laplace besagt:

$$\lim_{n \rightarrow \infty} P(\mu_n + a\sigma_n \leq X_n \leq \mu_n + b\sigma_n) = \frac{1}{\sqrt{2\pi}} \int_a^b \exp\left(-\frac{u^2}{2}\right) du.$$

In diesem Sinne wird die Binomialverteilung in der Grenze zur Normalverteilung.

Man beachte auch noch:

$$\begin{aligned} P(a \leq \frac{1}{\sigma_n}(X_n - \mu_n) \leq b) &= P(\mu_n + a\sigma_n \leq X_n \leq \mu_n + b\sigma_n) \\ &= P(np + a\sqrt{np(1-p)} \leq X_n \leq np + b\sqrt{np(1-p)}) . \end{aligned}$$

Bemerkung. Praktische Anwendung findet der Satz von de Moivre-Laplace zur approximativen Berechnung der Wahrscheinlichkeiten bei der Binomialverteilung wie folgt.

Hat man $P(k \leq X_n \leq \ell)$ anzunähern, mit ganzen Zahlen k und ℓ , so bestimme man zunächst $a, b \in \mathbb{R}$ mit

$$\begin{aligned} \mu_n + a\sigma_n &= k - 0,5 \\ \mu_n + b\sigma_n &= \ell + 0,5 , \end{aligned}$$

um den Diskretisierungsfehler auszugleichen.

Dann wird

$$P(k \leq X_n \leq \ell) \approx \frac{1}{\sqrt{2\pi}} \int_a^b \exp\left(-\frac{u^2}{2}\right) du$$

für große n . Näherungen für die Werte des Integrals auf der rechten Seite werden dabei Tabellen entnommen.

Beispiel. Eine Maschine arbeite mit Ausschußwahrscheinlichkeit $p = 0,1$.

Mit welcher Wahrscheinlichkeit produziert die Maschine bei $n = 1000$ Stück zwischen 50 und 80 Stück Ausschuß?

- (1) Man beantworte dies mittels Binomialverteilung.
- (2) Man beantworte dies näherungsweise mittels Normalverteilung.
- (3) Man beantworte dies näherungsweise mittels Poissonverteilung.

(1) Es wird

$$P(50 \leq X_n \leq 80) = \sum_{\ell=50}^{80} \binom{1000}{\ell} \cdot 0,1^\ell \cdot 0,9^{1000-\ell} \approx 0,017611 .$$

(2) Wir haben a und b zu bestimmen.

Es ist $\mu_n = np = 1000 \cdot 0,1 = 100$ und $\sigma_n = \sqrt{np(1-p)} = \sqrt{1000 \cdot 0,1 \cdot 0,9} \approx 9,4868$.

Aus

$$a\sigma_n + \mu_n = 50 - 0,5$$

folgt

$$a = \frac{50 - 0,5 - \mu_n}{\sigma_n} = \frac{50 - 0,5 - np}{\sqrt{np(1-p)}} = \frac{50 - 0,5 - 100}{\sqrt{1000 \cdot 0,1 \cdot 0,9}} \approx -5,373167.$$

Aus

$$b\sigma_n + \mu_n = 80 + 0,5$$

folgt

$$b = \frac{80 + 0,5 - \mu_n}{\sigma_n} = \frac{80 + 0,5 - np}{\sqrt{np(1-p)}} = \frac{80 + 0,5 - 100}{\sqrt{1000 \cdot 0,1 \cdot 0,9}} \approx -2,055480.$$

Mit obiger Bemerkung folgt nun

$$P(50 \leq X_n \leq 80) \approx \frac{1}{\sqrt{2\pi}} \int_a^b \exp(-u^2/2) \, du \approx 0,019916.$$

Das ist eine ordentliche Näherung an das Ergebnis aus (1).

(3) Bei Verwendung von Poissonverteilung als Näherung ist $\lambda = \mu_n = np = 100$. Es wird

$$P(50 \leq X_n \leq 80) \approx \sum_{k=50}^{80} \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=50}^{80} \frac{100^k}{k!} e^{-100} \approx 0,022649.$$

Das ist eine grobe Näherung an das Ergebnis aus (1).

Beispiel. (Münzwurf mit Normalverteilung).

Wir werfen eine Münze $n = 40$ mal. Es ist $p = 0,5$.

Mit welcher Wahrscheinlichkeit erhalten wir 20 mal "Zahl" ?

(1) Man beantworte dies mittels Binomialverteilung.

(2) Man beantworte dies näherungsweise mittels Normalverteilung.

(1) Es wird

$$P(X = 20) = \binom{40}{20} \cdot 0,5^{20} \cdot 0,5^{40-20} \approx 0,123537.$$

(2) Es ist $P(X_n = 20) = P(20 \leq X_n \leq 20)$. Dafür haben wir a und b zu bestimmen.

Es ist $\mu_n = np = 40 \cdot 0,5 = 20$ und $\sigma_n = \sqrt{np(1-p)} = \sqrt{40 \cdot 0,5 \cdot 0,5} = \sqrt{10}$.

Aus

$$a\sigma_n + \mu_n = 20 - 0,5$$

folgt

$$a = \frac{20 - 0,5 - \mu_n}{\sigma_n} = \frac{-0,5}{\sqrt{10}} \approx -0,158111.$$

Aus

$$b\sigma_n + \mu_n = 20 + 0,5$$

folgt

$$b = \frac{20+0,5-\mu_n}{\sigma_n} = \frac{0,5}{\sqrt{10}} \approx 0,158111 .$$

Mit obiger Bemerkung folgt nun

$$P(X_n = 20) = P(20 \leq X_n \leq 20) \approx \frac{1}{\sqrt{2\pi}} \int_a^b \exp(-u^2/2) \, du \approx 0,125633 .$$

Das ist eine ordentliche Näherung an das Ergebnis aus (1).

Kapitel 3

Induktive Statistik

3.1 Schätzung von Erwartungswert und Varianz

Sei $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsgröße.

Sei $\mu = E(X)$ der Erwartungswert. Sei $\sigma^2 = \text{Var}(X) = E((X - \mu)^2)$ die Varianz.

Um von statistischen Daten auf diese Parameter zu kommen, hat man zu schätzen.

3.1.1 Schätzung des Erwartungswerts

Sei eine Meßreihe $x_1, x_2, \dots, x_n$ für X gegeben.

Der Erwartungswert von X wird geschätzt zu

$$\mu_{\text{geschätzt}} := \frac{1}{n}(x_1 + x_2 + \dots + x_n) .$$

Seien $X_1, X_2, \dots, X_n$ unabhängige und wie X verteilte Zufallsgrößen.

Mit anderen Worten, es sind $X_1, X_2, \dots, X_n$ Kopien von X .

Insbesondere ist $E(X_i) = E(X) = \mu$ für $1 \leq i \leq n$.

Der zugehörige *Schätzer* ist

$$\bar{X}_n := \frac{1}{n}(X_1 + X_2 + \dots + X_n) .$$

Wir testen die Qualität dieses Schätzers. Es wird

$$E(\bar{X}_n) = E\left(\frac{1}{n}(X_1 + \dots + X_n)\right) = \frac{1}{n}(E(X_1) + \dots + E(X_n)) = \frac{1}{n} \cdot n\mu = \mu .$$

Man sagt: Der Schätzer $\bar{X}_n$ für den Erwartungswert μ ist *erwartungstreu*. Damit ist er geeignet.

3.1.2 Schätzung der Varianz

Sei eine Meßreihe $x_1, x_2, \dots, x_n$ für X gegeben.

Die Varianz von X wird geschätzt zu

$$\sigma_{\text{geschätzt}}^2 := \frac{1}{n-1} ((x_1 - \mu_{\text{geschätzt}})^2 + (x_2 - \mu_{\text{geschätzt}})^2 + \dots + (x_n - \mu_{\text{geschätzt}})^2) .$$

Die *Bessel-Korrektur*, statt des Vorfaktors $\frac{1}{n}$ vielmehr den Vorfaktor $\frac{1}{n-1}$ zu verwenden, ist nötig, um auch hier einen erwartungstreuen Schätzer zu bekommen. Würde man hier den Vorfaktor $\frac{1}{n}$ aus der Varianz der Meßreihe wie in §1.2 verwenden, dann würde man die Varianz der Zufallsgröße unterschätzen.

Seien $X_1, X_2, \dots, X_n$ unabhängige und wie X verteilte Zufallsgrößen.

Mit anderen Worten, es sind $X_1, X_2, \dots, X_n$ Kopien von X .

Insbesondere ist $E(X_i) = E(X) = \mu$ für $1 \leq i \leq n$.

Ferner ist $E((X_i - \mu)^2) = \text{Var}(X_i) = \text{Var}(X) = \sigma^2$ für $1 \leq i \leq n$.

Der zugehörige *Schätzer* ist

$$S_n^2 := \frac{1}{n-1} ((X_1 - \bar{X}_n)^2 + (X_2 - \bar{X}_n)^2 + \dots + (X_n - \bar{X}_n)^2) .$$

Wir testen die Qualität dieses Schätzers. Es wird

$$\begin{aligned} & (n-1) \cdot E(S_n^2) \\ = & (n-1) \cdot E\left(\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2\right) \\ = & \sum_{i=1}^n E((X_i - \bar{X}_n)^2) \\ = & \sum_{i=1}^n E(((X_i - \mu) - (\bar{X}_n - \mu))^2) \\ = & \sum_{i=1}^n E((X_i - \mu)^2 - 2(X_i - \mu)(\bar{X}_n - \mu) + (\bar{X}_n - \mu)^2) \\ = & \sum_{i=1}^n E((X_i - \mu)^2) - 2 \sum_{i=1}^n E((X_i - \mu)(\bar{X}_n - \mu)) + \sum_{i=1}^n E((\bar{X}_n - \mu)^2) \\ = & \sum_{i=1}^n \sigma^2 - 2 E((\sum_{i=1}^n (X_i - \mu))(\bar{X}_n - \mu)) + \sum_{i=1}^n E((\bar{X}_n - \mu)^2) \\ = & n\sigma^2 - 2 E((n\bar{X}_n - n\mu)(\bar{X}_n - \mu)) + nE((\bar{X}_n - \mu)^2) \\ = & n\sigma^2 - 2n E((\bar{X}_n - \mu)(\bar{X}_n - \mu)) + nE((\bar{X}_n - \mu)^2) \\ = & n\sigma^2 - nE((\bar{X}_n - \mu)^2) \\ = & n\sigma^2 - n \text{Var}(\bar{X}_n) \\ = & n\sigma^2 - \frac{1}{n} \text{Var}(\sum_{i=1}^n X_i) \\ \stackrel{\text{Unabh. der } X_i}{=} & n\sigma^2 - \frac{1}{n} \sum_{i=1}^n \text{Var}(X_i) \\ \stackrel{\text{Bem. aus §2.6}}{=} & n\sigma^2 - \frac{1}{n} \sum_{i=1}^n \sigma^2 \\ = & n\sigma^2 - \frac{1}{n} n\sigma^2 \\ = & (n-1)\sigma^2 . \end{aligned}$$

Damit ist

$$E(S_n^2) = \sigma^2 .$$

Man sagt: Der Schätzer S_n^2 für die Varianz σ^2 ist *erwartungstreu*. Damit ist er geeignet.

Beispiel. Wir haben die Meßreihe

$$x_1 = 10, \quad x_2 = 30, \quad x_3 = 40, \quad x_4 = 30$$

für die Zufallsgröße X vorliegen.

Der Schätzer $\bar{X}_4$ gibt den geschätzten Erwartungswert

$$\mu_{\text{geschätzt}} = \frac{1}{4}(10 + 30 + 40 + 30) = 27,5 .$$

Der Schätzer S_4^2 gibt die geschätzte Varianz

$$\sigma_{\text{geschätzt}}^2 = \frac{1}{4-1}((10-27,5)^2 + (30-27,5)^2 + (40-27,5)^2 + (30-27,5)^2) = \frac{1}{3} \cdot 475 \approx 158,33 .$$

Somit wird die geschätzte Standardabweichung

$$\sigma_{\text{geschätzt}} = \sqrt{\frac{1}{3} \cdot 475} \approx 12,5831 .$$

3.2 Parameterschätzung bei Verteilungen bekannter Art

3.2.1 Parameterschätzung bei Binomialverteilung

Sei X eine Binomialverteilung mit Erfolgswahrscheinlichkeit bei Einzelversuch p .

Seien aus einer Meßreihe Versuchsanzahl n und Erfolgsanzahl k bekannt, aber p unbekannt.

Mit anderen Worten, die Meßreihe mit n Versuchen ergab k Erfolge. Wir wollen p schätzen.

Sei

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

Sei $p_{\text{geschätzt}}$ die Maximalstelle von $L(p)$. Das nennt man auch den *Maximum-Likelihood-Ansatz*.

Ist $k = 0$, dann ist $L(p) = (1-p)^n$, also $p_{\text{geschätzt}} = 0$.

Ist $k = n$, dann ist $L(p) = p^n$, also $p_{\text{geschätzt}} = 1$.

Ist $1 \leq k \leq n-1$, dann rechnen wir dazu die kritische Stelle von $L(p)$ aus, bei welcher $0 < p < 1$ ist. Es wird

$$\begin{aligned} L'(p) &= \binom{n}{k} (kp^{k-1}(1-p)^{n-k} + p^k(n-k)(1-p)^{n-k-1} \cdot (-1)) \\ &= \binom{n}{k} p^{k-1}(1-p)^{n-k-1} (k(1-p) + p(n-k) \cdot (-1)) \\ &= \binom{n}{k} p^{k-1}(1-p)^{n-k-1} (k - kp - pn + pk) \\ &= \binom{n}{k} p^{k-1}(1-p)^{n-k-1} (k - pn) . \end{aligned}$$

Die Nullstelle ist bei $p_{\text{geschätzt}} = \frac{k}{n}$.

Für $0 \leq k \leq n$ ist also

$$p_{\text{geschätzt}} = \frac{k}{n} .$$

Beispiel. Ein Glücksrad, das wir nicht sehen können, liefert bei einer Versuchsreihe mit $n = 10$ Durchläufen insgesamt 3 Erfolge.

Wir schätzen die Erfolgswahrscheinlichkeit auf

$$p_{\text{geschätzt}} = \frac{k}{n} = \frac{3}{10} = 0,3 .$$

Auf diese Idee kann man auch ohne den Maximum-Likelihood-Ansatz kommen.

3.2.2 Parameterschätzung bei Poissonverteilung

Sei $X : \Omega \rightarrow \mathbb{R}$ eine Poisson-verteilte Zufallsgröße.

Es ist

$$P(X = k) = \begin{cases} \frac{\lambda^k}{k!} e^{-\lambda} & \text{falls } k \in \mathbb{Z}_{\geq 0} \\ 0 & \text{sonst.} \end{cases}$$

Anhand einer Meßreihe

$$k_1, \quad k_2, \quad \dots, \quad k_n$$

wollen wir die Rate λ schätzen.

Da $\lambda = \mu = E(X)$ ist, nehmen wir den Schätzer für den Erwartungswert aus §3.1.1:

$$\lambda_{\text{geschätzt}} := \frac{1}{n}(k_1 + k_2 + \dots + k_n) .$$

Beispiel. Im ersten Beispiel aus §2.7.2 hieß es:

Die Wahrscheinlichkeit, sich innerhalb eines Jahres in einer Stadt mit einem seltenen Virus zu infizieren, sei nicht bekannt.

Es gebe die Zufallsgröße X an, wieviele Personen sich dort in diesem Zeitraum infizieren.

Sei aus langjährigen Statistiken die plausible Annahme $E(X) = 10$ gewonnen. Sei X als poissonverteilt angenommen.

Wie wurde die plausible Annahme $E(X) = 10$ gewonnen?

Aus der Statistik wurde entnommen, daß in den letzten 5 Jahren die Infektionszahlen wie folgt waren:

Jahr 1:	8	Infizierte
Jahr 2:	10	Infizierte
Jahr 3:	11	Infizierte
Jahr 4:	11	Infizierte
Jahr 5:	10	Infizierte

Es wurde also

$$E(X) = \lambda_{\text{geschätzt}} = \frac{1}{5}(8 + 10 + 11 + 11 + 10) = 10$$

angenommen.

3.2.3 Parameterschätzung bei Normalverteilung

Sei $X : \Omega \rightarrow \mathbb{R}$ eine normalverteilte Zufallsgröße.

Sowohl der Erwartungswert $\text{Var}(X) = \mu$ als auch die Varianz $\text{Var}(X) = \sigma^2$ seien uns unbekannt.

Es liege eine Meßreihe

$$x_1, x_2, \dots, x_n$$

für X vor.

Wir schätzen den Erwartungswert wie in §3.1.1 zu

$$\mu_{\text{geschätzt}} := \frac{1}{n}(x_1 + x_2 + \dots + x_n).$$

Wir schätzen die Varianz wie in §3.1.2 zu

$$\sigma_{\text{geschätzt}}^2 := \frac{1}{n-1}((x_1 - \mu_{\text{geschätzt}})^2 + (x_2 - \mu_{\text{geschätzt}})^2 + \dots + (x_n - \mu_{\text{geschätzt}})^2).$$

Es wird also die *Bessel-Korrektur* berücksichtigt: als Vorfaktor nehmen wir $\frac{1}{n-1}$ statt $\frac{1}{n}$.

Beispiel. Wir haben die Meßreihe

$$x_1 = 10, x_2 = 30, x_3 = 40, x_4 = 30$$

für die Zufallsgröße X vorliegen, wie im Beispiel von §3.1.2.

Dort ergaben sich

$$\begin{aligned}\mu_{\text{geschätzt}} &= 27,5 \\ \sigma_{\text{geschätzt}} &= \sqrt{475/3} \approx 12,5831.\end{aligned}$$

Wollen wir auf der Grundlage dieser Schätzung die Wahrscheinlichkeit

$$P(25 \leq X \leq 35)$$

bestimmen, so setzen wir wie in §2.7.5 an:

$$P(\mu_{\text{geschätzt}} + z\sigma_{\text{geschätzt}} \leq X \leq \mu_{\text{geschätzt}} + w\sigma_{\text{geschätzt}}) = \frac{1}{\sqrt{2\pi}} \int_z^w \exp\left(-\frac{u^2}{2}\right) du$$

Wir müssen z und w passend bestimmen.

Aus

$$\mu_{\text{geschätzt}} + z\sigma_{\text{geschätzt}} = 25$$

ergibt sich

$$z = \frac{25 - \mu_{\text{geschätzt}}}{\sigma_{\text{geschätzt}}} \approx -0,1987$$

und

$$w = \frac{35 - \mu_{\text{geschätzt}}}{\sigma_{\text{geschätzt}}} \approx 0,5960 .$$

Somit wird

$$P(25 \leq X \leq 35) \approx \frac{1}{\sqrt{2\pi}} \int_z^w \exp\left(-\frac{u^2}{2}\right) du \approx 0,3032 .$$

3.2.4 Parameterschätzung bei Exponentialverteilung

Sei $X : \Omega \rightarrow \mathbb{R}$ eine exponentialverteilte Zufallsgröße mit Rate λ . Sei also

$$p_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{für } x \geq 0 \\ 0 & \text{für } x < 0 . \end{cases}$$

Dann ist, für $0 \leq a \leq b \leq +\infty$,

$$P(a \leq X \leq b) = \int_a^b p_X(x) dx = \int_a^b \lambda e^{-\lambda x} dx .$$

Sei die Rate λ nicht bekannt.

Es liege eine Meßreihe $x_1, x_2, \dots, x_n$ für X vor.

Wir schätzen die Rate zu

$$\lambda_{\text{geschätzt}} = \frac{1}{\frac{1}{n}(x_1 + x_2 + \dots + x_n)} .$$

Beispiel. Von Maschinen einer bestimmten Bauart seien folgende Betriebsdauern bis zum Ausfall festgestellt worden.

$$\begin{aligned} x_1 &= 5 \text{ Jahre} \\ x_2 &= 9 \text{ Jahre} \\ x_3 &= 12 \text{ Jahre} \\ x_4 &= 10 \text{ Jahre} \end{aligned}$$

Eine Maschine dieser Bauart wird neu in Betrieb genommen. Mit welcher Wahrscheinlichkeit läuft sie mindestens 12 Jahre?

Wir schätzen die Rate auf

$$\lambda_{\text{geschätzt}} = \frac{1}{\frac{1}{4}(5 + 9 + 12 + 10)} = \frac{1}{9}.$$

Damit wird

$$\begin{aligned} P(X \geq 12) &= P(12 \leq X \leq +\infty) \\ &= \int_{12}^{+\infty} \lambda_{\text{geschätzt}} e^{-\lambda_{\text{geschätzt}} x} dx \\ &= \int_{12}^{+\infty} \frac{1}{9} e^{-\frac{1}{9}x} dx \\ &= \lim_{\beta \rightarrow +\infty} \int_{12}^{\beta} \frac{1}{9} e^{-\frac{1}{9}x} dx \\ &= \lim_{\beta \rightarrow +\infty} [-e^{-\frac{1}{9}x}]_{12}^{\beta} \\ &= \lim_{\beta \rightarrow +\infty} -e^{-\frac{1}{9}\beta} + e^{-\frac{1}{9} \cdot 12} \\ &= e^{-\frac{4}{3}} \\ &\approx 0,2636. \end{aligned}$$

Also beläuft sich die Wahrscheinlichkeit, daß die Maschine mindestens 12 Jahre läuft, auf etwa 0,2636.

3.3 Hypothesentests

3.3.1 Hypothesentest bei binomialverteilten Größen

Sei p die Erfolgswahrscheinlichkeit bei einem Einzelversuch. Sei uns p nicht bekannt.

Sei $X = X_n$ die Zufallsgröße für die Anzahl der Erfolge bei n Versuchen. Es ist

$$P(X_n = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Der Erwartungswert ist

$$\mu_n = E(X_n) = np.$$

Die Standardabweichung ist

$$\sigma_n = \sqrt{\text{Var}(X_n)} = \sqrt{np(1-p)}.$$

Näherung der Binomial- durch Normalverteilung gibt, wie in §2.7.5, §2.7.6,

$$P(\mu_n - a\sigma_n \leq X_n \leq \mu_n + a\sigma_n) \approx \frac{1}{\sqrt{2\pi}} \int_{-a}^{+a} \exp\left(-\frac{u^2}{2}\right) du$$

wobei $a \geq 0$. Häufige Werte für spätere Verwendung:

$$\begin{aligned} P(\mu_n - \sigma_n \leq X_n \leq \mu_n + \sigma_n) &\approx \frac{1}{\sqrt{2\pi}} \int_{-1}^{+1} \exp\left(-\frac{u^2}{2}\right) du \approx 0,6827 \\ P(\mu_n - 2\sigma_n \leq X_n \leq \mu_n + 2\sigma_n) &\approx \frac{1}{\sqrt{2\pi}} \int_{-2}^{+2} \exp\left(-\frac{u^2}{2}\right) du \approx 0,9545 \\ P(\mu_n - 3\sigma_n \leq X_n \leq \mu_n + 3\sigma_n) &\approx \frac{1}{\sqrt{2\pi}} \int_{-3}^{+3} \exp\left(-\frac{u^2}{2}\right) du \approx 0,9973 . \end{aligned}$$

Also:

$$\begin{aligned} \frac{1}{\sqrt{2\pi}} \int_{-a}^{+a} \exp\left(-\frac{u^2}{2}\right) du &\approx P(\mu_n - a\sigma_n \leq X_n \leq \mu_n + a\sigma_n) \\ &= P(|X_n - \mu_n| \leq a\sigma_n) \\ &= P(|X_n - np| \leq a\sqrt{np(1-p)}) \\ &= P\left(\left|\frac{1}{n}X_n - p\right| \leq a\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right) . \end{aligned}$$

Sei nun n Einzelversuche durchgeführt, wobei sich H_n Erfolge eingestellt haben.

Sei $h_n := \frac{1}{n}H_n$ die relative Erfolgshäufigkeit aus dem Versuch. Es ist

$$P(|h_n - p| \leq a\frac{\sqrt{p(1-p)}}{\sqrt{n}}) \approx \frac{1}{\sqrt{2\pi}} \int_{-a}^{+a} \exp\left(-\frac{u^2}{2}\right) du .$$

Nun kennen wir p nicht. Also schätzen wir ab: $p(1-p) \leq \frac{1}{2} \cdot (1 - \frac{1}{2}) = \frac{1}{4}$. Damit ist

$$P(|h_n - p| \leq \frac{a}{2\sqrt{n}}) \geq \frac{1}{\sqrt{2\pi}} \int_{-a}^{+a} \exp\left(-\frac{u^2}{2}\right) du .$$

Bemerkung. Sei X_n die binomialverteilte Zufallsgröße für die Anzahl der Erfolge nach n Versuchen, mit Erfolgswahrscheinlichkeit p für den Einzelversuch.

Sei H_n die aus einem durchgeführten n -fach iterierten Versuch resultierende Anzahl von Erfolgen. Sei $h_n := \frac{1}{n}H_n$ die zugehörige relative Erfolgshäufigkeit.

(1) $a = 1$: Mit einer Wahrscheinlichkeit von $\geq 68,27\%$ ist

$$|h_n - p| \leq \frac{1}{2\sqrt{n}} .$$

(2) $a = 2$: Mit einer Wahrscheinlichkeit von $\geq 95,45\%$ ist

$$|h_n - p| \leq \frac{1}{\sqrt{n}} .$$

(3) $a = 3$: Mit einer Wahrscheinlichkeit von $\geq 99,73\%$ ist

$$|h_n - p| \leq \frac{3}{2\sqrt{n}} .$$

Beispiel. Der Anteil der Bahnhofsbaubefürworter in der Gesamtbevölkerung sei p .

- (1) Es werden $n = 100$ Personen befragt. Davon sprechen sich 55 für einen Bahnhofsbaubau aus.

Wir legen eine Wahrscheinlichkeit von $\geq 95,45\%$ als zureichende Sicherheit fest.

Somit ist mit zureichender Sicherheit

$$\left| \frac{55}{100} - p \right| \leq \frac{1}{\sqrt{100}} ,$$

also

$$|p - 0,55| \leq 0,1 ,$$

also

$$0,45 \leq p \leq 0,65 .$$

Somit ist das Umfrageergebnis, in üblicher Ausdrucksweise: Es sind 55% der Gesamtbevölkerung für den Bahnhofsbaubau, mit einer Fehlertoleranz von ± 10 Prozentpunkten.

- (2) Es werden $n = 400$ Personen befragt. Davon sprechen sich 224 für einen Bahnhofsbaubau aus.

Wir legen wieder eine Wahrscheinlichkeit von $\geq 95,45\%$ als zureichende Sicherheit fest.

Somit ist mit zureichender Sicherheit

$$\left| \frac{224}{400} - p \right| \leq \frac{1}{\sqrt{400}} ,$$

also

$$|p - 0,56| \leq 0,05 ,$$

also

$$0,51 \leq p \leq 0,61 .$$

Somit ist das Umfrageergebnis, in üblicher Ausdrucksweise: Es sind 56% der Gesamtbevölkerung für den Bahnhofsbaubau, mit einer Fehlertoleranz von ± 5 Prozentpunkten.

Beispiel. Eine Firma verkauft Schoko-Eier mit Überraschungen. In der Werbung behauptet sie, es sei in jedem vierten Ei eine Figur enthalten.

Die Behauptung der Firma sei unsere Nullhypothese: Die Wahrscheinlichkeit für eine Figur in einem Ei ist $p_0 = 0,25$.

Soll dieses Nullhypothese aufrechterhalten werden oder verworfen werden?

- (1) Erster Versuch. Die Tester kaufen 50 Eier und finden 7 Figuren.

Ist das schon Anlaß für eine Beschwerde?

Wir legen eine Wahrscheinlichkeit von $\geq 95,45\%$ als zureichende Sicherheit fest.

Mit zureichender Sicherheit ist *unter Annahme der Nullhypothese*

$$|\frac{7}{50} - p_0| = |\frac{7}{50} - 0,25| = 0,11 \leq \frac{1}{\sqrt{50}} \approx 0,1414 .$$

Das stellt kein Problem dar. Also kann die Nullhypothese aufrechterhalten werden.

Das Ergebnis des Testkaufs ist im Rahmen. Es gibt keine Beschwerde.

- (2) Zweiter Versuch. Die Tester kaufen 500 Eier und finden 70 Figuren.

Ist das schon Anlaß für eine Beschwerde?

Wir legen wieder eine Wahrscheinlichkeit von $\geq 95,45\%$ als zureichende Sicherheit fest.

Mit zureichender Sicherheit ist *unter Annahme der Nullhypothese*

$$|\frac{70}{500} - p_0| = |\frac{70}{500} - 0,25| = 0,11 \leq \frac{1}{\sqrt{500}} \approx 0,04472 .$$

Das trifft nicht zu. Also kann die Nullhypothese nicht aufrechterhalten werden.

Das Ergebnis des Testkaufs ist nicht im Rahmen. Es gibt eine Beschwerde.

Beispiel. Das Schlafmittel Dorma werde getestet.

Wir legen eine Sicherheit von $\geq 99,73\%$ als zureichende Sicherheit fest.

Wir testen mit 100 Patienten.

Jeder Patient testet eine Woche Dorma, eine andere Woche ein Placebo (“controlled”). Was jeweils wann geschieht, wird zufällig festgelegt (“randomized”). Weder er noch der durchführende Arzt wissen, in welcher Woche welches Mittel verabreicht wird (“double-blind”).

Es wird gemessen, wer in welcher Woche besser schläft.

Die Nullhypothese sei: Das Schlafmittel Dorma ist wirkungslos. Wenn wir eine Person herausgreifen und ihr Schlafverhalten messen, dann ist die Wahrscheinlichkeit, in der Dorma-Woche besser geschlafen zu haben, unter dieser Nullhypothese also

$$p_0 = 0,5 .$$

Mit anderen Worten, wenn das Schlafmittel Dorma wirkungslos ist, dann wirkt es genauso gut wie ein Placebo. Es ist also genauso wahrscheinlich, in der Placebo-Woche besser geschlafen zu haben wie in der Dorma-Woche. Wir erhalten diesenfalls eine Verteilung wie bei einem Münzwurf.

Nun haben $H_{100} = 70$ Patienten in der Dorma-Woche besser geschlafen. Es ist also $h_{100} = \frac{1}{100} \cdot 70 = 0,7$.

Mit zureichender Sicherheit, also mit $\geq 99,73\%$ Wahrscheinlichkeit, ist *unter Annahme der Nullhypothese*

$$|h_n - p_0| \leq \frac{3}{2\sqrt{n}} ,$$

also

$$|h_{100} - 0,5| \leq \frac{3}{2\sqrt{100}} ,$$

also

$$0,2 = |0,7 - 0,5| \leq \frac{3}{20} = 0,15 ,$$

Dies trifft nicht zu. Also ist die Nullhypothese, Dorma sei wirkungslos, zu verwerfen.

Das Testergebnis ist also: Das Schlafmittel Dorma wirkt.

3.3.2 Levene-Test

Seien k Meßreihen aus je mindestens 2 Werten gegeben:

$$\begin{array}{l} x_{1,1}, x_{1,2}, \dots, x_{1,n_1} \\ x_{2,1}, x_{2,2}, \dots, x_{2,n_2} \\ x_{3,1}, x_{3,2}, \dots, x_{3,n_3} \\ \vdots \\ x_{k,1}, x_{k,2}, \dots, x_{k,n_k} \end{array}$$

Hierbei sei $x_{i,1}, x_{i,2}, \dots, x_{i,n_i}$ eine Meßreihe für eine Zufallsvariable X_i , für $1 \leq i \leq k$.

Z.B. könnte die i -te Meßreihe von den Produkten der Maschine Nummer i stammen, und man möchte wissen, ob die Maschinen dieselbe Varianz aufweisen, d.h. ob ihre Resultate die gleiche Streuung um den Erwartungswert aufweisen.

Nullhypothese: Es ist

$$\text{Var}(X_1) = \text{Var}(X_2) = \dots = \text{Var}(X_k) .$$

Die Nullhypothese ist also: Es liegen übereinstimmende Varianzen vor.

Mit dem Levene-Test kann diese Nullhypothese überprüft werden.

Der Levene-Test besteht aus folgenden Schritten.

1. Wir bilden den Mittelwerte der i -ten Meßreihe, also

$$\bar{x}_{i,-} := \frac{1}{n_i} \sum_{j=1}^{n_i} x_{i,j}$$

für $1 \leq i \leq k$. (Dabei wird $\bar{x}_{i,-}$ als “ x quer i Strich” gelesen.)

2. Wir bilden die Abstände der Meßergebnisse zum jeweiligen Mittelwert, also

$$y_{i,j} := |x_{i,j} - \bar{x}_{i,-}|$$

für $1 \leq i \leq k$ und $1 \leq j \leq n_i$.

3. Wir bilden den jeweiligen Mittelwert dieser Abstände, also

$$\bar{y}_{i,-} := \frac{1}{n_i} \sum_{j=1}^{n_i} y_{i,j}$$

für $1 \leq i \leq k$.

4. Wir bilden den Mittelwert aller vorliegenden Abstände.

Sei $n := \sum_{i=1}^k n_i$.

Sei

$$\bar{y}_{-,-} := \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} y_{i,j} = \frac{1}{n} \sum_{i=1}^k n_i \bar{y}_{i,-}.$$

5. Wir bilden folgende Größen.

$$\begin{aligned} Z &:= \frac{1}{k-1} \sum_{i=1}^k n_i (\bar{y}_{i,-} - \bar{y}_{-,-})^2 \\ N &:= \frac{1}{n-k} \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{i,j} - \bar{y}_{i,-})^2 \\ W &:= \frac{Z}{N} \end{aligned}$$

6. Sei $\alpha = 0,05$ unser gewähltes Signifikanzlevel.

Falls $W \leq F_\alpha(k-1, n-k)$ ist, dann kann die Nullhypothese nicht mit zureichender Sicherheit verworfen werden, und $\text{Var}(X_1) = \text{Var}(X_2) = \dots = \text{Var}(X_k)$ wird als Arbeitshypothese aufrechterhalten.

Hierbei heißt $F_\alpha(a, b)$ der “upper critical value of the F-distribution”.

Die Werte von $F_\alpha(a, b)$ entnehmen wir folgender Tabelle.

Tabelle für $F_{0,05}(a, b)$										
$a \backslash b$	1	2	3	4	5	6	7	8	9	10
1	161,448	199,500	215,707	224,583	230,162	233,986	236,768	238,882	240,543	241,882
2	18,513	19,000	19,164	19,247	19,296	19,330	19,353	19,371	19,385	19,396
3	10,128	9,552	9,277	9,117	9,013	8,941	8,887	8,845	8,812	8,786
4	7,709	6,944	6,591	6,388	6,256	6,163	6,094	6,041	5,999	5,964
5	6,608	5,786	5,409	5,192	5,050	4,950	4,876	4,818	4,772	4,735
6	5,987	5,143	4,757	4,534	4,387	4,284	4,207	4,147	4,099	4,060
7	5,591	4,737	4,347	4,120	3,972	3,866	3,787	3,726	3,677	3,637
8	5,318	4,459	4,066	3,838	3,687	3,581	3,500	3,438	3,388	3,347
9	5,117	4,256	3,863	3,633	3,482	3,374	3,293	3,230	3,179	3,137
10	4,965	4,103	3,708	3,478	3,326	3,217	3,135	3,072	3,020	2,978
11	4,844	3,982	3,587	3,357	3,204	3,095	3,012	2,948	2,896	2,854
12	4,747	3,885	3,490	3,259	3,106	2,996	2,913	2,849	2,796	2,753
13	4,667	3,806	3,411	3,179	3,025	2,915	2,832	2,767	2,714	2,671
14	4,600	3,739	3,344	3,112	2,958	2,848	2,764	2,699	2,646	2,602
15	4,543	3,682	3,287	3,056	2,901	2,790	2,707	2,641	2,588	2,544
16	4,494	3,634	3,239	3,007	2,852	2,741	2,657	2,591	2,538	2,494
17	4,451	3,592	3,197	2,965	2,810	2,699	2,614	2,548	2,494	2,450
18	4,414	3,555	3,160	2,928	2,773	2,661	2,577	2,510	2,456	2,412
19	4,381	3,522	3,127	2,895	2,740	2,628	2,544	2,477	2,423	2,378
20	4,351	3,493	3,098	2,866	2,711	2,599	2,514	2,447	2,393	2,348

Es ist $\alpha = 0,05$ ein üblicher Wert. In Schritt 6 ist aber noch zu erkennen, wie vorzugehen ist, wenn man ein anderes Signifikanzlevel wählen möchte.

Beispiel.

Seien folgende Meßreihen zu den Produkten der Maschinen 1, 2, 3 gegeben. Es ist also $k = 3$.

$$\begin{aligned}
 x_{1,j} &: \underbrace{50}_{x_{1,1}}, \underbrace{60}_{x_{1,2}}, \underbrace{65}_{x_{1,3}}, \underbrace{55}_{x_{1,4}} & n_1 &= 4 \\
 x_{2,j} &: \underbrace{30}_{x_{2,1}}, \underbrace{40}_{x_{2,2}}, \underbrace{40}_{x_{2,3}}, \underbrace{35}_{x_{2,4}}, \underbrace{50}_{x_{2,5}}, \underbrace{45}_{x_{2,6}} & n_2 &= 6 \\
 x_{3,j} &: \underbrace{40}_{x_{3,1}}, \underbrace{60}_{x_{3,2}}, \underbrace{50}_{x_{3,3}}, \underbrace{40}_{x_{3,4}}, \underbrace{55}_{x_{3,5}} & n_3 &= 5
 \end{aligned}$$

Wir durchlaufen die Schritte des Levene-Tests.

1. Es wird

$$\begin{aligned}
 \bar{x}_{1,-} &= \frac{1}{4}(50 + 60 + 65 + 55) &= 57,5 \\
 \bar{x}_{2,-} &= \frac{1}{6}(30 + 40 + 40 + 35 + 50 + 45) &= 40 \\
 \bar{x}_{3,-} &= \frac{1}{5}(40 + 60 + 50 + 40 + 55) &= 49 .
 \end{aligned}$$

2. Als Abstände zum jeweiligen Mittelwert erhalten wir:

$$\begin{aligned}
 y_{1,j} &: 7,5, 2,5, 7,5, 2,5 \\
 y_{2,j} &: 10, 0, 0, 5, 10, 5 \\
 y_{3,j} &: 9, 11, 1, 9, 6
 \end{aligned}$$

3. Von den Abständen bilden wir jeweils den Mittelwert:

$$\begin{aligned}\bar{y}_{1,-} &= \frac{1}{4}(7,5 + 2,5 + 7,5 + 2,5) &= 5 \\ \bar{y}_{2,-} &= \frac{1}{6}(10 + 0 + 0 + 5 + 10 + 5) &= 5 \\ \bar{y}_{3,-} &= \frac{1}{5}(9 + 11 + 1 + 9 + 6) &= 7,2\end{aligned}$$

4. Von allen Abständen bilden wir den Mittelwert: Es ist $n = 4 + 6 + 5 = 15$. Unter Verwendung von Schritt 3 wird

$$\bar{y}_{-,-} = \frac{1}{15}(4 \cdot 5 + 6 \cdot 5 + 5 \cdot 7,2) = \frac{86}{15} \approx 5,7333.$$

5. Wir erhalten

$$Z = \frac{1}{3-1}(4 \cdot (5 - 5,7333)^2 + 6 \cdot (5 - 5,7333)^2 + 5 \cdot (7,2 - 5,7333)^2) \approx 8,0667$$

und

$$N = \frac{1}{15-3}((2,5^2+2,5^2+2,5^2+2,5^2)+(5^2+5^2+5^2+0^2+5^2+0^2)+(1,8^2+3,8^2+6,2^2+1,8^2+1,2^2)) \approx 15,4833.$$

Somit wird

$$W = \frac{Z}{N} \approx 0,5210$$

6. Der Tabelle entnehmen wir den Wert $F_{0,05}(3-1, 15-3) = 3,885$.

Es ist $W \approx 0,5210 \leq 3,885 = F_{0,05}(3-1, 15-3)$. Dies bestätigt die Nullhypothese.

Es liegt somit mit zureichender Sicherheit übereinstimmende Varianzen bei den Produkten der drei Maschinen vor.

3.3.3 t-Test für unabhängige Stichproben (Student)

Der t-Test stammt von W. S. GOSSET. Er arbeitete für die Guinness-Brauerei und hatte diesen Test entwickelt, um die Qualität des Stout-Biers zu überwachen. Da sein Arbeitgeber aber Publikationen untersagt hatte, veröffentlichte Gosset den Test unter dem Pseudonym STUDENT.

Seien zwei Meßreihen aus je mindestens 2 Werten gegeben:

$$\begin{aligned}x_{1,1}, x_{1,2}, \dots, x_{1,n_1} \\ x_{2,1}, x_{2,2}, \dots, x_{2,n_2}\end{aligned}$$

Hierbei sei $x_{i,1}, x_{i,2}, \dots, x_{i,n_i}$ eine Meßreihe für eine normalverteilte Zufallsvariable X_i , für $i \in \{1, 2\}$.

Z.B. könnte die erste Meßreihe von den Produkten der Maschine 1 stammen, die zweite Meßreihe von den Produkten der Maschine 2 stammen, und man möchte wissen, ob

die Maschinen denselben Erwartungswert aufweisen, d.h. ob sie im langfristigen Mittel dasselbe Resultat liefern.

Nullhypothese: Es ist

$$E(X_1) = E(X_2) .$$

Die Nullhypothese ist also: Es liegen übereinstimmende Erwartungswerte vor.

0. Wir überprüfen mit dem Levene-Test, ob $\text{Var}(X_1) = \text{Var}(X_2)$ nicht mit zureichender Sicherheit verworfen werden kann und damit als Arbeitshypothese aufrechterhalten wird.

Falls die Varianzen übereinstimmen, kann mit dem t-Test diese Nullhypothese überprüft werden.

Der t-Test besteht aus folgenden Schritten.

1. Wir bilden den Mittelwerte der i -ten Meßreihe, also

$$\bar{x}_{i,-} := \frac{1}{n_i} \sum_{j=1}^{n_i} x_{i,j}$$

für $i \in \{1, 2\}$.

2. Für die Standardabweichungen mit Bessel-Korrektur berechnen wir

$$(n_i - 1)s_i^2 := \sum_{j=1}^{n_i} (x_{i,j} - \bar{x}_{i,-})^2$$

für $i \in \{1, 2\}$.

3. Wir bilden

$$s_- := \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}} .$$

4. Wir bilden den t-Wert

$$t := \frac{\bar{x}_{1,-} - \bar{x}_{2,-}}{s_- \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} .$$

5. Sei $\alpha = 0,05$ unser gewähltes Signifikanzlevel.

Falls $|t| \leq t_{1-\frac{\alpha}{2}, (n_1-1)+(n_2-1)}$ ist, dann kann die Nullhypothese nicht mit zureichender Sicherheit verworfen werden, und $E(X_1) = E(X_2)$ wird als Arbeitshypothese aufrechterhalten.

Hierbei heißt $t_{1-\frac{\alpha}{2}, a}$ der “critical value of Student’s distribution with a degrees of freedom”.

Die Werte von $t_{1-\frac{\alpha}{2}, a} = t_{0,975, a}$ entnehmen wir folgender Tabelle.

Tabelle für $t_{0,975,a}$

a	$t_{0,975,a}$
1	12,706
2	4,303
3	3,182
4	2,776
5	2,571
6	2,447
7	2,365
8	2,306
9	2,262
10	2,228
11	2,201
12	2,179
13	2,160
14	2,145
15	2,131
16	2,120
17	2,110
18	2,101
19	2,093
20	2,086

Es ist $\alpha = 0,05$ ein üblicher Wert. In Schritt 5 ist aber noch zu erkennen, wie vorzugehen ist, wenn man ein anderes Signifikanzlevel wählen möchte.

Beispiel.

Seien folgende Meßreihen zu den Produkten der Maschinen 1, 2 gegeben.

$$\begin{array}{ll}
 x_{1,j} & : \underbrace{20}_{x_{1,1}}, \underbrace{30}_{x_{1,2}}, \underbrace{24}_{x_{1,3}}, \underbrace{30}_{x_{1,4}} \quad n_1 = 4 \\
 x_{2,j} & : \underbrace{30}_{x_{2,1}}, \underbrace{20}_{x_{2,2}}, \underbrace{30}_{x_{2,3}}, \underbrace{35}_{x_{2,4}}, \underbrace{35}_{x_{2,5}} \quad n_2 = 5
 \end{array}$$

0. Zunächst führen wir einen Levene-Test durch, um auf übereinstimmende Varianzen zu testen.

0.1. Es wird

$$\begin{aligned}
 \bar{x}_{1,-} &= \frac{1}{4}(20 + 30 + 24 + 30) = 26 \\
 \bar{x}_{2,-} &= \frac{1}{5}(30 + 20 + 30 + 35 + 35) = 30
 \end{aligned}$$

0.2. Als Abstände zum jeweiligen Mittelwert erhalten wir:

$$\begin{aligned}
 y_{1,j} &: 6, 4, 2, 4 \\
 y_{2,j} &: 0, 10, 0, 5, 5
 \end{aligned}$$

0.3. Von den Abständen bilden wir jeweils den Mittelwert:

$$\begin{aligned}
 \bar{y}_{1,-} &= \frac{1}{4}(6 + 4 + 2 + 4) = 4 \\
 \bar{y}_{2,-} &= \frac{1}{5}(0 + 10 + 0 + 5 + 5) = 4
 \end{aligned}$$

0.4. Von allen Abständen bilden wir den Mittelwert: Es ist $n = 4 + 5 = 9$. Unter Verwendung von Schritt 3 wird

$$\bar{y}_{-,-} = \frac{1}{9}(4 \cdot 4 + 5 \cdot 4) = 4.$$

0.5. Wir erhalten

$$Z = \frac{1}{2-1}(4 \cdot (4-4)^2 + 5 \cdot (4-4)^2) = 0$$

und

$$N = \frac{1}{9-2}((2^2 + 0^2 + 2^2 + 0^2) + (4^2 + 6^2 + 4^2 + 1^2 + 1^2)) \approx 11,1429.$$

Somit wird

$$W = \frac{Z}{N} = 0.$$

0.6. Der Tabelle entnehmen wir den Wert $F_{0,05}(2-1, 9-2) = 5,591$.

Es ist $W = 0 \leq 5,591 = F_{0,05}(2-1, 9-2)$. Dies bestätigt die Nullhypothese.

Es liegt somit mit zureichender Sicherheit übereinstimmende Varianzen bei den Produkten der zwei Maschinen vor.

Natürlich hätte man schon bei der Feststellung $Z = 0$ sofort auf mit zureichender Sicherheit übereinstimmende Varianzen schließen können.

Nachdem wir übereinstimmende Varianzen festgestellt haben, können wir nun die Schritte des t-Tests durchlaufen.

1. Aus dem Levene-Test kennen wir schon $\bar{x}_{1,-} = 26$ und $\bar{x}_{2,-} = 30$.

2. Es wird

$$\begin{aligned} (n_1 - 1)s_1^2 &= (20 - 26)^2 + (30 - 26)^2 + (24 - 26)^2 + (30 - 26)^2 &= 72 \\ (n_2 - 1)s_2^2 &= (30 - 30)^2 + (20 - 30)^2 + (30 - 30)^2 + (30 - 35)^2 + (30 - 35)^2 &= 150 \end{aligned}$$

3. Es wird

$$s_- = \sqrt{\frac{72 + 150}{(4-1) + (5-1)}} = \sqrt{\frac{222}{7}} \approx 5,6315.$$

4. Wir bilden den t-Wert

$$t := \frac{26 - 30}{5,6315 \cdot \sqrt{\frac{1}{4} + \frac{1}{5}}} \approx -1,0588$$

5. Der Tabelle entnehmen wir den Wert $t_{1-\frac{\alpha}{2}, (n_1-1)+(n_2-1)} = t_{0,975, 7} = 2,365$.

Es ist $|t| \approx 1,0588 \leq 2,365 = t_{0,975, 7}$.

Somit ist die Nullhypothese bestätigt, und es liegen mit zureichender Sicherheit übereinstimmende Erwartungswerte bei den Produkten der beiden Maschinen vor.

3.3.4 U-Test für unabhängige Stichproben (Wilcoxon-Mann-Whitney)

Der U-Test stammt von H. MANN und D. R. WHITNEY, und, unabhängig davon, von F. WILCOXON.

Definition. Sei $X : \Omega \rightarrow \mathbb{R}$ eine stetige Zufallsgröße.

Der *Median* M_X von X ist die reelle Zahl, für die $P(X \leq M_X) = \frac{1}{2}$ ist.

Diesfalls ist auch $P(X \geq M_X) = P(X > M_X) = \frac{1}{2}$.

Beispiel. Sei X exponentialverteilt mit Rate $\lambda > 0$. Es ist also

$$p_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{für } x \geq 0 \\ 0 & \text{für } x < 0. \end{cases}$$

Wir erinnern an den Erwartungswert $\mu_X = E(X) = \frac{1}{\lambda}$; vgl. §2.7.4.

Es ist $M_X = \frac{\ln(2)}{\lambda}$. Denn es wird

$$\begin{aligned} P(X \leq M_X) &= \int_{-\infty}^{M_X} p_X(x) \, dx \\ &= \int_0^{M_X} \lambda e^{-\lambda x} \, dx \\ &= [-e^{-\lambda x}]_0^{M_X} \\ &= -e^{-\lambda M_X} - (-1) \\ &= -e^{-\lambda \frac{\ln(2)}{\lambda}} + 1 \\ &= -e^{-\ln(2)} + 1 \\ &= -\frac{1}{2} + 1 \\ &= \frac{1}{2}. \end{aligned}$$

Insbesondere ist $M_X = \frac{\ln(2)}{\lambda} \neq \frac{1}{\lambda} = \mu_X$.

Verfahren. Seien $X_1 : \Omega \rightarrow \mathbb{R}$ und $X_2 : \Omega \rightarrow \mathbb{R}$ Zufallsgrößen.

Seien zwei Meßreihen gegeben.

$$\begin{array}{ll} x_{1,1}, x_{1,2}, \dots, x_{1,n_1} & \text{für die Zufallsgröße } X_1 \\ x_{2,1}, x_{2,2}, \dots, x_{2,n_2} & \text{für die Zufallsgröße } X_2 \end{array}$$

Mit dem U-Test wollen wir anhand dieser Meßreihen überprüfen, ob von übereinstimmenden Medianen ausgegangen werden kann, d.h. ob $M_{X_1} = M_{X_2}$ ist.

Unsere Nullhypothese ist also: Es ist $M_{X_1} = M_{X_2}$.

Der U-Test verläuft nun wie folgt.

1. Wir ordnen die Meßwerte an:

$$y_1 \leq y_2 \leq \dots \leq y_{n_1+n_2}.$$

Hierbei tauche jeder Meßwert aus den Reihen $x_{1,1}, \dots, x_{1,n_1}, x_{2,1}, \dots, x_{2,n_2}$ genau einmal unter den $y_1, \dots, y_{n_1+n_2}$ auf.

Der Rang von y_i sei wie folgt gegeben. Sei $\text{Gleich}(y_i) := \{1 \leq j \leq n_1 + n_2 : y_j = y_i\}$ die Menge der Indizes j , bei denen $y_j = y_i$ ist. Sei

$$\text{Rang}(y_i) := \frac{1}{2}(\min \text{Gleich}(y_i) + \max \text{Gleich}(y_i))$$

der *Rang* von y_i , wobei $1 \leq i \leq n_1 + n_2$.

Ist z.B. $y_{i-1} < y_i < y_{i+1}$, dann ist $\text{Gleich}(y_i) = \{i\}$ und also $\text{Rang}(y_i) = \frac{1}{2}(i + i) = i$.

Ist z.B. $y_{i-1} < y_i = y_{i+1} < y_{i+2}$, dann ist $\text{Gleich}(y_i) = \{i, i+1\}$ und also $\text{Rang}(y_i) = \frac{1}{2}(i + (i+1)) = i + \frac{1}{2}$. Genauso ist dann auch $\text{Rang}(y_{i+1}) = i + \frac{1}{2}$.

Ist z.B. $y_{i-1} < y_i = y_{i+1} = y_{i+2} < y_{i+3}$, dann ist $\text{Gleich}(y_i) = \{i, i+1, i+2\}$ und also $\text{Rang}(y_i) = \frac{1}{2}(i + (i+2)) = i+1$. Genauso ist dann auch $\text{Rang}(y_{i+1}) = \text{Rang}(y_{i+2}) = i+1$.

Wir bilden die Rangsummen

$$\begin{aligned} R_1 &:= \sum_{j=1}^{n_1} \text{Rang}(x_{1,j}) \\ R_2 &:= \sum_{j=1}^{n_2} \text{Rang}(x_{2,j}) . \end{aligned}$$

2. Sei

$$\begin{aligned} U_1 &:= n_1 \cdot n_2 + \frac{1}{2} n_1 \cdot (n_1 + 1) - R_1 \\ U_2 &:= n_1 \cdot n_2 + \frac{1}{2} n_2 \cdot (n_2 + 1) - R_2 . \end{aligned}$$

Sei

$$U := \min\{U_1, U_2\} .$$

Hier ist eine Probe möglich: Es muß $U_1 + U_2 = n_1 \cdot n_2$ sein.

3. Sei

$$\mu_U := \frac{1}{2} n_1 \cdot n_2 .$$

Sei

$$\sigma_U := \sqrt{\frac{1}{12} n_1 \cdot n_2 \cdot (n_1 + n_2 + 1)} .$$

Sei

$$z := \frac{U - \mu_U}{\sigma_U} .$$

4. *Unter Annahme der Nullhypothese* ist nun, wie bei der Normalverteilung:

Mit 68,27% Wahrscheinlichkeit : $z \in [-1, +1]$.

Mit 95,45% Wahrscheinlichkeit : $z \in [-2, +2]$.

Mit 99,73% Wahrscheinlichkeit : $z \in [-3, +3]$.

Legen wir z.B. 95,45% Wahrscheinlichkeit als zureichende Sicherheit fest, dann ist unter Annahme der Nullhypothese mit zureichender Sicherheit $z \in [-2, +2]$. Falls $z \notin [-2, +2]$ ist, so ist die Nullhypothese zu verwerfen.

Legen wir z.B. 99,93% Wahrscheinlichkeit als zureichende Sicherheit fest, dann ist unter Annahme der Nullhypothese mit zureichender Sicherheit $z \in [-3, +3]$. Falls $z \notin [-3, +3]$ ist, so ist die Nullhypothese zu verwerfen.

Die Nullhypothese zu verwerfen, bedeutet, zum Ergebnis zu kommen, daß mit zureichender Sicherheit $M_{X_1} \neq M_{X_2}$ ist, daß sich die Mediane also unterscheiden.

Die Nullhypothese aufrechtzuerhalten, bedeutet, man kann als Arbeitshypothese von $M_{X_1} = M_{X_2}$ ausgehen, also von übereinstimmenden Medianen.

Beispiel. Seien folgende Meßreihen gegeben.

$$\begin{array}{ll} x_{1,j} & : \underbrace{303}_{x_{1,1}}, \underbrace{305}_{x_{1,2}}, \underbrace{302}_{x_{1,3}}, \underbrace{302}_{x_{1,4}} \quad \text{für } X_1 \\ x_{2,j} & : \underbrace{310}_{x_{2,1}}, \underbrace{310}_{x_{2,2}}, \underbrace{307}_{x_{2,3}}, \underbrace{305}_{x_{2,4}}, \underbrace{304}_{x_{2,5}}, \underbrace{310}_{x_{2,6}} \quad \text{für } X_2 \end{array}$$

Es ist also $n_1 = 4$ und $n_2 = 6$.

Mit dem U-Test wollen wir anhand dieser Meßreihen überprüfen, ob von übereinstimmenden Medianen ausgegangen werden kann, d.h. ob $M_{X_1} = M_{X_2}$ ist.

Unsere Nullhypothese ist also: Es ist $M_{X_1} = M_{X_2}$.

Wir führen den U-Test durch. Als zureichende Sicherheit setzen wir eine Wahrscheinlichkeit von 95,45% fest.

1. Wir ordnen die Meßwerte an:

$$\underbrace{302}_{y_1 = x_{1,3}}, \underbrace{302}_{y_2 = x_{1,4}}, \underbrace{303}_{y_3 = x_{1,1}}, \underbrace{304}_{y_4 = x_{2,5}}, \underbrace{305}_{y_5 = x_{1,2}}, \underbrace{305}_{y_6 = x_{2,4}}, \underbrace{307}_{y_7 = x_{2,3}}, \underbrace{310}_{y_8 = x_{2,1}}, \underbrace{310}_{y_9 = x_{2,2}}, \underbrace{310}_{y_{10} = x_{2,6}}$$

Aus der Anordnung entnehmen wir:

$$\begin{aligned} R_1 &= \text{Rang}(x_{1,1}) + \text{Rang}(x_{1,2}) + \text{Rang}(x_{1,3}) + \text{Rang}(x_{1,4}) \\ &= 3 + 5,5 + 1,5 + 1,5 \\ &= 11,5 \end{aligned}$$

und

$$\begin{aligned} R_2 &= \text{Rang}(x_{2,1}) + \text{Rang}(x_{2,2}) + \text{Rang}(x_{2,3}) + \text{Rang}(x_{2,4}) + \text{Rang}(x_{2,5}) + \text{Rang}(x_{2,6}) \\ &= 9 + 9 + 7 + 5,5 + 4 + 9 \\ &= 43,5. \end{aligned}$$

2. Es wird

$$\begin{aligned} U_1 &= n_1 \cdot n_2 + \frac{1}{2} n_1 \cdot (n_1 + 1) - R_1 = 4 \cdot 6 + \frac{1}{2} \cdot 4 \cdot 5 - 11,5 = 22,5 \\ U_2 &= n_1 \cdot n_2 + \frac{1}{2} n_2 \cdot (n_2 + 1) - R_2 = 4 \cdot 6 + \frac{1}{2} \cdot 6 \cdot 7 - 43,5 = 1,5. \end{aligned}$$

Es wird

$$U = \min\{U_1, U_2\} = 1,5 .$$

Zur Probe rechnen wir: Es ist tatsächlich $U_1 + U_2 = 22,5 + 1,5 = 24 = n_1 \cdot n_2$.

3. Es wird

$$\mu_U = \frac{1}{2} n_1 \cdot n_2 = \frac{1}{2} \cdot 4 \cdot 6 = 12$$

und

$$\sigma_U = \sqrt{\frac{1}{12} n_1 \cdot n_2 \cdot (n_1 + n_2 + 1)} = \sqrt{\frac{1}{12} \cdot 4 \cdot 6 \cdot (4 + 6 + 1)} = \sqrt{22} \approx 4,6904 .$$

Es wird

$$z = \frac{U - \mu_U}{\sigma_U} = \frac{1,5 - 12}{4,6904} \approx -2,2386 .$$

4. Es ist

$$z \notin [-2, +2] .$$

Unter Annahme der Nullhypothese ist aber mit zureichender Sicherheit z in diesem Intervall enthalten.

Also kann die Nullhypothese nicht aufrechterhalten werden.

Der U-Test liefert also das Ergebnis: Die Mediane von X_1 und X_2 unterscheiden sich, es ist $M_{X_1} \neq M_{X_2}$.

Literatur

- [1] BEYER, O.; HACKEL, H.; PIEPER, V.; TIEDGE, J., *Wahrscheinlichkeitsrechnung und mathematische Statistik*, Teubner, 1999.
- [2] CROARKIN, C., TOBIAS, P., *NIST/SEMATECH e-Handbook of Statistical Methods*, <http://www.itl.nist.gov/div898/handbook/>, 2025.
- [3] DIPPON, J., *Statistik für Wirtschaftswissenschaftler*, Skript, Stuttgart, 2017.
- [4] EULER, B., *Fachdidaktik 1 Mathematik*, Skript, Stuttgart, 2025.
- [5] WEBER, H., *Einführung in die Wahrscheinlichkeit und Statistik für Ingenieure*, Teubner, 1992.