Kapitel 8

Grundbegriffe der Statistik und Wahrscheinlichkeitstheorie

8.1 Messreihen

Im Folgenden werden zwei Typen von *Merkmalen* betrachtet:

- *quantitativ-diskrete*, z.B. Alter in Jahren, Geschosszahl eines Gebäudes,...
  Die *Merkmalausprägungen* sind dann ganze Zahlen.
- *quantitativ-stetige*, z.B. Gebäudehöhe, Temperatur,...
  Die *Merkmalausprägungen* sind dann reelle Zahlen.

Am Beginn einer statistischen Untersuchung steht immer die mehrfache Beobachtung eines Merkmals. Das Beobachtungsergebnis ist dann eine *Messreihe* von n Zahlen

$$x_1, x_2, \ldots, x_n.$$

Definition 8.1.1 *Sei $x_1, x_2, \ldots, x_n$ eine Messreihe. Ordnet man die Werte der Messreihe der Größe nach, so ensteht die zugehörige* geordnete Messreihe

$$x_{(1)}, x_{(2)}, \ldots, x_{(n)}.$$

Sie besteht aus den gleichen Zahlen, aber so umgeordnet, dass gilt $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$.

Die empirische Verteilungsfunktion *einer Messreihe $x_1, x_2, \ldots, x_n$ ist die Funktion*

$$F(z; x_1, x_2, \ldots, x_n) = \frac{\textit{Zahl der } x_i \textit{ mit } x_i \leq z}{n} = \frac{\max\left\{i \,:\, x_{(i)} \leq z\right\}}{n}.$$

Wählt man $r-1$ Zahlen $a_1 < a_2 < \ldots < a_{r-1}$, so entsteht die Unterteilung von $\mathbb{R}$ in r Klassen

$$\mathbb{R} =]-\infty, a_1] \cup]a_1, a_2] \cup \cdots \cup]a_{r-2}, a_{r-1}] \cup]a_{r-1}, \infty[.$$

Mit der Abkürzung $F(z) = F(z; x_1, x_2, \ldots, x_n)$ ergeben sich dann die *relativen Klassenhäufigkeiten* für diese r Klassen zu

$$F(a_1), F(a_2) - F(a_1), \ldots, F(a_{r-1}) - F(a_{r-2}), 1 - F(a_{r-1}).$$

Wählt man noch zwei zusätzliche Zahlen

$$a_0 < \min\{a_1, x_{(1)}\}, \quad a_r > \max\{a_{r-1}, x_{(n)}\},$$

so können die relativen Klassenhäufigkeiten in einem *Histogramm* dargestellt werden: über jedem der Intervalle $]a_{j-1}, a_j]$, $j = 1, \ldots, r$, wird ein Rechteck erreichtet, das die jeweilige Klassenhäufigkeit als Fläche hat.

Beispiel:

Die zur Messreihe

$$2.2 \quad 4.5 \quad 0.8 \quad 1.7 \quad 5.8 \quad 1.2 \quad 5.6 \quad 2.5 \quad 3.9 \quad 1.7$$

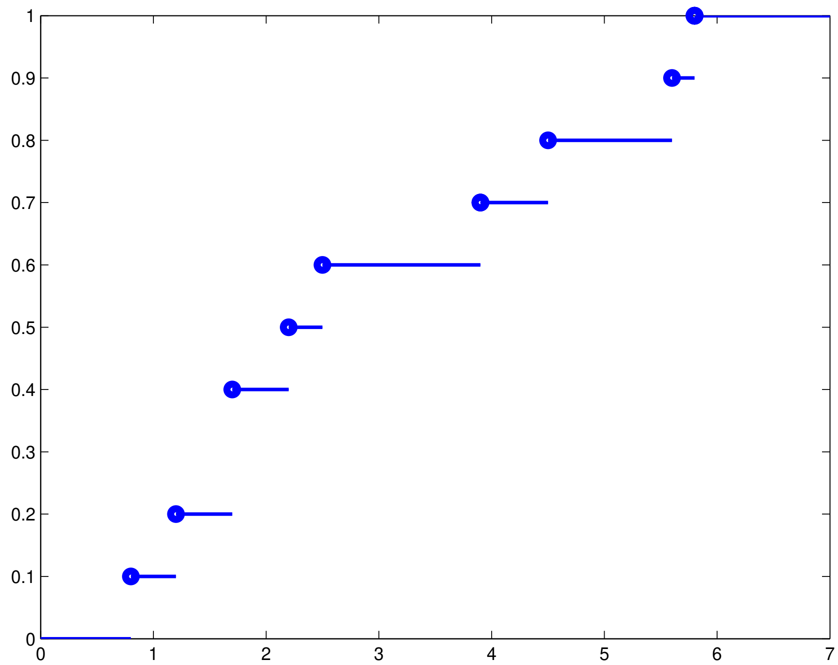
gehörige geordnete Messreihe ist

$$0.8 \quad 1.2 \quad 1.7 \quad 1.7 \quad 2.2 \quad 2.5 \quad 3.9 \quad 4.5 \quad 5.6 \quad 5.8.$$

Hierbei ist

$$x_{(1)} = x_3,\ x_{(2)} = x_6,\ x_{(3)} = x_{(4)} = x_4 = x_{10},\ x_{(5)} = x_1,\ x_{(6)} = x_8,$$
$$x_{(7)} = x_9,\ x_{(8)} = x_2,\ x_{(9)} = x_7,\ x_{(10)} = x_5.$$

Die empirische Verteilungsfunktion hat folgenden Graphen:



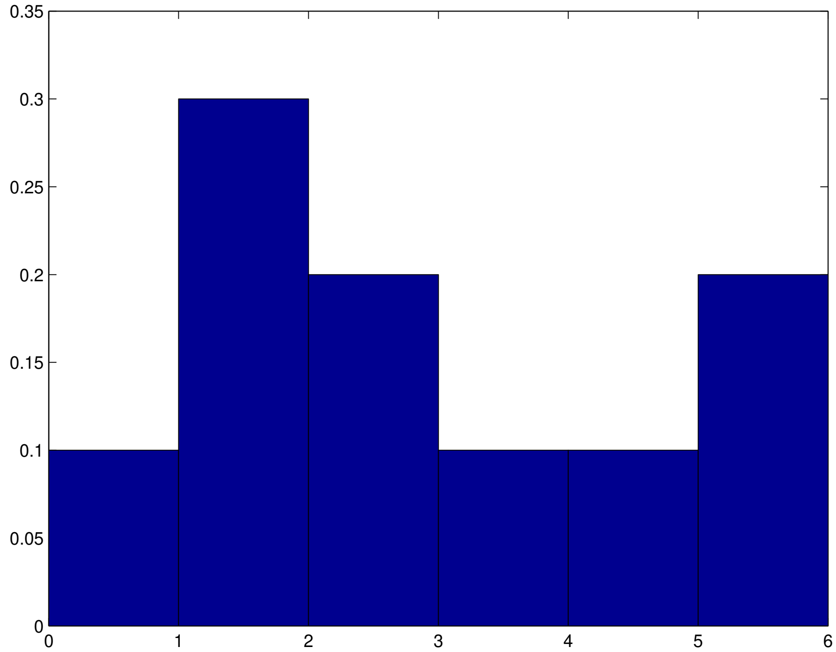
Zu der Unterteilung

$$0 < 1 < 2 < 3 < 4 < 5 < 6$$

ergeben sich die Klassenhäufigkeiten

$$\frac{1}{10}, \quad \frac{4-1}{10} = \frac{3}{10}, \quad \frac{6-4}{10} = \frac{2}{10}, \quad \frac{7-6}{10} = \frac{1}{10}, \quad \frac{8-7}{10} = \frac{1}{10}, 1 - \frac{8}{10} = \frac{2}{10},$$

also das Histogramm



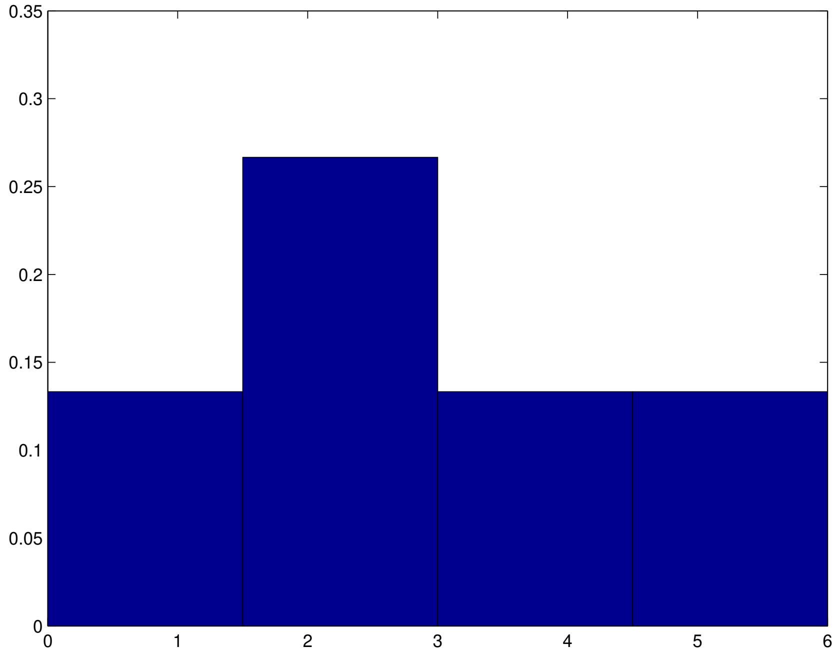
Zur Unterteilung

$$0 < 1.5 < 3 < 4.5 < 6$$

ergeben sich die Klassenhäufigkeiten

$$\frac{2}{10}, \quad \frac{6-2}{10} = \frac{4}{10}, \quad \frac{8-6}{10} = \frac{2}{10} \quad 1 - \frac{8}{10} = \frac{2}{10},$$

mit zugehörigem Histogramm



.

8.2 Lage- und Streumaßzahlen

Zur Beschreibung und Charakterisierung von Messreihen dienen Lage- und Streumaßzahlen.

Sei $x_1, \ldots, x_n$ eine Messreihe.

8.2.1 Lagemaßzahlen

Beispiele für Lagemaßzahlen sind

Arithmetisches Mittel:

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \ldots + x_n)$$

Median:

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

p**-Quantil** $(0 < p < 1)$**:**

$$x_p = \begin{cases} x_{(np)}, & \text{falls } np \text{ ganzzahlig,} \\ x_{([np]+1)}, & \text{falls } np \text{ nicht ganzzahlig.} \end{cases}$$

Hierbei ist

$$[x] = \max\{z \in \mathbb{Z} \,:\, z \leq x\} \quad (GaussscheKlammer)$$

die größte ganze Zahl $\leq x$.

α-gestutztes Mittel $(0 < \alpha < 0.5)$**:**

$$\bar{x}_\alpha = \frac{1}{n-2k}(x_{(k+1)} + \ldots + x_{(n-k)}), \quad k = [n\alpha].$$

Das 0.25-Quantil $x_{0.25}$ wird als *unteres Quartil*, das 0.75-Quantil $x_{0.75}$ wird als *oberes Quartil* bezeichnet.

8.2.2 Streuungsmaße

Beispiele für Streuungsmaße sind:

Empirische Varianz oder **empirische Stichprobenvarianz:**

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

Empirische Streuung:

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

Spannweite:

$$v = x_{(n)} - x_{(1)}$$

Quartilabstand:

$$q = x_{0.75} - x_{0.25}.$$

8.2.3 Zweidimensionale Messreihen

Werden bei einer statistischen Erhebung zwei verschiedene Merkmale gleichzeitig ermittelt, so entstehen *zweidimensionale Messreihen*, d.h. endliche Folgen

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n).$$

Analog wie oben definieren wir folgende Maßzahlen:

Arithmetische Mittel:

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \ldots + x_n), \quad \bar{y} = \frac{1}{n}(y_1 + y_2 + \ldots + y_n)$$

Empirische Varianzen:

$$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2, \quad s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

Empirische Streuungen:

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

Weiterhin sind auch folgende Maßzahlen zwischen den x_i und y_i von Bedeutung:

Empirische Kovarianz:

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

Empirischer Korrelationskoeffizient:

$$r_{xy} = \frac{s_{xy}}{s_x\, s_y}.$$

Bemerkung: Es gilt immer

$$-1 \le r_{xy} \le 1.$$

Beweis: Sei $u = (x_i - \bar{x})_{1\le i\le n}$, $v = (y_i - \bar{y})_{1\le i\le n}$. Dann gilt nach der Cauchy-Schwartz-Ungleichung

$$r_{xy} = \frac{u^T v}{\|u\|_2 \|v\|_2} \in [-1, 1].$$

□

Bemerkung: Die empirische Varianz s^2 lässt sich auch nach der Formel berechnen

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

Ebenso gilt für die empirische Kovarianz

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

Beweis: Es gilt

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

Die Formel für s_{xy} folgt ganz analog. □

Zur Veranschaulichung einer zweidimensionalen Messreihe dient das *Punktediagramm*, bei dem die Punkte (x_i, y_i), $i = 1, \ldots, n$, als Punkte in einem $x - y$-Koordinatensystem eingetragen werden.

8.2.4 Regressionsgerade

Der Korrelationskoeffizient r_{xy} gibt Hinweise, ob die y-Werte tendenziell monoton wachsend oder monoton fallend von den x-Werten abhängen. Für diesen Zusammenhang soll nun angenommen werden, dass er sich im wesentlichen durch eine lineare Gleichung der Form

$$y = ax + b$$

beschreiben lässt. Wir nehmen also an, dass sich die Datenpunkte um eine Gerade mit der Steigung a und Achsenabschnitt b gruppieren. Wir wollen nun a und b bestimmen, damit die Gerade möglichst gut zu den Datenpunkten passt. Das Quadrat des Abstands zwischen Datenpunkt (x_i, y_i) und einem Punkt $(x_i, ax_i + b)$ auf der Geraden mit demselben x-Wert ist $(y_i - ax_i - b)^2$. Steigung a und Achsenabschnitt b der Geraden sollen nun so bestimmt werden, dass die Summe all dieser Quadrate

$$S(a,b) = \sum_{i=1}^{n} (y_i - ax_i - b)^2$$

minimal wird. Wir erhalten dann die sogenannte *Regressionsgerade*. Wir suchen also eine Lösung des Problems

$$\min_{(a,b)\in\mathbb{R}^2} S(a,b). \tag{8.3}$$

Bestimmung der Minimalstelle:

$$\frac{\partial S(a,b)}{\partial a} = \sum_{i=1}^{n} 2(y_i - ax_i - b)\,(-x_i) = -2\sum_{i=1}^{n}(x_i y_i - ax_i^2 - bx_i) = 0.$$

$$\frac{\partial S(a,b)}{\partial b} = \sum_{i=1}^{n} 2(y_i - ax_i - b)\,(-1) = -2\sum_{i=1}^{n}(y_i - ax_i - b) = 0.$$

Die zweite Gleichung ergibt

$$n\bar{y} - an\bar{x} - nb = 0,$$

also

$$b = \bar{y} - a\bar{x}.$$

Einsetzen in die erste Gleichung ergibt

$$\sum_{i=1}^{n}(x_i y_i - ax_i^2 - \bar{y}x_i + a\bar{x}x_i) = 0$$

und somit

$$\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} = a\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right).$$

Insgesamt ergibt sich für die Lösung $(\hat{a}, \hat{b})$ von (8.3) unter Verwendung von (8.1), (8.2):

Berechnung der Regressionsgerade:

$$y = \hat{a}\,x + \hat{b},$$

mit

$$\hat{a} = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} = \frac{s_{xy}}{s_x^2}, \quad \hat{b} = \bar{y} - \hat{a}\bar{x}.$$

Wie bereits erwähnt, heisst die so gefundene Gerade *Regressionsgerade*. Die Abweichungen der Punkte (x_i, y_i) von der Regressionsgerade in vertikaler Richtung

$$r_i = y_i - \hat{a}x_i - \hat{b}, \quad i = 1, \ldots, n$$

heißen *Residuen*. Nach kurzer Rechnung erhält man folgende

Formel für das Residuenquadrat:

$$\sum_{i=1}^n r_i^2 = \sum_{i=1}^n (y_i - \bar{y})^2\,(1 - r_{xy}^2).$$

Die vertikale Abweichung von der Regressionsgerade hängt also eng mit dem Korrelationskoeffizienten r_{xy} zusammen. Für die extremen Werte $r_{xy} = 1$ bzw. $r_{xy} = -1$ verschwinden die Residuen, alle Punkte (x_i, y_i) liegen also auf der Regressionsgeraden.

Da die Werte

$$r_{xy} = \frac{s_{xy}}{s_x s_y} \quad \text{und} \quad \hat{a} = \frac{s_{xy}}{s_x^2}$$

gleiches Vorzeichen haben, ergibt sich also für $r_{xy} > 0$ eine streng monoton steigende, für $r_{xy} < 0$ eine streng monoton fallende und für $r_{xy} = 0$ eine horizontale Regressionsgerade.

Das Vorzeichen von r_{xy} gibt also den Trend der Abhängigkeit der y-Werten von den x-Werten an.

8.3 Zufallsexperimente und Wahrscheinlichkeit

8.3.1 Zufallsexperimente

Ein Vorgang, der so genau beschrieben ist, dass er als beliebig oft wiederholbar betrachtet werden kann, und dessen Ergebnisse vom Zufall abhängen, nennen wir *Zufallsexperiment*. Es wird angenommen, dass die Menge der möglichen Ergebnisse soweit bekannt ist, dass jedem Ergebnis ein Element ω einer Menge Ω zugeordnet werden kann.

Definition 8.3.1 *Ω heißt* Ergebnismenge, *seine Elemente ω* Ergebnisse. *Teilmengen $A \subset \Omega$ heißen* Ereignisse. *Ein Ereignis $A \subset \Omega$ tritt ein, falls ein Ergebnis $\omega \in A$ beobachtet wird.*

Beispiele:

1. Wurf eines Würfels: $\Omega = \{1, 2, 3, 4, 5, 6\}$. Das Ereignis $A = \{1, 3, 5\}$ tritt ein, falls eine ungerade Zahl gewürfelt wird.
2. Wurf zweier unterscheidbarer Würfel: $\Omega = \{(i, j) : i, j \in \{1, 2, 3, 4, 5, 6\}\}$. Ω hat $6 \cdot 6 = 36$ Elemente.
3. Wurf zweier nicht unterscheidbarer Würfel: $\Omega = \{(i, j) : i, j \in \{1, 2, 3, 4, 5, 6\}, \ i \leq j\}$. Ω hat 21 Elemente.
4. Lebensdauer eines Gerätes: $\Omega =]0, \infty[$. Das Ereignis $A = \{\omega \in \mathbb{R} : \omega > 100\}$ tritt ein, wenn das Gerät mehr als 100 Stunden fehlerfrei funktioniert.

Definition 8.3.2 *Das aus zwei Ereignissen A und B* zusammengesetzte Ereignis *$A \cup B$ tritt ein, falls ein Ergebnis ω mit $\omega \in A$ oder $\omega \in B$ (d.h. $\omega \in A \cup B$) beobachtet wird.*

Entsprechend tritt das Ereignis $A \cap B$ ein, falls ein Ergebnis ω mit $\omega \in A$ und $\omega \in B$ (d.h. $\omega \in A \cap B$) beobachtet wird.

$A^c = \Omega \setminus A$ heißt zu A komplementäres Ereignis.

Zwei Ereignisse A und B heißen unvereinbar, *falls $A \cap B = \emptyset$.*

Die leere Menge $\emptyset$ heißt unmögliches Ereignis *und Ω das* sichere Ereignis.

Die einelementigen Mengen $\{\omega\}$ von Ω heißen Elementarereignisse.

Auch für Folgen $A_1, A_2, \ldots$ von Ereignissen definieren wir das zusammengesetzte Ereignis $\bigcup_{i=1}^{\infty} A_i$, das eintritt, wenn mindestens ein A_i eintritt, und das Ereignis $\bigcap_{i=1}^{\infty} A_i$, das eintritt, wenn alle A_i zugleich eintreten.

8.3.2 Wahrscheinlichkeit

Fragt man im Falle der Betriebsdauer eines Gerätes danach, wie wahrscheinlich es ist, dass das Gerät exakt nach 100 Stunden (keinen Augenblick früher oder später!) seinen ersten Defekt hat, dann ist dies praktisch ausgeschlossen. Fragt man jedoch danach, dass der erste Defekt zwischen 90 und 100 Stunden auftritt, also nach der Wahrscheinlichkeit des Ereignisses $A = [90, 100]$, dann ist dies eine sachgerechte Fragestellung.

Dies zeigt, dass es sinnvoll ist, die Wahrscheinlichkeit des Eintretens von Ereignissen zu betrachten. Wir haben dabei die Vorstellung, dass die Wahrscheinlichkeit $P(A)$ für das

Eintreten des Ereignisses A immer genauer der relativen Häufigkeit des Eintretens von A in Versuchsserien entspricht, je länger die Versuchsreihe wird.

Dazu betrachten wir ein System $\mathcal{A}$ von Ereignissen (es muss nicht die Potenzmenge $\mathcal{P}(\Omega)$, also die Menge aller Teilmengen von Ω sein!), das folgende Eigenschaften hat:

Definition 8.3.3 *ein System $\mathcal{A} \subset \mathcal{P}(\Omega)$ von Ereignissen heißt σ-*Algebra*, wenn gilt:*

a) $\Omega \in \mathcal{A}$.

b) Falls $A \in \mathcal{A}$, dann gilt auch $A^c \in \mathcal{A}$.

c) Mit jeder Folge $A_1, A_2, \ldots \in \mathcal{A}$ gilt auch $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$.

Bemerkung: Sind $A, B \in \mathcal{A}$, dann ist wegen b) und c) auch

$$A \cap B = (A^c \cup B^c)^c \in \mathcal{A}.$$

□

Eine σ-Algebra erlaubt gerade die Verknüpfungen von Ereignissen, die in der Praxis nützlich sind. Um jedem Ereignis eine Wahrscheinlichkeit zuzuordnen, betrachtet man eine Abbildung $P : \mathcal{A} \to \mathbb{R}$ mit folgenden Eigenschaften:

Definition 8.3.4 *Eine Abbildung $P : \mathcal{A} \to \mathbb{R}$ heißt* Wahrscheinlichkeitsmaß*, wenn sie den folgenden* Axiomen von Kolmogorov *genügt:*

a) $P(A) \geq 0$ *für* $A \in \mathcal{A}$,

b) $P(\Omega) = 1$,

c) $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$ *für paarweise unvereinbare* $A_1, A_2, \ldots \in \mathcal{A}$.

Bemerkung: c) umfasst auch endliche disjunkte Vereinigungen $\bigcup_{i=1}^{n} A_i$ durch die Wahl $A_i = \emptyset$, $i \geq n+1$.

Aus diesen Axiomen folgen nützliche Regeln für das Rechnen mit Wahrscheinlichkeiten von Ereignissen $A, B, A_1, \ldots, A_n$:

$$P(\emptyset) = 0,$$
$$0 \leq P(A) \leq 1,$$
$$P(A^c) = 1 - P(A),$$
$$A \subset B \implies P(A) \leq P(B),$$
$$P(A \cup B) = P(A) + P(B) - P(A \cap B),$$
$$P(A_1 \cup \cdots \cup A_n) = \sum_{i=1}^{n} P(A_i), \quad \text{falls } A_1, \ldots, A_n \text{ paarweise unvereinbar (Additivität).}$$

Falls die Ergebnismenge endlich ist, also $\Omega = \{\omega_1, \ldots, \omega_n\}$, und man wie beim Würfelwurf annehmen kann, dass jedes Elementarereignis $\{\omega_i\}$ gleich wahrscheinlich ist, dann folgt aus der Additivität

$$P(\{\omega_i\}) = \frac{1}{n}, \quad i = 1, \ldots, n$$

und für beilebige Ereignisse mit Elementzahl $\#A$ gilt

$$P(A) = \frac{\text{Elementzahl von } A}{n} = \frac{\#A}{\#\Omega}.$$

Die Annahme gleicher Wahrscheinlichkeit für die Elementarereignisse heißt *Laplace-Annahme*.

Beispiel: Drei Würfel werden geworfen. Wie groß ist die Wahrscheinlichkeit, dass die Würfelsumme 11 ist?

Wir wählen $\Omega = \{(i, j, k) \,:\, i, j, k = 1, \ldots, 6\}$. Dann ist $\#\Omega = 6^3 = 216$.

A sei das Ereignis "Würfelsumme ist 11".

Möglichkeiten für die drei Summanden (der Größe nach geordnet):

$$11 = 1 + 4 + 6 = 1 + 5 + 5 = 2 + 3 + 6 = 2 + 4 + 5 = 3 + 3 + 5 = 3 + 4 + 4.$$

Wir summieren die Anzahl der Tripel auf, die auf die angegebenen Summanden führen:

$$\#A = 6 + 3 + 6 + 6 + 3 + 3 = 27.$$

Dies ergibt

$$P(A) = \frac{\#A}{\#\Omega} = \frac{27}{216} = \frac{1}{8} = 0,125.$$

8.3.3 Elementare Formeln der Kombinatorik

Zur Berechnung der Elementezahlen von Ereignissen werden häufig kombinatorische Formeln verwendet. Wir geben einige wichtige Formeln an:

Sei Ω eine Menge mit n Elementen und $k \in \mathbb{N}$.

Geordnete Probe mit Wiederholungen:
Ein k-Tupel $(x_1, \ldots, x_k)$ mit $x_i \in \Omega$, $i = 1, \ldots, k$, heißt *geordnete Probe* von Ω vom Umfang k *mit Wiederholungen*. Es gibt

$$n^k \quad \text{(Anzahl geordneter Proben mit Wiederholungen)}$$

solcher Proben (für jede Stelle gibt es n Möglichkeiten).

Geordnete Probe ohne Wiederholungen:
Ein k-Tupel $(x_1, \ldots, x_k)$, $k \leq n$, mit $x_i \in \Omega$, $i = 1, \ldots, k$, und $x_i \neq x_j$ für $i \neq j$ heißt *geordnete Probe* von Ω vom Umfang k *ohne Wiederholungen*. Es gibt

$$n(n-1)(n-2) \cdot \ldots \cdot (n-k+1) \quad \text{(Anzahl geordneter Proben ohne Wiederholungen)}$$

solcher Proben (für die erste Stelle gibt es n Möglichkeiten, für die zweite $n-1$, usw.).

Im Fall $k = n$ spricht man von einer *Permutation* der Menge Ω. Davon gibt es

$$n! = n(n-1)(n-2) \cdot \ldots \cdot 2 \cdot 1 \quad \text{(Anzahl von Permutationen)}$$

Ungeordnete Probe ohne Wiederholungen:
Eine Teilmenge $\{x_1, \ldots, x_k\}$, $k \leq n$, von Ω heißt *ungeordnete Probe* von Ω vom Umfang k *ohne Wiederholungen*. Es gibt

$$\binom{n}{k} = \frac{n(n-1)(n-2) \cdot \ldots \cdot (n-k+1)}{k!} = \frac{n!}{k!(n-k)!} \quad \text{(Anzahl } k\text{-elem. Teilmengen)}$$

solcher Proben (es gibt $n(n-1)(n-2) \cdot \ldots \cdot (n-k+1)$ geordnete Proben, aber jeweils $k!$ bestehen aus den gleichen k Elementen).

Beispiel:

1. Wie viele Möglichkeiten gibt es, k Einsen und $n-k$ Nullen anzuordnen?

   Lösung: Jede Anordnung der Einsen entspricht einer k-elementigen Teilmenge von $\{1, \ldots, n\}$, welche die Positionen der Einsen angibt. Also: $\binom{n}{k}$ Möglichkeiten.

2. Beim Austeilen gemischter Karten (32 Karten, davon 4 Asse) sei A das Ereignis "die ersten drei Karten sind Asse". Dann gilt unter der Laplace-Annahme

$$P(A) = \frac{4 \cdot 3 \cdot 2 \cdot 29!}{32!} = \frac{24}{32 \cdot 31 \cdot 30} = \frac{1}{1240}.$$

8.4 Bedingte Wahrscheinlichkeit, Unabhängigkeit

8.4.1 Bedingte Wahrscheinlichkeit

Seien A, B zwei Ereignisse mit $P(B) > 0$. Oft interessiert die Wahrscheinlichkeit von A unter der Bedingung, dass B eintritt. Man definiert diese *bedingte Wahrscheinlichkeit* $P(A|B)$ von A unter der Bedingung B durch

Bedingte Wahrscheinlichkeit von A unter der Bedingung B:

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

Beispiel: Beim Ziehen von einem gemischten Kartenstapel (32 Karten, 4 Asse) betrachte die Ereignisse A "die zweite Karte ist ein Ass" und B "die erste Karte ist ein Ass". Dann gilt

$$P(B) = \frac{4 \cdot 31!}{32!} = \frac{1}{8}, \quad P(A \cap B) = \frac{4 \cdot 3 \cdot 30!}{32!} = \frac{12}{32 \cdot 31}.$$

Dies ergibt
$$P(A|B) = \frac{12 \cdot 8}{32 \cdot 31} = \frac{3}{31}.$$
Direkte Rechnung: Wenn schon ein Ass gezogen ist, dann ist die Wahrscheinlichkeit, dass die zweite Karte wieder ein Ass ist
$$P(A|B) = \frac{3 \cdot 30!}{31!} = \frac{3}{31}.$$

Im Folgenden seien $A_1, \dots, A_n$ paarweise unvereinbare Ereignisse, d.h. $A_i \cap A_j = \emptyset$ für $i \neq j$, und es sei $\bigcup_{i=1}^n A_i = \Omega$. Man spricht von einer *vollständigen Ereignisdisjunktion*.

Es gelten die folgenden Rechenregeln:

Regel von der vollständigen Wahrscheinlichkeit:

$A_1, \dots, A_n$ sei eine vollständige Ereignisdisjunktion mit $P(A_i) > 0$, $i = 1, \dots, n$. Dann gilt:
$$P(B) = \sum_{i=1}^n P(A_i) \cdot P(B|A_i).$$

Beweis: Es gilt $P(A_i) \cdot P(B|A_i) = P(B \cap A_i)$. Die Mengen $C_i = B \cap A_i$ sind paarweise disjunkt mit $\bigcup_{i=1}^n C_i = B$. Wegen der Additivität gilt also
$$\sum_{i=1}^n P(A_i) \cdot P(B|A_i) = \sum_{i=1}^n P(C_i) = P\left(\bigcup_{i=1}^n C_i\right) = P(B).$$
□

Formel von Bayes:

$A_1, \dots, A_n$ sei eine vollständige Ereignisdisjunktion mit $P(A_i) > 0$, $i = 1, \dots, n$, und B sei ein Ereignis mit $P(B) > 0$. Dann gilt für $i = 1, \dots, n$:
$$P(A_i|B) = \frac{P(A_i) \cdot P(B|A_i)}{\sum_{k=1}^n P(A_k) \cdot P(B|A_k)}.$$

Beweis: Der Nenner ist $P(B)$ nach der Regel von der vollständigen Wahrscheinlichkeit. Also ist die rechte Seite gegebn durch
$$\frac{P(A_i) \cdot P(B|A_i)}{P(B)} = \frac{P(A_i) \cdot P(B \cap A_i)/P(A_i)}{P(B)} = \frac{P(B \cap A_i)}{P(B)} = P(A_i|B).$$
□

Beispiel: Bei einer Reihenuntersuchung sind die Ereignisse A: ”untersuchter Patient ist erkrankt” und B: ”Befund positiv” von Interesse. Es sei $P(A) = 0,001$ die Wahrscheinlichkeit, dass ein Patient erkrankt ist. Weiter seien $P(B|A) = 0,92$ und $P(B|A^c) = 0,01$

die Wahrscheinlichkeiten für einen positiven Befund bei einem erkrankten bzw. nicht erkrankten Patienten.

Gesucht ist die bedingte Wahrscheinlichkeit, dass ein Patient bei einem positiven Befund tatsächlich erkrankt ist, also $P(A|B)$.

Mit $A_1 = A$, $A_2 = A^c$ ergibt die Bayessche Formel

$$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A^c)P(B|A^c)} = \frac{0,001 \cdot 0,92}{0,001 \cdot 0,92 + 0,999 \cdot 0,01} = 0,0844.$$

Multiplikationsformel:

$A_1, \ldots, A_n$ seien Ereignisse mit $P(A_1 \cap A_2 \cap \cdots \cap A_n) > 0$. Dann gilt

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap \cdots \cap A_{n-1}).$$

Beweis: Vollständige Induktion nach n: Für $n = 2$ gilt

$$P(A_1) \cdot P(A_2|A_1) = P(A_1 \cap A_2).$$

Induktionsschritt:

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

□

8.4.2 Unabhängigkeit

Beim zweifachen Werfen eines Würfels erkennt man, dass die Ereignisse

$$A = \text{"1 beim zweiten Wurf"}, \quad B = \text{"1 beim ersten Wurf"}$$

von völlig unabhängig ablaufenden Teilexperimenten bestimmt wird und für die bedingte Wahrscheinlichkeit gilt

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/6 \cdot 1/6}{1/6} = 1/6 = P(A).$$

Wir haben also

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

Dies motiviert die

Definition 8.4.1 *Zwei Ereignisse A und B heißen* unabhängig, *falls gilt*

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

Ereignisse $A_1, \ldots, A_n$ heißen vollständig unabhängig, *falls für alle $\{i_1, \ldots, i_k\} \subset \{1, \ldots, n\}$ gilt*

$$P(A_{i_1} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) \cdot \ldots \cdot P(A_{i_k}).$$

Bemerkung: Aus der paarweisen Unabhängigkeit von mehr als zwei Ereignissen folgt nicht immer die vollständige Unabhängigkeit. □

8.5 Zufallsvariablen und Verteilungsfunktion

Es sei Ω die Ergebnismenge und $\mathcal{A}$ das Ereignissystem, auf dem die Wahrscheinlichkeit P erklärt ist. Oft ist man in der Statistik an einem dem Ergebnis $\omega \in \Omega$ zugeordneten Zahlenwert $X(\omega)$ interessiert.

Definition 8.5.1 *Eine* Zufallsvariable *ist eine Abbildung*

$$X : \Omega \to \mathbb{R}$$

mit der Eigenschaft, dass für jedes Intervall $I \subset \mathbb{R}$ die Urbildmenge

$$A = \{\omega \in \Omega \;:\; X(\omega) \in I\}$$

zum Ereignissystem $\mathcal{A}$ gehört. Die Wahrscheinlichkeit dieses Ereignisses "X nimmt Werte im Intervall I an" bezeichnet man abkürzend mit $P(X \in I)$ und schreibt entsprechend

$$P(a \leq X \leq b), \;\; P(X \leq x), \;\; P(X < x), \;\; P(|X - a| < b), \;\; P(X = b) \;\; \textit{usw.}$$

Beispiel: Zwei Würfel werden geworfen. Wir wählen $\Omega = \{(i,j) \;:\; i,j = 1, \ldots, 6\}$. Wir betrachten die Zufallsvariable $X : \Omega \to \mathbb{R}$,

$$X((i,j)) = i + j.$$

X beschreibt also die Summe der beiden gewürfelten Zahlen. Nun gilt zum Beipiel

$$P(X = 1) = 0, \;\; P(X = 2) = P(\{(1,1\}) = \frac{1}{36}, \;\; P(X = 3) = P(\{(1,2),(2,1)\}) = \frac{2}{36}, \ldots$$

Definition 8.5.2 *Sei $X : \Omega \to \mathbb{R}$ eine Zufallsvariable. Die Abbildung $F : \mathbb{R} \to \mathbb{R}$*

$$F(x) = P(X \leq x), \quad x \in \mathbb{R},$$

heißt Verteilungsfunktion *der Zufallsvariable X.*

Man kann zeigen, dass mit den Abkürzungen

$$F(x+) = \lim_{h \searrow 0} F(x+h), \quad F(x-) = \lim_{h \searrow 0} F(x-h),$$
$$F(-\infty) = \lim_{x \to -\infty} F(x), \quad F(\infty) = \lim_{x \to \infty} F(x)$$

gilt: Verteilungsfunktionen sind monoton wachsende Funktionen mit

$$F(-\infty) = 0, \quad F(\infty) = 1, \quad F(x+) = F(x) \quad \forall\, x \in \mathbb{R}.$$

Zudem lassen sich alle interessierenden Wahrscheinlichkeiten im Zusammenhang mit der Zufallsvariable X berechnen:

$$\begin{aligned} P(X = a) &= F(a) - F(a-), \\ P(a < X \leq b) &= F(b) - F(a), \\ P(a \leq X < b) &= F(b-) - F(a-), \\ P(a \leq X \leq b) &= F(b) - F(a-), \\ P(X > a) &= 1 - F(a). \end{aligned}$$

Eine Zufallsvariable X heißt *diskret verteilt*, wenn sie nur endlich viele oder abzählbar unendlich viele Werte $x_1, x_2, \ldots$ annimmt. Ihre Verteilungsfunktion ist eine monoton wachsende Treppenfunktion, die jeweils an den Stellen x_i Sprünge der Höhe $P(X = x_i)$ hat.

Eine Zufallsvariable X heißt *stetig verteilt mit der Dichte f*, wenn ihre Verteilungsfunktion F durch

$$F(x) = \int_{-\infty}^{x} f(t)\, dt, \quad x \in \mathbb{R},$$

gegeben ist. Die Dichte ist hierbei eine nichtnegative Funktion, die Verteilungsfunktion F ist stetig und es gilt $F'(x) = f(x)$ für alle Stetigkeitsstellen x von f.

8.5.1 Beispiele für diskrete Verteilungen

Geometrische Verteilung

Es sei $0 < p < 1$. Eine Zufallsvariable X mit dem Wertebereich $\mathbb{N} = \{1, 2, \ldots\}$ heißt *geometrisch verteilt* mit dem Parameter p, falls

$$P(X = i) = (1-p)^{i-1}\, p, \quad i = 1, 2, \ldots$$

Anwendung: Wird ein Zufallsexperiment, bei dem ein bestimmtes Ereignis mit Wahrscheinlichkeit p eintritt, so lange unabhängig wiederholt, bis zum ersten Mal dieses Ereignis eintritt, dann kann die Anzahl der dazu benötigten Versuche durch eine geometrisch verteilte Zufallsvariable modelliert werden (”Warten auf den ersten Erfolg”).

Binomialverteilung

Seine $n \in \mathbb{N}$ und $0 < p < 1$. Eine Zufallsvariable X mit dem Wertebereich $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$ heißt *binomialverteilt* mit Parametern n und p, kurz $B(n, p)$-verteilt, falls

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

Anwendung: Wird ein Zufallsexperiment, bei dem ein bestimmtes Ereignis mit Wahrscheinlichkeit p eintritt, n-mal unabhängig wiederholt, und dabei gezählt, wie oft dieses Ereignis eintritt, so kann diese zufällige Anzahl als $B(n, p)$-verteilte Zufallsvariable X beschrieben werden ("Anzahl der Erfolge bei n Versuchen").

Poissonverteilung

Sei $\lambda > 0$. Eine Zufallsvariable X mit dem Wertebereich $\mathbb{N} \cup \{0\}$ heißt *Poisson-verteilt*, falls gilt

$$P(X = i) = \frac{\lambda^i}{i!} e^{-\lambda}, \quad i = 0, 1, 2, \ldots$$

Sie eignet sich zur Modellierung von Zählergebnissen folgenden Typs: In einer Telefonzentrale wird die Anzahl der innerhalb von 10 Minuten eingehenden Anrufe gezählt. λ gibt die "mittlere Anzahl" der eingehenden Anrufe an.

8.5.2 Beispiele für stetige Verteilungen

Rechteckverteilung

Es sei $a < b$. Eine stetig verteilte Zufallsvariable mit der Dichte

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

heißt *rechteckverteilt* im Intervall $[a, b]$, kurz $R(a, b)$-verteilt. Für die Verteilungsfunktion ergibt sich

$$F(x) = \int_{-\infty}^{x} f(t)\, dt = \begin{cases} 0, & x \le a, \\ \frac{x-a}{b-a}, & a < x < b, \\ 1 & x \ge b. \end{cases}$$

Exponentialverteilung

Sei $\lambda > 0$. Eine stetig verteilte Zufallsvariable X mit der Dichte

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

heißt *exponentialverteilt* mit Parameter λ, kurz $Ex(\lambda)$-verteilt. Für die Verteilungsfunktion ergibt sich

$$F(x) = \int_{-\infty}^{x} f(t)\, dt = \begin{cases} 0, & x < 0, \\ 1 - e^{-\lambda x}, & x \geq 0. \end{cases}$$

Normalverteilung

E seien $\mu \in \mathbb{R}$ und $\sigma > 0$. Eine stetig verteilte Zufallsvariable X mit der Dichte

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

heißt *normalverteilt* mit Parameter μ und σ^2, kurz: $N(\mu, \sigma^2)$-verteilt.

Im Fall $\mu = 0, \sigma^2 = 1$ spricht man von einer *Standard-Normalverteilung* und bezeichnet ihre Verteilungsfunktion mit

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt.$$

Φ ist nicht geschlossen angebbar und muss tabelliert oder numerisch ausgewertet werden. Offensichtlich gilt

$$\Phi(0) = \frac{1}{2}, \quad \Phi(-x) = 1 - \Phi(x), \quad x \geq 0.$$

Ist X eine $N(\mu, \sigma^2)$-verteilte Zufallsvariable, dann rechnet man leicht nach, dass die Verteilungsfunktion durch

$$F_{\mu,\sigma^2}(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

gegeben ist. Tatsächlich ergibt sich durch die Substitution $s = \frac{t-\mu}{\sigma}$

$$F_{\mu,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}(\frac{t-\mu}{\sigma})^2}\, dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-\mu}{\sigma}} e^{-\frac{1}{2}s^2}\, ds = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

8.6 Erwartungswert und Varianz

Ist X eine diskret verteilte Zufallsvariable mit den Werten $x_1, x_2, \ldots$, so heißt

$$E(X) = \sum_i x_i P(X = x_i)$$

Erwartungswert von X, falls $\sum_i |x_i| P(X = x_i) < \infty$.

Ist X eine stetig verteilte Zufallsvariable mit Dichte f, so heißt

$$E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx$$

Erwartungswert von X, falls $\int_{-\infty}^{\infty} |x|\, f(x)\, dx < \infty$.

Beispiele:

1. Sei X Poisson-verteilt mit Parameter $\lambda > 0$.

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

2. Sei X exponentialverteilt. Dann gilt

$$E(X) = \int_0^{\infty} x\, \lambda e^{-\lambda x}\, dx = -x e^{-\lambda x}|_0^{\infty} + \int_0^{\infty} e^{-\lambda x}\, dx = 0 - \frac{1}{\lambda} e^{-\lambda x}|_0^{\infty} = \frac{1}{\lambda}.$$

Ist $h : \mathbb{R} \to \mathbb{R}$ eine stückweise stetige Funktion. Dann hat die Zufallsvariable $h(X)$ für eine diskret verteilt Zufallsvariable X den Erwartungswert (im Falle seiner Existenz)

$$E(h(X)) = \sum_i h(x_i) P(X = x_i).$$

Ist X stetig verteilt mit Dichte f, dann hat $h(X)$ den Erwartungswert (im Falle seiner Existenz)

$$E(h(X)) = \int_{-\infty}^{\infty} h(x)\, f(x)\, dx.$$

Der Erwartungswert der quadratischen Abweichung der Zufallsvariablen X von ihrem Erwartungswert $E(X)$ heißt *Varianz* von X:

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

Die *Standardabweichung von* X ist definiert durch $\sqrt{\text{Var}(X)}$.

8.6.1 Rechenregeln

Es gelten folgende Rechenregeln:

$$E(aX+b) = aE(X)+b$$
$$E(h_1(X)+h_2(X)) = E(h_1(X))+E(h_2(X)).$$

Daraus erhält man

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

Begründung: Es gilt

$$\begin{aligned}\text{Var}(X) &= E([X-E(X)]^2) = E(X^2-2E(X)X+E(X)^2)\\ &= E(X^2)-E(2E(X)X)+E(E(X)^2) = E(X^2)-2E(X)E(X)+E(X)^2\\ &= E(X^2)-E(X)^2.\end{aligned}$$

Außerdem gilt

$$\text{Var}(aX+b) = a^2\text{Var}(X),$$

da

$$\begin{aligned}\text{Var}(aX+b) &= E((aX+b)^2)-(E(aX+b))^2 = E(a^2X^2+2abX+b^2)-(aE(X)+b)^2\\ &= a^2E(X^2)+2abE(X)+b^2-a^2E(X)^2-2abE(X)-b^2\\ &= a^2(E(X^2)-E(X)^2) = a^2\text{Var}(X).\end{aligned}$$

Einige Beispiele:

Verteilung	$E(X)$	Var(X)
$N(\mu,\sigma^2)$	μ	σ^2
$Ex(\lambda)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
$B(n,p)$	np	$np(1-p)$

Die Tschebyschevsche Ungleichung stellt einen Zusammenhang zwischen Erwartungswert und Varianz her:

Tschebyschevsche Ungleichung:
Es gilt

$$P(|X-E(X)|\geq c) \leq \frac{\text{Var}(X)}{c^2}, \quad c>0.$$

Aus der Definition des Erwartungswerts ist klar, dass für Zufallsvariablen $X_1, X_2, \ldots, X_n$ gilt

$$E(X_1+\ldots+X_n) = E(X_1)+\ldots+E(X_n).$$

Die Frage, ob eine entsprechende Formel auch für die Varianz gilt, führt auf den Begriff der Unabhängigkeit von Zufallsvariablen $X_1, X_2, \ldots, X_n$.

Definition 8.6.1 *Seien $X_1, X_2, \ldots, X_n$ Zufallsvariablen mit Verteilungsfunktionen $F_1, \ldots, F_n$. Die* gemeinsame Verteilungsfunktion *von $X_1, X_2, \ldots, X_n$ ist definiert durch*

$$F(x_1, \ldots, x_n) = P(X_1 \leq x_1, \cdots, X_n \leq x_n), \quad (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

Die Zufallsvariablen heißen unabhängig, *wenn für alle $(x_1, \ldots, x_n) \in \mathbb{R}^n$ die Ereignisse*

$$\{X_1 \leq x_1\}, \cdots, \{X_n \leq x_n\}$$

vollständig unabhängig sind, also

$$P(X_1 \leq x_1, \cdots, X_n \leq x_n) = P(X_1 \leq x_1) \cdot \ldots \cdot P(X_n \leq x_n)$$

oder kurz

$$F(x_1, \ldots, x_n) = F_1(x_1) \cdot \ldots \cdot F_n(x_n).$$

Satz 8.6.2 *Die Zufallsvariablen $X_1, X_2, \ldots, X_n$ seien unabhängig. Dann gilt*

$$\mathit{Var}(X_1 + \ldots + X_n) = \mathit{Var}(X_1) + \ldots + \mathit{Var}(X_n).$$

8.7 Gesetz der großen Zahlen, zentraler Grenzwertsatz

8.7.1 Das schwache Gesetz der großen Zahlen

Durch die Mittelung vieler unabhängiger identisch verteilter Zufallsvariablen erhält man eine Zufallsvariable, die mit großer Wahrscheinlichkeit Werte nahe beim Erwartungswert liefert.

Satz 8.7.1 (Das schwache Gesetz der großen Zahlen)
Ist $X_1, X_2, \ldots$ eine Folge unabhängiger identisch verteilter Zufallsvariablen (d.h. je endlich viele von ihnen sind unabhängig) mit $E(X_i) = \mu$, $\mathit{Var}(X_i) = \sigma^2$, dann gilt

$$\lim_{n \to \infty} P\left(\left| \frac{1}{n} \sum_{i=1}^{n} X_i - \mu \right| \geq \varepsilon \right) = 0 \quad \forall \varepsilon > 0.$$

Beweis: Setze $Y_n = \frac{1}{n} \sum_{i=1}^{n} X_i$. Dann gilt $E(Y_n) = \mu$ und wegen der Unabhängigkeit $\mathrm{Var}(Y_n) = \frac{1}{n^2} \mathrm{Var}(X_1 + \ldots + X_n) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$. Die Tschebyschevsche Ungleichung ergibt nun

$$P(|Y_n - \mu| \geq \varepsilon) \leq \frac{\mathrm{Var}(Y_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \to 0 \quad \text{für } n \to \infty.$$

□

8.7.2 Zentraler Grenzwertsatz

Wir betrachten eine Zufallsvariable

$$Y = X_1 + \ldots + X_n$$

mit unabhängigen Summanden $X_1, \ldots, X_n$. Extrem große Werte von Y treten nur dann auf, wenn sehr viele X_i gleichzeitig große Werte annehmen. Wegen der Unabhängigkeit ist es sehr wahrscheinlich, dass große Werte eines Summanden durch kleine Werte eines anderen Summanden kompensiert werden. Es zeigt sich, dass die Verteilung von Y für großes n mehr und mehr einer Normalverteilung entspricht:

Satz 8.7.2 *(*Zentraler Grenzwertsatz*)*
Ist $X_1, X_2, \ldots$ eine Folge unabhängiger Zufallsvariablen mit

$$E(X_i) = \mu_i, \quad \mathit{Var}(X_i) = \sigma_i^2, \quad i = 1, 2, \ldots,$$

so gilt unter schwachen zusätzlichen Voraussetzungen, z.B. dass X_i identisch verteilt sind:

$$\lim_{n\to\infty} P\left(\frac{X_1 + \ldots + X_n - (\mu_1 + \ldots, \mu_n)}{\sqrt{\sigma_1^2 + \ldots + \sigma_n^2}} \le y\right) = \Phi(y) \quad \forall\, y \in \mathbb{R}.$$

Ein arithmetisches Mittel

$$\bar{X}_{(n)} = \frac{1}{n}(X_1 + \ldots + X_n)$$

ist für großes n also näherungsweise $N(\mu, \sigma^2)$-verteilt, wobei

$$\mu = \frac{1}{n}E(X_1 + \ldots + X_n) = \frac{1}{n}(\mu_1 + \ldots + \mu_n), \quad \sigma^2 = \frac{1}{n^2}\mathit{Var}(X_1 + \ldots + X_n) = \frac{1}{n^2}(\sigma_1^2 + \ldots + \sigma_n^2).$$

Bemerkung: Hat X Erwartungswert μ und Varianz σ^2, dann hat $\frac{X-\mu}{\sigma}$ den Erwartungswert 0 und Varianz 1. □

Wahrscheinlichkeitstheoretisches Modell für Messreihen

Als mathematisches Modell für das Entstehen von Messreihen werden im folgenden unabhängige, identisch verteilte Zufallsvariablen $X_1, \ldots, X_n$ verwendet.
Eine Messreihe $x_1, \ldots, x_n$ wird als Realisierung der Zufallsvariablen $X_1, \ldots, X_n$ angesehen, wir nehmen also an, dass ein Ergebnis $\omega \in \Omega$ existiert mit

$$x_1 = X_1(\omega), \ldots, x_n = X_n(\omega).$$

Haben die X_i Erwartungswert μ und Varianz σ^2, dann sagt Satz 8.7.2 insbesondere aus, dass dann das arithmetische Mittel $\bar{X}_{(n)} = \frac{1}{n}(X_1 + \ldots + X_n)$ für grosse n näherungsweise $N(\mu, \sigma^2/n)$-verteilt ist.

Die Verteilungsfunktion der Zufallsvariablen $X_1, \ldots, X_n$ sei mit F bezeichnet. Was sagt die Messreihe über F aus?

Es ist intuitiv einleuchtend, dass die empirische Verteilungsfunktion

$$F_n(z; x_1, x_2, \ldots, x_n) = \frac{\text{Zahl der } x_i \text{ mit } x_i \le z}{n}$$

in engem Zusammenhang zur Wahrscheinlichkeit

$$F(z) = P(X_1 \le z)$$

stehen muss. Es gilt:

Satz 8.7.3 (Zentralsatz der Statistik)
Sei $X_1, X_2, \ldots$ eine Folge unabhängiger identisch verteilter Zufallsvariablen mit der Verteilungsfunktion F und bezeichne

$$D_n(X_1, \ldots, X_n) = \sup_{z \in \mathbb{R}} |F_n(z; X_1, \ldots, X_n) - F(z)|$$

die zufällige Maximalabweichung zwischen empirischer und "wahrer" Verteilungsfunktion. Dann gilt

$$P(\lim_{n \to \infty} D_n(X_1, \ldots, X_n) = 0) = 1,$$

$D_n(X_1, \ldots, X_n)$ konvergiert also mit Wahrscheinlichkeit 1 gegen 0.

8.8 Testverteilungen und Quantilapproximationen

In der Statistik, insbesondere in der Testtheorie, werden die folgenden Verteilungen benötigt, die von der Normalverteilung abgeleitet werden:

Seien $Z_1, \ldots, Z_n$ unabhängige, identisch $N(0, 1)$-verteilte Zufallsgrößen.

χ^2_r-Verteilung:

Es sei $1 \le r \le n$. Eine Zufallsvariable X heißt χ^2_r-verteilt, falls sie die Verteilungsfunktion besitzt

$$F(x) = P(Z_1^2 + \ldots + Z_r^2 \le x), \quad x \in \mathbb{R}.$$

t_r-Verteilung:

Es sei $1 \le r \le n-1$. Eine Zufallsvariable X heißt t_r-verteilt, falls sie die Verteilungsfunktion besitzt

$$F(x) = P\left(\frac{Z_{r+1}}{\sqrt{(Z_1^2 + \ldots + Z_r^2)/r}} \le x\right), \quad x \in \mathbb{R}.$$

$F_{r,s}$**-Verteilung:**

Es sei $1 \leq r, s \leq n-1$ mit $r+s \leq n$. Eine Zufallsvariable X heißt F-verteilt mit r und s Freiheitsgraden, falls sie die Verteilungsfunktion besitzt

$$F(x) = P\left(\frac{(Z_1^2 + \ldots + Z_r^2)/r}{(Z_{r+1}^2 + \ldots + Z_{r+s}^2)/s} \leq x\right), \quad x \in \mathbb{R}.$$

Die Dichten dieser Verteilungen können unter Verwendung der Gamma-Funktion

$$\Gamma(x) = \int_0^\infty e^{-t}\, t^{x-1}, \quad x > 0$$

und der Beta-Funktion

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\, dt, \quad \alpha, \beta > 0$$

angegeben werden.

Bezeichnungen für Quantile: Allgemein ist das p-Quantil x_p für eine stetig verteilte Zufallsgröße mit Verteilungsfunktion F gegeben durch

$$F(x_p) = p.$$

Bezeichnungen:

u_p	p-Quantil der $N(0,1)$-Verteilung
$t_{r;p}$	p-Quantil der t_r-Verteilung
$\chi_{r;p}$	p-Quantil der χ_r^2-Verteilung
$F_{r,s;p}$	p-Quantil der $F_{r,s}$-Verteilung

Für gängige Werte von p existieren Tabellen für diese Quantile.

8.8.1 Wichtige Anwendungsbeispiele

Seien $X_1, \ldots, X_n$ unabhängige, identisch $N(\mu, \sigma^2)$-verteilte Zufallsvariable. Bilden wir ihr arithmetisches Mittel

$$\bar{X}_{(n)} := \frac{1}{n}\sum_{i=1}^n X_i$$

und die Stichprobenvarianz

$$S_{(n)}^2 := \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_{(n)})^2$$

dann gilt:

Satz 8.8.1 *Es seien $X_1, \ldots, X_n$ unabhängige, identisch $N(\mu, \sigma^2)$-verteilte Zufallsvariable. Dann gilt:*

- $\bar{X}_{(n)}$ *ist $N(\mu, \sigma^2/n)$-verteilt,*
- $\frac{n-1}{\sigma^2} S^2_{(n)}$ *ist χ^2_{n-1}-verteilt,*
- $\bar{X}_{(n)}$ *und* $S^2_{(n)}$ *sind unabhängig,*
- $\sqrt{n}\frac{\bar{X}_{(n)}-\mu}{\sqrt{S^2_{(n)}}}$ *ist t_{n-1}-verteilt.*