

Teil F

Stochastik

Kapitel 61

Grundbegriffe

61.1 Motivation

Stochastik (aus dem Griechischen: vermuten, erwarten) ist die Mathematik des Zufalls. Sie ist von großer Bedeutung in der Informatik. Beispiele:

- Analyse der Auslastung von Datennetzen
- Modellierung von Antwortzeiten im Rechner
- Zuverlässigkeitsanalyse von Hardware
- Raytracing in der Computergrafik (Monte-Carlo-Methoden)
- stochastische Optimierungsalgorithmen (genetische Algorithmen, simulated annealing)
- Analyse der mittleren Laufzeit von Algorithmen
- Kombinatorische Probleme in der Bioinformatik

61.2 Gebietsabgrenzung

Stochastik gliedert sich in zwei Gebiete:

(a) Wahrscheinlichkeitstheorie

Ausgehend von einem stochastischen Modell werden Wahrscheinlichkeiten berechnet.

Beispiel: Modellannahme:

Beim Wurf eines Würfels habe jede der Augenzahlen $1, \dots, 6$ die Wahrscheinlichkeit $\frac{1}{6}$.

Folgerung: Wahrscheinlichkeit für Augenzahl 1 oder 3 ist $\frac{1}{6} + \frac{1}{6} = \frac{1}{3}$.

(b) Statistik

Ausgehend von realen Daten/Messungen zieht man Schlussfolgerungen.

Beispiele:

- möglichst gute Approximationsgerade durch fehlerbehaftete Messwerte legen
- Hypothesentest: Ist ein neues Medikament wirksam?

61.3 Definition (Wahrscheinlichkeit)

Die Wahrscheinlichkeit (Wahrscheinlichkeit) eines Ereignisses beschreibt die zu erwartende relative Häufigkeit, mit der dieses Ereignis eintritt, wenn man den zu Grunde liegenden Prozess immer wieder unter den gleichen Bedingungen wiederholt.

Bsp.: Bei einer „fairen“ Münze beträgt die Wahrscheinlichkeit, Zahl zu würfeln, $\frac{1}{2}$.

61.4 Definition (Laplace-Experiment)

Ein Experiment heißt Laplace-Experiment, wenn es endlich viele, einander ausschließende Ausgänge hat, die alle gleich wahrscheinlich sind.

Bsp.:

- i) Der Wurf eines Würfels hat 6 gleich berechnigte Ausgänge \Rightarrow Laplace-Experiment.
- ii) Fällt ein Marmeladenbrot zu Boden, gibt es zwei Ausgänge, die nach Murphy nicht gleich berechnigt sind \Rightarrow kein Laplace-Experiment, falls dies stimmt.

61.5 Mengentheoretische Wahrscheinlichkeitsbeschreibung

Definition: (Ereignismenge, Ereignis, Wahrscheinlichkeitsverteilung)

Wir betrachten ein Zufallsexperiment mit endlich vielen Ausgängen und definieren drei Begriffe:

- (a) Ergebnismenge Ω (Stichprobenraum, Grundraum):
endliche, nicht leere Menge, deren Elemente ω_i die Versuchsausgänge beschreiben.
Bsp.: Würfelexperiment: $\Omega = \{1, 2, \dots, 6\}$
- (b) Ereignis:
Teilmenge von Ω .
Bsp.: Ergebnisse, bei denen 2 oder 5 gewürfelt werden \Rightarrow Ereignis $A = \{2, 5\}$
- (c) Wahrscheinlichkeitsverteilung, Wahrscheinlichkeitsmaß:
Abb. P von der Potenzmenge $\wp(\Omega)$ nach \mathbb{R} mit folgenden Eigenschaften:
 - i) Normiertheit: $P(\Omega) = 1$
 - ii) Nichtnegativität: $P(A) \geq 0 \quad \forall A \in \wp(\Omega)$
 - iii) Additivität: $P(A \cup B) = P(A) + P(B)$ für alle disjunkten $A, B \in \wp(\Omega)$.

Folgerung: Der Wertebereich von P liegt in $[0, 1]$.

Kapitel 62

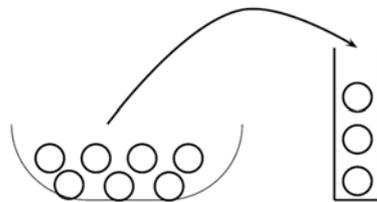
Kombinatorik

62.1 Motivation

Die Kombinatorik liefert wichtige Modelle zum Berechnen von Wahrscheinlichkeiten bei Laplace-Experimenten. Sie spielt eine grundlegende Rolle in der Informatik.

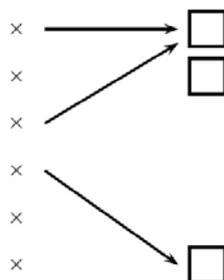
62.2 Zwei äquivalente Sprechweisen

(a) Stichprobensprechweise, Urnenmodell



Aus einer Urne mit n unterscheidbaren Kugeln werden k Kugeln gezogen. Dabei kann das Ziehen mit oder ohne Zurücklegen erfolgen, und die Reihenfolge eine oder keine Rolle spielen.

(b) Zuordnungssprechweise, Schubladenmodell



k Objekte werden auf n Schubladen verteilt. Dabei sind die Objekte entweder unterscheidbar oder nicht unterscheidbar, und die Schubladen dürfen einfach oder mehrfach besetzt werden. Urnen- und Schubladenmodell sind äquivalent:

Urnenmodell	Schubladenmodell
mit / ohne Zurücklegen	mit / ohne Mehrfachbesetzung
in / ohne Reihenfolge	unterscheidbare / ununterscheidbare Objekte

62.3 Produktregel der Kombinatorik

Bei einem k -stufigen Experiment habe der Ausgang einer Stufe keinen Einfluss auf die Anzahl der möglichen Ausgänge bei späteren Stufen. Haben die einzelnen Stufen n_1, \dots, n_k Ausgänge, so hat das Gesamtexperiment $n_1 \cdot \dots \cdot n_k$ Ausgänge.

Die Produktregel ist wichtig bei der Beschreibung der vier kombinatorischen Grundsituationen.

62.4 Die vier kombinatorischen Grundsituationen

(a) Geordnete Stichprobe mit Wiederholung

Bei k Ziehungen mit Zurücklegen aus einer Urne mit n Objekten gibt es n^k Möglichkeiten, wenn die Reihenfolge eine Rolle spielt.

Beispiel (aus einem älteren Stochastikbuch):

Herr Meier will seinen ungezogenen Sohn mit 10 Ohrfeigen bestrafen. Auf wie viele Arten kann er das tun, wenn er bei jedem Schlag zwei Möglichkeiten hat (rechts oder links)? Es gibt $2^{10} = 1024$ Möglichkeiten.

(b) Geordnete Stichprobe ohne Wiederholung

Urnenmodell: k Ziehungen aus n Objekten ohne Zurücklegen, aber mit Berücksichtigung der Reihenfolge.

$$\text{Möglichkeiten: } n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-k+1) = \frac{n!}{(n-k)!}.$$

\uparrow
 1. Ziehung

\uparrow
 k -te Ziehung

Spezialfall: $k = n$: $\Rightarrow n!$ Möglichkeiten.

Für große n approximiert man $n!$ mit der Stirling-Formel

$$n! \approx \sqrt{2\pi n} n^n e^{-n}$$

Beispiel: Herr Meier will seine 5 Kinder in einer Reihe anordnen für eine Gruppenaufnahme. Es gibt $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$ Möglichkeiten.

(c) Ungeordnete Stichprobe ohne Wiederholung

Wie in (b), jedoch müssen die $k!$ Permutationen der k gezogenen Elemente miteinander identifiziert werden. Daher gibt es nur $\frac{n!}{(n-k)!k!} = \binom{n}{k}$ Möglichkeiten.

Beispiel: Beim Lottoschein gibt es $\binom{49}{6} = \frac{49 \cdot 48 \cdot 47 \cdot 46 \cdot 45 \cdot 44}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6} = 13.983.816$ Möglichkeiten,

6 der 49 Zahlen anzukreuzen. Die Wahrscheinlichkeit, 6 Richtige zu tippen, ist daher $\frac{1}{13.983.816} \approx 7 \cdot 10^{-8}$.

(d) Ungeordnete Stichprobe mit Wiederholung

Hier ist das Schubladenmodell hilfreich: Es sollen k nicht unterscheidbare Objekte in n Schubladen verstaut werden, wobei Mehrfachbesetzung möglich ist:



Der Gesamtzustand wird beschrieben durch die Reihenfolge von k Objekten und $n - 1$ Trennungsstrichen.

$$\Rightarrow \frac{(k+n-1)!}{k!(n-1)!} = \binom{k+n-1}{k} = \binom{k+n-1}{n-1} \text{ Möglichkeiten.}$$

(Durch die Division durch $k!(n-1)!$ wurden die Permutationen identifiziert)

Beispiel: Auf wie viele Arten können 60 Parlamentssitze auf 3 Parteien verteilt werden?

$$k = 60, n = 3$$

$$\Rightarrow \binom{62}{60} = \binom{62}{2} = \frac{62 \cdot 61}{2 \cdot 1} = 1891 \text{ Möglichkeiten.}$$

Kapitel 63

Erzeugende Funktionen

63.1 Motivation

Erzeugende Funktionen wirken auf den ersten Blick etwas abstrakt, aber sie sind ein wichtiges Werkzeug, um kombinatorische Probleme systematischer und eleganter zu lösen. Sie sind zudem in verschiedenen anderen Gebieten der Stochastik nützlich.

63.2 Permutationen und Kombinationen

In §62 haben wir geordnete und ungeordnete k -elementige Stichproben einer n -elementigen Menge betrachtet. Im geordneten Fall nennt man eine solche Stichprobe auch k -Permutation, im ungeordneten Fall eine k -Kombination.

63.3 Beispiel einer erzeugenden Funktion

Nach dem Binomialsatz 19.7 gilt:

$$(x + 1)^n = \sum_{k=0}^n \binom{n}{k} x^k 1^{n-k} = \sum_{k=0}^n \binom{n}{k} x^k$$

Der Koeffizient $\binom{n}{k}$ von x^k beschreibt die Zahl der k -Kombinationen einer n -elementigen Menge ohne Wiederholung (vgl. 62.4.(c)). In den Koeffizienten der Funktion

$$f(x) = (x + 1)^n = \binom{n}{0} x^0 + \binom{n}{1} x^1 + \dots + \binom{n}{n} x^n$$

stecken somit alle Informationen über dieses kombinatorische Problem. Dies motiviert die folgende Definition:

63.4 Definition (Erzeugende Funktion)

Eine Funktion f mit

$$f(x) = \sum_{k=0}^n a_k x^k$$

heißt erzeugende Funktion für die Koeffizienten a_k . Lassen sich die Zahlen a_k mit einem kombinatorischen Problem identifizieren, so nennen wir f die erzeugende Funktion für dieses Problem.

63.5 Beispiel

$f(x) = (1+x)^n$ ist die erzeugende Funktion der Kombinationen ohne Wiederholung einer n -elementigen Menge.

Schubladeninterpretation

Jeder der n Faktoren $(1+x)$ wird als Schublade aufgefasst, in die null oder ein Element passt. Wählt man beim Ausmultiplizieren von $(1+x)$ den Faktor 1, bleibt die Schublade leer. Wählt man x , wird sie mit einem Element besetzt.

Beispielsweise beschreibt

$$f(x) = (1+x)^3 = 1 + 3x + 3x^2 + x^3$$

alle Kombinationen, 0,1,2 oder 3 Elemente auf 3 Schubladen zu verteilen. Dabei bedeuten die Summanden:

1	$= 1 \cdot 1 \cdot 1$	1 Möglichkeit bei 0 Elementen
$3x$	$= x \cdot 1 \cdot 1 + 1 \cdot x \cdot 1 + 1 \cdot 1 \cdot x$	3 Möglichkeiten bei 1 Element
$3x^2$	$= x \cdot x \cdot 1 + x \cdot 1 \cdot x + 1 \cdot x \cdot x$	3 Möglichkeiten bei 2 Elementen
x^3	$= x \cdot x \cdot x$	1 Möglichkeit bei 3 Elementen

Wie verallgemeinert man dieses Prinzip, wenn eine Schublade mit mehreren Objekten besetzt werden soll?

63.6 Beispiel

Lassen wir pro Schublade bis zu zwei Objekten zu (die sich wiederholen dürfen), lautet der Faktor $(1+x+x^2)$ statt $(1+x)$.

Z.B. können wir die Anzahl der Kombinationen mit Wdh. aus einer 2-elementigen Menge bestimmen, wenn jedes Element 0-,1- oder 2-mal ausgewählt werden kann:
 erzeugende Fkt.:

$$\begin{aligned} f(x) &= (1+x+x^2)^2 \\ &= (1+x+x^2)(1+x+x^2) \\ &= 1 \cdot 1 + 1 \cdot x + 1 \cdot x^2 + x \cdot 1 + x \cdot x + x \cdot x^2 + x^2 \cdot 1 + x^2 \cdot x + x^2 \cdot x^2 \\ &= 1 + 2x + 3x^2 + 2x^3 + x^4 \end{aligned}$$

Es gibt hier 1 Möglichkeit,	0 Objekte zu verteilen.
Es gibt hier 2 Möglichkeiten	1 Objekt zu verteilen.
Es gibt hier 3 Möglichkeiten	2 Objekte zu verteilen.
Es gibt hier 2 Möglichkeiten	3 Objekte zu verteilen.
Es gibt hier 1 Möglichkeit	4 Objekte zu verteilen.

Man kann sogar für die unterschiedlichen Elemente unterschiedliche Beschränkungen einführen:

63.7 Beispiel

Bestimme die Kombinationen einer 4-elementigen Menge $\{x_1, \dots, x_n\}$ mit den folgenden Beschränkungen:

Element:	Beschränkung	Polynom:
x_1	0-,1- oder 3-mal	$1 + x + x^3$
x_2	1- oder 2-mal	$x + x^2$
x_3	1-mal	x
x_4	0- oder 4-mal	$1 + x^4$

erzeugende Fkt.:

$$\begin{aligned} f(x) &= (1 + x + x^3)(x + x^2)x(1 + x^4) \\ &= \dots \\ &= x^2 + 2x^3 + x^4 + x^5 + 2x^6 + 2x^7 + x^8 + x^9 + x^{10} \end{aligned}$$

Es gibt also z.B. zwei 6-Kombinationen.

Diese Beispiele motivieren:

63.8 Satz: (Kombinationen mit vorgegebenen Wiederholungen)

Die erzeugende Funktion der Kombinationen mit Wiederholung aus einer n -elementigen Menge $\{x_1, \dots, x_n\}$, in der x_i in den Anzahlen $v_1^{(i)}, \dots, v_{k_i}^{(i)}$ auftreten darf, ist gegeben durch

$$\prod_{i=1}^n (x^{v_1^{(i)}} + x^{v_2^{(i)}} + \dots + x^{v_{k_i}^{(i)}}).$$

Läßt man beliebige Wiederholungen zu, erhält man mit

$$1 + x + x^2 + \dots = \sum_{k=0}^{\infty} x^k = \frac{1}{1-x} \quad (\text{für } |x| < 1)$$

den folgenden Satz:

63.9 Satz: (Kombinationen mit beliebigen Wiederholungen)

Die erzeugende Funktion der Kombinationen mit beliebigen Wiederholungen von n Elementen lautet

$$f(x) = \frac{1}{(1-x)^n} = (1 + x + x^2 + \dots)^n.$$

63.10 Bemerkungen

- Die gesuchten Koeffizienten vor x^k , $k \in \mathbb{N}_0$ ergeben sich durch eine formale Potenzreihenentwicklung. Dabei kann man Konvergenzbedingungen ignorieren, da man o.B.d.A. $|x| < 1$ annehmen darf.
- Muss jedes Element mind. p -mal auftreten, ergibt sich wegen

$$x^p + x^{p+1} + \dots = x^p \sum_{k=0}^{\infty} x^k = \frac{x^p}{1-x}$$

die erzeugende Funktion

$$f(x) = \frac{x^{np}}{(1-x)^n}.$$

Bisher hatten wir nur Kombinationen betrachtet. Sind erzeugende Funktionen auch bei Permutationen nützlich? Hierzu müssen wir den Begriff der erzeugenden Funktion durch den Begriff der exponentiell erzeugenden Funktion ersetzen:

63.11 Definition (Exponentiell erzeugende Funktion)

Eine Funktion f mit

$$f(x) = \sum_{k=0}^n a_k \frac{x^k}{k!}$$

heißt exponentiell erzeugende Funktion für die Koeffizienten a_k .

Bemerkung: Der Name wird motiviert durch die Exponentialreihe

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}.$$

63.12 Bedeutung für Permutationen

Nach 62.4.(b) gibt es bei einer n -elementigen Menge $n \cdot (n-1) \cdot \dots \cdot \frac{n!}{(n-k)!}$ k -Permutationen ohne Wiederholung ($k = 0, \dots, n$). Wegen

$$(x+1)^n = \sum_{k=0}^n \binom{n}{k} x^k = \sum_{k=0}^n \frac{n!}{(n-k)!k!} x^k = \sum_{k=0}^n \frac{n!}{(n-k)!} \frac{x^k}{k!}$$

sind dies die Koeffizienten der exponentiell erzeugenden Funktion

$$f(x) = (1+x)^n.$$

Die exponentiell erzeugende Funktion spielt also bei Permutationen die selbe Rolle wie die erzeugende Funktion bei Kombinationen.

Das Analogon zu Satz 63.8 lautet:

63.13 Satz: (Permutationen mit vorgegebenen Wiederholungen)

Die exponentiell erzeugende Funktion der Permutationen mit Wiederholung aus einer n -elementigen Menge $\{x_1, \dots, x_n\}$, in der x_i in den Anzahlen $v_1^{(i)}, \dots, v_{k_i}^{(i)}$ auftreten darf, lautet:

$$\prod_{i=1}^n \left(\frac{x^{v_1^{(i)}}}{v_1^{(i)}!} + \frac{x^{v_2^{(i)}}}{v_2^{(i)}!} + \dots + \frac{x^{v_{k_i}^{(i)}}}{v_{k_i}^{(i)}!} \right)$$

Läßt man beliebige Wiederholungen zu, folgt mit

$$1 + \frac{x}{1!} + \frac{x^2}{2!} + \dots = \sum_{k=0}^{\infty} \frac{x^k}{k!} = e^x$$

das Analogon zu Satz 63.9:

63.14 Satz: (Permutationen mit beliebigen Wiederholungen)

Die exponentiell erzeugende Funktion der Permutationen mit beliebigen Wiederholungen von n Elementen lautet

$$f(x) = e^{nx} = (e^x)^n = \left(1 + \frac{x}{1!} + \frac{x^2}{2!} + \dots\right)^n.$$

Kapitel 64

Bedingte Wahrscheinlichkeit

64.1 Motivation

In §61 haben wir den Begriff der Wahrscheinlichkeit (Wahrscheinlichkeit) kennengelernt. Oft hat man Zusatzinformationen, mit denen sich präzisere Wahrscheinlichkeitsaussagen machen lassen. Dies führt zu so genannten bedingten Wahrscheinlichkeiten.

64.2 Beispiel

Aufteilung der 1500 Angehörigen eines Betriebs nach Geschlecht und Rauchergewohnheiten:

	Frauen B	Männer \bar{B}
Raucher A	600	200
Nichtraucher \bar{A}	300	400

Ω : Menge der Betriebsangehörigen

A : Menge der Raucher ($\bar{A} = \Omega \setminus A$: Nichtraucher)

B : Menge der Frauen ($\bar{B} = \Omega \setminus B$: Männer)

Wir lösen eine Person zufällig aus. Dabei treffen folgende Wahrscheinlichkeit zu ($|A|$: Mächtigkeit von A):

$$\begin{aligned}P(A) &= \frac{|A|}{|\Omega|} = \frac{800}{1500} = \frac{8}{15} && \text{Raucheranteil} \\P(B) &= \frac{|B|}{|\Omega|} = \frac{900}{1500} = \frac{3}{5} && \text{Frauenanteil} \\P(A \cap B) &= \frac{|A \cap B|}{|\Omega|} = \frac{600}{1500} = \frac{2}{5} && \text{Anteil der rauchenden Frauen}\end{aligned}$$

Wie groß ist die Wahrscheinlichkeit $P(A | B)$, dass eine Person raucht, falls es sich um eine Frau handelt?

$$\begin{aligned}P(A | B) &= \frac{|A \cap B|}{|B|} = \frac{\frac{|A \cap B|}{|\Omega|}}{\frac{|B|}{|\Omega|}} = \frac{P(A \cap B)}{P(B)} && (*) \\&= \frac{2/5}{3/5} = \frac{2}{3}.\end{aligned}$$

Man nennt $P(A | B)$ die bedingte Wahrscheinlichkeit von A unter der Bedingung (Hypothese) B . Es gilt stets (vgl. (*))

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

Zwei Ereignisse A und B heißen unabhängig, wenn gilt:

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

64.3 Verallgemeinerung

Seien $A_1, \dots, A_n \subset \Omega$ und $P(A_1 \cap \dots \cap A_{n-1}) > 0$. Dann gilt:

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

Beweis:

Die rechte Seite läßt sich schreiben als

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

□.

64.4 Beispiel

Wie groß ist die Wahrscheinlichkeit, mit 6 Würfeln 6 verschiedene Zahlen zu würfeln?

- A_1 : irgend ein Ergebnis für 1. Würfel
- A_2 : ein vom 1. Ergebnis verschiedenes Ergebnis für Würfel 2
- \vdots
- A_6 : ein von A_1, \dots, A_5 verschiedenes Ergebnis für Würfel 6.

$$\begin{aligned} P(A_1 \cap \dots \cap A_6) &= P(A_1) \cdot P(A_2 | A_1) \cdot \dots \cdot P(A_6 | A_1 \cap \dots \cap A_5) \\ &= 1 \cdot \frac{5}{6} \cdot \dots \cdot \frac{1}{6} = \frac{6!}{6^6} \approx 0,015. \end{aligned}$$

64.5 Beispiel

Ein Krebstest ist mit 96% iger Sicherheit positiv, falls der Patient Krebs hat, mit 94%iger Sicherheit negativ, falls er keinen Krebs hat. Bei einem Patienten, in dessen Altersgruppe 0,5% aller Personen Krebs haben, verläuft der Test positiv. Wie groß ist die Wahrscheinlichkeit, dass er tatsächlich Krebs hat?

- Ereignis K : Patient hat Krebs
- Ereignis T : Test ist positiv

$$P(K | T) = \frac{P(K \cap T)}{P(T)} = \frac{0,005 \cdot 0,96}{0,005 \cdot 0,96 + 0,995 \cdot 0,06} \approx 0,074.$$

Der Patient kann also noch relativ beruhigt sein.

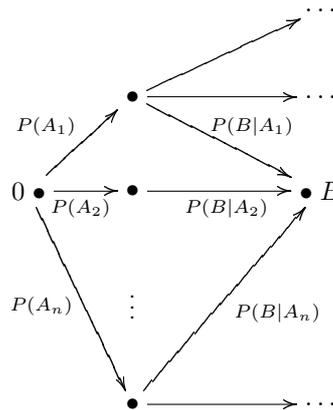
Fazit: Um eine seltene Krankheit zuverlässig zu erkennen, darf ein Test nur sehr wenige „false positives“ haben.

64.6 Satz: (Satz von der totalen Wahrscheinlichkeit)

Sei $\Omega = \bigcup_{i=1}^n A_i$ eine Partition von Ω in mögliche Ereignisse $A_i, i = 1, \dots, n$. Ferner sei $P(A_i) > 0$ für alle i . Dann gilt für jedes Ereignis B :

$$P(B) = \sum_{i=1}^n P(A_i) \cdot P(B | A_i)$$

Veranschaulichung:



Fahrer startet bei 0 und fährt mit Wahrscheinlichkeit $P(A_1), \dots, P(A_n)$ zu A_1, \dots, A_n . Die Wahrscheinlichkeit von dort nach B zu fahren, beträgt $P(B | A_1), \dots, P(B | A_n)$. Die Gesamtw., dass der Fahrer nach B gelangt, ist die Summe der Wahrscheinlichkeit aller Pfade von 0 nach B .

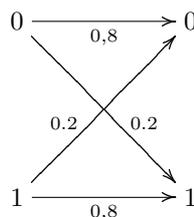
$$P(B) = \sum_{i=1}^n P(A_i)P(B | A_i). \tag{64.1}$$

Bemerkung: Im Beispiel 64.5 haben wir im Nenner bereits den Satz von der totalen Wahrscheinlichkeit verwendet.

64.7 Beispiel

Um einen binären Nachrichtenkanal robuster gegenüber Störungen zu machen, sendet man die Bitfolge 0000000 statt 0 und 1111111 statt 1. Störungen treten in 20% aller Fälle auf, und die Wahrscheinlichkeit für 0000000 sei 0,1.

Es wird 0100110 empfangen. Wie groß ist die Wahrscheinlichkeit, dass 0000000 gesendet wurde?



$$P(0000000 | 0100110) = \frac{0,1 \cdot 0,2^3 \cdot 0,8^4}{0,1 \cdot 0,2^3 \cdot 0,8^4 + 0,9 \cdot 0,2^4 \cdot 0,8^3} \approx 0,308$$

Man wird den Block also als 1 lesen, obwohl die Mehrzahl der Bits Nullen sind!

Eng verwandt mit dem Satz von der totalen Wahrscheinlichkeit ist

64.8 Satz: (Formel von Bayes)

Sei $P(B) > 0$ und seien die Voraussetzungen von Satz 64.6 erfüllt. Dann gilt:

$$P(A_k | B) = \frac{P(A_k)P(B | A_k)}{\sum_{i=1}^n P(A_i)P(B | A_i)}$$

Beweis:

Folgt aus $P(A_k | B) = \frac{P(A_k \cap B)}{P(B)}$, indem man

$$P(B) = \sum_{i=1}^n P(A_i)P(B | A_i) \quad (64.6)$$

$$P(A_k \cap B) = P(A_k) \cdot P(B|A_k) \quad (64.2)$$

einsetzt.

□.

Bemerkung: In Satz 64.8 wird $P(A_k | B)$ aus $P(B | A_i)$, $i = 1, \dots, n$ berechnet, d.h. Ursache und Wirkung kehren sich um. Eine typische Anwendung besteht darin, dass man eine Wirkung misst und nach der wahrscheinlichsten Ursache fragt (inverses Problem).

64.9 Anwendungsbeispiele

- Ein Arzt beobachtet bei einem Patienten ein Symptom B . Es kann von n verschiedenen Krankheiten A_k , $k = 1, \dots, n$ herrühren. Um die wahrscheinlichste Ursache zu finden, muss man also $P(A_k | B)$ abschätzen.
- Aus einem verrauschten Bild will man das wahrscheinlichste unverrauschte Bild rekonstruieren (vgl. auch Bsp. 64.7).
- In der Computertomographie schickt man Röntgenstrahlung in verschiedenen Richtungen durch den Patienten und mißt die durchgedrungene Intensität. Aus diesen Auswirkungen versucht man, Rückschlüsse auf die Ursache (Gewebe, Knochen, Tumor, ...) zu ziehen.

Kapitel 65

Zufallsvariable, Erwartungswert, Varianz

65.1 Motivation

Oft möchte man dem Resultat eines Zufallsexperiments eine reelle Zahl zuordnen. Der Gewinn bei einem Glücksspiel ist ein Beispiel hierfür. In diesem Fall interessiert man sich auch für den zu erwartenden Gewinn und für ein Maß für die statistischen Schwankungen. Dies führt uns auf Begriffe wie Zufallsvariable, Erwartungswert und Varianz. In der Informatik werden sie u.a. bei der Zuverlässigkeitsanalyse von Systemen benötigt.

65.2 Definition (Zufallsvariable)

Sei Ω ein Stichprobenraum. Eine Funktion X , die jedem Ergebnis $\omega \in \Omega$ eine reelle Zahl $X(\omega)$ zuordnet, heißt Zufallsvariable.

Bemerkung: Eine Zufallsvariable ist also weder zufällig noch eine Variable, sondern eine Funktion. Man kann sie stets als Gewinn bei einem Glücksspiel interpretieren.

65.3 Beispiel

Eine faire Münze mit Seiten 0 und 1 werde 3 Mal geworfen. Die Anzahl der Einsen sei der Gewinn. Man kann $\Omega = \{000, 001, \dots, 111\}$ als Stichprobenraum und den Gewinn als Zufallsvariable $X(\omega)$ auffassen.

Ergebnis ω	000	001	010	011	100	101	110	111
Gewinn $X(\omega)$	0	1	1	2	1	2	2	3
Ws. $P(\omega)$	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8

Verschiedene Ereignisse ω_1, ω_2 können zum selben Gewinn führen.

65.4 Definition (Verteilung)

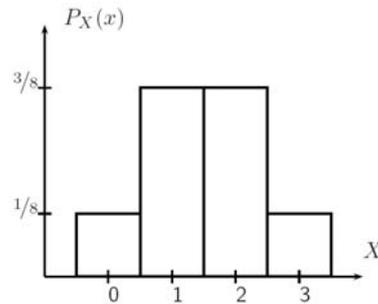
Die Verteilung P_X einer Zufallsvariablen X ordnet jedem Wert $x \in X$ eine Wahrscheinlichkeit $P_X(x)$ zu.

65.5 Beispiel

In Beispiel 65.3 können wir dem Gewinn folgende Wahrscheinlichkeiten zuordnen:

Gewinn X	0	1	2	3
Ws. $P_X(x)$	$1/8$	$3/8$	$3/8$	$1/8$

Oft veranschaulicht man die Verteilung einer Zufallsvariablen durch ein Histogramm:



Da wir hier meistens diskrete Zufallsvariablen betrachten, sind die Verteilungen ebenfalls diskret.

Interessiert man sich für den Durchschnittsgewinn je Versuchswiederholung, gelangt man zum Begriff des Erwartungswertes:

65.6 Definition (Erwartungswert)

Unter dem Erwartungswert $E(X)$ einer (diskreten) Zufallsvariablen X versteht man das gewichtete Mittel der Funktion X über Ω , wobei jeder Wert mit seiner Wahrscheinlichkeit gewichtet wird:

$$E(X) = \sum_{\omega \in \Omega} X(\omega)P(\omega)$$

65.7 Beispiel

Für die Zufallsvariable X aus 65.3 erhält man als Erwartungswert

$$\begin{aligned} E(X) &= 0 \cdot \frac{1}{8} + 1 \cdot \frac{1}{8} + 1 \cdot \frac{1}{8} + 2 \cdot \frac{1}{8} + 1 \cdot \frac{1}{8} + 2 \cdot \frac{1}{8} + 2 \cdot \frac{1}{8} + 2 \cdot \frac{1}{8} + 3 \cdot \frac{1}{8} \\ &= \frac{12}{8} = \frac{3}{2}. \end{aligned}$$

Bequemer hätte man den Erwartungswert mit Hilfe der Verteilung P_X berechnet (vgl. Tabelle aus Bsp. 65.5):

$$E(X) = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{3}{2}.$$

Es gilt also auch:

$$E(X) = \sum_{x \in X(\Omega)} x P_X(x)$$

Bemerkung: Bei kontinuierlichen Zufallsvariablen verwendet man Integrale statt Summen.

Mit dem Erwartungswert kann man gut arbeiten, denn es gilt:

65.8 Satz: (Linearität des Erwartungswerts)

Seien X, Y Zufallsvariablen und $\alpha, \beta \in \mathbb{R}$. Dann gilt:

$$E(\alpha X + \beta Y) = \alpha E(X) + \beta E(Y).$$

Beweis:

$$\begin{aligned} E(\alpha X + \beta Y) &= \sum_{\omega \in \Omega} (\alpha X(\omega) + \beta Y(\omega)) P(\omega) \\ &= \alpha \sum_{\omega \in \Omega} X(\omega) P(\omega) + \beta \sum_{\omega \in \Omega} Y(\omega) P(\omega) \\ &= \alpha E(X) + \beta E(Y) \end{aligned}$$

□.

65.9 Unabhängigkeit zweier Zufallsvariablen

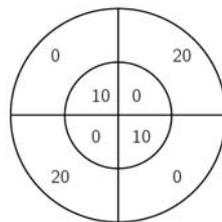
Mit 65.8 wissen wir, dass gilt: $E(X + Y) = E(X) + E(Y)$.

Gilt auch $E(XY) = E(X)E(Y)$?

Man kann zeigen, dass dies dann gilt, wenn X und Y unabhängig sind, d.h. (vgl. 64.2):

$$P((X = a) \cap (Y = b)) = P(X = a) \cdot P(Y = b) \quad \forall a, b.$$

65.10 Beispiele

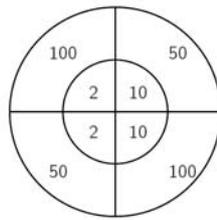


- (a) Glücksrad mit 4 Ergebnissen, auf denen die Zufallsvariable X (äußerer Ring) und Y (innerer Ring) definiert sind.

$$\left. \begin{aligned} E(X) &= \frac{1}{4} \cdot 20 + \frac{1}{4} \cdot 0 + \frac{1}{4} \cdot 20 + \frac{1}{4} \cdot 0 = 10 \\ E(Y) &= \frac{1}{4} \cdot 0 + \frac{1}{4} \cdot 10 + \frac{1}{4} \cdot 0 + \frac{1}{4} \cdot 10 = 5 \end{aligned} \right\} E(X)E(Y) = 50$$

aber: $E(XY) = \frac{1}{4} \cdot 20 \cdot 0 + \frac{1}{4} \cdot 0 \cdot 10 + \frac{1}{4} \cdot 20 \cdot 0 + \frac{1}{4} \cdot 0 \cdot 10 = 0 \neq E(X)E(Y)$.

X und Y sind nicht unabhängig: Das Ereignis $Y = 0$ hat die Ws. $\frac{1}{2}$. Weiß man jedoch, dass $X = 20$ eingetreten ist, dann hat $Y = 0$ die Ws. 1.



b)

$$\left. \begin{aligned} E(X) &= \frac{1}{4} \cdot 50 + \frac{1}{4} \cdot 100 + \frac{1}{4} \cdot 50 + \frac{1}{4} \cdot 100 = 75 \\ E(Y) &= \frac{1}{4} \cdot 10 + \frac{1}{4} \cdot 10 + \frac{1}{4} \cdot 2 + \frac{1}{4} \cdot 2 = 6 \end{aligned} \right\} \Rightarrow E(X)E(Y) = 450$$

$$E(XY) = \frac{1}{4} \cdot 50 \cdot 10 + \frac{1}{4} \cdot 100 \cdot 10 + \frac{1}{4} \cdot 50 \cdot 2 + \frac{1}{4} \cdot 100 \cdot 2 = 450 = E(X)E(Y).$$

X und Y sind unabhängig:

$Y = 2$ und $Y = 10$ sind gleich wahrscheinlich. Weiß man, z.B., dass $X = 50$ eingetreten ist, so sind $Y = 2$ und $Y = 10$ noch stets gleich wahrscheinlich. Oft ist man nicht nur am Erwartungswert interessiert. Man möchte auch wissen, wie stark die Verteilung um den Erwartungswert streut. Hierzu dienen die Begriffe Varianz und Standardabweichung:

65.11 Definition (Varianz)

Sei X eine Zufallsvariable mit Erwartungswert $\mu := E(X)$. Dann versteht man unter der Varianz $V(X) = \sigma^2$ den Erwartungswert von $(X - \mu)^2$:

$$\sigma^2 = V(X) := E((X - \mu)^2)$$

$\sigma := \sqrt{V(X)}$ nennt man die Standardabweichung (Streuung) von X .

65.12 Berechnung der Varianz

Wegen der Linearität des Erwartungswerts und wegen $E(\text{const.}) = \text{const.}$ gilt:

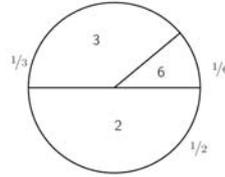
$$\begin{aligned} E((X - \mu)^2) &= E(X^2 - 2\mu X + \mu^2) \\ &= E(X^2) - 2\mu \underbrace{E(X)}_{\mu} + \mu^2 \\ &= E(X^2) - \mu^2. \end{aligned}$$

Wir erhalten somit eine wichtige Formel zur Berechnung der Varianz:

$$\boxed{\sigma^2 = E(x^2) - \mu^2} \quad (\text{Verschiebungssatz})$$

65.13 Beispiel

Sei X der Gewinn auf dem Glücksrad



$$\begin{aligned} \Rightarrow \mu &= \frac{1}{2} \cdot 2 + \frac{1}{3} \cdot 3 + \frac{1}{6} \cdot 6 = 3 && \text{mittlerer Gewinn} \\ E(X^2) &= \frac{1}{2} \cdot 4 + \frac{1}{3} \cdot 9 + \frac{1}{6} \cdot 36 = 11 \\ \sigma^2 &= E(X^2) - \mu^2 = 11 - 3^2 = 2 && \text{Varianz} \\ \sigma &= \sqrt{2} && \text{Standardabweichung} \end{aligned}$$

65.14 Satz: (Eigenschaften der Varianz)

Sei X eine Zufallsvariable und seien $\alpha, \beta \in \mathbb{R}$. Dann gilt:

i) $V(\alpha X) = \alpha^2 V(X)$

ii) $V(X + \beta) = V(X)$

Beweis:

i) $V(\alpha X) \stackrel{65.8}{=} E((\alpha X - \alpha\mu)^2) = E(\alpha^2(X - \mu)^2) \stackrel{65.8}{=} \alpha^2 V(X)$

ii) $V(X + \beta) = E((X + \beta - \mu - \beta)^2) = E((X - \mu)^2) = V(X)$

□.

65.15 Standardisierte Zufallsvariable

Ist X eine Zufallsvariable mit Erwartungswert μ und Varianz σ^2 , so nennt man

$$X^* := \frac{X - \mu}{\sigma}$$

die Standardisierte von X . Es gilt:

$$E(X^*) = \frac{1}{\sigma} \underbrace{E(X)}_{\mu} - \frac{\mu}{\sigma} = \frac{\mu}{\sigma} - \frac{\mu}{\sigma} = 0$$

$$V(X^*) \stackrel{65.14}{=} \frac{1}{\sigma^2} \underbrace{V(X)}_{\sigma^2} = \frac{\sigma^2}{\sigma^2} = 1.$$

Eine solche Standardisierung ist nützlich, wenn man die Verteilung einer Zufallsvariablen mit einer tabellierten Verteilung vergleichen möchte, da letzterer oft nur in standardisierter Form vorliegt.

Wir wollen nun weitere wichtige Eigenschaften des Erwartungswerts studieren. Aus dem Verschiebungssatz 65.12 folgt wegen $\sigma^2 \geq 0$:

$$E(X^2) - \underbrace{(E(X))^2}_{\mu} \geq 0 \quad \text{d.h. } E(X^2) \geq (E(X))^2.$$

Dies ist der Spezialfall eines allgemeineren Resultats:

65.16 Satz: (Ungleichung von Jensen)

Sei $r : \mathbb{R} \rightarrow \mathbb{R}$ eine konvexe (!) Funktion und X eine Zufallsvariable. Dann gilt:

$$E(r(X)) \geq r(E(X))$$

Beweis:

Sei X zunächst eine diskrete Zufallsvariable.

Dann fassen wir $E(X) = \sum_{\omega \in \Omega} X(\omega)P(\omega)$ als Konvexkombination der $X(\omega)$, $\omega \in \Omega$ mit Gewichten $P(\omega)$ auf (d.h. $P(\omega) \geq 0 \forall \omega \in \Omega$ und $\sum_{\omega \in \Omega} P(\omega) = 1$). Aus der Konvexität von r folgt mit 25.5:

$$\begin{aligned} E(r(X)) &= \sum_{\omega \in \Omega} r(X(\omega))P(\omega) \\ &\geq r\left(\sum_{\omega \in \Omega} X(\omega)P(\omega)\right) = r(E(X)). \end{aligned}$$

Ist X eine kontinuierliche Zufallsvariable, ersetzt man Summen durch Integrale. □.

65.17 Beispiele

(a) Die Funktion $r(t) = t^2$ ist konvex, da $r''(t) = 2 > 0$. Daher gilt:

$$E(X^2) \geq (E(X))^2.$$

(b) Sei $\Theta > 0$. Dann ist $r(t) = e^{\Theta t}$ konvex, und es gilt:

$$E(e^{\Theta X}) \geq e^{\Theta E(X)}.$$

Die Funktion $E(e^{\Theta X})$ wird später eine wichtige Rolle spielen.

65.18 Gleichung von Wald

Seien X_1, X_2, \dots unabhängige Zufallsvariablen. Manchmal interessiert man sich für die erste Zeit (Stopzeit) $N = n$, in der die Summe $X_1 + \dots + X_N$ einen vorgegebenen Wert y übersteigt, d.h.

$\sum_{i=1}^{n-1} X_i \leq y$ und $\sum_{i=1}^n i = y$. Dann ist N selbst eine Zufallsvariable. Für ihren Erwartungswert kann man zeigen:

Satz: (Gleichung von Wald)

Seien X_1, X_2, \dots unabhängige, gleich verteilte Zufallsvariablen mit endlichem Erwartungswert, und sei N eine Stopzeit für X_1, X_2, \dots . Ferner sei $S_N = \sum_{i=1}^N X_i$. Dann gilt:

$$E(S_N) = E(X) \cdot E(N).$$

65.19 Beispiel

Sei $X_i = 1$, falls beim i -ten Münzwurf Kopf eintritt, ansonsten 0. Wie ist der Erwartungswert $E(X)$ für die Stopzeit $N := \min\{n \mid X_1 + \dots + X_n \geq 10\}$?

$$\begin{aligned} \text{Es gilt: } E(X) &= \frac{1}{2} \quad \text{und} \quad E(S_N) = 10. \\ \Rightarrow E(N) &= 20. \end{aligned}$$

Eng verwandt mit dem Begriff der Varianz ist die Kovarianz. Sie ist ein Maß für die Korreliertheit zweier Zufallsvariablen.

65.20 Definition (Kovarianz, Korrelationskoeffizient)

Seien X, Y Zufallsvariablen mit Erwartungswert μ_X, μ_Y und Varianz σ_X^2, σ_Y^2 . Dann heißt

$$\text{Cov}(X, Y) := \sigma_{XY}^2 := E((X - \mu_X)(Y - \mu_Y))$$

die Kovarianz von X und Y , und

$$\rho_{XY} := \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

der Korrelationskoeffizient von X und Y .

Ist $\text{Cov}(X, Y) = 0$, so heißen X und Y unkorreliert.

65.21 Bemerkungen:

a) $V(X) = \text{Cov}(X, X)$

b) Sind $X - \mu_X, Y - \mu_Y$ Funktionen auf einem endlichen Stichprobenraum $\Omega = \{\omega_1, \dots, \omega_n\}$, so kann man sie auch als Vektoren im \mathbb{R}^n mit der i -ten Komponente $X(\omega_i) - \mu_X, Y(\omega_i) - \mu_Y$ ansehen. Dann bezeichnen (vgl. §41, §43):

- die Standardabweichungen die induzierten euklidischen Normen
- die Varianzen die quadrierten euklidischen Normen
- die Kovarianz das euklidische Skalarprodukt
- der Korrelationskoeffizient den Arcuscossinus des Winkels zwischen beiden Vektoren. Insbesondere ist $-1 \leq \rho_{XY} \leq 1$.
- die Unkorreliertheit die Orthogonalität

wenn wir das gewichtete euklidische Skalarprodukt $\langle u, v \rangle := \sum_{i=1}^n p_i u_i v_i$ zu Grunde legen. Dabei ist $p_i := P(\omega_i)$.

Man kann Folgendes zeigen:

65.22 Satz: (Rechenregeln für die Korrelation)

Seien X, Y Zufallsvariablen mit Erwartungswert μ_X, μ_Y . Dann gilt:

- a) $\text{Cov}(X, Y) = E(X, Y) - \mu_X \mu_Y$ (vgl. 65.12).
- b) $\text{Cov}(\alpha X + \beta, \gamma Y + \delta) = \alpha \gamma \text{Cov}(X, Y)$ (vgl. 65.14).
- c) $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
- d) Für m Zufallsvariablen X_1, \dots, X_m gilt:

$$V(X_1 + \dots + X_m) = \sum_{i=1}^m V(X_i) + \sum_{i \neq j} \text{Cov}(X_i, X_j).$$

- e) Sind X, Y unabhängig, so sind sie auch unkorreliert.
- f) Für paarweise unkorrelierte X_1, \dots, X_n gilt:

$$V(X_1 + \dots + X_n) = \sum_{i=1}^n V(X_i).$$

65.23 Bemerkungen

- (a) Die Umkehrung von 65.22.(e) gilt nicht! Beispiel:

Ergebnis ω	1	2	3	4
Zufallsvar. X	1	-1	2	-2
Zufallsvar. Y	-1	1	2	-2
Wahrscheinlichkeit $P(\omega)$	$2/5$	$2/5$	$1/10$	$1/10$

Dann ist $E(X) = 0 = E(Y)$ und

$$\text{Cov}(X, Y) = -1 \cdot \frac{2}{5} - 1 \cdot \frac{2}{5} + 4 \cdot \frac{1}{10} + 4 \cdot \frac{1}{10} = 0.$$

Aber: X und Y sind nicht unabhängig, denn $X(\omega)$ bestimmt ω und $Y(\omega)$ eindeutig.

- (b) In der Informatik tauchen Erwartungswerte und Varianzen z.B. bei der Zuverlässigkeitsanalyse von Systemen auf oder bei der Abschätzung von Wartezeiten bei Internetanfragen. Kovarianzen sind u.A. wichtig im Bereich des maschinellen Lernens.

65.24 Zuverlässigkeitsanalyse von Systemen

Ein System bestehe aus n Komponenten. Der Zustand der k -ten Komponente wird durch die Zufallsvariable (Indikator)

$$I_k := \begin{cases} 1 & (k\text{-te Komponente funktioniert}) \\ 0 & (k\text{-te Komponente funktioniert nicht}) \end{cases}$$

beschrieben. Ihr Erwartungswert beschreibt die Zuverlässigkeit der Komponente k . Wir setzen:

$$p_k := E(I_k), \quad q_k := 1 - p_k \quad (\text{Ausfallws.})$$

Interessiert man sich für die Zuverlässigkeit p des Gesamtsystems, muss man verschiedene Fälle unterscheiden:

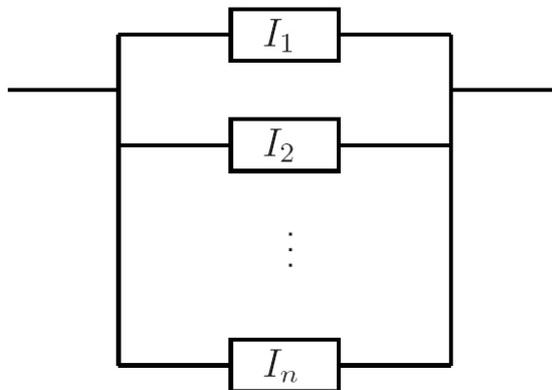
(a) Reihensysteme



Ein Reihensystem arbeitet, wenn alle Komponenten arbeiten: $p = p_1 \cdot \dots \cdot p_n$.

(b) Parallelsysteme

Ein Parallelsystem fällt aus, wenn alle Komponenten ausfallen:

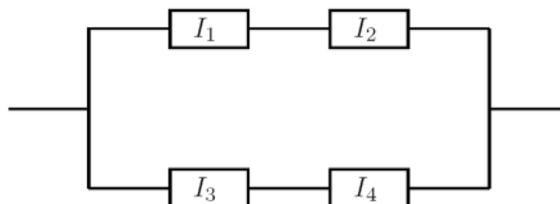


$$q = q_1 \cdot \dots \cdot q_n$$

$$\Rightarrow p = 1 - q = 1 - (1 - p_1) \cdot \dots \cdot (1 - p_n).$$

(c) Gemischte Systeme

werden hierarchisch in Reihensysteme und Parallelsystem zerlegt:



oberes Reihensystem: $p_1 \cdot p_2$

unteres Reihensystem: $p_3 \cdot p_4$

paralleles Gesamtsystem: $p = 1 - (1 - p_1 p_2)(1 - p_3 p_4)$.

Kapitel 66

Abschätzungen für Abweichungen vom Erwartungswert

66.1 Motivation

Mit der Varianz bzw. Standardabweichung kennen wir bereits ein Maß für die Fluktuation einer Zufallsvariablen um ihren Erwartungswert.

Gibt es weitere nützliche Maßzahlen hierfür?

Ist es möglich, Abschätzungen zu finden, die wahrscheinlichkeitstheoretische Aussagen darüber liefern, wie häufig eine Zufallsvariable außerhalb eines Intervalls um den Erwartungswert liegt?

66.2 Definition (k -tes (zentrales) Moment)

Sei X eine Zufallsvariable mit Erwartungswert $\mu = E(X)$. Dann bezeichnen wir mit $E(X^k)$, $k \in \mathbb{N}$ das k -te Moment von X . Ferner definiert $E((X - \mu)^k)$ das k -te zentrale Moment von X .

66.3 Bemerkungen

- Der Erwartungswert $\mu = E(X)$ ist das erste Moment.
- Die Varianz $\sigma^2 = E((X - \mu)^2)$ ist das zweite zentrale Moment.
- Mit den 3. und 4. zentralen Momenten lassen sich Aussagen über die Schiefe bzw. Flachheit der Verteilung einer Zufallsvariablen gewinnen.
- Höhere Momente sind anschaulich schwieriger zu interpretieren, liefern jedoch ebenfalls wichtige Aussagen.
- Momente haben eine große Bedeutung in der Mustererkennung bei der Analyse von Texturen und der Erkennung von Objekten unter Rotationen und Skalierungen.

Ähnlich wie wir die Menge aller k -Permutationen und k -Kombinationen einer n -elementigen Menge durch den Begriff der erzeugenden Funktion kompakt beschreiben konnten, gibt es eine Funktion, die sämtliche Momente einer Zufallsvariablen beinhaltet:

66.4 Definition (Momenten erzeugende Funktion)

Sei X eine Zufallsvariable. Falls $M_X(\Theta) := E(e^{\Theta X})$ existiert, nennen wir $M_X(\Theta)$ die Momenten erzeugende Funktion von X .

66.5 Satz: (Eigenschaften Momenten erzeugender Funktionen)

Die Momenten erzeugende Funktion $M_X(\Theta) = E(e^{\Theta X})$ einer Zufallsvariablen X hat folgende Eigenschaften:

a) Die n -te Ableitung in $\Theta = 0$ liefert das n -te Moment:

$$M_X^{(n)}(0) = E(X^n).$$

b) Skalierungsverhalten:

Sei $Y = aX + b$. Dann ist $M_Y(\Theta) = e^{b\Theta} M_X(a\Theta)$.

c) $M_{X+Y}(\Theta) = M_X(\Theta)M_Y(\Theta)$, falls X und Y unabhängige Zufallsvariablen sind.

Beweis: Wir zeigen nur (a):

Mit der Potenzreihenentwicklung von \exp und der Linearität des Erwartungswerts gilt:

$$\begin{aligned} M_X(\Theta) &= E(e^{\Theta X}) \\ &= E\left(\sum_{k=0}^{\infty} \frac{\Theta^k x^k}{k!}\right) \\ &= \sum_{k=0}^{\infty} \frac{\Theta^k}{k!} E(X^k) \end{aligned}$$

Gliedweise Differentiation liefert:

$$\begin{aligned} M_X'(\Theta) &= \sum_{k=1}^{\infty} \frac{\Theta^{k-1}}{(k-1)!} E(X^k) = \sum_{k=0}^{\infty} \frac{\Theta^k}{k!} E(X^{k+1}) \\ &\vdots \\ M_X^{(n)}(\Theta) &= \sum_{k=0}^{\infty} \frac{\Theta^k}{k!} E(X^{k+n}) \end{aligned}$$

und somit

$$M_X^{(n)}(0) = E(X^n).$$

□.Wir kommen nun zu den Abschätzungen für Fluktuationen jenseits eines vorgegebenen Abstands um den Erwartungswert einer Zufallsvariablen. Grundlegend ist der folgende Satz:

66.6 Satz: (Markow'sche Ungleichung)

Sei X eine Zufallsvariable und $h : \mathbb{R} \rightarrow \mathbb{R}$ eine nichtnegative und nichtfallende Funktion mit $h(t) > 0$. Dann gilt:

$$P(X \geq t) \leq \frac{E(h(X))}{h(t)}$$

Beweis (für eine diskrete Zufallsvariable)

Nach Def. 65.7 gilt:

$$E(h(X)) = \sum_{z \in X(\Omega)} h(z)P_X(z)$$

und wegen

$$\sum_{z \in X(\Omega)} \underbrace{h(z)}_{>0} \underbrace{P_X(z)}_{\geq 0} \geq \sum_{\substack{z \in X(\Omega) \\ z \geq t}} h(z)P_X(z) \stackrel{\text{Nicht-fallend}}{\geq} h(t) \underbrace{\sum_{\substack{z \in X(\Omega) \\ z \geq t}} P_X(z)}_{P(X \geq t)}$$

folgt

$$P(X \geq t) \leq \frac{1}{h(t)} E(h(X)).$$

□.

66.7 Folgerungen

- a) Setzt man $h(x) = \begin{cases} x & (x > 0) \\ 0 & (\text{sonst}) \end{cases}$, folgt die einfache Markow-Ungleichung

$$\boxed{P(X \geq t) \leq \frac{E(X)}{t}} \quad (t > 0)$$

- b) Mit $Y := (X - E(X))^2$ und $h(x) = x$ für $x > 0$ kann man die Tschebyschew-Ungleichung für den Erwartungswert μ und die Standardabweichung σ von X beweisen:

$$\boxed{P(|X - \mu| \geq t) \leq \frac{\sigma^2}{t^2}}$$

Alternativschreibweise:

$$P(|X - \mu| \geq c \cdot \sigma) \leq \frac{1}{c^2}$$

- c) Mit $h(X) = e^{\Theta X}$ für $\Theta \geq 0$ ergibt sich:

$$P(X \geq t) \leq e^{-\Theta t} M_X(\Theta)$$

Dies führt zur Chernoff-Schranke

$$P(X \geq t) \leq \inf_{\Theta \geq 0} e^{-\Theta t} M_X(\Theta)$$

Bemerkung: Für die einfache Markow-Ungleichung benötigt man das erste Moment (Erwartungswert μ), für die Tschebyschew-Ungleichung die ersten beiden Momente μ, σ^2 , und für die Chernoff-Ungleichung alle Momente (Momenten erzeugende Funktion). Je mehr Momente man kennt, desto mehr weiß man über die Verteilung von X und desto schärfere Abschätzungen kann man erwarten.

66.8 Beispiel

Eine faire Münze werde n Mal geworfen. Tritt beim k -ten Wurf Kopf auf, setzt man $Y_k := 1$, sonst $Y_k := 0$. Wir interessieren uns für die „Kopfhäufigkeit“ nach n Würfeln:

$$X_n := Y_1 + \dots + Y_n$$

Y_1, \dots, Y_n sind unabhängige Zufallsvariablen. Es gilt:

$$\begin{aligned}\mu &= E(X_n) = \frac{n}{2} \\ \sigma^2 &= E((X_n - \mu)^2) = \sum_{k=1}^n \left[\left(0 - \frac{1}{2}\right)^2 \cdot \frac{1}{2} + \left(1 - \frac{1}{2}\right)^2 \cdot \frac{1}{2} \right] = \frac{n}{4}\end{aligned}$$

und wegen

$$M_{Y_k}(\Theta) = E(e^{\Theta Y_k}) = e^{\Theta \cdot 0} \frac{1}{2} + e^{\Theta \cdot 1} \frac{1}{2} = \frac{1 + e^{\Theta}}{2}$$

folgt mit 66.5:

$$M_{X_n}(\Theta) = M_{Y_1 + \dots + Y_n} = \left(\frac{1 + e^{\Theta}}{2} \right)^n$$

Sei $\alpha = 0.8$. Wie groß ist die Wahrscheinlichkeit, nach $n=100$ Würfeln $X_n \geq \alpha \cdot n = 80$ zu erhalten? Wir vergleichen die 3 Ungleichungen aus 66.7:

- Einfache Markow-Ungleichung

Mit $\mu = 50$ und $t = 80$ ergibt sich

$$P(X_{100} \geq 80) \leq \frac{\mu}{t} = \frac{50}{80} = 0,625.$$

- Tschebyschew-Ungleichung

Mit $\mu = 50, t = 30$ und $\sigma^2 = \frac{100}{4} = 25$ ergibt sich

$$\begin{aligned}P(X_{100} \geq 80) &\leq P(|X_{100} - 50| \geq 30) \\ &\leq \frac{\sigma^2}{t^2} = \frac{25}{30^2} \approx 0,028.\end{aligned}$$

Obwohl Tschebyschew Abweichungen nach beiden Seiten berücksichtigt, ist die Abschätzung schärfer als bei der einfachen Markow-Ungleichung.

- Chernoff-Schranke

$$P(X_{100} \geq 80) \leq \inf_{\Theta \leq 0} \underbrace{e^{-80 \cdot \Theta} \left(\frac{1 + e^{\Theta}}{2} \right)^{100}}_{=: f(\Theta)}$$

Durch Ableiten zeigt man, dass $f(\Theta)$ minimiert wird für $\Theta = \ln 4$. Damit folgt

$$\begin{aligned}P(X_{100} \geq 80) &\leq 4^{-80} \cdot \left(\frac{1 + 4}{2} \right)^{100} \\ &\approx 4,26 \cdot 10^{-9}.\end{aligned}$$

Erwartungsgemäß ist dies eine wesentlich schärfere Schranke als (a) und (b).

66.9 Schwaches Gesetz der großen Zahlen

Mit Hilfe der Tschebyschew-Ungleichung kann man das schwache Gesetz der großen Zahlen beweisen:

Es werde ein Versuch n Mal wiederholt, bei dem das Ereignis A mit Ws. p eintritt. Dann strebt die Ws., dass sich die relative Häufigkeit $h_n(A)$ um weniger als ε von p unterscheidet, gegen 1 für $n \rightarrow \infty$:

$$\lim_{n \rightarrow \infty} P(|h_n(A) - p| < \varepsilon) = 1.$$

Dabei ist ε eine beliebig kleine positive Zahl.

Kapitel 67

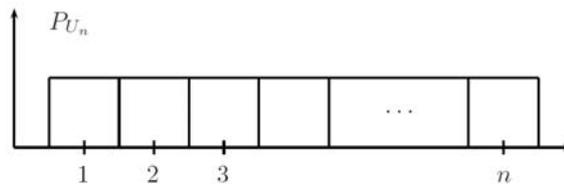
Wichtige diskrete Verteilungen

67.1 Motivation

Einige diskrete Verteilungen treten sehr häufig auf und tragen einen eigenen Namen. Wir wollen vier dieser Verteilungen genauer betrachten: Gleichverteilung, Binomialverteilung, Poisson-Verteilung und geometrische Verteilung.

67.2 Die Gleichverteilung

Sei $\Omega = \{\omega_1, \dots, \omega_n\}$ ein Stichprobenraum mit $P(\omega_k) = \frac{1}{n} \forall k$ (Laplace-Experiment). Ferner sei U_n eine Zufallsvariable mit $U_n(\omega_k) = k$. Dann ist $P_{U_n}(k) = \frac{1}{n}$ für $k = 1, \dots, n$. Eine solche Verteilung heißt (diskrete) Gleichverteilung auf der Menge $\{1, \dots, n\}$.



Es gilt:

$$\begin{aligned}\mu = E(U_n) &= \sum_{k=1}^n k P_{U_n}(k) = \sum_{k=1}^n k \cdot \frac{1}{n} = \frac{1}{n} \sum_{k=1}^n k \quad \text{arithm. Mittel} \\ \sigma^2 &= V(U_n) = \sum_{k=1}^n k^2 P_{U_n}(k) - \mu^2 = \frac{1}{n} \sum_{k=1}^n k^2 - \frac{1}{n^2} \left(\sum_{k=1}^n k \right)^2.\end{aligned}$$

67.3 Beispiel

Beim Würfeln liegt eine diskrete Gleichverteilung auf der Menge $\{1, \dots, 6\}$ vor mit:

$$\begin{aligned}\mu &= \frac{1}{6} \sum_{k=1}^6 k = \frac{21}{6} = 3,5 \\ \sigma^2 &= \frac{1}{6}(1^2 + 2^2 + \dots + 6^2) - 3,5^2 \approx 2,917 \quad \Rightarrow \sigma \approx 1,708.\end{aligned}$$

67.4 Die Binomialverteilung

Wir betrachten ein Experiment, das aus einer n -fachen Wiederholung eines Einzelexperimentes besteht, bei dem ein Ereignis A jeweils mit Wahrscheinlichkeit p eintritt. Ein solches Experiment heißt Bernoulli-Experiment.

Wir setzen $X_i(\omega) = \begin{cases} 1 & \text{falls } A \text{ beim } i\text{-ten Versuch eintritt} \\ 0 & \text{sonst.} \end{cases}$

A_k bezeichnet das Ereignis, das im Gesamtexperiment gilt:

$$X(\omega) := \sum_{i=1}^n X_i(\omega) = k$$

Nach 62.4.(c) (ungeordnete Stichprobe ohne Wiederholung) hat A_k insgesamt $\binom{n}{k}$ Realisierungen. Jedes solche Resultat tritt mit Wahrscheinlichkeit $p^k(1-p)^{n-k}$ auf. Die entsprechende Verteilung

$$b_{n,p}(k) := \binom{n}{k} p^k (1-p)^{n-k}$$

heißt Binomialverteilung mit Paramtern n und p .

Für den Erwartungswert von X_i gilt:

$$E(X_i) = 1 \cdot p + 0 \cdot (1-p) = p \quad \Rightarrow \quad E(X) = \sum_{i=1}^n E(X_i) = np.$$

Da die Einzelereignisse unabhängig sind, gilt:

$$V(X) = \sum_{i=1}^n V(X_i)$$

Mit

$$V(X_i) = E(X_i^2) - (E(X_i))^2 = 1^2 \cdot p + 0^2 \cdot (1-p) - p^2 = p(1-p)$$

folgt

$$V(X) = np(1-p).$$

67.5 Beispiel: Zufallsabhängigkeit sportlicher Resultate

Andreas und Bernd tragen eine Tischtennisturnier mit $n = 1, 3, 5, \dots, 2m+1$ Spielen aus. Wer die meisten Einzelspiele gewinnt, ist Sieger. Andreas gewinnt ein Einzelspiel mit Wahrscheinlichkeit $p = 0,6$. Wie groß

sind die Siegeschancen für den schlechteren Spieler Bernd?

Bernd scheidet, wenn Andreas $S_n \leq m$ Erfolge erzielt. Die Wahrscheinlichkeit hierfür beträgt:

$$\begin{aligned}
 P(S_n \leq m) &= b_{n,p}(0) + b_{n,p}(1) + \dots + b_{n,p}(m) \\
 &= \sum_{k=0}^m \binom{n}{k} p^k (1-p)^{n-k} \\
 n = 1: \quad P(S_1 \leq 0) &= \binom{1}{0} 0,6^0 \cdot 0,4^1 = 1 \cdot 0,4 = 0,4 \\
 n = 3: \quad P(S_3 \leq 1) &= \binom{3}{0} 0,6^0 \cdot 0,4^3 + \binom{3}{1} 0,6^1 \cdot 0,4^2 = 0,4^3 + 3 \cdot 0,6 \cdot 0,4^2 \approx 0,352 \\
 n = 5: \quad P(S_5 \leq 2) &= \binom{5}{0} 0,6^0 \cdot 0,4^5 + \binom{5}{1} 0,6^1 \cdot 0,4^4 + \binom{5}{2} 0,6^2 \cdot 0,4^3 \approx 0,317 \\
 n = 7: \quad P(S_7 \leq 3) &= \dots \approx 0,290 \\
 n = 9: \quad P(S_9 \leq 4) &= \dots \approx 0,267
 \end{aligned}$$

Es ist also gar nicht so unwahrscheinlich, dass der schlechtere Spieler das Turnier gewinnt.

67.6 Die Poissonverteilung

Für große n wird das Arbeiten mit der Binomialverteilung unhandlich. Ist p klein ($0 \leq p \leq 1$), gibt es eine gute Approximation, die Poisson-Verteilung zum Parameter λ :

$$p(k) := \frac{\lambda^k}{k!} e^{-\lambda}, \quad \lambda > 0; \quad k = 0, 1, 2, \dots$$

Man kann zeigen, dass $p(k)$ für $\lambda = np$ die Binomialverteilung $b_{n,p}(k)$ approximiert. Ferner hat eine Poisson-verteilte Zufallsvariable den Erwartungswert λ und die Varianz λ . Durch Umbenennen von „Erfolg“ und „Fehl Schlag“ ist die Poissonverteilung auch für $0,9 \leq p \leq 1$ eine gute Approximation an die Binomialverteilung. Generell beschreibt die Poisson-Verteilung Ereignisse, die im zeitlichen Verlauf zufällig und unabhängig von einander auftreten, z.B.:

- atomarer Zerfall
- das Eintreffen von Bedienwünschen an einem Server
- Anrufe in einem Call-Center
- das Auftreten von Softwarefehlern in einem Programmsystem.

67.7 Reales Beispiel: Der große Jubiläumstag

Genau in einem Jahr feiert ein großer Betrieb seinen 100. Geburtstag. Die Direktion beschließt, allen Kindern von Betriebsangehörigen, die an diesem Tag geboren werden, ein Sparkonto von 3000 Euro anzulegen. Da rund 730 Kinder pro Jahr geboren werden, erwartet man Auslagen von 6000 Euro. Um Zufallsschwankungen vorzubeugen, plant man 15.000 Euro ein. Wie groß ist die Wahrscheinlichkeit, dass das Geld nicht reicht?

$n = 730$ Kinder/Jahr

$p = \frac{1}{365}$ Wahrscheinlichkeit, dass Geburtstag auf Jubiläumstag fällt

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

Das Geld reicht nicht, falls $k \geq 6$ Kinder geboren werden.

$$\begin{aligned}
 p(k \geq 6) &= 1 - p(k \leq 5) = 1 - p(0) - p(1) - \dots - p(5) \\
 &= 1 - \frac{2^0}{0!} e^{-2} - \frac{2^1}{1!} e^{-2} - \dots - \frac{2^5}{5!} e^{-2} \\
 &\approx 0,0168
 \end{aligned}$$

Die Wahrscheinlichkeit einer unangenehmen Zufallsüberraschung ist also gering. Man rechnet nicht damit.

Anmerkung: Am Jubiläumstag wurden 36 Kinder geboren! Die Direktion hat es also verstanden, ihre Angestellten auch für außerbetriebliche Aktivitäten zu begeistern.

67.8 Die geometrische Verteilung

Eine diskrete Zufallsvariable X , die in einem Bernoulli-Experiment angibt, bei welchem Versuch ein Ereignis A mit Wahrscheinlichkeit $p(A) = p$ zum ersten Mal eintritt, heißt geometrisch verteilt mit Parameter p . Die entsprechende Verteilung lautet:

$$P_X(k) = p(1-p)^{k-1}$$

Mann kann zeigen:

$$\begin{aligned} E(X) &= \frac{1}{p} \\ V(X) &= \frac{1-p}{p^2}. \end{aligned}$$

Sie spielt in der Informatik eine große Rolle bei der Modellierung von Wartezeiten.

67.9 Beispiel

Wie lange muss man beim Würfeln warten, bis die Augenzahl 4 erstmalig auftritt?

Mit $p = \frac{1}{6}$ beträgt der Erwartungswert

$$E(X) = \frac{1}{p} = 6$$

und die Varianz

$$V(X) = \frac{1-p}{p^2} = \frac{5/6}{1/6^2} = 30.$$

Kapitel 68

Wichtige kontinuierliche Verteilungen

68.1 Motivation

Zufallsvariablen sind nicht immer diskret, sie können oft auch jede beliebige reelle Zahl in einem Intervall $[c, d]$ einnehmen. Beispiele für solche „kontinuierlichen“ Zufallsvariablen sind Größe, Gewicht oder Zeit. In diesen Fällen macht es wenig Sinn, die Wahrscheinlichkeit anzugeben, dass die Zufallsvariable einen bestimmten Wert annimmt (diese Wahrscheinlichkeit ist 0). Wir müssen Wahrscheinlichkeit für Intervalle betrachten. Hierzu sind Begriffe wie Dichten notwendig.

68.2 Definition (Dichte)

Sei X eine kontinuierliche Zufallsvariable. Existiert eine Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$ mit

$$(a) \quad f(x) \geq 0 \quad \forall x \in \mathbb{R}$$

$$(b) \quad \int_{-\infty}^{\infty} f(x) \, dx = 1$$

und

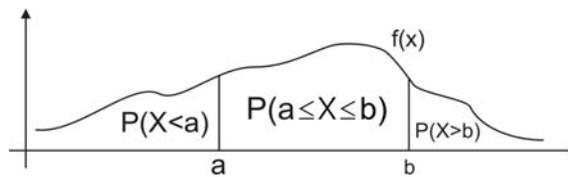
$$P(a \leq X \leq b) = \int_a^b f(x) \, dx.$$

So nennt man f die Dichte von X .

Erwartungswert und Varianz von X sind definiert durch

$$\begin{aligned} \mu &= E(X) = \int_{-\infty}^{\infty} f(x) \, dx, \\ \sigma^2 &= V(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) \, dx. \end{aligned}$$

68.3 Veranschaulichung



68.4 Definition (Verteilungsfunktion)

Sei X eine kontinuierliche Zufallsvariable mit Dichte f . Dann nennt man ihre Stammfunktion

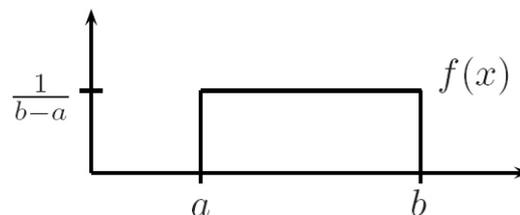
$$F(X) = P(X \leq x) = \int_{-\infty}^x f(u) du$$

die Verteilungsfunktion von X .

68.5 Beispiel: Kontinuierliche Gleichverteilung

Eine kontinuierliche Zufallsvariable, die auf $[c, d]$ gleich verteilt ist, hat die Dichte

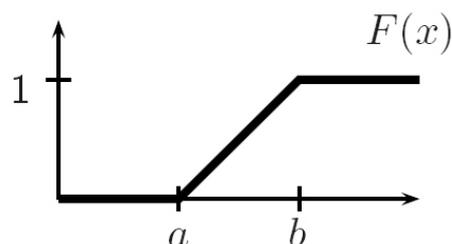
$$f(x) = \begin{cases} \frac{1}{b-a} & (a \leq x \leq b) \\ 0 & (\text{sonst}) \end{cases}$$



und die Verteilungsfunktion

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

Wir kommen nun zur wichtigsten kontinuierlichen Verteilung:



68.6 Die Standardnormalverteilung

Ist X eine kontinuierliche Zufallsvariable mit der Dichte

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

so kann man $P(a < X \leq b)$ mit der Stammfunktion

$$F(X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{u^2}{2}} du$$

berechnen:

$$P(a < X \leq b) = \Phi(b) - \Phi(a).$$

φ nennt man normale Dichte, und Φ ist die Standardnormalverteilung ($N(0, 1)$ -Verteilung).

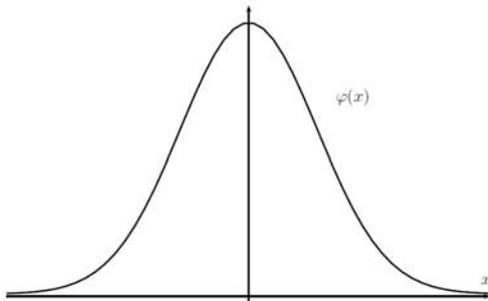


Abbildung 68.1: normale Dichte

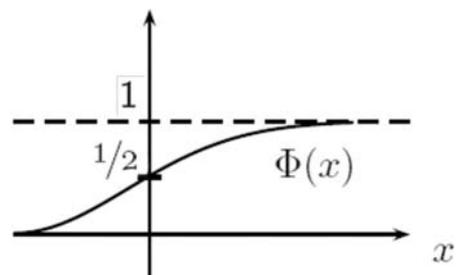


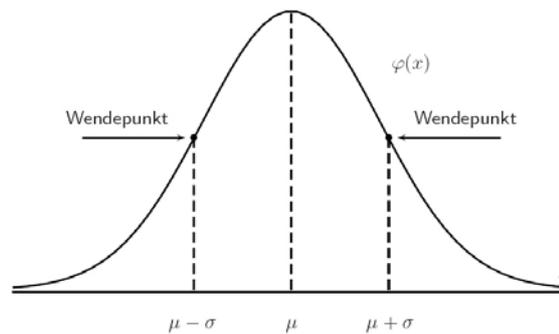
Abbildung 68.2: Standardnormalverteilung

$\Phi(x)$ ist nicht analytisch auswertbar, liegt aber tabelliert vor. Eine standardnormalverteilte Zufallsvariable hat Erwartungswert 0 und Varianz 1. Daher heißt sie $N(0, 1)$ -verteilt.

68.7 Die allgemeine Normalverteilung

Eine kontinuierliche Zufallsvariable X genügt einer allgemeinen Normalverteilung ($N(\mu, \sigma^2)$ -Verteilung, Gauß-Verteilung) mit Erwartungswert μ und Varianz σ^2 , wenn ihre Dichte gegeben ist durch

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



Es gilt:

$$\begin{aligned} P(\mu - \sigma \leq X \leq \mu + \sigma) &\approx 68\% \\ P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) &\approx 95,5\% \\ P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) &\approx 99,7\% \end{aligned}$$

Ist X $N(\mu, \sigma^2)$ -verteilt, so ist $Y := \frac{X - \mu}{\sigma} N(0, 1)$ -verteilt. Somit ist die Tabelle der Standardnormalverteilung ausreichend.

68.8 Approximation der Binomialverteilung durch die Gaußverteilung

Eine Binomialverteilung mit n Einzelexperimenten mit Wahrscheinlichkeit p kann man durch eine allgemeine Normalverteilung mit Erwartungswert np und Standardabweichung $\sigma = \sqrt{np(1-p)}$ approximieren.

$$b_{n,p}(k) = \binom{n}{k} p^k (1-p)^{n-k} \approx \underbrace{\Phi\left(\frac{k + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right)}_{\text{Fläche über dem Intervall } [k - \frac{1}{2}, k + \frac{1}{2}]}$$

Diese Approximation ist gut für $np > 5$ und $n(1-p) > 5$, d.h. insbesondere für große n oder $p \approx 0,5$.

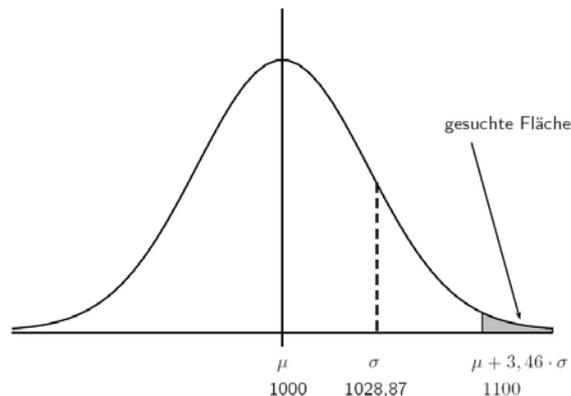
68.9 Beispiel

Wie groß ist die Wahrscheinlichkeit, dass in 6000 Würfeln eines fairen Würfels die Sechs mindestens 1100 Mal auftritt?

$$n = 6000, p = \frac{1}{6}$$

Wegen $np = 1000 > 5$ und $n(1 - p) = 5000 > 5$ ist die Approximation durch die Gaußverteilung sinnvoll:

$$\begin{aligned}\mu &= n \cdot p = 1000 \\ \sigma &= \sqrt{np(1-p)} = \sqrt{6000 \cdot \frac{1}{6} \cdot \frac{5}{6}} \approx 28,87 \\ 1100 &= \underbrace{1000}_{\mu} + 3,46 \cdot \underbrace{28,87}_{\sigma}\end{aligned}$$



Die Wahrscheinlichkeit, mehr als 1100 Sechsen zu würfeln, beträgt

$$1 - \underbrace{\Phi(3,46)}_{\substack{\text{in Tabelle} \\ \text{nachschlagen}}} \approx 0,00028.$$

Der wichtigste Grund für die Bedeutung der Gaußverteilung ist der folgende Satz:

68.10 Satz: (Zentraler Grenzwertsatz)

Seien X_1, X_2, \dots unabhängige Zufallsvariablen, die alle die gleiche Verteilung und somit auch den selben Erwartungswert μ und die selbe Varianz σ^2 besitzen. Ferner sei $Y_n := X_1 + \dots + X_n$ und

$$Z_n := \frac{Y_n - n\mu}{\sigma\sqrt{n}}$$

bezeichne die Standardisierte von Y_n (vgl. 65.15). Dann konvergiert die Verteilungsfunktion $F_n(x)$ von Z_n für $n \rightarrow \infty$ gegen die Standardnormalverteilung $\Phi(x)$.

68.11 Bemerkungen

- Der Beweis von Satz 68.10 ist aufwändig. Siehe z.B. R. Nelson: Probability, Stochastic Processes and Queueing Theory, Springer, New York, 1995, Abschnitt 5.5.6.
- Beachte, dass die einzelnen Zufallsvariablen nicht normalverteilt sein müssen. Ihre Verteilung kann beliebig sein!
- Die Normalverteilung ist also eine sinnvolle Approximation in allen Fällen, in denen sich eine Zufallsvariable aus vielen gleichartigen Einzeleinflüssen zusammensetzt.
 Beispiel: Eine Messung wird oft wiederholt. Dann approximieren die Ergebnisse eine Gaußverteilung.

Kapitel 69

Multivariate Verteilungen und Summen von Zufallsvariablen

69.1 Motivation

Manchmal möchte man das Zusammenwirken mehrerer Zufallsvariablen X_1, \dots, X_n studieren. Gibt es in diesem „multivariaten“ Fall Aussagen über die gemeinsame Verteilung? Lassen sich Aussagen über die Verteilung der Summe von Zufallsvariablen treffen?

69.2 Wichtige Definitionen

Mehrere Zufallsvariablen X_1, \dots, X_n fasst man zu einem Vektor $X = (X_1, \dots, X_n)^\top$ zusammen. Im kontinuierlichen Fall ist die resultierende Dichte eine Funktion mehrerer Variabler. Für diese gemeinsame Dichte gilt:

$$\begin{aligned} f(x_1, \dots, x_n) &\geq 0 \quad \forall x_1, \dots, x_n \in \mathbb{R} \\ \int_{\mathbb{R}^n} f(x_1, \dots, x_n) dx_1 \dots dx_n &= 1 \\ P(a_1 \leq X_1 \leq b_1, \dots, a_n \leq X_n \leq b_n) &= \int_{a_n}^{b_n} \dots \int_{a_1}^{b_1} f(x_1, \dots, x_n) dx_1 \dots dx_n. \end{aligned}$$

Ferner betrachtet man die multivariate Verteilungsfunktion

$$F(x_1, \dots, x_n) = \int_{-\infty}^{x_n} \dots \int_{-\infty}^{x_1} f(x_1, \dots, x_n) dx_1 \dots dx_n.$$

Statt eines einzelnen Erwartungswerts hat man einen Erwartungswertvektor

$$\mu = (E(X_1), \dots, E(X_n))^\top.$$

Varianzen und Kovarianzen fasst man zu einer symmetrischen und positiv definiten Kovarianzmatrix zusammen

$$\Sigma = \begin{pmatrix} V(X_1) & \text{Cov}(X_1, X_2) & \dots & \text{Cov}(X_1, X_n) \\ \text{Cov}(X_2, X_1) & V(X_2) & & \text{Cov}(X_2, X_n) \\ \vdots & & \ddots & \\ \text{Cov}(X_n, X_1) & & & V(X_n) \end{pmatrix}$$

Die wichtigste multivariate Verteilung ist die multivariate Normalverteilung ($N_n(\mu, \Sigma)$ -Verteilung):

Sei $X = (X_1, \dots, X_n)^\top$ ein Vektor von normalverteilten Zufallsvariablen mit Erwartungswert $\mu = (E(X_1), \dots, E(X_n))^\top$ und Kovarianzmatrix Σ , dann besitzt die multivariate Normalverteilung die Dichte

$$f(x) = \frac{1}{(2\pi)^{n/2} \sqrt{\det \Sigma}} e^{-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)} \quad (x \in \mathbb{R}^n).$$

69.3 Beispiel

Eine Apfelbaumplantage mit gleich alten Bäumen werde durch 3 normalisierte Zufallsvariablen beschrieben ($N(\mu, \sigma^2)$ -verteilt):

X_1 : Höhe eines Baumes [m] $N(4, 1)$ -verteilt
 X_2 : Ertrag [1 kg] $N(20, 100)$ -verteilt
 X_3 : Zahl der Blätter [1000 Stück] $N(20, 225)$ -verteilt.

Diese Zufallsvariablen seien korreliert mit

$$\begin{aligned} \text{Cov}(X_1, X_2) &= 9 \\ \text{Cov}(X_1, X_3) &= 12,75 \\ \text{Cov}(X_2, X_3) &= 120. \end{aligned}$$

Dann liegt eine $N_3(\mu, \Sigma)$ -Verteilung vor mit

$$\mu = \begin{pmatrix} 4 \\ 20 \\ 20 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1 & 9 & 12,75 \\ 9 & 100 & 120 \\ 12,75 & 120 & 225 \end{pmatrix}.$$

Kann man unter geeigneten Voraussetzungen die gemeinsame Dichte $f(x_1, \dots, x_n)$ aus den einzelnen Dichten $f_1(x_1), \dots, f_n(x_n)$ berechnen?

Man kann zeigen:

69.4 Satz: (Gemeinsame Dichte unabhängiger Zufallsvariablen)

Seien X_1, \dots, X_n unabhängige(!) Zufallsvariablen mit Dichten $f_1(x_1), \dots, f_n(x_n)$, so hat $X = (X_1, \dots, X_n)^\top$ die gemeinsame Dichte

$$f(x_1, \dots, x_n) = f_1(x_1) \cdot \dots \cdot f_n(x_n) \quad (*).$$

Hat umgekehrt $X = (X_1, \dots, X_n)^\top$ eine gemeinsame Dichte in der Produktdarstellung (*), so sind X_1, \dots, X_n unabhängig.

69.5 Beispiel

Zwei unabhängige Zufallsvariablen X_1, X_2 seien $N(\mu_1, \sigma_1^2)$ - bzw. $N(\mu_2, \sigma_2^2)$ -verteilt. Da Unabhängigkeit Unkorreliertheit impliziert (vgl. 65.22.(c)), hat die Kovarianzmatrix Diagonalgestalt:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$$

Mit $\det \Sigma = \sigma_1^2 \sigma_2^2$ und $\Sigma^{-1} = \begin{pmatrix} \frac{1}{\sigma_1^2} & 0 \\ 0 & \frac{1}{\sigma_2^2} \end{pmatrix}$ hat $X = (X_1, X_2)^\top$ nach 69.2 eine multivariate Normalverteilung mit Dichte

$$\begin{aligned} f(x_1, x_2) &= \frac{1}{2\pi \sqrt{\sigma_1^2 \sigma_2^2}} e^{-\frac{1}{2}[(x_1 - \mu_1)^2 / \sigma_1^2 + (x_2 - \mu_2)^2 / \sigma_2^2]} \\ &= \frac{1}{\sqrt{2\pi} \sigma_1} e^{-\frac{(x_1 - \mu_1)^2}{2\sigma_1^2}} \cdot \frac{1}{\sqrt{2\pi} \sigma_2} e^{-\frac{(x_2 - \mu_2)^2}{2\sigma_2^2}}. \end{aligned}$$

Dies ist gerade das Produkt der zwei Einzeldichten $f_1(x_1)$ und $f_2(x_2)$.

Gibt es Aussagen über die Dichte, wenn man die Summe zweier Zufallsvariablen betrachtet? Hierzu benötigen wir

69.6 Definition (Faltung)

Falls für die Funktionen $f, g : \mathbb{R} \rightarrow \mathbb{R}$ das Integral

$$(f * g)(x) := \int_{-\infty}^{\infty} f(x-y)g(y)dy$$

existiert, so nennen wir $f * g$ die Faltung von f und g (engl.: convolution).

69.7 Satz: (Summe unabhängiger kontinuierlicher Zufallsvariabler)

Seien X_1, X_2 unabhängige kontinuierliche Zufallsvariable mit den Dichten f_1, f_2 , so hat $X_1 + X_2$ die Dichte $f_1 * f_2$.

Beweis:

Mit $B := \{(x_1, x_2) \mid x_1 + x_2 \leq s\}$ ergibt sich für die Verteilung von $X_1 + X_2$:

$$P(X_1 + X_2 \leq s) = \iint_B \underbrace{f_1(x_1)f_2(x_2)}_{\text{Unabh.}} dx_1 dx_2$$

Mit der Substitution $u := x_1 + x_2$ folgt:

$$P(X_1 + X_2 \leq s) = \int_{-\infty}^s \underbrace{\left(\int_{-\infty}^{\infty} f_1(u-x_2)f_2(x_2) dx_2 \right)}_{(f_1 * f_2)(u)} du$$

□.

Hiermit läßt sich beweisen:

69.8 Satz: (Summe unabhängiger normalverteilter Zufallsvariabler)

Seien X_1, X_2 unabhängige kontinuierliche Zufallsvariablen mit $N(\mu_1, \sigma_1^2)$ - bzw. $N(\mu_2, \sigma_2^2)$ -Verteilung. Dann ist $X = X_1 + X_2$ $N(\mu, \sigma^2)$ -verteilt und es gilt:

$$\begin{aligned} \mu &= \mu_1 + \mu_2 \\ \sigma^2 &= \sigma_1^2 + \sigma_2^2. \end{aligned}$$

Auch im Fall diskreter Zufallsvariablen gibt es vergleichbare Aussagen zu 69.6-69.8:

69.9 Definition (Diskrete Faltung)

Für $f = (f_i)_{i \in \mathbb{Z}}, g = (g_i)_{i \in \mathbb{Z}}$ definiert man die diskrete Faltung von f und g durch

$$(f * g)_i := \sum_{j \in \mathbb{Z}} f_{i-j}g_j.$$

69.10 Satz: (Summe unabhängiger diskreter Zufallsvariabler)

Seien X_1, X_2 unabhängige diskrete Zufallsvariablen mit Verteilungen P_{X_1}, P_{X_2} . Dann hat $P_{X_1+X_2}$ die Verteilung $P_{X_1} * P_{X_2}$, wobei $*$ die diskrete Faltung bezeichnet.

69.11 Satz: (Summe unabhängiger Poisson-verteilter Zufallsvariabler)

Seien X_1, X_2 unabhängige diskrete Zufallsvariablen, die einer Poisson-Verteilung mit Parameter λ_1 bzw λ_2 genügen. (kurz: $P(\lambda_1)$ -, $P(\lambda_2)$ -verteilt). Dann ist $X_1 + X_2$ $P(\lambda_1 + \lambda_2)$ -verteilt.

69.12 Beispiel

Beim radioaktiven Zerfall einer Substanz werden ionisierende Teilchen frei. Mit einem Geiger-Müller-Zählrohr zählt man die innerhalb einer Minute eintreffenden Teilchen. Sie sind Poisson-verteilt. Hat man zwei radioaktive Substanzen mit Poisson-Verteilung $P(\lambda_1)$ bzw. $P(\lambda_2)$, so genügt die Gesamtheit der pro Zeitintervall produzierten Teilchen einer $P(\lambda_1 + \lambda_2)$ -Verteilung.

Kapitel 70

Parameterschätzung und Konfidenzintervalle

70.1 Motivation

Bisher sind wir stets von theoretischen Modellen (z.B. „fairer Würfel“) ausgegangen, die erlauben, Parameter wie Erwartungswert oder Varianz einer Verteilung exakt zu berechnen. In vielen realen Situationen kennt man jedoch nur den Verteilungstyp und muss auf Grund von Stichproben die Parameter schätzen. Wie geht man dabei vor? Die geschätzten Parameter sind i.A. fehlerhaft. Lässt sich ein Vertrauensintervall angeben, innerhalb dessen ein Parameter mit einer vorgegebenen Sicherheit liegt?

70.2 Definition (Stichprobenwerte)

Gegeben seien n Beobachtungswerte x_1, \dots, x_n eines Zufallsexperiments. Dann nennen wir $(x_1, \dots, x_n)^\top$ Stichprobe vom Umfang n . Die einzelnen x_i heißen Stichprobenwerte.

70.3 Beispiel

In einer Kiste befinden sich 10.000 Schrauben. Ein Teil davon ist fehlerhaft. Für eine Stichprobe werden 100 Schrauben entnommen. Die Zufallsvariable X_i beschreibt den Zustand der i -ten entnommenen Schraube:

$$X_i(\omega) = \begin{cases} 0 & \text{falls } i\text{-te Schraube in Ordnung} \\ 1 & \text{falls } i\text{-te Schraube defekt.} \end{cases}$$

Eine konkrete Realisierung des Zufallsvektors $(X_1, \dots, X_{100})^\top$ liefere die Stichprobe $(x_1, \dots, x_{100})^\top = (0, 1, 0, 0, 1, 0, \dots, 1, 0)^\top$.

So wie wir bei den Zufallsvariablen Parameter wie Erwartungswert oder Varianz zugeordnet haben, können wir auch für Stichproben Kenngrößen definieren:

70.4 Definition (Mittelwert, Varianz, Standardabweichung)

Für eine Stichprobe $(x_1, \dots, x_n)^\top$ definiert man:

- den Mittelwert (arithmetisches Mittel) durch:

$$\bar{x} := \frac{1}{n}(x_1 + \dots + x_n)$$

- die Varianz durch

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

- die Standardabweichung durch

$$s := \sqrt{s^2}.$$

70.5 Bemerkungen

- Man kann zeigen, dass der Mittelwert \bar{x} und die Varianz s^2 geeignete Approximationen an den Erwartungswert μ und die Varianz σ^2 einer Zufallsvariablen sind.
- Die Tatsache, dass im Nenner von s^2 die Größe $n-1$ statt n steht, hat tiefere theoretische Hintergründe, auf die wir hier nicht eingehen (siehe z.B. Hartmann, Satz ??)
- Ähnlich zum Verschiebungssatz 65.12 gibt eine häufig benutzte Formel zum Berechnen der Varianz einer Stichprobe:

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

Beweis:

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

□.

70.6 Beispiel

Bei einer Wahlumfrage geben 400 von 1000 Personen an, die Partei A wählen zu wollen. Das Umfrageinstitut prognostiziert auf Grund dieser Stichprobe einen Wahlausgang mit 40% aller Stimmen für Partei A .

70.7 Konfidenzintervalle

In Beispiel 70.6 werden verschiedene Stichproben zu leicht unterschiedlichen Resultaten führen, die wiederum i.A. vom tatsächlichen Wahlausgang abweichen.

Können wir statt eines einzelnen Werts $p = 0,4$ ein Vertrauensintervall (Konfidenzintervall) $[p_u, p_o]$ angeben, innerhalb dessen das Endresultat mit einer vorgegebenen Ws. (Konfidenzniveau) von z.B. 95% liegt?

70.8 Beispiel: Wahlumfrage aus 70.6

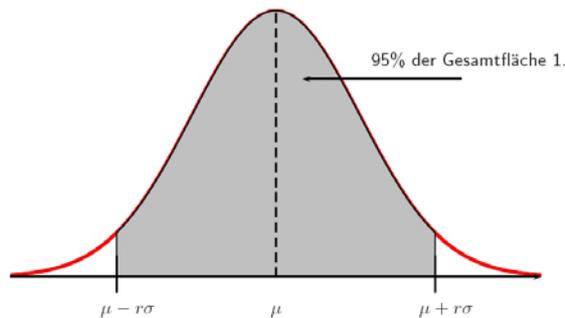
Wir gehen von einer Binomialverteilung aus und schätzen p durch $p = \frac{400}{1000} = 0,4$ ab. Sei

$$X_i = \begin{cases} 1 & \text{Befragte/r } i \text{ wählt } A \\ 0 & \text{Befragte/r } i \text{ wählt } A \text{ nicht} \end{cases}$$

und $X := \sum_{i=1}^{1000} X_i$. Dann gilt nach 67.4 mit $p = 0,4$ und $n = 1000$:

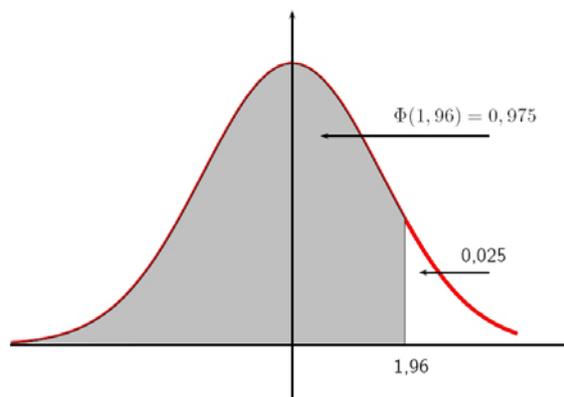
$$\begin{aligned} \mu &= E(X) = n \cdot p = 400 \\ \sigma^2 &= V(X) = n \cdot p \cdot (1 - p) = 240 \Rightarrow \sigma \approx 15,49. \end{aligned}$$

Wegen $np = 400 > 5$ und $n(1 - p) = 600 > 5$ können wir für X auch eine Normalverteilung mit $\mu = 400$ und $\sigma = 15,49$ annehmen. Wir suchen ein Intervall $[\mu - r\sigma, \mu + r\sigma]$, innerhalb dessen das Integral über die Dichtefunktion den Wert 0,95 annimmt:



Tabelliert ist die Standardnormalverteilung ($\mu = 0, \sigma = 1$)

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



Man findet: $\Phi(1,96) \approx 0,975$.
Somit ist aus Symmetriegründen

$$\frac{1}{\sqrt{2\pi}} \int_{-1,96}^{1,96} e^{-\frac{t^2}{2}} dt \approx 0,95$$

und $r = 1,96$.

Damit ergibt sich das Konfidenzintervall

$$\begin{aligned} [\mu - r\sigma, \mu + r\sigma] &= [400 - 1,96 \cdot 15,49; 400 + 1,96 \cdot 15,49] \\ &\approx [369,6; 430,4]. \end{aligned}$$

Bei einem Konfidenzniveau von 95% erzielt Partei A also zwischen 36,96% und 43,04% der Stimmen. Möchte man ein kleineres Konfidenzintervall, muss man mehr Personen befragen.

70.9 Beispiel: Überbuchung eines Flugzeugs

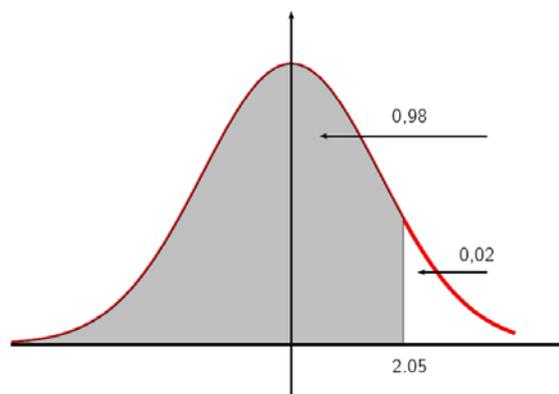
Ein Flugzeug hat 200 Sitze. Wie viele Reservierungen dürfen angenommen werden, wenn erfahrungsgemäß 5% aller Passagiere nicht erscheinen? Die Fluggesellschaft ist bereit, in 1 von 50 Fällen in Verlegenheit zu geraten.

Sei n die Anzahl der Reservierungen und X die Anzahl der tatsächlich erscheinenden Passagiere. Legt man eine Binomialverteilung zu Grunde mit $p = 0,95$, so gilt:

$$\begin{aligned} \mu &= E(X) = np = 0,95n \\ \sigma^2 &= np(1-p) \approx 0,0475 \Rightarrow \sigma = 0,2179\sqrt{n} \end{aligned}$$

Der Tabelle der Standardnormalverteilung entnimmt man:

$$\Phi(2,05) \approx 0,98 = 1 - \frac{1}{50}$$



Fordert man

$$\mu + 2,05\sigma \stackrel{!}{\leq} 200$$

ergibt sich:

$$0,95n + 2,05 \cdot 0,2179\sqrt{n} \leq 200$$

Man prüft leicht nach, dass dies für $n \leq 203$ erfüllt ist. (ausprobieren oder $Y := \sqrt{n}$ setzen und quadratische Gleichung lösen).

Bemerkung: Die Approximation durch die Normalverteilung war gerechtfertigt wegen $np > 5$ und $n(1-p) > 5$.

Kapitel 71

Hypothesentests

71.1 Motivation

Bei Hypothesentests will man eine gewisse Annahme über eine Zufallsvariable darauf hin überprüfen, ob sie korrekt ist. Beispiele:

- Ist eine Münze fair $\left(p = \frac{1}{2}\right)$?
- Sind die Rechner von Hersteller A zuverlässiger als von Hersteller B ?

Ein statistisches Experiment soll uns dabei eine Entscheidung mit einer vorgegebenen Irrtumswahrscheinlichkeit ermöglichen. Gegenüber den Verfahren aus §70 kann man in den Rechnungen die Hypothese mit verwenden, hat also mehr in der Hand.

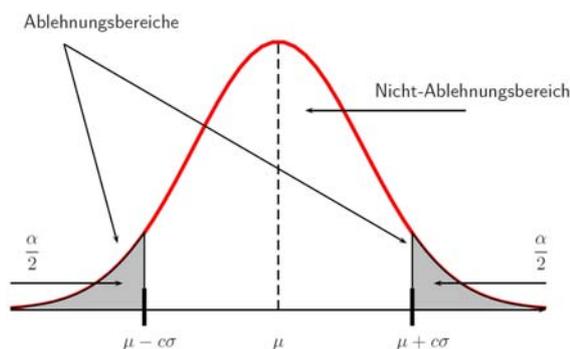
71.2 Parametertest am Beispiel eines Münzexperimentes

Wir beobachten das Ereignis $A = \text{„Münze zeigt Kopf“}$ und wollen die Hypothese $p_0 = p(A) = \frac{1}{2}$ überprüfen, indem wir 200 Münzwürfe durchführen:

$$X_i = \begin{cases} 1 & \text{(Münze zeigt Kopf beim } i\text{-ten Wurf)} \\ 0 & \text{(Münze zeigt Zahl beim } i\text{-ten Wurf)}. \end{cases}$$

Wie weit darf $S_{200} := \sum_{i=1}^{200} X_i$ sich vom Erwartungswert 100 unterscheiden, damit wir mit einer Irrtumsws.

$\alpha = 0,05$ (Signifikanzniveau von $1 - \alpha = 0,95$) die Hypothese $p_0 = \frac{1}{2}$ nicht ablehnen?



Wir legen eine Binomialverteilung mit $n = 200, p = 0,5$ zu Grunde, die wir durch eine Normalverteilung mit $\mu = np = 100, \sigma = \sqrt{np(1-p)} = \sqrt{50} \approx 7,07$ approximieren. Wegen $\Phi(1,96) \approx 0,975$ ist $c = 1,96$ für $\alpha = 0,05$. Tritt bei 200 Würfeln eine Kopffzahl S_n außerhalb $[\mu - c\sigma; \mu + c\sigma] \approx [86,1; 113,6]$ auf, wird man die Hypothese $p_0 = \frac{1}{2}$ auf einem Signifikanzniveau 0,95 ablehnen. Andernfalls wird man sie nicht ablehnen.

71.3 Bemerkungen

- (a) Eine Hypothese an einen Parameter (etwa $p_0 = \frac{1}{2}$) nennt man auch Nullhypothese H_0 , die Gegenannahme (etwa $p \neq \frac{1}{2}$) ist die Gegenhypothese H_1 .
- (b) Bei Hypothesentests können 2 Arten von Fehlern auftreten:
- **Fehler 1. Art:** Hypothese wird abgelehnt, obwohl sie richtig ist (wird durch Irrtumsws. α beschrieben).
 - **Fehler 2. Art:** Hypothese wird angenommen, obwohl sie falsch ist. Dieser Fehler kann insbesondere für kleines α sehr groß sein.

71.4 Der χ^2 -Test („Chi-Quadrat-Test“)

Der χ^2 -Test ist einer der wichtigsten Tests. Er wird bei folgenden Problem angewandt:
 Ein Versuch habe m mögliche Ausgänge. Wir testen die Hypothese H_0 , dass die Resultate mit vorgegebenen Ws. p_1, \dots, p_m auftreten.
 Trifft H_0 zu, erwarten wir bei n Versuchen als Häufigkeit für die einzelnen Ausgänge: np_1, \dots, np_m .
 In Wirklichkeit werden die Häufigkeiten X_1, \dots, X_m beobachtet.
 Als Maß für die Abweichung X_i und np_i verwendet man

$$\chi^2 := \sum_{i=1}^m \frac{(X_i - np_i)^2}{np_i}$$

Ist χ^2 „zu groß“, wird man H_0 ablehnen.
 Um zu beurteilen, was „zu groß“ bedeutet, ist es sinnvoll den Erwartungswert von χ^2 zu kennen. Ist jedes X_i $b_{n,p}$ -verteilt, gilt auch

$$V(X_i) = E((X_i - np_i)^2) = np_i(1 - p_i)$$

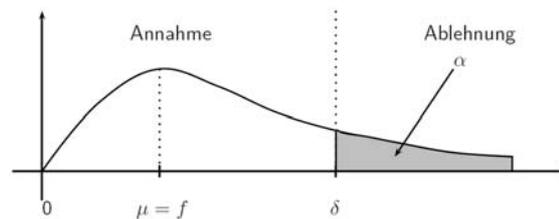
und aus der Linearität des Erwartungswerts folgt:

$$\begin{aligned} E(\chi^2) &= \sum_{i=1}^m \frac{1}{np_i} E((X_i - np_i)^2) = \sum_{i=1}^m \frac{1}{np_i} np_i(1 - p_i) \\ &= \sum_{i=1}^m 1 - \sum_{i=1}^m p_i = m - 1. \end{aligned}$$

$f := m - 1$ bezeichnet die Freiheitsgrade der χ^2 -Verteilung, d.h. $m - 1$ der p_i , $i = 1, \dots, m$ sind frei wählbar. Es gilt also

$$\mu = E(\chi^2) = f$$

Der typische Verlauf einer χ^2 -Verteilung sieht folgendermaßen aus:



Für einen gegebenen Freiheitsgrad f und eine Irrtumsws. α ist die χ^2_f -Verteilung tabelliert. Man nimmt H_0 an, falls der berechnete χ^2 -Wert $\leq \delta$ ist. Man lehnt H_0 ab, falls der berechnete χ^2 -Wert $> \delta$ ist.

71.5 Beispiel

Wir wollen mit 120 Würfeln nachprüfen, ob ein Würfel „fair“ ist, d.h. alle Augenzahlen sind gleich wahrscheinlich:

$$p_1 = \dots = p_6 = \frac{1}{6}.$$

Als Ergebnis erhalten wir:

Augenzahl i	1	2	3	4	5	6
beobachtete Häufigkeit X_i	15	21	25	19	14	26
erwartete Häufigkeit np_i	20	20	20	20	20	20

Wir erhalten:

$$\chi^2 = \frac{(15 - 20)^2}{20} + \frac{(21 - 20)^2}{20} + \dots + \frac{(26 - 20)^2}{20} \approx 6,2.$$

Wir haben $f = 6 - 1 = 5$ Freiheitsgrade. Geben wir eine Irrtumsws. von $\alpha = 0,1$ vor, so findet man in einer Tabelle

$$p(\chi^2 \leq \underbrace{9,24}_{\delta}) = \underbrace{0,9}_{1-\alpha}$$

Wegen $\chi^2 = 6,2 \leq 9,24 = \delta$ akzeptieren wir die Hypothese H_0 , dass alle Augenzahlen gleich wahrscheinlich sind.

71.6 Bemerkungen

Möchte man eine $N(\mu, \sigma^2)$ -Verteilung mit dem χ^2 -Test für ein gegebenes μ, σ^2 verifizieren, teilt man in m (z.B. gleich wahrscheinliche) Klassen ein:

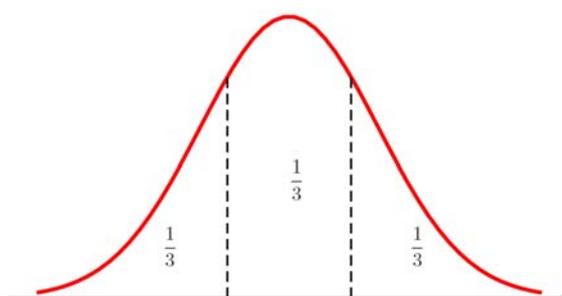
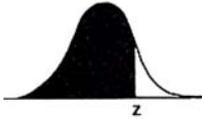
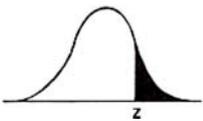
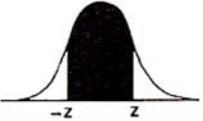
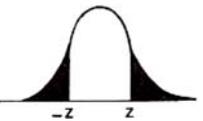


Abbildung 71.1: Bsp.: $m = 3$ Klassen

Dann überprüft man, ob die experimentell ermittelten Häufigkeiten in jeder Klasse das χ^2 -Kriterium zu einer vorgegebenen Irrtumsws. α bei $f = m - 1$ Freiheitsgraden erfüllen.

Wahrscheinlichkeiten der folgenden Ereignisse:

	$S_n - np \leq z\sigma$	$S_n - np \geq z\sigma$	$ S_n - np \leq z\sigma$	$ S_n - np \geq z\sigma$
				
z	$\Phi(z)$	$1 - \phi(z)$	$2\Phi(z) - 1$	$2 - 2\Phi(z)$
0.0	.500	.500	.0000	1.0000
0.1	.540	.460	.0797	.9203
0.2	.579	.421	.159	.841
0.3	.618	.382	.236	.764
0.4	.655	.345	.311	.689
0.5	.691	.309	.383	.617
0.6	.726	.274	.451	.549
0.7	.758	.242	.516	.484
0.8	.788	.212	.576	.424
0.9	.816	.184	.632	.368
1.0	.841	.159	.683	.317
1.1	.864	.136	.729	.271
1.2	.885	.115	.770	.230
1.3	.9032	.0968	.806	.194
1.4	.9192	.0808	.838	.162
1.5	.9332	.0668	.866	.134
1.6	.9452	.0548	.890	.110
1.7	.9554	.0446	.9109	.0891
1.8	.9641	.0359	.9281	.0719
1.9	.9713	.0287	.9425	.0575
2.0	.9772	.0228	.9545	.0455
2.1	.9821	.0179	.9643	.0357
2.2	.9861	.0139	.9722	.0278
2.3	.9893	.0107	.9786	.0217
2.4	.99180	.00820	.9836	.0164
2.5	.99379	.00621	.9876	.0124
2.6	.99534	.00466	.99068	.00932
2.7	.99653	.00347	.99307	.00693
2.8	.99744	.00256	.99489	.00511
2.9	.99813	.00187	.99627	.00373
3.0	.99865	.00135	.99730	.00270
3.1	.999032	.000968	.99806	.00194
3.2	.999313	.000687	.99863	.00137
3.3	.999517	.000483	.999033	.000967
3.4	.999663	.000337	.999326	.000674
3.5	.999767	.000233	.999535	.000465
3.6	.999841	.000159	.999682	.000318
3.7	.999892	.000108	.999784	.000216
3.8	.9999277	.0000723	.999855	.000145
3.9	.9999519	.0000481	.9999038	.0000962
4.0	.9999683	.0000317	.9999367	.0000633

Die Stetigkeitskorrektur kann man berücksichtigen, indem man die Grenzen des schwarzen Gebiets um $\frac{1}{2\sigma}$ ins weiße Gebiet verlegt. Wahrscheinlichkeiten für $z < 0$ erhält man durch Symmetrie.

Schranken für χ^2 bei f Freiheitsgraden

$f \setminus P$	0,99	0,975	0,95	0,90	0,10	0,05	0,025	0,01
1	0,00016	0,00098	0,00393	0,01579	2,70554	3,84146	5,02389	6,63490
2	0,00201	0,00506	0,10259	0,21072	4,60517	5,99147	7,37776	9,21034
3	0,11483	0,21580	0,35185	0,58438	6,25139	7,81473	9,34840	11,3449
4	0,29711	0,48442	0,71072	1,06362	7,77944	9,48773	11,1433	13,2767
5	0,55430	0,83121	1,14548	1,61031	9,23635	11,0705	12,8325	15,0863
6	0,87209	1,23735	1,63539	2,20413	10,6446	12,5916	14,4494	16,8119
7	1,23904	1,68987	2,16735	2,83311	12,0170	14,0671	16,0128	18,4753
8	1,64648	2,17973	2,73264	3,48954	13,3616	15,5073	17,5346	20,0902
9	2,08781	2,70039	3,32511	4,16816	14,6837	16,9190	19,0228	21,6660
10	2,55821	3,24697	3,94030	4,86518	15,9871	18,3070	20,4831	23,2093
11	3,0535	3,8158	4,5748	5,5778	17,275	19,675	21,920	24,725
12	3,5706	4,4038	5,2260	6,3038	18,549	21,026	23,337	26,217
13	4,1069	5,0087	5,8919	7,0415	19,812	22,362	24,736	27,688
14	4,6604	5,6287	6,5706	7,7895	21,064	23,685	26,119	29,143
15	5,2294	6,2621	7,2604	8,5468	22,307	24,996	27,488	30,578
16	5,812	6,908	7,962	9,312	23,54	26,30	28,85	32,00
17	6,408	7,564	8,672	10,09	24,77	27,59	30,19	33,41
18	7,015	8,231	9,390	10,86	25,99	28,87	31,53	34,81
19	7,633	8,907	10,12	11,65	27,20	30,14	32,85	36,19
20	8,260	9,591	10,85	12,44	28,41	31,41	34,17	37,57
21	8,897	10,28	11,59	13,24	29,62	32,67	35,48	38,93
22	9,542	10,98	12,34	14,04	30,81	33,92	36,78	40,29
23	10,20	11,69	13,09	14,85	32,00	35,17	38,08	41,64
24	10,86	12,40	13,85	15,66	33,20	36,42	39,36	42,98
25	11,52	13,12	14,61	16,47	34,38	37,65	40,65	44,31
26	12,20	13,84	15,38	17,29	35,56	38,89	41,92	45,64
27	12,88	14,57	16,15	18,11	36,74	40,11	43,19	46,96
28	13,56	15,31	16,93	18,94	37,92	41,34	44,46	48,28
29	14,26	16,05	17,71	19,77	39,09	42,56	45,72	49,59
30	14,95	16,79	18,49	20,60	40,26	43,77	46,98	50,89

Erläuterungen der Tafel. Z.B. die 5. Zeile der Tafel bedeutet: Bei 5 Freiheitsgraden ist

$$\begin{aligned}
 P(\chi^2 \geq 0,5543) &= 0,99, \\
 P(\chi^2 \geq 0,83121) &= 0,975, \\
 &\vdots \\
 P(\chi^2 \geq 15,0863) &= 0,01.
 \end{aligned}$$

Kapitel 72

Methode der kleinsten Quadrate

72.1 Problemstellung

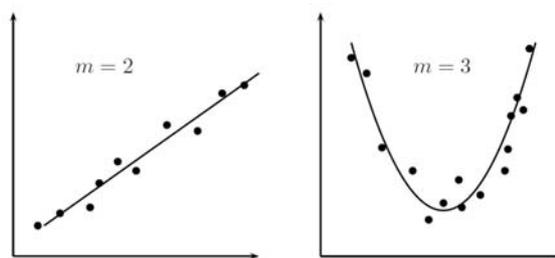
In einem Experiment interessiert man sich für die Beziehung zwischen zwei Variablen x und y . Hierzu hat man viele Wertepaare

$$(x_1, y_1), \dots, (x_n, y_n)$$

gemessen. Die Messungen können Fehler in Form von statistischen Fluktuationen enthalten. Man möchte nun die Beziehung zwischen x und y durch ein einfaches Polynom

$$y = f(x) = \sum_{k=1}^m a_k x^{k-1}$$

approximieren (z.B. $m = 2$: Gerade, $m = 3$: Parabel) und sucht die „optimalen“ Koeffizienten a_1, \dots, a_m .



72.2 Methode der kleinsten Quadrate

Jede der n Messungen (x_i, y_i) beschreibt eine lineare Gleichung für die Unbekannten a_1, \dots, a_m :

$$\begin{aligned} a_1 + a_2 x_1 + \dots + a_m x_1^{m-1} &= y_1 \\ &\vdots \\ a_1 + a_2 x_n + \dots + a_m x_n^{m-1} &= y_n \end{aligned}$$

Im Allgemeinen hat man sehr viel mehr Gleichungen als Unbekannte ($n \gg m$), und das lineare Gleichungssystem

$$\underbrace{\begin{pmatrix} 1 & x_1 & \dots & x_1^{m-1} \\ \vdots & & & \vdots \\ 1 & x_n & \dots & x_n^{m-1} \end{pmatrix}}_{M \in \mathbb{R}^{n \times m}} \underbrace{\begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix}}_{a \in \mathbb{R}^m} = \underbrace{\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}}_{y \in \mathbb{R}^n}$$

ist inkonsistent. Beispielsweise kann man nicht erwarten, dass 50 Messwerte exakt auf einer Geraden liegen ($n = 50, m = 2$).

Da $Ma = y$ nicht exakt lösbar ist, sucht man statt dessen eine „Lösung“ a^* , die den quadratischen Fehler

$$|Ma - y|^2 = \left(\sum_{k=1}^m a_k x_1^{k-1} - y_1 \right)^2 + \dots + \left(\sum_{k=1}^m a_k x_n^{k-1} - y_n \right)^2$$

minimiert (Ausgleichskurve, Regressionskurve).

72.3 Minimierung des quadratischen Fehlers

Wir suchen das Minimum der Funktion

$$\begin{aligned} f(a_1, \dots, a_m) &= (Ma - y)^\top (Ma - y) \\ &= a^\top M^\top Ma - a^\top M^\top y - \underbrace{y^\top Ma}_{a^\top M^\top y} + y^\top y \\ &= a^\top M^\top Ma - 2a^\top M^\top y + y^\top y. \end{aligned}$$

Notwendige Bedingung:

$$\begin{aligned} 0 \stackrel{!}{=} \nabla_a f &:= \begin{pmatrix} \frac{\partial f}{\partial a_1} \\ \vdots \\ \frac{\partial f}{\partial a_m} \end{pmatrix} = M^\top Ma + \underbrace{a^\top M^\top M}_{\substack{(M^\top M)a, \\ \text{da } M^\top M \\ \text{symmetrisch}}} - 2M^\top y \\ &= 2M^\top Ma - 2M^\top y. \end{aligned}$$

Die Lösung a^* löst also die so genannte Normalengleichung

$$\boxed{M^\top Ma = M^\top y}$$

Dies ist ein System aus m Gleichungen mit m Unbekannten a_1, \dots, a_m . Ist $M^\top M$ invertierbar, gilt:

$$\boxed{a^* = (M^\top M)^{-1} M^\top y}$$

Man nennt

$$\boxed{M^+ := (M^\top M)^{-1} M^\top}$$

die Pseudoinverse (Moore-Peurose-Inverse) der (nicht invertierbaren!) $n \times m$ -Matrix M .

72.4 Bemerkungen

- (a) a^* ist tatsächlich ein Minimum:
Die Hesse-Matrix $H_a f = 2M^T M$ ist positiv semidefinit:

$$x^T \cdot 2M^T M x = 2(Mx)^T M x \geq 0 \quad \forall x \in \mathbb{R}^n.$$

Da $M^T M$ invertierbar sein soll, ist $H_a f$ sogar positiv definit. Nach 57.5 folgt also: a^* ist ein Minimum.

- (b) Man kann zeigen, dass $M^T M$ invertierbar ist, falls $\text{rang}(M) = m$, d.h. es gibt m der n Gleichungen des Systems $Ma = b$, die linear unabhängig sind.
- (c) Wir haben also am Beispiel der Ausgleichsrechnung ein allgemeines Verfahren hergeleitet, um ein überbestimmtes (und i.A. inkonsistentes) Gleichungssystem zu „lösen“:

72.5 Satz: (Pseudolösung überbestimmter Gleichungssysteme)

Sei $n > m$, $A \in \mathbb{R}^{n \times m}$, $\text{rang}(A) = m$ und $b \in \mathbb{R}^n$. Falls das überbestimmte lineare Gleichungssystem

$$Ax = b$$

inkonsistent ist, ist es nicht exakt lösbar. Es gibt jedoch eine eindeutig bestimmte Pseudolösung x^* , die den quadratischen Fehler $|Ax - b|^2$ minimiert:

$$x^* = A^+ b$$

mit der Pseudoinversen $A^+ = (A^T A)^{-1} A^T$.

72.6 Bemerkung

Pseudolösungen spielen auch in der Informatik eine große Rolle. Überbestimmte Gleichungssysteme treten z.B. bei der Suche in Internetdatenbanken auf (\rightarrow Prof. Weikum, Informationssysteme)

72.7 Beispiel

Bestimme mit der Methode der kleinsten Quadrate die Regressionsgerade durch die 4 Punkte

$$(0, 1), (1, 3), (2, 4), (3, 4).$$

Lösung: Wir suchen die Koeffizienten a_1, a_2 der Geradengleichung

$$y = a_1 + a_2 x$$

Hierzu haben wir 4 Bestimmungsgleichungen:

$$a_1 + 0 \cdot a_2 = 1$$

$$a_1 + 1 \cdot a_2 = 3$$

$$a_1 + 2 \cdot a_2 = 4$$

$$a_1 + 3 \cdot a_2 = 4$$

Das überbestimmte Gleichungssystem lautet $Ma = y$ mit

$$M = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix}, \quad a = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \quad y = \begin{pmatrix} 1 \\ 3 \\ 4 \\ 4 \end{pmatrix}$$

$$\Rightarrow M^T M = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 4 & 6 \\ 6 & 14 \end{pmatrix}.$$

$(M^T M)^{-1}$ existiert, da $\det(M^T M) = 4 \cdot 14 - 6 \cdot 6 = 20 \neq 0$.

Nach kurzer Rechnung erhält man:

$$(M^T M)^{-1} = \frac{1}{10} \begin{pmatrix} 7 & -3 \\ -3 & 2 \end{pmatrix}$$

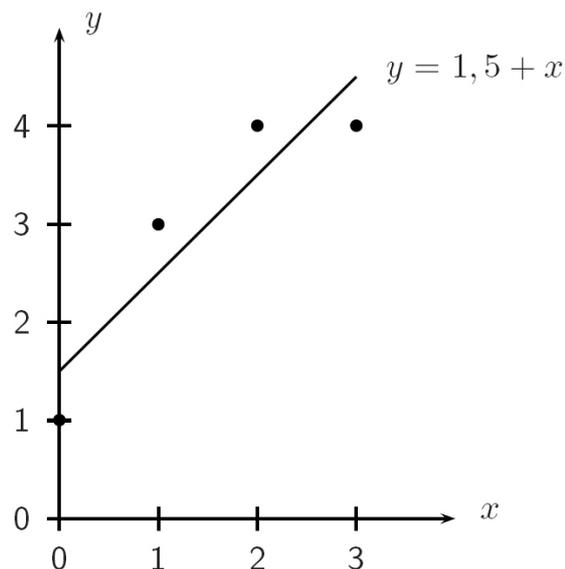
Damit lautet die Pseudolösung von $Ma = y$:

$$a^* = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = (M^T M)^{-1} M^T y$$

$$= \frac{1}{10} \begin{pmatrix} 7 & -3 \\ -3 & 2 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \\ 4 \\ 4 \end{pmatrix}$$

$$= \frac{1}{10} \begin{pmatrix} 7 & -3 \\ -3 & 2 \end{pmatrix} \begin{pmatrix} 12 \\ 23 \end{pmatrix} = \frac{1}{10} \begin{pmatrix} 15 \\ 10 \end{pmatrix} = \begin{pmatrix} 1,5 \\ 1 \end{pmatrix}.$$

Die Ausgleichgerade lautet somit $y = 1,5 + x$.



Kapitel 73

Robuste Statistik

73.1 Motivation

Will man aus realen Daten statistische Parameter schätzen (z.B. arithmetisches Mittel als Schätzer für den Erwartungswert; Parameter einer Regressionskurve), kann es sein, dass das Ergebnis auf Grund von Ausreißern stark verfälscht wird.

Beispiel:

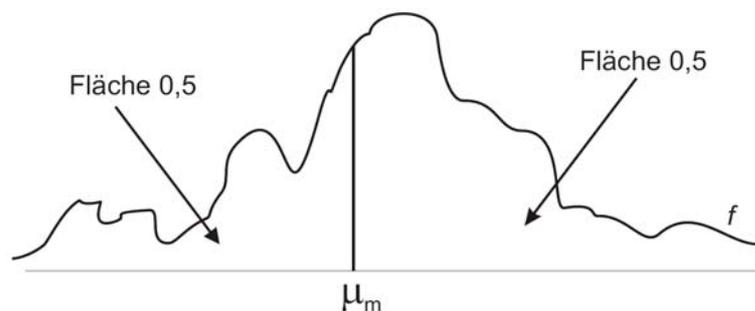
9 Studierende benötigen 10 Semester für ihr Studium, 1 benötigt 40 Semester. Das arithmetische Mittel ergibt eine mittlere Studiendauer von 13 Semestern. Sie ist jedoch nicht repräsentativ für die Mehrzahl der Studierenden.

Gibt es statistische Verfahren, die robuster gegenüber Ausreißern sind?

73.2 Median

Sei X eine Zufallsvariable. Dann nennt man jede Zahl μ_m mit $P(X \geq \mu_m) \geq \frac{1}{2}$ und $P(X \leq \mu_m) \geq \frac{1}{2}$ einen Median von X .

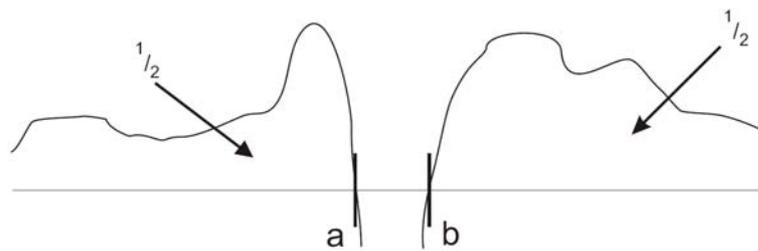
Veranschaulichung für kontinuierliche Zufallsvariable mit Dichte f :



Für die Verteilungsfunktion F gilt: $F(\mu_m) = \frac{1}{2}$.

73.3 Bemerkung

- (a) Nicht immer gibt es einen eindeutigen Median. Gibt es ein Intervall $[a, b]$ mit $P(X \leq a) = \frac{1}{2}$ und $P(X \geq b) = \frac{1}{2}$, so ist jede Zahl aus $[a, b]$ ein Median.



(b) I.A. stimmen Erwartungswert und Median nicht überein.

73.4 Empirischer Erwartungswert

Hat man $2k + 1$, der Größe nach geordnete Messwerte

$$x_1 \leq x_2 \leq \dots \leq x_{2k+1},$$

dann nennt man $\hat{\mu}_m := x_{k+1}$ den (empirischen) Median dieser Daten. Es sind 50% der Daten $\geq \hat{\mu}_m$, und 50% sind $\leq \hat{\mu}_m$. Bei einer geraden Anzahl von Messungen

$$x_1 \leq x_2 \leq \dots \leq x_{2k}$$

gilt für jedes $\hat{\mu} \in [x_k, x_{k+1}]$: $\geq 50\%$ der Daten sind $\geq \hat{\mu}$, $\geq 50\%$ sind $\leq \hat{\mu}$. Man definiert in diesem Fall:

$$\hat{\mu}_m := \frac{1}{2}(x_{k+1} + x_k)$$

als „den“ (empirischen) Median.

73.5 Beispiele

- (a) Der Median der Studiendauer in 73.1 beträgt 10 Semester. Der Ausreißer mit 40 Semestern hat somit keinen Einfluss auf den Median.
- (b) In der Bildverarbeitung ersetzt der Medianfilter einen Grauwert durch seinen Median innerhalb eines $(2k + 1) \times (2k + 1)$ -Fensters:

32	17	24
35	251	21
12	24	25

Ordnen der Grauwerte:

$$12 \leq 17 \leq 21 \leq 24 \leq \underbrace{24}_{\text{Median}} \leq 25 \leq 32 \leq 35 \leq 251$$

Der Grauwert 251 (Ausreißer) wird durch den Median 24 ersetzt. Medianfilter sind robust gegenüber Impulsrauschen (Ausreißern nach oben oder unten) und erhalten Kanten.

73.6 M-Schätzer

Seien x_1, \dots, x_n Messwerte und $\Psi : [0, \infty) \rightarrow \mathbb{R}$ eine monoton wachsende Strafffunktion (engl.: penaliser). Dann nennt man dasjenige μ , das

$$E(X) = \sum_{i=1}^n \Psi(|x - x_i|)$$

minimiert, den M-Schätzer von x_1, \dots, x_n .

73.7 Beispiele

(a) Beliebige ist die Familie $\Psi(s) = s^p$ mit $p \geq 0$. Man kann zeigen:

(i) $p = 2$ liefert das arithmetische Mittel \bar{x} . Es minimiert den quadratischen Abstand

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

(ii) $p = 1$ liefert den Median $\hat{\mu}$. Er minimiert die Abstandssumme

$$E(x) = \sum_{i=1}^n |x_i - x|$$

(iii) $p \rightarrow 0$ liefert als Minimierer die Modalwerte (Maxima des Histogramms)

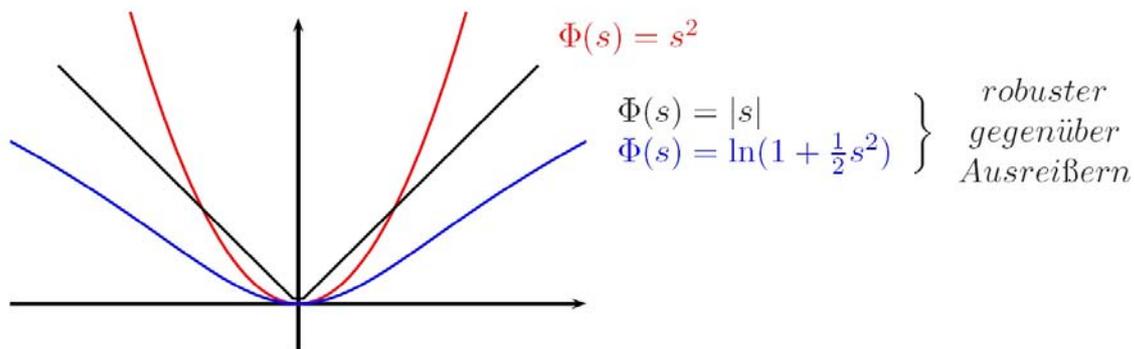
(iv) $p \rightarrow \infty$ ergibt den Midrange

$$\frac{\max\{x_i\} + \min\{x_i\}}{2}.$$

Kleinere Werte für p liefern robustere M-Schätzer, da sie Ausreißer x_i , für die $\Psi(|x_i - x|) = |x_i - x|^p$ groß wird, weniger stark bestrafen.

(b) Eine andere Straffunktion, die robuster als die übliche quadratische Straffunktion $\Psi(s) = s^2$ ist, ist z.B. die Lorentz-Strafffunktion

$$\Psi(s) = \ln\left(1 + \frac{1}{2}s^2\right).$$



Kapitel 74

Fehlerfortpflanzung

74.1 Motivation

Es werden 2 physikalische Größen x und y mehrmals gemessen. Man erhält als Mittelwert \bar{x}, \bar{y} und als (empirische) Standardabweichung s_x, s_y (70.4). Nun will man hieraus eine neue Größe $z = f(x, y)$ berechnen. Als Schätzer für den Erwartungswert verwendet man

$$\bar{z} := f(\bar{x}, \bar{y}).$$

Wie gehen jedoch die Meßfehler s_x, s_y in die Standardabweichung s_z ein? Man kann zeigen:

74.2 Satz: (Fehlerfortpflanzungsgesetz von Gauß)

Schätzt man aus zwei Fehler behafteten Größen x, y mit Mittelwert \bar{x}, \bar{y} und Standardabweichung s_x, s_y eine neue Größe $z = f(x, y)$, so ist der Schätzer für den Erwartungswert gegeben durch

$$\bar{z} := f(\bar{x}, \bar{y})$$

und für die Varianz von z gilt:

$$s_z^2 = \left(\frac{\partial f}{\partial x} \right)^2 \Big|_{\bar{z}} s_x^2 + \left(\frac{\partial f}{\partial y} \right)^2 \Big|_{\bar{z}} s_y^2$$

Beweisidee: Taylorentwicklung.

74.3 Beispiel

Zwei Widerstände R_1, R_2 werden mehrmals gemessen. Man erhält (in Ohm)

$$\begin{array}{ll} \bar{R}_1 = 100 & s_{R_1} = 0,8 \quad (\text{Schreibweise: } R_1 = 100 \pm 0,8) \\ \bar{R}_2 = 200 & s_{R_2} = 1 \quad (R_2 = 200 \pm 1). \end{array}$$

Wie groß sind Gesamtwiderstand R und sein Fehler bei Parallelschaltung?

$$\begin{aligned} \frac{1}{R} &= \frac{1}{R_1} + \frac{1}{R_2} = \frac{R_2 + R_1}{R_1 R_2} \Rightarrow R = \frac{R_1 R_2}{R_1 + R_2} \\ \bar{R} &= \frac{\bar{R}_1 \cdot \bar{R}_2}{\bar{R}_1 + \bar{R}_2} = \frac{100 \cdot 200}{100 + 200} \approx 66,67 \\ \frac{\partial R}{\partial R_1} &= \frac{(R_1 + R_2)R_2 - R_1 R_2}{(R_1 + R_2)^2} = \frac{R_2^2}{(R_1 + R_2)^2} \\ \frac{\partial R}{\partial R_2} &= \frac{R_1^2}{(R_1 + R_2)^2} \\ \left. \frac{\partial R}{\partial R_1} \right|_{\bar{R}} &\approx 0,44 \\ \left. \frac{\partial R}{\partial R_2} \right|_{\bar{R}} &\approx 0,11 \\ s_R &= \sqrt{\left(\left. \frac{\partial R}{\partial R_1} \right|_{\bar{R}} \right)^2 s_{R_1}^2 + \left(\left. \frac{\partial R}{\partial R_2} \right|_{\bar{R}} \right)^2 s_{R_2}^2} \\ &= \sqrt{0,44^2 \cdot 0,8^2 + 0,11^2 \cdot 1^2} \approx 0,37. \\ R &= \bar{R} \pm s_R \approx 66,67 \pm 0,37. \end{aligned}$$

Kapitel 75

Markowketten

75.1 Motivation

Der Zustand eines Systems zur Zeit $n \in \mathbb{N}$ werde durch eine Zufallsvariable X_n beschrieben und soll nur von X_{n-1} abhängen (nicht jedoch von früheren Zuständen X_{n-2}, X_{n-3}, \dots). Wir möchten das zeitliche Verhalten dieses Systems studieren, insbesondere das Langzeitverhalten für $n \rightarrow \infty$.

Prozesse dieser Art sind in der Informatik z.B. bei der Untersuchung der Auslastung von Servern wichtig (Warteschlangenmodelle).

75.2 Definition (Markowkette, Stochastischer Prozess)

Ein stochastischer Prozess ist eine Familie (X_t) von Zufallsvariablen mit $t \in \mathbb{R}$ oder $t \in \mathbb{N}$. Wir denken dabei an t als Zeitparameter, der kontinuierlich oder diskret ist. Ein diskreter stochastischer Prozess (X_n) , $n \in \mathbb{N}$ heißt Markowkette, wenn die Verteilung von X_n bei gegebenen X_{n-1} nicht von den früheren Verteilungen X_k , $k < n - 1$ abhängt:

$$\begin{aligned} P(X_n = i_n \mid X_{n-1} = i_{n-1}, X_{n-2} = i_{n-2}, \dots) \\ = P(X_n = i_n \mid X_{n-1} = i_{n-1}). \end{aligned}$$

Bemerkung: Wichtig sind insbesondere Markowketten, die nur endlich viele Zustände annehmen:

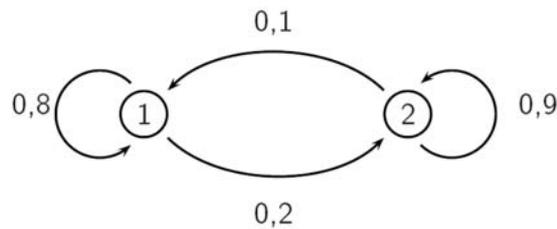
$$P(X_n \in \{1, \dots, k\}) = 1.$$

75.3 Beispiel mit Definition

Jedes Jahr ziehen 10 % der Bevölkerung außerhalb Kaliforniens nach Kalifornien, und 20 % der Bevölkerung Kaliforniens zieht aus. Eine Person befindet sich im Jahr $n - 1$ in einem von 2 Zuständen:

$$X_{n-1} := \begin{cases} 1 & \text{(Person wohnt in Kalifornien)} \\ 2 & \text{(Person wohnt nicht in Kalifornien)}. \end{cases}$$

Der Zustand im Jahr n läßt sich dann durch ein graphisches Modell mit Übergangswahrscheinlichkeiten beschreiben:



Sei p_{ij}^n die Ws., dass ein Zustand j zur Zeit $n - 1$ in den Zustand i übergeht (z.B. $p_{12}^n = 0,1$):

$$p_{ij}^n = P(X_n = i \mid X_{n-1} = j).$$

Wir definieren die Matrix der Übergangswahrscheinlichkeiten (Übergangsmatrix) durch

$$M_n := (p_{ij}^n) \in \mathbb{R}^{k \times k} \quad (\text{bei } k \text{ Zuständen})$$

Im Beispiel:

$$M_n = \begin{pmatrix} 0,8 & 0,1 \\ 0,2 & 0,9 \end{pmatrix}$$

Die Verteilung von X_n auf die Zustände $i = 1, \dots, k$ werde durch einen Vektor $u_n \in \mathbb{R}^k$ beschrieben.

$$u_n = \begin{pmatrix} u_{n1} \\ \vdots \\ u_{nk} \end{pmatrix} \in \mathbb{R}^k.$$

Dann berechnet sich u_n aus u_{n-1} durch:

$$u_n = M_n u_{n-1}$$

Sind im Beispiel zur Zeit $n - 1$ 60 % der Bevölkerung außerhalb Kaliforniens, gilt:

$$\begin{aligned} u_{n-1} &= \begin{pmatrix} 0,4 \\ 0,6 \end{pmatrix} \\ u_n &= \begin{pmatrix} 0,8 & 0,1 \\ 0,2 & 0,9 \end{pmatrix} \begin{pmatrix} 0,4 \\ 0,6 \end{pmatrix} = \begin{pmatrix} 0,38 \\ 0,62 \end{pmatrix} \end{aligned}$$

Im Jahr n sind somit 62 % außerhalb Kaliforniens.

75.4 Definition (Kette mit stationären Übergangswahrscheinlichkeiten)

Eine Markowkette (X_n) heißt homogen, oder Kette mit stationären Übergangswahrscheinlichkeiten, wenn die Übergangsmatrix M_n unabhängig von der Zeit n ist:

$$M_n = M \quad \forall n \in \mathbb{N}.$$

75.5 Bemerkungen

- (a) Beispiel 75.3 beschreibt eine homogene Markowkette.

- (b) Wir nennen eine Matrix $A = (a_{ij}) \in \mathbb{R}^{k \times k}$ eine stochastische Matrix, wenn alle Einträge nichtnegativ sind und die Spaltensummen 1 sind:

$$a_{ij} \geq 0 \quad \forall i, j \in \{1, \dots, n\}$$

$$\sum_{i=1}^k a_{ij} = 1 \quad \forall j \in \{1, \dots, n\}.$$

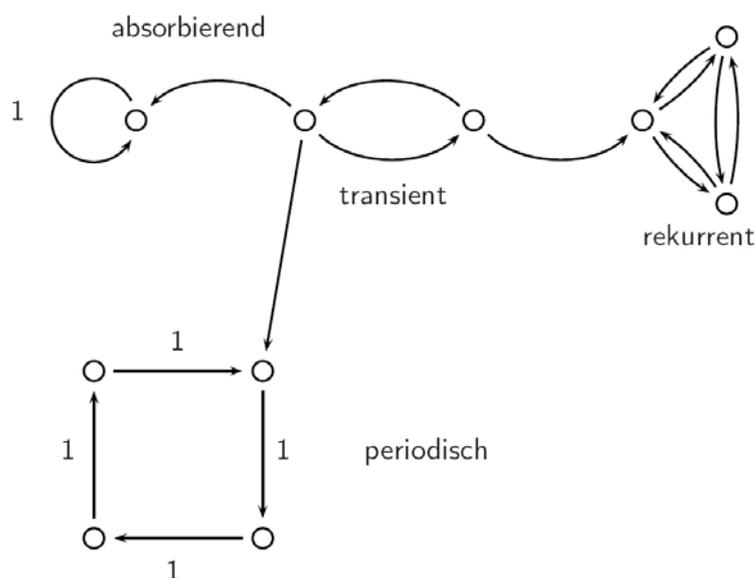
Übergangsmatrizen sind Beispiele für stochastische Matrizen.

- (c) In der Stochastikliteratur werden oft Zeilenvektoren u_n betrachtet, und man definiert p_{ij}^n als die Übergangswahrscheinlichkeit von Zustand i nach Zustand j . Dann ist

$$u_n = u_{n-1} M_n$$

und stochastische Matrizen haben Zeilensumme 1.

75.6 Zustandsbeschreibung endlicher homogener Markowketten



Definition: (*transient, rekurrent, periodisch*)

Sei (X_n) eine endliche homogene Markowkette. Ein Zustand i heißt

- transient, wenn $P(X_m \neq i \quad \forall m > n \mid X_n = i) > 0$
(kann verlassen werden).
- rekurrent, wenn $P(X_m \neq i \quad \forall m > n \mid X_n = i) = 0$
(mit Ws. 1 kehren wir zurück).
- periodisch mit Periode l , wenn $P(X_{n+l} = i \mid X_n = i) = 1$.

Eine Menge I heißt absorbierend, wenn

$$P(X_{n+1} \in I \mid X_n \in I) = 1.$$

Um das Zeitverhalten von Markowketten zu verstehen, müssen wir stochastische Matrizen näher untersuchen.

75.7 Satz: (Eigenwerte stochastischer Matrizen)

Sei $M = (p_{ij}) \in \mathbb{R}^{k \times k}$ eine stochastische Matrix. Dann gilt:

- (a) $\lambda = 1$ ist Eigenwert von M^\top
- (b) $|\lambda| \leq 1$ für alle Eigenwerte λ von M und M^\top .
- (c) $\lambda = 1$ ist einziger Eigenwert von M^\top mit $|\lambda| = 1$, falls $\min_j p_{jj} > 0$.

Beweis

- (a) Ist M stochastisch, so hat $A = M^\top$ Zeilensumme 1.

$$\Rightarrow A \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} \sum_j a_{1j} \\ \vdots \\ \sum_j a_{kj} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}.$$

Somit ist $\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ Eigenvektor von A zum Eigenwert $\lambda = 1$.

- (b) Betrachte die Spaltensummennorm

$$\|M\|_s := \max_j \left(\sum_{i=1}^k |p_{ij}| \right) = 1.$$

Dann gilt nach Satz 50.8 für jeden Eigenwert λ von M :

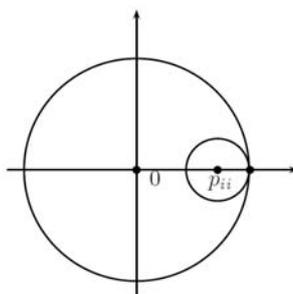
$$|\lambda| \leq \|M\|_s = 1.$$

Für M^\top betrachtet man die Zeilensummennorm.

- (c) Nach dem Satz von Gerschgorin (50.10) gibt es zu jedem Eigenwert λ von $M^\top = (a_{ij})$:

$$|\lambda - p_{ii}| \leq \sum_{\substack{j=1 \\ j \neq i}}^k |a_{ij}| = 1 - p_{ii}.$$

Also liegt λ in dem Kreis mit Mittelpunkt p_{ii} und Radius $1 - p_{ii}$. Er berührt den Einheitskreis von innen in $(1, 0)$:



Aus $|\lambda| = 1$ folgt somit: $\lambda = 1$.

Bemerkung: Aussage 75.7.(c) gilt auch für M statt M^\top .

75.8 Bedeutung des Eigenwerts $\lambda = 1$

Wir interessieren uns für das Verhalten einer endlichen homogenen Markowkette (X_n) für $n \rightarrow \infty$. Für einen Anfangszustand u_0 und eine Übergangsmatrix $M = (p_{ij})$ gilt:

$$\begin{aligned} u_1 &= Mu_0 \\ u_2 &= Mu_1 = M^2u_0 \\ &\vdots \\ u_n &= M^n u_0. \end{aligned}$$

Ist u_0 Eigenvektor von M zum Eigenwert $\lambda = 1$, gilt:

$$u_n = M^n u_0 = \lambda u_0 = u_0 \quad \forall n \in \mathbb{N},$$

d.h. der Zustand u_0 ist stabil.

Darüberhinaus kann man Folgendes zeigen:

75.9 Satz: (Potenzen stochastischer Matrizen)

Sei M eine stochastische Matrix. Genau dann existiert $\lim_{n \rightarrow \infty} M^n$, wenn 1 der einzige Eigenwert von M mit Betrag 1 ist.

Beweis eines Teilresultats:

Wir zeigen:

Sei λ ein Eigenwert von M mit $|\lambda| = 1$. Falls $\lim_{n \rightarrow \infty} M^n$ ex., ist 1 einziger Eigenwert von M (vom Betrag 1).

Sei $v \neq 0$ ein Eigenvektor von M zum Eigenwert λ mit $|\lambda| = 1$:

$$Mv = \lambda v \quad \Rightarrow \quad M^n v = \lambda^n v \quad \forall n \in \mathbb{N}.$$

Würde $\lim_{n \rightarrow \infty} M^n$ existieren, hätten wir

$$\left(\lim_{n \rightarrow \infty} M^n \right) v = \lim_{n \rightarrow \infty} (M^n v) = \lim_{n \rightarrow \infty} (\lambda^n v) = \left(\lim_{n \rightarrow \infty} \lambda^n \right) v$$

Also würde auch $\mu := \lim_{n \rightarrow \infty} \lambda^n$ existieren. Dann wäre auch

$$\mu = \lim_{n \rightarrow \infty} \lambda^{n+1} = \lambda \lim_{n \rightarrow \infty} \lambda^n = \lambda \cdot \mu.$$

Wegen $|\lambda| = 1$ ist auch $|\lambda^n| = 1$ und somit $|\mu| = 1 \neq 0$. Aus $\mu = \lambda\mu$ folgt dann (nach Division durch μ): $\lambda = 1$. □.

75.10 Gegenbeispiel

Wir betrachten eine stochastische Matrix, die Eigenwert λ mit $|\lambda| = 1$, aber $\lambda \neq 1$ besitzt:

$$M = \begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 1 & & & 0 \end{pmatrix} \in \mathbb{R}^{k \times k}.$$

Sie beschreibt eine zyklische Vertauschung.

Sei $\alpha := e^{2\pi i/k} = \cos \frac{2\pi}{k} + i \sin \frac{2\pi}{k}$ ($\Rightarrow \alpha^k = e^{2\pi i} = 1$)

Setzen wir

$$v_j := \begin{pmatrix} 1 \\ \alpha^j \\ \alpha^{2j} \\ \vdots \\ \alpha^{(k-1)j} \end{pmatrix}, \quad 0 \leq j \leq k-1,$$

so folgt wegen $\alpha^k = 1$:

$$Mv_j = \begin{pmatrix} \alpha^j \\ \alpha^{2j} \\ \vdots \\ \alpha^{kj} \end{pmatrix} = \alpha^j v_j.$$

Also sind $1, \alpha^1, \alpha^2, \dots, \alpha^{k-1} \in \mathbb{C}$ sämtliche Eigenwerte von M . Alle haben Betrag 1.

75.11 Markowketten im Gleichgewicht

Eine endliche homogene Markowkette (X_n) ist im Gleichgewicht, falls zu ihrer Übergangsmatrix $M = (p_{ij})$ ein Zustand u existiert mit

$$Mu = u,$$

d.h. u ist Eigenvektor von M zum Eigenwert 1, und es gilt:

$$\sum_{i=1}^k u_i = 1, \quad u_i \geq 0 \quad \text{für } i = 1, \dots, k.$$

75.12 Beispiel

Im Beispiel 75.3 war die Übergangsmatrix gegeben durch

$$M = \begin{pmatrix} 0,8 & 0,1 \\ 0,2 & 0,9 \end{pmatrix}$$

Berechnung der Eigenwerte:

$$\begin{aligned} 0 &\stackrel{!}{=} \begin{vmatrix} 0,8 - \lambda & 0,1 \\ 0,2 & 0,9 - \lambda \end{vmatrix} = (0,8 - \lambda)(0,9 - \lambda) - 0,02 \\ &= 0,72 - 0,8\lambda - 0,9\lambda + \lambda^2 - 0,02 \\ &= \lambda^2 - 1,7\lambda + 0,7 \\ \lambda_{1/2} &= \frac{1,7 \pm \sqrt{1,7^2 - 4 \cdot 0,7}}{2} = \frac{1,7 \pm \sqrt{2,89 - 2,8}}{2} \\ &= \frac{1,7 \pm 0,3}{2} \\ \lambda_1 &= 1, \\ \lambda_2 &= 0,7 \end{aligned}$$

Eigenvektor zu $\lambda_1 = 1$:

$$\begin{aligned} \begin{pmatrix} -0,2 & 0,1 \\ 0,2 & -0,1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \Rightarrow \quad -0,2x + 0,1y = 0 \\ &\quad \Rightarrow \quad x = \frac{y}{2} \\ v &= \begin{pmatrix} \alpha \\ 2\alpha \end{pmatrix}, \quad \alpha \neq 0. \end{aligned}$$

Da v eine Verteilung auf die Zustände beschreiben soll, muss gelten:

$$v_i \geq 0, \quad \sum_{i=1}^n v_i = 1$$

$$\Rightarrow v = \begin{pmatrix} 1/3 \\ 2/3 \end{pmatrix}.$$

Dies beschreibt den Gleichgewichtszustand.

$\frac{1}{3}$ der Personen wohnt in Kalifornien, $\frac{2}{3}$ außerhalb.

Wann ist es möglich, bei einer Markowkette von einem Zustand in einen beliebigen anderen zu gelangen?

75.13 Definition (irreduzibel)

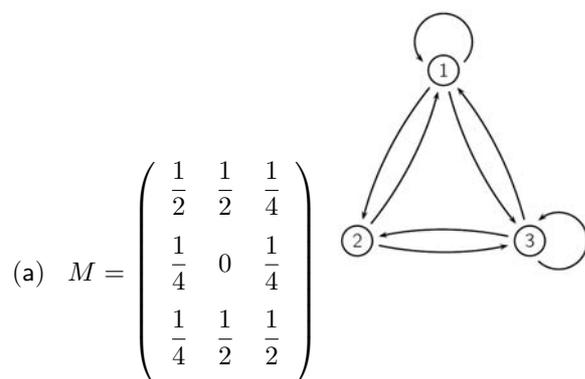
Eine stochastische Matrix $M = (p_{ij}) \in \mathbb{R}^{k \times k}$ heißt transitiv (irreduzibel), wenn man von jedem Zustand n zu jedem anderen Zustand m mit positiver Wahrscheinlichkeit in endlich vielen Schritten gelangen kann: Es gibt ein r , so dass für $M^r = B = (b_{ij})$ gilt:

$$b_{mn} > 0.$$

75.14 Praktisches Kriterium für Irreduzibilität

Man zeichnet zur Übergangsmatrix $M = (p_{ij})$ einen gewichteten Graphen: Ist $p_{ij} > 0$, zeichnet man einen Pfeil von j nach i . Falls man von jedem Zustand längs solcher Pfeile zu jedem anderen Zustand gelangt, ist M transitiv, andernfalls nicht.

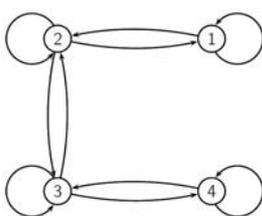
75.15 Beispiele



irreduzibel: Es gibt von jedem Punkt einen Weg zu jedem anderen Punkt.

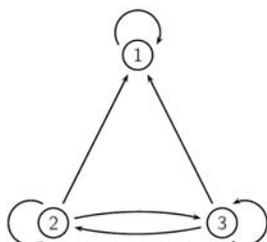
Beachte: Der Weg darf mehrere Pfeile umfassen. Daher führt auch ein Weg von ② nach ②.

(b) $M = \begin{pmatrix} \frac{3}{4} & \frac{1}{4} & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 \\ 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & \frac{1}{4} & \frac{3}{4} \end{pmatrix}$



irreduzibel.

(c) $M = \begin{pmatrix} 1 & \frac{2}{5} & \frac{1}{4} \\ 0 & \frac{1}{5} & \frac{1}{4} \\ 0 & \frac{2}{5} & \frac{1}{2} \end{pmatrix}$



reduzibel: Es führt z.B. kein Weg von ① nach ②.

Kapitel 76

Verborgene Markowmodelle

76.1 Motivation

Verborgene Markowmodelle (hidden Markov models, HMMs) spielen eine große Rolle in der Bioinformatik, der Spach- und der Mustererkennung. Mit Hilfe der aus dem Bereich der Markowketten bekannten Übergangsmatrizen (§75) sucht man nach der wahrscheinlichsten Zustandsfolge, die eine gegebene Beobachtung erzeugt haben könnte.

76.2 Beispiel: Spieler mit fairer und unfairer Münze

Ein Spieler besitzt eine faire Münze, bei der Kopf und Zahl gleich wahrscheinlich sind ($p^+(0) = p^+(1) = \frac{1}{2}$) und eine gezinkte Münze, bei der Zahl wahrscheinlicher ist ($p^-(0) = \frac{1}{4}, p^-(1) = \frac{3}{4}$). Wir beobachten n Münzwürfe und wollen entscheiden, ob er die faire oder die unfaire Münze genommen hat. Sei k die Zahl der Ergebnisse, bei denen Zahl beobachtet wurde. Die einzelnen Ergebnisse seien x_1, \dots, x_n .

Fall 1: Münze war fair.

Wahrscheinlichkeit, dass die Beobachtung $x = (x_1, \dots, x_n)$ eintritt, ist

$$P(x \mid \text{faire Münze}) = \prod_{i=1}^n p^+(x_i) = \frac{1}{2^n}.$$

Fall 2: Münze war unfair.

$$\underbrace{P(x \mid \text{unfaire Münze})}_{\prod_{i=1}^n p^-(x_i)} = \underbrace{\left(\frac{1}{4}\right)^{n-k}}_{n-k \text{ mal Kopf}} \underbrace{\left(\frac{3}{4}\right)^k}_{k \text{-mal Zahl}} = \frac{3^k}{4^n}.$$

Wir vermuten, dass die Münze fair war, falls

$$\begin{aligned} P(x \mid \text{faire Münze}) &> P(x \mid \text{unfaire Münze}) \\ \frac{1}{2^n} &> \frac{3^k}{4^n} && | \cdot 4^n \\ 2^n &> 3^k && | \cdot \log_2 \\ n &> k \log_2 3 \\ \frac{k}{n} &< \frac{1}{\log_2 3} && (*). \end{aligned}$$

76.3 Schwierigeres Beispiel

Wir nehmen an, dass der Spieler mit einer geringen Wahrscheinlichkeit von 0,1 die Münze innerhalb des Spiels austauscht. Wir wollen wissen, wann er welche Münze genommen hat.

Naiver Ansatz:

Wir betrachten die Beobachtungen innerhalb eines Fensters und überprüfen ob (*) zutrifft oder nicht.

Problem:

Lösung hängt von Fenstergröße ab und wir wissen nicht, wann der Spieler die Münze wechselt.

Verborgene Markowmodelle bieten einen alternativen stochastischen Zugang, um dieses Problem zu lösen.

76.4 Definition (Verborgenes Markowmodell)

Ein verborgenes Markowmodell (hidden Markov model, HMM) ist ein Tupel $\mathcal{M} = (\Sigma, Q, M, E)$. Dabei bezeichnen:

- Σ : Alphabet von Symbolen (Ann.: $|\Sigma| = r$)
- Q : Menge von Zuständen mit Symbolen aus Σ (Ann.: $|Q| = k$)
- $M = (p_{ij})$: stochastische Matrix mit Übergangsws.
 $M \in \mathbb{R}^{k \times k}$
- $E \in \mathbb{R}^{k \times r}$: Matrix der Emissionswahrscheinlichkeiten
 $E = (e, (b))$.

76.5 Beispiel

In Beispiel 76.3 bezeichnen:

- $\Sigma = \{0, 1\}$ Kopf (0) oder Zahl (1)
- $Q = \{f, b\}$ faire (f) oder unfaire (b=biased) Münze
- $M = \begin{pmatrix} 0,9 & 0,1 \\ 0,1 & 0,9 \end{pmatrix}$ Übergangsmatrix für Münzwechsel
- $e_f(0) = \frac{1}{2}$ $e_f(1) = \frac{1}{2}$ Emissionsws. für faire Münze
- $e_b(0) = \frac{1}{4}$ $e_b(1) = \frac{3}{4}$ Emissionsws. für unfaire Münze

Die vom Spieler tatsächlich verwendete Zustandsfolge ist $\pi = (\pi_1, \dots, \pi_n)$ mit $\pi_i \in Q$. Sie bleibt uns „verborgen“. Wir beobachten nur den Ergebnisvektor $x(x_1, \dots, x_n)$ mit $x_i \in \Sigma$.

Die Wahrscheinlichkeit, dass die beobachtete Sequenz x durch einen Pfad π erzeugt wurde, ist

$$P(x | \pi) = \prod_{i=1}^n P(x_i | \pi_i) \cdot \underbrace{P(\pi_i | \pi_{i+1})}_{\text{Übergangsws.}} \quad \text{mit } P(\pi_n | \pi_{i+1}) := 1.$$

76.6 Problemstellung

Gelöst werden soll nun das Dekodierungsproblem:

Finde zu einem gegebenen HMM $\mathcal{M} = (\Sigma, Q, M, E)$ und einer Beobachtung $x = (x_1, \dots, x_n)$ einen optimalen Pfad $\pi^* = (\pi_1^*, \dots, \pi_n^*)$, der $P(x | \pi)$ maximiert:

$$\pi^* = \arg \max_{\pi} P(x | \pi)$$

76.7 Der Viterbi-Algorithmus

Grundidee:

optimaler Pfad $(\pi_1^*, \dots, \pi_{i+1}^*)$ zur Beobachtung (x_1, \dots, x_{i+1}) ist optimal unter allen Pfaden, die in dem unbekanntem Zustand $\pi_i^* = m \in Q$ enden.

$s_m(i)$: Wahrscheinlichkeit, dass optimaler Pfad $(\pi_1^*, \dots, \pi_i^*)$ zur Beobachtung (x_1, \dots, x_i) in m endet ($1 \leq i \leq n$).

$$\Rightarrow s_l(i+1) = \underbrace{e_l(x_{i+1})}_{\text{Emissionsw. für } l} \cdot \underbrace{\max_{m \in Q} \{s_m(i) \cdot p_{lm}\}}_{\text{Übergang } m \rightarrow l} \quad (**)$$

Initialisiere:

$$\begin{aligned} s_o(0) &= 1 \\ 2_m(0) &= 0 \quad m \in \{1, \dots, n\}. \end{aligned}$$

Im optimalen π^* gilt $P(x | \pi^*) = \max_{m \in Q} \{s_m(i) \cdot p_{nm}\}$.

Das Viterbi-Verfahren benötigt $O(nk)$ Rechenoperationen.

Kapitel 77

Pseudozufallszahlen und Monte-Carlo-Simulation

77.1 Motivation

Es gibt Anwendungsgebiete (z.B. Ray Tracing in der Computergrafik, Strömungssimulation im Bereich Computational Fluid Dynamics, Berechnung hochdimensionaler Integrale), bei denen stochastische Simulationen einfache oder effiziente Alternativen zu deterministischen Algorithmen sind. Solche Simulationen nennt man auch Monte-Carlo-Verfahren. Sie benötigen die Erzeugung von Zufallszahlen auf dem Rechner. Da funktionierende Computer jedoch deterministisch arbeiten, verwendet man statt dessen Algorithmen, die Zahlen liefern, die echten Zufallszahlen ähneln. Solche Zahlen heißen Pseudozufallszahlen.

77.2 Erzeugung gleichverteilter Pseudozufallszahlen

Ziel: Erzeugung einer Sequenz Z_n von gleichverteilten Pseudozufallszahlen aus $[0, 1]$.
Beliebte Vorgehensweise: Lineare Kongruenzmethoden (in vielen Compilern verwendet)

	$m \in \mathbb{N}$	Modus
Geg.:	$a \in \{1, \dots, m-1\}$	Multiplikator
	$b \in \{0, \dots, m-1\}$	Inkrement
	$x_0 \in \{0, \dots, m-1\}$	Startwert

Verfahren:

$$\begin{aligned}x_n &:= a \cdot x_{n-1} + b && (\text{modulo } m) \\Z_n &:= \frac{x_n}{m}\end{aligned}$$

Dann approximiert die Sequenz (Z_n) eine Folge von stochastisch unabhängigen Zufallsvariablen, die auf $[0, 1]$ gleichverteilt sind. Die Approximationsgüte hängt von der Parameterwahl ab.

Klar: Nach spätestens m Schritten wiederholt sich die Folge.

Häufig verwendet, aber schlecht:

$$\begin{aligned}m &= 2^{16} = 65536, \\a &= 25173, \\b &= 13849\end{aligned}$$

Besser (Minimalstandard):

$$\begin{aligned} m &= 2^{31} - 1 = 2147483647, \\ a &= 7^5 = 16807, \\ b &= 0 \end{aligned}$$

77.3 Erzeugung von $N(0, 1)$ -verteilten Pseudozufallszahlen

Die Standardnormalverteilung im \mathbb{R}^2 hat die Dichte

$$\varphi(x, y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}$$

Für eine Kreisscheibe $K_t(0)$ um 0 mit Radius t und zwei stochastisch unabhängig $N(0, 1)$ -verteilte Variablen X, Y gilt:

$$P(X^2 + Y^2 \leq t^2) = \int_{K_t(0)} \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} dx dy$$

Geht man zu Polarkoordinaten (r, φ) über, ergibt sich mit

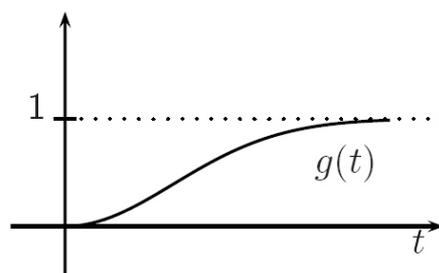
$$\begin{aligned} x &= r \cos \varphi = f_1(r, \varphi) \\ y &= r \sin \varphi = f_2(r, \varphi) \\ Jf(r, \varphi) &= \begin{pmatrix} \frac{\partial f_1}{\partial r} & \frac{\partial f_1}{\partial \varphi} \\ \frac{\partial f_2}{\partial r} & \frac{\partial f_2}{\partial \varphi} \end{pmatrix} = \begin{pmatrix} \cos \varphi & -r \sin \varphi \\ \sin \varphi & r \cos \varphi \end{pmatrix} \\ \det(Jf(r, \varphi)) &= r \cos^2 \varphi + r \sin^2 \varphi = r. \end{aligned}$$

Mit der Transformationsregel (§60):

$$\begin{aligned} P(X^2 + Y^2 \leq t^2) &= \int_{\varphi=0}^{2\pi} \int_{r=0}^t \frac{1}{2\pi} e^{-\frac{r^2}{2}} r dr d\varphi \\ &= 2\pi \int_{r=0}^t \frac{1}{2\pi} e^{-\frac{r^2}{2}} r dr \\ &= \left[e^{-\frac{r^2}{2}} \right]_0^t = 1 - e^{-t^2/2} \end{aligned}$$

Der Radius R des Zufallszahlenvektors (X, Y) erfüllt also:

$$P(R \leq t) = 1 - e^{-t^2/2} =: g(t)$$



Wertebereich von g : $[0, 1)$. Umkehrfunktion zu $g(t)$:

$$g^{-1}(z) = \sqrt{-2 \ln(1-z)}$$

Hat man eine in $[0, 1]$ gleichverteilte Zufallsvariable Z_1 , ergibt sich als Zufallsvariable R für den Radius der 2D-Standardnormalverteilung

$$R = \sqrt{-2 \ln(1-Z_1)}$$

Aus einer zweiten, auf $[0, 1]$ gleichverteilten Zufallsvariablen Z_2 erhält man als Zufallsvariable für einen auf $[0, 2\pi]$ gleichverteilten Winkel T :

$$T = 2\pi Z_2.$$

Dies motiviert folgenden Algorithmus (Box-Muller-Verfahren) zur Erzeugung zweier $N(0, 1)$ -verteilter Zufallszahlen X, Y aus zwei auf $[0, 1]$ gleichverteilten Zufallszahlen Z_1, Z_2 :

R	$:=$	$\sqrt{-2 \ln(1-Z_1)}$
T	$:=$	$2\pi Z_2$
X	$:=$	$R \cos T$
Y	$:=$	$R \sin T$

Bem.: Benötigt man $N(\mu, \sigma^2)$ -verteilte Zufallszahlen \tilde{X}, \tilde{Y} , setzt man (vgl. 65.15):

$$\begin{aligned}\tilde{X} &= \mu + \sigma X \\ \tilde{Y} &= \mu + \sigma Y\end{aligned}$$

(Pseudo-)Zufallszahlen benötigt man z.B. bei probabilistischen Algorithmen:

77.4 Beispiel: Quicksort

Sortieren einer Liste mit Quicksort:

- Suche ein zufälliges Element z der Liste (mit gleichverteilten Pseudozufallszahlen aus 77.2).
- Sortiere Teilliste mit Elementen $\leq z_1$ mit Quicksort
- Sortiere Teilliste mit Elementen $> z_1$ mit Quicksort

usw.

Quicksort hat eine mittlere Laufzeit von $O(n \log n)$, ist also nicht schlecht.

77.5 Beispiel: Buffon'sches Nadelexperiment

Ein probabilistischer Algorithmus zur Bestimmung von π :

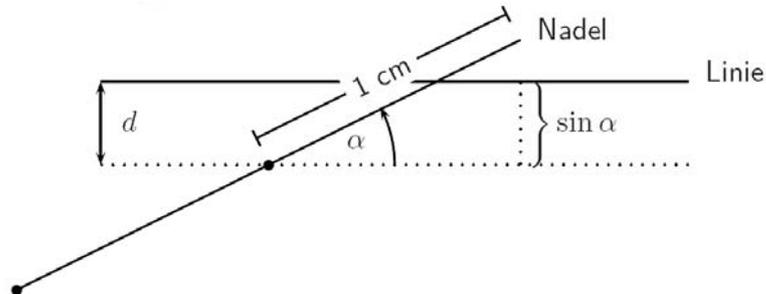
- Zeichne auf einem Blatt Papier parallele Linien im Abstand einer Stecknadellänge
- Lasse die Stecknadel auf das Papier fallen und überprüfe, ob sie eine der Linien trifft. Denke dabei fest an π .
- Zähle die Zahl N der Versuche und die Zahl T der Treffer. Dann gilt:

$$\frac{N}{T} \rightarrow \frac{\pi}{2} \quad \text{für } N \rightarrow \infty.$$

Warum funktioniert das?

Ann.: Nadellänge und Linienabstand seien 2 [cm].

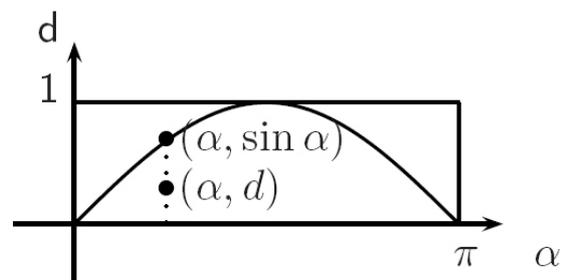
Die Nadel kann nur die nächst gelegene Linie schneiden und nur dann, wenn ihr Abstand d zwischen Nadelmittle und Linie $d < 1$ erfüllt:



Nadel schneidet Linie $\Leftrightarrow d < \sin \alpha$.

Zu jedem Wurf gehört ein Wertepaar $(\alpha, d) \in [0, \pi] \times [0, 1]$.

$d < \sin \alpha \Leftrightarrow (\alpha, d)$ liegt unter dem Graphen von $\sin x$



Im Zufallsexperiment sind die Punkte (α, d) gleich verteilt auf $[0, \pi] \times [0, 1]$:

$$\Rightarrow \frac{T}{N} \rightarrow \frac{\int_0^{\pi} \sin x \, dx}{\pi \cdot 1} = \frac{[-\cos x]_0^{\pi}}{\pi} = \frac{2}{\pi}$$

Unser Experiment war also ein stochastisches Integrationsverfahren für $\int_0^{\pi} \sin x \, dx$.

Im 1D-Fall sind solche Verfahren ineffizient, aber bei hochdimensionalen Integrationsproblemen haben sie eine bessere Komplexität als deterministische numerische Ansätze.