
Kapitel 11

Wahrscheinlichkeitsrechnung und Statistik

- Deskriptive Statistik
 - Wahrscheinlichkeitsrechnung
 - Zufallsvariable und Verteilungsfunktion
 - Funktionen in mehreren Veränderlichen
 - Anwendungen
-

Große Datenmengen

In vielen Anwendungsbereichen treten große Datenmengen auf. Z.B.:

- statistische Bevölkerungsdaten bei Landesämtern,
- Bonitätsdaten von Unternehmen und Privatpersonen bei Hermes oder Schufa,
- Messdaten bei technischen Prozessen, usw.

Solche Datenmengen lassen sich wegen ihres Umfangs nur mit Hilfe von Kennziffern und Graphiken charakterisieren.

Zur Definition und Erläuterung der eben erwähnten Instrumente ist jedoch ein Beispiel mit einer kleineren Datenmenge sinnvoller.

Beispiel

Bei einer Aktie wurden 25 Wochenschlusskurse beobachtet und festgehalten. Es ergaben sich in Euro:

47.7; 50.8; 50.4; 52.2; 48.2; 49.3; 50.9; 50.3; 49.1; 52.4; 49.6;
50.8; 50.0; 48.9; 51.4; 48.7; 48.8; 49.9; 50.2; 49.0; 51.8; 49.6;
48.6; 51.3; 50.1.

Anleger interessieren sich nun natürlich dafür, welchen durchschnittlichen Kurs die Aktie im Beobachtungszeitraum hatte, wie stark die Abweichungen vom Mittel (=Volatilität) waren und wie sich diese verteilen.

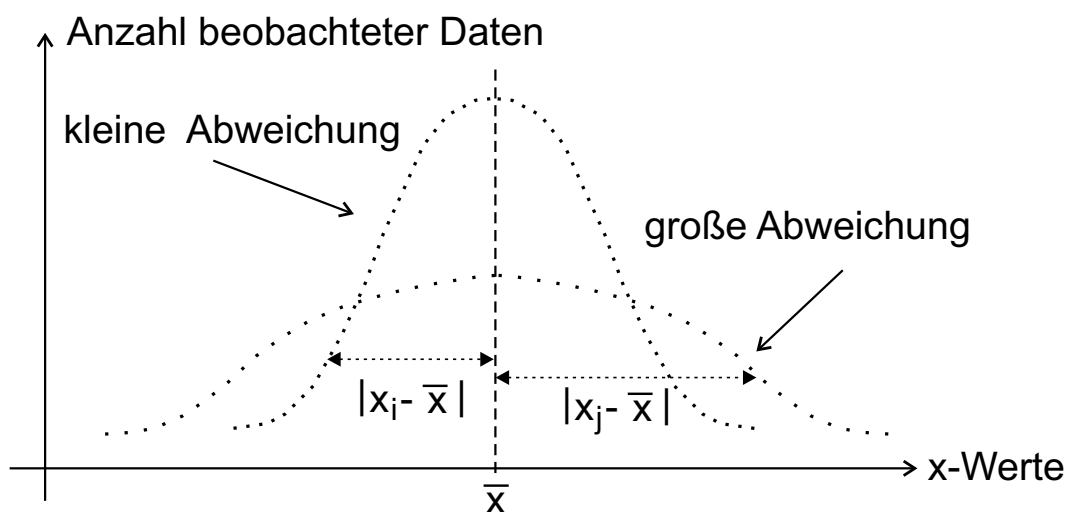
Mittelwert

Definition

Der Mittelwert (arithmetisches Mittel) der Datenwerte x_1, \dots, x_n ist definiert durch

$$\bar{x} := \frac{1}{n} \sum_{i=1}^n x_i.$$

Der Mittelwert \bar{x} alleine sagt nun nichts darüber aus, ob die Beobachtungswerte eng beieinander oder aber weit auseinander liegen.



Varianz

Für die Größe der einzelnen Abweichungen vom Mittelwert \bar{x} könnte man $|x_i - \bar{x}|$ als Maß wählen. Wegen der Unhandlichkeit des Betrages hat sich als Maß jedoch $(x_i - \bar{x})^2$ durchgesetzt. Natürlich interessiert man sich auch hier für das arithmetische Mittel dieser Abweichungen, das als *empirische Varianz* bezeichnet wird:

Definition

Die empirische Varianz (mittlere quadratische Abweichung) der Datenwerte x_1, \dots, x_n vom Mittelwert \bar{x} ist definiert durch

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

$S := \sqrt{S^2}$ heißt empirische Standardabweichung.

Beispiel

Für die Aktienkursdaten des Beispiels ergibt sich der Mittelwert zu

$$\bar{x} = \frac{1}{25} (47.7 + 50.8 + 50.4 + 52.2 + 48.2 + \dots \\ \dots + 51.8 + 49.6 + 48.6 + 51.3 + 50.1) = \frac{1250}{25} = 50.$$

Die empirische Varianz beträgt

$$s^2 = \frac{1}{25} \sum_{i=1}^{25} (x_i - 50)^2 \\ = \frac{1}{25} [(47.7 - 50)^2 + (50.8 - 50)^2 + (50.4 - 50)^2 + \dots \\ \dots + (51.3 - 50)^2 + (50.1 - 50)^2] = 1.4936.$$

Verschiebungsformel

Die empirische Varianz s^2 lässt sich häufig einfacher mit der so genannten *Verschiebungsformel* berechnen:

$$s^2 = \overline{x^2} - \bar{x}^2.$$

Dabei bezeichnet

$$\overline{x^2} := \frac{1}{n} \sum_{i=1}^n x_i^2$$

den Mittelwert der „quadrierten“ Daten.

Beispiel

Mit der Verschiebungsformel ergibt sich die Varianz des früheren Beispiels analog zu

$$\begin{aligned} s^2 &= \frac{1}{25} (47.7^2 + \dots + 50.1^2) - 50^2 \\ &= 2501.4936 - 50^2 = 1.4936. \end{aligned}$$

Übung

Die Marketingabteilung eines Unternehmens beobachtet in 40 aufeinanderfolgenden Monaten für ein Produkt folgende Absatzmengen:

499	484	493	487	500	493	504	507	485	501
495	497	490	510	494	502	494	491	502	494
509	488	500	498	505	508	492	504	493	503
514	503	488	486	493	496	499	487	498	497

Berechnen Sie Mittelwert \bar{x} , empirische Varianz S^2 und empirische Standardabweichung S .

Lösung

Der Mittelwert ergibt sich zu:

$$\begin{aligned}\bar{x} &= \frac{1}{40} \sum_{i=1}^{40} x_i = \frac{1}{40} (499 + 484 + 493 + \dots + 487 + 498 + 497) \\ &= 497.075.\end{aligned}$$

Für die empirische Varianz erhält man

$$\begin{aligned}s^2 &= \frac{1}{40} \sum_{i=1}^{40} (x_i - \bar{x})^2 = \frac{1}{40} \left((499 - 497.075)^2 + \dots \right. \\ &\quad \left. \dots + (497 - 497.075)^2 \right) \\ &= 53.8194,\end{aligned}$$

und daraus die empirische Standardabweichung

$$s = \sqrt{53.8194} = 7.3362.$$

Klasseneinteilung und Histogramme

Die Charakterisierung von Datenmengen unterstützt man in der Praxis auch durch die graphische Darstellung in Histogrammen. Dazu wird der Wertebereich der Daten x_1, \dots, x_n durch die Wahl von $k + 1$ Punkten $\xi_0 < \xi_1 < \dots < \xi_k$ in Intervalle $(\xi_{i-1}, \xi_i]$, die man *Klassen* nennt, eingeteilt. Dabei ist

$$\xi_0 < \min_{j=1, \dots, n} x_j \quad \text{und} \quad \xi_k > \max_{j=1, \dots, n} x_j$$

zu wählen. Wir bezeichnen mit n_i die Anzahl der Daten, die in der Klasse $(\xi_{i-1}, \xi_i]$ liegen, und nennen den Quotienten

$$h_i = \frac{n_i}{n}$$

relative Häufigkeit.

Diese Kennzahl gibt Auskunft über die Anordnungs-dichte der Daten für jede Klasse. Ihre Werte gehen in eine Funktion ein, die *empirische Dichte* genannt wird.

Empirische Dichte, Histogramm

Definition

Sind $0 \leq h_i \leq 1$, $1 \leq i \leq k$, relative Häufigkeiten bzgl. einer Klasseneinteilung $\xi_0 < \xi_1 < \dots < \xi_k$, dann heißt die Funktion

$$\hat{f}(x) := \begin{cases} \frac{h_i}{\xi_i - \xi_{i-1}} & , \text{ falls } x \in (\xi_{i-1}, \xi_i], i = 1, \dots, k, \\ 0 & , \text{ sonst} \end{cases}$$

empirische Dichte. Ihre graphische Darstellung wird Histogramm genannt.

Begriffe bei Histogrammen

Im Zusammenhang mit Histogrammen werden einige Begriffe benutzt, die wir hier definieren wollen:

- Der kleinste Wert ξ_0 heißt *Reduktionslage*.
- Die gesamte Länge $\xi_k - \xi_0$ nennt man *Variationsbreite*.
- Als *Klassenbreite* bezeichnet man die Werte $\xi_i - \xi_{i-1}$. Sie wird häufig konstant gewählt.
- Der Wert k heißt *Klassenzahl*. Diese wird häufig ungerade und ungefähr gleich \sqrt{n} gewählt (n Anzahl Datenwerte), es sollte jedoch $5 \leq k \leq 25$ gelten.

Beispiel

Für die Aktienkurse des früheren Beispiels wollen wir ein Histogramm mit Reduktionslage 47.5, konstanter Klassenbreite 1 und Variationsbreite 5 zeichnen.

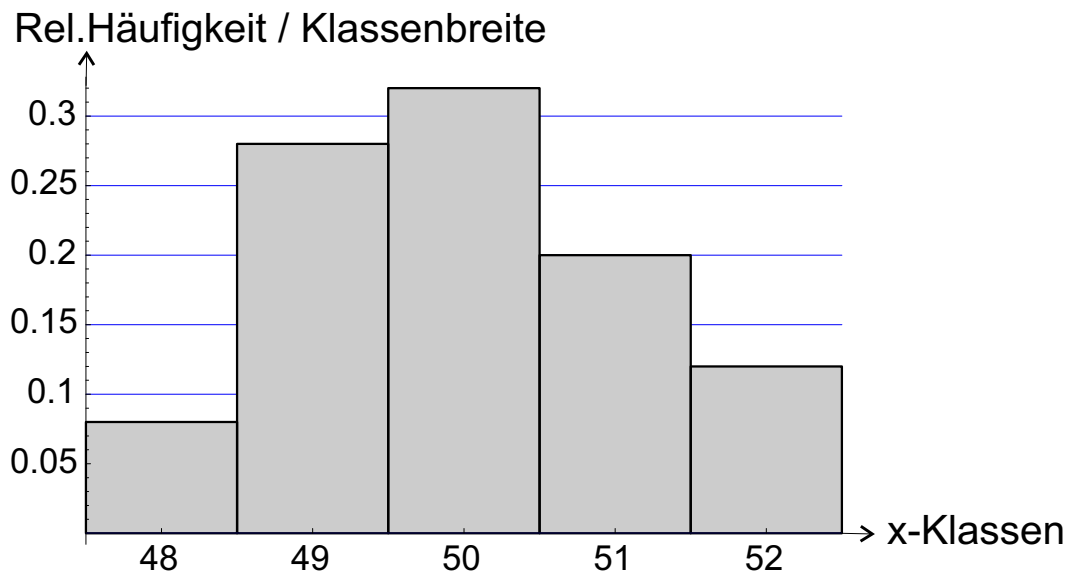
Der Wertebereich der Daten erstreckt sich von 47.7 bis 52.4. Die erste Klasse ergibt sich zu $(\xi_0, \xi_1] = (47.5; 48.5]$. Damit gilt $n_1 = 2$, denn nur die beiden Kurse 47.7 und 48.2 liegen in diesem Intervall. Es ist $n = 25$ und die relative Häufigkeit für diese Klasse ergibt sich zu $h_1 = \frac{2}{25} = 0.08$.

Analog erhält man alle weiteren Klassen mit ihren relativen Häufigkeiten:

<i>Klasse</i>	<i>Anzahl Kurse</i>	<i>rel. Häufigkeit</i>
$47.5 < x \leq 48.5$	2	0.08
$48.5 < x \leq 49.5$	7	0.28
$49.5 < x \leq 50.5$	8	0.32
$50.5 < x \leq 51.5$	5	0.20
$51.5 < x \leq 52.5$	3	0.12

Beispiel – Fortsetzung

Das sich daraus ergebende Histogramm sieht folgendermaßen aus:



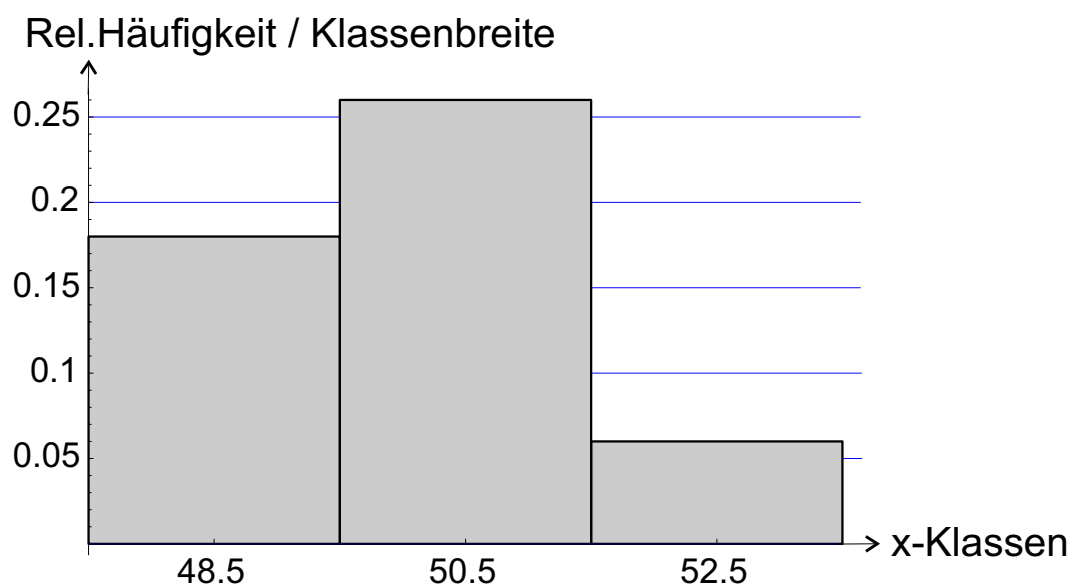
Die Säulenhöhen des Histogramms entsprechen hier wegen der konstanten Klassenbreite 1 der relativen Häufigkeit (was i. Allg. nicht der Fall ist!).

Beispiel – Fortsetzung

Wählt man bei gleicher Reduktionslage beispielsweise die konstante Klassenbreite 2, so erhält man:

Klasse	Anzahl Kurse	rel. Häufigkeit	$h_i/2$
$47.5 < x \leq 49.5$	9	0.36	0.18
$49.5 < x \leq 51.5$	13	0.52	0.26
$51.5 < x \leq 53.5$	3	0.12	0.06

und daraus ein Histogramm, bei dem die Säulen nur halb so hoch ($h_i/2!$) sind wie die relativen Häufigkeiten.



Säulenfläche und relative Häufigkeit

Das letzte Beispiel zeigt, dass die Säulenfläche und *nicht* die Säulenhöhe die relative Häufigkeit symbolisiert!

Summiert man über die gesamte Fläche — also über alle relativen Häufigkeiten —, so muss sich der Wert 1 ergeben, d.h. es gilt stets

$$\int_{\mathbb{R}} \hat{f}(x) dx = \int_{-\infty}^{\infty} \hat{f}(x) dx = 1.$$

Empirische Verteilungsfunktion

„Wie oft lag der Aktienkurs über 50 Euro?“ „In wie vielen Monaten wurde das Absatzmengenziel von mindestens 500 nicht erreicht?“

Zur Beantwortung dieser Fragen benötigen wir *kumulierte Häufigkeiten*, die die *empirische Verteilungsfunktion* zur Verfügung stellt:

Für die empirische Verteilungsfunktion $\hat{F}(x) := \int_{-\infty}^x \hat{f}(t) dt$ gilt ($i = 1, \dots, k$):

$$\hat{F}(x) = \begin{cases} 0 & , \text{ falls } x \leq \xi_0, \\ \sum_{j=1}^{i-1} h_j + \frac{x - \xi_{i-1}}{\xi_i - \xi_{i-1}} h_i & , \text{ falls } \xi_{i-1} < x \leq \xi_i, \\ 1 & , \text{ falls } x > \xi_k. \end{cases}$$

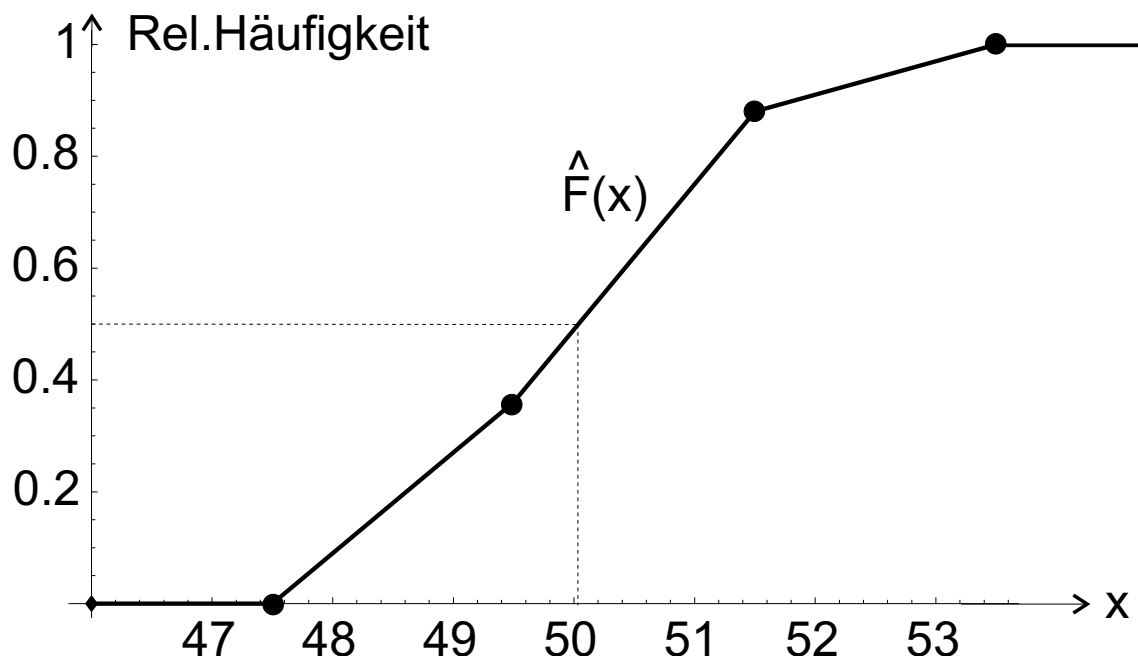
$\hat{F}(x)$ ist ein Maß für die Anzahl der Daten, deren Messwert kleiner gleich x ist.

Empirische Verteilungsfunktion des Beispiels

Die Formel ist leicht einzusehen:

Man kumuliert die Häufigkeiten und interpoliert dann stückweise linear.

Nachfolgende Abb. zeigt die empirische Verteilungsfunktion des früheren Beispiels zur Klassenbreite 2.



Weitere Lageparameter

Definition

Das x_p -Quantil ist der Wert x_p , bis zu dem $100 \cdot p$ % der Daten liegen, d.h. es muss gelten:

$$\int_{-\infty}^{x_p} \hat{f}(x) dx = p \quad \text{bzw.} \quad \hat{F}(x_p) = p \quad \text{mit} \quad 0 \leq p \leq 1.$$

Das Quantil $x_{0.5}$ heißt Median, da es die Datenmenge „halbiert“: 50% aller Daten liegen links bis zum Median, die anderen rechts davon. Die Quantile $x_{0.25}$ und $x_{0.75}$ bezeichnet man als Quartile.

Übung

Für das Aktienkursbeispiel sei die bereits besprochene zweite Klasseneinteilung

$$x_0 = 47.5 < 49.5 < 51.5 < 53.5 = x_3$$

angenommen.

- a) *Bestimmen Sie die zugehörige Verteilungsfunktion $\hat{F}(x)$.*

- b) *Wo liegt unter Zugrundelegung von $\hat{F}(x)$ der Median?*

Lösung

- a) Entnimmt man die relativen Häufigkeiten der Tabelle, so ergibt sich $\hat{F}(x)$ zu:

$$\hat{F}(x) = \begin{cases} 0 & , \text{ falls } x \leq 47.5, \\ (x - 47.5)/2 \cdot 0.36 & , \text{ falls } 47.5 < x \leq 49.5, \\ 0.36 + (x - 49.5)/2 \cdot 0.52 & , \text{ falls } 49.5 < x \leq 51.5, \\ 0.88 + (x - 51.5)/2 \cdot 0.12 & , \text{ falls } 51.5 < x \leq 53.5, \\ 1 & , \text{ falls } x > 53.5. \end{cases}$$

- b) Die Bestimmungsgleichung für den Median $x_{0.5}$ ist :

$$0.5 = \hat{F}(x_{0.5}) = 0.36 + (x - 49.5)/2 \cdot 0.52.$$

Auflösen dieser Gleichung liefert $x_{0.5} \approx 50.04$. Tatsächlich liegen 13 Aktienkurse unter diesem Wert, 12 darüber.

Regression und Korrelation

In der Praxis interessiert häufig nicht — wie bisher — eine (eindimensionale) Datenmenge, sondern eine Liste von Datenwerten, die zweidimensional sind:

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

Untersucht werden sollen funktionale Zusammenhänge zwischen den Variablen, d.h. wir betrachten x als *unabhängige Variable* und y als *abhängige Variable*. Die Notwendigkeit solcher Untersuchungen tritt bei vielen Fragestellungen auf:

- Wie stark hängt die Dauer der Arbeitslosigkeit vom Alter ab?
- Inwiefern beeinflusst eine Preissenkung die Absatzmenge eines Produktes?
- In welchem Umfang wirkt sich der Stand eines Aktienindex auf den Kurs einer Aktie aus?

Lineare Regressionsaufgabe

Interessant ist in diesem Zusammenhang, ob die y_i durch eine Funktion der x_i genügend genau dargestellt werden können, etwa $y_i = f(x_i)$. Wir wollen uns dabei auf den einfachsten Fall, nämlich den einer linearen Funktion f , beschränken:

Können die y_i durch $bx_i + a$ mit geeignet zu bestimmenden a, b genügend genau dargestellt werden? Gilt also $y_i \approx a + bx_i$?

Dies führt auf die so genannte *Lineare Regressionsaufgabe* der Bestimmung von a und b im Ansatz

$$y_i = bx_i + a + E_i,$$

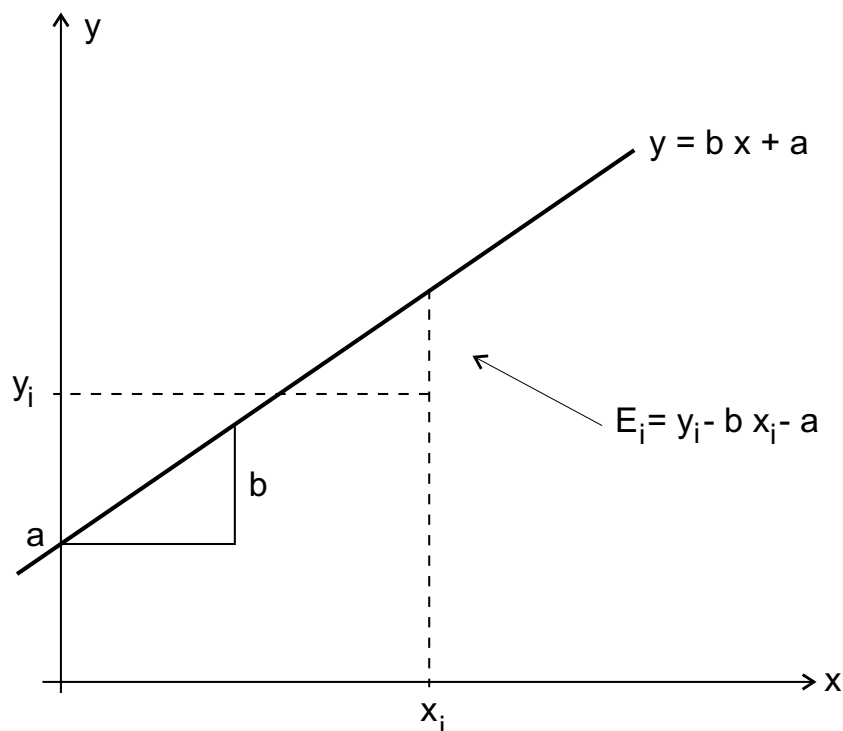
wobei

$$E_i := y_i - (bx_i + a)$$

eine Fehlergröße ist. Die Koeffizienten a und b der Geraden werden so berechnet, dass die Summe der quadrierten Fehler minimiert wird. Man nennt dieses Vorgehen deshalb die *Methode der kleinsten Quadrate*.

Methode der kleinsten Quadrate

Da natürlich i. Allg. die y -Werte nicht streng von den x -Werten abhängen, bildet die Datenmenge eine „Punktwolke“. Bezüglich dieser „Punktwolke“ sucht man nun eine Gerade, die die Abweichungen E_i zwischen den beobachteten Werten y_i und den zu x_i gehörenden Geradenwerten $bx_i + a$ „optimal ausgleicht“. Naheliegender ist, zu verlangen, dass die Summe der Abweichungsquadrate möglichst klein wird.



Gefunden werden muss also das Minimum der Funktion

$$S_E^2 = \sum_{i=1}^n E_i^2 = \sum_{i=1}^n (y_i - bx_i - a)^2.$$

Eine notwendige Bedingung hierfür ist das Verschwinden der partiellen Ableitungen nach b und a :

$$\begin{aligned}\frac{\partial S_E^2}{\partial a} &= -2 \sum_{i=1}^n (y_i - bx_i - a) \stackrel{!}{=} 0, \\ \frac{\partial S_E^2}{\partial b} &= -2 \sum_{i=1}^n x_i (y_i - bx_i - a) \stackrel{!}{=} 0.\end{aligned}$$

Division durch -2 und Zusammenfassung geeigneter Summen führen auf die so genannten *Normalgleichungen*, aus denen sich a und b *eindeutig* bestimmen lassen:

$$\begin{aligned}n \cdot a + \left(\sum_{i=1}^n x_i\right) \cdot b &= \sum_{i=1}^n y_i, \\ \left(\sum_{i=1}^n x_i\right) \cdot a + \left(\sum_{i=1}^n x_i^2\right) \cdot b &= \sum_{i=1}^n x_i y_i.\end{aligned}$$

Die partiellen Ableitungen 2. Ordnung zeigen, dass die Lösung der Normalgleichungen tatsächlich ein Minimum von S_E^2 liefert.

Beispiel

Für nachfolgende Messpunkte x_i mit zugehörigen Messwerten y_i bestimmen wir die Regressionsgerade durch Lösen der Normalgleichungen:

i	1	2	3	4
x_i	1.5	2.1	3.3	3.9
y_i	4.1	5.0	8.1	8.7

Zunächst ermittelt man $\sum_{i=1}^n x_i = 1.5 + 2.1 + 3.3 + 3.9 = 10.8$ und $\sum_{i=1}^n y_i = 4.1 + 5.0 + 8.1 + 8.7 = 25.9$. Mittels Taschenrechner berechnet sich ebenfalls leicht: $\sum_{i=1}^n x_i^2 = 1.5^2 + \dots + 3.9^2 = 32.76$ und $\sum_{i=1}^n x_i y_i = 1.5 \cdot 4.1 + \dots + 3.9 \cdot 8.7 = 77.31$. Die Normalgleichungen lauten damit:

$$\begin{aligned}4a + 10.8b &= 25.9, \\10.8a + 32.76b &= 77.31.\end{aligned}$$

Aus diesen folgt unmittelbar die Lösung $b = 2.05$ und $a = 0.94$. Die Regressionsgerade ergibt sich zu

$$y(x) = 2.05x + 0.94.$$

Lineare Regression

Definition

Die Gerade $y(x) = bx + a$ zu den Werten $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ mit den Koeffizienten

$$b = \frac{S_{xy}}{S_x^2} \quad \text{und} \quad a = \bar{y} - b\bar{x}, \quad (\text{falls } S_x > 0)$$

heißt Regressionsgerade zu y bzgl. x . Dabei stehen \bar{x}, \bar{y} für die Mittelwerte der x - bzw. y -Werte, S_x^2 für die Varianz der x -Werte und

$$S_{xy} := \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \cdot \bar{y}$$

für die empirische Kovarianz der x - und y -Werte.

Kovarianz und Korrelationskoeffizient

Die neu eingeführte Kovarianz kann wegen

$$\begin{aligned}\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) &= \frac{1}{n} \sum_{i=1}^n (x_i y_i - x_i \bar{y} - \bar{x} y_i + \bar{x} \bar{y}) \\ &= \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \bar{y} - \bar{x} \bar{y} + \bar{x} \bar{y} = S_{xy}\end{aligned}$$

auch aufgefasst werden als der Mittelwert des Produktes der Abweichungen der einzelnen Daten von ihrem jeweiligen Mittel. Üblicherweise wird dieses Maß noch normiert.

Kovarianz und Korrelationskoeffizient

Die (empirische) Kovarianz S_{xy} ergibt sich auch zu

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

Die entsprechende normierte Größe

$$r_{xy} := \frac{S_{xy}}{S_x S_y} \quad \text{mit} \quad -1 \leq r_{xy} \leq 1$$

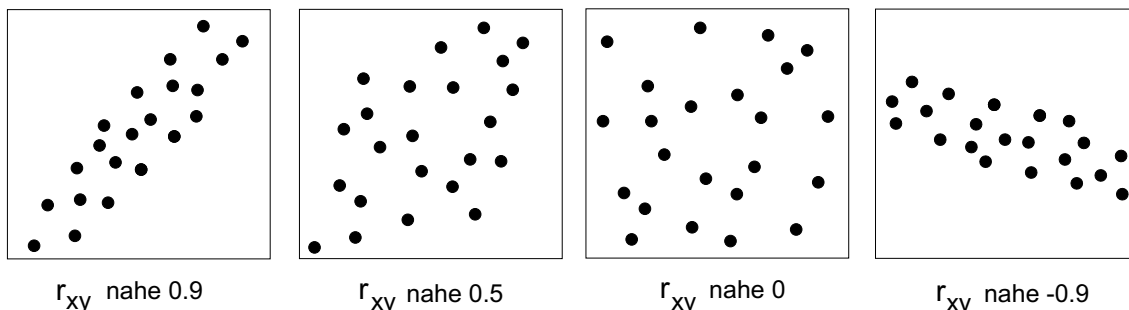
nennt man (empirischen) Korrelationskoeffizienten. S_x, S_y sind dabei die Standardabweichungen der x - bzw. y -Werte.

Der Korrelationskoeffizient

Da eine große Kovarianz auch durch eine große Streuung der x - und y -Datenmengen verursacht sein kann, benutzt man als Maß dafür, wie gut der lineare Zusammenhang ist, den Korrelationskoeffizienten r_{xy} .

Ist dieser nahe bei 0, dann sind die x - und y -Werte fast *unkorreliert*, für Werte nahe bei $+1$ bzw. -1 sind sie sehr gut *positiv* bzw. *negativ* korreliert.

Die Bedeutung von r_{xy} veranschaulicht auch folgende Abbildung:



Beispiel

Für die Datenmenge des früheren Beispiels soll die Korrelation ermittelt werden. Mittels Taschenrechner berechnet man die Standardabweichungen

$$S_x \approx 0.948683, S_y \approx 1.962619$$

und die Kovarianz

$$S_{xy} = 1.845.$$

Der Korrelationskoeffizient ergibt sich somit zu

$$\begin{aligned} r_{xy} &= \frac{S_{xy}}{S_x S_y} = \frac{1.845}{0.948683 \cdot 1.962619} \\ &= 0.990922. \end{aligned}$$

Da der Korrelationskoeffizient nahe bei 1 liegt, ist die Anpassung durch die Regression sehr gut.

Ereignisse und Wahrscheinlichkeiten

In der *Wahrscheinlichkeitsrechnung* untersucht man zufällige (stochastische) Erscheinungen. Diese zeichnen sich dadurch aus, dass sie unter bestimmten Bedingungen eintreten können, es aber nicht müssen.

So schwanken bei der Wiederholung eines Experimentes — darunter wollen wir Beobachtungen, Messungen, Proben, Tests, etc. verstehen — unter ansonsten gleichen Voraussetzungen die Ergebnisse in der Regel mehr oder weniger stark. Man spricht dann von einem *Zufallsexperiment*.

Wiederholt man dieses genügend oft, so ergeben sich trotz der Schwankungen der jeweiligen Einzelergebnisse gewisse *Gesetzmäßigkeiten* für die *Gesamtheit* aller Experimente. Die Grundlagen, um diese Gesetzmäßigkeiten aufzuspüren, werden jetzt bereitgestellt.

Ergebnismenge und Ereignisraum

Definition

Bei einem Zufallsexperiment sind stets mehrere so genannte Elementarereignisse möglich, die sich gegenseitig ausschließen. Die Menge aller Elementarereignisse

$$\Omega = \{\omega_1, \omega_2, \omega_3, \dots\}$$

heißt Ergebnismenge. Durch Zusammenfassung beliebiger Elementarereignisse erhält man Teilmengen von Ω , die man als Ereignisse bezeichnet. Die Menge aller Ereignisse, die sich aus Ω bilden lässt, heißt Ereignisraum E (= Potenzmenge von Ω).

Zu beachten ist, dass insbesondere die leere Menge \emptyset bzw. Ω selbst Ereignisse sind, die man als *unmögliches Ereignis* bzw. als *sicheres Ereignis* bezeichnet.

Weitere Ereignisse

Ausgehend von Ereignissen $A, B \subseteq \Omega$ kann man mit Hilfe der üblichen Mengenoperatoren folgende neue Ereignisse definieren:

- a) Das Ereignis $A \cap B$ (in der Literatur häufig auch als $A \cdot B$ notiert) tritt genau dann ein, wenn A und B eintreten. Man sagt, dass sich die beiden Ereignisse A, B *gegenseitig ausschließen* bzw. *unvereinbar sind*, falls $A \cap B = \emptyset$ gilt.

- b) Das Ereignis $A \cup B$ (bzw. $A + B$) trifft genau dann zu, wenn mindestens eines der beiden Ereignisse A oder B eintritt.

- c) Mit \bar{A} bezeichnet man das zu A *komplementäre Ereignis*, welches sich aus der Differenz $\Omega \setminus A$ ergibt. \bar{A} tritt genau dann ein, wenn A *nicht* eintritt.

Beispiel

- a) *Beim Werfen einer Münze gibt es zwei mögliche Elementarereignisse, nämlich „Kopf“ ($\omega_1 = K$) oder „Zahl“ ($\omega_2 = Z$). Damit gilt*

$$\Omega = \{K, Z\} \quad \text{und} \quad E = \{\emptyset, \{K\}, \{Z\}, \Omega\}.$$

- b) *Verkehrssampel liefert vier möglichen Elementarereignissen:*

Grün (G), Orange/Gelb (O), Rot (R) oder — zur Einleitung der Grünphase — gleichzeitig Rot und Orange (RO).

Ergebnismenge: $\Omega = \{G, O, R, RO\}$, Ereignisraum besteht aus 2^4 Ereignissen.

Steht die Ampel nicht auf Grün, so kann man dafür $\overline{\{G\}}$ schreiben oder kürzer \overline{G} .

Wahrscheinlichkeitsmaß

Ob bei der Durchführung eines Versuches ein ganz bestimmtes Elementarereignis ω_i eintreten wird, weiß man nicht. Also benötigt quantitatives Maß für die Wahrscheinlichkeit des Eintretens:

Je stärker man vom Eintreten des Ereignisses überzeugt ist, desto größer sollte die Zahl sein. Hierzu muss man aber nicht alle reellen Zahlen benutzen, sondern kann sich auf das Intervall $[0, 1]$ beschränken.

Bei der historisch bedingten *klassischen Definition der Wahrscheinlichkeit* setzt man die *Gleichwahrscheinlichkeit (Laplace-Annahme)* aller Elementarereignisse voraus.

Abzählregel

Ist die Ergebnismenge

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

endlich, so ordnet man jedem Elementarereignis als

Wahrscheinlichkeit die Zahl $\frac{1}{n}$

zu. Außerdem definiert man die Wahrscheinlichkeit $P(A)$ eines Ereignisses A aus dem Ereignisraum durch die so genannte *Abzählregel*

$$P(A) = \frac{|A|}{n} = \frac{\text{Anzahl der Elemente von } A}{\text{Anzahl der Elemente von } \Omega}.$$

Diese Definition entspricht unserer Vorstellung, dass ein Ereignis A umso wahrscheinlicher ist, je größer die Anzahl der dieses Ereignis auslösenden Elementarereignisse ist.

$P(A)$ nennt man auch *Laplace'sche Wahrscheinlichkeit*.

Beispiel

Beim Wurf mit einer ausgewogenen Münze (= Laplace-Annahme) haben wir zwei gleichwahrscheinliche Elementarereignisse, d.h.

$$P(\text{Zahl}) = P(\text{Kopf}) = 1/2.$$

Übung

Eine Münze und ein Würfel werden gemeinsam geworfen.

Wie groß ist die Laplace'sche Wahrscheinlichkeit dafür, dass die Münze „Zahl“ und der Würfel eine gerade Augenzahl anzeigt (=Ereignis A)?

Lösung

Zunächst muss die Ergebnismenge geeignet konstruiert werden, d.h. es müssen gleichwahrscheinliche Elementarereignisse definiert werden. Mit Z für „Zahl“ und K für „Kopf“ erhalten wir:

$$\Omega = \{ (Z, 1), (Z, 2), (Z, 3), (Z, 4), (Z, 5), (Z, 6), (K, 1), (K, 2), (K, 3), (K, 4), (K, 5), (K, 6) \}.$$

Jedes Elementarereignis hat damit die Wahrscheinlichkeit $1/12$. Das Ereignis

$$A = \{ (Z, 2), (Z, 4), (Z, 6) \}$$

besteht aus drei Elementarereignissen und somit gilt

$$P(A) = 3/12 = 1/4.$$

Relative Häufigkeit

Das Prinzip der Gleichwahrscheinlichkeit hat seine Grenzen. So sind bei der Ampel die vier Elementarereignisse G, O, R, RO offensichtlich nicht gleichwahrscheinlich.

Allgemein wird man einen Versuch mit einer Menge Ω von Elementarereignissen n -mal wiederholen. Ist nun A ein zum Versuch gehörendes Ereignis und tritt dieses m -mal ein, so ist

$$h_n = \frac{m}{n} \in [0, 1]$$

die *relative Häufigkeit* von A . Startet man eine neue Versuchsreihe mit n Versuchen, so kann sich h_n natürlich ändern. Die Praxis zeigt jedoch, dass sich h_n bei hinreichend groß gewähltem n nur unwesentlich ändert. Man nimmt daher oft an, dass die Folge $(h_n)_{n \in \mathbb{N}}$ einen Grenzwert hat und bezeichnet

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

als *statistische Wahrscheinlichkeit*. Mit dieser Vorgehensweise erhält man *keinen exakten* Wert für $P(A)$, aber eine *gute Näherung*, die für praktische Berechnungen ausreichend ist.

Das Axiomensystem von Kolmogoroff

Ereignissen können Wahrscheinlichkeiten in Form reeller Zahlen zugeordnet werden. Dabei sind jedoch gewisse Regeln einzuhalten, die man als *Axiomensystem von Kolmogoroff* bezeichnet:

Eine reellwertige Funktion $P : E \rightarrow \mathbb{R}$, die jedem Ereignis A aus dem Ereignisraum E eine Zahl $P(A)$ (=Wahrscheinlichkeit von A) zuordnet, heißt Wahrscheinlichkeitsfunktion, Wahrscheinlichkeitsmaß bzw. Verteilungsgesetz, wenn die folgenden Kolmogoroff'schen Axiome erfüllt sind:

- a) Für jedes Ereignis A gilt: $P(A) \geq 0$.
- b) Die Wahrscheinlichkeit des sicheren Ereignisses Ω ist 1, d.h. $P(\Omega) = 1$.
- c) Für paarweise unvereinbare Ereignisse A_1, A_2, A_3, \dots gilt:

$$P(A_1 \cup A_2 \cup A_3 \cup \dots) = P(A_1) + \\ + P(A_2) + P(A_3) + \dots$$

Rechenregeln für Wahrscheinlichkeiten

Aus den Kolmogoroff'schen Axiomen folgt:

a) Für das zu A komplementäre Ereignis \bar{A} gilt:

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

b) Für das unmögliche Ereignis \emptyset gilt:

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

c) Für das Ereignis $A \setminus B$ (Differenzmenge) gilt:

$$P(A \setminus B) = P(A) - P(A \cap B).$$

d) Für die Wahrscheinlichkeit der Summe zweier Ereignisse A, B gilt der sog. Additionssatz:

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

e) Für jedes Ereignis A gilt:

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

f) Das Wahrscheinlichkeitsmaß ist monoton:

$$A \subseteq B \implies P(A) \leq P(B).$$

Dass sich diese Regeln unmittelbar aus den Axiomen ergeben, sieht man leicht ein:

a) A und \bar{A} sind unvereinbar, zudem gilt

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

Also ergibt sich die Behauptung unmittelbar aus:

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

b) Unter Beachtung von $\emptyset = \bar{\Omega}$ und Regel a) gilt:

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

Die Regeln c)-f) lassen sich ebenfalls leicht verifizieren (siehe Aufgabe)

Die mit der *Abzählregel* bzw. mit Hilfe von *relativen Häufigkeiten* gebildeten Wahrscheinlichkeiten erfüllen übrigens das oben angeführte Axiomensystem.

Übung

Gegeben sei eine Ampel mit der Ergebnismenge $\Omega = \{G, O, R, RO\}$ (Grün (G), Orange/Gelb (O), Rot (R) oder — zur Einleitung der Grünphase — gleichzeitig Rot und Orange (RO)). Durch Messung wurden die Wahrscheinlichkeiten

$$P(G) = 1/3, P(R) = 1/2, P(O) = 1/9$$

festgestellt.

- a) Bestimmen Sie $P(RO)$ so, dass P eine Wahrscheinlichkeitsfunktion auf Ω darstellt.

- b) Wie groß ist die Wahrscheinlichkeit dafür, dass „Grün“ und „Orange“ gleichzeitig leuchten?

Lösung

a) Wenn P eine Wahrscheinlichkeitsfunktion ist, dann müssen die Axiome und Rechenregeln gelten. Man erhält deshalb:

$$\begin{aligned} P(RO) &= 1 - P(\overline{RO}) \\ &= 1 - P(G \cup R \cup O) \\ &= 1 - (P(G) + P(R) + P(O)) \\ &= 1 - 17/18 = 1/18. \end{aligned}$$

b) Das zu untersuchende Ereignis lässt sich als $G \cap O \subseteq \Omega$ schreiben. Es ist

$$P(G \cap O) = P(\emptyset) = 0.$$

Wahrscheinlichkeitsraum

Bisher hatten wir zur Beschreibung von Versuchen als Ereignisraum E die Potenzmenge $\mathcal{P}(\Omega)$ benutzt.

Eine Übung hat gezeigt, dass man einen Ereignisraum Ω mit einem Wahrscheinlichkeitsmaß P „ausstatten“ kann.

Im Allg. darf man dann aber nur solche Teilmengen von Ω als Ereignisse zulassen, die „vernünftig messbar“ sind. Ist nun \mathcal{A} eine Teilmenge von $\mathcal{P}(\Omega)$, die gewisse Messbarkeitseigenschaften erfüllt, dann nennt man das Tripel (Ω, \mathcal{A}, P) einen *Wahrscheinlichkeitsraum*.

Wir werden nur Ergebnismengen Ω betrachten, für die man stets $\mathcal{A} = \mathcal{P}(\Omega)$ wählen kann.

Bedingte Wahrscheinlichkeiten

Eine wichtige Rolle in der Wahrscheinlichkeitsrechnung spielt die Fragestellung:

„Wie groß ist die Wahrscheinlichkeit des Ereignisses A unter der *Bedingung*, dass das Ereignis B eingetreten ist?“

Ein Beispiel soll dies illustrieren:

Eine Losbude verkauft 10 000 Lose mit gleich vielen grünen und gelben Losen, wobei 9 000 Lose Nieten sind. Die Gewinnwahrscheinlichkeit beträgt nach der Abzählregel daher $1000/10000 = 0.1$.

Wir nehmen nun an, dass 70% der Gewinne in grünen Losen stecken. Kauft man also ein grünes Los (Bedingung!), dann steigt die Gewinnwahrscheinlichkeit auf $700/5000 = 0.14$.

Bedingte Wahrscheinlichkeiten

Es stehe das Ereignis A für das Ziehen eines Gewinns, B für das Ziehen eines grünen Loses.

Das Ereignis $A \cap B$ steht dann für das Ziehen eines Glücksloses, das die Farbe grün hat.

Es seien insgesamt n Lose vorhanden, wobei sich in den g grünen Losen r Gewinne befinden sollen.

Dann können wir die Wahrscheinlichkeit für „ A unter der Bedingung B “ (=Gewinnwahrscheinlichkeit unter der Voraussetzung, dass ein grünes Los gezogen wurde!), im Zeichen $P(A|B)$, folgendermaßen berechnen:

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

In unserem Beispiel war $n = 10\,000$, $g = 5\,000$ und $r = 700$, somit $P(A \cap B) = 700/10000 = 0.07$ und $P(B) = 5000/10000 = 0.5$. Damit folgt $P(A|B) = 0.07/0.5 = 0.14$.

Bedingte Wahrscheinlichkeiten

Definition

Die Wahrscheinlichkeit für das Eintreten des Ereignisses A unter der Bedingung, dass das Ereignis B eingetreten ist, kurz $P(A|B)$, heißt **bedingte Wahrscheinlichkeit** und ist gegeben durch

$$P(A|B) := \frac{P(A \cap B)}{P(B)}, \text{ wenn } P(B) > 0.$$

Umformung der Gleichung $P(A|B) = \frac{P(A \cap B)}{P(B)}$ bzw. der analogen Gleichung

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

liefert den sog. **Multiplikationssatz für Wahrscheinlichkeiten**:

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

Multiplikationssatz

Mit dem Multiplikationssatz für Wahrscheinlichkeiten

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

kann man die Wahrscheinlichkeit für das gleichzeitige Eintreten zweier Ereignisse A und B berechnen.

Direkt aus dem Multiplikationssatz folgt:

Der Satz von Bayes lautet:

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

Beispiel

Die Durchfallquote liegt bei 30%. $\frac{5}{8}$ der Studenten bereiteten sich auf Prüfungen vor. 80% der Studenten, die die Prüfung bestanden haben, waren vorbereitet.

Wie hoch war die Durchfallquote bei den vorbereiteten Prüflingen?

Ist A das Ereignis, dass der Studierende vorbereitet war, so gilt $P(A) = \frac{5}{8}$. Für das Ereignis B , dass der Student die Prüfung bestanden hat, ergibt sich $P(B) = 0.7$. Die bedingte Wahrscheinlichkeit, dass der Student vorbereitet war, wenn er bestanden hat, ist $P(A|B) = 0.8$.

Zu berechnen ist nun die bedingte Wahrscheinlichkeit $P(B|A)$, die sich aus dem Multiplikationssatz zu

$$P(B|A) = \frac{P(B)}{P(A)}P(A|B) = 0.7 \cdot \frac{8}{5} \cdot 0.8 = 0.896$$

ergibt. Die Durchfallquote unter den Vorbereiteten beträgt damit lediglich 10.4%.

Unabhängige Ereignisse

Der Multiplikationssatz liefert die Wahrscheinlichkeit dafür, dass zwei Ereignisse A , B gleichzeitig eintreten. Nun ist es natürlich möglich, dass die Wahrscheinlichkeit für das Eintreten von Ereignis A nicht davon beeinflusst ist, ob das Ereignis B eingetreten ist oder nicht. Man definiert daher:

Definition

Zwei Ereignisse A und B heißen unabhängig, wenn

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

gilt. Speziell folgt für zwei unabhängige Ereignisse A und B aus dem Multiplikationssatz:

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

Beispiel (Bernoulli'sches Zufallsexperiment)

Ein Gerät bestehe aus $n = 100$ Bauteilen. Das Gerät sei so aufgebaut, dass ein defektes Bauteil das ganze Gerät funktionsuntüchtig macht. Ein Bauteil funktioniere mit der Wahrscheinlichkeit $p = 0.98$.

Wie groß ist die Wahrscheinlichkeit für die Funktionsfähigkeit des gesamten Gerätes?

Ereignisdefinitionen:

- A_i : „ i -tes Bauteil funktioniert“, $i = 1, \dots, n$,
- A : „Gerät funktioniert“.

Es gilt dann $P(A_1) = \dots = P(A_n) = p$. Unter der Annahme, dass die A_i paarweise unabhängig sind, gilt nach mehrmaliger Anwendung der Multiplikationsformel

$$\begin{aligned} P(A) &= P(A_1 \cap A_2 \cap \dots \cap A_n) \\ &= P(A_1) \cdot P(A_2) \cdot \dots \cdot P(A_n) = p^n \\ &= (0.98)^{100} \approx 0.1326. \end{aligned}$$

Bernoulli'sches Zufallsexperiment – Fortsetzung

Wir wollen nun die Wahrscheinlichkeit $P_k(n; p)$ dafür berechnen, dass genau k der Ereignisse A_1, \dots, A_n eintreten (d.h. genau k der n Bauteile funktionieren). In diesem Fall treten $(n - k)$ der komplementären Ereignisse $\bar{A}_1, \dots, \bar{A}_n$ (defekte Bauteile) auf. Für diese gilt $P(\bar{A}_1) = \dots = P(\bar{A}_n) = 1 - p$. Wegen der Unabhängigkeit der Ereignisse ergibt sich

$$P(A_1 \cap \dots \cap A_{k-1} \cap A_k \cap \bar{A}_{k+1} \cap \bar{A}_{k+2} \cap \dots \cap \bar{A}_n) = p^k (1-p)^{n-k},$$

$$P(A_1 \cap \dots \cap A_{k-1} \cap \bar{A}_k \cap A_{k+1} \cap \bar{A}_{k+2} \cap \dots \cap \bar{A}_n) = p^k (1-p)^{n-k},$$

usw. Alle diese Ereignisse sind unvereinbar. Ihre Wahrscheinlichkeiten addieren sich daher. Aus der Kombinatorik ist bekannt, dass es genau $\binom{n}{k}$ solche Summanden (Kombinationen) gibt. Somit gilt

$$P_k(n; p) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Bernoulli'sches Zufallsexperiment – Fortsetzung

Wahrscheinlichkeiten für Bernoulli-Experiment:

$$P_k(n; p) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

Die Wahrscheinlichkeit dafür, dass sich unter den 100 Bauteilen genau 3 defekte befinden, ist also gegeben durch

$$P_{97}(100; p) = \binom{100}{97} p^{97} (1 - p)^3 \approx 0.1823.$$

Auch die Wahrscheinlichkeit für die Funktionstüchtigkeit des Gerätes ergibt sich aus dieser Formel:

$$\begin{aligned} P(A) &= P_{100}(100; p) \\ &= \binom{100}{100} p^{100} (1 - p)^0 \\ &= p^{100} \approx 0.1326. \end{aligned}$$

Grundbegriffe

Ein adäquater Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) ist ein mathematisches Modell für ein Zufallsexperiment. Den Elementarereignissen ordnet man nun aus praktischen Gründen reelle Zahlen zu:

Definition

Eine Zufallsvariable bzw. Zufallsgröße ist eine reelle Funktion

$$X : \Omega \rightarrow \mathbb{R},$$

die jedem Elementarereignis $\omega \in \Omega$ genau eine reelle Zahl $X(\omega)$ zuordnet.

Ist der Wertebereich von X endlich oder abzählbar unendlich, so heißt X diskret, andernfalls stetig.

Beispiel

- a) *Beim Werfen einer Münze kann man dem Ergebnis „Kopf“ (ω_1) den Wert 0 und dem Ergebnis „Zahl“ (ω_2) den Wert 1 zuordnen. Man erhält dann die diskrete Zufallsvariable*

$$X(\omega_i) = i - 1$$

für $i = 1, 2$ mit den möglichen Werten 0, 1.

- b) *Wir zählen die Anzahl X der Autos, die in einem bestimmten Zeitintervall eine Kreuzung überqueren. X ist dabei eine diskrete Zufallsvariable mit den abzählbar unendlich vielen Werten*

$$0, 1, 2, \dots$$

- c) *Wiederholte Messungen der Wassertemperatur eines Flusses an einer bestimmten Stelle führen auf eine stetige Zufallsvariable, deren Werte sich in einem bestimmten reellen Intervall bewegen.*

Verteilungsfunktion einer Zufallsvariablen

Um mit Zufallsvariablen arbeiten zu können, benötigt man deren so genannte *Wahrscheinlichkeitsverteilung*:

D.h. man muss die Wahrscheinlichkeit P dafür, dass die Zufallsvariable X einen bestimmten Wert annimmt (diskrete Variable) bzw. in einem bestimmten Intervall liegt (stetige Variable), ermitteln können.

Definition

Die Verteilungsfunktion $F(x)$ einer Zufallsvariablen X gibt die Wahrscheinlichkeit dafür an, dass X einen Wert annimmt, der kleiner oder gleich einer vorgegebenen Zahl x ist, d.h. es gilt

$$F(x) := P(X \leq x).$$

Diskrete Verteilungen

Nimmt die Zufallsvariable X nur diskrete Werte x_1, x_2, \dots mit den Wahrscheinlichkeiten

$$P(X = x_i) = p_i, \quad i = 1, 2, \dots$$

an, dann nennt man die diskrete Funktion

$$f(x) = \begin{cases} p_i & \text{für } x = x_i \\ 0 & \text{für } x \neq x_i \end{cases} \quad i = 1, 2, \dots$$

Wahrscheinlichkeitsfunktion der Zufallsvariablen X .

Für diese gilt offensichtlich $f(x) \geq 0$ und

$$\sum_i f(x_i) = 1.$$

Die zugehörige Verteilungsfunktion ergibt sich zu

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i),$$

wobei die Summation jeweils über alle x_i zu erstrecken ist, die die Ungleichung $x_i \leq x$ erfüllen.

Beispiel

Beim Werfen eines Würfels haben wir 6 Elementarereignisse ω_i (Augenzahl i). Die zugehörige Zufallsvariable X kann somit die Werte

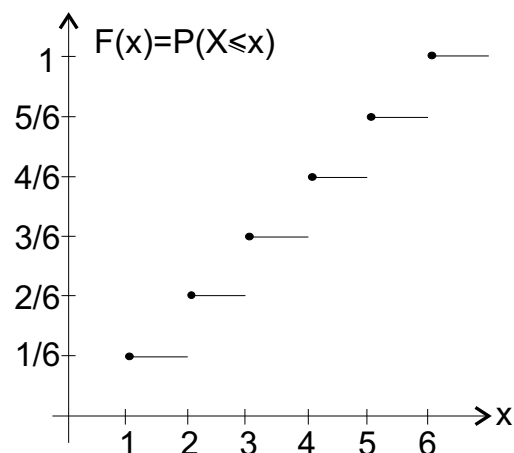
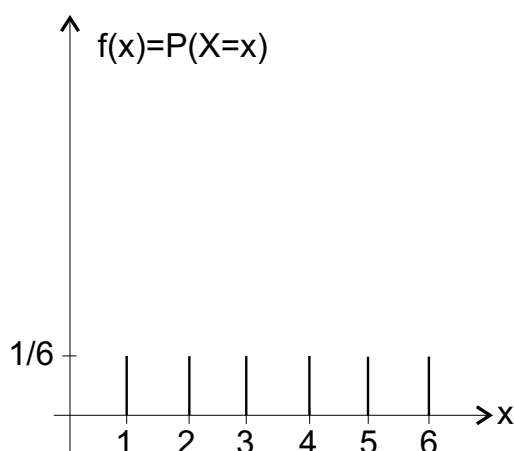
$$x_i = i, \quad i = 1, \dots, 6,$$

annehmen. Damit gilt

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

also $p_i = \frac{1}{6}$ für alle $i = 1, \dots, 6$.

Graphische Darstellung für die Wahrscheinlichkeitsfunktion $f(x)$ und die Verteilungsfunktion $F(x)$:



Stetige Verteilungen

In diesem Fall kann die Zufallsvariable X in einem Intervall I beliebig viele Werte annehmen.

Die Summation bzgl. der Wahrscheinlichkeitsfunktion, die im diskreten Fall die Definition der Verteilungsfunktion ermöglicht, muss jetzt durch eine Integration über die so genannte **Dichtefunktion** $f(x)$ ersetzt werden:

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt.$$

Beispiel

Eine stetige Verteilung sei durch die folgende Dichte festgelegt:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{für } a \leq x \leq b, \\ 0 & \text{sonst.} \end{cases}$$

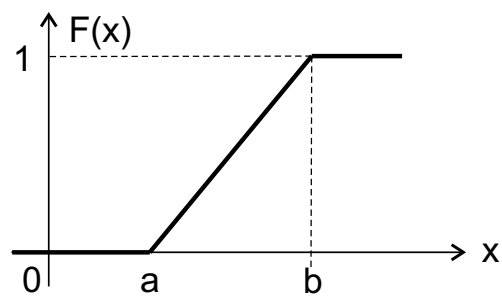
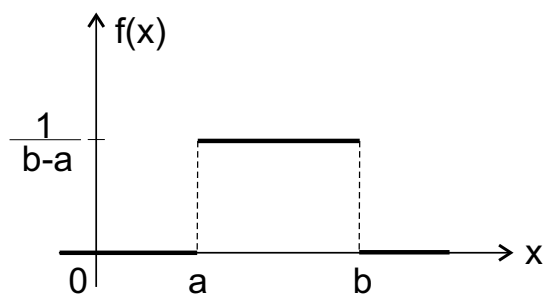
Dies ist eine so genannte Gleichverteilung über dem Intervall $[a, b]$ (synonym: Rechteckverteilung).

Für $x \in [a, b]$ gilt:

$$\int_{-\infty}^x f(t) dt = \int_a^x \frac{dt}{b-a} = \frac{t}{b-a} \Big|_{t=a}^x = \frac{x-a}{b-a}.$$

Daraus ergibt sich die Verteilungsfunktion zu:

$$F(x) = \begin{cases} 0 & \text{für } x < a, \\ \frac{x-a}{b-a} & \text{für } a \leq x \leq b, \\ 1 & \text{für } x > b. \end{cases}$$



Eigenschaften einer Verteilungsfunktionen

Wegen $F(x) = P(X \leq x)$ ergibt sich:

Definition

Für jede Verteilungsfunktion $F(x)$ gelten die folgenden Aussagen:

- $F(x)$ ist monoton wachsend,
- $0 \leq F(x) \leq 1$,
- $\lim_{x \rightarrow -\infty} F(x) = 0$
(unmögliches Ereignis),
- $\lim_{x \rightarrow \infty} F(x) = 1$
(sicheres Ereignis),
- $P(a < X \leq b) = F(b) - F(a)$.

Speziell für stetiges $F(x)$ mit Dichte $f(x)$ gilt:

$$F(b) - F(a) = \int_a^b f(x) dx.$$

Weitere Eigenschaften

Als Dichtefunktionen geeignet sind nur Funktionen, welche die Forderungen

$$f(x) \geq 0 \quad \text{und} \quad \int_{-\infty}^{\infty} f(x) dx = 1$$

erfüllen: Da F monoton steigend ist, muss nämlich $F'(x) = f(x) \geq 0$ gelten. Die Fläche unter f stellt die Gesamtwahrscheinlichkeit dar, so dass das Integral den Wert 1 ergeben muss.

Man beachte ferner, dass ein Ereignis mit der Wahrscheinlichkeit Null nicht unmöglich ist, sondern nur sehr wenig wahrscheinlich;

Andererseits ist ein Ereignis mit der Wahrscheinlichkeit Eins nicht sicher, aber sehr wahrscheinlich. So gilt beispielsweise für das nicht sicher eintretende Ereignis $X \neq x_0$:

$$P(X \neq x_0) = \int_{\mathbb{R} \setminus \{x_0\}} f(x) dx = 1.$$

Erwartungswert und Varianz

Sei X eine Zufallsvariable mit Wahrscheinlichkeits- bzw. Dichtefunktion $f(x)$. Dann sind Erwartungswert μ_X (auch $E[X]$ genannt) und Varianz σ_X^2 (auch $Var[X]$ genannt) von X

a) für diskretes X gegeben durch:

- $\mu_X = E[X] := \sum_i x_i f(x_i),$
- $\sigma_X^2 = Var[X] := \sum_i (x_i - \mu_X)^2 f(x_i);$

b) für stetiges X gegeben durch:

- $\mu_X = E[X] := \int_{-\infty}^{\infty} x f(x) dx,$
- $\sigma_X^2 = Var[X] := \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x) dx.$

Falls eine Summe bzw. ein Integral divergiert, dann ist die betroffene Größe nicht definiert. Als Standardabweichung von X bezeichnet man den Ausdruck $\sigma_X := \sqrt{Var[X]}$.

Beispiel

- a) Beim Werfen eines Würfels ergibt sich der Erwartungswert von X (wegen der Gleichwahrscheinlichkeit gilt $f(x_i) = 1/6$ für alle $i = 1, \dots, 6$) zu:

$$\mu_X = E[X] = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = 3.5.$$

Die Varianz errechnet sich folgendermaßen:

$$\begin{aligned}\sigma_X^2 = \text{Var}[X] &= \frac{1}{6} \left[(1 - 3.5)^2 + (2 - 3.5)^2 + (3 - 3.5)^2 + \right. \\ &\quad \left. + (4 - 3.5)^2 + (5 - 3.5)^2 + (6 - 3.5)^2 \right] \\ &= \frac{1}{6} \left[2.5^2 + 1.5^2 + 0.5^2 + 0.5^2 + 1.5^2 + 2.5^2 \right] \\ &= 17.5/6 = 2.91\bar{6}.\end{aligned}$$

Für die Standardabweichung gilt $\sigma_X = \sqrt{2.91\bar{6}} \approx 1.7078$.

Beispiel – Fortsetzung

- b) Der Erwartungswert der gleichverteilten (stetigen) Zufallsvariablen X mit Dichte

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{für } a \leq x \leq b, \\ 0 & \text{sonst.} \end{cases}$$

ergibt sich zu:

$$\begin{aligned} E[X] &= \frac{1}{b-a} \int_a^b x \, dx \\ &= \frac{1}{b-a} \cdot \frac{x^2}{2} \Big|_{x=a}^b \\ &= \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}. \end{aligned}$$

Ihre Varianz

$$Var[X] = \frac{1}{b-a} \int_a^b \left(x - \frac{a+b}{2} \right)^2 dx$$

lässt sich einfacher mittels der Verschiebungsfelme, die noch vorgestellt wird, ausrechnen.

Transformations- und Verschiebungsformel

Sei X eine Zufallsvariable und $a, b \in \mathbb{R}$, dann gelten die Transformationsformeln:

- $E[aX + b] = aE[X] + b$,
- $Var[aX + b] = a^2 \cdot Var[X]$.

Für die Varianz gilt auch die Verschiebungsformel:

$$Var[X] = E[X^2] - (E[X])^2.$$

Beispiel

Die Varianz der gleichverteilten Zufallsvariablen X mit Dichte

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{für } a \leq x \leq b, \\ 0 & \text{sonst.} \end{cases}$$

und dem Erwartungswert

$$E[X] = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}$$

lässt sich nun mittels Verschiebungsformel berechnen:

$$\begin{aligned} \text{Var}[X] &= \frac{1}{b-a} \int_a^b x^2 dx - \left(\frac{a+b}{2}\right)^2 \\ &= \frac{1}{b-a} \cdot \frac{x^3}{3} \Big|_{x=a}^b - \frac{1}{4}(a+b)^2 \\ &= \frac{1}{3} \cdot \frac{b^3 - a^3}{b-a} - \frac{1}{4}(a+b)^2 \\ &= \frac{1}{3}(b^2 + ab + a^2) - \frac{1}{4}(b^2 + 2ab + a^2) \\ &= \frac{(b-a)^2}{12}. \end{aligned}$$

Übung

Berechnen Sie die Varianz beim Werfen eines Würfels, d.h. für die zugehörige Zufallsvariable X , die die Werte

$$x_i = i, \quad i = 1, \dots, 6,$$

*annehmen kann mittels Verschiebungsformel.
(Es gilt: $P(X = i) = \frac{1}{6}$ für alle $i = 1, \dots, 6$.)*

Lösung

Aus der Verschiebungsformel

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

folgt sofort:

$$\begin{aligned}\text{Var}[X] &= \frac{1}{6}[1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2] - 3.5^2 \\ &= 2.9\bar{16}.\end{aligned}$$

Man beachte, dass es in vielen Fällen einfacher ist, die Varianz nicht mit der sie definierenden Formel, sondern durch die Verschiebungsregel zu berechnen.

Spezielle Verteilungen

Beim Bernoulli-Experiment haben wir die Wahrscheinlichkeit $P_k(n; p)$ dafür berechnet, dass genau k der *unabhängigen* Ereignisse A_1, \dots, A_n eintreten:

$$P_k(n; p) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

Falls eine Zufallsgröße X nach dieser Formel verteilt ist, so spricht man von Binomialverteilung (oder binomischer Verteilung):

Definition

Eine diskrete Zufallsvariable X heißt binomialverteilt mit den Parametern n und p , kurz

$$X \sim B(n, p),$$

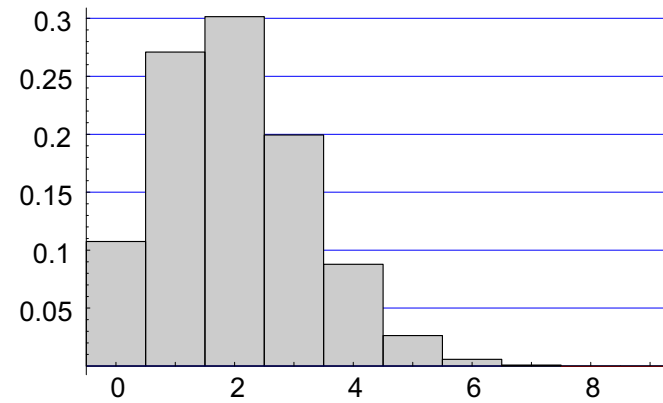
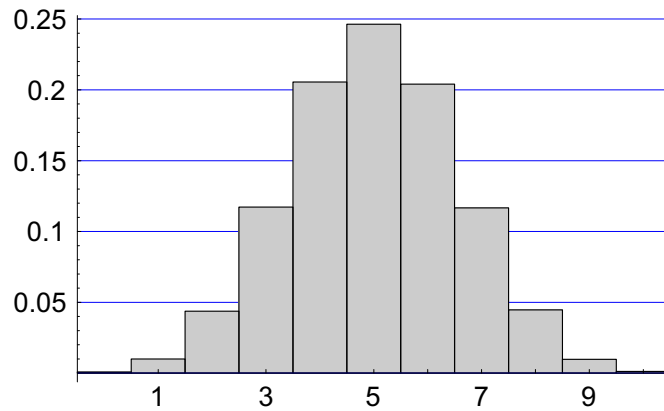
wenn ihre Wahrscheinlichkeitsfunktion gemäß

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

für $k = 0, 1, \dots, n$ gegeben ist.

Binomialverteilung

Die Abbildung zeigt die Wahrscheinlichkeitsfunktion für die Parameter $n = 10$ und $p = 0.5$ bzw. $p = 0.2$:



Für eine binomialverteilte Zufallsvariable X gilt:

$$E[X] = np \quad \text{und} \quad \text{Var}[X] = np(1 - p).$$

Poisson-Verteilung

Eine weitere wichtige Verteilung leitet sich als Grenzfall aus der Binomialverteilung ab. Die Zahl n der Wiederholungen wird sehr groß ($n \rightarrow \infty$), während die Wahrscheinlichkeit p für den einzelnen Ereigniseintritt sehr klein wird ($p \rightarrow 0$). Gleichzeitig fordert man aber ein konstantes Produkt $n \cdot p = \lambda$:

Definition

Eine diskrete Zufallsvariable X , die die Werte $k = 0, 1, 2, \dots$ annehmen kann, heißt poisson-verteilt mit Parameter $\lambda > 0$, kurz

$$X \sim \Pi(\lambda),$$

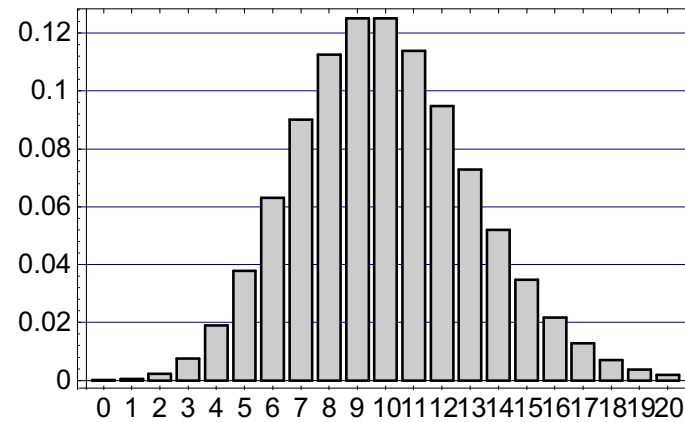
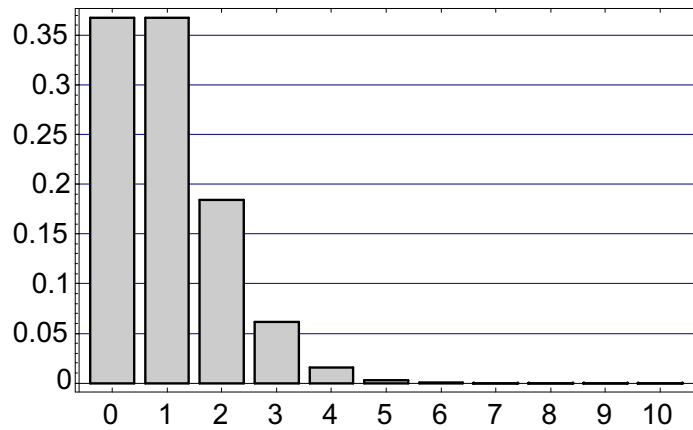
wenn ihre Wahrscheinlichkeitsfunktion gemäß

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

für $k = 0, 1, 2, \dots$ gegeben ist.

Poisson-Verteilung

Die Abbildung zeigt für die Parameterwerte $\lambda = 1$ bzw. $\lambda = 10$ die Wahrscheinlichkeitsfunktion der Poissonverteilung.



Poisson-Verteilung in der Praxis

In der Praxis hat man es bei Zufallsexperimenten häufig mit Ereignissen zu tun, die nur *sehr selten*, d.h. mit *geringer Wahrscheinlichkeit* auftreten. Solche Ereignisse genügen meist der Poisson-Verteilung. Früher war deren Anwendungsbereich auf recht ausgefallene Ereignisse beschränkt, wie z.B.

- auf Kinderselbstmorde oder
- auf durch Huftritt verursachte Todesfälle in der preußischen Armee.

Heutzutage spielt diese Verteilung jedoch eine wichtige Rolle z.B.

- im Fernsprechverkehr,
- in der statistischen Qualitätskontrolle,
- beim Zerfall von radioaktiven Substanzen,
- in der Biologie, der Meteorologie und vor allem in der Warteschlangentheorie.

Kenngrößen der Poisson-Verteilung

Für eine poisson-verteile Zufallsvariable X gilt:

$$E[X] = \lambda \quad \text{und} \quad Var[X] = \lambda.$$

Man benutzt die Poisson-Verteilung in der Praxis häufig zur Annäherung der Binomialverteilung, da diese für große n ($n \geq 100$) sehr unhandlich wird.

Beispiel

Im Bernoulli-Experiment hatten wir die Wahrscheinlichkeit für das Auftreten von genau 3 defekten Bauteilen zu ≈ 0.1823 berechnet.

Die Defekt-Wahrscheinlichkeit für ein Bauteil lag bei 0.02.

Setzen wir jetzt $p = 0.02$, so ergibt sich der Erwartungswert von X (X sei die Defektanzahl) bei 100 Bauteilen zu $np = 100 \cdot 0.02 = 2$.

Die — wesentlich einfacher zu berechnende — Approximation mit der Poisson-Verteilung mit

$$\lambda = n \cdot p = 2$$

lautet daher

$$P(X = 3) \approx \frac{2^3}{3!} e^{-2} \approx 0.1804.$$

Normalverteilung

Definition

Eine stetige Zufallsvariable X heißt normalverteilt mit Parametern $\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}^+$, kurz

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

falls sie die Dichtefunktion

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

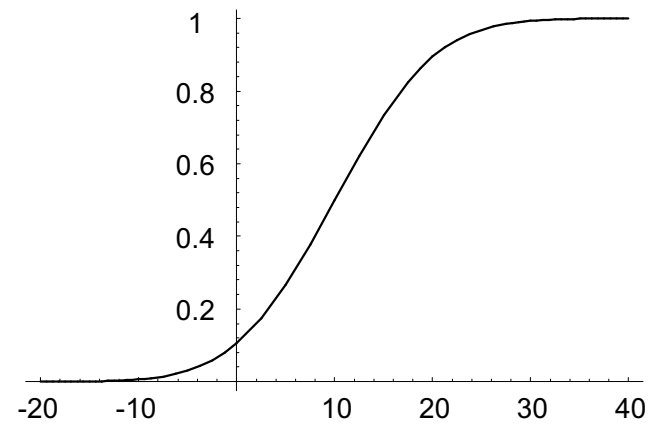
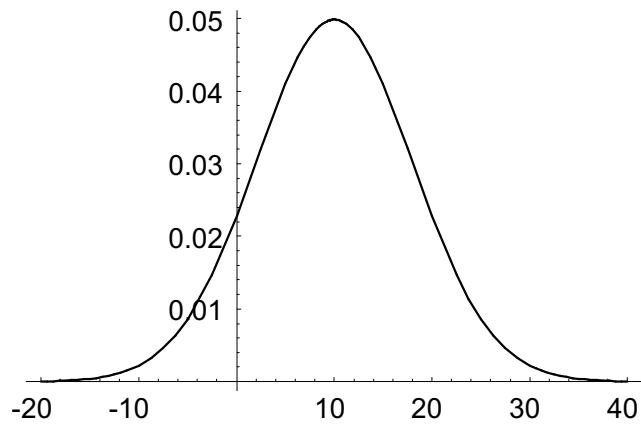
besitzt. Der Graph von φ heißt Gauß'sche Glockenkurve. Die Funktion

$$\Phi(x, \mu, \sigma) := \int_{-\infty}^x \varphi(t, \mu, \sigma) dt$$

wird als Gauß'sches Fehlerintegral bezeichnet.

Normalverteilung

Die Abbildung zeigt Dichte- und Verteilungsfunktion der Normalverteilung für die Parameter $\mu = 10$ und $\sigma^2 = 64$.



Kenngrößen der Normalverteilung

Die zentrale Rolle der Normalverteilung beruht darauf, dass viele in der Praxis auftretende Zufallsvariablen normalverteilt sind und andere (diskrete wie stetige) Verteilungen in der Praxis für große n oft durch Normalverteilungen angenähert werden können.

Außerdem bildet die Normalverteilung die Grundlage vieler Schätz- und Testverfahren. Die Bedeutung der Parameter μ und σ sei ohne Beweis aufgeführt:

Für eine normal-verteilte Zufallsvariable X gilt:

$$E[X] = \mu \quad \text{und} \quad Var[X] = \sigma^2.$$

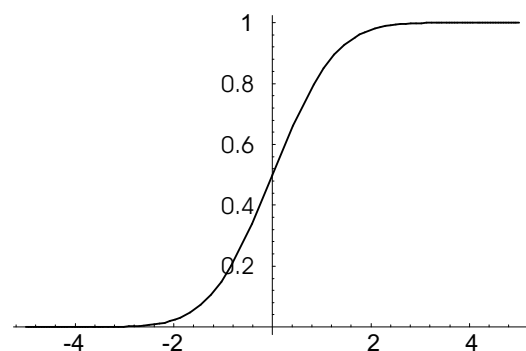
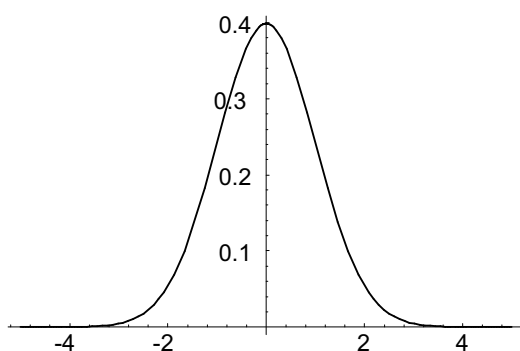
Standardnormalverteilung

Da die Werte der Normalverteilung nur mittels numerischer Integrationsverfahren ermittelt werden können, benutzt man in der Praxis die so genannte *Standardnormalverteilung*, deren Werte in fast allen Formelsammlungen tabelliert sind:

Definition

Für $\mu = 0$ und $\sigma = 1$ erhält man die Standardnormalverteilung:

$$\Phi(x) := \Phi(x, 0, 1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt.$$



Eigenschaften der Normalverteilung

Aufgrund der Achsensymmetrie der Dichtefunktion ergibt sich die wichtige Gleichung

$$\Phi(-x) = 1 - \Phi(x).$$

Da sich wegen dieser Symmetriegleichung alle Werte von $\Phi(-x)$ für negatives $-x$ stets auf die Auswertung von $\Phi(x)$ mit positivem Argument x zurückführen lassen, ist das Gauß'sche Fehlerintegral nur für positive Argumente tabelliert!

Hat man eine $N(\mu, \sigma)$ -verteilte Zufallsvariable X , so kann man diese durch die Transformation

$$Y = \frac{X - \mu}{\sigma}$$

immer zu einer $N(0, 1)$ -verteilten Zufallsvariablen Y normieren. Offensichtlich gilt dann

$$P(a < X \leq b) = P\left(\frac{a - \mu}{\sigma} < Y \leq \frac{b - \mu}{\sigma}\right).$$

Nun lässt sich die gesuchte Wahrscheinlichkeit mit Werten der Standardnormalverteilung berechnen:

$$P(a < X \leq b) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right).$$

Beispiel

Eine Zufallsgröße X sei normalverteilt mit den Parametern $\mu = 4$ und $\sigma = 2$.

Gesucht ist die Wahrscheinlichkeit dafür, dass X zwischen 2.5 und 8 liegt:

$$\begin{aligned} P(2.5 \leq X \leq 8) &= \Phi\left(\frac{8-4}{2}\right) - \Phi\left(\frac{2.5-4}{2}\right) \\ &= \Phi(2) - \Phi(-0.75) \\ &= \Phi(2) - (1 - \Phi(0.75)) \\ &= 0.9773 - 1 + 0.7734 \\ &= 0.7507. \end{aligned}$$

Bei der Berechnung haben wir die Symmetriegleichung ausgenutzt und die Werte

$$\Phi(2) = 0.9773, \quad \Phi(0.75) = 0.7734$$

einer Tabelle für die Standardnormalverteilung entnommen.

Eine sonderbare Ziffern-Verteilung und die Steuerrevision

Wettmöglichkeit:

Wenn die erste Zahl, die in der ARD-Tagesschau ab 20 Uhr genannt wird, mit einer Ziffer zwischen 1 und 3 beginnt, verlieren Sie 50 Euro. Liegt die erste Ziffer dagegen im Bereich 4 bis 9, dann gewinnen Sie 50 Euro.

Würden Sie diese Wette für die nächsten 365 Tage annehmen?

Vor einer Entscheidung wird man den erwarteten Gewinn berechnen: Unter der Annahme, dass alle Ziffern von 1 bis 9 als Anfangsziffer der (vom Nachrichtensprecher zufällig genannten) Zahl gleichwahrscheinlich sind, beträgt das Verlustrisiko $3/9 = 1/3$. Dem steht eine Gewinnchance von $6/9 = 2/3$ gegenüber. Damit berechnet sich der erwartete Gewinn zu

$$\mu = -50 \cdot \frac{1}{3} + 50 \cdot \frac{2}{3} = 16.67 \text{ Euro.}$$

Gutes Geschäft???

Schon nach ein paar Monaten werden Sie feststellen, dass Ihnen diese Wette erhebliche finanzielle Einbußen beschert. Warum eigentlich?

Weil der Erwartungswert *nicht* bei +16.67 Euro liegt, sondern in Wirklichkeit –10.21 Euro beträgt. Im Mittel verlieren Sie also mehr als 10 Euro pro Tag!! Wie lässt sich das erklären?

Die erste Ziffer bestimmter Zahlenmengen, besonders von „dimensions-behafteten“, wie z.B. Flächen von Gewässern, Stromverbrauchsdaten von Privathaushalten, Aktienkursen oder Naturkonstanten ist nämlich *nicht* gleichverteilt. Hier gilt ein Gesetz, das Ende des 19. Jahrhunderts auf kuriose Weise entdeckt wurde. Der Astronom *Simon Newcomb* vermutete, dass sich die Häufigkeit der bei einer Zahl auftauchenden führenden Ziffer durch das Gesetz

$$\text{Ziffernhäufigkeit} = \log_{10} \left(1 + \frac{1}{\text{jeweilige Ziffer}} \right)$$

beschreiben lässt.

Benford'sches Gesetz

Die Entdeckung von Newcomb geriet wieder in Vergessenheit.

Erst 1938 machte der Physiker *Frank Benford* unabhängig von Newcomb dieselbe Beobachtung und verifizierte sie anhand zahlreicher Statistiken aus ganz unterschiedlichen Lebensbereichen:

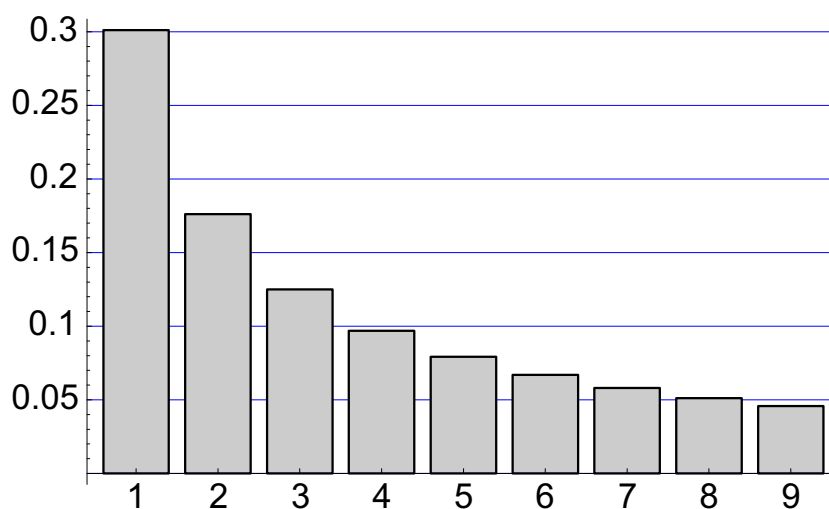
- physikalische Tabellen (spezifische Wärme, Atom- und Molekulargewichte),
- Hausnummern zufällig ausgewählter Personen,
- Einwohnerzahlen von US-Counties,
- Baseballergebnisse etc.

Die Verteilung wurde nach ihm benannt, sie ist heute unter dem Namen *Benford'sches Gesetz* bekannt.

Benford'sches Gesetz

Das Histogramm zeigt die Häufigkeitsverteilung für die erste Ziffer einer Zahl:

$$h(\text{1.Ziffer der Zahl} = i) = \log_{10}(1 + 1/i)$$



Die niedrigeren Ziffern sind also wesentlich wahrscheinlicher als die hohen. So liegt die Wahrscheinlichkeit dafür, dass die erste Ziffer 1, 2 oder 3 ist, bei

$$0.30103 + 0.176091 + 0.124939 = 0.60206.$$

Damit ist auch der ungünstige Ausgang unserer Wette klar: Deren Erwartungswert ergibt sich zu

$$\begin{aligned}\mu &= (-50) \cdot 0.60206 + 50 \cdot (1 - 0.60206) \\ &= -10.206.\end{aligned}$$

Benford'sches Gesetz

Der Mathematiker *Roger Pinkham* konnte 1961 zeigen, dass das Benford'sche Gesetz die *einzigste Verteilung* ist, die *invariant* gegenüber positiven Skalierungen ist.

Es ist also egal, ob man beispielsweise

- Flächen in Quadratkilometern oder in Quadratmetern misst,
- Gewichte in Tonnen oder Kilogramm angibt,
- bzw. monetäre Größen in DM oder Euro notiert.

Natürlich genügen nicht alle Statistiken dem Benford'schen Gesetz, es gibt ja auch binomial-verteilte, normalverteilte Zufallsvariablen, usw.

Geeignet sind aber Zahlen, die Größenordnungen repräsentieren, nicht der Identifikation dienen (wie z.B. Pass- oder Telefonnummern) und keine inhärenten Grenzen aufweisen.

Benford'sches Gesetz und Fibonacci-Zahlen

Die Folge $(f_n)_{n \in \mathbb{N}}$ der bekannten *Fibonacci-Zahlen*, deren Bildungsgesetz

$$f_0 = f_1 = 1, f_{n+1} = f_n + f_{n-1}$$

für $n \geq 1$ lautet, lässt sich als natürlicher Wachstumsprozess interpretieren (z.B. Größe von Kaninchenpopulationen).

Den relativen Häufigkeiten gemäß Benford'schem Gesetz sind in nachfolgender Tabelle die Häufigkeiten der Anfangsziffern für die ersten 100 bzw. 1000 Fibonacci-Zahlen gegenübergestellt:

Ziffer	Benford	100 Fib.-Zahlen	1000 Fib.Zahlen
1	0.3010300	0.30	0.301
2	0.1760910	0.18	0.177
3	0.1249390	0.13	0.125
4	0.0969100	0.09	0.096
5	0.0791812	0.08	0.080
6	0.0669468	0.06	0.067
7	0.0579919	0.05	0.056
8	0.0511525	0.07	0.053
9	0.0457575	0.04	0.045

Ziffernverteilung und Steuerrevision

Um die Jahrtausendwende hat man nun das Benford'sche Gesetz erstmals benutzt, um Steuerbetrug und Bilanzfälschungen aufzudecken: Echte Zahlen aus diesen Bereichen gehorchen nämlich der Benford-Verteilung, gefälschte Daten weichen davon ab.

Bei gefälschten Daten kommen bestimmte Ziffern (z.B. Lieblingszahlen des Fälschers) viel häufiger vor als bei korrekten Angaben.

So hat der Mathematiker *Mark Nigrini* umfangreiche Tests durchgeführt (z.B. fast 170 000 Steuererklärungen ausgewertet), die das Gesetz bestätigt haben. Nigrini entwickelte deshalb nach dem Benford'schen Gesetz ein Programm namens „Digital Analyser“, das bereits viele Bilanzfälscher und Steuerehinterzieher aufspüren konnte.

Anfang 2004 sind auch erste Untersuchungen mit deutschen Steuererklärungen durchgeführt worden.

HIV-Tests

Stellen Sie sich vor, Ihre Ärztin hat Ihnen gerade Ihr positives Resultat bei einem Aidsstest mitgeteilt. Haben Sie dann Aids?

Paradoxerweise ist es sogar wahrscheinlicher, dass Sie *nicht* an Aids erkrankt sind — trotz positivem Test! Dies werden wir im Folgenden mit Hilfe des Satzes von Bayes exakt nachrechnen.

Bekanntlich sind HIV-Tests Verfahren, die eine stattgefundenene Infektion mit einem menschlichen Immunschwächevirus (HIV) nachweisen können.

In Deutschland werden als Suchtests meist ELISA-Tests verwandt, die Antikörper gegen bestimmte Varianten des Virus nachweisen.

Testeigenschaften

Sie weisen diese Antikörper sehr „empfindlich“ nach, was besagt, dass möglichst viele Infektionen auch erkannt werden („hohe Sensitivität“ des Testverfahrens).

Ihr Nachteil ist aber, dass sie manchmal reagieren, obwohl keine Infektion besteht: Es liegen also in einigen Fällen falsche positive Testergebnisse vor („geringe Spezifität“ des Tests).

Grob gesagt kann man deshalb *bei negativem Testergebnis* ziemlich sicher sein, dass man *nicht* mit HIV infiziert ist, während bei positivem Ergebnis *immer* mit einem zweiten Bestätigungstest das Ergebnis des ersten überprüft werden muss.

Als Bestätigungstest wird in Deutschland meist ein Immunoblot genanntes Verfahren eingesetzt. Erst wenn mit dem Bestätigungstest Antikörper nachgewiesen werden, gilt die Diagnose „HIV positiv“ als sicher und sollte dem Patienten mitgeteilt werden.

HIV-Wahrscheinlichkeiten

Ende 2003 betrug der Anteil HIV-Positiver bei der erwachsenen Bevölkerung in Westeuropa etwa

0.3%.

Mathematisch gesprochen kann man dem Ereignis „HIV positiv“ also die Wahrscheinlichkeit zuordnen:

$$P(\text{HIV}+) = 0.003.$$

Gehen wir davon aus, dass ein Testverfahren in

99.9%

aller Fälle eine bestehende HIV-Infektion auch nachweist, dass also die bedingte Wahrscheinlichkeit „positives Testergebnis bei vorliegender HIV-Infektion“

$$P(\text{Test}+|\text{HIV}+) = 0.999$$

ist. Die Rate der wahren positiven Testergebnisse („Sensitivität“ des Testverfahrens) beträgt damit 99.9%.

HIV-Wahrscheinlichkeiten

Nun gibt es auch eine kleine Anzahl von Fällen, in denen falsche positive Testergebnisse ausgegeben werden, in denen also eine Person, die HIV negativ ist, fälschlicherweise ein positives Testergebnis erhält.

Nehmen wir an, dies geschieht in 2 % aller Fälle. Dann ist die bedingte Wahrscheinlichkeit eines positiven Testergebnisses, falls keine HIV-Infektion vorliegt, gleich

$$P(\text{Test+}|\text{HIV-}) = 0.02.$$

Wegen $P(\text{Test+}|\text{HIV+}) = 0.999$ gilt entsprechend

$$\begin{aligned} P(\text{Test-}|\text{HIV+}) &= 1 - P(\text{Test+}|\text{HIV+}) \\ &= 0.001 \end{aligned}$$

und wegen $P(\text{Test+}|\text{HIV-}) = 0.02$ ist

$$P(\text{Test-}|\text{HIV-}) = 1 - 0.02 = 0.98.$$

HIV-Wahrscheinlichkeiten

Mit Hilfe des Satzes von Bayes ergeben sich die folgenden Wahrscheinlichkeiten:

	HIV+	HIV-	
Test+	$P(\text{Test+} \cap \text{HIV+})$	$P(\text{Test+} \cap \text{HIV-})$	$P(\text{Test+})$ = 0.022937
	= $P(\text{Test+} \text{HIV+}) \cdot P(\text{HIV+})$	= $P(\text{Test+} \text{HIV-}) \cdot P(\text{HIV-})$	
	= $0.999 \cdot 0.003$	= $0.02 \cdot 0.997$	
	= 0.002997	= 0.019940	
Test-	$P(\text{Test-} \cap \text{HIV+})$	$P(\text{Test-} \cap \text{HIV-})$	$P(\text{Test-})$ = 0.977063
	= $P(\text{Test-} \text{HIV+}) \cdot P(\text{HIV+})$	= $P(\text{Test-} \text{HIV-}) \cdot P(\text{HIV-})$	
	= $0.001 \cdot 0.003$	= $0.98 \cdot 0.997$	
	= 0.000003	= 0.977060	
	$P(\text{HIV+}) = 0.003$	$P(\text{HIV-}) = 0.997$	1

HIV-Wahrscheinlichkeiten

Der Tabelle entnimmt man, dass für die Wahrscheinlichkeit, trotz positivem Testergebnis *nicht* an HIV erkrankt zu sein, gilt:

$$\begin{aligned} P(\text{HIV-}|\text{Test+}) &= \frac{P(\text{Test+} \cap \text{HIV-})}{P(\text{Test+})} \\ &= \frac{0.019940}{0.022937} = 0.8693378. \end{aligned}$$

Dies bedeutet, dass man selbst bei positivem Testergebnis in 87% aller Fälle gesund und nur in 13% wirklich HIV-positiv ist!

Und selbst wenn der Test wirklich zuverlässig *alle* HIV-Infizierten erkennt, wenn also

$$P(\text{Test+}|\text{HIV+}) = 1$$

ist, ändern sich obige Werte nur geringfügig.

HIV-Wahrscheinlichkeiten

Wir wollen nun versuchen, dieses auf den ersten Blick sehr überraschende Ergebnis zu verstehen:

Ausgangspunkt war $P(\text{HIV}+) = 0.003$, was bedeutet, dass von 1000 Personen durchschnittlich 3 mit HIV infiziert sind.

Wenn wir nun aber eine Population von 1 000 000 Menschen zugrunde legen, so sind 3 000 Infizierte darunter. Davon werden korrekt 2 997 Infizierte durch den Test entdeckt, 3 Personen werden fälschlicherweise nicht identifiziert.

Von den 997 000 Gesunden werden allerdings ganze 19 940 Personen fälschlicherweise als positiv getestet. Gehört man nun zu den 22 937 Personen mit positivem Testergebnis, so ist es viel wahrscheinlicher zur Gruppe der 19 940 Gesunden als zur Gruppe der 2 997 Kranken zu gehören.

HIV-Wahrscheinlichkeiten

Liefert also ein positiver Aidstest überhaupt keine Information?

Doch: Mit einem positiven Testergebnis ist die Wahrscheinlichkeit, an HIV erkrankt zu sein, im obigen Beispiel von 3 zu 1000 auf ca. 13 zu 100 angewachsen.

Die Chancen stehen schlechter. Aber dennoch besagt ein positiver Test keinesfalls, wirklich HIV-positiv zu sein.

Zugehörigkeit zu einer Risikogruppe

Als letztes wollen wir noch untersuchen, welchen Einfluss die Zugehörigkeit zu einer Risikogruppe auf die obigen Wahrscheinlichkeiten hat. Wir setzen wieder

$$P(\text{Test+}|\text{HIV+}) = 0.999 \quad \text{richtige Positive,}$$

$$P(\text{Test+}|\text{HIV-}) = 0.02 \quad \text{falsche Positive.}$$

Für die Wahrscheinlichkeit, HIV-positiv zu sein, setzen wir $P(\text{HIV+}) = x$, wobei x von der Verbreitung von AIDS in der jeweiligen Risikogruppe abhängt. Wir erhalten:

	HIV+	HIV-	
Test+	$0.999 \cdot x$	$0.02 \cdot (1 - x)$	$0.979 \cdot x + 0.02$
Test-	$0.001 \cdot x$	$0.98 \cdot (1 - x)$	$-0.979 \cdot x + 0.98$
	x	$1 - x$	1

Zugehörigkeit zu einer Risikogruppe

Die Wahrscheinlichkeit, HIV-negativ zu sein trotz positivem Testergebnis, ist dann

$$\begin{aligned} P(\text{HIV-}|\text{Test+}) &= \frac{P(\text{HIV-} \cap \text{Test+})}{P(\text{Test+})} \\ &= \frac{0.02 \cdot (1 - x)}{0.979 \cdot x + 0.02} \end{aligned}$$

Für verschiedene x -Werte erhält man also verschiedene Wahrscheinlichkeiten:

	x	$P(\text{HIV-} \text{Test+})$
geringes Risiko	→ 0.0003	0.9852
normales Risiko	→ 0.003	0.8693
hohes Risiko	→ 0.03	0.3930
sehr hohes Risiko	→ 0.3	0.0446

HIV-Wahrscheinlichkeiten

In einer Nicht-Risikogruppe mit geringem HIV-Risiko (nur 0.03% dieser Bevölkerungsgruppe sind HIV-infiziert) ist man bei Vorliegen eines positiven Testergebnisses in 99% der Fälle *nicht* infiziert.

Gehört man hingegen zu einer Risikogruppe (3% dieser Gruppe ist infiziert), so ist man nur in 39% nicht HIV-positiv.

In einer Gruppe höchsten Risikos mit 30% Infizierten ist die Wahrscheinlichkeit, bei positivem HIV-Test *nicht* infiziert zu sein, nur noch 4%. (Hier ist die Wahrscheinlichkeit, HIV-infiziert zu sein, aber auch ohne Test schon sehr hoch.)

Als überraschende Aussage bleibt: Wenn man einer Nicht-Risikogruppe angehört, ist es sogar äußerst unwahrscheinlich, bei positivem AIDS-Test wirklich infiziert zu sein!