

Statistik I für WInf und WI

Prof. Dr. Wilhelm Stannat

Inhalt:

I Deskriptive Statistik

1. Grundbegriffe
2. Auswertung eindimensionaler Datensätze
3. Auswertung zwei- und mehrdimensionaler Messreihen

II Wahrscheinlichkeitstheorie

1. Zufallsexperimente und Wahrscheinlichkeitsräume
2. Zufallsvariablen und Verteilungen
3. Erwartungswert und Varianz
4. Stetige Verteilungen
5. Grenzwertsätze

III Induktive Statistik

1. Schätzen
2. Testen

Das vorliegende Skript ist eine Zusammenfassung des zweiten Teils der Vorlesung Statistik I für WInf und WI, die im Wintersemester 2007/08 an der TU Darmstadt gehalten wurde. Die Lektüre des Skriptes ist kein gleichwertiger Ersatz für den Besuch der Vorlesung.

Korrekturen bitte per Email an: stannat@mathematik.tu-darmstadt.de

Teil II Wahrscheinlichkeitsrechnung

1. Zufallsexperimente und Wahrscheinlichkeitsräume

Unter einem **Zufallsexperiment** versteht man zunächst einmal einen zeitlich wie örtlich fest umrissenen Vorgang mit unbestimmtem Ausgang.

Beispiele

- Werfen eines Würfels oder Werfen einer Münze
- Wahlergebnis der nächsten Landtagswahl
- Temperatur oder Windgeschwindigkeit am Luisenplatz am 1. Dezember 2007, 12:00
- Körpergröße oder Kopfumfang eines Neugeborenen

Die Gesamtheit aller möglichen Ausgänge eines Zufallsexperiments heißt **Ergebnismenge** oder auch **Stichprobenraum** und wird mit Ω bezeichnet.

Ein Element $\omega \in \Omega$ heißt **Elementarereignis** oder **Stichprobe**. Es stellt einen möglichen Ausgang des zugrundeliegenden Zufallsexperiments dar.

Beispiele

- (i) einmaliges Würfeln: $\Omega = \{1, 2, \dots, 6\}$, $|\Omega| = 6$
 (Hierbei bezeichnet $|\Omega|$ die **Mächtigkeit der Menge** Ω , also die Anzahl der Elemente in Ω .)
- (ii) zweimaliges Würfeln:

$$\Omega = \{(i, j) : i, j \in \{1, \dots, 6\}\} = \{1, 2, \dots, 6\} \times \{1, 2, \dots, 6\} = \{1, 2, \dots, 6\}^2$$
 also $|\Omega| = 36$.
- (iii) Münzwurf: $\Omega = \{ \text{Kopf}, \text{Zahl} \}$.
- (iv) Autos am Darmstädter Kreuz am 25. August 2007: $\Omega = \{0, 1, 2, 3, \dots\} = \mathbb{N} \cup \{0\}$
- (v) Temperatur in Grad Kelvin am Luisenplatz am 1. Dezember 2007, 12 Uhr Mittags:
 $\Omega = [0, \infty[$ oder realistischer $[250, 290]$ ($0^\circ\text{C} = 273.15^\circ\text{K}$)

In den ersten vier Fällen sind die Ergebnisräume **endlich** oder **abzählbar unendlich**. Solche Ergebnisräume nennt man auch **diskret**. Im fünften Fall ist der Ergebnisraum nicht mehr abzählbar, sondern eine **kontinuierliche** Menge.

Die Wahrscheinlichkeitstheorie zu kontinuierlichen Ergebnisräumen ist mathematisch anspruchsvoller als die zu diskreten Ergebnisräumen. Daher betrachten wir **zunächst nur diskrete** Ergebnisräume Ω .

Ereignisse

Teilmengen $A \subset \Omega$ von Ω heißen **Ereignisse**. Die Gesamtheit aller Ereignisse ist somit nichts weiter als $\mathcal{P}(\Omega)$, also die **Potenzmenge** von Ω . Unter der Potenzmenge von Ω versteht man

die Gesamtheit aller Teilmengen von Ω einschließlich der leeren Menge \emptyset und der Menge Ω selber.

Beachten Sie: Ereignisse sind Elemente der Potenzmenge $\mathcal{P}(\Omega)$ von Ω , also Teilmengen von Ω , während Elementarereignisse Elemente von Ω sind.

Beispiele

- (i) $A = \{1, 3, 5\} = \text{Augenzahl ungerade}$
- (ii) $A = \{(5, 6), (6, 5), (6, 6)\} = \text{Augensumme} > 10$
- (iv) $A = \{22.000, 22.001, \dots\} = \{n : n \geq 22.000\} = \text{ungewöhnlich hohes Verkehrsaufkommen}$

Zwei Ereignisse sind besonders hervorzuheben:

- $\Omega = \text{das sichere Ereignis}$
- $\emptyset = \text{das unmögliche Ereignis.}$

Die bekannten Mengenoperationen lassen sich als **Operationen auf Ereignissen** interpretieren:

$A \cup B = A \text{ oder } B \text{ tritt ein}$

$A_1 \cup A_2 \cup \dots \cup A_n =: \bigcup_{k=1}^n A_k = \text{mind. eines der } A_k \text{ tritt ein}$

$A \cap B = A \text{ und } B \text{ treten ein}$

$A_1 \cap A_2 \cap \dots \cap A_n =: \bigcap_{k=1}^n A_k = \text{alle } A_k \text{ treten ein}$

$A^c := \Omega \setminus A := \{\omega \in \Omega : \omega \notin A\} = A \text{ tritt nicht ein}$

A^c heißt **Komplement** der Menge A (in Ω). Es gilt

$$\Omega^c = \emptyset \text{ und } \emptyset^c = \Omega.$$

Wahrscheinlichkeitsmaße

Für jedes Ereignis A legen wir im nächsten Schritt eine Wahrscheinlichkeit $P(A)$ zwischen 0 und 1 fest. $P(A)$ soll ein Maß dafür sein, dass das Ereignis A eintritt:

- tritt A niemals ein, so setzt man $P(A) = 0$. Insbesondere $P(\emptyset) = 0$.
- tritt A sicher ein, so setzt man $P(A) = 1$. Insbesondere $P(\Omega) = 1$.

Zusätzlich sollte gelten: Sind A und B disjunkte Ereignisse, d.h. A und B besitzen keine gemeinsamen Elementarereignisse, also $A \cap B = \emptyset$, so ist

$$P(A \cup B) = P(A) + P(B). \quad (2.1)$$

Diese Eigenschaft von P bezeichnet man als **Additivität**.

Aus (2.1) folgt unmittelbar: sind A_1, \dots, A_n paarweise disjunkte Ereignisse, d.h. $A_k \cap A_l = \emptyset$ für $k \neq l$, so folgt:

$$P(A_1 \cup \dots \cup A_n) = P(A_1) + \dots + P(A_n). \quad (2.2)$$

Gilt schließlich auch für jede **unendliche** Folge (A_n) paarweiser disjunkter Ereignisse

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k) \quad (2.3)$$

so spricht man von **σ -Additivität**.

Definition Ein **diskreter Wahrscheinlichkeitsraum** ist ein Paar (Ω, P) , wobei

- Ω eine nichtleere, diskrete (d.h. endliche oder abzählbar unendliche) Menge
- P ein **diskretes Wahrscheinlichkeitsmaß** auf Ω , d.h. eine Abbildung

$$P : \mathcal{P}(\Omega) \rightarrow \mathbb{R}$$

mit folgenden Eigenschaften:

- $P(A) \geq 0 \forall A \in \mathcal{P}(\Omega)$ (Nichtnegativität)
- $P(\Omega) = 1$ (Normiertheit)
- $P(\bigcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} P(A_k)$ für jede Folge (A_k) paarweise disjunkter Ereignisse (σ -Additivität).

Rechenregeln für P

- P ist (insbesondere) **endlich additiv**, d.h. für A_1, \dots, A_n paarweise disjunkt, ist

$$P(A_1 \cup \dots \cup A_n) = P(A_1) + \dots + P(A_n) = \sum_{k=1}^n P(A_k).$$

- $P(A^c) = 1 - P(A)$, denn A und A^c sind disjunkt, $A \cup A^c = \Omega$, also

$$1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c).$$

- $P(\emptyset) = 0$, denn $\emptyset^c = \Omega$, also

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

- $A \subset B$ impliziert $P(A) \leq P(B)$

denn $B = A \cup (B \cap A^c)$ und A und $B \cap A^c$ sind disjunkt, also

$$P(B) = P(A) + P(B \cap A^c) \geq P(A).$$

Konstruktion von Wahrscheinlichkeitsmaßen mit Hilfe von Wahrscheinlichkeitsfunktionen

Eine **Wahrscheinlichkeitsfunktion** (auf Ω) ist eine Funktion $p : \Omega \rightarrow [0, 1]$ mit

$$\sum_{\omega \in \Omega} p(\omega) = 1 \quad (2.4)$$

Bemerkung Beachten Sie, dass es sich bei (2.4) um eine unendliche Summe handelt, falls Ω unendlich viele Elemente enthält. Gemeint ist mit (2.4) also, dass die (möglicherweise unendliche) Reihe $\sum_{\omega \in \Omega} p(\omega)$ konvergiert und ihr Wert gleich 1 ist. Hierbei kommt es auf die **Reihenfolge**, in der die Wahrscheinlichkeiten $p(\omega)$ aufsummiert werden, **nicht** an, denn die Reihe ist wegen der Nichtnegativität der Summanden $p(\omega)$ absolut konvergent.

Zu gegebener Wahrscheinlichkeitsfunktion p definieren wir die Wahrscheinlichkeit $P(A)$ eines Ereignisses A durch

$$P(A) := \sum_{\omega \in A} p(\omega). \quad (2.5)$$

Die Wahrscheinlichkeit von A ist also gleich der Summe der Wahrscheinlichkeiten aller Elementarereignisse ω die in A liegen. Die so definierte Abbildung P ist ein diskretes Wahrscheinlichkeitsmaß auf Ω , d.h. nichtnegativ, normiert und σ -additiv.

Umgekehrt können wir zu jedem diskreten Wahrscheinlichkeitsmaß P auf Ω durch

$$p(\omega) := P(\{\omega\}), \omega \in \Omega \quad (2.6)$$

eine **Wahrscheinlichkeitsfunktion** auf Ω definieren.

Durch (2.5) und (2.6) ist also eine 1-1 Beziehung zwischen allen Wahrscheinlichkeitsmaßen über Ω und allen Wahrscheinlichkeitsfunktionen über Ω gegeben.

Beispiele

- (i) Beim Würfeln mit einem fairen Würfel ist jede der sechs möglichen Augenzahlen gleichwahrscheinlich. Man setzt daher

$$p(\omega) = \frac{1}{6} \text{ für } \omega \in \Omega = \{1, 2, 3, 4, 5, 6\}.$$

Es folgt z.B.

$$P(\text{Augenzahl ungerade}) = P(\{1, 3, 5\}) = \frac{3}{6} = \frac{1}{2}.$$

- (ii) Beim zweimaligen Würfeln mit einem fairen Würfel ist wiederum jedes der 36 Elementarereignisse aus $\Omega = \{1, 2, 3, 4, 5, 6\}^2$ gleichwahrscheinlich, also $p(\omega) = \frac{1}{36} \forall \omega \in \Omega$. Es folgt z.B.

$$P(\text{Augensumme} > 10) = P(\{(5, 6), (6, 5), (6, 6)\}) = \frac{3}{36} = \frac{1}{12}.$$

Beide Beispiele sind Spezialfälle eines Laplaceschen Wahrscheinlichkeitsraumes.

Laplacescher Wahrscheinlichkeitsraum

Ist Ω eine endliche Menge, so definiert

$$p(\omega) := \frac{1}{|\Omega|}, \quad \omega \in \Omega$$

eine Wahrscheinlichkeitsfunktion auf Ω . Für die Wahrscheinlichkeit $P(A)$ eines beliebigen Ereignisses folgt hieraus sofort

$$P(A) = \sum_{\omega \in A} \frac{1}{|\Omega|} = \frac{|A|}{|\Omega|}. \quad (2.7)$$

$P(A)$ heißt **Laplace-Wahrscheinlichkeit von A** . Da jedes Elementarereignis gleichwahrscheinlich ist, spricht man von P auch als der **Gleichverteilung auf Ω** .

Die Berechnung der Wahrscheinlichkeit $P(A)$ in (2.7) führt auf das Problem der **Abzählung der Elemente in A** , also auf ein **Abzählproblem**. Die wichtigsten Abzählprobleme sollen im folgenden anhand von einfachen **Urnenmodellen** illustriert werden:

Eine Urne enthalte n unterscheidbare Kugeln $1, 2, \dots, n$. Wir unterscheiden dann das k -malige Ziehen einer Kugel aus der Urne mit/ohne Zurücklegen, wobei es auf die Reihenfolge der gezogenen Kugeln ankommt/nicht ankommt:

1) in Reihenfolge mit Zurücklegen

$$\Omega = \{\omega = (x_1, \dots, x_k) : x_i \in \{1, \dots, n\}\}, |\Omega| = n^k$$

d.h., ein Elementarereignis $\omega = (x_1, \dots, x_k)$ ist ein **k -Tupel**, d.h. eine geordnete Menge der Länge k , wobei x_i für die Nummer der i -ten gezogenen Kugel steht.

2) in Reihenfolge ohne Zurücklegen

$$\Omega = \{\omega = (x_1, \dots, x_k) : x_i \in \{1, \dots, n\}, x_i \neq x_j \text{ für } i \neq j\}$$

$$|\Omega| = n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-k+1) = \frac{n!}{(n-k)!}.$$

Zur Erinnerung: Fakultätsfunktion

$$m! := m(m-1) \cdot (m-2) \cdot \dots \cdot 2 \cdot 1 = \prod_{k=1}^m k, \quad \text{und } 0! := 1.$$

Insbesondere

$$n! = n \cdot (n-1)! = n \cdot (n-1) \cdot (n-2)! = \dots = n \cdot (n-1) \cdot \dots \cdot (n-k+1) \cdot (n-k)!,$$

also

$$\frac{n!}{(n-k)!} = n \cdot (n-1) \cdot \dots \cdot (n-k+1).$$

Für $k = n$ erhält man als Spezialfall

$$|\Omega| = \frac{n!}{(n-n)!} = \frac{n!}{0!} = n!$$

$n!$ ist also gleich der Anzahl aller möglichen Anordnungen (oder auch **Permutationen**) der n -elementigen Menge $\{1, \dots, n\}$.

3) ohne Reihenfolge ohne Zurücklegen

$$\Omega = \{\omega = \{x_1, \dots, x_k\} : x_i \in \{1, 2, \dots, n\}, x_i \neq x_j \text{ für } i \neq j\}$$

Im Unterschied zum Ziehen ohne Zurücklegen werden nun alle k -Tupel (x_1, \dots, x_k) , die zu derselben Menge der gezogenen Kugeln führen, zu einem Elementarereignis zusammengefasst. Insgesamt gibt es $k!$ solcher Tupel (das entspricht also gerade der Anzahl der Permutationen der Menge der k gezogenen Kugeln), also erhalten wir insgesamt

$$\frac{n!}{(n-k)!} \cdot \frac{1}{k!} = \binom{n}{k}$$

Elementarereignisse. Es gilt also

$$|\Omega| = \binom{n}{k}.$$

Insbesondere: $\binom{n}{k}$ ist gleich der Anzahl aller k -elementigen Teilmengen aus einer n -elementigen Grundmenge.

Alternative Darstellung von Ω : Unter allen k -Tupeln, die zur selben Menge $\{x_1, \dots, x_k\}$ führen, gibt es genau ein Tupel $(x_{(1)}, \dots, x_{(k)})$, in dem die Elemente ihrer Größe nach angeordnet sind:

$$x_{(1)} < x_{(2)} < \dots < x_{(k)}.$$

Wir können auch schreiben

$$\Omega = \{(x_1, \dots, x_k) : x_i \in \{1, \dots, n\}, x_1 < x_2 < \dots < x_k\}.$$

4) ohne Reihenfolge mit Zurücklegen

Analog zu 3) ordnen wir wieder die Nummern der gezogenen Kugeln der Größe nach an:

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(k)} \tag{2.8}$$

wobei wegen des Zurücklegens Kugeln mehrfach gezogen werden können.

Durch Übergang von $x_{(i)}$ zu $x_{(i)} + i - 1$ erhält man aus (2.8) eine streng monoton aufsteigende Folge

$$x_{(1)} < x_{(2)} + 1 < x_{(3)} + 2 < \dots < x_{(k)} + k - 1.$$

Wir erhalten als Stichprobenraum in diesem Falle also

$$\Omega = \{(x_1, \dots, x_k) : x_i \in \{1, \dots, n, n+1, \dots, n+k-1\}, x_1 < x_2 < \dots < x_k\}.$$

Für die Mächtigkeit $|\Omega|$ von Ω ergibt sich nach 3)

$$|\Omega| = \binom{n+k-1}{k}.$$

Bedingte Wahrscheinlichkeiten und Unabhängigkeit

Ist über den Ausgang eines Zufallsexperiments bereits eine Teilinformation verfügbar, ändern sich entsprechend die Wahrscheinlichkeiten der Elementarereignisse.

Beispiel

Zweimaliges Würfeln eines fairen Würfels

$$P(\text{Augensumme} > 10) = \frac{1}{12}.$$

Wie ändert sich diese Wahrscheinlichkeit, wenn bereits bekannt ist, dass beim ersten Würfeln eine 6 gewürfelt wurde? Unter dieser Annahme bleiben nur noch sechs gleichwahrscheinliche Möglichkeiten für die zweite Augenzahl übrig, von denen die Augenzahlen 5 und 6 insgesamt zu einer Augensumme größer als 10 führen. Für die Wahrscheinlichkeit des Ereignisses $\text{Augenzahl} > 10$ unter der Bedingung 1. Augenzahl 6 ergibt sich somit

$$P(\text{Augensumme} > 10 \mid \text{1. Augenzahl } 6) = \frac{2}{6} = \frac{1}{3}.$$

Die bedingte Wahrscheinlichkeit ist also viermal höher als die ursprüngliche "a priori" Wahrscheinlichkeit.

Definition Für Ereignisse A, B mit $P(B) > 0$ heißt

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

die **bedingte Wahrscheinlichkeit von A unter der Bedingung B** (oder auch: **die bedingte Wahrscheinlichkeit von A gegeben B**). Im Falle $P(B) = 0$ setzen wir einfach $P(A \mid B) := 0$.

Eigenschaften der bedingten Wahrscheinlichkeit

- $P(A \mid B) \in [0, 1]$
- $P(\emptyset \mid B) = 0$

- Gilt $P(B) > 0$, so ist $P(\Omega | B) = 1$ und

$$P(\cdot | B) : \mathcal{P}(\Omega) \rightarrow [0, 1], A \mapsto P(A | B)$$

ist wieder eine diskrete Wahrscheinlichkeitsverteilung auf Ω . $P(\cdot | B)$ heißt **bedingte Wahrscheinlichkeitsverteilung unter der Bedingung B** .

Beispiel (Laplacescher Wahrscheinlichkeitsraum)

Ω endlich, $P(A) = \frac{|A|}{|\Omega|}$ sei die Gleichverteilung auf Ω . Dann folgt für $B \neq \emptyset$

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

Insbesondere: Die bedingte Wahrscheinlichkeitsverteilung ist im Falle des Laplaceschen Wahrscheinlichkeitsraumes gerade die Gleichverteilung auf B .

Beispiel

Bedingte Wahrscheinlichkeiten bilden die Grundlage für das Tarifsystem von Versicherungen. Verunglücken etwa mehr Männer als Frauen, sollten entsprechende Prämien einer Versicherung gegen Arbeitsunfälle für Männer höher als für Frauen sein, etwa:

$$P(\text{Unfall} | V \text{ weiblich}) = 0.002$$

$$P(\text{Unfall} | V \text{ männlich}) = 0.005.$$

Kennt man noch den Anteil der männlichen und weiblichen Versicherungsnehmer, etwa

$$P(V \text{ weiblich}) = \frac{2}{5} = 1 - P(V \text{ männlich}),$$

so kann man hieraus die totale Wahrscheinlichkeit eines Arbeitsunfalls errechnen:

$$\begin{aligned} P(\text{Unfall}) &= P(\text{Unfall und } V \text{ weiblich}) + P(\text{Unfall und } V \text{ männlich}) \\ &= P(\text{Unfall} | V \text{ weiblich})P(V \text{ weiblich}) \\ &\quad + P(\text{Unfall} | V \text{ männlich})P(V \text{ männlich}) \\ &= 0.002 \frac{2}{5} + 0.005 \frac{3}{5} = 0.0038. \end{aligned}$$

Die Berechnung der "totalen" Wahrscheinlichkeit für einen Arbeitsunfall ist ein Spezialfall des ersten Teils des folgenden Satzes.

Satz

Es seien B_1, \dots, B_n disjunkte Teilmengen von Ω und $A \subset B_1 \cup \dots \cup B_n$. Dann folgt:

- (i) **(Formel von der totalen Wahrscheinlichkeit)**

$$P(A) = \sum_{k=1}^n P(A | B_k)P(B_k). \quad (2.9)$$

(ii) **(Formel von Bayes)** Für $P(A) > 0$ gilt

$$P(B_i | A) = \frac{P(A | B_i)P(B_i)}{\sum_{k=1}^n P(A | B_k)P(B_k)}. \quad (2.10)$$

Beispiel

In obigem Beispiel kennt man bereits die totale Wahrscheinlichkeit eines Arbeitsunfalls

$$P(\text{Arbeitsunfall}) = 0.0038.$$

Die Formel von Bayes liefert nun für die "umgekehrte" bedingte Wahrscheinlichkeit

$$\begin{aligned} P(\text{V männlich} | \text{Arbeitsunfall}) &= \frac{P(\text{Arbeitsunfall} | \text{V männlich})P(\text{V männlich})}{P(\text{Arbeitsunfall})} \\ &= \frac{0.003}{0.0038} = 0.789. \end{aligned}$$

Wie zu erwarten handelt es sich bei einer verunglückten Person in fast 80% aller Fälle um Männer. Dieses Verhältnis kann sich aber ins Gegenteil verkehren, wenn entweder der Anteil der weiblichen Versicherungsnehmer den Anteil der männlichen Versicherungsnehmer weit übersteigt oder die bedingte Wahrscheinlichkeit $P(\text{Arbeitsunfall} | \text{V weiblich})$ für einen Arbeitsunfall eines weiblichen Versicherungsnehmers die entsprechende Wahrscheinlichkeit eines Arbeitsunfalles eines männlichen Versicherungsnehmers weit übersteigt.

Beispiel

Mitunter liefert die Formel von Bayes scheinbar überraschende Aussagen wie im Falle des folgenden Tests auf eine seltene Krankheit.

Angenommen, 5 Promille der Bevölkerung haben eine seltene Krankheit K , d.h.

$$P(K) = 0.005.$$

Ein medizinischer Test zeigt bei 99% der Erkrankten eine positive Reaktion, d.h.

$$P(\text{Test positiv} | K) = 0.99.$$

Allerdings zeigt besagter Test auch bei 2% der Gesunden eine positive Reaktion, d.h.

$$P(\text{Test positiv} | K^c) = 0.02.$$

Von besonderem Interesse ist nun offenbar folgende

Frage: Angenommen, der Test ist positiv. Wie groß ist die Wahrscheinlichkeit, dass die getestete Person tatsächlich an K erkrankt ist? Wie groß ist also die bedingte Wahrscheinlichkeit

$$P(K | \text{Test positiv})?$$

Die Formel von Bayes liefert

$$\begin{aligned} P(K | \text{Test positiv}) &= \frac{P(\text{Test positiv} | K)P(K)}{P(\text{Test positiv} | K) \cdot P(K) + P(\text{Test positiv} | K^c)P(K^c)} \\ &= \frac{0.99 \cdot 0.005}{0.99 \cdot 0.005 + 0.02 \cdot 0.995} = \frac{495}{2485} \sim 0.2. \end{aligned}$$

Also: Nur in 2 von 10 Fällen mit positivem Testergebnis ist die getestete Person auch wirklich an K erkrankt.

Unabhängigkeit

Ist $P(A) = P(A|B)$, d.h. die Wahrscheinlichkeit von A **unabhängig** davon, ob das Ereignis B eingetreten ist oder nicht, so folgt:

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

und damit

$$P(A \cap B) = P(A) \cdot P(B). \quad (2.11)$$

Zwei Ereignisse A und B mit (2.11) heißen (**stochastisch**) **unabhängig**.

Allgemeiner

Definition Die Ereignisse A_1, \dots, A_n heißen (stochastisch) unabhängig, falls für jede nicht-leere Teilmenge $\{i_1, \dots, i_k\} \subset \{1, \dots, n\}$ gilt:

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

Man beachte, dass zum Nachweis der Unabhängigkeit dreier Ereignisse A , B und C , der Nachweis der **paarweisen Unabhängigkeit** je zweier Ereignisse nicht ausreicht. Als Beispiel betrachten wir beim zweimaligen Werfen einer fairen Münze die Ereignisse

$$A = \text{1. Wurf Zahl}$$

$$B = \text{2. Wurf Zahl}$$

$$C = \text{1. und 2. Wurf gleich.}$$

Diese sind paarweise unabhängig aber nicht unabhängig, denn $P(A) = P(B) = P(C) = \frac{1}{2}$, $P(A \cap B) = P(A \cap C) = P(B \cap C) = \frac{1}{4}$, aber

$$P(A \cap B \cap C) = \frac{1}{4} \neq P(A)P(B)P(C).$$

Beispiel Beim zweimaligen Würfeln eines fairen Würfels ist die erste Augenzahl offenbar "unabhängig" von der zweiten Augenzahl, also jedes Ereignis A , das nur von der ersten Zahl abhängt, unabhängig von jedem Ereignis B , das nur von der zweiten Augenzahl abhängt, etwa:

$$A = \text{1. Augenzahl gerade}, \quad P(A) = \frac{1}{2}$$

$$B = \text{2. Augenzahl} \geq 5, \quad P(B) = \frac{1}{3}.$$

Dann gilt

$$\begin{aligned} P(A \cap B) &= P(\{(2, 5), (2, 6), (4, 5), (4, 6), (6, 5), (6, 6)\}) \\ &= \frac{6}{36} = \frac{1}{6} = \frac{1}{2} \cdot \frac{1}{3} = P(A) \cdot P(B). \end{aligned}$$

2. Zufallsvariablen und Verteilungen

Im ganzen Abschnitt sei (Ω, P) ein diskreter Wahrscheinlichkeitsraum. Eine Funktion

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

heißt **Zufallsvariable (auf Ω)**. Da Ω abzählbar, ist auch das Bild

$$X(\Omega) = \{X(\omega) : \omega \in \Omega\} \subset \mathbb{R}$$

abzählbar.

Für $x \in \mathbb{R}$ betrachten wir insbesondere das Ereignis

$$\{X = x\} := \{\omega \in \Omega : X(\omega) = x\} = X \text{ nimmt den Wert } x \text{ an}$$

Durch

$$p_X(x) := P(X = x), \quad x \in X(\Omega)$$

wird dann eine neue Wahrscheinlichkeitsfunktion auf $X(\Omega)$ definiert. Das zugehörige diskrete Wahrscheinlichkeitsmaß P_X auf $\mathcal{P}(X(\Omega))$ heißt **Verteilung von X (unter P)**.

Für beliebige Ereignisse $A \subset X(\Omega)$ gilt offenbar

$$\begin{aligned} P_X(A) &= \sum_{x \in A} p_X(x) = \sum_{x \in A} P(X = x) \\ &= P\left(\underbrace{\bigcup_{x \in A} \{\omega : X(\omega) = x\}}_{=\{\omega : X(\omega) \in A\}}\right) = P(X \in A). \end{aligned}$$

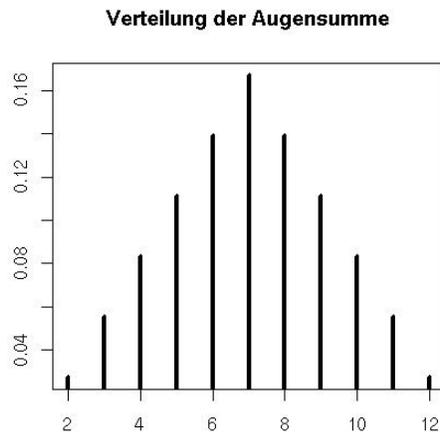
Beispiel Beim zweimaligen Würfeln eines fairen Würfels sei X die Augensumme. X ist eine Zufallsvariable mit Werten in der Menge $\{2, 3, \dots, 12\}$, von denen aber nicht alle Werte mit gleicher Wahrscheinlichkeit von X angenommen werden. Vielmehr gilt:

$$\begin{aligned} p_X(2) &= P(\{(k, l) \in \Omega : k + l = 2\}) = P(\{(1, 1)\}) = \frac{1}{36} \\ p_X(12) &= P(\{6, 6\}) = \frac{1}{36} \end{aligned}$$

und für die übrigen Werte

$$\begin{aligned} p_X(3) &= p_X(11) = \frac{2}{36}, p_X(4) = p_X(10) = \frac{3}{36} \\ p_X(5) &= p_X(9) = \frac{4}{36}, p_X(6) = p_X(8) = \frac{5}{36} \\ p_X(7) &= \frac{6}{36}. \end{aligned}$$

Graphische Veranschaulichung der Verteilung von X mit Hilfe eines **Stabdiagramms**



Die Verteilungsfunktion einer Zufallsvariablen

Die Funktion

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

heißt **Verteilungsfunktion** von X . Sie besitzt wie jede empirische Verteilungsfunktion (siehe Abschnitt 1.2) folgende Eigenschaften:

- F ist monoton wachsend
- $0 \leq F \leq 1$, $\lim_{x \rightarrow -\infty} F(x) = 0$, $\lim_{x \rightarrow \infty} F(x) = 1$
- F ist rechtsseitig stetig.

Unabhängigkeit von Zufallsvariablen

Definition Es seien X_1, X_2, \dots, X_n Zufallsvariablen auf dem Wahrscheinlichkeitsraum (Ω, P) . X_1, \dots, X_n heißen (**stochastisch**) **unabhängig**, falls für alle Teilmengen B_1, \dots, B_n von \mathbb{R} gilt:

$$P(X_1 \in B_1, \dots, X_n \in B_n) = P(X_1 \in B_1) \cdot \dots \cdot P(X_n \in B_n). \quad (2.12)$$

Die Zufallsvariablen X_1, \dots, X_n sind also genau dann (stochastisch) unabhängig, wenn für beliebige Teilmengen B_1, \dots, B_n die Ereignisse

$$\{X_1 \in B_1\}, \dots, \{X_n \in B_n\}$$

(stochastisch) unabhängig sind.

Äquivalent zu (2.12) ist folgende, in der Praxis einfacher zu überprüfende Bedingung: Für alle $x_1, \dots, x_n \in \mathbb{R}$ ist

$$P(X_1 = x_1, \dots, X_n = x_n) = P(X_1 = x_1) \cdot \dots \cdot P(X_n = x_n). \quad (2.13)$$

Beachten Sie, dass $P(X_k = x_k) = 0$ für die weitaus meisten Werte $x_k \in \mathbb{R}$, nämlich mindestens für alle $x_k \in \mathbb{R} \setminus X_k(\mathbb{R})$.

Beispiel Beim zweimaligen Würfeln sei X_1 die erste Augenzahl und X_2 die zweite. Mit (2.13) ist dann einfach zu sehen, dass X_1 und X_2 unabhängig sind.

Spezielle Verteilungen

Bernoulli-Verteilung

Fixiere eine Teilmenge $A \subset \Omega$ und definiere

$$X(\omega) := \begin{cases} 1 & \text{für } \omega \in A \\ 0 & \text{für } \omega \in A^c. \end{cases}$$

Wir interpretieren das Ereignis $\{X = 1\} = A$ als "Erfolg". Dementsprechend bezeichnen wir

$$p := P(X = 1) = P(A)$$

als **Erfolgswahrscheinlichkeit**. Entsprechend gilt für die Wahrscheinlichkeit eines Mißerfolges

$$P(X = 0) = P(A^c) = 1 - P(A) = 1 - p.$$

Definition Es sei $p \in [0, 1]$. Das durch die Wahrscheinlichkeitsfunktion $p : \{0, 1\} \rightarrow [0, 1]$

$$p(1) = p, \text{ und } p(0) = 1 - p$$

definierte Wahrscheinlichkeitsmaß auf $\{0, 1\}$ heißt **Bernoulli-Verteilung zu p** . Zufallsexperimente, die nur zwei mögliche Ausgänge kennen, nennt man entsprechend **Bernoulli-Experimente**.

Beispiele für Bernoulli-Experimente

- Werfen einer fairen Münze: $P(\text{Kopf}) = P(\text{Zahl}) = \frac{1}{2}$
- Geschlecht eines Neugeborenen: $P(\text{weiblich}) = 0.47$, $P(\text{männlich}) = 0.53$
- Ziehen einer Kugel aus einer Urne mit s schwarzen und w weißen Kugeln:

$$P(\text{gez. Kugel schwarz}) = \frac{s}{s+w}$$

Binomialverteilung

Es seien X_1, \dots, X_n unabhängige Zufallsvariablen, die alle Bernoulli-verteilt sind zu p .

Wir können X_i als Ausgang eines Bernoulli Experiments mit Erfolgswahrscheinlichkeit p interpretieren, wobei die Folge der n Experimente unabhängig ist. Dann zählt die Zufallsvariable

$$S_n := X_1 + \dots + X_n \in \{0, \dots, n\}$$

die **Gesamtanzahl der Erfolge**.

Für die Verteilung P_{S_n} der Summe S_n gilt dann

$$p_{S_n}(k) = P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k} =: b(k; n, p)$$

Hierbei ist $\binom{n}{k}$ gerade die Anzahl der n -Tupel mit genau k Einsen (und $n - k$ Nullen), p^k die Wahrscheinlichkeit für k Erfolge und $(1 - p)^{n-k}$ die Wahrscheinlichkeit für $n - k$ Mißerfolge.

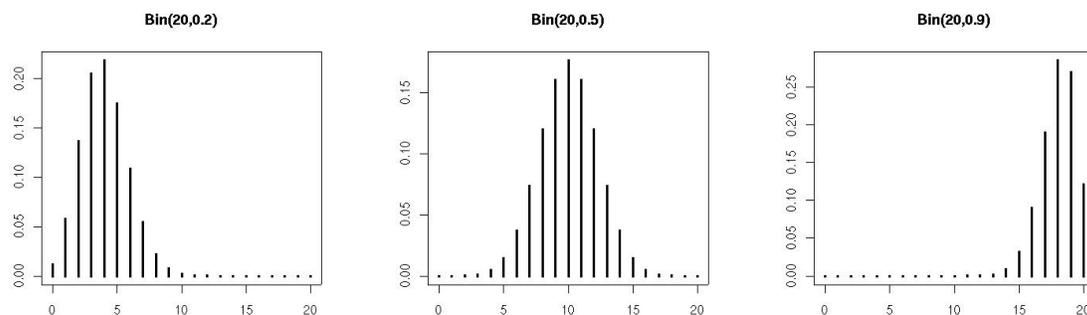
Definition Es sei $n \in \mathbb{N}$ und $p \in [0, 1]$. Das durch die Wahrscheinlichkeitsfunktion

$$b(\cdot; n, p) : \{0, \dots, n\} \rightarrow [0, 1]$$

$$k \mapsto \binom{n}{k} p^k (1 - p)^{n-k}$$

definierte Wahrscheinlichkeitsmaß auf $\{0, \dots, n\}$ heißt **Binominalverteilung zu n und p** und wird mit $\text{Bin}(n, p)$ bezeichnet.

Wir haben insbesondere gesehen: Bei einer Folge von n unabhängigen Bernoulli-Experimenten mit Erfolgswahrscheinlichkeit p ist die Summe der Erfolge binominalverteilt mit Parameter n und p .



Geometrische Verteilung

Wie groß ist die Wahrscheinlichkeit, dass man mit einem fairen Würfel genau k Versuche benötigt, bis zum ersten Mal eine 6 gewürfelt wird?

Für $k = 1$ ist die gesuchte Wahrscheinlichkeit offensichtlich $\frac{1}{6}$, für $k = 2$ ist sie gleich $\frac{5}{6} \cdot \frac{1}{6}$, denn die gesuchte Wahrscheinlichkeit ist aufgrund der Unabhängigkeit der beiden Würfel gleich dem Produkt aus der Wahrscheinlichkeit, beim ersten Würfeln keine 6 zu würfeln ($= \frac{5}{6}$), und der Wahrscheinlichkeit, beim zweiten Würfeln eine 6 zu würfeln ($= \frac{1}{6}$).

Für allgemeines k können wir wie folgt vorgehen: Wir definieren eine Folge von Zufallsvariablen X_1, X_2, X_3, \dots durch

$$X_k := 1 \text{ falls beim } k\text{-ten Wurf eine 6 gewürfelt wird}$$

und $X_k := 0$ sonst. Offenbar sind die Zufallsvariablen X_1, X_2, X_3, \dots unabhängig Bernoulli-verteilt mit Erfolgswahrscheinlichkeit $p = \frac{1}{6}$. Das Ereignis A_k , im k -ten Wurf zum ersten Mal eine 6 zu würfeln, kann mit Hilfe dieser Zufallsvariablen nun wie folgt beschrieben werden:

$$A_k = \{X_1 = 0, X_2 = 0, \dots, X_{k-1} = 0, X_k = 1\}.$$

Aufgrund der Unabhängigkeit der Zufallsvariablen ergibt sich für die gesuchte Wahrscheinlichkeit

$$\begin{aligned} P(A_k) &= P(X_1 = 0, X_2 = 0, \dots, X_{k-1} = 0, X_k = 1) \\ &= P(X_1 = 0) \cdot P(X_2 = 0) \cdot \dots \cdot P(X_{k-1} = 0) \cdot P(X_k = 1) \\ &= \frac{5}{6} \cdot \frac{5}{6} \cdot \dots \cdot \frac{5}{6} \cdot \frac{1}{6} = \left(\frac{5}{6}\right)^{k-1} \frac{1}{6}. \end{aligned}$$

Allgemeiner Gegeben eine Folge von unabhängigen Zufallsvariablen X_1, X_2, X_3, \dots , die alle Bernoulli-verteilt sind zu $p > 0$. Definiere die **Wartezeit auf den ersten Erfolg**

$$T := \min\{k \geq 1 : X_k = 1\}.$$

Wie in obigem Fall der Wartezeit auf die erste 6 beim Würfeln mit einem fairen Würfel, erhalten wir für die Verteilung von T

$$\begin{aligned} P(T = k) &= P(X_1 = 0, X_2 = 0, \dots, X_{k-1} = 0, X_k = 1) \\ &= P(X_1 = 0) \cdot P(X_2 = 0) \cdot \dots \cdot P(X_{k-1} = 0) \cdot P(X_k = 1) \\ &= (1 - p)^{k-1} \cdot p \end{aligned}$$

für $k = 1, 2, 3, \dots$

Definition Es sei $p \in]0, 1]$ Das durch die Wahrscheinlichkeitsfunktion

$$\begin{aligned} g_p : \mathbb{N} &\mapsto [0, 1] \\ k &\mapsto (1 - p)^{k-1} p \end{aligned}$$

definierte Wahrscheinlichkeitsmaß auf \mathbb{N} heißt **geometrische Verteilung zu p** und wird mit $\text{Geom}(p)$ bezeichnet.

Poissonverteilung

Für $\lambda > 0$ definiert

$$\pi_\lambda(k) := e^{-\lambda} \frac{\lambda^k}{k!}, \quad k \in \mathbb{N}_0$$

eine Wahrscheinlichkeitsfunktion auf \mathbb{N}_0 , denn aus der Reihenentwicklung der Exponentialfunktion

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}, \quad x \in \mathbb{R}$$

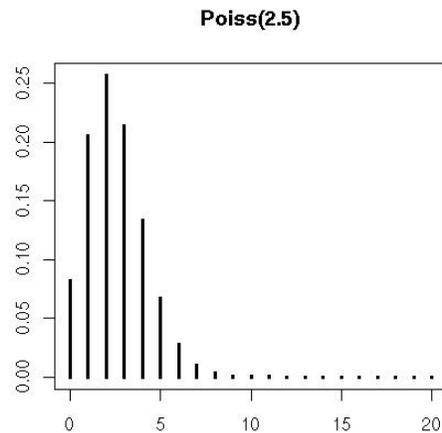
folgt

$$\sum_{k=0}^{\infty} \pi_\lambda(k) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot e^\lambda = e^0 = 1.$$

Definition Es sei $\lambda > 0$. Das durch die Wahrscheinlichkeitsfunktion

$$\begin{aligned} \pi_\lambda : \mathbb{N}_0 &\rightarrow [0, 1] \\ k &\mapsto e^{-\lambda} \frac{\lambda^k}{k!} \end{aligned}$$

definierte Wahrscheinlichkeitsmaß auf \mathbb{N}_0 heißt **Poissonverteilung zu λ** und wird mit $\text{Poiss}(\lambda)$ bezeichnet.



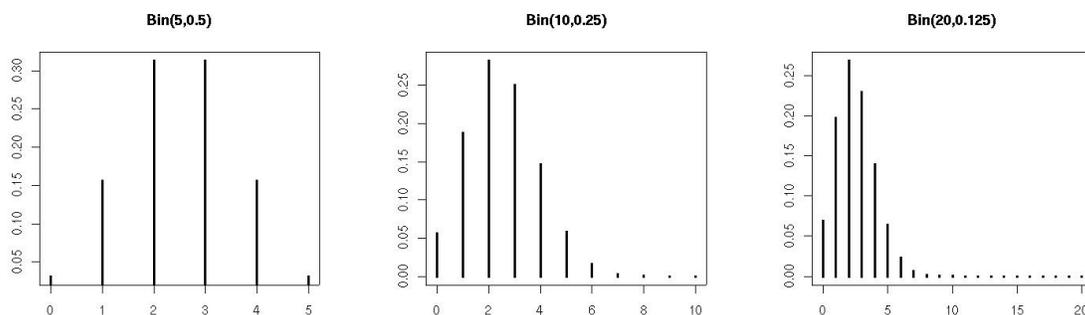
Die Poissonverteilung empfiehlt sich als Näherung der Binomialverteilung $\text{Bin}(n, p)$ für große n und kleine p . Die Approximation ist umso besser, je kleiner der Wert np^2 ist. Diese Näherung wird gerechtfertigt durch die folgende Beobachtung:

Poissonscher Grenzwertsatz

Es sei $(p_n) \subset [0, 1]$ eine Folge von Erfolgsparametern mit $\lim_{n \rightarrow \infty} np_n = \lambda > 0$. Dann folgt

$$\lim_{n \rightarrow \infty} b(k; n, p_n) = \pi_\lambda(k) \quad \text{für alle } k \in \mathbb{N}_0.$$

Mit anderen Worten: Die Wahrscheinlichkeitsfunktion der Binomialverteilung $\text{Bin}(n, p_n)$ konvergiert punktweise gegen die Wahrscheinlichkeitsfunktion der Poissonverteilung $\text{Poiss}(\lambda)$. Im folgenden eine Illustration dieser Konvergenz für $\lambda = 2.5$.



Eine näherungsweise Berechnung von Wahrscheinlichkeiten gewisser Ereignisse mit Hilfe einer Poissonverteilung ist immer dann gerechtfertigt, wenn es sich um seltene Ereignisse handelt.

Beispiel Bei der Herstellung von DVD-Scheiben ist ein Anteil von $p = 0.002$ bereits bei der Produktion defekt. Wie groß ist die Wahrscheinlichkeit, dass in einem Warenposten mit $n = 1.000$ DVD-Scheiben mindestens fünf Scheiben defekt sind?

Zur Beantwortung dieser Frage sei X die Anzahl der defekten DVD-Scheiben. Da es sich bei der Produktion einer defekten DVD-Scheibe (eher) um ein seltenes Ereignis handelt, empfiehlt sich eine Näherung der Verteilung von X mit Hilfe einer Poissonverteilung. Den Parameter λ wählt man gemäß der Regel

$$\lambda = np = 1000 \cdot 0.002 = 2.$$

Damit folgt für die gesuchte Wahrscheinlichkeit

$$\begin{aligned} P(X \geq 5) &= 1 - P(X \leq 4) = 1 - e^{-2} \left(\frac{2^0}{0!} + \frac{2^1}{1!} + \frac{2^2}{2!} + \frac{2^3}{3!} + \frac{2^4}{4!} \right) \\ &= 1 - e^{-2} \left(1 + 2 + 2 + \frac{4}{3} + \frac{2}{3} \right) \approx 0.05. \end{aligned}$$

Hypergeometrische Verteilung

Es sei eine Grundgesamtheit mit N Elementen gegeben, von denen K Elemente die Eigenschaft E besitzen. Aus dieser Grundgesamtheit werde n -mal ohne Zurücklegen gezogen. Wir sind interessiert an der Anzahl k der gezogenen Elemente, die die Eigenschaft E besitzen. Hierzu definieren wir

$$X = \text{Anzahl der gezogenen Elemente mit Eigenschaft } E.$$

Ist $\frac{n}{N}$ klein, so gibt es keinen großen Unterschied zwischen dem Ziehen ohne Zurücklegen und dem Ziehen mit Zurücklegen. Daher empfiehlt sich in diesem Falle eine Approximation der Verteilung von X durch die Binomialverteilung $\text{Bin}(n, p)$ mit $p = \frac{K}{N}$, also

$$P(X = k) \approx b(k; n, \frac{K}{N}).$$

Ist $\frac{n}{N}$ jedoch vergleichsweise groß, so muss die gesuchte Verteilung exakt berechnet werden:

$$P(X = k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}, \quad k = 0, \dots, n. \quad (2.14)$$

Zur Herleitung der Formel (2.14) für die gesuchte Wahrscheinlichkeit beachte man, dass $\binom{K}{k}$ (bzw. $\binom{N-K}{n-k}$) gerade die Anzahl der k (bzw. $n-k$)-elementigen Teilmengen einer K (bzw. $N-K$)-elementigen Grundmenge ist, während $\binom{N}{n}$ die Anzahl aller n -elementigen Teilmengen der Grundgesamtheit aus N Elementen ist.

Definiton Es sei $K \leq N$, $n \leq N$. Das durch die Wahrscheinlichkeitsfunktion

$$\begin{aligned} H(\cdot; n, N, K) &: \{0, \dots, n\} \rightarrow [0, 1] \\ k &\mapsto \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}} \end{aligned}$$

definierte Wahrscheinlichkeitsmaß auf $\{0, \dots, n\}$ heißt **Hypergeometrische Verteilung zu n, N und K** und wird mit $\text{Hyp}(n, N, K)$ bezeichnet.

3. Erwartungswert und Varianz

Erwartungswert und Varianz sind die beiden wichtigsten Kennzahlen einer Zufallsvariablen. Im ganzen Abschnitt sei (Ω, P) ein diskreter Wahrscheinlichkeitsraum, p die zugehörige Wahrscheinlichkeitsfunktion.

Der **Erwartungswert** $E(X)$ einer Zufallsvariablen X wird definiert als der Mittelwert

$$E(X) := \sum_{\omega \in \Omega} X(\omega)p(\omega) \quad (2.15)$$

der Funktionswerte $X(\omega)$ gewichtet mit den Einzelwahrscheinlichkeiten $p(\omega) = P(X = x)$.

Ist Ω endlich, so bereitet diese Definition keine Probleme. Im Falle Ω unendlich muss man noch Sorge tragen, dass die Reihe (2.15) absolut konvergiert. Dies ist dann der Fall, wenn die Reihe

$$\sum_{\omega \in \Omega} |X(\omega)|p(\omega)$$

konvergiert, und man sagt in diesem Fall, dass der Erwartungswert $E(X)$ von X existiert.

Beispiel X sei die Augenzahl beim Würfeln eines fairen Würfels

Dann gilt

$$E(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{7}{2}.$$

Der Erwartungswert stimmt also in diesem Falle mit dem arithmetischen Mittel der Funktionswerte überein.

Es sei X eine Zufallsvariable, deren Erwartungswert existiert. Ist x_1, x_2, \dots eine Aufzählung des Bildes $X(\Omega)$ von X , so folgt

$$\begin{aligned} E(X) &= \sum_{\omega \in \Omega} X(\omega)p(\omega) = \sum_k \sum_{\omega \in \Omega: X(\omega)=x_k} X(\omega)p(\omega) \\ &= \sum_k x_k P(X = x_k) = \sum_k x_k p_X(x_k). \end{aligned}$$

Insbesondere gilt also, dass der Erwartungswert einer Zufallsvariablen X **nur von ihrer Verteilung P_X abhängt!**

Rechenregeln für Erwartungswerte

Es seien X, Y Zufallsvariablen, deren Erwartungswerte existieren. Dann gilt:

- Linearität $E(aX + bY) = aE(X) + bE(Y)$ für alle $a, b \in \mathbb{R}$.
- Nichtnegativität $X \geq 0$ (d.h. $X(\omega) \geq 0$ für alle $\omega \in \Omega$)

$$\implies E(X) \geq 0.$$

- Monotonie $X \leq Y$ (d.h. $Y - X \geq 0$)

$$\implies E(X) \leq E(Y).$$

- Ist X konstant, also $X = c$ für eine Konstante c (d.h. $X(\omega) = c$ für alle $\omega \in \Omega$), so folgt

$$E(X) = c.$$

- Transformationssatz: Ist $h : \mathbb{R} \rightarrow \mathbb{R}$ eine stückweise stetige Funktion und ist x_1, x_2, x_3, \dots eine Aufzählung des Bildes $X(\Omega)$, so gilt: Der Erwartungswert der Zufallsvariablen $h(X)$ existiert, genau dann wenn die Summe $\sum_k |h(x_k)| p_X(x_k) < \infty$ konvergiert und in diesem Fall ist

$$E(h(X)) = \sum_k h(x_k) p_X(x_k) \quad (2.16)$$

- Sind X, Y unabhängig, so existiert auch der Erwartungswert von XY , und es gilt

$$E(XY) = E(X) E(Y).$$

Beispiele

- (i) Sind X_1, \dots, X_n unabhängig Bernoulli-verteilt mit Erfolgswahrscheinlichkeit p , so folgt

$$E(X_k) = 0 \cdot P(X_k = 0) + 1 \cdot P(X_k = 1) = p.$$

Insbesondere gilt für den Erwartungswert der Summe

$$S_n = X_1 + \dots + X_n$$

$$E(S_n) = E(X_1) + \dots + E(X_n) = p + \dots + p = np.$$

Da S_n binomialverteilt ist mit Parameter n und p , folgt insbesondere: Für den Erwartungswert einer binomialverteilten Zufallsvariablen S_n mit Parametern n und p gilt:

$$E(S_n) = np.$$

- (ii) Ist X Poiss(λ)-verteilt, so folgt

$$\begin{aligned} E(X) &= \sum_{k=0}^{\infty} k P(X = k) = \sum_{k=0}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} \\ &= \sum_{k=1}^{\infty} e^{-\lambda} \frac{\lambda \lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \lambda. \end{aligned}$$

Ein Maß für die Streuung der Funktionswerte $X(\omega)$ um ihren Erwartungswert $E(X)$ ist die mittlere quadratische Abweichung

$$\text{Var}(X) := E((X - E(X))^2) = \sum_{\omega \in \Omega} (X(\omega) - E(X))^2 p(\omega). \quad (2.17)$$

Sie heißt **Varianz** von X .

Damit der Ausdruck (2.17) wohldefiniert ist, müssen die Erwartungswerte $E(X)$ und $E((X - E(X))^2)$ existieren. Man kann zeigen, dass beide existieren, falls der Erwartungswert $E(X^2)$ von X^2 existiert.

Unter der **Standardabweichung** von X versteht man die Größe

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

Wie der Erwartungswert, so hängt auch die Varianz (und damit auch die Standardabweichung) nur von der Verteilung P_X von X unter P ab. Ist nämlich $x_1, x_2, x_3 \dots$ eine Aufzählung der Werte von X , so folgt

$$\begin{aligned} \text{Var}(X) &= \sum_{\omega \in \Omega} (X(\omega) - E(X))^2 p(\omega) \\ &= \sum_k \sum_{\omega \in \Omega: X(\omega)=x_k} (x_k - E(X))^2 p(\omega) \\ &= \sum_k (x_k - E(X))^2 p_X(x_k). \end{aligned}$$

Beispiel X sei die Augenzahl beim Würfeln eines fairen Würfels. Dann folgt

$$\text{Var}(X) = \left(1 - \frac{7}{2}\right)^2 \cdot \frac{1}{6} + \left(2 - \frac{7}{2}\right)^2 \cdot \frac{1}{6} + \dots + \left(6 - \frac{7}{2}\right)^2 \cdot \frac{1}{6} = \frac{35}{12}.$$

Rechenregeln für Varianzen

Es seien X, Y, X_1, \dots, X_n Zufallsvariablen, für die die Erwartungswerte $E(X^2)$, $E(Y^2)$, $E(X_1^2), \dots, E(X_n^2)$ existieren. Dann gilt:

- $\text{Var}(aX + b) = a^2 \text{Var}(X)$ für alle $a, b \in \mathbb{R}$.
- Verschiebungssatz $\text{Var}(X) = E(X^2) - (E(X))^2$.
- X, Y unabhängig $\Rightarrow \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$.

Beweis

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

Da X und Y unabhängig, folgt $E(XY) = E(X)E(Y)$, und damit verschwindet der dritte Term auf der rechten Seite.

Allgemeiner gilt die **Identität von Bienaymé**

Sind X_1, \dots, X_n unabhängig, so folgt

$$\text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n).$$

Beispiele

- (i) Sind X_1, \dots, X_n unabhängig Bernoulli-verteilt mit Erfolgswahrscheinlichkeit p , so folgt für die Varianz der Summe $S_n = X_1 + \dots + X_n$

$$\text{Var}(S_n) = \text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n).$$

Für die Varianz der Bernoulli-verteilten Zufallsvariablen X_k errechnet man sofort

$$\text{Var}(X_k) = E(X_k^2) - (E(X_k))^2 = p - p^2 = p(1 - p),$$

so dass

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

Da S_n binomialverteilt ist mit Parameter n und p , folgt insbesondere: Für die Varianz einer binomialverteilten Zufallsvariablen S_n mit Parameter n und p gilt

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

- (ii) Ist X Poiss(λ)-verteilt, so folgt

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

also

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

Kovarianz

Sind X und Y zwei Zufallsvariablen, deren Varianzen existieren, so ist die **Kovarianz**

$$\text{Cov}(X, Y) := E((X - E(X))(Y - E(Y)))$$

wohldefiniert. Sie ist das Analogon zur empirischen Kovarianz einer zweidimensionalen Messreihe. Die Größe

$$\rho(X, Y) := \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}}$$

heißt dementsprechend der **Korrelationskoeffizient** von X und Y . Ist $\rho(X, Y) = 0$, so heißen X und Y **unkorreliert**.

Die Kovarianz hängt nur von der **gemeinsamen Verteilung** P_{XY} der Zufallsvariablen X und Y unter P ab. Hierunter versteht man die diskrete Wahrscheinlichkeitsverteilung zur Wahrscheinlichkeitsfunktion

$$p_{XY}(x, y) := P(X = x, Y = y), \quad x \in X(\Omega), y \in Y(\Omega)$$

auf dem Produktraum $X(\Omega) \times Y(\Omega) := \{(x, y) : x \in X(\Omega), y \in Y(\Omega)\} \subset \mathbb{R}^2$.

Ist nämlich x_1, x_2, x_3, \dots eine Aufzählung der Werte von X und y_1, y_2, y_3, \dots eine Aufzählung der Werte von Y , so folgt

$$\begin{aligned} \text{Cov}(X, Y) &= \sum_{\omega \in \Omega} (X(\omega) - E(X)) (Y(\omega) - E(Y)) \\ &= \sum_k \sum_l \sum_{\omega \in \Omega : X(\omega) = x_k, Y(\omega) = y_l} (x_k - E(X)) (y_l - E(Y)) \\ &= \sum_k \sum_l (x_k - E(X)) (y_l - E(Y)) p_{XY}(x_k, y_l). \end{aligned}$$

Rechenregeln für Kovarianzen

- $\text{Cov}((a_X X + b_X), (a_Y Y + b_Y)) = a_X a_Y \text{Cov}(X, Y)$ für alle $a_X, a_Y, b_X, b_Y \in \mathbb{R}$.
- Verschiebungssatz $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$.
- X, Y unabhängig $\Rightarrow \text{Cov}(X, Y) = 0$. Die Umkehrung gilt im allgemeinen nicht.

4. Stetige Verteilungen

In vielen Fällen kann der Wertebereich einer Zufallsvariablen X nicht diskret gewählt werden (z.B. Wartezeiten, Laufzeiten, Körpergröße, Lufttemperatur,...) sondern muss als Intervall $[a, b]$ oder gleich ganz \mathbb{R} gewählt werden. Eine solche Zufallsvariable kann natürlich nicht auf einem diskreten Wahrscheinlichkeitsraum (Ω, P) definiert sein. Es bedarf hierzu also einer Erweiterung des Begriffes des Wahrscheinlichkeitsraumes auf überabzählbare Ergebnismengen Ω . Die mathematische Theorie zur rigorosen Durchführung dieser Erweiterung sprengt eindeutig den Rahmen dieser Vorlesung, man findet sie in Büchern zur Wahrscheinlichkeitstheorie.

Im folgenden betrachten wir nur den für die Anwendungen enorm wichtigen Fall stetig verteilter Zufallsvariablen X . Dabei heißt X **stetig verteilt mit Dichte f** , falls gilt

$$P(X \leq b) = \int_{-\infty}^b f(x) dx \quad \text{für alle } b \in \mathbb{R}. \quad (2.18)$$

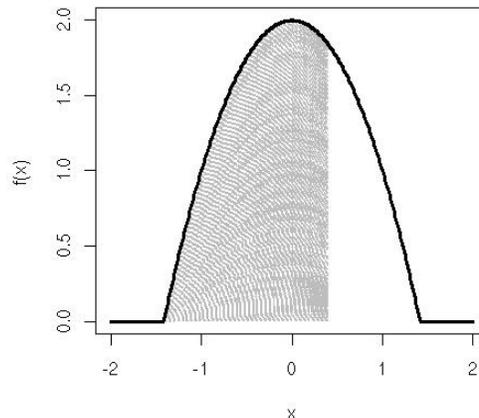
Hierbei ist $f : \mathbb{R} \rightarrow \mathbb{R}$ eine uneigentlich Riemann-integrierbare Funktion mit

- $f(x) \geq 0$ für alle $x \in \mathbb{R}$,
- $\int_{-\infty}^{+\infty} f(x) dx = 1$.

Für eine mit Dichte f stetig verteilte Zufallsvariable X wird also die Wahrscheinlichkeit der Ereignisse

$$\{\omega : X(\omega) \leq b\}$$

durch die schraffierte Fläche angegeben.



Wie im Falle diskreter Zufallsvariablen heißt die Funktion

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

die Verteilungsfunktion von X . Sie besitzt genau dieselben Eigenschaften wie im diskreten Fall.

Mit Hilfe von (2.18) können wir dann auch sofort die Wahrscheinlichkeit des Ereignisses $\{\omega : a < X(\omega) \leq b\}$ berechnen, denn

$$\begin{aligned} P(a < X \leq b) &= P(X \leq b) - P(X \leq a) \\ &= \int_{-\infty}^b f(x) dx - \int_{-\infty}^a f(x) dx \\ &= \int_a^b f(x) dx. \end{aligned} \quad (2.19)$$

Für eine stetig verteilte Zufallsvariable X gilt

$$P(X = x) = 0 \quad \forall x \in \mathbb{R},$$

d.h. X nimmt einen **bestimmten** Wert x nur mit Wahrscheinlichkeit 0 an. Dies ist ein fundamentaler Unterschied zu diskreten Zufallsvariablen. Damit gilt insbesondere

$$P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b) \quad \forall a, b \in \mathbb{R}. \quad (2.20)$$

Stochastische Unabhängigkeit

Der Begriff der stochastischen Unabhängigkeit lässt sich unmittelbar auf stetig verteilte Zufallsvariablen übertragen: zwei (stetig verteilte) Zufallsvariablen X und Y heißen **stochastisch unabhängig**, falls

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y) \quad \forall x, y \in \mathbb{R}.$$

Allgemeiner: Die (stetig verteilten) Zufallsvariablen X_1, \dots, X_n heißen **stochastisch unabhängig**, falls

$$P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n) = P(X_1 \leq x_1) \cdot P(X_2 \leq x_2) \cdot \dots \cdot P(X_n \leq x_n) \quad (2.21)$$

für alle $x_1, x_2, \dots, x_n \in \mathbb{R}$.

Die Analogie zum diskreten Fall erkennt man wie folgt: Ist $B_i :=] - \infty, x_i]$, so kann man (2.21) in der Form

$$P(X_1 \in B_1, X_2 \in B_2, \dots, X_n \in B_n) = P(X_1 \in B_1) \cdot P(X_2 \in B_2) \cdot \dots \cdot P(X_n \in B_n)$$

schreiben.

Erwartungswert, Varianz und Kovarianz

Ist X stetig verteilt mit Dichte f , so sagen wir, dass der Erwartungswert $E(X)$ von X existiert, falls die Funktion $|x|f(x)$ uneigentlich Riemann-integrierbar ist (dann ist auch $xf(x)$ uneigentlich Riemann-integrierbar) und man setzt in diesem Falle

$$E(X) := \int_{-\infty}^{+\infty} xf(x) dx.$$

Ist zusätzlich auch die Funktion $(x - E(X))^2 f(x)$ uneigentlich Riemann-integrierbar, so definiert man die Varianz $\text{Var}(X)$ durch

$$\text{Var}(X) := \int_{-\infty}^{+\infty} (x - E(x))^2 f(x) dx$$

und die Standardabweichung wie im diskreten Fall durch

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

Die Rechenregeln für Erwartungswerte und Varianz diskret verteilter Zufallsvariablen (siehe Abschnitt II.3) übertragen sich unmittelbar auf den Fall stetig verteilter Zufallsvariablen. Der Transformationssatz überträgt sich dabei wie folgt: Ist $h : \mathbb{R} \rightarrow \mathbb{R}$ eine stückweise stetige Funktion so gilt: Der Erwartungswert der Zufallsvariablen $h(X)$ existiert genau dann wenn die Funktion $|h(x)|f(x)$ uneigentlich Riemann-integrierbar ist und in diesem Fall ist

$$E(h(X)) = \int_{-\infty}^{+\infty} h(x)f(x) dx. \quad (2.22)$$

Zwei Zufallsvariablen X und Y heißen **gemeinsam stetig verteilt** mit **gemeinsamer stetiger Dichte** f_{XY} , falls gilt

$$P(X \leq a, Y \leq b) = \int_{-\infty}^a \int_{-\infty}^b f_{XY}(x, y) dx dy \quad \forall a, b \in \mathbb{R}$$

für eine integrierbare Funktion $f_{XY} : \mathbb{R}^2 \rightarrow \mathbb{R}$ mit

- $f_{XY}(x, y) \geq 0$ für alle $x, y \in \mathbb{R}$
- $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{XY}(x, y) dx dy = 1.$

Die Berechnung der Kovarianz $\text{Cov}(X, Y)$ erfolgt dann über die gemeinsame Dichte mit Hilfe der Formel

$$\text{Cov}(X, Y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - E(X))(y - E(Y))f_{XY}(x, y) dx dy.$$

Die Rechenregeln für die Kovarianzen für diskret verteilte Zufallsvariablen übertragen sich Wort für Wort auf den gemeinsam stetig verteilten Fall.

Wichtige stetige Verteilungen

Gleichverteilung

Für $a < b$ heißt eine Zufallsvariable mit Dichte

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

(stetig) gleichverteilt auf $[a, b]$. Für alle Teilintervalle $[c, d]$ folgt aus (2.19) und (2.20)

$$P(c \leq X \leq d) = P(c < X \leq d) = \int_c^d \frac{1}{b-a} dx = \frac{d-c}{b-a}.$$

Mit anderen Worten: X überdeckt Teilintervalle derselben Länge $d - c$ mit jeweils derselben Wahrscheinlichkeit. Dies erklärt die Bezeichnung Gleichverteilung.

X nimmt mit Wahrscheinlichkeit 1 nur Werte in $[a, b]$ an, denn

$$P(X \in [a, b]) = P(a \leq X \leq b) = \frac{b - a}{b - a} = 1.$$

Für Erwartungswert und Varianz einer auf $[a, b]$ gleichverteilten Zufallsvariablen gilt

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx = \int_a^b x \frac{1}{b-a} dx = \frac{1}{2} \frac{x^2}{b-a} \Big|_a^b = \frac{1}{2}(a+b)$$

$$\text{Var}(X) = \int_{-\infty}^{+\infty} \left(x - \frac{1}{2}(a+b)\right)^2 f(x) dx$$

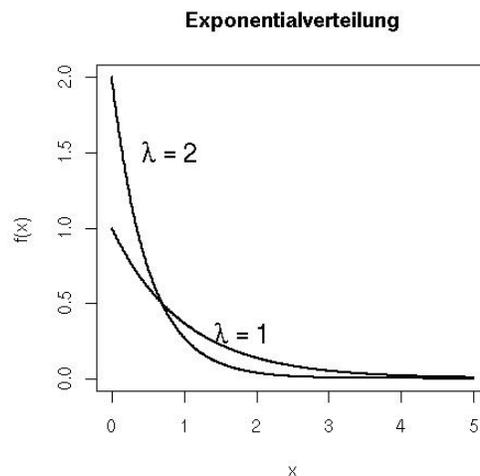
$$= \int_a^b \left(x - \frac{1}{2}(a+b)\right)^2 \frac{1}{b-a} dx = \frac{1}{12}(b-a)^2.$$

Exponentialverteilung

Für $\lambda > 0$ ist

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

eine Dichte. Die zugehörige Verteilung heißt **Exponentialverteilung** zum Parameter λ . Sie wird mit $\text{Exp}(\lambda)$ bezeichnet.



Die Exponentialverteilung ist das stetige Analogon der geometrischen Verteilung, die ja die Verteilung von Wartezeiten auf den ersten Erfolg in einer Folge von unabhängigen Bernoulli Experimenten beschreibt. Dementsprechend verwendet man die Exponentialverteilung zur Modellierung von stetig verteilten Wartezeiten.

Ist $X \text{Exp}(\lambda)$ verteilt, so gilt

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

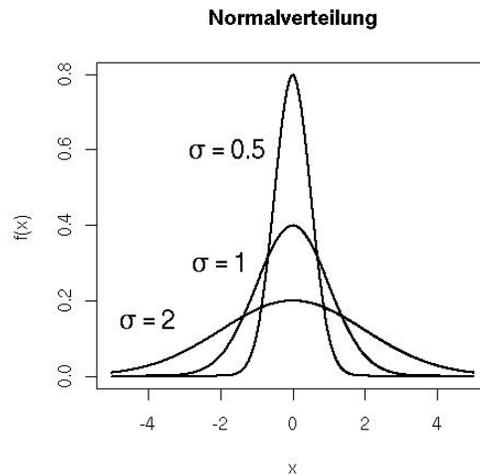
$$\text{Var}(X) = \lambda \int_0^{+\infty} \left(x - \frac{1}{\lambda}\right)^2 e^{-\lambda x} dx = \frac{1}{\lambda^2}.$$

Normalverteilung

Für $m \in \mathbb{R}$ und $\sigma > 0$ ist

$$f_{m,\sigma^2}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{(x-m)^2}{\sigma^2}}$$

eine Dichte. Die zugehörige Verteilung heißt **Normalverteilung** mit Mittel m und Varianz σ^2 . Sie wird mit $N(m, \sigma^2)$ bezeichnet. Im Falle $m = 0$ und $\sigma^2 = 1$ spricht man von der **Standardnormalverteilung**.



f_{m,σ^2} besitzt ein absolutes Maximum in $x = m$ und Wendepunkte in $\pm\sigma$. Wegen ihrer Form wird f auch als **Gaußsche Glockenkurve** bezeichnet. σ bestimmt **Breite** und **Höhe** der Glockenkurve.

Eine Zufallsvariable X mit Dichte f_{m,σ^2} heißt normalverteilt mit Mittel m und Varianz σ^2 , denn es gilt

$$E(X) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} x e^{-\frac{1}{2}\frac{(x-m)^2}{\sigma^2}} dx = m$$

$$\text{Var}(X) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} (x-m)^2 e^{-\frac{1}{2}\frac{(x-m)^2}{\sigma^2}} dx = \sigma^2.$$

Eigenschaften normalverteilter Zufallsvariablen

- Ist X eine $N(m, \sigma^2)$ -verteilte Zufallsvariable, so ist

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

eine $N(0, 1)$ -verteilte, also standardnormalverteilte, Zufallsvariable. Man kann also die Berechnung der Wahrscheinlichkeiten $P(X \leq b)$ zurückführen auf die Berechnung entsprechender Wahrscheinlichkeiten einer standardnormalverteilten Zufallsvariablen

$$P(X \leq b) = P\left(\frac{X - m}{\sigma} \leq \frac{b - m}{\sigma}\right) = P\left(Y \leq \frac{b - m}{\sigma}\right). \quad (2.23)$$

Die Werte der Verteilungsfunktion der Standardnormalverteilung

$$\Phi(x) := P(Y \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt \text{ für } x \geq 0$$

findet man tabelliert in Formelsammlungen und in jeder guten Programmbibliothek. Mit Hilfe der Verteilungsfunktion Φ der Standardnormalverteilung berechnet man dann

$$\begin{aligned} P(a \leq X \leq b) &= P\left(\frac{a-m}{\sigma} \leq Y \leq \frac{b-m}{\sigma}\right) \\ &= \Phi\left(\frac{b-m}{\sigma}\right) - \Phi\left(\frac{a-m}{\sigma}\right). \end{aligned} \quad (2.24)$$

- Sind X_i , $i = 1 \dots n$, unabhängig normalverteilt mit Mittel m_i und Varianz σ_i^2 , so ist die Summe $S_n = X_1 + \dots + X_n$ wieder normalverteilt mit Mittel $\sum_{i=1}^n m_i$ und Varianz $\sum_{i=1}^n \sigma_i^2$.

Anwendung: Konfidenzschätzungen

Im Vorgriff auf das nächste Kapitel wollen wir im folgenden eine der wichtigsten Anwendungen der Normalverteilung in der Statistik diskutieren.

Eine Messreihe X_1, \dots, X_n unterliegt in der Regel zufälligen Mess- oder Beobachtungsfehlern. Daher können X_1, \dots, X_n auch als Zufallsvariablen angesehen werden. Als Verteilung empfiehlt sich in der Regel eine Normalverteilung $N(m, \sigma^2)$ für unbekannte m und σ^2 . Als Schätzungen für m und σ^2 wählt man naheliegenderweise das

- empirische Mittel $\bar{X} := \frac{1}{n} \sum_{i=1}^n X_i$ für m und die
- Stichprobenvarianz $s_X^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ für σ^2 .

Aussagen über **Genauigkeit** und **Sicherheit** dieser Schätzung liefern **Konfidenzschätzungen**: Von zentraler Bedeutung ist die Wahrscheinlichkeit

$$P\left(|\bar{X} - m| \leq t \cdot \frac{s_X}{\sqrt{n}}\right) \quad (2.25)$$

dafür, dass das Mittel m in einem vorgegebenen **Vertrauensbereich** (bzw. **Konfidenzintervall**) der Form

$$\left[\bar{X} - t \frac{s_X}{\sqrt{n}}, \bar{X} + t \frac{s_X}{\sqrt{n}}\right]$$

liegt. Für große Stichproben ($n \geq 30$) wird die gesuchte Wahrscheinlichkeit angenähert durch die Standardnormalverteilung

$$P\left(|\bar{X} - m| \leq t \frac{s_X}{\sqrt{n}}\right) \sim 2\Phi(t) - 1.$$

Man spricht in diesem Zusammenhang auch von einer Normalapproximation.

In der Praxis geht man von einem Vertrauensniveau γ aus (z. B. $\gamma = 95\%$) und fragt nach dem **Vertrauensbereich** für m . Zum Beispiel für $\gamma = 95\%$ ist $t = 1.96$. Mit einer Sicherheit von 95% liegt also der unbekannte Erwartungswert m im Intervall

$$\left[\bar{X} - 1.96 \frac{s_X}{\sqrt{n}}, \bar{X} + 1.96 \frac{s_X}{\sqrt{n}}\right].$$

Für $n < 30$ muss obige Wahrscheinlichkeit (2.25) mit Hilfe der t -Verteilung approximiert werden (s.u.). Man erhält z.B. für $\gamma = 95\%$ und $n = 10$ den Wert $t = 2.26$. Mit einer Sicherheit von 95% liegt der unbekannte Erwartungswert m im Intervall $\left[\bar{X} - \frac{2.26}{\sqrt{10}}s_X, \bar{X} + \frac{2.26}{\sqrt{10}}s_X\right]$.

Zum Abschluss dieses Abschnitts noch einige weitere für die induktive Statistik wichtige stetige Verteilungen in einer Übersicht.

χ^2 -Verteilung

Es seien X_1, \dots, X_n unabhängig $N(0, 1)$ -verteilte Zufallsvariablen. Dann heißt die Verteilung der Zufallsvariablen

$$Z_n = X_1^2 + \dots + X_n^2$$

χ_n^2 -Verteilung (oder χ^2 -Verteilung mit n Freiheitsgraden).

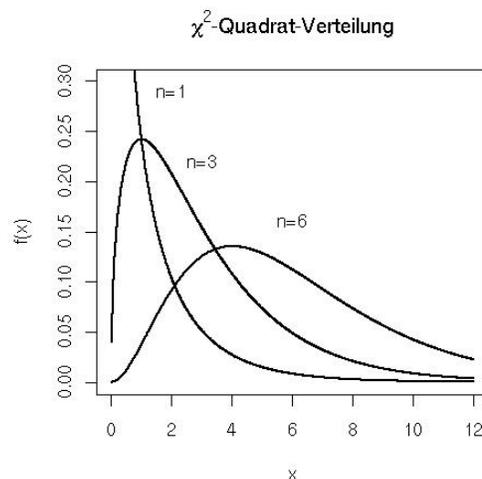
Aus den Rechenregeln für Erwartungswert und Varianz folgt sofort

$$E(Z_n) = n, \text{Var}(Z_n) = \underbrace{\text{Var}(X_1^2)}_{=2} + \dots + \underbrace{\text{Var}(X_1^2)}_{=2} = 2n.$$

Die Dichte g_n der χ_n^2 -Verteilung hat die Form

$$g_n(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} & \text{für } x > 0 \\ 0 & \text{sonst} \end{cases}$$

Für wachsendes n nähern sich die Dichten g_n der Gaußschen Glockenkurve an, weshalb man ab $n > 30$ eine Normalverteilungsapproximation wählt.



t-Verteilung

Es seien X und Z_n unabhängig, X $N(0, 1)$ -verteilt und Z_n χ_n^2 -verteilt. Dann heißt die Verteilung der Zufallsvariablen

$$T_n := \frac{X}{\sqrt{Z_n/n}}$$

t_n -Verteilung (oder t -Verteilung mit n Freiheitsgraden).

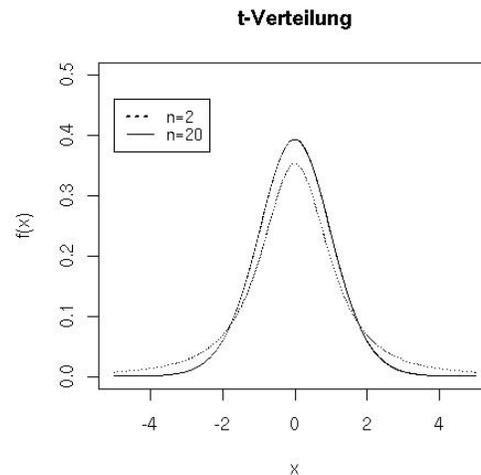
Es gilt

$$E(T_n) = 0, \text{Var}(T_n) = \frac{n}{n-2} \text{ für } n \geq 3.$$

Die Dichte h_n der t_n -Verteilung ist gegeben durch

$$h_n(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{1}{2}\right)} \frac{1}{\sqrt{n}} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}.$$

Die Dichte h_n hat eine ähnliche Form wie die Gaußsche Glockenkurve, jedoch für kleine n breitere Enden als die Standardnormalverteilung. Für $n > 30$ ist jedoch eine Approximation durch die Standardnormalverteilung bereits sehr gut.



F-Verteilung (Fisher-Verteilung)

Es seien Z_m und \tilde{Z}_n unabhängig, Z_m χ_m^2 -verteilt, \tilde{Z}_n χ_n^2 -verteilt. Dann heißt die Verteilung der Zufallsvariablen

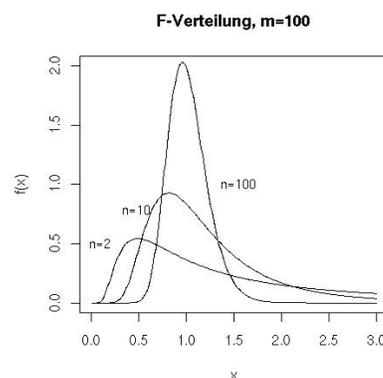
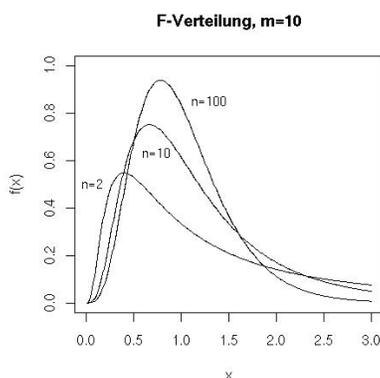
$$Z_{m,n} := (Z_m/m) / (\tilde{Z}_n/n)$$

$F_{m,n}$ -Verteilung (oder F -Verteilung mit m und n Freiheitsgraden).

Es gilt

$$E(Z_{m,n}) = \frac{n}{n-2} \text{ für } n \geq 3$$

$$\text{Var}(Z_{m,n}) = \frac{2n^2(n+m-2)}{m(n-4)(n-2)^2} \text{ für } n \geq 5.$$



5. Grenzwertsätze

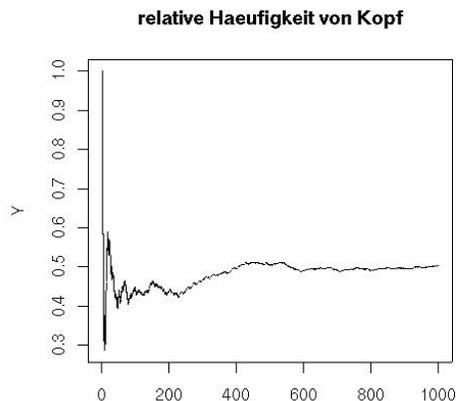
(A) Gesetz der großen Zahlen und der Hauptsatz der Statistik

Werfen wir eine faire Münze n mal und setzen wir $X_k = 1$ (bzw. $X_k = 0$) falls beim k -ten Münzwurf Kopf (bzw. Zahl) oben liegt, so nähert sich die relative Häufigkeit für Kopf

$$\frac{1}{n} \sum_{k=1}^n X_k(\omega)$$

für wachsendes n mit großer Wahrscheinlichkeit der *theoretischen* Wahrscheinlichkeit $\frac{1}{2}$ für Kopf. Man bezeichnet $\frac{1}{n} \sum_{k=1}^n X_k(\omega)$ auch als *empirisches Mittel* und $m = E(X_k) = \frac{1}{2}$ als *theoretisches Mittel*. Bei vielfacher Wiederholung des Münzwurfes stellt man fest, dass sich das empirische Mittel für wachsende n dem theoretischen Mittel annähert.

In der folgenden Grafik ist als Illustration die Folge der empirischen Mittel für insgesamt 1000 Münzwürfe aufgetragen.



Diese Beobachtung gilt ganz allgemein für die relativen Häufigkeiten eines beliebigen Ereignisses in einer unabhängigen Wiederholung ein und desselben Zufallsexperimentes. Sie wird als Gesetz der großen Zahlen bezeichnet.

Satz (Gesetz der großen Zahlen) Es sei X_1, X_2, \dots eine Folge unabhängiger Zufallsvariablen mit gemeinsamem Erwartungswert $E(X_k) = m$ und gemeinsamer Varianz $\text{Var}(X_k) = \sigma^2$. Dann folgt für alle $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P \left(\left\{ \omega : \left| \frac{1}{n} \sum_{k=1}^n X_k(\omega) - m \right| \geq \varepsilon \right\} \right) = 0.$$

Die obige Aussage zur Asymptotik der relativen Häufigkeiten eines Ereignisses A leitet sich aus dem Satz wie folgt ab: Es sei

$$X_k(\omega) := \begin{cases} 1 & \text{falls } A \text{ in der } k\text{-ten Wiederholung eintritt} \\ 0 & \text{sonst.} \end{cases}$$

Dann sind die X_1, X_2, \dots eine Folge unabhängig Bernoulli-verteilter Zufallsvariablen mit Parameter $p := P(A) = E(X_k)$. Für die relativen Häufigkeiten $f_{n,\omega}(A) := \frac{1}{n} \sum_{k=1}^n X_k(\omega)$

des Ereignisses A in n Wiederholungen gilt dann die Aussage des Gesetzes der großen Zahlen:

$$\lim_{n \rightarrow \infty} P(\{\omega : |f_{n,\omega}(A) - P(A)| \geq \varepsilon\}) = 0 \quad \forall \varepsilon > 0.$$

Exkurs: Beweis des Gesetzes der großen Zahlen

Der Beweis des Gesetzes beruht im wesentlichen auf folgender Ungleichung:

Satz (Tschebychevsche Ungleichung)

Es sei X eine Zufallsvariable für die der Erwartungswert von X^2 existiert. Dann gilt für alle $c > 0$

$$P(\{\omega : |X(\omega) - E(X)| \geq c\}) \leq \frac{1}{c^2} \text{Var}(X).$$

Beweis Wir geben den Beweis nur im Falle eines diskreten Wahrscheinlichkeitsraumes. Offenbar gilt

$$\begin{aligned} P(\{\omega : |X(\omega) - E(X)| \geq c\}) &= \sum_{\omega \in \Omega : |X(\omega) - E(X)| \geq c} p(\omega) \\ &\leq \sum_{\omega \in \Omega : |X(\omega) - E(X)| \geq c} \left(\frac{X(\omega) - E(X)}{c} \right)^2 p(\omega) \\ &\leq \sum_{\omega \in \Omega} \left(\frac{X(\omega) - E(X)}{c} \right)^2 p(\omega) = E \left(\left(\frac{X - E(X)}{c} \right)^2 \right) \\ &= \frac{1}{c^2} \text{Var}(X). \end{aligned}$$

Dabei haben wir in der zweiten Ungleichung verwendet, dass $\left(\frac{X(\omega) - E(X)}{c} \right)^2 \geq 0$ für alle $\omega \in \Omega$, und damit die Summe von $\left(\frac{X(\omega) - E(X)}{c} \right)^2 p(\omega)$ über alle $\omega \in \Omega$ nicht kleiner sein kann als die Teilsumme. \square

Beweis des Gesetzes der großen Zahlen

Es sei n fest gewählt. Wir definieren die Zufallsvariable

$$Y := \frac{1}{n} \sum_{k=1}^n X_k.$$

Da $E(X_k) = m$ für alle k , folgt aus der Linearität des Erwartungswertes

$$E(Y) = E \left(\frac{1}{n} \sum_{k=1}^n X_k \right) = \frac{1}{n} \sum_{k=1}^n E(X_k) = m.$$

Die Zufallsvariablen X_1, \dots, X_n sind nach Annahme unabhängig, also besagt die Identität von Bienaymé, dass

$$\text{Var}(Y) = \text{Var} \left(\frac{1}{n} \sum_{k=1}^n X_k \right) = \frac{1}{n^2} \sum_{k=1}^n \text{Var}(X_k) = \frac{1}{n^2} n \sigma^2 = \frac{\sigma^2}{n}.$$

Die Tschebychevsche Ungleichung, angewandt auf Y , ergibt die Abschätzung

$$P\left(\left\{\omega : \left|\frac{1}{n}\sum_{k=1}^n X_k(\omega) - m\right| \geq \varepsilon\right\}\right) \leq P(\{\omega : |Y(\omega) - E(Y)| \geq \varepsilon\}) \\ \leq \frac{\text{Var}(Y)}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2 n}.$$

Da $\frac{\sigma^2}{\varepsilon^2 n} \rightarrow 0$ für $n \rightarrow \infty$, folgt schließlich auch

$$\lim_{n \rightarrow \infty} P\left(\left\{\omega : \left|\frac{1}{n}\sum_{k=1}^n X_k(\omega) - m\right| \geq \varepsilon\right\}\right) = 0. \quad \square$$

Der Hauptsatz der Statistik

Satz Es sei X eine Zufallsvariable mit Verteilungsfunktion F und es seien X_1, X_2, \dots eine Folge von unabhängig und identisch verteilten Zufallsvariablen mit derselben Verteilungsfunktion F . Dann gilt für die Folge der zugehörigen **empirischen Verteilungsfunktionen**

$$F_{n,\omega}(x) := \frac{1}{n} \#\{i \in \{1, \dots, n\} : X_i(\omega) \leq x\} \\ = \text{relative Häufigkeit } f_{n,\omega}(A) \text{ des Ereignisses } A = \{X \leq x\},$$

der ersten n Realisierungen $X_1(\omega), \dots, X_n(\omega)$, dass

$$\lim_{n \rightarrow \infty} P\left(\left\{\omega : \sup_{x \in \mathbb{R}} |F_{n,\omega}(x) - F(x)| \geq \varepsilon\right\}\right) = 0 \quad \forall \varepsilon > 0,$$

d.h. die maximale Abweichung zwischen empirischer Verteilungsfunktion $F_{n,\omega}$ und theoretischer Verteilungsfunktion F konvergiert mit wachsendem n mit großer Wahrscheinlichkeit gegen 0.

(B) Der zentrale Grenzwertsatz

Betrachtet man in der Situation des Gesetzes der großen Zahlen mit $m := E(X_k)$ und $\sigma^2 = \text{Var}(X_k)$ die **standardisierten Summen**

$$S_n^* := \frac{S_n - E(S_n)}{\sqrt{\text{Var}(S_n)}} = \frac{\sum_{k=1}^n X_k - nm}{\sqrt{n\sigma^2}} = \sqrt{n} \frac{\bar{X}_n - m}{\sigma}$$

so stellt man fest, dass die zugehörigen Verteilungsfunktionen

$$F_{S_n^*}(x) := P(S_n^* \leq x) = P\left(\sqrt{n} \frac{\bar{X}_n - m}{\sigma} \leq x\right)$$

punktweise für alle x gegen die Verteilungsfunktion der Standardnormalverteilung konvergieren, d.h. es gilt

$$\lim_{n \rightarrow \infty} F_{S_n^*}(x) = \lim_{n \rightarrow \infty} P\left(\sqrt{n} \frac{\bar{X}_n - m}{\sigma} \leq x\right) \lim_{n \rightarrow \infty} P\left(\frac{S_n - E(S_n)}{\sqrt{\text{Var}(S_n)}} \leq x\right) \\ = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt = \Phi(x) \quad \forall x \in \mathbb{R}.$$

Man sagt auch, dass die standardisierten Summen **asymptotisch normalverteilt** sind und bezeichnet die Aussage als **zentralen Grenzwertsatz**.

Man kann dieses Resultat wiederum insbesondere auf die n -fache unabhängige Wiederholung ein und desselben Zufallsexperimentes anwenden:

Ist A ein Ereignis mit Wahrscheinlichkeit $p := P(A)$ und

$$X_k(\omega) := \begin{cases} 1 & \text{falls } A \text{ in der } k\text{-ten Wiederholung eintritt} \\ 0 & \text{sonst,} \end{cases}$$

so sind X_1, X_2, \dots unabhängig Bernoulli verteilt mit $m = E(X_k) = p$ und $\sigma^2 = \text{Var}(X_k) = p(1-p)$ und die standardisierten Häufigkeiten

$$S_n^* = \frac{S_n - np}{\sqrt{np(1-p)}} = \sqrt{n} \frac{\bar{X}_n - p}{\sqrt{p(1-p)}}$$

sind asymptotisch normalverteilt, d.h.

$$\lim_{n \rightarrow \infty} P(S_n^* \leq x) = \Phi(x) \quad \text{für alle } x \in \mathbb{R}.$$

Die Bedeutung des zentralen Grenzwertsatzes für die induktive Statistik besteht vor allem darin, dass man aufgrund der Aussage dieses Satzes die Verteilung einer standardisierten Summe S_n^* von unabhängig und identisch verteilten Zufallsvariablen (in der induktiven Statistik: die Stichprobenvariablen) mit wachsendem n (in der induktiven Statistik: mit wachsender Stichprobenlänge) zunehmend besser durch eine Standardnormalverteilung approximieren kann. Diese Approximation heißt **Normalapproximation**.

Literatur

- [1] G. Bamberg, F. Baur, M. Krapp, Statistik, 13. Auflage, R. Oldenbourg Verlag, 2007.
- [2] L. Fahrmeir, R. Künstler, I. Pigeot, G. Tutz, Statistik, 6. Auflage, Springer Verlag, 2007.

Weitere Literatur

- [3] J. Bley Müller, G. Gehlert, H. Gülicher, Statistik für Wirtschaftswissenschaftler, 14. Auflage, Verlag Vahlen, 2004.
- [4] L. Fahrmeir, R. Künstler, I. Pigeot, G. Tutz, A. Caputo, S. Lang, Arbeitsbuch Statistik, 4. Auflage, Springer Verlag, 2004.
- [5] J. Schira, Statistische Methoden der VWL und BWL: Theorie und Praxis, 2. Auflage, Pearson Studium, 2005.