SS 2013

Diskrete Wahrscheinlichkeitstheorie

Javier Esparza

Fakultät für Informatik
TU München

http://www7.in.tum.de/um/courses/dwt/ss13

Sommersemester 2013

Vorlesungsinhalt

- Diskrete Wahrscheinlichkeitsräume
  - Ereignisse, Zufallsvariablen, Unabhängigkeit
  - Binomial-, geometrische-, Poisson-Verteilung
  - Ungleichungen von Markov und Chebyshev, Chernoff-Schranken
- Kontinuierliche Wahrscheinlichkeitsräume
  - Normalverteilung, Exponentialverteilung
  - Zentraler Grenzwertsatz
- Induktive Statistik
  - Schätzvariablen
  - Konfidenzintervalle
  - Testen von Hypothesen
- Markov-Ketten
  - Ankunftsw'keiten, stationäre Verteilung
  - Warteschlangen

Literatur I

- N. Henze:
  Stochastik für Einsteiger, 9. Auflage
  Vieweg+Teubner, 2012
- T. Schickinger, A. Steger:
  Diskrete Strukturen - Band 2,
  Springer Verlag 2001
- M. Greiner, G. Tinhofer:
  Stochastik für Informatiker,
  Carl Hanser Verlag, 1996
- H. Gordon:
  Discrete Probability,
  Springer-Verlag, 1997

Literatur II

- R. Motwani, P. Raghavan:
  Randomized Algorithms,
  Cambridge University Press, 1995
- L. Fahrmeir, R. Künstler, I. Pigeot, G. Tutz:
  Statistik - Der Weg zur Datenanalyse,
  Springer-Verlag, 1997

Was bedeutet Zufall?

- Der Heilige Augustinus über die Zeit:

  > *Was also ist die Zeit? Wenn niemand mich danach fragt, weiß ich es; wenn ich es jemand auf seine Frage hin erklären will, weiß ich es nicht.*

- Pragmatische Sicht: **Zufall = mangelndes Wissen**.
  Bei gewissen Vorgängen wissen wir für eine sichere Vorhersage nicht genug: Ziehung von Lottozahlen, Regenfall über München, Absturz eines informatischen Systems.
- Es können trotzdem probabilistische Vorhersagen gemacht werden.
- Ziel der Vorlesung: lernen, korrekte Vorhersagen dieser Art zu machen und zu identifizieren.

Es wird gewürfelt . . . I

Am Ende der Vorlesung werden Sie in der Lage sein, folgende Fragen zu beantworten:

(1) Ein fairer Würfel wird 10 Mal geworfen. Mit welcher W'keit kriegt man
   - keine Sechs? (≈ 0.162)
   - genau 1 Mal eine Sechs? (≈ 0.323)
   - genau 3 Mal eine Sechs? (≈ 0.155)

(2) Ein fairer Würfel wird 600000 Mal geworfen. Mit welcher W'keit kriegt man mindestens 100100 Sechser? (≈ 0.36)

(3) Drei faire Würfel werden gleichzeitig geworfen. Mit welcher W'keit
   - beträgt die Summe der Augenzahlen mindestens 14? ($35/216$)
   - ist die größte Augenzahl höchstens 3? ($1/8$)
   - sind alle drei Augenzahlen verschieden? ($5/9$)

Es wird gewürfelt . . . II

(4) Sechs faire Würfel werden mit den Zahlen 1 bis 6 nummeriert und dann gleichzeitig geworfen. Wieviele Würfel fallen im Mittel mit der eigenen Nummer als Augenzahl? (1.2)

(5) Hundert faire Würfel werden gleichzeitig geworfen. Mit welcher W'keit beträgt die Summe der Augenzahlen höchstens 310? (≈ 0.01)

(6) Drei faire Würfel werden 1000 Mal gleichzeitig geworfen. Mit welcher W'keit kriegt man kein einziges Mal drei Sechser? (≈ 0.01)

(7) 10 faire Würfel werden mit den Zahlen 1 bis 10 numeriert und dann 10000 Mal gleichzeitig geworfen. Das Ergebnis eines Wurfes wird als Vektor $(a_1, \ldots, a_{10})$ mit $1 \leq a_i \leq 6$ dargestellt. Mit welcher W'keit wiederholt sich kein Ergebnis? (≈ 0.437)

Es wird gewürfelt ... III

(8) Wie oft muss man einen fairen Würfel im Mittel werfen, um

- die erste Sechs
- drei Sechser hintereinander
- alle sechs Augenzahlen mindestens einmal

zu kriegen? ($6, 258, 14.7$)

(9) Anna versucht, das Ergebnis eine Reihe von Würfen eines fairen Würfels zu raten. Der Würfel wird 600 Mal geworfen. Wie oft muss Anna das richtige Ergebnis raten, so dass die W'keit, das es aus Zufall geschieht, höchstens 1% beträgt? (150)

(10) Ein Würfel wird 15 Mal geworfen und man bekommt 1 Mal die eins, 3 Mal die 2, 4 Mal die 3, 3 Mal die 4, 4 Mal die 5 und 0 Mal die 6. Mit welcher Konfidenz kann behauptet werden, der Würfel sei nicht fair? (0.95)

Es wird gewürfelt . . . IV

Zwei der Ergebnisse in (1)-(10) sind falsch.
Sie können schon heute diese zwei
Ergebnisse experimentell bestimmen.

Zufall in der Informatik

- Hardware-Fehler können durch Zufallsvorgänge (z.B. Strahlung) eintreten.

  Zuverlässigkeitsanalyse

- Download-Zeiten von Web-Seiten, Antwort-Zeiten von Diensten können nicht präzise vorhergesagt werden.

  Leistungsanalyse

- Viele Programme verwenden Zufallsbits, um mit großer Wahrscheinlichkeit gewisse Problemsituationen (z.B. Verklemmungen) zu vermeiden:

  Randomisierte Algorithmen

- Viele Sicherheitsprotokolle beruhen auf Annahmen über die W'keit, dass ein Angreifer Information über Schlüsseln gewinnt:

  Sicherheitsanalyse

Thema der Vorlesung: Zufallsexperimente

Das Experiment wird unter genau festgelegten Versuchsbedingungen durchgeführt.

- Ungenau: zwei Zahlen werden zufällig gewählt.
- Genau: ein Paar $(x, y) \in [0, 2^{32} - 1] \times [0, 2^{32} - 1]$ wird zufällig gewählt, jedes Paar hat die gleiche W'keit.

Die Menge der möglichen Ergebnisse (Ausgänge) ist vor der Durchführung des Experimentes bekannt.

Das Experiment kann zumindest prinzipiell beliebig oft unter gleichen Bedingungen wiederholt werden.

Mathematische Modellierung von Zufallsexperimenten

Drei faire Würfel werden geworfen.
Frage: Was ist die W'keit, dass die Summe der Augenzahlen größer als 12 ist?

- Die Menge der möglichen Ausgänge wird aus der Beschreibung gewonnen.
  Ausgänge: $(1,1,1),\ldots,(6,6,6)$
- Die W'keiten der Ausgänge werden ebenfalls aus der Beschreibung gewonnen.
  $\Pr[(1,1,1)] = \ldots = \Pr[(6,6,6)] = \frac{1}{216}$
- Die Frage, die uns interessiert, wird auf die Berechnung der W'keit einer Menge von möglichen Ausgängen reduziert.
  $\Pr[\{(5,4,4),(4,5,5),\ldots,(6,6,6)\}]$
- Zwei Fehlerquellen: falsche Modelle und falsche Berechnungen!

Mathematische Modellierung von Zufallsexperimenten II

Die Reduktion der Frage auf W'keiten ist nicht immer so direkt.

Drei faire Würfel werden geworfen.
Frage: Sollte man folgende Wette eingehen? Wenn insgesamt über 12 Augenzahlen angezeigt werden, gewinnt man 1 Euro. Sonst verliert man 2 Euro.

Teil I

Grundbegriffe

1. Diskrete Wahrscheinlichkeitsräume

Diskrete Wahrscheinlichkeitskeitsräume

Definition 1

1. Ein diskreter W'keitsraum ist durch eine (endliche oder abzählbar unendliche) Ergebnismenge $\Omega = \{\omega_1, \omega_2, \ldots\}$ von Elementarereignissen gegeben.
2. Jedem Elementarereignis ω_i ist eine Wahrscheinlichkeit $\Pr[\omega_i]$ zugeordnet, wobei wir fordern, dass $0 \le \Pr[\omega_i] \le 1$ und
$$\sum_{\omega \in \Omega} \Pr[\omega] = 1.$$
3. Eine Menge $E \subseteq \Omega$ heißt Ereignis. Die Wahrscheinlichkeit $\Pr[E]$ eines Ereignisses ist definiert durch
$$\Pr[E] := \sum_{\omega \in E} \Pr[\omega].$$

Beispiel 2

Zwei faire Würfel werden geworfen.

Frage: Mit welcher W'keit beträgt die Gesamtzahl der Augen mindestens 10?

Modell 1: $\Omega = \{\ (1,1), \ldots, (6,6)\ \}$, $\Pr[(i,j)] = \dfrac{1}{36}$.

mind-10 $= \{(4,6), (5,5), \ldots, (6,6)\}$, $\Pr[\text{mind-10}] = \dfrac{1}{6}$.

Modell 2: $\Omega = \{2, 3, 4, \ldots, 10, 11, 12\}$,
$\Pr[2] = \dfrac{1}{36}, \ldots, \Pr[7] = \dfrac{1}{6}, \ldots, \Pr[12] = \dfrac{1}{36}$.

mind-10 $= \{10, 11, 12\}$, $\Pr[\text{mind-10}] = \dfrac{3}{26} + \dfrac{2}{36} + \dfrac{1}{36} = \dfrac{1}{6}$.

Modelle II

Modell 3: $\Omega = \{2, 3, 4, \ldots, 10, 11, 12\}$, $\Pr(i) = \frac{1}{11}$.

$\Pr[\text{mind-10}] = \frac{3}{11}$.

Model 3 ist aus mathematischer Sicht zulässig, jedoch falsch. Es entspricht nicht dem, was man unter "faire Würfel" versteht!

Physikalische Analogie

Elementarereignis ω	$\rightarrow$	Objekt
W'keit $\Pr[\omega]$	$\rightarrow$	Gewicht des Objekts
$\sum_{\omega\in\Omega}\Pr[\omega]=1$	$\rightarrow$	Gesamtgewicht aller Objekte: 1 Kg.
Ereignis E	$\rightarrow$	Objektmenge
$\Pr[E]:=\sum_{\omega\in E}\Pr[\omega]$	$\rightarrow$	Gesamtgewicht der Objektmenge

Vereinigung, Schnitt, Komplement von Ereignissen

Wenn A und B Ereignisse sind, dann sind auch $A \cup B$, $A \cap B$, $A \setminus B$ etc. Ereignisse.

Das Ereignis $\bar{E} = \Omega \setminus E$ heißt komplementäres Ereignis zu E oder komplement von E.

Zwei Ereignisse A und B heißen disjunkt wenn $A \cap B = \emptyset$ gilt.

Einfache Eigenschaften I

Die folgenden Eigenschaften sind einfache Konsequenzen des Axioms $\sum_{\omega\in\Omega} \Pr[\omega] = 1$ und der Definition $\Pr[E] := \sum_{\omega\in E} \Pr[\omega]$.

Für Ereignisse $A, B, A_1, A_2 \ldots$ gilt:

- $\Pr[\emptyset] = 0$, $\Pr[\Omega] = 1$.
- $0 \leq \Pr[A] \leq 1$.
- $\Pr[\bar{A}] = 1 - \Pr[A]$.
- Wenn $A \subseteq B$, so folgt $\Pr[A] \leq \Pr[B]$.
- Wenn $A \cap B = \emptyset$, so folgt $\Pr[A \cup B] = \Pr[A] + \Pr[B]$.
- Allgemein: $\Pr[A \cup B] = \Pr[A] + \Pr[B] - \Pr[A \cap B]$.

  (Intuitiv: $\Pr[A] + \Pr[B]$ zählt das "Gewicht" der Objekte in $A \cap B$ zweimal. Korrektur: $\Pr[A \cap B]$ abziehen.)

Einfache Eigenschaften II

- (Additionssatz) Wenn $A_1, \ldots, A_n$ paarweise disjunkt sind (also wenn für alle Paare $i \neq j$ gilt, dass $A_i \cap A_j = \emptyset$), so folgt
$$\Pr\left[\bigcup_{i=1}^{n} A_i\right] = \sum_{i=1}^{n} \Pr[A_i].$$

Für eine unendliche Menge von paarweise disjunkten Ereignissen $A_1, A_2, \ldots$ gilt analog
$$\Pr\left[\bigcup_{i=1}^{\infty} A_i\right] = \sum_{i=1}^{\infty} \Pr[A_i]\ .$$

Einfache Eigenschaften III

- (Boole'sche Ungleichung) Für beliebige $A_1, \ldots, A_n$ gilt

$$\Pr\left[\bigcup_{i=1}^{n} A_i\right] \leq \sum_{i=1}^{n} \Pr[A_i] \; .$$

Analog gilt für eine unendliche Folge $A_1, A_2, \ldots$, dass

$$\Pr\left[\bigcup_{i=1}^{\infty} A_i\right] \leq \sum_{i=1}^{\infty} \Pr[A_i] \; .$$

Die Siebformel I

Der Additionssatz gilt nur für paarweise disjunkte Ereignisse. Für den allgemeinen Fall erhalten wir folgenden

Satz 3 (Siebformel, Prinzip der Inklusion/Exklusion)

Für Ereignisse $A_1, \ldots, A_n$ $(n \geq 2)$ *gilt:*

$$\begin{aligned}
\Pr\left[\bigcup_{i=1}^{n} A_i\right] &= \sum_{i=1}^{n} \Pr[A_i] - \sum_{1 \leq i_1 < i_2 \leq n} \Pr[A_{i_1} \cap A_{i_2}] + - \ldots \\
&+ (-1)^{l-1} \sum_{1 \leq i_1 < \ldots < i_l \leq n} \Pr[A_{i_1} \cap \ldots \cap A_{i_l}] + - \ldots \\
&+ (-1)^{n-1} \cdot \Pr[A_1 \cap \ldots \cap A_n]\,.
\end{aligned}$$

Die Siebformel II

Insbesondere gilt für zwei Ereignisse A und B

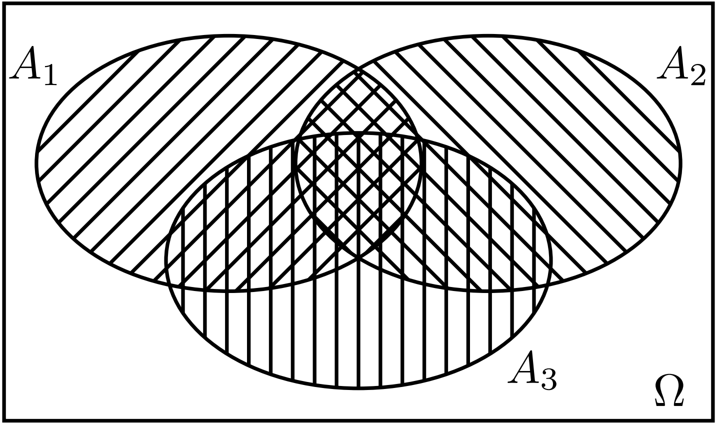
$$\Pr[A \cup B] = \Pr[A] + \Pr[B] - \Pr[A \cap B]\ .$$

Für drei Ereignisse A_1, A_2 und A_3 erhalten wir

$$\begin{aligned}
\Pr[A_1 \cup A_2 \cup A_3] \quad = \quad & \Pr[A_1] + \Pr[A_2] + \Pr[A_3] \\
& -\Pr[A_1 \cap A_2] - \Pr[A_1 \cap A_3] \\
& -\Pr[A_2 \cap A_3] \\
& +\Pr[A_1 \cap A_2 \cap A_3]\ .
\end{aligned}$$

Die Siebformel III

Beweis: Nur für $n = 3$:



Durch die im Satz angegebene Summe wird jedes Flächenstück genau einmal gezählt.

Laplacesche W'keitsräume

Definition 4

Ein diskreter W'keitsraum $(\Omega, \Pr)$ heißt Laplacescher W'keitsräum, wenn $\Pr[\omega] = \Pr[\omega']$ für alle $\omega, \omega' \in \Omega$.

Sei $(\Omega, \Pr)$ ein Laplacescher W'keitsräum. Dann:

- Ω ist endlich.
- Für jedes Ereignis E gilt $\Pr[E] = \dfrac{|E|}{|\Omega|}$

Bemerkung: Unwissen über die W'keiten bedeutet nicht, dass ein W'keitsraum ein Laplacescher W'keitsraum sein muss!

Wenn nach unserem Kenntnis der Gesetze (physikalische, psychologische, soziologische ...), die das Ergebnis des Experiments bestimmen, kein Elementarereignis "bevorzugt" wird, dann kann die Laplace-Annahme sinnvoll sein.

Experimentelle Wahl der W'keiten

Definition 5

$$\begin{aligned}\text{relative Häufigkeit von } E \quad &:= \quad \frac{\text{absolute Häufigkeit von } E}{\text{Anzahl aller Beobachtungen}} \\ &= \quad \frac{\text{Anzahl Eintreten von } E}{\text{Anzahl aller Beobachtungen}}.\end{aligned}$$

Um die W'keiten zu wählen: Wiederhole das Experiment und wähle die relativen Häufigkeiten.

Basiert auf das Prinzip („Gesetz der großen Zahlen“):

> *Wenn man das gleiche Zufallsexperiment sehr oft wiederholt, konvergieren die relativen Häufigkeiten der Ereignisse gegen deren Wahrscheinlichkeiten.*

Historische Anfänge der Wahrscheinlichkeitstheorie

Die ersten Hinweise auf mathematische Untersuchungen zu Problemen der Wahrscheinlichkeitstheorie finden sich in einem Briefwechsel zwischen den französischen Mathematikern

Pierre Fermat (1601–1665)

und

Blaise Pascal (1623–1662).

Pascal beschäftigte sich neben der Mathematik auch mit Fragestellungen aus dem Bereich der Physik und auch aus der Informatik! Sein Vater hatte als Steuerinspektor in Rouen umfangreiche Rechnungen durchzuführen und so wurde Pascal zum Bau einer mechanischen Rechenmaschine, der so genannten *Pascaline*, motiviert.

Historische Anfänge der Wahrscheinlichkeitstheorie

In dem Briefwechsel taucht bereits der Ansatz $\Pr[E] = |E|/|\Omega|$ zur Berechnung der Wahrscheinlichkeit von E auf. Auch den Begriff des Erwartungswerts kann man dort schon finden. Weder Fermat noch Pascal publizierten ihre Überlegungen zur Wahrscheinlichkeitstheorie. Der Niederländer

Christiaan Huygens (1629–1695)

entwickelte ebenfalls Methoden zum Arbeiten mit Wahrscheinlichkeiten aus. Er publizierte im Jahre 1657 auch eine kleine Arbeit mit dem Titel „De ratiociniis in ludo aleae" (Über die Gesetzmäßigkeiten beim Würfelspiel).

2. Mehrstufige Experimente

Mehrstufige Experimente I

Experimente bestehen oft aus Teilexperimenten, die der Reihe nach durchgeführt werden.

Welches Teilexperiment als nächstes durchgeführt wird, kann vom Ausgang des vorigen Teils abhängen.

Elementarereignisse $=$ mögliche Sequenzen von “Teilereignissen”.

Die W’keit eines Elementarereignisses ist das Produkt der W’keiten der Teilereignisse.

Beispiel 6

Eine faire Münze wird geworfen. Zeigt sie Kopf, werden zwei faire Würfel geworfen. Zeigt sie Zahl, wird nur ein fairer Würfel geworfen.

Frage: Was ist die W'keit, höchstens 4 Augen zu bekommen?

Modell:

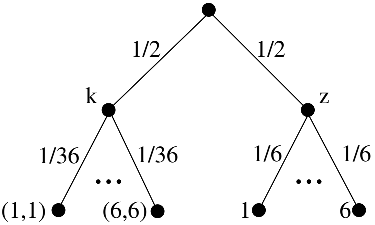
$$\Omega = \{\mathsf{k}\,(1,1), \ldots, \mathsf{k}\,(6,6), \mathsf{z}\,1, \ldots, \mathsf{z}\,6\}$$

$$\Pr[\mathsf{k}\,(i,j)] = \frac{1}{2} \cdot \frac{1}{36} \qquad \Pr[\mathsf{z}\,i] = \frac{1}{2} \cdot \frac{1}{6}$$

W'keit, höchstens 4 Augen zu bekommen: $\frac{1}{2} \cdot \frac{6}{36} + \frac{1}{2} \cdot \frac{4}{6} = \frac{5}{12}$

Mehrstufige Experimente III

Mehrstufige Experimente können als Bäume visualisiert werden.
Die Elementarereignisse sind die Pfade des Baumes.



Jeder Knoten repräsentiert ein Ereignis (die Menge der Pfade, die den Knoten besuchen).

Die W'keit dieses Ereignisses ist das Produkt der W'keiten im Pfad von der Würzel zum Knoten.

Beispiel 7

Eine faire Münze wird so lange geworfen, bis die gleiche Seite zweimal hintereinander fällt.

Frage: Mit welcher W'keit ist die Anzahl der Würfe gerade?

Modell:

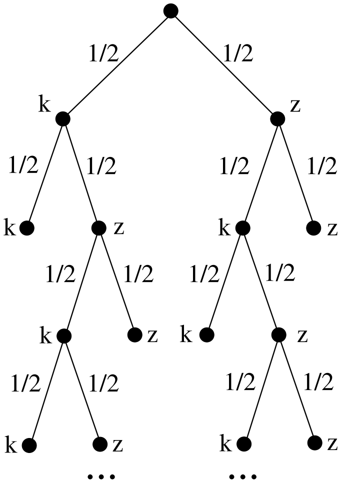
$$\Omega = \{\textsf{kk, zz, kzz, zkk, kzkk, zkzz} \ldots\} \qquad \Pr[\omega] = 2^{-|\omega|}$$

Es gilt:

$$\Pr[\textsf{gerade}] = \Pr[\textsf{kk}] + \Pr[\textsf{zz}] + \Pr[\textsf{kzkk}] + \ldots$$

Mehrstufige Experimente V

Visualisierung als Baum:



Mehrstufige Experimente VI

"Abstrakteres" Modell: die Elementarereignisse sind die Anzahlen der Würfe.

$$\Omega = \{2, 3, 4 \ldots\} .$$

W'keiten der Elementarereignisse:

$$\begin{aligned}
\Pr[2i] &= \Pr[\ (\mathsf{kz})^{(i-1)}\mathsf{kk}\ ] + \Pr[\ (\mathsf{zk})^{(i-1)}\mathsf{zz}\ ] \\
&= \left(\frac{1}{2}\right)^{2i} + \left(\frac{1}{2}\right)^{2i} = \left(\frac{1}{2}\right)^{2i-1} \\
\Pr[\mathsf{gerade}] &= \sum_{i=1}^{\infty} \Pr[2i] = \sum_{i=1}^{\infty} \left(\frac{1}{2}\right)^{2i-1} = \frac{1}{2} \sum_{i=0}^{\infty} \left(\frac{1}{4}\right)^{i} \\
&= \frac{1}{2} \cdot \frac{1}{3/4} = \frac{2}{3}
\end{aligned}$$

Mehrstufige Experimente VII

Beispiel 8

Sie nehmen an einer Spielshow im Fernsehen teil, bei der Sie eine von drei verschlossenen Türen auswählen sollen. Hinter einer Tür wartet der Preis, ein Auto, hinter den beiden anderen stehen Ziegen. Sie zeigen auf eine Tür. Sie bleibt vorerst geschlossen. Der Moderator weiss, hinter welcher Tür sich das Auto befindet; mit den Worten "Ich zeige Ihnen mal was" öffnet er eine andere Tür, und eine meckernde Ziege schaut ins Publikum. Er fragt: "Bleiben Sie bei Nummer eins, oder wählen sie Nummer zwei? "

Frage: Welche Tür hat bessere Chancen?

- Nummer eins.
- Nummer zwei.
- Beide gleich.

Mehrstufige Experimente VIII

Wir vergleichen zwei Experimente: Im ersten Experiment bleiben Sie immer bei der Tür, die Sie gewählt haben. Im zweiten Experiment wechseln Sie immer die Tür.

Annahmen (genaue Beschreibung der Experimente):

- Das Auto wird mit W'keit $1/3$ hinter Tür 1, 2, oder 3 gestellt.
- Sie wählen eine Tür mit W'keit $1/3$.
- Wenn der Moderator zwei Türen aufmachen darf, dann wählt er eine Tür mit W'keit $1/2$.

Elementarereignis: (Tür-des-Autos
Ihre-erste-Wahl
Wahl-des-Moderators
Ihre-zweite-Wahl
Ergebnis)

Mehrstufige Experimente IX

- Experiment "Ich bleibe bei meiner ersten Wahl"
  Modell:

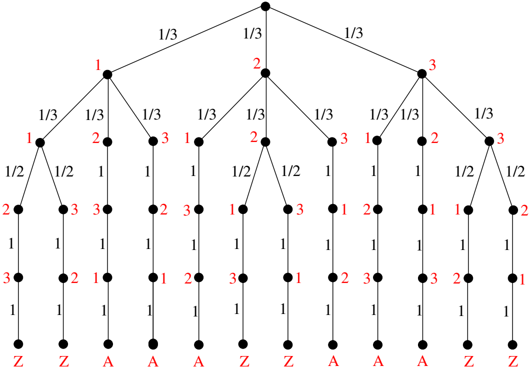
$$\Omega_b = \{\, 1\,1\,2\,1\,A\,,\ 1\,1\,3\,1\,A\,,\ 1\,2\,3\,2\,Z\,, \ldots,\ 3\,3\,2\,3\,A\,\}$$

$$\Pr[1\,1\,2\,1\,A] = \frac{1}{3}\cdot\frac{1}{3}\cdot\frac{1}{2} \qquad \Pr[1\,2\,3\,1\,A] = \frac{1}{3}\cdot\frac{1}{3}$$

$$\Pr[\text{Auto}] = \frac{1}{3}$$

- Experiment "Ich wechsle"

$$\Omega_w = \{\ 1\,1\,2\,3\,Z\ ,\ 1\,1\,3\,2\,Z\ ,\ 1\,2\,3\,1\,A\ , \ldots ,\ 3\,3\,2\,1\,Z\ \}$$



$$\Pr[\text{Auto}] = \frac{2}{3}$$

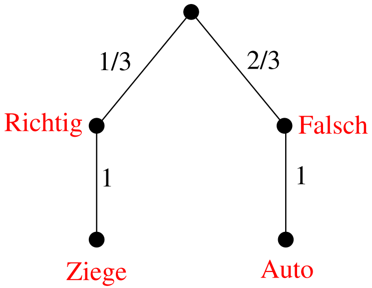
Mehrstufige Experimente XI

- Wir gewinnen ein abstrakteres Modell des Experiments “Ich wechsle” durch “Aggregation”.
- Alle Sequenzen, in denen Sie eine falsche Tür am Anfang wählen, werden aggregiert. In diesen Sequenzen gewinnen Sie immer das Auto.
- Alle Sequenzen, in denen Sie die richtige Tür am Anfang wählen, werden aggregiert. In diesen Sequenzen gewinnen Sie nie das Auto.

Mehrstufige Experimente XI

Ein viel einfacheres Modell:

$$\Omega_w = \{\text{Richtig.Ziege}, \text{Falsch.Auto}\}$$



$$\Pr[\text{Auto}] = \Pr[\text{Falsch.Auto}] = \frac{2}{3}.$$

3. Zufallsvariablen, Erwartungswert, Varianz

Zufallsvariablen I

Anstatt der Ereignisse selbst sind wir oft an „Auswirkungen“ oder „Merkmalen“ der (Elementar-)Ereignisse interessiert.

Definition 9

Sei ein Wahrscheinlichkeitsraum auf der Ergebnismenge Ω gegeben. Eine Abbildung

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

heißt (numerische) Zufallsvariable mit Wertebereich

$$W_X := X(\Omega) = \{x \in \mathbb{R} \mid \exists\, \omega \in \Omega \text{ mit } X(\omega) = x\}$$

Eine Zufallsvariable über einer endlichen oder abzählbar unendlichen Ergebnismenge Ω heißt diskret.

$\Pr[X = a]$ ist eine Abkürzung für $\Pr[\,\{\omega \in \Omega \mid X(\omega) = a\}\,]$.

Definition 10

- Die Funktion

$$\begin{aligned} f_X : \mathbb{R} &\to [0,1] \\ x &\mapsto f_X(x) := \Pr[X = x] \end{aligned}$$

nennt man (diskrete) Dichte oder Zähldichte der Zufallsvariablen X.

- Die Funktion

$$\begin{aligned} F_X : \mathbb{R} &\to [0,1] \\ x &\mapsto F_X(x) := \sum_{x' \in W_X \,:\, x' \leq x} \Pr[X = x'] \end{aligned}$$

heißt (kumulative) Verteilung der Zufallsvariablen X.

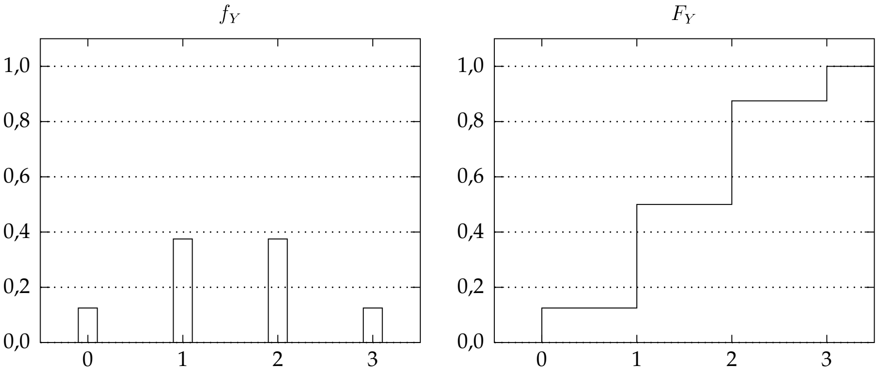
Zufallsvariablen III

Beispiel 11

Wir werfen drei faire Münzen. Sei Y die Zufallsvariable, die die Anzahl der Köpfe angibt. Für die Zufallsvariable Y haben wir

$$
\begin{aligned}
\Pr[Y = 0] &= \Pr[\,\{zzz\}\,] &&= \frac{1}{8}\\
\Pr[Y = 1] &= \Pr[\,\{kzz, zkz, zzk\}\,] &&= \frac{3}{8}\\
\Pr[Y = 2] &= \Pr[\,\{kkz, kzk, zkk\}\,] &&= \frac{3}{8}\\
\Pr[Y = 3] &= \Pr[\,\{kkk\}\,] &&= \frac{1}{8}
\end{aligned}
$$

Zufallsvariablen IV



Dichte und Verteilung von Y

Bernoulli-Verteilung I

Beispiel 12

Ein Würfel wird geworfen. Sei X die Zufallsvariable mit $W_X = \{0, 1\}$ und $X(n) = 1$ falls die Augenzahl $n = 3, 6$ ist und $X(n) = 0$ sonst. Es gilt

$$f_X(1) = 1/3 \qquad f_X(0) = 2/3$$

Definition 13

Eine Zufallssvariable X mit $W_X = \{0, 1\}$ und Dichte

$$f_X(1) = p \qquad f_X(0) = 1 - p$$

für einen Wert $0 \leq p \leq 1$ heißt Bernoulli-verteilt.
Den Parameter p nennen wir Erfolgswahrscheinlichkeit.

Bernoulli-Verteilung II

Der Name der Bernoulli-Verteilung geht zurück auf den Schweizer Mathematiker Jakob Bernoulli (1654–1705). Wie viele andere Mathematiker seiner Zeit hätte auch Bernoulli nach dem Wunsch seines Vaters ursprünglich Theologe werden sollen. Sein Werk *ars conjectandi* stellt eine der ersten Arbeiten dar, die sich mit dem Teil der Mathematik beschäftigen, den wir heute als Wahrscheinlichkeitstheorie bezeichnen.

Binomialverteilung I

Beispiel 14

Eine Münze wird 10-Mal geworfen. Die W'keit von 'Kopf' beträgt $1/3$ bei jedem Wurf.

Modell:

$$\Omega = \{k, z\}^{10} \qquad \Pr[\omega] = \left(\frac{1}{3}\right)^{\mathcal{K}(\omega)} \left(\frac{2}{3}\right)^{10-\mathcal{K}(\omega)}$$

wobei $\mathcal{K}(\omega)$ die Anzahl der Köpfe in ω darstellt.

Es gilt z.B. $\mathcal{K}(kkzzkzkzzz) = 4$ und $\mathcal{K}(z^{10}) = 0$.

Bemerkung: Alle Elementarereignisse mit derselben Anzahl von Köpfen haben die gleiche W'keit.

Binomialverteilung II

$\mathcal{K}\colon \Omega \to \mathbb{R}$ ist eine Zufallsvariable mit $W_{\mathcal{K}} = \{0, \ldots n\}$.

Es gibt $\binom{10}{x}$ Elementarereignisse in Ω mit $\mathcal{K}(\omega) = x$.

Für die Dichte von $\mathcal{K}$ gilt also

$$f_{\mathcal{K}}(x) = \binom{10}{x} \left(\frac{1}{3}\right)^{x} \left(\frac{2}{3}\right)^{10-x}$$

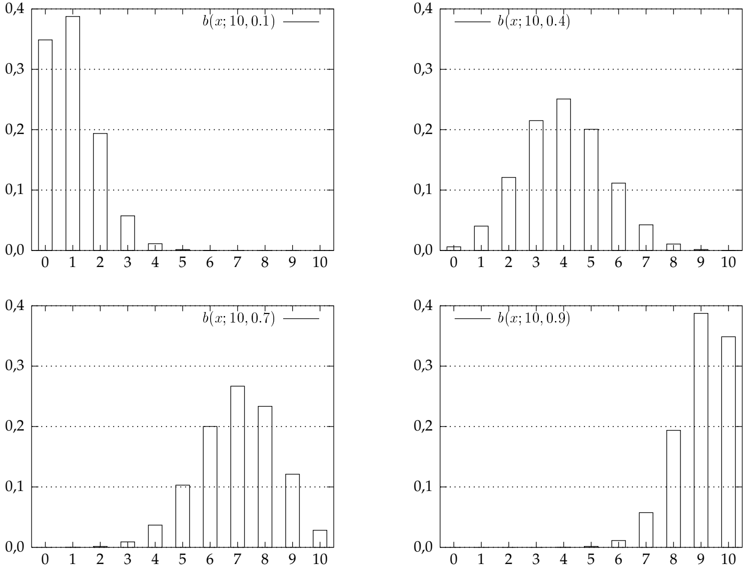
Definition 15

Eine Zufallssvariable X mit $W_X = \{0, \ldots, n\}$ und Dichte

$$f_X(x) = \binom{n}{x} p^x q^{n-x}$$

für einen Wert $0 \leq p \leq 1$ und $q := 1 - p$ heißt binomialverteilt.

Binomialverteilung III



Dichte der Binomialverteilung

Geometrische Verteilung I

Beispiel 16

Eine Münze wird geworfen bis 'Kopf' fällt. Die W'keit von Kopf beträgt $1/3$.

Modell:

$$\Omega = Z^* K \qquad \Pr[Z^n K] = \left(\frac{2}{3}\right)^n \cdot \frac{1}{3}$$

Sei $X : \Omega \to \mathbb{N}$ die Zufallsvariable, die die Anzahl der Würfe angibt, d.h. $X(Z^n K) = n + 1$.

Es gilt:

$$f_X(x) = \left(\frac{1}{3}\right)^{x-1} \cdot \frac{2}{3}$$

Geometrische Verteilung II

Definition 17

Eine Zufallssvariable X mit $W_X = \mathbb{N}$ und Dichte

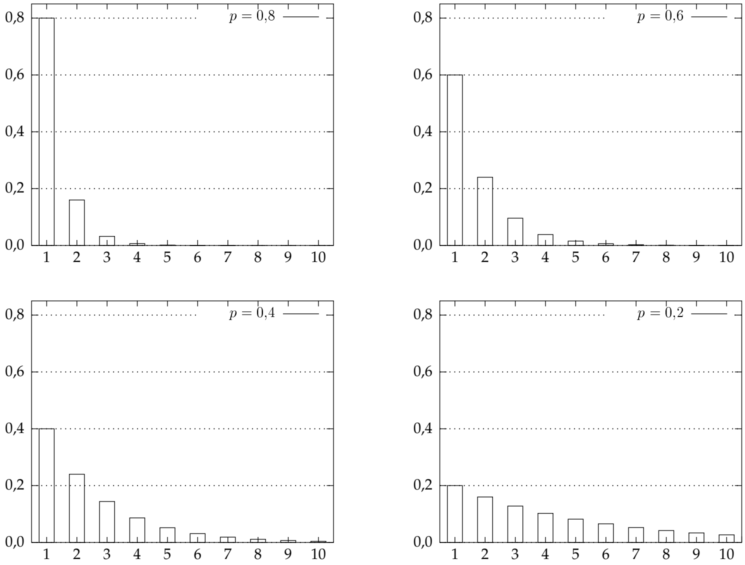
$$f_X(i) = pq^{i-1}$$

für ein $p \in [0,1]$ und $q := 1 - p$ ist geometrisch Verteilt.

Die kumulative Verteilung einer geometrisch verteilten Zufallsvariablen X mit Erfolgswahrscheinlichkeit p ist die Funktion

$$F_X(x) = \Pr[X \leq x] = \sum_{i=1}^{x} pq^{i-1} = p \cdot \sum_{i=0}^{x-1} q^i = p \cdot \left(\frac{1-q^x}{1-q} \right) = 1 - q^x$$

Geometrische Verteilung III



Dichte der geometrischen Verteilung

Mehrere Zufallsvariablen I: Gemeinsame Dichte

Seien $X, Y : \Omega \to \mathbb{R}$ zwei Zufallsvariablen über demselben W'keitsraum.

Wir definieren

$$\Pr[X = x, Y = y] := \Pr[\ \{\omega \mid X(\omega) = x \wedge Y(\omega) = y\}\ ]\,.$$

Die Funktion

$$f_{X,Y}(x, y) := \Pr[X = x, Y = y]$$

heißt gemeinsame (diskrete) Dichte der Zufallsvariablen X und Y.

Mehrere Zufallsvariablen II: Gemeinsame Dichte

Beispiel 18

Die Studierenden von Prof. Evilsparza haben zwei Klausuren geschrieben. Die möglichen Noten sind $1, 2, \ldots, 5$. Ein Studierender besteht, wenn seine Bestnote eine 1 ist oder seine Schnittnote mindestens 3 beträgt. Prof. Evilsparza hat keine Lust zu korrigieren, und verteilt die Noten zufällig.

Modell:

$$\Omega = \{(1,1), \ldots, (5,5)\} \qquad \Pr[(i,j)] = \frac{1}{25}$$

Seien B, S zwei Zufallsvariablen, die die Bestnote und die Schnittnote angeben.

Frage: Welche gemeinsame Dichte haben B und S?

Mehrere Zufallsvariablen III: Gemeinsame Dichte

Die gemeinsame Dichte $f_{B,S}$ von B und S ist:

$B \setminus S$	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
1	$\frac{1}{25}$	$\frac{2}{25}$	$\frac{2}{25}$	$\frac{2}{25}$	$\frac{2}{25}$	0	0	0	0
2	0	0	$\frac{1}{25}$	$\frac{2}{25}$	$\frac{2}{25}$	$\frac{2}{25}$	0	0	0
3	0	0	0	0	$\frac{1}{25}$	$\frac{2}{25}$	$\frac{2}{25}$	0	0
4	0	0	0	0	0	0	$\frac{1}{25}$	$\frac{2}{25}$	0
5	0	0	0	0	0	0	0	0	$\frac{1}{25}$

Mehrere Zufallsvariablen IV: Randdichte

Die Funktionen f_X und f_Y, definiert durch

$$f_X(x) = \sum_{y \in W_Y} f_{X,Y}(x,y) \quad \text{bzw.} \quad f_Y(y) = \sum_{x \in W_X} f_{X,Y}(x,y)$$

nennen wir Randdichten.

Die Ereignisse „$Y = y$“ bilden eine Partitionierung des Wahrscheinlichkeitsraumes, und es gilt daher

$$\Pr[X = x] = \sum_{y \in W_Y} \Pr[X = x, Y = y] = f_X(x)\,,$$

Die Dichten der einzelnen Zufallsvariablen entsprechen also genau den Randdichten.

Mehrere Zufallsvariablen V: Randdichte

Die Randdichten f_B, f_S des Beispiels sind:

$B \setminus S$	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	f_B
1	$\frac{1}{25}$	$\frac{2}{25}$	$\frac{2}{25}$	$\frac{2}{25}$	$\frac{2}{25}$	0	0	0	0	$\frac{9}{25}$
2	0	0	$\frac{1}{25}$	$\frac{2}{25}$	$\frac{2}{25}$	$\frac{2}{25}$	0	0	0	$\frac{7}{25}$
3	0	0	0	0	$\frac{1}{25}$	$\frac{2}{25}$	$\frac{2}{25}$	0	0	$\frac{5}{25}$
4	0	0	0	0	0	0	$\frac{1}{25}$	$\frac{2}{25}$	0	$\frac{3}{25}$
5	0	0	0	0	0	0	0	0	$\frac{1}{25}$	$\frac{1}{25}$
f_S	$\frac{1}{25}$	$\frac{2}{25}$	$\frac{3}{25}$	$\frac{4}{25}$	$\frac{5}{25}$	$\frac{4}{25}$	$\frac{3}{25}$	$\frac{2}{25}$	$\frac{1}{25}$	1

Mehrere Zufallsvariablen VI: Randdichte

Die W'keit zu Bestehen beträgt

$$\begin{aligned}
\Pr[B = 1 \cup S \leq 3] &= \Pr[B = 1] + \Pr[S \leq 3] - \Pr[B = 1, S \leq 3] \\
&= f_B(1) + \sum_{i=1}^{3} f_S(i) - \sum_{i=1}^{3} f_{B,S}(1, i) \\
&= \frac{9}{25} + \left(\frac{1}{25} + \frac{2}{25} + \frac{3}{25} + \frac{4}{25} + \frac{5}{25}\right) \\
&\quad - \left(\frac{1}{25} + \frac{2}{25} + \frac{2}{25} + \frac{2}{25} + \frac{2}{25}\right) \\
&= \frac{15}{25} = \frac{3}{5}
\end{aligned}$$

Bemerkung: Die gemeinsame Dichte ist nicht notwendig durch die beiden Randdichten eindeutig bestimmt!

Mehrere Zufallsvariablen VII: Gemeinsame Verteilung

Für zwei Zufallsvariablen X, Y definiert man die gemeinsame (kumulative) Verteilung

$$\begin{aligned} F_{X,Y}(x,y) &= \Pr[X \leq x, Y \leq y] = \Pr[\ \{\omega \mid X(\omega) \leq x, Y(\omega) \leq y\}\ ] \\ &= \sum_{x' \leq x} \sum_{y' \leq y} f_{X,Y}(x', y')\,. \end{aligned}$$

Die (kumulative) Randverteilung oder Marginalverteilung ergibt sich gemäß

$$F_X(x) = \sum_{x' \leq x} f_X(x') = \sum_{x' \leq x} \sum_{y \in W_Y} f_{X,Y}(x', y)$$

sowie

$$F_Y(y) = \sum_{y' \leq y} f_Y(y') = \sum_{y' \leq y} \sum_{x \in W_X} f_{X,Y}(x, y')\,.$$

Mehrere Zufallsvariablen VIII: Gemeinsame Verteilung

Die gemeinsame Verteilung $F_{B,S}$ und die Randdichten F_B, F_S von B und S sind:

$F_{B,S}$	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	F_B
1	$\frac{1}{25}$	$\frac{3}{25}$	$\frac{5}{25}$	$\frac{7}{25}$	$\frac{9}{25}$	$\frac{9}{25}$	$\frac{9}{25}$	$\frac{9}{25}$	$\frac{9}{25}$	$\frac{9}{25}$
2	$\frac{1}{25}$	$\frac{3}{25}$	$\frac{6}{25}$	$\frac{10}{25}$	$\frac{14}{25}$	$\frac{16}{25}$	$\frac{16}{25}$	$\frac{16}{25}$	$\frac{16}{25}$	$\frac{16}{25}$
3	$\frac{1}{25}$	$\frac{3}{25}$	$\frac{6}{25}$	$\frac{10}{25}$	$\frac{15}{25}$	$\frac{19}{25}$	$\frac{21}{25}$	$\frac{21}{25}$	$\frac{21}{25}$	$\frac{21}{25}$
4	$\frac{1}{25}$	$\frac{3}{25}$	$\frac{6}{25}$	$\frac{10}{25}$	$\frac{15}{25}$	$\frac{19}{25}$	$\frac{22}{25}$	$\frac{24}{25}$	$\frac{24}{25}$	$\frac{24}{25}$
5	$\frac{1}{25}$	$\frac{3}{25}$	$\frac{6}{25}$	$\frac{10}{25}$	$\frac{15}{25}$	$\frac{19}{25}$	$\frac{22}{25}$	$\frac{24}{25}$	$\frac{25}{25}$	$\frac{25}{25}$
F_S	$\frac{1}{25}$	$\frac{3}{25}$	$\frac{6}{25}$	$\frac{10}{25}$	$\frac{15}{25}$	$\frac{19}{25}$	$\frac{22}{25}$	$\frac{24}{25}$	$\frac{25}{25}$	1

Mehrere Zufallsvariablen IX: Gemeinsame Verteilung

Beispiel 19

Aus einem Skatblatt mit 32 Karten ziehen wir zufällig eine Hand von zehn Karten sowie einen Skat von zwei Karten. Unter den Karten gibt es vier Buben. Die Zufallsvariable X zählt die Anzahl der Buben in der Hand, während Y die Anzahl der Buben im Skat angibt. Wir berechnen die Dichten von X und Y.

Es gilt für $0 \leq x \leq 4$ und $0 \leq y \leq 2$

$$\Pr[X = x] = \frac{\binom{4}{x}\binom{28}{10-x}}{\binom{32}{10}} \qquad \Pr[Y = y] = \frac{\binom{4}{y}\binom{28}{2-y}}{\binom{32}{2}}$$

Mehrere Zufallsvariablen X: Gemeinsame Verteilung

Für die gemeinsame Dichte $f_X, Y \;\; X$ gilt z.B.

$$f_{X,Y}(4,1) = \Pr[X = 4, Y = 1] = 0,$$

und allgemein

$$f_{X,Y}(x,y) = \frac{\binom{4}{x}\binom{28}{10-x}\binom{4-x}{y}\binom{28-(10-x)}{2-y}}{\binom{32}{10}\binom{22}{2}}.$$

Zusammengesetzte Zufallsvariablen I

Definition 20

Seien $X, Y : \Omega \to \mathbb{R}$ Zufallsvariablen. Die Zufallsvariable $X + Y$ ist die Abbildung

$$\begin{array}{rcl} X + Y : \Omega & \to & \mathbb{R} \\ \omega & \mapsto & X(\omega) + Y(\omega) \end{array}$$

Weitere Zufallsvariablen wie z. B.

$$3X - 2Y + 2 \qquad X \cdot Y \qquad \min\{X, Y\} \qquad X^Y$$

usw. werden analog definiert.

Allgemein: für Zufallsvariablen $X_1, \ldots, X_n$ und $f : \mathbb{R}^n \to \mathbb{R}$ bezeichnet $f(X_1, \ldots, X_n)$ die Zufallsvariable mit

$$f(X_1, \ldots, X_n)(\omega) := f(X_1(\omega), \ldots, X_n(\omega))$$

Zusammengesetzte Zufallsvariablen II

Beispiel 21 (Klausuren nochmal)

Die Studierenden von Prof. Evilsparza haben zwei Klausuren geschrieben. Ein Studierender besteht, wenn seine Bestnote eine 1 ist oder seine Schnittnote mindestens 3 beträgt. Prof. Evilsparza hat keine Lust zu korrigieren und verteilt die Noten zufällig.

Seien X_1, X_2 die Zufallsvariablen, die die Noten der ersten und der zweiten Klausur angeben, z.B. $X_1(1,3) = 1$ und $X_2(1,3) = 3$.

Seien B, S die Zufallsvariablen, die die beste und die Schnittnote angeben.

Es gilt

$$B = \max\{X_1, X_2\} \qquad S = \frac{X_1 + X_2}{2}$$

Erwartungswert I

Drei faire Münzen werden geworfen.
Frage: Wieviele Köpfe fallen im Schnitt ? (Vive la France!)

Eine Münze wird so lange geworfen, bis sie zum ersten Mal „Kopf“ zeigt.
Frage: Wie oft wirft man im Schnitt?

Beim Roulettespiel (Zahlen 0 bis 36) wetten wir immer wieder 1 Euro auf Nummer 5.

Frage: Wieviel gewinnen wir im Schnitt pro Spiel?

Erwartungswert II

Definition 22

Zu einer Zufallsvariablen X definieren wir den Erwartungswert $\mathbb{E}[X]$ durch

$$\mathbb{E}[X] := \sum_{x \in W_X} x \cdot \Pr[X = x] = \sum_{x \in W_X} x \cdot f_X(x) \,,$$

sofern $\sum_{x \in W_X} |x| \cdot \Pr[X = x]$ konvergiert.

Erwartungswert II

Beispiel 23

Drei faire Münzen werden geworfen. Wieviele Köpfe fallen in Schnitt?

Modell:

$$\Omega = \{K, Z\}^3 \qquad \Pr[\omega] = \frac{1}{8}$$

Sei X die Zufallsvariable, die die Anzahl der Köpfe angibt.

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{i=0}^{3} i \cdot \Pr[X = i] \\
&= 1 \cdot \Pr[X = 1] \; + \; 2 \cdot \Pr[X = 2] \; + \; 3 \cdot \Pr[X = 3] \\
&= 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} \; = \; \frac{3}{2}\,.
\end{aligned}$$

Erwartungswert III

Beispiel 24

Eine Münze wird so lange geworfen, bis sie zum ersten Mal „Head“ zeigt. Sei k die Anzahl der durchgeführten Würfe. Wenn k ungerade ist, zahlt der Spieler an die Bank k Euro. Andernfalls (k gerade) zahlt die Bank k Euro an den Spieler. Sei G die Zufallsvariable mit

$$G(\omega) := \begin{cases} |\omega| & \text{falls } |\omega| \text{ ungerade,} \\ -|\omega| & \text{falls } \omega \text{ gerade.} \end{cases}$$

Es gilt

$$\Pr[\text{„Anzahl Würfe} = k\text{“}] = (1/2)^k \,.$$

Damit erhalten wir

$$\mathbb{E}[G] = \sum_{k=1}^{\infty} (-1)^{k-1} \cdot k \cdot \left(\frac{1}{2}\right)^k \,.$$

Erwartungswert IV

Da

$$\sum_{k=1}^{\infty} |(-1)^{k-1} \cdot k| \cdot \left(\frac{1}{2}\right)^k \leq \sum_{k=1}^{\infty} k \cdot \left(\frac{1}{2}\right)^k ,$$

existiert der Erwartungswert $\mathbb{E}[G]$.

Es gilt

$$\begin{aligned}
\mathbb{E}[G] &= \sum_{j=1}^{\infty}(2j-1) \cdot \left(\frac{1}{2}\right)^{2j-1} - \sum_{j=1}^{\infty} 2j \cdot \left(\frac{1}{2}\right)^{2j} \\
&= \sum_{j=1}^{\infty} \left(\frac{1}{2}\right)^{2j-1} \cdot [(2j-1) - j] \\
&= \frac{1}{2} \cdot \sum_{j=1}^{\infty}(j-1) \cdot \left(\frac{1}{4}\right)^{j-1} = \ldots = \frac{2}{9} .
\end{aligned}$$

Erwartungswert V

Wird jedoch, um das Risiko zu steigern, der zu zahlende Betrag von k Euro jeweils auf 2^k Euro erhöht, also

$$G'(\omega) := \begin{cases} 2^{|\omega|} & \text{falls } |\omega| \text{ ungerade,} \\ 2^{-|\omega|} & \text{falls } \omega \text{ gerade.} \end{cases}$$

dann existiert $\mathbb{E}[G']$ nicht, da

$$\begin{aligned} \mathbb{E}[G'] &= \sum_{i=1}^{\infty} (-1)^{k-1} \cdot 2^k \cdot \left(\frac{1}{2}\right)^k \\ &= \sum_{i=1}^{\infty} (-1)^{k-1} = +1 - 1 + 1 - 1 + - \ldots . \end{aligned}$$

Erwartungswert VI: Verteilungen

- $X \sim \text{Ber}(p)$ bezeichnet, dass X Bernoulli-verteilt ist mit Erfolgsw'keit p.
- $X \sim \text{Bin}(n, p)$ bezeichnet, dass X binomialverteilt ist mit n Versuchen und Erfolgsw'keit p.
- $X \sim \text{Geo}(p)$ bezeichnet, dass X geometrisch verteilt ist mit Erfolgsw'keit p.

Satz 25

- *Wenn $X \sim \text{Ber}(p)$, dann $\mathbb{E}[X] = p$.*
- *Wenn $X \sim \text{Bin}(n, p)$, dann $\mathbb{E}[X] = np$.*
- *Wenn $X \sim \text{Geo}(p)$, dann $\mathbb{E}[X] = \frac{1}{p}$.*

Erwartungswert VII: Bernoulli- und geometrische Verteilung

Bernoulli-Verteilung:

$$\mathbb{E}[X] = 1 \cdot p + 0 \cdot (1-p) = p$$

Geometrische Verteilung:

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{i=1}^{\infty} i \cdot pq^{i-1} = p \cdot \sum_{i=1}^{\infty} i \cdot q^{i-1} \\
&= p \cdot \frac{\mathsf{d}}{\mathsf{d}\, q} \left(\sum_{i=0}^{\infty} q^i \right) \\
&= p \cdot \frac{\mathsf{d}}{\mathsf{d}\, q} \left(\frac{1}{1-q} \right) \\
&= p \cdot \frac{1}{(1-q)^2} = \frac{1}{p}.
\end{aligned}$$

Erwartungswert VIII: Binomialverteilung

Binomialverteilung:

$$
\begin{aligned}
\mathbb{E}[X] &= \sum_{i=0}^{n} i \cdot \binom{n}{i} p^i q^{n-i} \\
&= pq^{n-1} \sum_{i=1}^{n} \binom{n}{i} \cdot i \cdot \left(\frac{p}{q}\right)^{i-1} \\
&= pq^{n-1} \cdot S(p/q) \qquad \text{mit} \qquad S(z) := \sum_{i=1}^{n} \binom{n}{i} \cdot i \cdot z^{i-1}
\end{aligned}
$$

Wir betrachten die Funktion $S(z)$.

Erwartungswert IX: Binomialverteilung

$$
\begin{aligned}
S(z) &= \sum_{i=1}^{n} \binom{n}{i} \cdot i \cdot z^{i-1} &&= \sum_{i=0}^{n} \binom{n}{i} \cdot \frac{\mathsf{d}}{\mathsf{d}\, z} z^{i} \\
&= \frac{\mathsf{d}}{\mathsf{d}\, z} \left(\sum_{i=0}^{n} \binom{n}{i} z^{i} \right) &&= \frac{\mathsf{d}}{\mathsf{d}\, z} (1+z)^{n} = n\,(1+z)^{n-1}
\end{aligned}
$$

Und damit

$$
\begin{aligned}
\mathbb{E}[X] &= pq^{n-1} \cdot S(p/q) \\
&= pq^{n-1} \cdot n \left(1 + \frac{p}{q} \right)^{n-1} \\
&= pq^{n-1} n \left(\frac{1}{q} \right)^{n-1} = np
\end{aligned}
$$

Varianz I

Wir betrachten die beiden folgenden Zufallsexperimente:

1. Wir würfeln (mit einem fairen Würfel): bei gerader Augenzahl erhalten wir 1 Euro, bei ungerader Augenzahl müssen wir 1 Euro bezahlen.
2. Wir würfeln (mit einem fairen Würfel): bei 6 Augen erhalten wir 5 Euro, ansonsten müssen wir 1 Euro bezahlen.

In beiden Fällen ist der erwartete Gewinn $= 0$.

Dennoch sind die „Schwankungen“ im ersten Fall geringer als im zweiten.

Wie können die „Schwankungen“ quantifiziert werden?

Varianz II

Erste Idee: „Schwankungsgrad“ durch $\mathbb{E}[|X-\mu|]$ berechnen, wobei $\mu = \mathbb{E}[X]$ sei. Die Betragsfunktion ist jedoch mathematisch „unhandlich“.

Zweite Idee: „Schwankungsgrad“ berechnen durch

$$\mathbb{E}[(X-\mu)^2] \quad \text{oder} \quad \sqrt{\mathbb{E}[(X-\mu)^2]}\,.$$

Definition 26

Für eine Zufallsvariable X mit $\mu = \mathbb{E}[X]$ definieren wir die Varianz $\mathrm{Var}[X]$ durch

$$\mathrm{Var}[X] := \mathbb{E}[(X-\mu)^2] = \sum_{x \in W_X} (x-\mu)^2 \cdot \Pr[X=x]\,.$$

Die Größe $\sigma := \sqrt{\mathrm{Var}[X]}$ heißt Standardabweichung von X.

Varianz III

Satz 27

Für eine beliebige Zufallsvariable X gilt

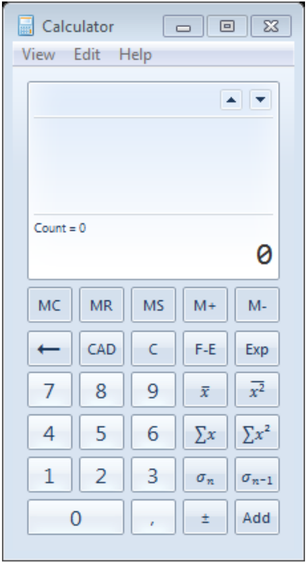
$$\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \,.$$

Beweis:

Sei $\mu := \mathbb{E}[X]$. Nach Definition gilt

$$\begin{aligned}
\mathrm{Var}[X] &= \mathbb{E}[(X-\mu)^2] = \mathbb{E}[X^2 - 2\mu \cdot X + \mu^2] \\
&= \sum_{\omega\in\Omega} \left(X(\omega)^2 - 2\mu X(\omega) + \mu^2\right) \Pr[\omega] \\
&= \sum_{\omega\in\Omega} X(\omega)^2 \Pr[\omega] - 2\mu \sum_{\omega\in\Omega} X(\omega) \Pr[\omega] + \sum_{\omega\in\Omega} \mu^2 \Pr[\omega] \\
&= \mathbb{E}[X^2] - 2\mu \cdot \mathbb{E}[X] + \mu^2 \\
&= \mathbb{E}[X^2] - 2\mu \cdot \mu + \mu^2 \\
&= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \,.
\end{aligned}$$

Varianz IV



Calculator
View Edit Help
Count = 0
0
MC MR MS M+ M-
CAD C F-E Exp
7 8 9
4 5 6
1 2 3
0 , ± Add

Varianz V

Beispiel 28

1. Wir würfeln (mit einem fairen Würfel): bei gerader Augenzahl erhalten wir 1 Euro, bei ungerader Augenzahl müssen wir 1 Euro bezahlen. Es ist

$$\mu = 0 \quad \text{und} \quad \mathrm{Var}[X] = \frac{1}{2} \cdot 1^2 + \frac{1}{2} \cdot (-1)^2 = 1\,.$$

2. Wir würfeln (mit einem fairen Würfel): bei 6 Augen erhalten wir 5 Euro, ansonsten müssen wir 1 Euro bezahlen.
   Es ist

$$\mu = 0 \quad \text{und} \quad \mathrm{Var}[X] = \frac{1}{6} \cdot 5^2 + \frac{5}{6} \cdot (-1)^2 = 5\,.$$

Varianz VI: Bernouilli-Verteilung

Sei X eine Bernoulli-verteilte Zufallsvariable mit Erfolgswahrscheinlichkeit p.

$$
\begin{aligned}
\mathrm{Var}[X] &= \sum_{x \in W_X} (x - \mathbb{E}[X])^2 \cdot \Pr[X = x] \\
&= (0 - p)^2 (1 - p) + (1 - p)^2 p \\
&= p - p^2
\end{aligned}
$$

Varianz VII: Geometrische Verteilung

Sei X eine geometrisch verteilte Zufallsvariable mit Erfolgswahrscheinlichkeit p.

$$\begin{aligned}
\mathbb{E}[X^2] &= \sum_{i=1}^{\infty} i^2 \cdot pq^{i-1} = \sum_{i=0}^{\infty} (i+1)^2 \cdot pq^i \\
&= p \cdot \left(\sum_{i=0}^{\infty} (i+2)(i+1) \cdot q^i - \sum_{i=0}^{\infty} (i+1) \cdot q^i \right) \\
&= p \cdot \left(\frac{\mathsf{d}^2}{\mathsf{d}\, q^2} \left(\sum_{i=0}^{\infty} q^{i+2} \right) - \frac{\mathsf{d}}{\mathsf{d}\, q} \left(\sum_{i=0}^{\infty} q^{i+1} \right) \right) \\
&= p \cdot \left(\frac{\mathsf{d}^2}{\mathsf{d}\, q^2} \left(\frac{q^2}{1-q} \right) - \frac{\mathsf{d}}{\mathsf{d}\, q} \left(\frac{q}{1-q} \right) \right) \\
&= p \cdot \left(\frac{2}{(1-q)^3} - \frac{1}{(1-q)^2} \right) \\
&= \frac{2-p}{p^2}
\end{aligned}$$

Momente einer Zufallsvariable

Der Erwartungswert und die Varianz gehören zu den so genannten Momenten einer Zufallsvariablen:

Definition 29

Für eine Zufallsvariable X nennen wir $\mathbb{E}[X^k]$ das k-te Moment und $\mathbb{E}[(X - \mathbb{E}[X])^k]$ das k-te zentrale Moment.

Der Erwartungswert ist also identisch zum ersten Moment, während die Varianz dem zweiten zentralen Moment entspricht.

Teil II

Weitere Grundbegriffe

4. Markov-Diagramme

Markov-Diagramme I

Viele mehrstufige Zufallsexperimente können graphisch mit Markov-Diagrammen dargestellt werden.

Definition 30

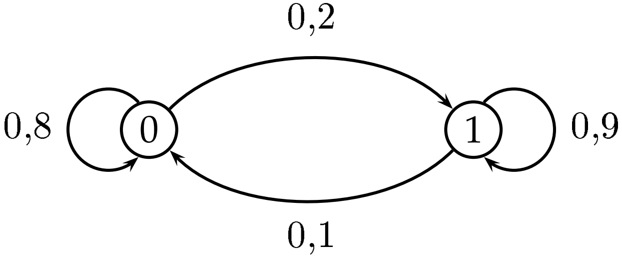
Ein Markov-Diagramm $D = (Q, T, \delta)$ besteht aus

- einer endlichen Menge Q von Zuständen,
- einer Menge $T \subseteq Q \times Q$ von Transitionen, und
- einer W'keitsfunktion $\delta \colon T \to (0, 1]$, die Folgendes erfüllt für jeden Zustand q:

$$\sum_{(q,q') \in T} \delta(q, q') = 1 \; .$$

(d.h., die W'keiten der aus q ausgehenden Transitionen addieren sich zu 1, wie es sich gehört!)

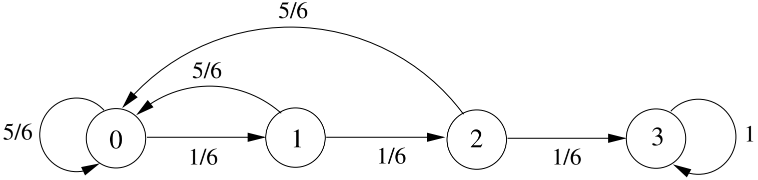
Markov-Diagramme II: Beispiel



Markov-Diagramme III: Noch ein Beispiel

Beispiel 31

Ein Würfel wird geworfen, bis dreimal hintereinander eine Sechs geworfen wird. Die Sequenzen von Würfen, die mit drei Sechsen Enden, können durch ein Markov-Diagramm dargestellt werden.



Markov-Diagramme IV

Ein (endlicher) Pfad eines Markov-Diagramms ist eine endliche Sequenz $q_0 q_1 \ldots q_k$ von Zuständen mit $k \geq 0$ und $(q_i, q_{i+1}) \in T$ für alle $0 \leq i \leq k-1$. (Ein Zustand darf mehrmals in einem Pfad vorkommen.)

Die Länge von $q_0 q_1 \ldots q_k$ ist k.

Einen Pfad $\pi = q_0 q_1 \ldots q_k$ von D ordnen wir eine W'keit zu:

$$\Pr[\pi] = \begin{cases} 1 & \text{falls } k = 0 \\ \displaystyle\prod_{i=0}^{k-1} \delta(q_i, q_{i+1}) & \text{falls } k > 0 \end{cases}$$

Für eine präfix-freie Menge Π von Pfaden definieren wir

$$\Pr[\Pi] = \sum_{\pi \in \Pi} \Pr[\pi] \ .$$

Markov-Diagramme V

Satz 32

Seien q_1, q_2 *verschiedene Zustände eines Markov-Diagramms* D *mit der Eigenschaft: jeder Zustand von* D *liegt auf mindestens einem Pfad von* q_1 *nach* q_2*.*

Sei $[q_1 \rightsquigarrow q_2]^{q_2=1}$ *die Menge aller Pfade des Diagramms von* q_1 *nach* q_2*, die* q_2 *genau einmal besuchen. Es gilt*

$$\Pr[\ [q_1 \rightsquigarrow q_2]^{q_2=1}\ ] = 1\ .$$

Markov-Diagramme VI

Beweisskizze:

Wir benötigen einige Definitionen. Für jeden Zustand q und jedes $k \geq 1$, sei

- $[q \rightsquigarrow]_k$ die Menge der Pfade aus q der Länge k.
- $[q \rightsquigarrow]_k^{q' \geq 1}$ die Menge der Pfade aus q der Länge k, die q' mindestens einmal besuchen.
- $[q \rightsquigarrow]_k^{q'=0}$ die Menge der Pfade aus q der Länge k, die q' nicht besuchen.
- $[q \rightsquigarrow q']_{\leq k}^{q'=1}$ die Menge der Pfaden von q nach q' der Länge höchstens k

usw.

Markov-Diagramme VII

Für jedes q, q' und jedes $k \geq 1$:

(1) $\Pr[\ [q \rightsquigarrow]_k\ ] = 1.$ („Massenerhaltung“)
Beweis: Einfache Induktion über k.

(2) Sei n die Anzahl der Knoten von D.
Es gibt ein $\epsilon > 0$, so dass $\Pr[\ [q \rightarrow]_n^{q_2 \geq 1}\ ] \geq \epsilon$.
Beweis: weil $[q \rightarrow]_n$ immer mindestens einen Pfad enthält, der q_2 besucht (denn q_2 ist aus jedem Zustand erreichbar).

(3) $\Pr[\ [q \rightsquigarrow]_n^{q_2 = 0}\ ] \leq 1 - \epsilon$ und $\Pr[\ [q \rightsquigarrow]_{2n}^{q_2 = 0}\ ] \leq (1 - \epsilon)^2$ und
…und $\Pr[\ [q \rightsquigarrow]_{kn}^{q_2 = 0}\ ] \leq (1 - \epsilon)^k$.
Beweis: aus (1) und (2) durch Induktion über $k \geq 0$.

(4) $\Pr[\ [q \rightsquigarrow q']_{\leq k}^{q' = 1}\ ] = \Pr[\ [q \rightsquigarrow]_k^{q' \geq 1}\ ]$.

Beweis: Folgt aus (1) und der Tatsache, dass wenn $\pi \in [q \rightsquigarrow q']_{\leq k}^{q' = 1}$ und $\pi\,\pi'$ Länge k hat, dann $\pi\,\pi' \in [q \rightsquigarrow]_k^{q' \geq 1}$.

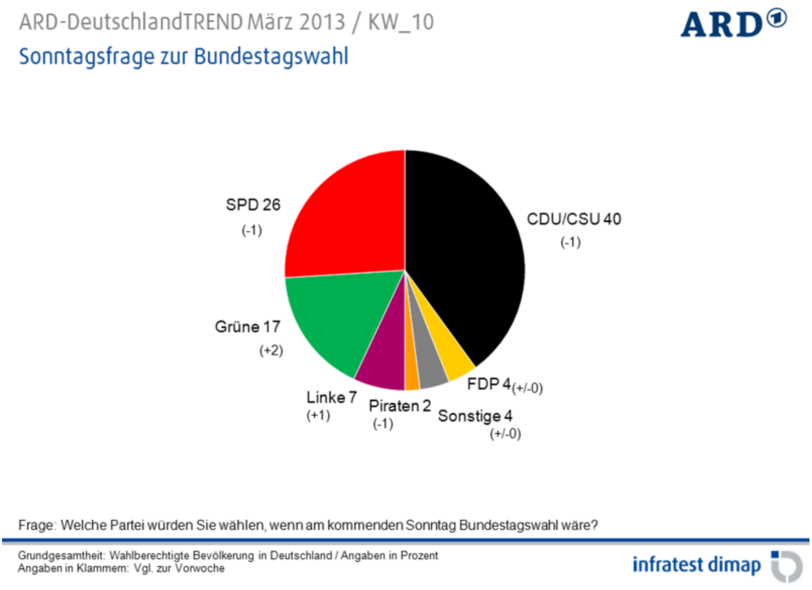
Markov-Diagramme VIII

Aus (1)-(4) folgt:

$$
\begin{array}{lll}
 & \Pr[\ [q_1 \rightsquigarrow q_2]^{q_2=1}\ ] & \\
= & \lim\limits_{k\to\infty} \Pr[\ [q_1 \rightsquigarrow q_2]^{q_2=1}_{\leq k}\ ] & \text{Def. von } [q_1 \rightsquigarrow q_2]^{q_2=1} \\
= & \lim\limits_{k\to\infty} \Pr[\ [q_1 \rightsquigarrow]^{q_2\geq 1}_{k}\ ] & (4) \\
= & \lim\limits_{k\to\infty} (1 - \Pr[\ [q_1 \rightsquigarrow]^{q_2=0}_{k}\ ]) & \Pr[\ [q_1 \rightsquigarrow]_k\ ] = 1,\ (1) \\
= & \lim\limits_{k\to\infty} (1-(1-\epsilon)^k) & \Pr[\ [q_1 \rightsquigarrow]^{q_2=0}_{kn}\ ] \leq (1-\epsilon)^k,\ (3) \\
= & 1 &
\end{array}
$$

5. Bedingte Wahrscheinlichkeiten

Bedingte Wahrscheinlichkeiten I



Bedingte Wahrscheinlichkeiten II

Beispiel 33 (ARD-Deutschlandtrend, März 2013)

Die Zustimmung für die CDU fiel um einen Punkt auf 40%.

66% der Wähler sprechen sich für die Homo-Ehe (sorry, für die rechtliche Gleichstellung gleichgeschlechtlicher Lebenspartnerschaften), bei den CDU-Anhänger jedoch nur 55%.

Eine zufällig gewählte Person wird gefragt, welche Partei sie wählen und ob sie die Gleischstellung befürworten würde.

Modell:

$$\Omega = \{(\text{CDU, Ja}), (\text{CDU, Nein}), (\text{SPD, Ja}), \ldots, (\text{Andere, Nein})\}$$

$$\begin{aligned}
\Pr[\text{CDU}] &= \Pr[(\text{CDU, Ja})] + \Pr[(\text{CDU, Nein})] && = 0.40 \\
\Pr[\text{Ja}] &= \Pr[(\text{CDU, Ja})] + \ldots + \Pr[(\text{Andere, Ja})] && = 0.66
\end{aligned}$$

Aber was ist die Zahl 0.55?

Bedingte Wahrscheinlichkeiten III

55% der CDU-Anhänger sprechen sich für die Homo-Ehe aus (antworten Ja auf die zweite Frage).

0.55 ist die „W'keit des Ereignisses Ja unter der Bedingung, dass CDU eingetreten ist“. Schreibweise:

$$\Pr[\text{Ja} \mid \text{CDU}] \; .$$

Welche Beziehung gibt es zwischen bedingten und normalen W'keiten?

- Was ist die W'keit von $(\text{CDU, Ja}) = \text{CDU} \cap \text{Ja}$?

$$\left.\begin{array}{l}\text{40\% CDU-Anhänger} \\ \text{55\% davon für die Homo-Ehe}\end{array}\right\} \Longrightarrow 0.40 \cdot 0.55 = 0.22$$

- Es gilt also:

$$\Pr[\text{Ja} \mid \text{CDU}] \cdot \Pr[\text{CDU}] = \Pr[\text{CDU} \cap \text{Ja}]$$

Bedingte Wahrscheinlichkeiten IV

Definition 34

A und B seien Ereignisse mit $\Pr[B] > 0$. Die bedingte Wahrscheinlichkeit $\Pr[A|B]$ von A gegeben B ist definiert als

$$\Pr[A|B] := \frac{\Pr[A \cap B]}{\Pr[B]}\ .$$

Bedingte Wahrscheinlichkeiten V: Eigenschaften

Einige einfache Eigenschaften:

1. $\Pr[B|B] = 1$.
2. $\Pr[A|\Omega] = \Pr[A]$.
3. $\Pr[A \cap B] = \Pr[A \mid B]\Pr[B] = \Pr[B \mid A]\Pr[A]$.
4. (Satz der totalen W'keit) Seien $A_1, \ldots, A_n$ paarweise disjunkt und es gelte $B \subseteq A_1 \cup \ldots \cup A_n$. Dann folgt

$$\Pr[B] = \sum_{i=1}^{n} \Pr[B|A_i] \cdot \Pr[A_i]$$

(aus den bedingungen folgt $\Pr[B] = \sum_{i=1}^{n} \Pr[B \cap A_i]$)

Bedingte Wahrscheinlichkeiten VI

Beispiel 35 (Zweikinderproblem)

Wir nehmen an, dass bei der Geburt eines Kindes beide Geschlechter gleich wahrscheinlich sind.

Prof. Evilsparza braucht zwei kleine Kinder für bösartige Mathe-Experimente. Er besorgt sich eine Liste aller Familien in Garching mit zwei Kindern und wählt zufällig eine davon. Als er an der Tür klingelt macht ein Mädchen auf, schreit und flieht.

Frage: Mit welcher W'keit ist das andere Kind der Familie auch ein Mädchen ?

Modell: $\Omega := \{mm, mj, jm, jj\}$

$$\Pr[mm] = \Pr[mj] = \Pr[jm] = \Pr[jj] = \frac{1}{4}$$

Bedingte Wahrscheinlichkeiten VII

Wir wissen, dass das Ereignis „mindestens ein Mädchen“

$$M := \{mm, mj, jm\}$$

eingetreten ist.

Wir suchen also nach der bedingten W'keit $\Pr[mm|M]$:

$$\Pr[mm|M] = \frac{\Pr[\{mm\} \cap M]}{\Pr[M]} = \frac{1/4}{3/4} = \frac{1}{3}.$$

Bedingte Wahrscheinlichkeiten VIII

Weitere Eigenschaften:

Sei B ein Ereignis mit $\Pr[B] > 0$.

Die bedingten Wahrscheinlichkeiten $\Pr[\bullet|B]$ bilden einen neuen Wahrscheinlichkeitsraum über Ω, denn

$$\sum_{\omega\in\Omega} \Pr[\omega|B] = \sum_{\omega\in\Omega} \frac{\Pr[\omega \cap B]}{\Pr[B]} = \sum_{\omega\in B} \frac{\Pr[\omega]}{\Pr[B]} = \frac{\Pr[B]}{\Pr[B]} = 1.$$

Damit gelten alle Rechenregeln für Wahrscheinlichkeiten auch für bedingte Wahrscheinlichkeiten. Beispielsweise:

$$\begin{aligned} \Pr[A \cup C|B] &\leq \Pr[A|B] + \Pr[C|B] \\ \Pr[\bar{A}|B] &= 1 - \Pr[A|B]\,. \end{aligned}$$

Bedingte Wahrscheinlichkeiten IX: Der Multiplikationssatz

Satz 36 (Multiplikationssatz)

Wenn $\Pr[A_1 \cap \ldots \cap A_n] > 0$, *dann*

$$\begin{aligned} \Pr[A_1 \cap \ldots \cap A_n] &= \Pr[A_1] \cdot \Pr[A_2|A_1] \cdot \Pr[A_3|A_1 \cap A_2] \cdot \ldots \\ &\quad \ldots \cdot \Pr[A_n|A_1 \cap \ldots \cap A_{n-1}]\,. \end{aligned}$$

Beweis:

Alle bedingten Wahrscheinlichkeiten sind wohldefiniert, da
$\Pr[A_1] \geq \Pr[A_1 \cap A_2] \geq \ldots \geq \Pr[A_1 \cap \ldots \cap A_n] > 0$.

Die rechte Seite der Aussage können wir umschreiben zu

$$\frac{\Pr[A_1]}{1} \cdot \frac{\Pr[A_1 \cap A_2]}{\Pr[A_1]} \cdot \frac{\Pr[A_1 \cap A_2 \cap A_3]}{\Pr[A_1 \cap A_2]} \cdot \ldots \cdot \frac{\Pr[A_1 \cap \ldots \cap A_n]}{\Pr[A_1 \cap \ldots \cap A_{n-1}]}.$$

Offensichtlich kürzen sich alle Terme bis auf $\Pr[A_1 \cap \ldots \cap A_n]$. $\square$

Beispiel 37 (Geburtstagsproblem)

Wie groß ist die Wahrscheinlichkeit, dass in einer m-köpfigen Gruppe mindestens zwei Personen am selben Tag Geburtstag haben?

Umformulierung:
Man werfe b Bälle zufällig und gleich wahrscheinlich in k Körbe.

Frage: Wie groß ist die Wahrscheinlichkeit, dass nach dem Experiment jeder Ball allein in seinem Korb liegt?

Modell:

$$\Omega = \{1, \ldots, b\}^k \qquad \Pr[i_1, \ldots, i_k] = \frac{1}{k^b}$$

Für das Geburtstagsproblem: $k = 365$.

Bedingte Wahrscheinlichkeiten XI

Annahme: $b \leq k$ (sonst ist die W'keit gleich 0).

Wir werfen die Bälle nacheinander. Seien

- $A_i =$ „Ball i landet in einem noch leeren Korb"
- $A =$ „Alle Bälle liegen allein in einem Korb" .

Nach dem Multiplikationssatz gilt

$$\begin{aligned} \Pr[A] &= \Pr[A_1 \cap \dots \cap A_b] \\ &= \Pr[A_1] \cdot \Pr[A_2 | A_1] \cdot \ldots \cdot \Pr[A_i | \cap_{i=1}^{b-1} A_i]. \end{aligned}$$

Unter der Bedingung, dass die ersten $j-1$ Bälle jeweils in einem leeren Korb landen, bedeutet A_j, dass der j-te Ball in einen der $k-(j-1)$ leeren Körben fällt. Daraus folgt

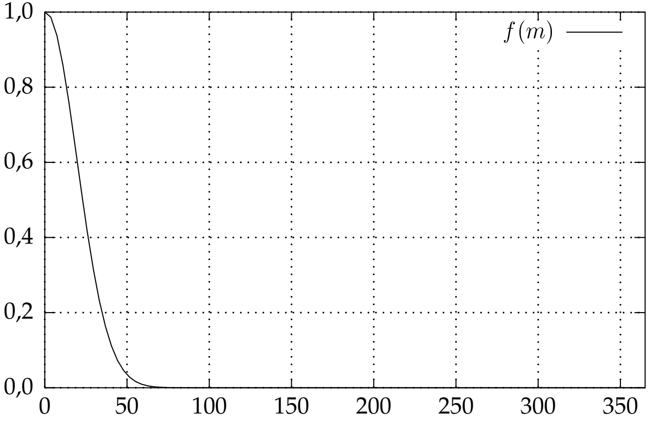
$$\Pr[A_j | A_1 \cap \dots \cap A_{j-1}] = \frac{k-(j-1)}{k} = 1 - \frac{j-1}{k}.$$

Bedingte Wahrscheinlichkeiten XII

Mit $\Pr[A_j|A_1 \cap \cdots \cap A_{j-1}] = 1 - \frac{j-1}{n}$ erhalten wir

$$
\begin{aligned}
\Pr[A] &= \Pr[A_1] \cdot \Pr[A_2|A_1] \cdot \ldots \cdot \Pr[A_b| \cap_{i=1}^{b-1} A_i] \\
&= \prod_{j=1}^{b} \left(1 - \frac{j-1}{k}\right) \\
&\leq \prod_{j=1}^{b} e^{-\frac{j-1}{k}} \qquad \left(\text{wegen } 1 - x \leq e^{-x}\right) \\
&= e^{-\frac{1}{n} \cdot \sum_{j=1}^{b-1} j} \\
&= e^{-\frac{b(b-1)}{2k}}
\end{aligned}
$$

Bedingte Wahrscheinlichkeiten XIII



Verlauf von $f(m) := e^{-b(b-1)/(2k)}$ für $n = 365$

Beispiel 38 (Die Lottosensation am 29.6.1995)

Stuttgart (dpa/lsw). Die Staatliche Toto-Lotto GmbH in Stuttgart hat eine Lottosensation gemeldet: Zum ersten Mal in der 40-jährigen Geschichte das deutschen Zahlenlottos wurden zwei identische Gewinnreihen festgestellt. Am 21. Juni dieses Jahres [3016te Ausspielung] kam im Lotto am Mittwoch in der Ziehung A die Gewinnreihe 15-25-27-30-42-48 heraus. Genau die selben Zahlen wurden bei der 1628. Ausspielung im Samstaglotto schon einmal gezogen, nämlich am 20. Dezember 1986. Welch ein Lottozufall: Unter den 49 Zahlen sind fast 14 Millionen verschiedene Sechserreihen möglich.

Bedingte Wahrscheinlichkeiten XV

Frage: Was ist die Wahrscheinlichkeit, dass die erste Wiederholung einer Gewinnreihe ($k = 13,983,816$ verschiedene Sechserreihen) spätestens bei der $b = 3016$-te Ausspielung auftritt?

Antwort: das ist die W'keit, dass mindestens zwei Bälle im selben Korb landen, i.e., $\Pr[\overline{A}]$.

$$
\begin{aligned}
\Pr[\overline{A}] &= 1 - \Pr[A] \\
&\approx 1 - e^{-\frac{b(b-1)}{2k}} \\
&= 1 - e^{-\frac{3016 \cdot 3015}{2 \cdot 13983816}} \\
&\approx 1 - e^{-0.325} \\
&\approx 0,278
\end{aligned}
$$

6. Stochastische Unabhängigkeit

Unabhängige Ereignisse I

Beispiel 39 (Zweimaliges Würfeln)

Zwei faire Würfel werden gleichzeitig gewürfelt.

$$\Omega = \{(i,j) \mid 1 \le i,j \le 6\} \qquad \Pr[(i,j)] = \frac{1}{36}\,.$$

Wir definieren die Ereignisse

$$A := \text{Augenzahl der ersten Würfel ist gerade}$$
$$B := \text{Augenzahl der zweiten Würfel ist gerade}$$

Nach unserer Intuition gewinnen wir durch das Eintreten von A keine Information in Bezug auf B hinzu.

Im Modell erwarten wir $\Pr[B \mid A] = \Pr[B]$
oder äquivalent $\Pr[A \cap B] = \Pr[A] \cdot \Pr[B]$.

Unabhängige Ereignisse II

$$\begin{aligned} A \cap B &= \{(2,2),(2,4),(2,6),\ldots,(6,2),(6,4),(6,6)\} \\ \Pr[A \cap B] &= \frac{9}{36} = \frac{1}{4} \\ \Pr[A] \cdot \Pr[B] &= \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} \end{aligned}$$

Unabhängige Ereignisse III

Intuition: Zwei Ereignisse A, B sind stochastisch unabhängig wenn das Vorwissen, dass B eintritt, keinen Einfluss auf die Wahrscheinlichkeit von A hat.

Definition 40 (Erste Definition)

Zwei Ereignisse $A, B \subseteq \Omega$ mit $\Pr[B] > 0$ sind unabhängig wenn

$$\Pr[A|B] = \Pr[A] \; .$$

Definition 41 (Allgemeinere Definition)

Zwei Ereignisse $A, B \subseteq \Omega$ sind unabhängig wenn

$$\Pr[A \cap B] = \Pr[A] \cdot \Pr[B].$$

Unabhängige Ereignisse IV

Beide Definitionen sind Äquivalent wenn $\Pr[B] > 0$:

- Wenn $\Pr[A|B] = \Pr[A]$ dann

$$\Pr[A \cap B] = \Pr[A|B] \cdot \Pr[B] = \Pr[A] \cdot \Pr[B]$$

- Wenn $\Pr[A \cap B] = \Pr[A] \cdot \Pr[B]$ und $\Pr[B] > 0$ dann

$$\Pr[A|B] = \frac{\Pr[A \cap B]}{\Pr[B]} = \Pr[A]$$

Unabhängige Ereignisse V

$A :=$ Augenzahl im ersten Wurf ist gerade

$B :=$ Augenzahl im zweiten Wurf ist gerade

$C :=$ Summe der Augenzahlen beider Würfe beträgt 7

Sind A und C unabhängig ?

$$\begin{aligned}
C &= \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\} \\
A \cap C &= \{(2,5),(4,3),(6,1)\} \\
\Pr[A \cap C] &= \frac{3}{36} = \frac{1}{12} \\
\Pr[A] \cdot \Pr[C] &= \frac{1}{2} \cdot \frac{1}{6} = \frac{1}{12}
\end{aligned}$$

Unabhängige Ereignisse VI

$A :=$ Augenzahl im ersten Wurf ist gerade

$B :=$ Augenzahl im zweiten Wurf ist gerade

$C :=$ Summe der Augenzahlen beider Würfe beträgt 7

A und B sind physikalisch unabhängig und stochastisch unabhängig.

A und C sind physikalisch abhängig aber stochastisch unabhängig.

- Physikalische Unabhängigkeit impliziert stochastische Unabhängigkeit.
- Physikalische Abhängigkeit schliesst stochastische Unabhängigkeit nicht aus.

Unabhängige Ereignismenge I

Wann sind drei Ereignisse A, B, C unabhängig ?

Intuition: Information über den Eintritt von A und B beeinflusst die W'keit von C nicht.

Erster Versuch einer Formalisierung: A, B, C sind unabhängig gdw. $\Pr[A \cap B] = \Pr[A] \cdot \Pr[B]$, $\Pr[A \cap C] = \Pr[A] \cdot \Pr[C]$, und $\Pr[B \cap C] = \Pr[B] \cdot \Pr[C]$.

Beispiel 42 (Zweimaliges Würfeln)

$A :=$ Augenzahl im ersten Wurf ist gerade,

$B :=$ Augenzahl im zweiten Wurf ist gerade,

$C :=$ Summe der Augenzahlen beider Würfe beträgt 7.

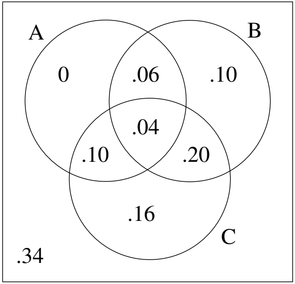
A, B, C sind paarweise unabhängig, aber z.B.

$$\Pr[C|A \cap B] = 0 \neq \Pr[C]$$

Unabhängige Ereignismenge II

Zweiter Versuch einer Formalisierung: A, B, C sind unabhängig gdw. $\Pr[A \cap B \cap C] = \Pr[A] \cdot \Pr[B] \cdot \Pr[C]$.

Beispiel 43 (Glyn George)



$$
\begin{aligned}
\Pr[A] &= 0.20\\
\Pr[B] &= 0.40\\
\Pr[A] &= 0.50\\
\Pr[A \cap B \cap C] &= 0.04\\
\Pr[A|B \cap C] &= 1.00 \neq \Pr[A]
\end{aligned}
$$

Unabhängige Ereignismenge III

Definition 44

Die Ereignisse $A_1, \ldots, A_n$ heißen unabhängig, wenn für alle $1 \leq k \leq n$ und für alle $1 \leq i_1, \ldots, i_k \leq n$ gilt, dass

$$\Pr[A_{i_1} \cap \ldots \cap A_{i_k}] = \Pr[A_{i_1}] \cdot \ldots \cdot \Pr[A_{i_k}]. \tag{1}$$

Eine unendliche Familie von Ereignissen A_i mit $i \in \mathbb{N}$ heißt unabhängig, wenn (1) für jede endliche Teilmenge $I \subseteq \mathbb{N}$ erfüllt ist.

Unabhängige Ereignismenge IV

Lemma 45

Die Ereignisse $A_1, \ldots, A_n$ sind genau dann unabhängig, wenn für alle $(s_1, \ldots, s_n) \in \{0,1\}^n$ gilt, dass

$$\Pr[A_1^{s_1} \cap \ldots \cap A_n^{s_n}] = \Pr[A_1^{s_1}] \cdot \ldots \cdot \Pr[A_n^{s_n}], \tag{2}$$

wobei $A_i^0 = \bar{A}_i$ und $A_i^1 = A_i$.

Unabhängige Ereignismenge V

Beweis (1) $\Rightarrow$ (2):

Induktion über die Anzahl der Nullen in $s_1, \ldots, s_n$.

Wenn $s_1 = \ldots = s_n = 1$ gilt, so ist nichts zu zeigen.

Sonst gelte o.E. $s_1 = 0$ und sei $B = A_2^{s_2} \cap \ldots \cap A_n^{s_n}$.

$$
\begin{aligned}
& \Pr[A_1^{s_1} \cap \cdots \cap A_n^{s_n}] \\
&= \Pr[\bar{A}_1 \cap B] && (s_1 = 0 \text{ und Def. } B) \\
&= \Pr[B] - \Pr[A_1 \cap B] \\
&= \textstyle\prod_{i=2}^{n} \Pr[A_i^{s_i}] - \Pr[A_1] \cdot \prod_{i=2}^{n} \Pr[A_i^{s_i}] && (\text{Ind.Vor}) \\
&= (1 - \Pr[A_1]) \cdot \textstyle\prod_{i=2}^{n} \Pr[A_i^{s_i}] \\
&= \Pr[\bar{A}_1] \cdot \textstyle\prod_{i=2}^{n} \Pr[A_i^{s_i}] \\
&= \textstyle\prod_{i=1}^{n} \Pr[A_i^{s_i}]
\end{aligned}
$$

Unabhängige Ereignismenge VI

Beweis (2) $\Rightarrow$ (1):

Wir zeigen nur (2) $\Rightarrow \Pr[A_1 \cap A_2] = \Pr[A_1] \cdot \Pr[A_2]$.

Es gilt (Satz der totalen Wahrscheinlichkeit):

$$\begin{aligned}
\Pr[A_1 \cap A_2] &= \sum_{s_3,\ldots,s_n \in \{0,1\}} \Pr[A_1 \cap A_2 \cap A_3^{s_3} \cap \ldots \cap A_n^{s_n}] \\
&= \sum_{s_3,\ldots,s_n \in \{0,1\}} \Pr[A_1] \cdot \Pr[A_2] \cdot \Pr[A_3^{s_3}] \cdot \ldots \cdot \Pr[A_n^{s_n}] \\
&= \Pr[A_1] \cdot \Pr[A_2] \cdot \sum_{s_3=0,1} \Pr[A_3^{s_3}] \cdot \ldots \cdot \sum_{s_n=0,1} \Pr[A_n^{s_n}] \\
&= \Pr[A_1] \cdot \Pr[A_2]
\end{aligned}$$

Unabhängige Ereignismenge VII

Lemma 46

Seien A, B und C unabhängige Ereignisse.

1. *A und $\bar{B}$ sind unabhängig, sowie $\bar{A}, B$, und $\bar{A}, \bar{B}$.*
2. *$A \cap B$ und C sind unabhängig.*
3. *$A \cup B$ und C sind unabhängig*

Beweis:

(1) Folgt aus Lemma 45 mit $n = 2$.

(2) Zu zeigen: $\Pr[(A \cap B) \cap C] = \Pr[A \cap B] \cdot \Pr[C]$.

$$\begin{aligned}
& \Pr[(A \cap B) \cap C] \\
= \; & \Pr[A \cap B \cap C] \\
= \; & \Pr[A] \cdot \Pr[B] \cdot \Pr[C] \quad (A, B, C \text{ unabh.}) \\
= \; & \Pr[A \cap B] \cdot \Pr[C] \qquad (A, B \text{ unabh.})
\end{aligned}$$

Unabhängige Ereignismenge VIII

(3) Zu zeigen: $\Pr[(A \cup B) \cap C] = \Pr[A \cup B] \cdot \Pr[C]$.

$$
\begin{aligned}
& \Pr[(A \cup B) \cap C] \\
= \quad & \Pr[(A \cap C) \cup (B \cap C)] \\
= \quad & \Pr[A \cap C] + \Pr[B \cap C] - \Pr[A \cap B \cap C] \\
= \quad & \Pr[C] \cdot (\Pr[A] + \Pr[B] - \Pr[A \cap B]) \qquad (A, B, C \text{ unabh.}) \\
= \quad & \Pr[A \cup B] \cdot \Pr[C]
\end{aligned}
$$

Unabhängige Zufallsvariablen I

Definition 47

Die Zufallsvariablen $X_1, \ldots, X_n$ heißen unabhängig, wenn für alle $(x_1, \ldots, x_n) \in W_{X_1} \times \ldots \times W_{X_n}$ gilt

$$\Pr[X_1 = x_1, \ldots, X_n = x_n] = \Pr[X_1 = x_1] \cdot \ldots \cdot \Pr[X_n = x_n]\,.$$

Alternative Definition:

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdot \ldots \cdot f_{X_n}(x_n)\,.$$

Noch eine Äquivalente Definition:

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = F_{X_1}(x_1) \cdot \ldots \cdot F_{X_n}(x_n)\,.$$

Unabhängige Zufallsvariablen II

Satz 48

Seien $X_1, \ldots, X_n$ unabhängige Zufallsvariablen und $S_1, \ldots, S_n$ beliebige Mengen mit $S_i \subseteq W_{X_i}$. Dann sind die Ereignisse „$X_1 \in S_1$“, ..., „$X_n \in S_n$“ unabhängig.

Beweis:

$$
\begin{aligned}
& \Pr[X_1 \in S_1, \ldots, X_n \in S_n] \\
= \quad & \sum_{x_1 \in S_1} \ldots \sum_{x_n \in S_n} \Pr[X_1 = x_1, \ldots, X_n = x_n] \\
\overset{\text{Unabh.}}{=} \quad & \sum_{x_1 \in S_1} \ldots \sum_{x_n \in S_n} \Pr[X_1 = x_1] \cdot \ldots \cdot \Pr[X_n = x_n] \\
= \quad & \left(\sum_{x_1 \in S_1} \Pr[X_1 = x_1] \right) \cdot \ldots \cdot \left(\sum_{x_n \in S_n} \Pr[X_n = x_n] \right) \\
= \quad & \Pr[X_1 \in S_1] \cdot \ldots \cdot \Pr[X_n \in S_n] \,.
\end{aligned}
$$

Unabhängige Zufallsvariablen III

Satz 49

$f_1, \ldots, f_n$ seien reellwertige Funktionen ($f_i : \mathbb{R} \to \mathbb{R}$ für $i = 1, \ldots, n$). Wenn die Zufallsvariablen $X_1, \ldots, X_n$ unabhängig sind, dann gilt dies auch für $f_1(X_1), \ldots, f_n(X_n)$.

Beweis:

Sei $z_i \in W_{f(X_i)}$ für $i = 1, \ldots, n$ und $S_i = \{x;\ f(x) = z_i\}$.

$$
\begin{aligned}
& \Pr[f_1(X_1) = z_1, \ldots, f_n(X_n) = z_n] \\
= \quad & \Pr[X_1 \in S_1, \ldots, X_n \in S_n] \\
\overset{\text{Unabh.}}{=} \quad & \Pr[X_1 \in S_1] \cdot \ldots \cdot \Pr[X_n \in S_n] \\
= \quad & \Pr[f_1(X_1) = z_1] \cdot \ldots \cdot \Pr[f_n(X_n) = z_n] \,.
\end{aligned}
$$

$\square$

Unabhängige Zufallsvariablen IV

Satz 50

Seien X, Y unabhängige Zufallsvariablen und $Z := X + Y$. Es gilt

$$f_Z(z) = \sum_{x \in W_X} f_X(x) \cdot f_Y(z - x)\,.$$

Beweis: Mit Hilfe des Satzes von der totalen Wahrscheinlichkeit:

$$\begin{aligned}
f_Z(z) &= \Pr[Z = z] = \sum_{x \in W_X} \Pr[X + Y = z | X = x] \cdot \Pr[X = x] \\
&= \sum_{x \in W_X} \Pr[Y = z - x \mid X = x] \cdot \Pr[X = x] \\
&\overset{\text{Unab.}}{=} \sum_{x \in W_X} \Pr[Y = z - x] \cdot \Pr[X = x] \\
&= \sum_{x \in W_X} f_X(x) \cdot f_Y(z - x)\,.
\end{aligned}$$

Beispiel 51

Zwei faire Würfel werden geworfen. Seien X, Y die Zufallsvariablen, die die Augenzahl des ersten bzw. des zweiten Würfels angeben. Wir berechnen die Dichte von $Z := X + Y$:

$$\begin{aligned}\Pr[Z = z] &= \sum_{x \in W_X} \Pr[X = x] \cdot \Pr[Y = z - x] \\ &= \sum_{x=1}^{6} \frac{1}{6} \cdot \Pr[Y = z - x] = \sum_{x=\max\{1,z-6\}}^{\min\{6,z-1\}} \frac{1}{36}.\end{aligned}$$

Für $2 \leq z \leq 7$ erhalten wir

$$\Pr[Z = z] = \sum_{i=1}^{z-1} \frac{1}{36} = \frac{z-1}{36} \,.$$

Unabhängige Zufallsvariablen VI

Und für $7 < z \leq 12$:

$$\Pr[Z = z] = \frac{13 - z}{36}\,.$$

Den Ausdruck

$$\sum_{x \in W_X} f_X(x) \cdot f_Y(z - x)$$

aus Satz 50 nennt man in Analogie zu den entsprechenden Begriffen bei Potenzreihen auch Faltung oder Konvolution der Dichten f_X und f_Y.

7. Satz von Bayes

Beispiel 52 (HIV-Test)

Krankheitstests erkennen machmal kranke Menschen als gesund ("falsch negativ", sehr gefährlich) oder gesunde Menschen als krank ("falsch positiv", weniger gefährlich).

Für den ELISA-Test zur Erkennung von Antikörpern gegen das HIV beträgt die W'keit eines falsch negativen Resultates etwa 0.1%. Die W'keit eines falsch positiven Resultates beträgt 0.2%.

In Deutschland leben ca. 80,000 HIV-Infizierte, etwa 0.1% der Bevölkerung.

Eine zufällig gewählte Person (in Deutschland lebend) wird getestet. Das Ergebnis ist positiv.

Frage: Wie hoch ist die W'keit einer Infektion?

Satz von Bayes II

Seien I, G die Ereignisse “die Person ist infiziert” und “die Person ist gesund”.

$$\Pr[I] \approx 0.001 \qquad \Pr[G] = 1 - \Pr[I] \approx 0.999$$

Seien P, N die Ereignisse “das Ergebnis des Tests ist positiv” und “das Ergebnis des Tests ist negativ”.

$$\Pr[P|I] = 0.999 \qquad \Pr[N|G] = 0.998$$

Wir suchen $\Pr[I|P]$.

Satz von Bayes III

$$\Pr[I] \approx 0.001 \quad \Pr[G] \approx 0.999 \quad \Pr[P|I] = 0.999 \quad \Pr[N|G] = 0.998$$

Es gilt $\Pr[I|P] \cdot \Pr[P] = \Pr[I \cap P] = \Pr[P|I] \cdot \Pr[I]$.

$$\begin{aligned}
\Pr[I|P] &= \frac{\Pr[P|I] \cdot \Pr[I]}{\Pr[P]} \\
&= \frac{\Pr[P|I] \cdot \Pr[I]}{\Pr[P|G] \cdot P[G] + \Pr[P|I] \cdot P[I]} \\
&= \frac{\Pr[P|I] \cdot \Pr[I]}{(1 - \Pr[N|G]) \cdot P[G] + \Pr[P|I] \cdot P[I]} \\
&= \frac{0.999 \cdot 0.001}{0.002 \cdot 0.999 + 0.999 \cdot 0.001} \\
&= \frac{0.001}{0.003} = \frac{1}{3}
\end{aligned}$$

Satz von Bayes IV

Satz 53 (Satz von Bayes)

Die Ereignisse $A_1, \ldots, A_n$ *seien paarweise disjunkt, mit* $\Pr[A_j] > 0$ *für alle* j*. Ferner sei* $B \subseteq A_1 \cup \ldots \cup A_n$ *ein Ereignis mit* $\Pr[B] > 0$*. Dann gilt für ein beliebiges* $i = 1, \ldots, n$

$$\Pr[A_i|B] = \frac{\Pr[A_i \cap B]}{\Pr[B]} = \frac{\Pr[B|A_i] \cdot \Pr[A_i]}{\sum_{j=1}^{n} \Pr[B|A_j] \cdot \Pr[A_j]}\ .$$

Analog gilt für paarweise disjunkte Ereignisse $A_1, A_2, \ldots$ *mit* $B \subseteq \bigcup_{i=1}^{\infty} A_i$*, dass*

$$\Pr[A_i|B] = \frac{\Pr[A_i \cap B]}{\Pr[B]} = \frac{\Pr[B|A_i] \cdot \Pr[A_i]}{\sum_{j=1}^{\infty} \Pr[B|A_j] \cdot \Pr[A_j]}\ .$$

Satz von Bayes V

Seien $\mathcal{K}_1, \ldots, \mathcal{K}_n$ Krankheiten und $\mathcal{S}_1, \ldots \mathcal{S}_m$ Symptome.

Wir definieren die Ereignisse:

$$
\begin{aligned}
K_i &= \textit{Der Patient hat die Krankheit } \mathcal{K}_i \\
S_j &= \textit{Der Patient zeigt das Symptom } \mathcal{S}_j
\end{aligned}
$$

Eine Belegung der Symptome ist ein Element von $\{0,1\}^m$.

Jedem Patient kann eine Belegung $\mathcal{B}$ zugeordnet werden, die angibt, welche Symptome im Patient vorhanden ($\mathcal{B}(j) = 1$) und abwesend ($\mathcal{B}(j) = 0$) sind.

Satz von Bayes VI

Wir definieren das Ereignis:

$B =$ *Der Patient zeigt die Belegung* $\mathcal{B}$*.*

Es gilt $B = \bigcap_{j=1}^{m} U_j$, wobei

$$U_j = \begin{cases} S_j & \text{falls } \mathcal{B}(j) = 1 \\ \overline{S}_j & \text{sonst} \end{cases}$$

Der Satz von Bayes kann nun mit $A_i := K_i$ angewendet werden, unter der Annahme, dass der Patient eine und genau eine der Krankheiten $\mathcal{K}_1, \ldots, \mathcal{K}_n$ hat:

$$\Pr[K_i|B] = \frac{\Pr[B|K_i] \cdot \Pr[K_i]}{\Pr[B]} \; \frac{\Pr[B|K_i] \cdot \Pr[K_i]}{\sum_{j=1}^{\infty} \Pr[B|K_j] \cdot \Pr[K_j]} \, .$$

Satz von Bayes VII

Wenn $S_1 \cap K_j, \ldots, S_m \cap K_j$ unabhängig sind für alle $1 \le j \le n$:

$$\begin{aligned}
\Pr[B|K_j] \cdot \Pr[K_j] &= \Pr[B \cap K_j] \\
&= \Pr[\bigcap_{k=1}^{m} U_k \cap K_j] \\
&= \Pr[\bigcap_{k=1}^{m} (U_k \cap K_j)] \\
&\overset{\text{Unabh.}}{=} \prod_{k=1}^{m} \Pr[U_k \cap K_j] \\
&= \prod_{k=1}^{m} (\Pr[U_k|K_j] \cdot \Pr[K_j]) \\
&= \Pr[K_j]^m \cdot \prod_{k=1}^{m} \Pr[U_k|K_j]
\end{aligned}$$

Satz von Bayes VIII

Durch Einsetzen erhalten wir:

$$\Pr[K_i|B] = \frac{\Pr[K_i]^m \cdot \prod_{k=1}^{m} \Pr[U_k|K_i]}{\sum_{j=1}^{n} \Pr[K_j]^m \cdot \prod_{k=1}^{m} \Pr[U_k|K_j]}$$

mit

$$\Pr[U_k|K_i] = \begin{cases} \Pr[S_k|K_i] & \text{falls } \mathcal{B}(k) = 1 \\ \Pr[\overline{S}_k|K_i] = 1 - \Pr[S_k|K_i] & \text{sonst} \end{cases}$$

Beispiel 54 (Diagnose von Hirntumoren)

Microscopic description: The H&E sections show a tumor with a predominately papillary or filiform architecture (Figures 2 and 3). The cells are arranged in a pseudocolumnar fashion forming perivascular palisades about a fibrovascular core. There is marked nuclear pleomorphism with many bizarre nuclei, and a brisk mitotic rate (Figure 4), the nucleoli are indistinct. In areas, the papillary configuration gives way to patternless sheets of cells interrupted by zonal necrosis. Focally, the tumor cells surround microcavities which are either optically empty or contain "wisps" of eosinophilic fibrillary material. Confluent calcospherites are focally prominent. Vascular endothelial proliferation is present. Foci of glial frame indicate invasiveness. Rests of native plexus are occasionally encountered.

Satz von Bayes X

A priori Wahrscheinlichkeiten:

Diagnose	Wahrscheinlichkeit
Grade I	0.25
Grade II	0.25
Grade III	0.20
Grade IV	0.30

Satz von Bayes XI

Bedingte Wahrscheinlichkeiten:

		Gr. I	Gr. II	Gr. III	Gr. IV
Diffuse infiltration	DI	0.364	0.878	0.896	0.819
Necrosis	NE	0.221	0.246	0.716	0.850
Vascular abnormalities	VA	0.648	0.686	0.849	0.926
Vascular occlusions	VO	0.113	0.121	0.518	0.640
Nuclear polymorphism	NP	0.702	0.825	0.971	0.968
Cellular polymorphism	CP	0.362	0.706	0.801	0.833
Visible perycarion	VP	0.494	0.865	0.896	0.887
Typical mitoses	TM	0.121	0.109	0.726	0.853
Atipical mitoses	AM	0.005	0.029	0.556	0.733
Undifferentiated cells	UC	0.413	0.450	0.651	0.943

Satz von Bayes XII

Thomas Bayes (1702–1761) war ein bekannter Theologe und Mitglied der Royal Society. Als sein bedeutendstes Werk gilt sein Beitrag zur Wahrscheinlichkeitstheorie „Essay Towards Solving a Problem in the Doctrine of Chances“. Diese Arbeit wurde erst 1763 publiziert.

Teil III

Rechnen mit Wahrscheinlichkeiten

8. Rechenregeln für Erwartungswert und Varianz

Rechenregeln für den Erwartungswert I

Satz 55

Für jede Zufallsvariable X gilt

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \cdot \Pr[\omega]\,.$$

Beweis:

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{x \in W_X} x \cdot \Pr[X = x] \\
&= \sum_{x \in W_X} x \sum_{\{\omega \in \Omega \mid X(\omega) = x\}} \Pr[\omega] \\
&= \sum_{\omega \in \Omega} X(\omega) \cdot \Pr[\omega]\,.
\end{aligned}$$

Rechenregeln für den Erwartungswert II

Satz 56

Sei X eine Zufallsvariable mit $W_X \subseteq \mathbb{N}_0$. Dann gilt

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} \Pr[X \geq i]\,.$$

Beweis:

$$\begin{aligned}\mathbb{E}[X] &= \sum_{i=0}^{\infty} i \cdot \Pr[X = i] = \sum_{i=1}^{\infty} \sum_{j=1}^{i} \Pr[X = i] \\ &= \sum_{j=1}^{\infty} \sum_{i \geq j}^{\infty} \Pr[X = i] = \sum_{j=1}^{\infty} \Pr[X \geq j]\,.\end{aligned}$$

Rechenregeln für den Erwartungswert III

Beispiel 57

Zwei faire Würfel werden geworfen. Der Spieler gewinnt k Euro, wobei k die kleinere der zwei Augenzahlen ist.

Frage: Wieviel Euro gewinnt der Spieler in Durchschnitt?

$$\Omega = \{(1,1),(1,2),\ldots,(6,6)\} \qquad \Pr[(i,j)] = \frac{1}{36}$$

Wir definieren zwei Zufallsvariablen $X_1, X_2 \colon \Omega \to \mathbb{R}$:

$$X_1(i,j) := i \qquad X_2(i,j) := j$$

Wir müssen $\mathbb{E}[\min\{X_1, X_2\}]$ berechnen.

Rechenregeln für den Erwartungswert IV

Wir verwenden Satz 56.

$$\begin{aligned}
\mathbb{E}[\min\{X_1, X_2\}] &= \sum_{i=1}^{6} \Pr[\min\{X_1, X_2\} \geq i] \\
&= \sum_{i=1}^{6} \Pr[„X_1 \geq i\text{“} \cap „X_2 \geq i\text{“}] \\
&\overset{\text{Unabh.}}{=} \sum_{i=1}^{6} \Pr[X_1 \geq i] \cdot \Pr[X_2 \geq i] \\
&= \sum_{i=1}^{6} (\Pr[X_1 \geq i])^2 \\
&= \sum_{i=1}^{6} \left(\frac{(7-i)}{6}\right)^2 \\
&= \frac{91}{36} \approx 2.528
\end{aligned}$$

Definition 58

Sei X eine Zufallsvariable und A ein Ereignis mit $\Pr[A] > 0$. Die bedingte Zufallsvariable $X|A$ besitzt die Dichte

$$f_{X|A}(x) := \Pr[X = x \mid A] = \frac{\Pr[\text{„}X = x\text{“} \cap A]}{\Pr[A]}\,.$$

Die Definition von $f_{X|A}$ ist zulässig, da

$$\sum_{x \in W_X} f_{X|A}(x) = \sum_{x \in W_X} \frac{\Pr[\text{„}X = x\text{“} \cap A]}{\Pr[A]} = \frac{\Pr[A]}{\Pr[A]} = 1\,.$$

Der Erwartungswert $\mathbb{E}[X|A]$ der Zufallsvariablen $X|A$ berechnet sich entsprechend:

$$\mathbb{E}[X|A] = \sum_{x \in W_X} x \cdot f_{X|A}(x)\,.$$

Rechenregeln für den Erwartungswert VI

Satz 59

Sei X eine Zufallsvariable. Für paarweise disjunkte Ereignisse $A_1, \ldots, A_n$ mit $A_1 \cup \ldots \cup A_n = \Omega$ und $\Pr[A_1], \ldots, \Pr[A_n] > 0$ gilt

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X|A_i] \cdot \Pr[A_i]\,.$$

Für paarweise disjunkte Ereignisse $A_1, A_2, \ldots$ mit $\bigcup_{i=1}^{\infty} A_k = \Omega$ und $\Pr[A_1]$, $\Pr[A_2], \ldots > 0$ gilt analog

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} \mathbb{E}[X|A_i] \cdot \Pr[A_i],$$

sofern die Erwartungswerte auf der rechten Seite alle existieren und die Summe $\sum_{i=1}^{\infty} |\mathbb{E}[X|A_i]| \cdot \Pr[A_i]$ konvergiert.

Rechenregeln für den Erwartungswert VII

Beweis:

$$
\begin{aligned}
\mathbb{E}[X] &= \sum_{x \in W_X} x \cdot \Pr[X = x] \\
&= \sum_{x \in W_X} x \cdot \sum_{i=1}^{n} \Pr[X = x | A_i] \cdot \Pr[A_i] \quad \text{(Tot. W'keit)} \\
&= \sum_{i=1}^{n} \Pr[A_i] \sum_{x \in W_X} x \cdot \Pr[X = x | A_i] \\
&= \sum_{i=1}^{n} \Pr[A_i] \cdot \mathbb{E}[X | A_i].
\end{aligned}
$$

Der Beweis für den unendlichen Fall verläuft analog.

Rechenregeln für den Erwartungswert VIII

Beispiel 60

Wir werfen eine Münze so lange, bis zum ersten Mal „Kopf" erscheint. Dies geschehe in jedem Wurf unabhängig mit Wahrscheinlichkeit p. Wir definieren dazu die Zufallsvariable $X :=$ „Anzahl der Würfe". Wir haben bereits gesehen, dass

$$\Pr[X = k] = p(1-p)^{k-1}$$

und damit

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} k \cdot p(1-p)^{k-1} = p \cdot \frac{1}{(1-(1-p))^2} = \frac{1}{p}\,.$$

Rechenregeln für den Erwartungswert IX

Alternative Berechnungsmethode: (gestützt auf Satz 59)
Sei $K_1 :=$ „Im ersten Wurf fällt Kopf“.

Offensichtlich gilt $\mathbb{E}[X|K_1] = 1$.

Nehmen wir nun an, dass im ersten Wurf nicht „Kopf“ gefallen ist. Wir starten das Experiment neu. Sei X' die Anzahl der Würfe bis zum ersten Auftreten von „Kopf“ im neu gestarteten Experiment.

Wegen der Gleichheit der Experimente gilt $\mathbb{E}[X'] = \mathbb{E}[X]$.

Damit schließen wir $\mathbb{E}[X|\bar{K}_1] = 1 + \mathbb{E}[X'] = 1 + \mathbb{E}[X]$

und erhalten eine einfache Gleichung für $\mathbb{E}[X]$:

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}[X|K_1] \cdot \Pr[K_1] + \mathbb{E}[X|\bar{K}_1] \cdot \Pr[\bar{K}_1] \\ &= 1 \cdot p + (1 + \mathbb{E}[X]) \cdot (1-p)\,.\end{aligned}$$

mit Lösung $\mathbb{E}[X] = 1/p$.

Rechenregeln für den Erwartungswert X

Formale Ableitung von $\mathbb{E}[X|\bar{K}_1] = 1 + \mathbb{E}[X]$:

$$\begin{aligned}
\mathbb{E}[X|\bar{K}_1] &= \sum_{n=1}^{\infty} n \cdot \frac{\Pr[X = n \cap \bar{K}_1]}{\Pr[\bar{K}_1]} \\
&= \sum_{n=2}^{\infty} n \cdot \frac{\Pr[z^{n-1}k]}{\Pr[\bar{K}_1]} = \sum_{n=1}^{\infty} (n+1) \cdot \frac{\Pr[z^n k]}{\Pr[\bar{K}_1]} \\
&= \sum_{n=1}^{\infty} (n+1) \cdot \frac{\Pr[\bar{K}_1] \cdot \Pr[z^{n-1}k]}{\Pr[\bar{K}_1]} \\
&= \sum_{n=1}^{\infty} (n+1) \cdot \Pr[z^{n-1}k] \\
&= \sum_{n=1}^{\infty} n \cdot \Pr[z^{n-1}k] + \sum_{n=1}^{\infty} \Pr[z^{n-1}k] = \mathbb{E}[X] + 1
\end{aligned}$$

Rechenregeln für den Erwartungswert XI

Beispiel 61

Wenn man eine Sechs wirft dann darf man bei Mensch-ärgere-dich-nicht bekanntermaßen nochmal werfen. In Spanien gilt jedoch noch folgende Regel: wirft man dreimal hintereinander eine Sechs, muss man den Stein, den man zuletzt gezogen hat, zurück zum Startpunkt bringen.

Frage: Wie oft muss man einen fairen Würfel im Mittel werfen, um dreimal hintereinander eine Sechs zu würfeln?

Rechenregeln für den Erwartungswert XII

Als Elementarereignisse für jede Stufe des Experiments (einmal würfeln) nehmen wir 6 (eine Sechs wird gewürfelt) mit $\Pr[6] = 1/6$ und $\overline{6}$ (keine Sechs wird gewürfelt) mit $\Pr[\overline{6}] = 5/6$.

Modell:

$$\Omega = \left((\epsilon + 6 + 66)\overline{6}\right)^* 666 \qquad \Pr[\omega] = \left(\frac{1}{6}\right)^{s(\omega)} \cdot \left(\frac{5}{6}\right)^{|\omega| - s(\omega)}$$

mit $s(\omega) =$ Anzahl von Sechsen in ω.

Wir definieren drei Ereignisse:

$$\begin{aligned} \mathbf{\overline{6}} &= \text{Elemente aus } \Omega\text{, die mit } \overline{6} \text{ anfangen} \\ \mathbf{6\overline{6}} &= \text{Elemente aus } \Omega\text{, die mit } 6\overline{6} \text{ anfangen} \\ \mathbf{66\overline{6}} &= \text{Elemente aus } \Omega\text{, die mit } 66\overline{6} \text{ anfangen} \end{aligned}$$

Es gilt

$$\Omega = \mathbf{\overline{6}} \cup \mathbf{6\overline{6}} \cup \mathbf{66\overline{6}} + \{666\}$$

Rechenregeln für den Erwartungswert XIII

Sei X die Zufallsvariable, die die Gesamtzahl der Würfel angibt.
Mit Satz 59 gilt

$$\begin{aligned} E[X] &= E[X \mid \overline{\mathbf{6}}] \cdot \Pr[\overline{6}] + E[X \mid \mathbf{6}\overline{\mathbf{6}}] \cdot \Pr[\mathbf{6}\overline{\mathbf{6}}] \\ &\quad + E[X \mid \mathbf{66}\overline{\mathbf{6}}] \cdot \Pr[\mathbf{66}\overline{\mathbf{6}}] + E[X \mid 666] \cdot \Pr[666] \end{aligned}$$

Wir haben

$$\begin{aligned} E[X \mid \overline{\mathbf{6}}] &= 1 + E[X] & \qquad E[X \mid \mathbf{66}\overline{\mathbf{6}}] &= 3 + E[X] \\ E[X \mid \mathbf{6}\overline{\mathbf{6}}] &= 2 + E[X] & \qquad E[X \mid 666] &= 3 \end{aligned}$$

und damit

$$E[X] = \frac{5}{6}(1 + E[X]) + \frac{5}{36}(2 + E[X]) + \frac{5}{216}(3 + E[X]) + \frac{3}{216}$$

mit Lösung $E[X] = 258$.

Rechenregeln für den Erwartungswert XIV

Satz 62 (Linearität des Erwartungswerts)

Für beliebige Zufallsvariablen $X_1, \ldots, X_n$ und $a_1, \ldots, a_n \in \mathbb{R}$ gilt

$$\mathbb{E}[a_1 X_1 + \cdots + a_n X_n] = a_1 \mathbb{E}[X_1] + \cdots + a_n \mathbb{E}[X_n]\,.$$

Beweis:

$$\begin{aligned}
&\mathbb{E}[a_1 X_1 + \ldots + a_n X_n] \\
&= \sum_{\omega \in \Omega} (a_1 \cdot X_1(\omega) + \ldots + a_n \cdot X_n(\omega)) \cdot \Pr[\omega] \\
&= a_1 \cdot \left(\sum_{\omega \in \Omega} X_1(\omega) \cdot \Pr[\omega] \right) + \cdots + a_n \cdot \left(\sum_{\omega \in \Omega} X_n(\omega) \cdot \Pr[\omega] \right) \\
&= a_1 \cdot \mathbb{E}[X_1] + \ldots + a_n \cdot \mathbb{E}[X_n]\,.
\end{aligned}$$

$\square$

Beispiel 63

n betrunkene Seeleute torkeln nach dem Landgang in ihre Kojen. Sie haben völlig die Orientierung verloren, weshalb wir annehmen, dass jede Zuordnung der Seeleute zu den n Betten gleich wahrscheinlich ist (genau ein Seemann pro Bett).

Frage: Wie viele Seeleute liegen im Mittel im richtigen Bett?

Die Anzahl der Seeleute im richtigen Bett zählen wir mit der Zufallsvariablen X, die als Summe der Zufallsvariablen $X_1, \ldots, X_n$ dargestellt wird, wobei

$$X_i := \begin{cases} 1 & \text{falls Seemann } i \text{ in seinem Bett liegt,} \\ 0 & \text{sonst.} \end{cases}$$

Offenbar gilt $X := X_1 + \cdots + X_n$.

Rechenregeln für den Erwartungswert XVI

Für die Variablen X_i erhalten wir

$$\Pr[X_i = 1] = \frac{1}{n}$$

da jedes Bett von Seemann i mit gleicher Wahrscheinlichkeit aufgesucht wird.

Daraus folgt

$$\mathbb{E}[X_i] = 0 \cdot \Pr[X_i = 0] \; + \; 1 \cdot \Pr[X_i = 1] = \frac{1}{n},$$

und somit

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=1}^{n} \frac{1}{n} = 1\,.$$

Im Mittel sucht also nur ein Seemann sein eigenes Bett auf.

Satz 64 (Multiplikativität des Erwartungswerts)

Für unabhängige Zufallsvariablen $X_1, \ldots, X_n$ gilt

$$\mathbb{E}[X_1 \cdot \ldots \cdot X_n] = \mathbb{E}[X_1] \cdot \ldots \cdot \mathbb{E}[X_n]\,.$$

Beweis: Wir beweisen den Fall $n = 2$. Der allgemeine Fall ist analog.

$$\begin{aligned}
\mathbb{E}[X \cdot Y] &= \sum_{x \in W_X} \sum_{y \in W_Y} xy \cdot \Pr[X = x, Y = y] \\
&\overset{\text{Unabh.}}{=} \sum_{x \in W_X} \sum_{y \in W_Y} xy \cdot \Pr[X = x] \cdot \Pr[Y = y] \\
&= \sum_{x \in W_X} x \cdot \Pr[X = x] \sum_{y \in W_Y} y \cdot \Pr[Y = y] \\
&= \mathbb{E}[X] \cdot \mathbb{E}[Y]\,.
\end{aligned}$$

Rechenregeln für den Erwartungswert XVIII

Für die Gültigkeit von Satz 64 ist die Unabhängigkeit der Zufallsvariablen wirklich notwendig.

Sei $Y = -X$ für eine Zufallsvariable mit einer von Null verschiedenen Varianz. Dann gilt

$$\mathbb{E}[X \cdot Y] = -\mathbb{E}[X^2] \neq -(\mathbb{E}[X])^2 = \mathbb{E}[X] \cdot \mathbb{E}[Y]\,.$$

Rechenregeln für die Varianz I

Satz 65

Für eine beliebige Zufallsvariable X und $a, b \in \mathbb{R}$ gilt

$$\mathrm{Var}[a \cdot X + b] = a^2 \cdot \mathrm{Var}[X]\,.$$

Beweis:

Aus der in Satz 62 gezeigten Linearität des Erwartungswerts folgt $\mathbb{E}[X + b] = \mathbb{E}[X] + b$.

Zusammen mit der Definition der Varianz ergibt sich damit sofort

$$\mathrm{Var}[X + b] = \mathbb{E}[(X + b - \mathbb{E}[X + b])^2] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathrm{Var}[X]\,.$$

Weiter folgt mit Satz 27:

$$\mathrm{Var}[a \cdot X] = \mathbb{E}[(aX)^2] - \mathbb{E}[aX]^2 = a^2\mathbb{E}[X^2] - (a\mathbb{E}[X])^2 = a^2 \cdot \mathrm{Var}[X]\,,$$

und daraus zusammen die Behauptung.

Rechenregeln für die Varianz II

Satz 66 (Additivität der Varianz)

Für unabhängige Zufallsvariablen $X_1, \ldots, X_n$ und $X := X_1 + \ldots + X_n$ gilt

$$\mathrm{Var}[X] = \mathrm{Var}[X_1] + \ldots + \mathrm{Var}[X_n]\,.$$

Beweis:

Fall $n = 2$ mit Zufallsvariablen X und Y.

$$\mathbb{E}[(X+Y)^2] = \mathbb{E}[X^2 + 2XY + Y^2] = \mathbb{E}[X^2] + 2\mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[Y^2]$$
$$\mathbb{E}[X+Y]^2 = (\mathbb{E}[X] + \mathbb{E}[Y])^2 = \mathbb{E}[X]^2 + 2\mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[Y]^2$$

Wir ziehen die zweite Gleichung von der ersten ab und erhalten

$$\mathbb{E}[(X+Y)^2] - \mathbb{E}[X+Y]^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2 + \mathbb{E}[Y^2] - \mathbb{E}[Y]^2\,.$$

Mit Hilfe von Satz 27 folgt die Behauptung. □

Rechenregeln für die Varianz IV

Für abhängige Zufallsvariablen $X_1, \ldots, X_n$ gilt Satz 66 im Allgemeinen nicht. Als Beispiel funktioniert wiederum der Fall $X = -Y$:

$$\mathrm{Var}[X + Y] = 0 \neq 2 \cdot \mathrm{Var}[X] = \mathrm{Var}[X] + \mathrm{Var}[Y]\,.$$

9. Indikatorvariablen

Indikatorvariablen I

Indikatorvariablen erlauben die Darstellung von W'keiten als Erwartungswerte.

Dadurch können Rechenregeln für den Erwartungswert auch für die Berechnung von W'keiten angewendet werden.

Definition 67

Zu einem Ereignis A heißt die Zufallsvariable

$$I_A := \begin{cases} 1 & \text{falls } A \text{ eintritt,} \\ 0 & \text{sonst} \end{cases}$$

Indikatorvariable des Ereignisses A.

(Die Variablen $X_1, \ldots, X_n$ im Matrosen-Beispiel sind Indikatorvariablen.)

Indikatorvariablen II

Lemma 68

Seien $A, A_1, \ldots, A_n$ Ereignisse (nicht notwendig unabhängig!).

(1) $\Pr[A] = \mathbb{E}[I_A]$.

(2) $\Pr[A_1 \cap \ldots \cap A_n] = \mathbb{E}[I_{A_1} \cdot \ldots \cdot I_{A_n}]$.

Beweis:

(1) $\mathbb{E}[I_A] = 1 \cdot \Pr[A] + 0 \cdot \Pr[\bar{A}] = \Pr[A]$.

(2) Das Produkt von Indikatorvariablen ist gleich 1 genau dann, wenn alle entsprechenden Ereignisse eintreten. $\square$

Indikatorvariablen III

Beispiel 69

Wir betrachten wieder das Beispiel der betrunkenen Matrosen.

Frage: Welche Varianz hat die Variable $X := X_1 + \ldots + X_n$?

Wir haben $\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[X^2] - 1$.

Mit $X = X_1 + \cdots + X_n$ gilt:

$$\begin{aligned}
\mathbb{E}[X^2] &= \mathbb{E}\left[\sum_{i=1}^{n} X_i^2 + \sum_{i=1}^{n}\sum_{j\neq i} X_i X_j\right] \\
&= \sum_{i=1}^{n} \mathbb{E}\left[X_i^2\right] + \sum_{i=1}^{n}\sum_{j\neq i} \mathbb{E}\left[X_i X_j\right] && \text{(Linearität)} \\
&= n \cdot \mathbb{E}\left[X_1^2\right] + n(n-1) \cdot \mathbb{E}\left[X_1 X_2\right] && \text{(Symmetrie)}
\end{aligned}$$

Indikatorvariablen IV

Sei A_i das Ereignis, dass der i-te Seemann im richtigen Bett liegt. Mit der Notation der Indikatorvariablen sei $X_i = I_{A_i}$. Dann gilt:

$$\mathbb{E}[X_1 X_2] = \mathbb{E}[I_{A_1} I_{A_2}] = \Pr[A_1 \cap A_2] = \frac{1}{n(n-1)}$$

sowie

$$\mathbb{E}[X_1^2] = 0^2 \cdot \Pr[\bar{A}_1] + 1^2 \cdot \Pr[A_1] = \Pr[A_1] = \frac{1}{n}\,.$$

Wir erhalten

$$\begin{aligned}
\mathbb{E}[X^2] &= n \cdot \mathbb{E}\left[X_1^2\right] + n(n-1) \cdot \mathbb{E}\left[X_1 X_2\right] \\
&= n \cdot \frac{1}{n} + n(n-1) \cdot \frac{1}{n(n-1)} = 2\,.
\end{aligned}$$

und

$$\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = 2 - 1 = 1.$$

Indikatorvariablen V

Einfacher Beweis für Satz 3 mit Hilfe von Indikatorvariablen:

Zur Erinnerung:

Satz 3 (Siebformel, Prinzip der Inklusion/Exklusion)

Für Ereignisse $A_1, \ldots, A_n$ $(n \geq 2)$ gilt:

$$\begin{aligned}
\Pr\left[\bigcup_{i=1}^{n} A_i\right] &= \sum_{i=1}^{n} \Pr[A_i] - \sum_{1\leq i_1<i_2\leq n} \Pr[A_{i_1} \cap A_{i_2}] + - \ldots \\
&\quad + (-1)^{l-1} \sum_{1\leq i_1<\ldots<i_l\leq n} \Pr[A_{i_1} \cap \ldots \cap A_{i_l}] + - \ldots \\
&\quad + (-1)^{n-1} \cdot \Pr[A_1 \cap \ldots \cap A_n]\,.
\end{aligned}$$

Indikatorvariablen VI

Beweis: Sei $B := A_1 \cup \ldots \cup A_n$. Wir betrachten die Indikatorvariablen $I_i := I_{A_i}$ der Ereignisse $A_1, \ldots, A_n$ und die Indikatorvariable $I_{\bar{B}}$ des Ereignisses $\bar{B}$.

Es gilt

$$\prod_{i=1}^{n}(1 - I_i) = \begin{cases} 1 & \text{falls } I_1 = \ldots = I_n = 0, \\ & \text{d.h. wenn } B \text{ nicht eintritt} \\ 0 & \text{sonst} \end{cases} = I_{\bar{B}}$$

Wir erhalten:

$$I_{\bar{B}} = 1 - \sum_{1 \le i \le n} I_i + \sum_{1 \le i_1 < i_2 \le n} I_{i_1} I_{i_2} - \ldots + (-1)^n I_1 \cdot \ldots \cdot I_n.$$

Indikatorvariablen VII

Es folgt

$$
\begin{aligned}
\Pr[B] &= 1 - \Pr[\overline{B}] = 1 - \mathbb{E}[I_{\bar{B}}] && \text{(Lem.68(1))} \\
&= 1 - \mathbb{E}\left[1 - \sum_{1 \le i \le n} I_i + \ldots + (-1)^n I_1 \cdots I_n\right] \\
&= \sum_{1 \le i \le n} \mathbb{E}[I_i] - \sum_{1 \le i_1 < i_2 \le n} \mathbb{E}[I_{i_1} I_{i_2}] \\
&\quad + \ldots + (-1)^n \mathbb{E}[I_1 \cdot \ldots \cdot I_n] && \text{(Linear.)} \\
&= \sum_{1 \le i \le n} \Pr[A_i] - \sum_{1 \le i_1 < i_2 \le n} \Pr[A_{i_1} \cap A_{i_2}] \\
&\quad + \ldots + (-1)^n \Pr[A_1 \cap \ldots \cap A_n] && \text{(Lem.68(2))}
\end{aligned}
$$

10. Formelsammlung

Gesetze zum Rechnen mit Ereignissen I

Im Folgenden seien A und B, sowie $A_1, \ldots, A_n$ Ereignisse. Die Notation $A \uplus B$ steht für $A \cup B$ und zugleich $A \cap B = \emptyset$ (disjunkte Vereinigung). $A_1 \uplus \ldots \uplus A_n = \Omega$ bedeutet also, dass die Ereignisse $A_1, \ldots, A_n$ eine Partition der Ergebnismenge Ω bilden.

$\Pr[\emptyset] = 0$
$0 \leq \Pr[A] \leq 1$
$\Pr[\bar{A}] = 1 - \Pr[A]$
$A \subseteq B \implies \Pr[A] \leq \Pr[B]$

Gesetze zum Rechnen mit Ereignissen II

$\forall i \neq j \,:\, A_i \cap A_j = \emptyset \implies$ $\Pr\left[\bigcup_{i=1}^{n} A_i\right] = \sum_{i=1}^{n} \Pr[A_i]$	Additionssatz
$\Pr[A \cup B] = \Pr[A] + \Pr[B] - \Pr[A \cap B]$	Inklusion/Exklusion, Siebformel
$\Pr\left[\bigcup_{i=1}^{n} A_i\right] \leq \sum_{i=1}^{n} \Pr[A_i]$	Boolesche Ungleichung
$\Pr[A \mid B] = \dfrac{\Pr[A \cap B]}{\Pr[B]}$ für $\Pr[B] > 0$	Def. bedingte Ws.

Gesetze zum Rechnen mit Ereignissen III

$B \subseteq A_1 \uplus \ldots \uplus A_n \implies$ $\Pr[B] = \sum_{i=1}^{n} \Pr[B\|A_i] \cdot \Pr[A_i]$	Satz von der totalen Wahrscheinlichkeit
$\Pr[B] > 0,\ B \subseteq A_1 \uplus \ldots \uplus A_n \implies$ $\Pr[A_i\|B] = \dfrac{\Pr[B\|A_i] \cdot \Pr[A_i]}{\sum_{i=1}^{n} \Pr[B\|A_i] \cdot \Pr[A_i]}$	Satz von Bayes
$\Pr[A_1 \cap \ldots \cap A_n] = \Pr[A_1] \cdot \Pr[A_2\|A_1] \cdot \ldots \cdot \Pr[A_n\|A_1 \cap \ldots \cap A_{n-1}]$	Multiplikationssatz
A und B unabhängig $\iff$ $\Pr[A \cap B] = \Pr[A] \cdot \Pr[B]$	Definition Unabhängigkeit

Gesetze zum Rechnen mit Zufallsvariablen I

Sei X eine diskrete Zufallsvariable. Wenn $\mathbb{E}[X]$ und $\mathrm{Var}[X]$ existieren, dann gilt:

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{x \in W_X} x \cdot \Pr[X = x] \\
&= \sum_{\omega \in \Omega} X(\omega) \cdot \Pr[\omega] \\
\Big(&= \sum_{i=1}^{\infty} \Pr[X \geq i], \quad \text{falls } W_X \subseteq \mathbb{N}_0 \Big)
\end{aligned}$$

Erwartungswert

$$\begin{aligned}
\mathrm{Var}[X] &= \mathbb{E}[(X - \mathbb{E}[X])^2] \\
&= \sum_{x \in W_X} \Pr[X = x] \cdot (x - \mathbb{E}[X])^2
\end{aligned}$$

Varianz

Gesetze zum Rechnen mit Zufallsvariablen II

Seien a, b, a_1, …, $a_n \in \mathbb{R}$, $f_1, \ldots, f_n : \mathbb{R} \to \mathbb{R}$.

$X_1, \ldots, X_n$ unabhängig $\iff$	für alle $(a_1, \ldots, a_n)$: $\Pr[X_1 = a_1, \ldots, X_n = a_n] = \Pr[X_1 = a_1] \cdot \ldots \cdot \Pr[X_n = a_n]$
$X_1, \ldots, X_n$ unabhängig $\implies$	$f_1(X_1), \ldots, f_n(X_n)$ unabhängig

Gesetze zum Rechnen mit Zufallsvariablen III

$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X \mid A_i] \cdot \Pr[A_i]$
$\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$
$\mathrm{Var}[a \cdot X + b] = a^2 \cdot \mathrm{Var}[X]$

Gesetze zum Rechnen mit Zufallsvariablen IV

$\mathbb{E}[a_1 X_1 + \ldots + a_n X_n]$ $= a_1 \mathbb{E}[X_1] + \ldots + a_n \mathbb{E}[X_n]$	Linearität des Erwartungswerts
$X_1, \ldots, X_n$ unabhängig $\Longrightarrow$ $\mathbb{E}[X_1 \cdot \ldots \cdot X_n] = \mathbb{E}[X_1] \cdot \ldots \cdot \mathbb{E}[X_n]$	Multiplikativität des Erwartungswerts
$X_1, \ldots, X_n$ unabhängig $\Longrightarrow$ $\mathrm{Var}[X_1 + \ldots + X_n] = \mathrm{Var}[X_1] + \ldots + \mathrm{Var}[X_n]$	Varianz einer Summe

Wichtige Reihen

$\sum_{k=0}^{\infty} 1 \cdot x^k$	$=$	$\frac{1}{1-x}$	$\lvert x\rvert < 1$	Geometrische Reihe
$\sum_{k=1}^{\infty} k \cdot x^{k-1}$	$=$	$\frac{1}{(1-x)^2}$	$\lvert x\rvert < 1$	Ableitung der geometrische Reihe
$\sum_{k=0}^{\infty} \binom{n}{k} \cdot x^k$	$=$	$(1+x)^n$	$\lvert x\rvert < 1$	Binomialreihe
$\sum_{k=0}^{\infty} \frac{1}{k!} \cdot x^k$	$=$	e^x		Exponentialreihe

11. Wichtige Verteilungen

Erwartungswert und Varianz der Binomialverteilung

Sei $X \sim \mathrm{Bin}(n, p)$. Wir haben $\mathbb{E}[X] = np$ berechnet. Für die Varianz haben wir noch keinen Ausdruck berechnet.

Sind $X_1, \ldots, X_n$ unabhängig Bernoulli-verteilte Zufallsvariablen mit der gleichen Erfolgsw'keit p, dann gilt $X_1 + \ldots + X_n \sim \mathrm{Bin}(n, p)$. Mit der Linearität des Erwartungswertes folgt:

$$\mathbb{E}[X] = \mathbb{E}[X_1] + \ldots + \mathbb{E}[X_n] = np$$

Mit der Rechenregel für die varianz einer Summe von unabhängigen Zufalsvariablen folgt:

$$\mathrm{Var}[X] = \mathrm{Var}[X_1] + \ldots + \mathrm{Var}[X_n] = npq$$

Summe von Binomialverteilugen

Satz 70

Wenn $X \sim \mathrm{Bin}(n_x, p)$ *und* $Y \sim \mathrm{Bin}(n_y, p)$ *unabhängig sind, dann gilt* $(X + Y) \sim \mathrm{Bin}(n_x + n_y, p)$.

Beweis:

Seien $X_1, \ldots, X_{n_x}, Y_1, \ldots, Y_{n_y}$ unabhängige Bernoulli-verteilte Zufallsvariablen mit der gleichen Erfolgsw'keit p. Es gilt

$$X_1 + \ldots + X_{n_x} \sim \mathrm{Bin}(n_x, p) \qquad Y_1 + \ldots + Y_{n_y} \sim \mathrm{Bin}(n_y, p)$$

und

$$X_1 + \ldots + X_{n_x} + Y_1 + \ldots + Y_{n_y} \sim \mathrm{Bin}(n_x + n_y, p)\ .$$

$\square$

Gedächtnislosigkeit der geometrischen Verteilung I

Beispiel 71

Bei mündlichen Prüfungen stellt Prof. Evilsparza irgendeine beliebige Frage und, während der Kandidat antwortet, denkt er an seinen nächsten Urlaub auf den Malediven. Anschließend wirft er eine faire Würfel, und der Kandidat besteht nur dann, wenn der Würfel eine 1 zeigt. An einem Tag prüft Prof. Evilsparza Kandidaten nur, bis einer besteht (danach hat er keine Lust mehr).

Heute prüft Prof. Evilsparza. Kandidat Nr. 8 kommt in den Warteraum und erfährt, dass bisher 3 Kandidaten geprüft wurden und alle durchgefallen sind.

Frage: Mit welcher W'keit wird er noch geprüft ?

Gedächtnislosigkeit der geometrischen Verteilung II

Modell: $\Omega = F^*B \qquad \Pr[F^kB] = \left(\frac{5}{6}\right)^k \left(\frac{1}{6}\right)$

Sei X die Zufallsvariable, die angibt, wieviele Kandidaten insgesamt geprüft werden (z.B. gilt $X(F^kB) = k+1$). Gesucht wird:

$$\Pr[X > 7 \mid X > 3]$$

Intuitiv ist die Situation im Experiment äquivalent zu: Kandidat Nr. 5 kommt in den Warteraum und erfährt, dass noch kein Kandidat geprüft wurde.

In diesem Fall ist die gesuchte W'keit $\Pr[X > 4]$, und daher erwarten wir

$$\Pr[X > 7 \mid X > 3] = \Pr[X > 4]$$

und im Allgemeinen:

$$\Pr[X > x + y \mid X > x] = \Pr[X > y]\ .$$ Stimmt es?

Gedächtnislosigkeit der geometrischen Verteilung III

Die Variable X is geometrisch verteilt mit Erfolgsw'keit $p = 1/6$.

$$
\begin{aligned}
\Pr[X > y] &= \sum_{i=y+1}^{\infty} \Pr[X = i] = \sum_{i=y+1}^{\infty} (1-p)^{i-1} p \\
&= (1-p)^y p \cdot \sum_{i=0}^{\infty} (1-p)^i \\
&= (1-p)^y p \cdot \frac{1}{1-(1-p)} = (1-p)^y \\
\Pr[X > y + x \mid X > x] &= \frac{\Pr[X > y + x, X > x]}{\Pr[X > x]} \\
&= \frac{\Pr[X > y + x]}{\Pr[X > x]} = \frac{(1-p)^{y+x}}{(1-p)^x} \\
&= (1-p)^y
\end{aligned}
$$

Gedächtnislosigkeit der geometrischen Verteilung IV

Definition 72

Eine Zufallsvariable X heißt gedächtnislos wenn

$$\Pr[X > x + y \mid X > x] = \Pr[X > y]$$

gilt für alle $x, y \in \mathbb{R}$.

Intuition: wenn eine Variable gedächtnislos ist, dann kann man die Vergangenheit „vergessen", da sie keine nützliche Information für die Zukunft gibt.

Lemma 73

Geometrisch verteilte Zufallsvariablen sind gedächtnislos.

Es gilt sogar die Umkehrung: eine gedächtnislose, Zufallsvariable X mit $W_X = \mathbb{N}_0$ ist geometrisch verteilt!

Beispiel 74 (Das Coupon-Collector-Problem)

Manche Firmen legen den Verpackungen ihrer Produkte oft kleine Bilder bei, um den Käufer zum Sammeln anzuregen.

Ein Kind kauft Packungen so lange ein, bis es die vollständige Sammlung besitzt. Wir nehmen an, dass jede Packung nur ein Bild enthält und bei jedem Kauf jedes Bild mit gleicher Wahrscheinlichkeit auftritt.

Frage: Wie viele Packungen muss das Kind im Mittel erwerben?

Gedächtnislosigkeit der geometrischen Verteilung VI

Modell bei n Beilagen:

$$\begin{aligned}\Omega &= \{w\,k \mid \ w \in \{1,\ldots,n\}^*, k \in \{1,\ldots,n\}, \text{ und} \\ &\qquad\qquad \text{alle Zahlen bis auf } k \text{ kommen in } w \text{ vor}\} \\ \Pr[w\,k] &= \left(\frac{1}{n}\right)^{|w|+1}\end{aligned}$$

Sei X die Zufallsvariable, die angibt, wie viele Bilder gekaut werden (z.B. $X(1122123) = 7$). Gesucht wird $\mathbb{E}[X]$.

Gedächtnislosigkeit der geometrischen Verteilung VII

Bezeichne Phase i die Schritte vom Erwerb des $(i-1)$-ten Bilds (ausschließlich) bis zum Erwerb der i-ten Bilds (einschließlich).

Sei etwa $n = 4$, und seien die Beilagen mit den Zahlen $1, 2, 3, 4$ identifiziert. Ein Experiment ist z.B.:

$$\underbrace{2}_{1}, \underbrace{2, 1}_{2}, \underbrace{2, 2, 3}_{3}, \underbrace{1, 3, 2, 3, 1, 4}_{4} .$$

Beobachtung: Phase i endet genau dann, wenn wir eine der $n - i + 1$ Beilagen erhalten, die wir noch nicht besitzen.

Sei X_i die variable, die angibt, wieviele Bilder in der i-ten Phase gekauft werden. Mit der Gedächtnislosigkeit der geometrischen Verteilung folgt, dass X_i geometrisch verteilt mit

$$p = \frac{n - i + 1}{n} \qquad \mathbb{E}[X_i] = \frac{n}{n - i + 1} .$$

Gedächtnislosigkeit der geometrischen Verteilung VIII

Wir haben

$$\begin{aligned}\mathbb{E}[X] &= \sum_{i=1}^{n} \mathbb{E}[X_i] \\ &= \sum_{i=1}^{n} \frac{n}{n-i+1} \\ &= n \cdot \sum_{i=1}^{n} \frac{1}{i} = n \cdot H_n,\end{aligned}$$

wobei $H_n := \sum_{i=1}^{n} \frac{1}{i}$ die n-te harmonische Zahl bezeichnet.

Da $H_n = \ln n + O(1) + O\left(\frac{1}{n}\right)$, folgt

$$\mathbb{E}[X] = n \ln n + O(n) \ .$$

Negative Binomialverteilung I

Beispiel 75

Eine Münze wird solange geworfen, bis zum n-te Mal Kopf vorkommt. Die W'keit von Kopf beträgt p.

Frage: Mit welcher W'keit wird die Münze k-Mal geworfen?

Modell:

$$\begin{aligned} \Omega &= \{w \in \{K,Z\}^* K \mid K \text{ kommt in } w \ (n-1)\text{-Mal vor}\} \\ \Pr[w] &= p^n q^{|w|-n} \end{aligned}$$

Sei Z die Zufallsvariable, die angibt, wie oft insgesamt geworfen wird. Gesucht wird die Dichte $f_Z(k)$.

Negative Binomialverteilung II

Falls $Z(w) = k$ ist, so kommt in w genau n-Mal Kopf und $(k-n)$-Mal Zahl.

Dafür gibt es genau

$$\binom{k-1}{n-1}$$

Möglichkeiten, von denen jede mit W'keit

$$p^n (1-p)^{k-n}$$

eintritt. Es gilt also

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

Die Zufallsvariable Z nennt man negativ binomialverteilt mit Ordnung n.

Hypergeometrische Verteilung I

Beispiel 76

In der Legislaturperiode 2008-2013 hat der bayerischer Landtag 187 Abgeordnete, 92 davon bilden die CSU-Fraktion. 23 Abgeordnete haben Familienangehörige beschäftigt. 17 davon gehören zur CSU-Fraktion.

Frage: Kann die Anhäufung bei der CSU ein Zufall sein?

Mathematische Formulierung: 187 Bälle in einem Sack. 92 Bälle sind schwarz, der Rest weiß. 23 Bälle werden zufällig gezogen (ohne Zurücklegen). Sei X die Zufallsvariable, die die Anzahl der schwarzen Bällen, die gezogen wurden, angibt. Gesucht ist $\Pr[X \geq 17]$.

$$\Pr[X = k] = \frac{\binom{92}{k}\binom{95}{23-k}}{\binom{187}{23}} \qquad \Pr[X \geq 17] = \sum_{k=17}^{23} \frac{\binom{92}{k}\binom{95}{23-k}}{\binom{187}{23}} \approx 0.0097$$

Hypergeometrische Verteilung II

Beispiel 77

In der Legislaturperiode 2008-2013 hat der bayerischer Landtag 187 Abgeordnete. 15 Abgeordnete bilden die FDP-Fraktion. 23 Abgeordnete haben Familienangehörige beschäftigt. Keiner gehört zur FDP-Fraktion.

Frage: Kann die „weisse Weste“ der FDP ein Zufall sein?

Mathematische Formulierung: 187 Bälle in einem Sack. 15 Bälle sind gelb, der Rest weiß. 23 Bälle werden zufällig gezogen (ohne zurücklegen). Sei Y die Zufallsvariable, die die Anzahl der gelben Bällen, die gezogen wurden, angibt. Gesucht ist $\Pr[Y \leq 0] = \Pr[Y = 0]$.

$$\Pr[Y = k] = \frac{\binom{15}{k}\binom{172}{23-k}}{\binom{187}{23}} \qquad \Pr[Y = 0] = \frac{\binom{172}{23}}{\binom{187}{23}} = 0.128$$

Hypergeometrische Verteilung III

Eine Zufallsvariable X mit $W_X = \mathbb{N}_0$ ist hypegeometrisch verteilt mit Parametern N, M, n wenn ihre Dichte folgende Gestalt hat:

$$f_X(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

Es gilt $\mathbb{E}[X] = n\dfrac{M}{N}$

Beispiel 78

Beim Kundenservice einer Firma rufen im Durschnitt k Kunden pro Tag an.

Wir betrachten ein diskretes Modell, in dem der Tag in $n \geq k$ gleich lange Zeitintervallen unterteilt wird (jeweils $24/n$ Stunden lang). Wir nehmen an, dass Kunden in jedem Intervall mit der selben W'keit anrufen und dass höchstens ein Kunde pro Zeitintervall anruft.

Damit ruft ein Kunde in ein Intervall mit W'keit $\dfrac{k}{n}$ an. Die Anzahl X der Anrufe an einem Tag ist binomialverteilt:

$$\Pr[X \leq a] = \sum_{i=0}^{a} \binom{n}{i} \left(\frac{k}{n}\right)^{i} \left(1 - \frac{k}{n}\right)^{(n-i)}$$

Poisson-Verteilung II

Die folgende Tabelle zeigt $\Pr[X \le 5]$ für $k = 3$ und verschiedene Werte von n:

n	$\Pr[X \le 5]$
5	1
6	0.9844
8	0.9640
24	0.9297
$24 * 60$	0.9163

Frage: Zu welchem Wert konvergiert diese Folge?

Poisson-Verteilung III

Sei X_n die Anzahl der Anrufe bei einer Unterteilung in n Zeitabschnitten. Es gilt

$$X_n \sim \mathrm{Bin}(n, p_n) \qquad \text{mit } p_n = \lambda/n\ .$$

Für ein beliebiges a mit $0 \leq a \leq n$ ist die W'keit $b(a; n, p_n)$, dass X_n den Wert a annimmt, gleich

$$\begin{aligned} b(a; n, p_n) &= \binom{n}{a} \cdot p_n^a \cdot (1 - p_n)^{n-a} \\ &= \frac{n^{\underline{a}}}{a!} \cdot \frac{\lambda^a}{n^a} \cdot \left(1 - \frac{\lambda}{n}\right)^{n-a} \\ &= \frac{\lambda^a}{a!} \cdot \frac{n^{\underline{a}}}{n^a} \cdot \left(1 - \frac{\lambda}{n}\right)^{n} \cdot \left(1 - \frac{\lambda}{n}\right)^{-a}\ . \end{aligned}$$

Poisson-Verteilung IV

Wir betrachten nun $n \to \infty$ und erinnern uns, dass

$$\lim_{n\to\infty} \frac{n^{\underline{k}}}{n^k} = 1 \quad \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda} \quad \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{-k} = 1$$

Damit folgt

$$\begin{aligned}
\lim_{n\to\infty} b(k;n,p_n) &= \lim_{n\to\infty} \left(\frac{\lambda^a}{a!} \cdot \frac{n^{\underline{a}}}{n^a} \cdot \left(1 - \frac{\lambda}{n}\right)^n \cdot \left(1 - \frac{\lambda}{n}\right)^{-a}\right) \\
&= \frac{\lambda^k}{k!} \cdot 1 \cdot e^{-\lambda} \cdot 1 \\
&= \frac{e^{-\lambda}\lambda^k}{k!}.
\end{aligned}$$

Poisson-Verteilung V

Definition 79

Eine Zufallsvariable X mit $W_X = \mathbb{N}_0$ und Dichte

$$f_X(i) = \frac{e^{-\lambda}\lambda^i}{i!} \quad \text{für alle } i \in \mathbb{N}_0, \text{ wobei } \lambda > 0$$

heißt Poisson-verteilt mit dem Parameter λ. Wir schreiben $X \sim \mathrm{Po}(\lambda)$.

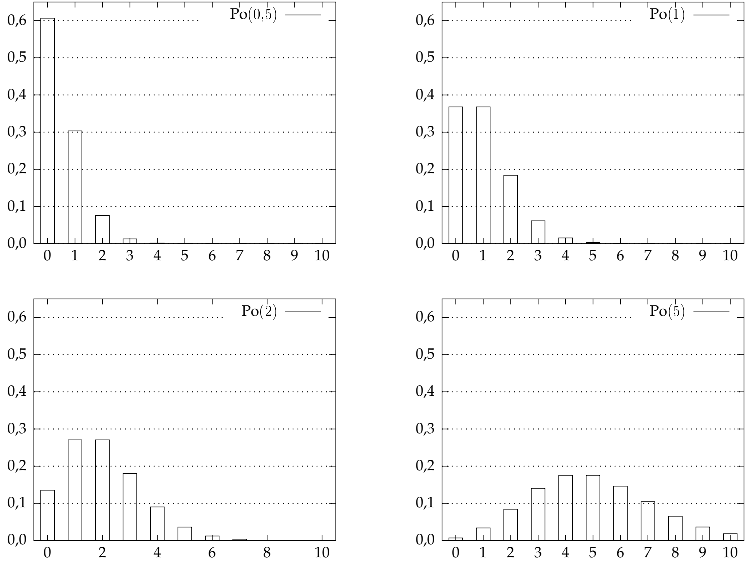
f_X ist eine zulässige Dichte, da

$$\sum_{i=0}^{\infty} f_X(i) = \sum_{i=0}^{\infty} \frac{e^{-\lambda}\lambda^i}{i!} = e^{-\lambda} \cdot e^{\lambda} = 1\,.$$

Es gilt also

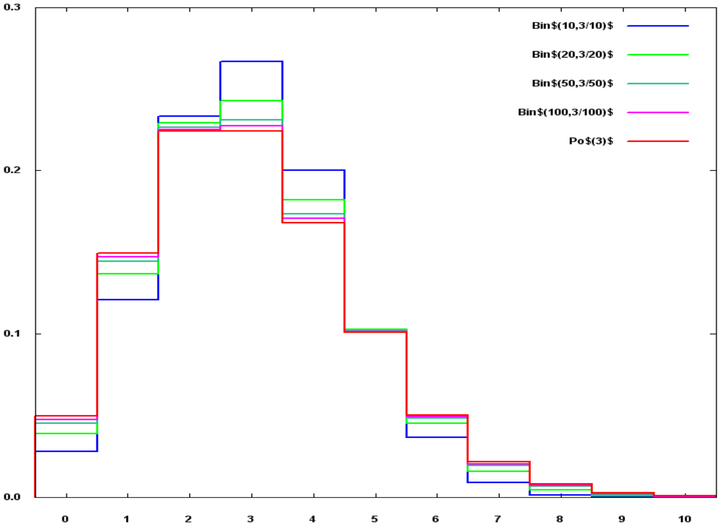
$$\lim_{n\to\infty} b(k; n, \frac{\lambda}{n}) = f_X(k) \qquad \text{für } X \sim \mathrm{Po}(\lambda).$$

Poisson-Verteilung VI



Dichte der Poisson-Verteilung

Poisson-Verteilung VII



Vergleich von Binomial- und Poisson-Verteilung

Poisson-Verteilung VIII

Aus

$$\lim_{n\to\infty} b(k; n, \frac{\lambda}{n}) = f_X(k) \qquad \text{für } X \sim \mathrm{Po}(\lambda).$$

folgt:

Ist p in Vergleich zu n hinreichend klein, so kann man $\mathrm{Po}(np)$ als Approximation der Binomialverteilung $\mathrm{Bin}(n, p)$ verwenden.

Diese Tatsache wird als Gesetz seltener Ereignisse bezeichnet, da die Wahrscheinlichkeit eines einzelnen Treffers relativ klein sein muss, wenn die Approximation gute Ergebnisse liefern soll.

Poisson-Verteilung IX

Die folgenden Voraussetzungen müssen erfüllt sein, damit die Annahme der Poisson-Verteilung gerechtfertigt ist:

- Die Ereignisse treten nie zur gleichen Zeit auf.
- Die Wahrscheinlichkeit, dass ein Ereignis in einem (kleinen) Zeitintervall δt auftritt, ist proportional zur Länge von δt.
- Die Anzahl der Ereignisse in einem festen Zeitintervall hängt nur von dessen Länge ab, nicht aber von der Lage auf der Zeitachse.
- Wenn man zwei disjunkte Zeitintervalle betrachtet, so sind die Anzahlen der Ereignisse in diesen Zeiträumen voneinander unabhängig.

Beispiel 80 (FRM II in Bezug auf Erdbeben in Japan)

FRM II ist analog zu allen deutschen KKWs nach einem Ermessenserdbeben nach der MSK Skala ausgelegt. Die Bemessungsintensität für FRM II lautet I(MSK) = VI–VII mit Eintrittswahrscheinlichkeit $10^{-5}/a$.
(VI: leichte Verputzschäden an Gebäuden. VII: Risse im Verputz, in Wänden und an Schornsteinen.5-0-6.3 auf der Richterskala).

In der seismographischen Region Bayerische Molasse sind aus den vergangenen Jahrhunderten insgesamt 6 tektonische Erdbeben berichtet. Das stärkste mit I(MSK)=VI ereignete sich am 9.10.1935 bei St. Martin in Österreich 135 km östlich von Garching. Dieses und alle anderen Erdbeben lösten nur sehr schwache Bodenbewegungen in Garching aus mit I(MSK) = III–IV. (III: nur von wenigen Personen gespürt. IV: von vielen Personen gespürt; Geschirr und Fenster klirren.)

Poisson-Verteilung XI

Frage: Mit welcher Wahrscheinlichkeit wird Garching in 2014 einmal oder mehr von einem Erdbeben mit I(MSK)$\geq$VI–VII heimgesucht?

Die Anzahl X der Katastrophen in Garching wird durch eine Poisson-Verteilung mit Parameter $\lambda = 10^{-5}$ pro Jahr modelliert. (Vergleich: nach Wikipedia $\lambda \approx 1000$ für die ganze Welt.)

Damit gilt

$$\begin{aligned}
\Pr[X \geq 1] &= 1 - \Pr[X = 0] = 1 - e^{-\lambda} \\
&\approx 1 - 0{,}99999000005 \approx 10^{-5}. \\
\Pr[X \geq 2] &= 1 - \Pr[X = 0] - \Pr[X = 1] = 1 - e^{-\lambda} - \lambda e^{-\lambda} \\
&\approx 1 - 0{,}99999000005 - 0{,}00000999990 = 5 \cdot 10^{-11}.
\end{aligned}$$

Poisson-Verteilung XII: Erwartungswert

Erwartungswert und Varianz

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{i=0}^{\infty} i \cdot \frac{e^{-\lambda}\lambda^i}{i!} \\
&= \lambda e^{-\lambda} \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} \\
&= \lambda e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} \\
&= \lambda e^{-\lambda} e^{\lambda} = \lambda\,.
\end{aligned}$$

Poisson-Verteilung XIII: Varianz

Es gilt

$$\begin{aligned}\mathrm{Var}[X] &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[X^2] - \mathbb{E}[X] + \mathbb{E}[X] - \mathbb{E}[X]^2\\ &= \mathbb{E}[X(X-1)] + \mathbb{E}[X] - \mathbb{E}[X]^2\end{aligned}$$

und

$$\begin{aligned}\mathbb{E}[X(X-1)] &= \sum_{i=0}^{\infty} i(i-1) \cdot \frac{e^{-\lambda}\lambda^i}{i!}\\ &= \lambda^2 e^{-\lambda} \sum_{i=2}^{\infty} \frac{\lambda^{i-2}}{(i-2)!} = \lambda^2 e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!}\\ &= \lambda^2 e^{-\lambda} e^{\lambda} = \lambda^2\end{aligned}$$

Damit erhalten wir

$$\mathrm{Var}[X] = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - \mathbb{E}[X]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Poisson-Verteilung XIV: Summe

Satz 81

Sind X *und* Y *unabhängige Zufallsvariablen mit* $X \sim \mathrm{Po}(\lambda)$ *und* $Y \sim \mathrm{Po}(\mu)$*, dann gilt*

$$Z := X + Y \sim \mathrm{Po}(\lambda + \mu)\,.$$

Poisson-Verteilung XV: Summe

Beweis:

$$
\begin{aligned}
f_Z(z) &= \sum_{x=0}^{\infty} f_X(x) \cdot f_Y(z-x) = \sum_{x=0}^{z} \frac{e^{-\lambda}\lambda^x}{x!} \cdot \frac{e^{-\mu}\mu^{z-x}}{(z-x)!} \\
&= e^{-(\lambda+\mu)} \cdot \frac{(\lambda+\mu)^z}{z!} \cdot \sum_{x=0}^{z} \frac{z!}{x!(z-x)!} \left(\frac{\lambda}{\lambda+\mu}\right)^x \left(\frac{\mu}{\lambda+\mu}\right)^{z-x} \\
&= e^{-(\lambda+\mu)} \cdot (\lambda+\mu)^z \frac{1}{z!} \cdot \sum_{x=0}^{z} \binom{z}{x} p^x (1-p)^{z-x},
\end{aligned}
$$

wobei $p := \dfrac{\lambda}{\lambda+\mu}$.

Da die Summe gleich 1 ist, folgt

$$
f_Z(z) = e^{-(\lambda+\mu)} \cdot (\lambda+\mu)^z \frac{1}{z!} \, .
$$

12. Abschätzen von Wahrscheinlichkeiten

Markov-Ungleichung I

Satz 82 (Markov-Ungleichung)

Sei X eine Zufallsvariable, die nur nicht-negative Werte annimmt. Dann gilt für alle $t \in \mathbb{R}$ mit $t > 0$, dass

$$\Pr[X \geq t] \leq \frac{\mathbb{E}[X]}{t}.$$

Äquivalent dazu:

$$\Pr[X \geq t \cdot \mathbb{E}[X]] \leq 1/t\,.$$

Markov-Ungleichung II

Beweis:

$$\begin{aligned} t \cdot \Pr[X \geq t] &= t \cdot \sum_{x \in W_X,\, x \geq t} \Pr[X = x] \\ &\leq \sum_{x \in W_X,\, x \geq t} x \cdot \Pr[X = x] \\ &\leq \sum_{x \in W_X} x \cdot \Pr[X = x] \\ &= \mathbb{E}[X]\,. \end{aligned}$$

$\square$

Markov-Ungleichung III

Alternativer Beweis:

Es gilt

$$\mathbb{E}[X] = \mathbb{E}[X|X < t] \cdot \Pr[X < t] + \mathbb{E}[X|X \geq t] \cdot \Pr[X \geq t]\,.$$

Wegen $\mathbb{E}[X|X < t] \cdot \Pr[X < t] \geq 0$ und $\mathbb{E}[X|X \geq t] \geq t$ folgt sofort

$$\mathbb{E}[X] \geq t \cdot \Pr[X \geq t]\,.$$

Markov-Ungleichung IV

Die Markov-Ungleichung ist nach Andrey Andreyevich Markov (1856–1922) benannt, der an der Universität von St. Petersburg bei Chebyshev studierte und später dort arbeitete. Neben seiner mathematischen Tätigkeit fiel Markov durch heftige Proteste gegen das Zaren-Regime auf, und nur sein Status als vermeintlich harmloser Akademiker schützte ihn vor Repressalien durch die Behörden. Im Jahr 1913 organisierte er parallel zum dreihundertjährigen Geburtstag der Zarenfamilie Romanov eine Feier zum zweihundertjährigen Geburtstag des Gesetzes der großen Zahlen (s.u.).

Chebyshev-Ungleichung I

Die folgende Abschätzung ist nach Pavnuty Lvovich Chebyshev (1821–1894) benannt, der ebenfalls an der Staatl. Universität in St. Petersburg wirkte.

Satz 83 (Chebyshev-Ungleichung)

Sei X eine Zufallsvariable, und sei $t \in \mathbb{R}$ mit $t > 0$. Dann gilt

$$\Pr[|X - \mathbb{E}[X]| \geq t] \;\leq\; \frac{\mathrm{Var}[X]}{t^2}\,.$$

Äquivalent dazu:

$$\Pr[|X - \mathbb{E}[X]| \geq t\sqrt{\mathrm{Var}[X]}] \;\leq\; 1/t^2\,.$$

Chebyshev-Ungleichung II

Beweis:

Wir stellen fest, dass

$$\Pr[|X - \mathbb{E}[X]| \geq t] \;=\; \Pr[(X - \mathbb{E}[X])^2 \geq t^2]\,.$$

Setze

$$Y := (X - \mathbb{E}[X])^2\,.$$

Dann gilt $\mathbb{E}[Y] = \mathrm{Var}[X]$, und damit mit der Markov-Ungleichung:

$$\Pr[|X - \mathbb{E}[X]| \geq t] \;=\; \Pr[Y \geq t^2] \;\leq\; \frac{\mathbb{E}[Y]}{t^2} \;=\; \frac{\mathrm{Var}[X]}{t^2}\,.$$

$\square$

Chebyshev-Ungleichung III

Beispiel 84

Wir werfen 1000-mal eine ideale Münze und ermitteln die Anzahl X der Würfe, in denen „Kopf“ fällt.

Frage: Wie groß ist die W'keit, dass mehr als 550-mal „Kopf“ fällt?

Mit $X \sim \mathrm{Bin}(1000, \frac{1}{2})$ gilt

$$\mathbb{E}[X] = \frac{1}{2}n = 500 \qquad \text{und} \qquad \mathrm{Var}[X] = \frac{1}{4}n = 250.$$

Chebyshev-Ungleichung $\Pr[|X - \mathbb{E}[X]| \geq t] \;\leq\; \frac{\mathrm{Var}[X]}{t^2}$:

$$\begin{aligned}\Pr[X \geq 550] &\leq \Pr[|X - 500| \geq 50] \\ &\leq \frac{250}{50^2} = 0{,}1\,.\end{aligned}$$

Chebyshev-Ungleichung IV

Setze nun $n = 10000$ und betrachte wieder eine maximal 10%-ige Abweichung vom Erwartungswert:

$$\mathbb{E}[X] = 5000 \qquad \text{und} \qquad \mathrm{Var}[X] = 2500$$

$$\begin{aligned}\Pr[X \geq 5500] &\leq \Pr[|X - 5000| \geq 500] \\ &\leq \frac{2500}{500^2} = 0{,}01\,.\end{aligned}$$

Satz 85 (Gesetz der großen Zahlen)

Sei X eine Zufallsvariable und seien $X_1, \ldots, X_n$ unabhängige Zufallsvariablen mit derselben Verteilung wie X.

Sei

$$Z := \frac{X_1 + \ldots + X_n}{n},$$

Seien $\varepsilon, \delta > 0$ beliebig aber fest. Dann gilt für alle $n \geq \frac{\mathrm{Var}[X]}{\varepsilon\delta^2}$:

$$\Pr[\,|Z - \mathbb{E}[X]| \geq \delta\,] \;\leq\; \varepsilon.$$

Gesetz der großen Zahlen II

Beweis:

Für Z gilt

$$\mathbb{E}[Z] = \frac{1}{n} \cdot (\mathbb{E}[X_1] + \ldots + \mathbb{E}[X_n]) = \frac{1}{n} \cdot n \cdot \mathbb{E}[X] = \mathbb{E}[X],$$

sowie (Unabhängigkeit)

$$\mathrm{Var}[Z] = \frac{1}{n^2} \cdot (\mathrm{Var}[X_1] + \ldots + \mathrm{Var}[X_n]) = \frac{1}{n^2} \cdot n \cdot \mathrm{Var}[X] = \frac{\mathrm{Var}[X]}{n}.$$

Mit der Chebyshev-Ungleichung erhalten wir

$$\Pr[|Z - \mathbb{E}[X]| \geq \delta] = \Pr[|Z - \mathbb{E}[Z]| \geq \delta] \leq \frac{\mathrm{Var}[Z]}{\delta^2} = \frac{\mathrm{Var}[X]}{n\delta^2} \leq \varepsilon,$$

nach Wahl von n. $\square$

Gesetz der großen Zahlen III

Sei X eine Indikatorvariable für ein Ereignis A mit $\Pr[A] = p$.

Somit ist X Bernoulli-verteilt mit $\mathbb{E}[X] = p$.

$Z = \frac{1}{n}(X_1 + \ldots + X_n)$ gibt die relative Häufigkeit an, mit der A bei n Wiederholungen des Versuchs eintritt, denn

$$Z = \frac{\text{Anzahl der Versuche, bei denen } A \text{ eingetreten ist}}{\text{Anzahl aller Versuche}}.$$

Mit Hilfe des obigen Gesetzes der großen Zahlen folgt

$$\Pr[\,|Z - p| \geq \delta\,] \leq \varepsilon,$$

für genügend großes n.

Also nähert sich die relative Häufigkeit von A bei hinreichend vielen Wiederholungen des Experiments mit beliebiger Sicherheit und beliebige Genauigkeit an die „wahre“ Wahrscheinlichkeit p an.

Gesetz der großen Zahlen IV

Die obige Variante eines Gesetzes der großen Zahlen geht auf Jakob Bernoulli zurück, der den Satz in seinem Werk ars conjectandi zeigte.

Es soll betont werden, dass das Gesetz der großen Zahlen die

$$\text{relative Abweichung } \left|\frac{1}{n}\sum_i X_i - p\right|$$

und nicht die

$$\text{absolute Abweichung } \left|\sum_i X_i - np\right|$$

abschätzt!

Chernoff-Schranken I

Die hier betrachtete Art von Schranken ist nach *Herman Chernoff* (*1923) benannt. Sie finden in der komplexitätstheoretischen Analyse von Algorithmen eine sehr häufige Verwendung.

Satz 86

Seien $X_1, \ldots, X_n$ *unabhängige Bernoulli-verteilte Zufallsvariablen mit* $\Pr[X_i = 1] = p_i$ *und* $\Pr[X_i = 0] = 1 - p_i$.

Seien $X := \sum_{i=1}^{n} X_i$ *und* $\mu := \mathbb{E}[X] = \sum_{i=1}^{n} p_i$.

Für jedes $\delta > 0$ *gilt*

$$\Pr[X \geq (1+\delta)\mu] \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}} \right)^{\mu}.$$

Chernoff-Schranken II

Beweis:

Für $t > 0$ gilt $X \geq (1+\delta)\mu$ iff $e^{tX} \geq e^{t(1+\delta)\mu}$ und so

$$\Pr[X \geq (1+\delta)\mu] = \Pr[e^{tX} \geq e^{t(1+\delta)\mu}]\,.$$

Mit der Markov-Ungleichung folgt

$$\Pr[X \geq (1+\delta)\mu] = \Pr[e^{tX} \geq e^{t(1+\delta)\mu}] \leq \frac{\mathbb{E}[e^{tX}]}{e^{t(1+\delta)\mu}}.$$

Wegen der Unabhängigkeit der Zufallsvariablen $X_1, \ldots, X_n$ gilt

$$\mathbb{E}[e^{tX}] = \mathbb{E}\left[\exp\left(\sum_{i=1}^{n} tX_i\right)\right] = \mathbb{E}\left[\prod_{i=1}^{n} e^{tX_i}\right] = \prod_{i=1}^{n} \mathbb{E}[e^{tX_i}].$$

Chernoff-Schranken III

Weiter ist für $i \in \{1, \ldots, n\}$:

$$\mathbb{E}[e^{tX_i}] = e^{t\cdot 1}p_i + e^{t\cdot 0}(1-p_i) = e^t p_i + 1 - p_i = 1 + p_i(e^t - 1)\,,$$

und damit

$$\begin{aligned}
\Pr[X \geq (1+\delta)\mu] &\leq \frac{\prod_{i=1}^n (1 + p_i(e^t - 1))}{e^{t(1+\delta)\mu}} \\
&\leq \frac{\prod_{i=1}^n \exp(p_i(e^t - 1))}{e^{t(1+\delta)\mu}} \\
&= \frac{\exp(\sum_{i=1}^n p_i(e^t - 1))}{e^{t(1+\delta)\mu}} = \frac{e^{(e^t-1)\mu}}{e^{t(1+\delta)\mu}} =: f(t)\,.
\end{aligned}$$

Wir wählen nun t so, dass $f(t)$ minimiert wird, nämlich

$$t := \ln(1+\delta)\,.$$

und erhalten

$$\Pr[X \geq (1+\delta)\mu] \leq \frac{e^{(e^t-1)\mu}}{e^{t(1+\delta)\mu}} = \frac{e^{\delta\mu}}{(1+\delta)^{(1+\delta)\mu}}\,.$$

Chernoff-Schranken III

Weiter ist für $i \in \{1, \ldots, n\}$:

$$\mathbb{E}[e^{tX_i}] = e^{t \cdot 1} p_i + e^{t \cdot 0}(1 - p_i) = e^t p_i + 1 - p_i = 1 + p_i(e^t - 1)\,,$$

und damit

$$\begin{aligned}
\Pr[X \geq (1+\delta)\mu] &\leq \frac{\prod_{i=1}^n (1 + p_i(e^t - 1))}{e^{t(1+\delta)\mu}} \\
&\leq \frac{\prod_{i=1}^n \exp(p_i(e^t - 1))}{e^{t(1+\delta)\mu}} \\
&= \frac{\exp(\sum_{i=1}^n p_i(e^t - 1))}{e^{t(1+\delta)\mu}} = \frac{e^{(e^t-1)\mu}}{e^{t(1+\delta)\mu}} =: f(t)\,.
\end{aligned}$$

Wir wählen nun t so, dass $f(t)$ minimiert wird, nämlich

$$t := \ln(1+\delta)\,.$$

und erhalten

$$\Pr[X \geq (1+\delta)\mu] \leq \frac{e^{(e^t-1)\mu}}{e^{t(1+\delta)\mu}} = \frac{e^{\delta\mu}}{(1+\delta)^{(1+\delta)\mu}}\,.$$

Chernoff-Schranken IV

Beispiel 87

Wir werfen eine faire Münze n-mal. Wir suchen eine obere Schranke für die W'keit, mit der „Kopf“

$$\frac{n}{2}\,(1 + 10\%) \quad \text{Mal}$$

oder öfter fällt.

n	Chebyshev	Chernoff
1000	$0{,}1$	$0{,}0889$
10000	$0{,}01$	$0{,}308 \cdot 10^{-10}$
n	$\frac{\frac{1}{4}n}{(0{,}1 \cdot \frac{1}{2}n)^2} = \frac{100}{n}$	$\left(\frac{e^{0{,}1}}{(1+0{,}1)^{1+0{,}1}}\right)^{n/2} \approx 0.9975^n$

Satz 88

Seien $X_1, \ldots, X_n$ *unabhängige Bernoulli-verteilte Zufallsvariablen mit* $\Pr[X_i = 1] = p_i$ *und* $\Pr[X_i = 0] = 1 - p_i$*.*

Seien $X := \sum_{i=1}^{n} X_i$ *und* $\mu := \mathbb{E}[X] = \sum_{i=1}^{n} p_i$*.*

Für jedes $0 < \delta < 1$ *gilt*

$$\Pr[X \leq (1-\delta)\mu] \leq \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu}.$$

Beweis:

Analog zum Beweis von Satz 86. $\square$

Chernoff-Schranken VI

Abschätzungen, wie sie in Satz 86 und Satz 88 angegeben sind, nennt man auch

tail bounds,

da sie Schranken für die tails, also die vom Erwartungswert weit entfernten Bereiche angeben.

Man spricht hierbei vom

upper tail (vgl. Satz 86) und vom lower tail (vgl. Satz 88).

Die Chernoff-Schranken hängen exponentiell von μ ab!

Lemma 89

Für $0 \le \delta < 1$ *gilt*

$$(1-\delta)^{1-\delta} \ge e^{-\delta+\delta^2/2} \quad \textit{und} \quad (1+\delta)^{1+\delta} \ge \delta + \delta^2/3.$$

Beweis:

Nur für die erste Ungleichung. Sei

$$\begin{aligned} f(\delta) &= \ln\left((1-\delta)^{1-\delta}\right) = (1-\delta)\ln(1-\delta) \\ g(\delta) &= \ln\left(e^{-\delta+\delta^2/2}\right) = -\delta + \delta^2/2 \end{aligned}$$

Es reicht zu zeigen $f(\delta) \ge g(\delta)$ für $0 \le \delta < 1$.

Chernoff-Schranken V

Mit $f(0) = 0 = g(0)$ reicht es sogar $f'(\delta) \geq f'(\delta)$, d.h.

$$((1-\delta)\ln(1-\delta))' = -1 - \ln(1-\delta) \geq -1 + \delta = \left(-\delta + \delta^2/2\right)'$$

Wir haben

$$\begin{array}{rrcl} & -1 - \ln(1-\delta) & \geq & -1 + \delta \\ \iff & \ln(1-\delta) & \leq & -\delta \\ \iff & (1-\delta) & \leq & e^{-\delta} \end{array}$$

Die letzte Ungleichung folgt aus

$$e^{-x} = 1 - x + \frac{1}{2}x^2 - \cdots$$

Chernoff-Schranken VIII

Korollar 90

Seien $X_1, \ldots, X_n$ *unabhängige Bernoulli-verteilte Zufallsvariablen mit* $\Pr[X_i = 1] = p_i$ *und* $\Pr[X_i = 0] = 1 - p_i$*. Dann gelten folgende Ungleichungen für* $X := \sum_{i=1}^n X_i$ *und* $\mu := \mathbb{E}[X] = \sum_{i=1}^n p_i$*:*

1. $\Pr[X \geq (1+\delta)\mu] \leq e^{-\mu\delta^2/3}$ *für alle* $0 < \delta \leq 1,81$*,*
2. $\Pr[X \leq (1-\delta)\mu] \leq e^{-\mu\delta^2/2}$ *für alle* $0 < \delta \leq 1$*,*
3. $\Pr[|X - \mu| \geq \delta\mu] \leq 2e^{-\mu\delta^2/3}$ *für alle* $0 < \delta \leq 1$*,*
4. $\Pr[X \geq (1+\delta)\mu] \leq \left(\frac{e}{1+\delta}\right)^{(1+\delta)\mu}$ *und*
5. $\Pr[X \geq t] \leq 2^{-t}$ *für* $t \geq 2e\mu$*.*

Chernoff-Schranken IX

Beweis:

1 und 2 folgen direkt aus Satz 86 bzw. 88 und Lemma 89.

Aus 1 und 2 zusammen folgt 3.

Die Abschätzung 4 erhalten wir direkt aus Satz 86, da für den Zähler gilt

$$e \leq e^{(1+\delta)} \,.$$

5 folgt aus 4, indem man $t = (1+\delta)\mu$ setzt, $t \geq 2e\mu$:

$$\left(\frac{e}{1+\delta}\right)^{(1+\delta)\mu} \leq \left(\frac{e}{t/\mu}\right)^{t} \leq \left(\frac{1}{2}\right)^{t} \,.$$

$\square$

Beispiel 91

Wir werfen $c \cdot n$ Bälle unabhängig und gleichverteilt in n Körbe.

Frage: Wie groß sollen die Körbe sein, so dass mit großer W'keit kein Korb überläuft ?

Sei $Y_i :=$ Anzahl der Bälle im i-ten Korb für $i = 1, \ldots, n$ und sei

$$Y := \max_{1 \leq i \leq n} Y_i$$

Wir suchen nach einer möglichst langsam wachsenden Funktion $f(n)$ mit

$$\lim_{n \to \infty} \Pr[Y \geq f(n)] = 0$$

Chernoff-Schranken XI

Wir verwenden Aussage 5 von Korollar 90.

Es gilt $p_1 = \ldots = p_{cn} = \dfrac{1}{n}$ und $\mu = \sum_{i=1}^{cn} p_i = c$.

Mit $t = 2 \log n$ folgt

$$\Pr[Y_i \geq 2 \ln n] \leq 1/n^2 \qquad \text{für } \ln n \geq ec.$$

Daraus ergibt sich

$$\begin{aligned}
\Pr[Y \geq 2 \ln n] &= \Pr[\, Y_1 \geq 2 \ln n \;\cup \ldots \cup\; Y_{cn} \geq 2 \ln n \,] \\
&\leq \sum_{i=1}^{cn} \Pr[\, Y_1 \geq 2 \ln n \,] \\
&\leq cn \cdot \frac{1}{n^2} = \frac{c}{n}
\end{aligned}$$

Chernoff-Schranken XII

Literatur:

Torben Hagerup, Christine Rüb:
A guided tour of Chernoff bounds
Inf. Process. Lett. **33**, pp. 305–308 (1990)

13. Erzeugende Funktionen

Erzeugende Funktionen I

Definition 92

Für eine Zufallsvariable X mit $W_X \subseteq \mathbb{N}_0$ ist die (wahrscheinlichkeits-)erzeugende Funktion definiert durch

$$G_X(s) := \sum_{k=0}^{\infty} \Pr[X = k] \cdot s^k = \mathbb{E}[s^X]\,.$$

Bei w'keitserzeugenden Funktionen haben wir kein Problem mit der Konvergenz, da für $|s| < 1$ gilt

$$\begin{aligned}
|G_X(s)| &= \left| \sum_{k=0}^{\infty} \Pr[X = k] \cdot s^k \right| \\
&\le \sum_{k=0}^{\infty} \Pr[X = k] \cdot |s^k| \le \sum_{k=0}^{\infty} \Pr[X = k] = 1\,.
\end{aligned}$$

Erzeugende Funktionen II

Gegeben $G_X(s)$ können wir die W'keiten $\Pr[X = i]$ "ablesen".

Mit

$$G'_X(s) = \sum_{k=1}^{\infty} k \cdot \Pr[X = k] \cdot s^{k-1}$$

gilt

$$\Pr[X = 1] = G'_X(0)$$

Analog erhalten wir

$$G_X^{(i)}(0) = \Pr[X = i] \cdot i!$$

also

$$\Pr[X = i] = \frac{G_X^{(i)}(0)}{i!}$$

Erzeugende Funktionen III: Verteilungen

Bernoulli-Verteilung Sei X eine Bernoulli-verteilte Zufallsvariable mit $\Pr[X=0]=1-p$ und $\Pr[X=1]=p$. Dann gilt

$$G_X(s)=\mathbb{E}[s^X]=(1-p)\cdot s^0+p\cdot s^1=1-p+ps\,.$$

Gleichverteilung auf $\{0,\ldots,n\}$

Sei X auf $\{0,\ldots,n\}$ gleichverteilt, d.h. für $0\le k\le n$ ist $\Pr[X=k]=1/(n+1)$. Dann gilt

$$G_X(s)=\mathbb{E}[s^X]=\sum_{k=0}^{n}\frac{1}{n+1}\cdot s^k=\frac{s^{n+1}-1}{(n+1)(s-1)}\,.$$

Erzeugende Funktionen IV: Verteilungen

Binomialverteilung Für $X \sim \mathrm{Bin}(n, p)$ gilt nach der binomischen Formel

$$G_X(s) = \mathbb{E}[s^X] = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \cdot s^k = (1 - p + ps)^n \,.$$

Geometrische Verteilung

Sei X eine geometrisch verteilte Zufallsvariable mit Erfolgswahrscheinlichkeit p. Dann gilt

$$\begin{aligned} G_X(s) = \mathbb{E}[s^X] &= \sum_{k=1}^{\infty} p(1-p)^{k-1} \cdot s^k \\ &= ps \cdot \sum_{k=1}^{\infty} ((1-p)s)^{k-1} = \frac{ps}{1-(1-p)s} \,. \end{aligned}$$

Erzeugende Funktionen V: Verteilungen

Poisson-Verteilung Für $X \sim \mathrm{Po}(\lambda)$ gilt

$$G_X(s) = \mathbb{E}[s^X] = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} \cdot s^k = e^{-\lambda + \lambda s} = e^{\lambda(s-1)}.$$

Sei X binomialverteilt mit $X \sim \mathrm{Bin}(n, \lambda/n)$. Für $n \to \infty$ folgt

$$G_X(s) = \left(1 - \frac{\lambda}{n} + \frac{\lambda s}{n}\right)^n = \left(1 + \frac{\lambda(s-1)}{n}\right)^n \to e^{\lambda(s-1)}.$$

Man kann beweisen, dass aus der Konvergenz der erzeugenden Funktion die Konvergenz der Verteilung folgt.

Erzeugende Funktionen VI: Erwartungswert und Varianz

Gegeben $G_X(s)$ können wir auch $\mathbb{E}[X]$, $\mathrm{Var}[X]$, und andere Momente "ablesen".

Da

$$G_X(s) := \sum_{k=0}^{\infty} \Pr[X = k] \cdot s^k = \mathbb{E}[s^X]\,,$$

gilt

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} k \cdot \Pr[X = k] = G_X'(1)\,.$$

Erzeugende Funktionen VII: Erwartungswert und Varianz

Beispiel 93

Sei X binomialverteilt mit $X \sim \mathrm{Bin}(n, p)$, also

$$G_X(s) = (1 - p + ps)^n \,.$$

Dann gilt

$$G'_X(s) = n \cdot (1 - p + ps)^{n-1} \cdot p$$

und somit

$$\mathbb{E}[X] = G'_X(1) = np \,.$$

Erzeugende Funktionen VIII: Erwartungswert und Varianz

Ebenso ergibt sich

$$\mathbb{E}[X(X-1)\dots(X-k+1)] = G_X^{(k)}(1)\,,$$

und damit

$$\begin{aligned}\mathrm{Var}[X] &= \mathbb{E}[X(X-1)] + \mathbb{E}[X] - \mathbb{E}[X]^2\\ &= G''_X(1) + G'_X(1) - (G'_X(1))^2\,.\end{aligned}$$

Erzeugende Funktionen IX: Summe

Mit dem folgenden Satz gewinnen wir die erzeugenden Funktionen für Summen von unabhängigen Variablen.

Satz 94

Seien $X_1, \ldots, X_n$ *unabhängige Zufallsvariablem mit* $W_{X_1}, \ldots, W_{X_n} \subseteq \mathbb{N}$ *und sei* $Z := X_1 + \ldots + X_n$. *Es gilt*

$$G_Z(s) = \prod_{i=1}^{n} G_{X_i}(s)$$

Beweis: Nur für $n = 2$

Erzeugende Funktionen IX: Summe

$$
\begin{aligned}
G_Z(s) &= \sum_{k=0}^{\infty} \Pr[Z = k] \cdot s^k \\
&= \sum_{k=0}^{\infty} \left(\sum_{\{(i,j)\in\mathbb{N}^2 \mid i+j=k\}} \Pr[X_1 = i, X_2 = j] \right) \cdot s^k \\
&= \sum_{k=0}^{\infty} \left(\sum_{\{(i,j)\in\mathbb{N}^2 \mid i+j=k\}} \Pr[X_1 = i] \cdot \Pr[X_2 = j] \right) \cdot s^{i+j} \\
&= \sum_{i,j=0}^{\infty} \Pr[X_1 = i] \cdot \Pr[X_2 = j] \cdot s^{i+j} \\
&= \left(\sum_{i=0}^{\infty} \Pr[X_1 = i] \cdot s^i \right) \cdot \left(\sum_{j=0}^{\infty} \Pr[X_2 = j] \cdot s^j \right) \\
&= G_{X_1}(s) \cdot G_{X_2}(s)
\end{aligned}
$$

Teil IV

Kontinuierliche Wahrscheinlichkeitsräume

14. Kontinuierliche Wahrscheinlichkeitsräume

Informelle Einführung I

Wir suchen nach einem mathematischen Modell für dieses Zufallsexperiment:

> *Wir wählen eine reelle Zahl $x \in [0,1]$ zufällig.*
> *Alle Zahlen haben die gleiche W'keit, gewählt zu werden.*

Wir erwarten z.B.:

$\Pr[\ [0,1/2]\ ] = 1/2 \quad \Pr[\ \{1/2\}\ ] = 0 \quad \Pr[\ [0,1/2] \cup [2/3,5/6]\ ] = 2/3$

Da $\Pr[\{x\}] = 0$ für jedes $x \in [0,1]$ gelten muß, kann die W'keit von $A \subseteq [0,1]$ nicht als $\Pr[A] := \sum_{\omega \in A} \Pr[\omega]$ definiert werden.

Lösung: Physikalische Analogie

Informelle Einführung II: Physikalische Analogie

W'keit	$\rightarrow$	1 Kg Masse
Diskreter W'keitsraum	$\rightarrow$	Masse konzentriert auf Punkte (Punktmassen)
Kontinuierlicher W'keitsraum	$\rightarrow$	Masse verteilt im ausgedehnten Körper

Ausgedehnter (eindimensionaler) Körper:

- Masse an einem beliebigen Punkt: 0 gr.
- Masse eines kleinen Bereiches um einen Punkt $x \approx$ Dichte am Punkt $x \times$ Volumen des Bereiches.
- Dichte modelliert durch Dichtefunktion $f(x)$.
- Dichtefunktion erfüllt $\int_0^1 f(x)dx = 1$
- Masse im Subintervall $[a, b]$: $\int_a^b f(x)dx$.

Informelle Einführung III: Physikalische Analogie

In einem kontinuierlichen W'keitsraum:

- W'keit eines beliebigen Punktes: 0.
- W'keit eines kleinen Bereiches um den Punkt $x \approx$ W'keitsdichte am Punkt $x \times$ Volumen des Bereiches.
- W'keitsdichte modelliert durch W'keitsdichtefunktion $f(x)$.
- W'keitsdichtefunktion erfüllt $\displaystyle\int_0^1 f(x)dx = 1$
- W'keit eines Subintervalls $[a, b]$: $\displaystyle\int_a^b f(x)dx$.

Informelle Einführung IV: Beispiele

Beispiel 1: $\Omega = [0,3]$ (es muß nicht immer $[0,1]$ sein ...)

$f(x) = 1/3$ für alle $x \in [0,3]$.

- $\displaystyle\int_0^3 1/3\, dx = \left[\frac{x}{3}\right]_0^3 = 1$
- $\displaystyle\int_1^2 1/3\, dx = \left[\frac{x}{3}\right]_1^2 = 1/3$

Beispiel 2: $\Omega = [0,\infty)$

$f(x) = e^{-x}$ für alle $x \in [0,\infty)$.

- $\displaystyle\int_0^\infty e^{-x}\, dx = \left[-e^{-x}\right]_0^\infty = -0 - (-1) = 1$
- $\displaystyle\int_0^1 e^{-x}\, dx = \left[-e^{-x}\right]_0^1 = -e^{-1} - (-1) = 1 - 1/e$

Informelle Einführung V: Beispiele

Beispiel 3: $\Omega = [0,\infty) \times [0,\infty)$

$f(x) = e^{-(x+y)}$ für alle $x, y \in [0,\infty)$.

$$\begin{aligned}
\int_0^\infty \int_0^\infty e^{-(x+y)} dx dy &= \int_0^\infty e^{-x} \left(\int_0^\infty e^{-y} dy \right) dx \\
&= \int_0^\infty e^{-x} \left[-e^{-y} \right]_0^\infty dx \\
&= \int_0^\infty e^{-x} dx \\
&= 1
\end{aligned}$$

Informelle Einführung VI: Probleme

Wir wählen eine reelle Zahl $x \in [0,1]$ zufällig. Alle Zahlen haben die gleiche W'keit, gewählt zu werden.
Frage: Was ist die W'keit, daß x irrational ist?

Sei A die Menge der irrationalen Zahlen in $[0,1]$.
Intuitiv erwarten wir $\Pr[A] = 1$.

Modell: $\Omega = [0,1]$, W'keitsdichte $f(x) = 1$ für alle $x \in [0,1]$.

Nach dem bisherigen Ansatz $\Pr[A] := \int_0^1 f(x) \cdot I_A \cdot dx$ mit

$$I_A(x) = \begin{cases} 1 & \text{falls } x \text{ irrational} \\ 0 & \text{sonst} \end{cases}$$

Dieses Integral ist jedoch nicht definiert im Sinne der Schulmathematik! (Riemman'sches Integral)

W'keitsräume I

Können wir eine Funktion $\Pr\colon 2^{[0,1]} \to [0,1]$ definieren, die jeder Menge $A \subseteq [0,1]$ die W'keit $\Pr[A]$ zuordnet?

Einige Anforderungen an $\Pr$:

(a) $\Pr[\,[0,1]\,] = 1$.

(b) Wenn $A_1, A_2, \ldots \subseteq [0,1]$ paarweise disjunkt sind, dann gilt

$$\Pr\left[\bigcup_{i=1}^{\infty} A_i\right] = \sum_{i=1}^{\infty} \Pr[A_i].$$

(c) Für $A \subseteq [0,1]$, sei $A + x = \{a + x \mid a \in A\}$.
Wenn $A + x \subseteq [0,1]$, dann $\Pr[A] = \Pr[A + x]$.
(Translationsinvarianz, muß gelten wenn alle Zahlen die gleiche W'keit haben!)

Satz 95 (Vitali)

Keine Abbildung $\Pr : 2^{[0,1]} \to [0,1]$ *erfüllt (a)-(c).*

Beweisskizze: (für $[-1,2]$ statt $[0,1]$ da etwas einfacher.)

- Definiere: $x \sim y$ gdw. $x - y$ eine Rationalzahl ist.
- Zeige: $\sim$ ist Äquivalenzrelation und partitioniert $[0,1]$ in Äquivalenzklassen.
- Sei $V \subseteq [0,1]$ Menge, die genau einen Repräsentanten jeder Äquivalenzklasse enthält (Auswahlaxiom!).
- Sei $q_1, q_2 \ldots$ Aufzählung der Rationalzahlen in $[0,1]$. Betrachte Mengen $V_k = \{v + q_k \mid v \in V\}$.
- Zeige: $[0,1] \subseteq \bigcup_{k=1}^{\infty} V_k \subseteq [-1,2]$ und V_k paarweise disjunkt.
- Zeige: mit (a)-(c) gilt $\Pr[V_k] = \Pr[V]$ und $1 \leq \sum_{i=1}^{\infty} V_k \leq 3$, Widerspruch.

W'keitsräume III

Problem: Die Anforderungen (a)-(c) sind ''nicht verhandelbar"

Man muß akzeptieren, daß $\Pr$ nicht für jede Menge definiert werden kann. D.h., nicht jede Menge kann ein Ereignis sein!

Neues Ziel: Eine "gute" Familie $\mathcal{A} \subseteq 2^{[0,1]}$ von Ereignissen finden, für die wir gleichzeitig eine "gute" Abbildung $\Pr\colon \mathcal{A} \to [0,1]$ definieren können.

Was bedeutet "gut"?

Antwort (Kolmogorov):

$\mathcal{A}$ soll eine σ-Algebra bilden.

$\Pr$ soll die Kolmogorov-Axiome erfüllen.

Dann sagen wir: $\mathcal{A}$ und $\Pr$ bilden einen W'keitsraum.

W'keitsräume IV: σ-Algebren

Definition 96

Sei Ω eine beliebige Menge. Eine Menge $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ heißt σ-Algebra über Ω, wenn folgende Eigenschaften erfüllt sind:

(E1) $\Omega \in \mathcal{A}$.

(E2) Wenn $A \in \mathcal{A}$, dann folgt $\bar{A} \in \mathcal{A}$.

(E3) Für $n \in \mathbb{N}$ sei $A_n \in \mathcal{A}$. Dann gilt auch $\displaystyle\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.

Die Elemente von $\mathcal{A}$ heißen Ereignisse.

- Die Axiome erlauben die Zusammensetzung von Ereignissen aus boole'schen Kombinationen von einfacheren Ereignissen.
- Unendliche Vereinigungen oder Schnitte sind notwendig für die Analyse von mehrstufigen Experimenten mit beliebig vielen oder unendlich vielen Stufen.
- Beliebige (überabzählbare) Vereinigungen sind verboten!

W'keitsräume V: Kolmogorov-Axiome und Definition

Definition 97 (Kolmogorov-Axiome, Wahrscheinlichkeitsraum)

Sei $\mathcal{A}$ eine σ-Algebra über Ω. Eine Abbildung $\Pr\colon \mathcal{A} \to [0,1]$ erfüllt die Kolmogorov-Axiome wenn

(W1) $\Pr[\Omega] = 1$.

(W2) Wenn $A_1, A_2, \ldots$ paarweise disjunkte Ereignisse sind, dann

$$\Pr\left[\bigcup_{i=1}^{\infty} A_i\right] = \sum_{i=1}^{\infty} \Pr[A_i].$$

Ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \Pr)$ besteht aus einer beliebigen Menge Ω, einer σ-Algebra $\mathcal{A}$ über Ω und einer Abbildung $\Pr\colon \mathcal{A} \to [0,1]$, die W1 und W2 erfüllt.

W'keitsräume VI: Einfache Eigenschaften

Aus der Definition folgt:

- Es gelten alle einfache Eigenschaften von W'keiten: Boole'sche Ungleichung, Siebregel, etc.
- Diskrete W'keitsräume sind der Spezialfall in dem Ω eine abzählbare Menge ist und $\mathcal{A} = \mathcal{P}(\Omega)$.

W'keitsräume VII: Ausblick auf die nächsten Folien

Ein W'keitsraum, in dem z.B. die W'keit, dass die Zahl x irrational ist, definiert ist, findet man mit Hilfe des folgenden Satzes:

Satz 98

Sei $\Omega \subseteq \mathbb{R}^n$ *ein Intervall.*
Sei $\mathcal{B}(\Omega)$ *die Menge der Borel'schen Mengen über* Ω*.*
Sei $f\colon \Omega \to \mathbb{R}_0^+$ *eine W'keitsdichte über* Ω*.*
Sei $\Pr\colon \mathcal{B}(\Omega) \to \mathbb{R}_0^+$ *mit*

$$\Pr[A] := \int_A f \, dx$$

wobei $\int_A f \, dx$ *das Lebesgue-Integral von* f *über* A *bezeichnet.*
Dann bildet $(\Omega, \mathcal{B}(\Omega), \Pr)$ *einen W'keitsraum.*

In den kommenden Folien führen wir diese Begriffe ein.

Intervalle

Definition 99

Seien $a, b \in Rr$. Das geschlossene Intervall $[a, b] \subseteq \mathbb{R}$ ist die Menge der reellen Zahlen x mit $a \leq x \leq b$.
Si $n \geq 1$. Ein (n-dimensionales) geschlossenes Intervall ist das kartesische Produkt von n geschlossenen Intervallen.

Offene und halboffene Intervalle werden analog definiert.

Borel'sche Mengen I

Definition 100

Sei $\Omega \subseteq \mathbb{R}^n$ ein (geschlossenes, halboffenes, offenes) Intervall. Die Menge $\mathcal{B}(\Omega)$ der Borel'schen Mengen über Ω ist die kleinste Menge, die folgende Bedingungen erfüllt:

- Jedes geschlossene Subintervall von Ω gehört zu $\mathcal{B}(\Omega)$.
- Wenn $A \in \mathcal{B}(\Omega)$, dann $\bar{A} \in \mathcal{B}(\Omega)$.
- Wenn $A_n \in \mathcal{B}(\Omega)$ für alle $n \in \mathbb{N}$, dann $\displaystyle\bigcup_{n=1}^{\infty} A_n \in \mathcal{B}(\Omega)$.

Borel'sche Mengen II

Nicht-Borel'sche Mengen existieren, aber es ist sehr schwer, eine zu finden!

- $\{x\}$ bildet eine Borel'sche Menge für jedes $x \in \Omega$.
- Jede abzählbare Teilmenge von Ω ist eine Borel'sche Menge.
- Die Irrationalzahlen bilden eine Borel'sche Menge.
- Der Einheitskreis ist eine Borel'sche Menge über $[0,1] \times [0,1]$.

Lemma 101

$\mathcal{B}(\Omega)$ ist die kleinste σ-Algebra, die alle geschlossene Subintervalle von Ω enthält.

Borel'sche Mengen III: Volumen einer Borel'sche Menge

Borel'schen Mengen kann man eine Länge, Fläche, Volumen etc. zuordnen.

Definition 102

Eine Funktion $\mu\colon \mathcal{B}(\Omega) \to \mathbb{R}^n \cup \{\infty\}$ ist eine Maßfunktion, falls folgende Bedingungen erfüllt sind:

- $\mu(A) \geq 0$ für alle $A \in \mathcal{B}(\Omega)$,
- $\mu(\emptyset) = 0$ und
- wenn $A_1, A_2, \ldots \in \mathcal{B}(\Omega)$ paarweise disjunkt sind, dann

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i).$$

Borel'sche Mengen IV: Volumen einer Borel'sche Menge

Satz 103 (Caratheodory)

Sei Ω ein Intervall. Es gibt eine und nur eine Maßfunktion $\mu\colon \mathcal{B}(\Omega) \to \mathbb{R}^n$ *mit*

$$\mu\left(\prod_{i=1}^{n}[a_i, b_i]\right) = \prod_{i=1}^{n}(b_i - a_i) \; .$$

Intuition: Jede Borel'sche Menge hat eine wohldefinierte Länge, Fläche, Volumen, ...

- $\mu(\{x\}) = 0$ für jedes $x \in \Omega$.
- $\mu(\mathbb{Q}) = 0$, denn $\mathbb{Q}$ ist abzählbar.
- $\mu([0, 1] \setminus \mathbb{Q}) = 1$.

Lebesgue Integral I: Messbare Funktionen

Wir suchen nach Dichtefunktionen mit der Eigenschaft

> Für jeden Dichtwert ist das Gesamtgewicht der Punkte mit dieser Dichte wohldefiniert.

Noch allgemeiner:

> Für jede Borel'sche Menge M von Dichtwerten ist das Gesamtgewicht der Punkte mit Dichte in M wohldefiniert.

Lebesgue Integral II: Messbare Funktionen

Definition 104

Sei Ω ein Intervall. Eine Funktion $f : \Omega \to \mathbb{R}$ heißt messbar, falls das Urbild jeder Borel'schen Menge ebenfalls eine Borel'sche Menge ist.

- Indikatorfunktionen von Borel'schen Mengen sind messbar.
- Stetige Funktionen sind messbar.
- Summen und Produkte von messbaren Funktionen sind wiederum messbar.
- Die Verkettung messbarer Funktionen ist messbar.
- Der punktweise Grenzwert einer Folge von messbaren Funktionen ist wiederum messbar.

Lebesgue-Integral III: Informelle Definition

Jeder messbaren Funktion $f\colon \Omega \to \mathbb{R}_0^+$ kann ein Integral, das so genannte Lebesgue-Integral, geschrieben $\int f \, \mathrm{d}x$, zugeordnet werden.

$\int f \, \mathrm{d}x$ ist eine Funktion $\mathcal{B}(\Omega) \to \mathbb{R}$, die jedem Borel'schen Menge A eine reelle Zahl $\displaystyle\int_A f \, \mathrm{d}x$ zuordnet.

Informelle Idee für die Definition von $\displaystyle\int_A f \, \mathrm{d}x$ (nicht korrekt!):

- Partitioniere $\mathbb{R}_0^+$ in Subintervallen $[0, a_1], [a_1, a_2] \ldots [a_n, \infty)$.
- Für jedes $[a_i, a_{i+1}]$, berechne $a_i \cdot \mu(f^{-1}(\,[a_i, a_{i+1}]\,) \cap A)$

  ($a_i \times$ Maß der Punkte x mit $f(x) \in [a_i, a_{i+1}]$)
- Addiere die Ergebnisse.
- Nehme den Grenzwert wenn $n \to \infty$.

Lebesgue-Integral IV: Bezug zur Riemann-Integral

Satz 105

Sei $A = [a, b] \subseteq \Omega$ mit $a, b \in \mathbb{R}$ und sei $f \colon A \to \mathbb{R}$ beschränkt (d.h., es gibt $c \in \mathbb{R}$ mit $f(x) \leq c$ für alle $x \in A$).
Wenn $\int_a^b f(x)dx$ existiert, dann gilt

$$\int_A f \, \mathrm{d}x = \int_a^b f(x)dx$$

Der Satz gilt auch für Intervalle $(-\infty, b]$, $[a, \infty)$, oder $(-\infty, \infty)$ und kann auch auf mehrere Dimensionen erweitert werden.

W'keitsdichte

Definition 106

Sei $\Omega \subseteq \mathbb{R}^n$ ein Intervall. Eine W'keitsdichte über Ω ist eine messbare Funktion $f : \Omega \to \mathbb{R}$ mit $\displaystyle\int_\Omega f \,\mathrm{d}x = 1$.

Physikalische Analogie: jede Borel'sche Menge hat ein wohldefiniertes Gewicht und Ω wiegt 1 Kg.

Satz 107

Sei $\Omega \subseteq \mathbb{R}^n$ ein Intervall und sei $f : \Omega \to \mathbb{R}$ eine W'keitsdichte. Die Abbildung $\Pr\colon \mathcal{B}(\Omega) \to \mathbb{R}_0^+$ mit $\Pr[A] = \displaystyle\int_A f \cdot dx$ erfüllt die Kolmogorov-Axiome und die Triple $(\Omega, \mathcal{B}(\Omega), \Pr)$ bildet einen W'keitsraum.

W'keitsräume: Zusammenfassung

Der Ansatz der informellen Einführung ist ausreichend für fast alle Beispiele dieser Vorlesung .

Es hat jedoch Probleme, insbesondere für Fragen, die nicht-terminierenden Zufallsexperimenten betreffen.

Die formale Definition von W'keitsraum erlaubt eine fundierte Betrachtung dieser Fragen.

Wir haben die formale Definition nur für den Fall betrachtet, in dem Ω ein Intervall ist. Sie kann jedoch viel allgemeiner formuliert werden.

Bertrand'sches Paradoxon I

In diskreten W'keisträume gibt es nur eine mögliche Modellierung von „alle Ausgänge sind gleichwahrscheinlich“.

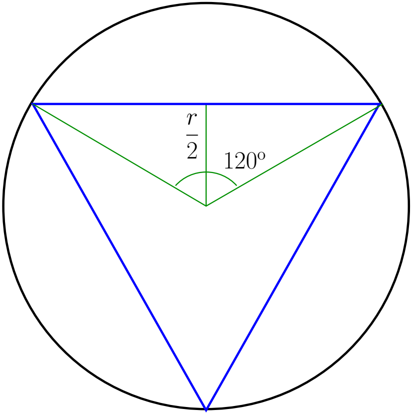
Im kontinuierlichen Fall kann es verschiedene Möglichkeiten geben, „gleichwahrscheinliche Ausgänge“ zu modellieren, die zu verschiedenen Ergbenissen führen!

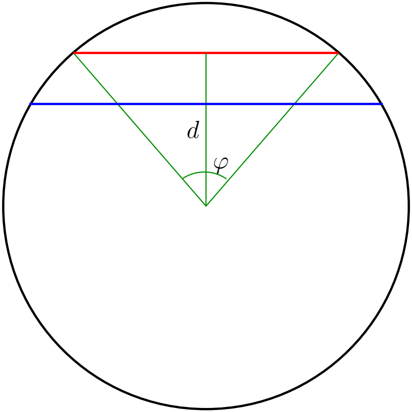
Beispiel 108 (Bertrand'sches Paradoxon)

Wir betrachten einen Kreis von Radius r mit einem eingeschriebenen gleichseitigen Dreieck (Spitze nach unten). Wir wählen eine waagerechte Sehne des oberen Halbkreises zufällig. Alle diese Sehnen seien “gleichwahrscheinlich”.

Frage: Was ist die Wahrscheinlichkeit, mit der die Länge der Sehne die Seitenlänge des Dreiecks übersteigt?

Bertrand'sches Paradoxon II





Bertrand'sches Paradoxon II

Die Seiten des Dreiecks haben Abstand $\frac{r}{2}$ vom Mittelpunkt.

Die Lage der Sehne ist

- durch den Abstand d zum Kreismittelpunkt, oder
- durch den Winkel φ mit dem Kreismittelpunkt

eindeutig festgelegt.

Modell 1: $\Omega = [0, r]$, W'keitsdichte von d: $f(x) = 1/r$.

A tritt ein, wenn $d < \frac{r}{2}$, und es folgt $\Pr[A] = \frac{1}{2}$.

Modell 2: $\Omega = [0, 180]$, W'keitsdichte von φ: $f(x) = 1/180$.

A tritt ein, wenn $\varphi \in]120°, 180°]$, und es folgt somit $\Pr[A] = \frac{1}{3}$.

15. (Kontinuierliche) Zufallsvariablen

Zufallsvariablen I: Informelle Einführung

Beispiel 109

Wir vergleichen zwei Probleme:

Problem I: Eine Zahl $x \in \{1, \ldots, 10\}$ wird zufällig gewählt.
Alle Zahlen haben die gleiche W'keit, gewählt zu werden.
Wenn x gewählt wird, dann gewinnt man x^2 Euro.

Frage: Mit welcher W'keit gewinnt man höchstens 9 Euro?

Problem II: Eine reelle Zahl $x \in [1, 10]$ wird zufällig gewählt.
Alle Zahlen haben die gleiche W'keit, gewählt zu werden.
Wenn x gewählt wird, dann gewinnt man x^2 Euro.

Frage: Mit welcher W'keit gewinnt man mindestens 9 Euro?

Zufallsvariablen II: Informelle Einführung

Modell I: $\Omega = \{1, \dots, 10\}$ $\qquad \Pr[i] = 1/10$.

Zufallsvariable $G\colon \Omega \to \mathbb{R}$ mit $G(\omega) = \omega^2$.

$\Pr[G \leq 9] = \Pr[\{1, 2, 3\}] = 3/10$.

Modell II: $\Omega = [1, 10]$ $\qquad$ W'keitsdichte $f(x) = 1/9$.

Zufallsvariable $G\colon \Omega \to \mathbb{R}$ mit $G(\omega) = \omega^2$.

$$\Pr[G \leq 9] = \Pr[\ [1, 3]\ ] = \int_1^3 \frac{1}{9}\,\mathrm{d}x = 2/9$$

Und wenn „$G \leq 9$“ keine Borel'sche Menge wäre ?

Zufallsvariablen III: Definition, Verteilung, Dichte

Definition 110

Sei $(\Omega, \mathcal{A}, \Pr)$ ein W'keitsraum. Eine kontinuierliche Zufallsvariable ist eine messbare Abbildung $X : \Omega \to \mathbb{R}$.

Eine Zufallsvariable $X : \Omega \to \mathbb{R}$ ist stetig wenn es eine W'keitsdichtefunktion $f_X : \mathbb{R} \to \mathbb{R}_0^+$ (die Dichte von X) gibt mit

$$\Pr[X \in A] = \int_A f_X(x)\,\mathsf{d}x$$

Die Verteilung oder Verteilungsfunktion von X ist die Funktion:

$$F_X(x) := \Pr[X \leq x] = \int_{-\infty}^{x} f_X(t)\,\mathsf{d}t.$$

In den kommenden Folien: Zufallsvariable = stetige Zufallsvariable.

Zufallsvariablen IV: Eigenschaften

Sei X eine Zufallsvariable:

- F_X ist monoton steigend und stetig.
- $\Pr[a \leq X \leq b] = F_X(b) - F_X(a)$.
- $\lim\limits_{x \to -\infty} F_X(x) = 0$ und $\lim\limits_{x \to \infty} F_X(x) = 1$.
- Zwischen den Ereignissen „$a < X \leq b$“, „$a \leq X \leq b$“, „$a \leq X < b$“ und „$a < X < b$“ besteht kein wesentlicher Unterschied, da

$$\int_{[a,b]} f(t)\,\mathrm{d}t = \int_{(a,b]} f(t)\,\mathrm{d}t = \int_{[a,b[} f(t)\,\mathrm{d}t = \int_{]a,b[} f(t)\,\mathrm{d}t.$$

- $f_X(x) = \dfrac{\mathrm{d}\,F_X(x)}{\mathrm{d}\,x}$

Zufallsvariablen V: Gleichverteilung

Definition 111 (Gleichverteilung)

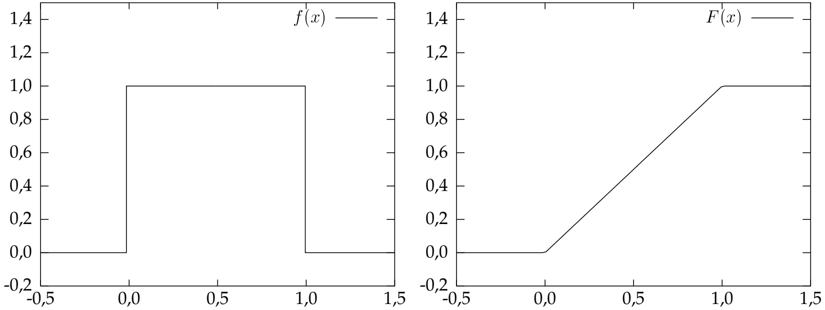
Eine kontinuierliche Zufalsvariable X ist gleichverteilt auf dem Intervall $[a, b]$ wenn

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

Verteilungsfunktion der Gleichverteilung:

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t = \begin{cases} 0 & \text{für } x < a, \\ \dfrac{x-a}{b-a} & \text{für } a \le x \le b, \\ 1 & \text{für } x > b. \end{cases}$$

Zufallsvariablen VI: Gleichverteilung



Gleichverteilung über dem Intervall $[0, 1]$

Zufallsvariablen VII: Exponentialverteilung

Definition 112

Eine kontinuierliche Zufalsvariable X ist exponentialverteilt wenn

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

Verteilungsfunktion der Exponentialverteilung:

$$\begin{aligned} F_X(x) &= \int_0^x e^{-\lambda t}\, \mathrm{d}t = \\ &= \left[-e^{-\lambda t}\right]_0^x \\ &= \begin{cases} 1 - e^{-\lambda y} & \text{für } y \geq 0, \\ 0 & \text{sonst.} \end{cases} \end{aligned}$$

W'keitsräume als Zufallsvariablen

- Sei $R = (\Omega, \mathcal{B}(\mathbb{R}), \Pr)$ ein W'keitsraum mit W'keitsdichte $f : \Omega \to \mathbb{R}_0^+$.
  Sei $X : \Omega \to \mathbb{R}$ eine Zufallsvariable.
- $R' = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \Pr')$ mit $\Pr'[A] = \int_A f_X \, \mathrm{d}x$ bildet ebenfalls einen W'keitsraum.
- Wenn $\Omega = \mathbb{R}$ und $X(\omega) = \omega$, dann sind die zwei Räume identisch.
- W'keitsräume mit $\Omega = \mathbb{R}$ werden daher oft direkt als Zufallsvariablen modelliert.

Beispiel: Experiment: Eine Zahl $x \in [0, 1]$ wird zufällig gewählt.

Modell: Sei X gleichverteilt auf $[0, 1]$.

Zusammengesetzte Zufallsvariablen I

Beispiel 113

Die Zeit X, die ein Server für eine Anfrage braucht sei exponentialverteilt mit Dichte $f_X(t) = \lambda e^{-\lambda t}$.
Die Gesamtkosten einer Anfrage seien $at + b$, wenn die Anfrage in t Sekunden beantwortet wird.
Frage: Mit welcher W'keit Kostet eine Anfrage mindestens k Euro ($k \geq b$)?

Die Kosten werden von der Zufallsvariable $Y := aX + b$ modelliert.
Wir müssen $\Pr[Y \leq k]$ berechnen.

Zusammengesetzte Zufallsvariablen II

- Allgemein gilt für $Y = g(X)$:

$$F_Y(y) = \Pr[Y \leq y] = \Pr[g(X) \leq y] = \int_C f_X(t)\,\mathrm{d}t.$$

mit $C := \{t \in \mathbb{R} \mid g(t) \leq y\}$.

- Wenn g messbar, dann ist C eine Borel'sche Menge und das Integral ist definiert.
- Wenn $g(X) = aX + b$ dann

$$\Pr[g(X) \leq y] = \Pr\left[X \leq \frac{y-b}{a}\right] = F_X\left(\frac{y-b}{a}\right)$$

In unserem Beispiel:

$$\Pr[Y \leq k] = F_X\left(\frac{k-b}{a}\right) = 1 - e^{\lambda(k-b)/a}$$

Simulation von Zufallsvariablen I: Die Inversionsmethode

Die Simulation einer Zufallsvariablen X mit Dichte f_X ist die algorithmische Erzeugung von Zufallswerten, deren Verteilung der Verteilung von X entspricht.

Zufallsgeneratoren simulieren eine Zufallsvariable U gleichverteilt über $[0, 1]$.

Wie können Variablen mit anderen Verteilungen simuliert werden?

Simulation von Zufallsvariablen II: Die Inversionsmethode

Satz 114

Sei X eine Zufallsvariable mit einer streng monoton wachsenden Verteilungsfunktion F_X. Dann hat F_X eine (eindeutige) inverse Funktion F_X^{-1} mit $F_X(F_X^{-1}(x)) = x$ für alle $x \in (0,1)$.

Sei $\tilde{X} = F_X^{-1}(U)$. Es gilt $F_{\tilde{X}}(t) = F_X(t)$ für alle $t \in (0,1)$.

Beweis:

$$\begin{aligned}
F_{\tilde{X}}(t) &= \Pr[\tilde{X} \le t] \\
&= \Pr[F_X^{-1}(U) \le t] \\
&= \Pr[U \le F_X(t)] \\
&= F_U(F_X(t)) \\
&= F_X(t)\,.
\end{aligned}$$

$\square$

Beispiel 115

Wir simulieren die Variable X mit Exponentialverteilung

$$F_X(x) = 1 - e^{-\lambda t} \text{ für } t \geq 0 \; .$$

Wir erhalten auf $(0,1)$ die Umkehrfunktion

$$F_X^{-1}(t) = -\frac{1}{\lambda}\ln(1-t) \; .$$

und damit

$$\tilde{X} = -\frac{1}{\lambda}\ln(1-U) \; .$$

Kontinuierliche ZV als Grenzwerte diskreter ZV I

Sei X eine kontinuierliche Zufallsvariable.

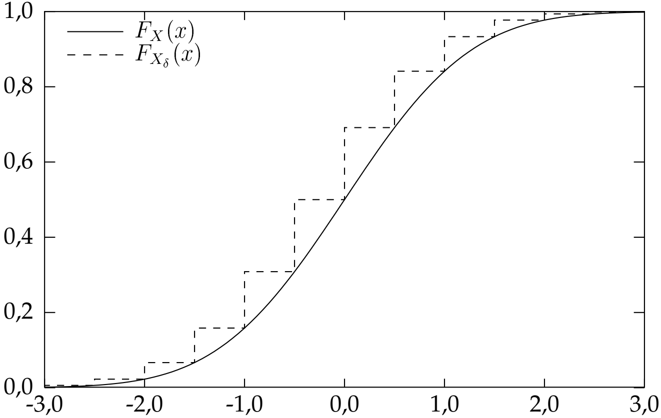
Wir konstruieren eine diskrete Zufallsvariable X_δ, indem wir für ein festes $\delta > 0$ definieren

$$X_\delta = n\delta \iff X \in [n\delta, (n+1)\delta[\text{ für } n \in \mathbb{Z}.$$

Es gilt

$$\Pr[X_\delta = n\delta] = F_X((n+1)\delta) - F_X(n\delta)\,.$$

Kontinuierliche ZV als Grenzwerte diskreter ZV II



Für $\delta \to 0$ nähert sich die Verteilung von X_δ der Verteilung von X immer mehr an.

Erwartungswert und Varianz I

Definition 116

Für eine kontinuierliche Zufallsvariable X ist der Erwartungswert definiert durch

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} t \cdot f_X(t)\, \mathrm{d}t,$$

sofern das Integral $\int_{-\infty}^{\infty} |t| \cdot f_X(t)\, \mathrm{d}t$ endlich ist.

Für die Varianz gilt entsprechend

$$\mathrm{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \int_{-\infty}^{\infty} (t - \mathbb{E}[X])^2 \cdot f_X(t)\, \mathrm{d}t,$$

wenn $\mathbb{E}[(X - \mathbb{E}[X])^2]$ existiert.

Erwartungswert und Varianz II

Beispiel 117

Für Erwartungswert und Varianz der Gleichverteilung ergibt sich

$$\begin{aligned}
\mathbb{E}[X] &= \int_a^b t \cdot \frac{1}{b-a}\,\mathrm{d}t = \frac{1}{b-a} \cdot \int_a^b t \cdot \mathrm{d}t \\
&= \frac{1}{2(b-a)} \cdot [t^2]_a^b \\
&= \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}, \\
\mathbb{E}[X^2] &= \frac{1}{b-a} \cdot \int_a^b t^2 \cdot \mathrm{d}t = \frac{b^2+ba+a^2}{3}, \\
\mathrm{Var}[X] &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \ldots = \frac{(a-b)^2}{12}\,.
\end{aligned}$$

Erwartungswert einer zusammengesetzten Variable I

Lemma 118

Sei X eine kontinuierliche Zufallsvariable, und sei $Y := g(X)$. Dann gilt

$$\mathbb{E}[Y] = \int_{-\infty}^{\infty} g(t) \cdot f_X(t)\,\mathrm{d}t\,.$$

Beweis:

Nur für $Y := a \cdot X + b$ mit $a, b \in \mathbb{R}$ und $a > 0$. Es gilt

$$\mathbb{E}[a \cdot X + b] = \int_{-\infty}^{\infty} t \cdot f_Y(t)\,\mathrm{d}t = \int_{-\infty}^{\infty} t \cdot f_X\left(\frac{t-b}{a}\right) \cdot \frac{1}{a}\,\mathrm{d}t.$$

Durch die Substitution $u := (t-b)/a$ mit $\mathrm{d}\,u = (1/a)\,\mathrm{d}\,t$ erhalten wir

$$\mathbb{E}[a \cdot X + b] = \int_{-\infty}^{\infty} (au+b) f_X(u)\,\mathrm{d}u.$$

$\square$

Gemeinsame Dichte und Verteilung

Definition 119

Die gemeinsame Dichte zweier kontinuierlichen Zufallsvariablen X, Y ist eine integrierbare Dichtefunktion $f_{X,Y} : \mathbb{R}^2 \to \mathbb{R}_0^+$ mit

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y = 1$$

Für ein Ereignis $A \in \mathcal{B}(\mathbb{R}^2)$ gilt

$$\Pr[A] = \int_A f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

Die (gemeinsame) Verteilung $F_{X,Y} : \mathbb{R}^2 \to [0,1]$ von X und Y ist definiert durch:

$$F_{X,Y}(x,y) = \Pr[X \le x, Y \le y] = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(u,v)\,\mathrm{d}u\,\mathrm{d}v.$$

Randverteilungen

Definition 120

Sei $f_{X,Y}$ die gemeinsame Dichte der Zufallsvariablen X und Y.
Die Randverteilung der Variablen X ist gegeben durch

$$F_X(x) = \Pr[X \le x] = \int_{-\infty}^{x} \left[\int_{-\infty}^{\infty} f_{X,Y}(u,v) \,\mathrm{d}v \right] \mathrm{d}u.$$

Analog nennen wir

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,v) \,\mathrm{d}v$$

die Randdichte von X. Entsprechende Definitionen gelten symmetrisch für Y.

Unabhängigkeit I

Definition 121

Zwei kontinuierliche Zufallsvariablen X und Y heißen unabhängig, wenn

$$\Pr[X \leq x, Y \leq y] = \Pr[X \leq x] \cdot \Pr[Y \leq y]$$

für alle $x, y \in \mathbb{R}$ gilt.

Dies ist gleichbedeutend mit

$$F_{X,Y}(x, y) = F_X(x) \cdot F_Y(y)\,.$$

Differentiation ergibt

$$f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y)\,.$$

Unabhängigkeit II

Definition 122

Die kontinuierlichen Zufallsvariablen $X_1, \ldots, X_n$ sind unabhängig, wenn

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = F_{X_1}(x_1) \cdot \ldots \cdot F_{X_n}(x_n)$$

bzw.

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdot \ldots \cdot f_{X_n}(x_n)$$

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

Summen von unabhängigen Zufallsvariablen I

Satz 123

Seien X und Y unabhängige kontinuierliche Zufallsvariablen. Für die Dichte von $Z := X + Y$ gilt

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) \cdot f_Y(z - x) \, \mathrm{d}x \, .$$

Beweis:

Nach Definition der Verteilungsfunktion gilt

$$F_Z(t) = \Pr[Z \leq t] = \Pr[X + Y \leq t] = \int_{A(t)} f_{X,Y}(x, y) \, \mathrm{d}x\mathrm{d}y$$

wobei $A(t) = \{(x, y) \in \mathbb{R}^2 \mid x + y \leq t\}$.

Summen von unabhängigen Zufallsvariablen II

Beweis (Forts.):

Aus der Unabhängigkeit von X und Y folgt

$$F_Z(t) = \int_{A(t)} f_X(x) \cdot f_Y(y)\,\mathrm{d}x\mathrm{d}y = \int_{-\infty}^{\infty} f_X(x) \cdot \left(\int_{-\infty}^{t-x} f_Y(y)\,\mathrm{d}y \right) \mathrm{d}x.$$

Mittels der Substitution $z := x + y$, $\mathrm{d}\,z = \mathrm{d}\,y$ ergibt sich

$$\int_{-\infty}^{t-x} f_Y(y)\,\mathrm{d}y = \int_{-\infty}^{t} f_Y(z - x)\,\mathrm{d}z$$

und somit

$$F_Z(t) = \int_{-\infty}^{t} \left(\int_{-\infty}^{\infty} f_X(x) f_Y(z - x)\,\mathrm{d}x \right) \mathrm{d}z\,.$$

$\square$

16. Die Normalverteilung

Die Normalverteilung I

Die Normalverteilung ist in gewisser Weise das kontinuierliche Analogon zur Binomialverteilung wenn $n \to \infty$ bei konstanter Erfolgswahrscheinlichkeit p.

Definition 124

Eine Zufallsvariable X mit Wertebereich $W_X = \mathbb{R}$ heißt normalverteilt mit den Parametern $\mu \in \mathbb{R}$ und $\sigma \in \mathbb{R}^+$, wenn sie die Dichte

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) =: \varphi(x;\mu,\sigma)$$

besitzt.

Die Normalverteilung II

In Zeichen schreiben wir $X \sim \mathcal{N}(\mu, \sigma^2)$.

$\mathcal{N}(0,1)$ heißt Standardnormalverteilung.

Die Dichte $\varphi(x; 0, 1)$ kürzen wir durch $\varphi(x)$ ab.

Die Verteilungsfunktion zu $\mathcal{N}(\mu, \sigma^2)$ ist

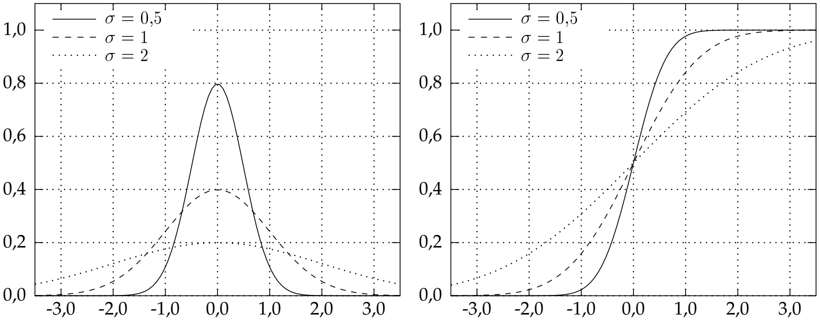
$$F(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \int_{-\infty}^{x} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) \mathrm{d}t =: \Phi(x; \mu, \sigma)\,.$$

Diese Funktion heißt Gauß'sche Φ-Funktion.

Die Verteilung $\varphi(x; 0, 1)$ kürzen wir durch $\Phi(x)$ ab.

Es gibt keinen geschlossenen Ausdruck für $\Phi(x; \mu, \sigma)$.

Die Normalverteilung III



Dichte und Verteilung von $\mathcal{N}(0, \sigma^2)$

Die Normalverteilung IV: Erwartungswert und Varianz

Wir berechnen zunächst Erwartungswert und Varianz einer Zufallsvariable $X \sim \mathcal{N}(0, 1)$. (Satz 126).

(D.h., wir betrachten den Spezialfall $\mu = 0, \sigma = 1$.)

Dann berechnen wir Erwartungswert und Varianz für beliebige Werte der Parameter μ und σ. (Satz 128).

Erwartungswert und Varianz von $\mathcal{N}(0,1)$ I

Lemma 125

$$I := \int_{-\infty}^{\infty} e^{-x^2/2}\,\mathrm{d}x = \sqrt{2\pi}.$$

Beweis:

Wir berechnen zunächst I^2:

$$\begin{aligned} I^2 &= \left(\int_{-\infty}^{\infty} e^{-x^2/2}\,\mathrm{d}x\right)\left(\int_{-\infty}^{\infty} e^{-y^2/2}\,\mathrm{d}y\right) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)/2}\,\mathrm{d}x\,\mathrm{d}y\,. \end{aligned}$$

Erwartungswert und Varianz von $\mathcal{N}(0,1)$ II

Beweis (Forts.):

Wir gehen nun zu Polarkoordinaten über durch die Substitution:

$$x := r\cos\phi \qquad\qquad y := r\sin\phi$$

Die Derterminante der Jacobi-Matrix der Substitution ist

$$\begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial y}{\partial r} \\ \frac{\partial x}{\partial \phi} & \frac{\partial y}{\partial \phi} \end{vmatrix} = \begin{vmatrix} \cos\phi & \sin\phi \\ -r\sin\phi & r\cos\phi \end{vmatrix} = r(\cos^2\phi + \sin^2\phi) = r$$

und wir erhalten

$$I^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)/2}\,\mathrm{d}x\,\mathrm{d}y \;=\; \int_0^{2\pi}\int_0^{\infty} e^{-r^2/2}\, r \;\mathrm{d}r\,\mathrm{d}\phi$$

$$= \int_0^{2\pi}\left[-e^{-r^2/2}\right]_0^{\infty}\mathrm{d}\phi \;=\; \int_0^{2\pi} 1\,\mathrm{d}\phi \;=\; 2\pi.$$

Erwartungswert und Varianz von $\mathcal{N}(0,1)$ III

Satz 126

X sei $\mathcal{N}(0,1)$-verteilt. Dann gilt

$$\mathbb{E}[X] = 0 \textit{ und } \quad \mathrm{Var}[X] = 1.$$

Beweis:

Es gilt

$$\mathbb{E}[X] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x \cdot e^{-x^2/2} \, \mathrm{d}x.$$

Sei $f(x) = x \cdot e^{-x^2/2}$.

Es gilt $f(-x) = -f(x)$ und damit $\mathbb{E}[X] = 0$.

Erwartungswert und Varianz von $\mathcal{N}(0,1)$ IV

Für die Varianz berechnen wir $\mathbb{E}[X^2]$.

Mit Lemma 125 gilt

$$\int_{-\infty}^{\infty} e^{-x^2/2}\,\mathrm{d}x = \sqrt{2\pi}.$$

Durch partielle Integration erhalten wir

$$\begin{aligned}\sqrt{2\pi} &= \int_{-\infty}^{\infty} e^{-x^2/2}\,\mathrm{d}x \\ &= xe^{-x^2/2}\Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} x^2 \cdot e^{-x^2/2}\,\mathrm{d}x \\ &= 0 \quad + \quad \sqrt{2\pi}\cdot\mathbb{E}[X^2]\end{aligned}$$

Daraus folgt $\mathbb{E}[X^2] = 1$ und somit $\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = 1$.

Erwartungswert und Varianz von $\mathcal{N}(\mu, \sigma)$ I

Lemma 127 (Lineare Transformation der Normalverteilung)

Wenn $X \sim \mathcal{N}(\mu, \sigma^2)$ und $Y = aX + b$ dann

$$Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)\,.$$

Beweis: Fall „$a > 0$“ („$a < 0$“ analog):

$$\begin{aligned}\Pr[Y \le y] &= \Pr[aX + b \le y] = \Pr\left[X \le \frac{y-b}{a}\right] \\ &= \frac{1}{\sqrt{2\pi}\sigma} \cdot \int_{-\infty}^{(y-b)/a} \exp\left(-\frac{(u-\mu)^2}{2\sigma^2}\right) \mathrm{d}u.\end{aligned}$$

Mit $u = (v - b)/a$ und $\mathrm{d}u = (1/a)\,\mathrm{d}v$:

$$\Pr[Y \le y] = \frac{1}{\sqrt{2\pi}a\sigma} \cdot \int_{-\infty}^{y} \exp\left(-\frac{(v-(a\mu+b))^2}{2(a\sigma)^2}\right) \mathrm{d}v\,.$$

Also $Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.

Erwartungswert und Varianz von $\mathcal{N}(\mu, \sigma)$ II

Satz 128

Sei $X \sim \mathcal{N}(\mu, \sigma^2)$*. Dann gilt*

$$\mathbb{E}[X] = \mu \textit{ und } \mathrm{Var}[X] = \sigma^2 \,.$$

Beweis:

Sei $Y := \frac{1}{\sigma}(X - \mu)$ (und so $X = \sigma Y + \mu$).

Mit Lemma 127 gilt $Y \sim \mathcal{N}\left(\left(\frac{1}{\sigma}\right)\mu - \frac{\mu}{\sigma}, \left(\frac{1}{\sigma}\right)^2 \sigma^2\right) \sim \mathcal{N}(0,1)$

und so

$$\mathbb{E}[X] = \mathbb{E}[\sigma Y + \mu] = \sigma \cdot \mathbb{E}[Y] + \mu = \mu$$

und

$$\mathrm{Var}[X] = \mathrm{Var}[\sigma Y + \mu] = \sigma^2 \cdot \mathrm{Var}[Y] = \sigma^2.$$

$\square$

Normalverteilung als Grenzwert der Binomialverteilung I

Seien $X_1, \ldots, X_n$ unabhängige Bernoulli-verteilte Zufallsvariablen mit gleicher Erfolgswahrscheinlichkeit p.

Sei $H_n := X_1 + \ldots + X_n$. Es gilt $H_n \sim \mathrm{Bin}(n; p)$ mit

$$\mathbb{E}[H_n] = np \qquad \text{und} \qquad \mathrm{Var}[H_n] = npq\ .$$

Die standardisierte Variable $H_n^* := \dfrac{H_n - np}{\sqrt{npq}}$ erfüllt

$$\mathbb{E}[H_n^*] = \frac{\mathbb{E}[H_n] - np}{\sqrt{npq}} = 0 \quad \text{und} \quad \mathrm{Var}[H_n^*] = \left(\frac{1}{\sqrt{npq}}\right)^2 \cdot \mathrm{Var}[H_n] = 1\ .$$

Wir betrachten die Sequenz $H_1^*, H_2^*, H_3^* \ldots$ und untersuchen den Grenzwert von $F_{H_1^*}, F_{H_2^*}, F_{H_3^*} \ldots$.

Normalverteilung als Grenzwert der Binomialverteilung II

Satz 129 (Grenzwertsatz von de Moivre)

Seien $X_1, \ldots, X_n$ unabhängige Bernoulli-verteilte Zufallsvariablen mit gleicher Erfolgswahrscheinlichkeit p.

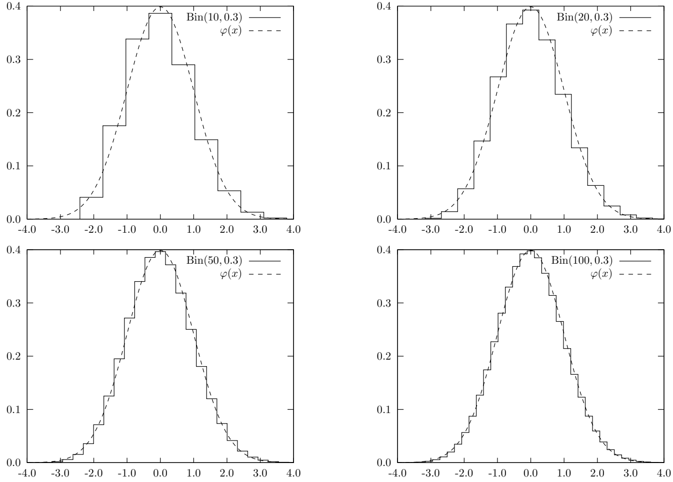
Sei $H_n := X_1 + \ldots + X_n$. Die Verteilung $F_{H_n^*}$ der Zufallsvariable

$$H_n^* := \frac{H_n - np}{\sqrt{npq}}$$

konvergiert gegen die Standardnormalverteilung $\mathcal{N}(0,1)$ für $n \to \infty$.

Der Satz wird später als Spezialfall des zentralen Grenzwertsatzes bewiesen. (Satz 134 und Korollar 139).

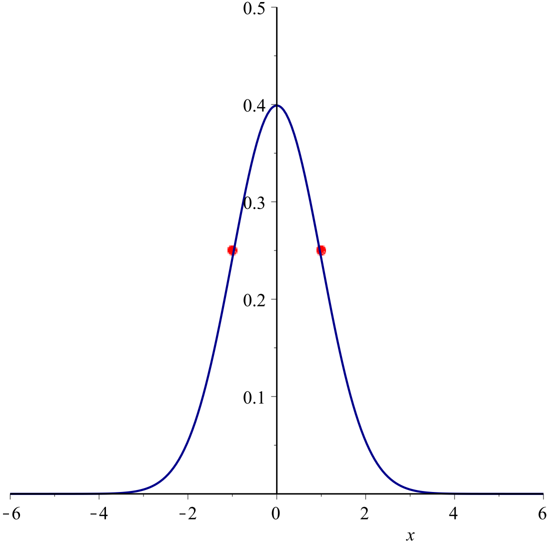
Normalverteilung als Grenzwert der Binomialverteilung III



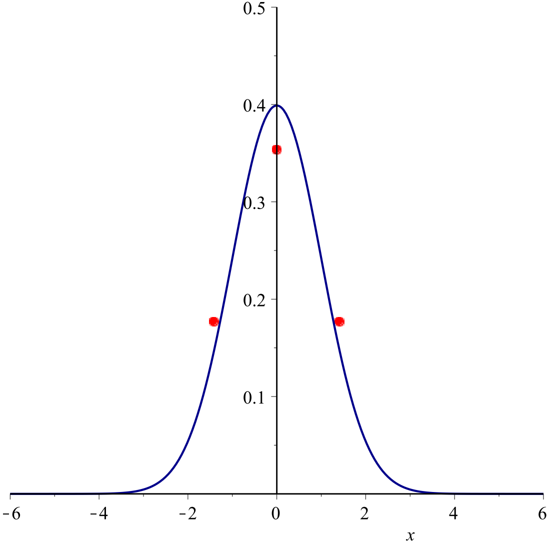
Vergleich von Binomial- und Normalverteilung

$\mathrm{Bin}(n, 0.3)$ bei $0.3n$ zentriert, mit $\sqrt{0.3 \cdot 0.7n}$ horizontal gestaucht und vertikal gestreckt

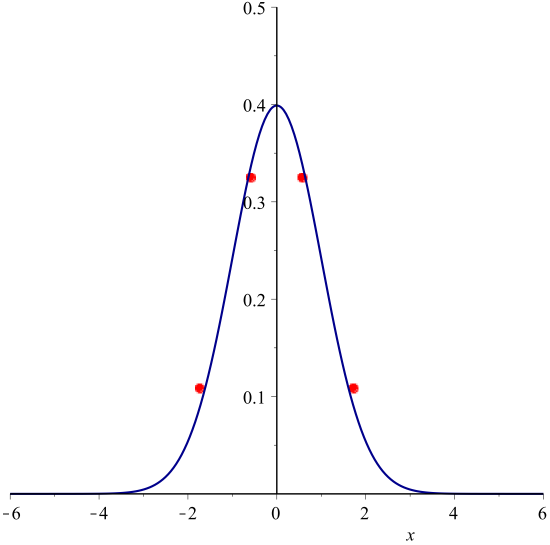
Binomial → Normal I



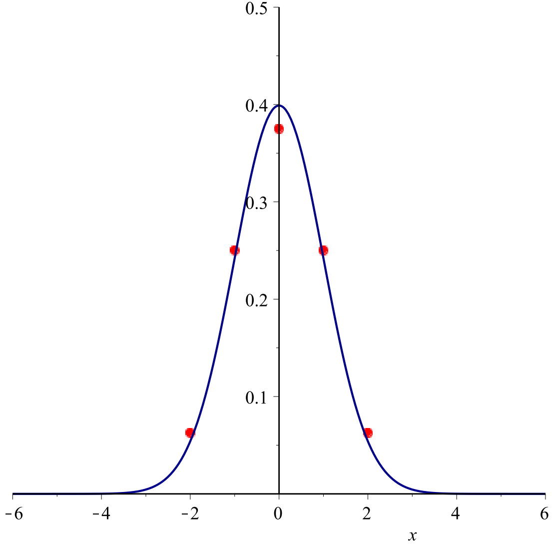
Binomial → Normal II



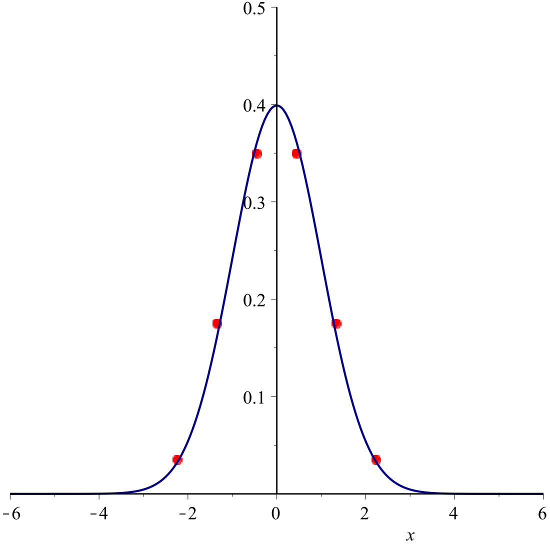
Binomial $\to$ Normal III



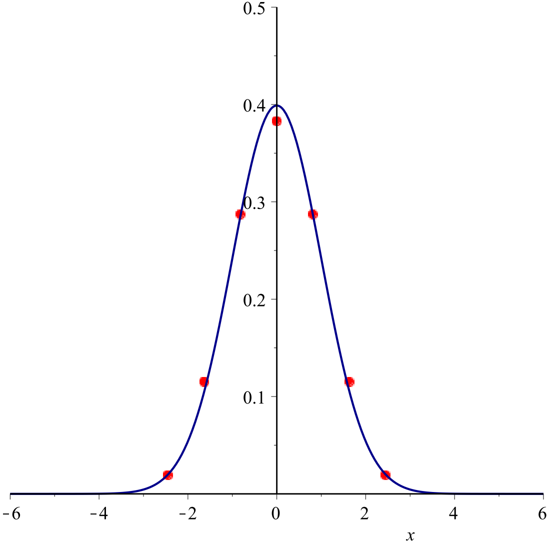
Binomial $\rightarrow$ Normal IV



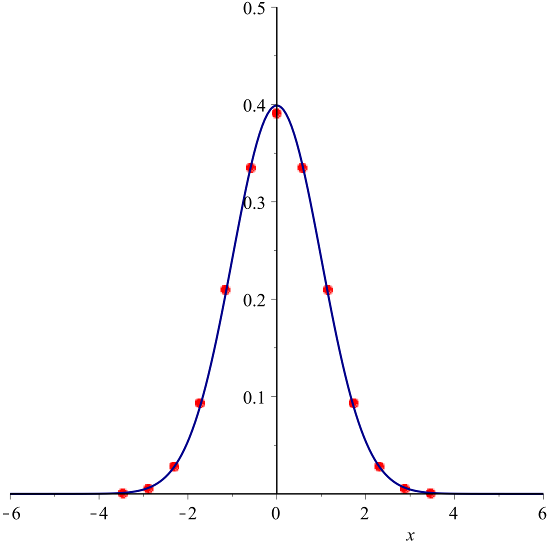
Binomial → Normal V



Binomial $\to$ Normal VI



Binomial → Normal XII



Normalverteilung als Grenzwert der Binomialverteilung III

H_n beschreibt die Anzahl der Erfolge bei n Versuchen.

$\frac{H_n}{n}$ beschreibt die Frequenz der Erfolge.

Die folgende Aussage ist eine Konsequenz von Satz 129:

Korollar 130

Sei $H_n \sim \mathrm{Bin}(n, p)$ eine binomialverteilte Zufallsvariable.

Die Verteilung von $\frac{H_n}{n}$ konvergiert gegen

$$\mathcal{N}\left(p, \frac{pq}{n}\right)$$

für $n \to \infty$. Insbesondere gilt $\lim_{n\to\infty} \mathrm{Var}\left[\frac{H_n}{n}\right] = 0$.

Beispiel 131

25.05.2012, ZDF: Die Umfragen zu diesem Politbarometer wurden wie immer von der Mannheimer Forschungsgruppe Wahlen durchgeführt. Die Interviews wurden in der Zeit vom 22. bis 24. Mai 2012 bei 1312 zufällig ausgewählten Wahlberechtigten telefonisch erhoben. Die Befragung ist repräsentativ für die wahlberechtigte Bevölkerung in Deutschland. Der Fehlerbereich beträgt bei einem Parteianteil von 40 Prozent rund +/- drei Prozentpunkte und bei einem Parteianteil von zehn Prozent rund +/- zwei Prozentpunkte. Daten zur politischen Stimmung: CDU/CSU: 38 Prozent, SPD: 34 Prozent, FDP: zwei Prozent, Linke: vier Prozent, Grüne: 14 Prozent, Piraten: sechs Prozent.

Wie werden diese Fehlerbereiche berechnet?
Warum sind sie vom Parteilanteil abhängig?

Normalverteilung und Stichproben II

Jede Umfrage muss zwei Parameter angeben:

- Konfidenz- oder Vertrauensintervall.
  Intervall um den geschätzten Wert s.
  Für die CDU/DSU: $s = 0.38$, Intervall $[s - 0.03, s + 0.03]$.
  („Fehlerbereich“ ist nicht der übliche Fachbegriff.)
- Konfidenzniveau.
  Untere Schranke für die W'keit, dass der unbekannte wahre Wert p im Konfidenzintervall liegt.
  Das übliche Konfidenzniveau bei diesen Studien beträgt 95%.

Ohne die Angabe beider Parameter ist eine Umfrage Wertlos!

Normalverteilung und Stichproben III

Zu berechnen ist die Zahl δ, für die die W'keit, dass bei 1312 Befragten das Ergebnis der Umfrage im Intervall $[p-\delta, p+\delta]$ liegt, 0.95 beträgt.

Dazu modellieren wir die Befragung durch folgendes n-stufiges Experiment ($n = 1312$).

Normalverteilung und Stichproben IV

Umfrage als mehrstufiges Experiment:

- Eine Urne enthält weisse und schwarze Bälle.
- Das Experiment besteht aus n Stufen.
- In jeder Stufe wird ein zufälliger Ball aus der Urne extrahiert und zurück in die Urne gestellt.
- Sei p die (unbekannte!) W'keit, dass der extrahierter Ball schwarz ist.
- Sei k die Anzahl der gezogenen Bällen, die Schwarz sind $(0 \leq k \leq n)$.
- Ergebnis s des Experiments: $s = k/n$.
- Die Umfrage hat ein Konfidenzniveau von 95%:
  Die W'keit, dass $|s - p| \leq 0.03$ beträgt (mindestens) 95%.

Normalverteilung und Stichproben V

Es gilt $\Pr[-2 \le \mathcal{N}(0,1) \le 2] \approx 0.9554$ (Tabelle, Internet)

Wir haben

$$\begin{aligned}
0.95 &\approx \Pr[-2 \le H_n^* \le 2] \\
&= \Pr[np - 2\sqrt{np(1-p)} \le H_n \le np + 2\sqrt{np(1-p)}] \\
&= \Pr\left[p - 2\sqrt{\frac{p(1-p)}{n}} \le \frac{H_n}{n} \le p + 2\sqrt{\frac{p(1-p)}{n}}\right]
\end{aligned}$$

Damit gilt: bei n befragten und 95% Konfidenz erhählt man einen Fehlerbereich von

$$\delta = 2\sqrt{\frac{p(1-p)}{n}}$$

.

Normalverteilung und Stichproben VI

Mit $p(1-p) \leq 1/4$ für alle $0 \leq p \leq 1$ erhalten wir

$$0.95 \approx \Pr[-2 \leq H_n^* \leq 2] \leq \Pr\left[p - \sqrt{\frac{1}{n}} \leq \frac{H_n}{n} \leq p + \sqrt{\frac{1}{n}}\right]$$

Damit gilt $\delta \leq \dfrac{1}{\sqrt{n}}$ für alle p. In anderen Worten:

unabhängig vom Wert von p erhält man bei n befragten und 95% Konfidenz einen Fehlerbereich von ca. $\dfrac{1}{\sqrt{n}}$.

Bei $n = 1312$ beträgt der Fehlerbereich $\frac{1}{\sqrt{1312}} \approx 0.0276$ oder 2,8 Prozentpunkte für alle p.

Für $p \leq 0.4$: $2\sqrt{\frac{0.24}{1312}} \approx 0.0271$ oder 2,7 Prozentpunkte.

Für $p \leq 0.1$: $2\sqrt{\frac{0.09}{1312}} \approx 0.0166$ oder 1,7 Prozentpunkte.

17. Momenterzeugende Funktionen

Momenterzeugende Funktionen I

Definition 132

Zu einer Zufallsvariablen X ist die momenterzeugende Funktion gemäß

$$M_X(s) = \mathbb{E}\left[e^{Xs}\right] = \int_{-\infty}^{\infty} e^{ts} \cdot f_X(t)\,\mathrm{d}t$$

definiert.

Momenterzeugende Funktionen II: Gleichverteilung

Für eine auf $[a, b]$ gleichverteilte Zufallsvariable U gilt

$$\begin{aligned} M_U(t) = \mathbb{E}[e^{tX}] &= \int_a^b e^{tx} \cdot \frac{1}{b-a} \, \mathsf{d}x \\ &= \left[\frac{e^{tx}}{t(b-a)} \right]_a^b \\ &= \frac{e^{tb} - e^{ta}}{t(b-a)} \, . \end{aligned}$$

Momenterzeugende Funktionen III: Normalverteilung

Für eine standardnormalverteilte Zufallsvariable $N \sim \mathcal{N}(0,1)$ gilt

$$M_N(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{t\xi} e^{-\xi^2/2} \, \mathrm{d}\xi$$
$$= e^{t^2/2} \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-(t-\xi)^2/2} \, \mathrm{d}\xi = e^{t^2/2} \, .$$

Daraus ergibt sich für $Y \sim \mathcal{N}(\mu, \sigma^2)$

$$\begin{aligned} M_Y(t) &= \mathbb{E}[e^{tY}] \\ &= e^{t\mu} \cdot \mathbb{E}[e^{(t\sigma)\cdot \frac{Y-\mu}{\sigma}}] \\ &= e^{t\mu} \cdot M_N(t\sigma) \qquad \text{(wegen } \tfrac{Y-\mu}{\sigma} \sim \mathcal{N}(0,1)\text{)} \\ &= e^{t\mu + (t\sigma)^2/2} \, . \end{aligned}$$

Summe von Normalverteilungen

Satz 133 (Additivität der Normalverteilung)

Seien $X_1, \ldots, X_n$ unabhängige Zufallsvariablen mit $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ für alle $1 \leq i \leq n$.

Sei $Z := a_1 X_1 + \ldots + a_n X_n$. Es gilt $Z \sim \mathcal{N}(\mu, \sigma^2)$ mit

$$\mu = a_1\mu_1 + \ldots + a_n\mu_n \quad \textit{und} \quad \sigma^2 = a_1^2\sigma_1^2 + \ldots + a_n^2\sigma_n^2\,.$$

Beweis: Mit $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ gilt $M_{X_i}(t) = e^{t\mu_i + (t\sigma_i)^2/2}$.
Aus der Unabhängigkeit der X_i folgt

$$\begin{aligned} M_Z(t) = \mathbb{E}[e^{t(a_1X_1 + \cdots + a_nX_n)}] &= \prod_{i=1}^n \mathbb{E}[e^{(a_it)X_i}] \\ &= \prod_{i=1}^n M_{X_i}(a_it) = \prod_{i=1}^n e^{a_it\mu_i + (a_it\sigma_i)^2/2} = e^{t\mu + (t\sigma)^2/2}\,. \end{aligned}$$

18. Der zentraler Grenzwertsatz

Der zentraler Grenzwertsatz I

Dieser Satz ist von großer Bedeutung für die Anwendung der Normalverteilung in der Statistik.

Informell lautet die Aussage des Satzes:

Die Verteilung einer Summe unabhängiger
identisch ABER BELIEBIG verteilter Zufallsvariablen
nähert der Normalverteilung umso mehr an,
je mehr Zufallsvariablen an der Summe beteiligt sind.

(OK, nicht ganz: Erwartungswert und Varianz der Variablen müssen endlich sein.)

Zentraler Grenzwertsatz II

Satz 134 (Zentraler Grenzwertsatz)

Die Zufallsvariablen $X_1, \ldots, X_n$ *besitzen jeweils dieselbe Verteilung und seien unabhängig.*

Erwartungswert und Varianz von X_i *existieren für* $i = 1, \ldots, n$ *und seien mit* μ *bzw.* σ^2 *bezeichnet* ($\sigma^2 > 0$)*.*

Die Zufallsvariablen Y_n *seien definiert durch* $Y_n := X_1 + \ldots + X_n$ *für* $n \geq 1$*. Dann folgt, dass die Zufallsvariablen*

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

asymptotisch standardnormalverteilt sind, also $Z_n \sim \mathcal{N}(0, 1)$ *für* $n \to \infty$*.*

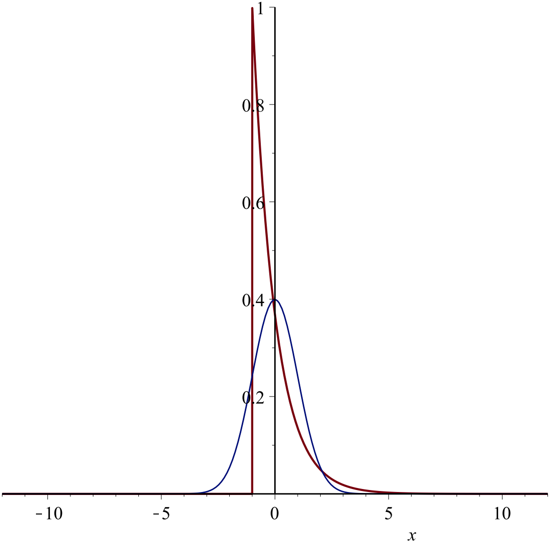
Zentraler Grenzwertsatz III

Etwas formaler ausgedrückt gilt: Die Folge der zu Z_n gehörenden Verteilungsfunktionen F_n hat die Eigenschaft

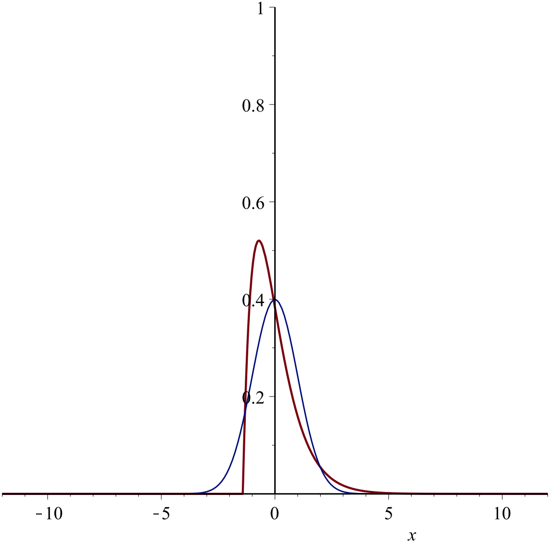
$$\lim_{n \to \infty} F_n(x) = \Phi(x) \text{ für alle } x \in \mathbb{R}.$$

Wir sagen dazu auch: Die Verteilung von H_n^* konvergiert gegen die Standardnormalverteilung für $n \to \infty$.

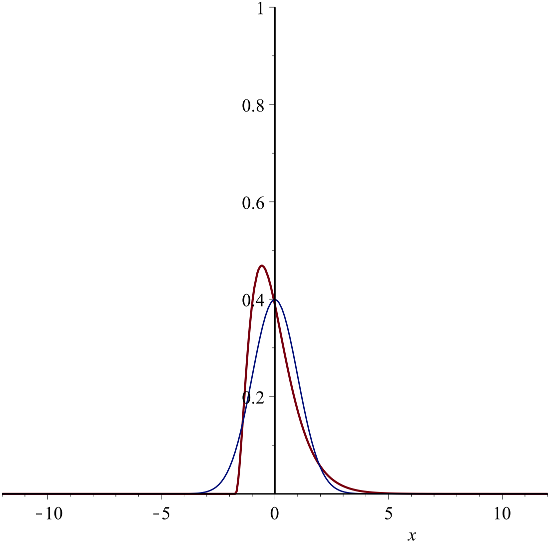
Exponential → Normal I



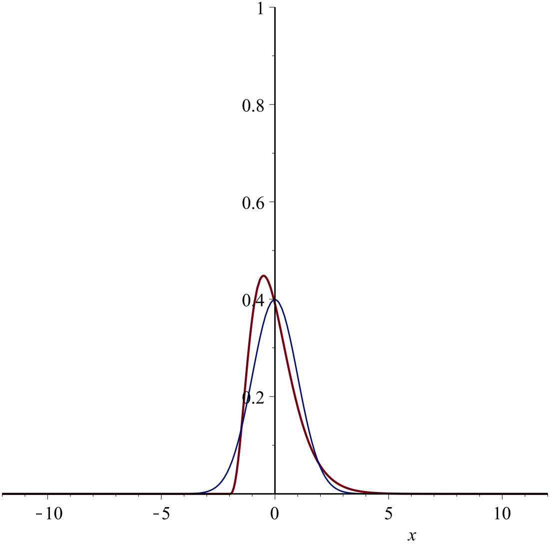
Exponential $\rightarrow$ Normal II



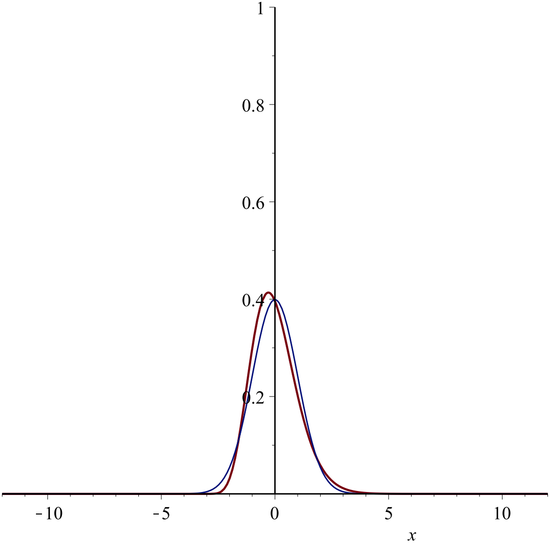
Exponential → Normal III



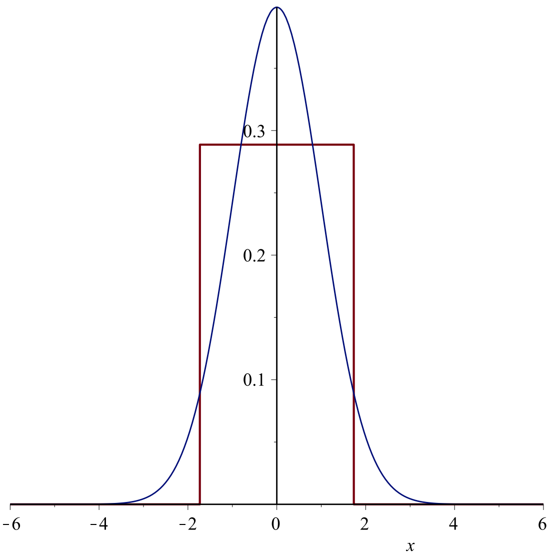
Exponential → Normal IV



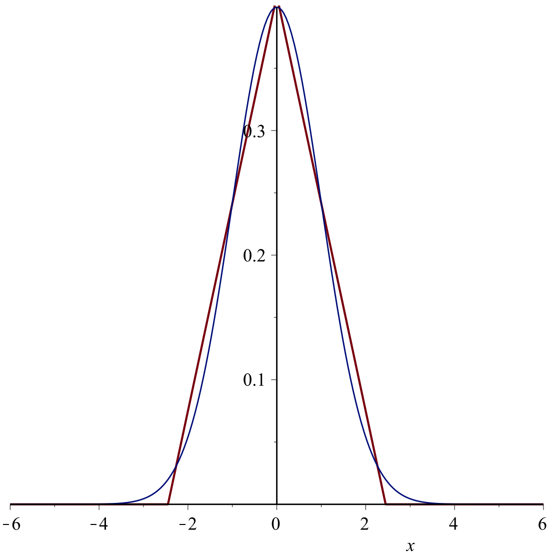
Exponential → Normal XII



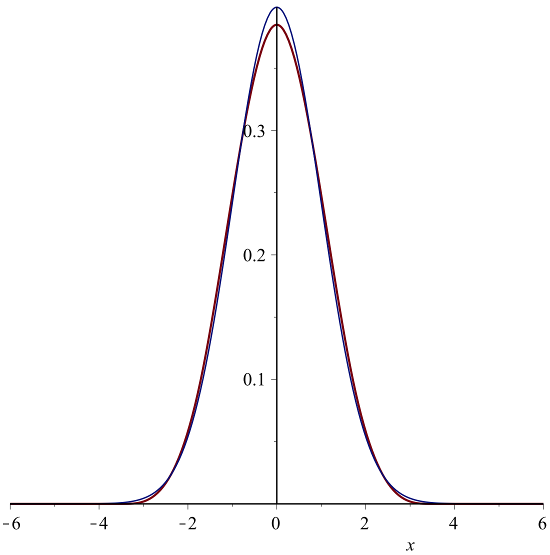
Gleich $\rightarrow$ Normal I



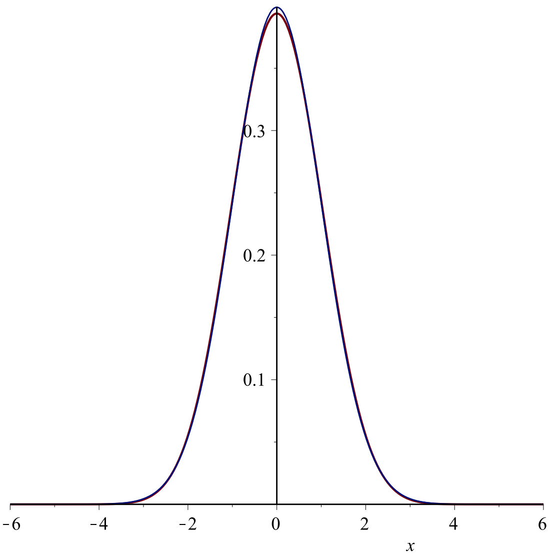
Gleich $\rightarrow$ Normal II



Gleich → Normal IV



Gleich $\to$ Normal XII



Zentraler Grenzwertsatz IV

Beweis:

Sei $X_i^* := (X_i - \mu)/\sigma$ für $i = 1, \ldots, n$.

Es gilt $\mathbb{E}[X_i^*] = 0$ und $\mathrm{Var}[X_i^*] = 1$.

Damit haben wir

$$
\begin{aligned}
M_Z(t) \quad &= \mathbb{E}[e^{tZ}] = \mathbb{E}[e^{t(X_1^* + \ldots + X_n^*)/\sqrt{n}}] \\
&= \mathbb{E}\left[e^{tX_1^*/\sqrt{n}} \cdot \ldots \cdot e^{tX_1^*/\sqrt{n}}\right] \\
&= \mathbb{E}\left[e^{tX_1^*/\sqrt{n}}\right] \cdot \ldots \cdot \mathbb{E}\left[e^{tX_1^*/\sqrt{n}}\right] && \text{(Unabh.)} \\
&= M_{X_1^*}(t/\sqrt{n}) \cdot \ldots \cdot M_{X_n^*}(t/\sqrt{n})\,. \\
&= \left(M_{X_1^*}(t/\sqrt{n})\right)^n && \text{(X_i's gleich verteilt)}
\end{aligned}
$$

Zentraler Grenzwertsatz V

Wir betrachten die Taylorentwicklung von $M_{X_i^*}(t) =: h(t)$ mit Entwicklungsstelle 0:

$$h(t) = h(0) + h'(0) \cdot t + \frac{h''(0)}{2} \cdot t^2 + \mathcal{O}(t^3).$$

Es gilt

$$\begin{aligned}
h'(t) &= \frac{\mathsf{d}}{\mathsf{d}\,t}\mathbb{E}[e^{tX_i^*}] = \frac{\mathsf{d}}{\mathsf{d}\,t}\mathbb{E}\left[\sum_{i=0}^{\infty} \frac{(tX_i^*)^n}{n!}\right] \\
&= \sum_{i=0}^{\infty} n \cdot t^{n-1}\mathbb{E}\left[\frac{(X_i^*)^n}{n!}\right] \qquad \text{(Lin. und Arith)} \\
&= \sum_{i=1}^{\infty} \mathbb{E}\left[X_i^* \frac{(tX_i^*)^{n-1}}{(n-1)!}\right] = \mathbb{E}\left[X_i^* \sum_{i=0}^{\infty} \frac{(tX_i^*)^n}{n!}\right] \\
&= \mathbb{E}[e^{tX_i^*} \cdot X_i^*]
\end{aligned}$$

Zentraler Grenzwertsatz VI

Analog erhalten wir $h''(t) = \mathbb{E}[e^{tX_i^*} \cdot (X_i^*)^2]$.

Damit gilt

$$h'(0) = \mathbb{E}[X_i^*] = 0 \qquad h''(0) = \mathbb{E}[(X_i^*)^2] = \mathrm{Var}[X] = 1\ .$$

Durch Einsetzen in die Taylorreihe folgt

$$h(t) = 1 + t^2/2 + \mathcal{O}(t^3)$$

und damit

$$\begin{aligned} M_Z(t) &= \left(M_{X_1^*}(t/\sqrt{n})\right)^n = \left(1 + \frac{t^2}{2n} + \mathcal{O}\left(\frac{t^3}{n^{3/2}}\right)\right)^n \\ &\to e^{t^2/2} \text{ für } n \to \infty. \end{aligned}$$

Aus der Konvergenz der momenterzeugenden Funktion folgt auch die Konvergenz der Verteilung. Damit ist Z asymptotisch normalverteilt.

Zentraler Grenzwertsatz VII

Die momenterzeugende Funktion existiert nicht bei allen Zufallsvariablen.

Der Beweis ist deshalb unvollständig.

Man umgeht dieses Problem, indem man statt der momenterzeugenden Funktion die charakteristische Funktion

$$\tilde{M}_X(t) = \mathbb{E}[e^{\mathrm{i}tX}]$$

betrachtet.

Der ZGWS und experimentelle Messungen I

Beispiel 135

Physical constants, National Institute of Standards and Technology (http://physics.nist.gov/constants) (June 2012)

Newtonian constant of gravitation G

Value	6.673 84 $\times 10^{-11}$ m^3 kg^{-1} s^{-2}
Standard uncertainty	0.000 80 $\times 10^{-11}$ m^3 kg^{-1} s^{-2}
Concise form	6.673 84(80) $\times 10^{-11}$ m^3 kg^{-1} s^{-2}

Was ist die "standard uncertainty" ?

Beispiel 136

New Scientist Physics & Math, 09.02.2012: The Higgs boson is the missing piece of the standard model of physics ... In December, the LHC's two main particle detectors, CMS and ATLAS, each reported excesses of events, such as the appearance of a pair of photons in the shrapnel from particle collisions. These excess events could be due to a Higgs with a mass of around 125 gigaelectron volts ... By convention, researchers only declare a discovery when an anomaly reaches a statistical significance known as 5 sigma, which means there is less than a 1-in-a-million chance it is just a fluke. The size of the anomalies reported in December was 1.9 sigma for CMS and 2.5 sigma for ATLAS, which indicate a probability of a fluke of roughly 1 per cent.

Wieso bedeutet „5 sigma“, dass die W'keit eines Zufallstreffers unter 10^{-6} liegt?

Der ZGWS und experimentelle Messungen III

Messfehler: Abweichung eines aus Messungen gewonnenen Wertes vom wahren Wert der Messgröße.

Systematische Fehler: Messfehler, die sich bei wiederholter Messung nicht im Mittel aufheben.
(Z.B. Fehler durch falsche Kalibrierung von Messgeräten.

Systematische Fehler werden durch sorgfältige Analyse und unabhängige Reproduktion des Experimentes beseitigt.

Zufallsfehler: Messfehler, die sich bei wiederholter Messung im Mittel aufheben.
(Z.B. Fehler durch nicht beherrschbare Einflüsse der Messgeräte oder nicht beherrschbare Änderungen des Wertes der Messgröße.)

Der ZGWS und experimentelle Messungen IV

Annahme: systematische Fehler sind beseitigt worden.

Zufallsfehler werden durch Wiederholung mit Hilfe des Zentralen Grenzwertsatzes beherrscht:

- Ergebnis des Experiments $\rightarrow$ Zufallsvariable X.
- $\mu := \mathbb{E}[X]$ ist der wahre Wert der Messgröße.
  (Folgt aus der Annahme).
- Experiment wird N-Mal wiederholt $\rightarrow$ Variablen $X_1, \ldots, X_N$, identisch verteilt.
- Die Verteilung von X ist nicht bekannt, aber aus dem ZGWS folgt: $Z_n \approx \mathcal{N}(0, 1)$ wenn $n \rightarrow \infty$.

Annahme: $Z_N \approx \mathcal{N}(0, 1)$.

Der ZGWS und experimentelle Messungen V

Sei σ_N^2 die Varianz von Y_N/N (σ_N die Standardabweichung).

Es gilt

$$\sigma_N = \mathrm{Var}\left[\tfrac{Y_N}{N}\right] = \mathrm{Var}\left[\tfrac{X_1+\cdots+X_N}{N}\right] = \tfrac{1}{N^2}\left(N \cdot \mathrm{Var}[X]\right) = \frac{\sigma}{\sqrt{N}}$$

$$\begin{aligned}
\Phi(k) - \Phi(-k) &= \Pr[-k \le Z_N \le k] \\
&= \Pr\left[\mu - k\frac{\sigma}{\sqrt{N}} \le \frac{Y_N}{N} \le \mu + k\frac{\sigma}{\sqrt{N}}\right] \\
&= \Pr\left[\mu - k\sigma_N \le \frac{Y_N}{N} \le \mu + k\sigma_N\right]
\end{aligned}$$

Die W'keit, dass Y_N/N (der geschätzte Wert) mehr als k Standardabweichungen vom wahren Wert liegt, beträgt höchstens $\Phi(k) - \Phi(-k)$.

Der ZGWS und experimentelle Messungen VI

Die W'keit, dass Y_N/N *mehr als* k *Standardabweichungen vom wahren Wert liegt, beträgt höchstens* $\Phi(k) - \Phi(-k)$.

k	$\Phi(k) - \Phi(-k)$
1	0.6826895
2	0.9544997
3	0.9973003
4	0.9999367
5	0.9999995

Der ZGWS und experimentelle Messungen I

Beispiel 137

Physical constants, National Institute of Standards and Technology (http://physics.nist.gov/constants) (June 2012)

Newtonian constant of gravitation G

Value	6.673 84 $\times$ 10^{-11} m^3 kg^{-1} s^{-2}
Standard uncertainty	0.000 80 $\times$ 10^{-11} m^3 kg^{-1} s^{-2}
Concise form	6.673 84(80) $\times$ 10^{-11} m^3 kg^{-1} s^{-2}

Was ist die “standard uncertainty” ?

Die standard uncertainty ist (im Wesentlichen) σ_N.

Beispiel 138

New Scientist Physics & Math, 09.02.2012: In December, the LHC's two main particle detectors, CMS and ATLAS, each reported excesses of events, such as the appearance of a pair of photons in the shrapnel from particle collisions. ... By convention, researchers only declare a discovery when an anomaly reaches a statistical significance known as 5 sigma, which means there is less than a 1-in-a-million chance it is just a fluke. The size of the anomalies reported in December was 1.9 sigma for CMS and 2.5 sigma for ATLAS, which indicate a probability of a fluke of roughly 1 per cent.

Wieso bedeutet „5 sigma“, dass die W'keit eines Zufallstreffers unter 10^{-6} liegt?

Weil $\Pr[-5 \leq Z_N \leq 5] \approx 10^{-6}$ wenn $Z_N \sim \mathcal{N}(0,1)$.

Korollar 139 (Grenzwertsatz von de Moivre)

$X_1, \ldots, X_n$ seien unabhängige Bernoulli-verteilte Zufallsvariablen mit gleicher Erfolgsw'keit p. Dann gilt für $H_n := X_1 + \ldots + X_n$ und $n \geq 1$, dass die Verteilung der Zufallsvariablen

$$H_n^* := \frac{H_n - np}{\sqrt{np(1-p)}}$$

für $n \to \infty$ gegen die Standardnormalverteilung konvergiert.

Beweis:

Wenn $X_i \sim Ber(p)$ dann $\mu = \mathbb{E}[X_i] = p$, $\sigma^2 = \mathrm{Var}[X_i] = pq$.
Einsetzen im ZGWS ergibt $Z_n = H_n^*$ und so $H_n^* \sim \mathcal{N}(0,1)$. $\square$

Grenzwertsatz von de Moivre II

Historisch gesehen entstand Korollar 139 vor Satz 134.

Für den Fall $p = 1/2$ wurde Korollar 139 bereits von Abraham de Moivre (1667–1754) bewiesen. De Moivre war gebürtiger Franzose, musste jedoch aufgrund seines protestantischen Glaubens nach England fliehen. Dort wurde er unter anderem Mitglied der Royal Society, erhielt jedoch niemals eine eigene Professur.

Die allgemeine Formulierung von Korollar 139 geht auf Pierre Simon Laplace (1749–1827) zurück. Allerdings vermutet man, dass die Lösung des allgemeinen Falls $p \neq 1/2$ bereits de Moivre bekannt war.

Elementarer Beweis des GWSdM für $p = 1/2$ I

Wir schreiben $f(n) \sim_\infty g(n)$ für $\lim_{n\to\infty} f(n)/g(n) = 1$.

Zu zeigen: $\Pr[a \le H_{2n}^* \le b] \sim_\infty \int_a^b \varphi(t)\,\mathrm{d}t$.

Mit $\mathbb{E}[H_{2n}] = n$ und $\mathrm{Var}[H_{2n}] = n/2$ ($p = 1/2$!) erhalten wir

$$H_{2n}^* = \frac{H_{2n} - n}{\sqrt{n/2}}$$

Es folgt

$$\begin{aligned}
& \Pr[a \le H_{2n}^* \le b] \\
= & \sum_{i=\alpha}^{\beta} \Pr[H_{2n} = n + i] \quad \alpha = \left\lceil a\sqrt{n/2} \right\rceil, \beta = \left\lfloor b\sqrt{n/2} \right\rfloor \\
= & \sum_{i=\alpha}^{\beta} \binom{2n}{n+i} \cdot \left(\frac{1}{2}\right)^{2n} =: \sum_{i=\alpha}^{\beta} p_{n,i}
\end{aligned}$$

Elementarer Beweis des GWSdM für $p = 1/2$ II

Es gilt

$$\max_i p_{n,i} \leq p_n^* := \binom{2n}{n} \cdot \left(\frac{1}{2}\right)^{2n} = \frac{(2n)!}{(n!)^2} \cdot \left(\frac{1}{2}\right)^{2n}$$

Wir setzen

$$q_{n,i} := \frac{p_{n,i}}{p_n^*}$$

d.h.

$$\Pr[a \leq H_{2n}^* \leq b] = \sum_{i=\alpha}^{\beta} p_n^* \cdot q_{n,i}$$

Wir schätzen erst p_n^*, dann $q_{n,i}$ und zuletzt $\Pr[a \leq H_{2n}^* \leq b]$ ab.

Elementarer Beweis des GWSdM für $p = 1/2$ III

Abschätzung von p_n^*:

Mit der Stirling'schen Approximation für $n!$ gilt

$$p_n^* \sim_\infty \frac{(2n)^{2n} \cdot e^{-2n} \cdot \sqrt{2\pi \cdot 2n}}{(n^n \cdot e^{-n} \cdot \sqrt{2\pi n})^2} \cdot \left(\frac{1}{2}\right)^{2n} = \frac{1}{\sqrt{\pi n}}.$$

Elementarer Beweis des GWSdM für $p = 1/2$ IV

Abschätzung von $q_{n,i}$:

Für $i > 0$ gilt

$$q_{n,i} = \frac{\binom{2n}{n+i} \cdot \left(\frac{1}{2}\right)^{2n}}{\binom{2n}{n} \cdot \left(\frac{1}{2}\right)^{2n}} = \frac{(2n)! \cdot n! \cdot n!}{(n+i)! \cdot (n-i)! \cdot (2n)!}$$

$$= \frac{\prod_{j=0}^{i-1}(n-j)}{\prod_{j=1}^{i}(n+j)} = \prod_{j=1}^{i} \frac{n-j+1}{n+j} = \prod_{j=1}^{i} \left(1 - \frac{2j-1}{n+j}\right).$$

Für $i < 0$ gilt $q_{n,-i} = q_{n,i}$.

Wir schätzen $\ln \prod_{j=1}^{i} \left(1 - \frac{2j-1}{n+j}\right)$ ab.

Elementarer Beweis des GWSdM für $p = 1/2$ V

Mit $1 - 1/x \leq \ln x \leq x - 1$ für $x > 0$ gilt

$$
\begin{aligned}
\ln\left(\prod_{j=1}^{i}\left(1 - \frac{2j-1}{n+j}\right)\right) &= \sum_{j=1}^{i} \ln\left(1 - \frac{2j-1}{n+j}\right) \\
&\leq -\sum_{j=1}^{i} \frac{2j-1}{n+j} \;\leq\; -\sum_{j=1}^{i} \frac{2j-1}{n+i} \\
&= -\frac{i(i+1)-i}{n+i} \;=\; -\frac{i^2}{n} + \frac{i^3}{n(n+i)} \\
&= -\frac{i^2}{n} + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right),
\end{aligned}
$$

da $i = \mathcal{O}(\sqrt{n})$ für $\alpha \leq i \leq \beta$.

Elementarer Beweis des GWSdM für $p = 1/2$ VI

Ebenso erhalten wir

$$\ln\left(\prod_{j=1}^{i}\left(1-\frac{2j-1}{n+j}\right)\right) \geq \sum_{j=1}^{i}\left(1-\left(1-\frac{2j-1}{n+j}\right)^{-1}\right)$$

$$= \sum_{j=1}^{i}\frac{-2j+1}{n-j+1} \geq -\sum_{j=1}^{i}\frac{2j-1}{n-i}$$

$$= -\frac{i^2}{n-i} \;=\; -\frac{i^2}{n} - \mathcal{O}\left(\frac{1}{\sqrt{n}}\right) .$$

Zusammen gilt

$$e^{-\frac{i^2}{n-i} = -\frac{i^2}{n} - \mathcal{O}\left(\frac{1}{\sqrt{n}}\right)} \leq q_{n,i} \leq e^{-\frac{i^2}{n} + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right)}$$

Wegen $e^{\pm\mathcal{O}(1/\sqrt{n})} = 1 \pm o(1)$ folgt daraus

$$q_{n,i} \;\sim_\infty\; e^{-i^2/n}$$

Elementarer Beweis des GWSdM für $p = 1/2$ VII

Abschätzung von $\Pr[a \le H_{2n}^* \le b] = \sum_{i=\alpha}^{\beta} p_n^* \cdot q_{n,i}$:

$$\Pr[a \le H_{2n}^* \le b] \;\sim_\infty\; \frac{1}{\sqrt{\pi n}} \cdot \sum_{i=\alpha}^{\beta} e^{-i^2/n}.$$

Elementarer Beweis des GWSdM für $p = 1/2$ VIII

Letzter Schritt:

Mit $\delta := \sqrt{2/n}$ erhalten wir

$$\Pr[a \leq H_{2n}^* \leq b] \ \sim_\infty \ \frac{1}{\sqrt{2\pi}} \cdot \sum_{i=\alpha}^{\beta} \delta e^{-(i\delta)^2 \cdot \frac{1}{2}}.$$

Die rechte Seite entspricht einer Näherung für

$$\int_a^b \varphi(t)\,\mathrm{d}t = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-t^2/2}\,\mathrm{d}t$$

durch Aufteilung der integrierten Fläche in Balken der Breite δ.

Für $n \to \infty$ konvergiert die Fläche der Balken gegen das Integral und wir erhalten

$$\Pr[a \leq H_{2n}^* \leq b] \ \sim_\infty \ \int_a^b \varphi(t)\,\mathrm{d}t$$

q. e. d.

19. Die Exponentialverteilung

Die Exponentialverteilung I

Beispiel 140

Beim Kundenservice einer Firma rufen im Durschnitt k Kunden pro Tag an.

Wir betrachten ein diskretes Modell, in dem der Tag in $n \geq k$ gleich lange Zeitintervallen unterteilt wird (jeweils $24/n$ Stunden lang). Wir nehmen an, dass Kunden in jedem Intervall mit der selben W'keit anrufen und dass höchstens ein Kunde pro Zeitintervall anruft.

Damit ruft ein Kunde in ein Intervall mit W'keit $p_n = \dfrac{k}{n}$ an. Die Zeit X_n bis zum ersten Anruf ist geometrisch verteilt:

$$\Pr[X_n \leq a] = \sum_{i=1}^{a} (1 - p_n)^{(i-1)} \cdot p_n = 1 - (1 - p_n)^a$$

Die Exponentialverteilung II

Wir berechnen die W'keit, dass für $k = 3$ der erste Anruf in der ersten Hälfte des Tages stattfindet.

Bei einer Einteilung in n Intervallen ist diese W'keit gleich $\Pr[X_n \leq n/2]$.

Die folgende Tabelle zeigt $\Pr[X_n \leq n/2]$ für $k = 3$ und verschiedene Werte von n:

n	$\Pr[X \leq n/2]$
6	0.8750
12	0.8221
24	0.7986
48	0.7875
$24 * 60$	0.7772

Frage: Zu welchem Wert konvergiert diese Folge?

Die Exponentialverteilung III

Definition 141

Eine Zufallsvariable X heißt exponentialverteilt mit dem Parameter $\lambda > 0$, wenn sie die Dichte

$$f(x) = \begin{cases} \lambda \cdot e^{-\lambda x} & \text{falls } x \geq 0, \\ 0 & \text{sonst} \end{cases}$$

besitzt.

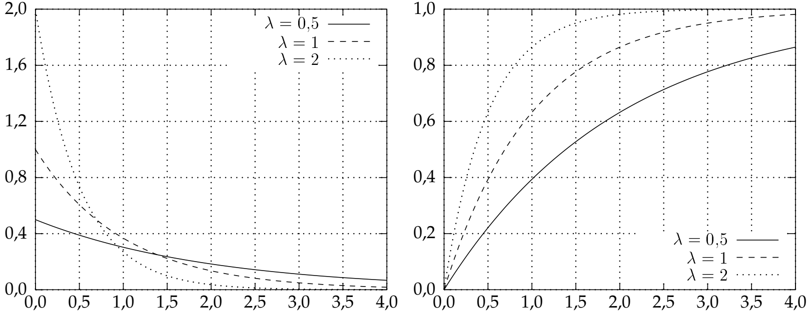
Wir schreiben auch $X \sim Exp(\lambda)$.

Für die Verteilungsfunktion gilt für $x \geq 0$

$$F(x) = \int_0^x \lambda \cdot e^{-\lambda t} \, \mathrm{d}t = \left[-e^{-\lambda t} \right]_0^x = 1 - e^{-\lambda x}.$$

sowie $F(x) = 0$ für $x < 0$.

Die Exponentialverteilung IV



Dichte und Verteilung der Exponentialverteilung

Satz 142

Sei $\lambda > 0$ und sei $X_n \sim Geo(\lambda/n)$ eine Folge von Zufallsvariablen. Die Folge $Y_n = X_n/n$ ist asymptotisch exponentialverteilt mit Parameter λ, d.h. $Y_n \sim Exp(\lambda)$ für $n \to \infty$.

Vergleiche mit: $\mathrm{Po}(\lambda) = \lim_{n\to\infty} \mathrm{Bin}(n, \lambda/n)$.

In unserem Beispiel: $\lambda = 3$ und

$$\Pr[X \leq 1/2] = F(1/2) = 1 - e^{-3/2} \approx 0.7768\ .$$

Die Exponentialverteilung VI

Beweis: Für ein beliebiges $k \in \mathbb{N}$ gilt

$$\begin{aligned}\Pr[X_n \le kn] &= \sum_{i=1}^{kn} (1-p_n)^{i-1} \cdot p_n \;=\; p_n \cdot \sum_{i=0}^{kn-1} (1-p_n)^i \\ &= p_n \cdot \frac{1-(1-p_n)^{kn}}{p_n} \;=\; 1 - \left(1 - \frac{\lambda}{n}\right)^{kn}.\end{aligned}$$

und damit

$$\begin{aligned}\lim_{n\to\infty} \Pr[Y_n \le t] &= \lim_{n\to\infty} \Pr[X_n \le t \cdot n] = \lim_{n\to\infty} \left[1 - \left(1 - \frac{\lambda}{n}\right)^{tn}\right] \\ &= 1 - e^{-\lambda t}.\end{aligned}$$

Die Folge Y_n geht also für $n \to \infty$ in eine exponentialverteilte Zufallsvariable mit Parameter λ über.

Die Exponentialverteilung VII: Erwartungswert

$$\begin{aligned}\mathbb{E}[X] &= \int_0^\infty t \cdot \lambda \cdot e^{-\lambda t}\,\mathsf{d}t \\ &= \Big[t \cdot (-e^{-\lambda t})\Big]_0^\infty + \int_0^\infty e^{-\lambda t}\,\mathsf{d}t \\ &= 0 + \left[-\frac{1}{\lambda} \cdot e^{-\lambda t}\right]_0^\infty = \frac{1}{\lambda}\,.\end{aligned}$$

Die Exponentialverteilung VIII: Varianz

Analog erhalten wir

$$\begin{aligned}\mathbb{E}[X^2] &= \int_0^\infty t^2 \cdot \lambda \cdot e^{-\lambda t} \,\mathsf{d}t \\ &= \Big[t^2 \cdot (-e^{-\lambda t})\Big]_0^\infty + \int_0^\infty 2t \cdot e^{-\lambda t} \,\mathsf{d}t \\ &= 0 + \frac{2}{\lambda} \cdot \mathbb{E}[X] = \frac{2}{\lambda^2}\end{aligned}$$

und somit

$$\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{1}{\lambda^2}\,.$$

Die Exponentialverteilung: Gedächtnislosigkeit I

Lemma 143 (Skalierung exponentialverteilter Variablen)

Wenn $X \sim Exp(\lambda)$ *und* $Y := aX$ *für* $a > 0$ *dann* $Y \sim Exp(\lambda/a)$.

Beweis:

$$\begin{aligned} F_Y(x) &= \Pr[Y \le x] = \Pr[aX \le x] \\ &= \Pr\left[X \le \frac{x}{a}\right] = F_X\left(\frac{x}{a}\right) \\ &= 1 - e^{-\lambda x/a}\,. \end{aligned}$$

□

Die Exponentialverteilung: Gedächtnislosigkeit II

Satz 144 (Gedächtnislosigkeit)

Eine kontinuierliche Zufallsvariable X mit Wertebereich $\mathbb{R}^+$ ist genau dann exponentialverteilt, wenn für alle $x, y > 0$ gilt, dass

$$\Pr[X > x + y \mid X > y] = \Pr[X > x]\,. \qquad (*)$$

Beweis:

Sei X exponentialverteilt mit Parameter λ. Dann gilt

$$\begin{aligned}
\Pr[X > x + y \mid X > y] &= \frac{\Pr[X > x + y, X > y]}{\Pr[X > y]} \\
&= \frac{\Pr[X > x + y]}{\Pr[X > y]} \\
&= \frac{e^{-\lambda(x+y)}}{e^{-\lambda y}} = e^{-\lambda x} = \Pr[X > x]\,.
\end{aligned}$$

Die Exponentialverteilung: Gedächtnislosigkeit III

Beweis (Forts.):

Sei umgekehrt X eine kontinuierliche Zufallsvariable, die die Gleichung $(*)$ erfüllt. Definiere $g(x) := \Pr[X > x]$.

Wir zeigen $g(x) = e^{-\lambda x}$ für alle $x \in \mathbb{Q}^+$.

Aufgrund der Stetigkeit folgt daraus $g(x) = e^{-\lambda x}$ für alle $x \in \mathbb{R}^+$.

Aus (*) folgt

$$\begin{aligned} g(x+y) &= \Pr[X > x+y] \\ &= \Pr[X > x+y \mid X > y] \cdot \Pr[X > y] \\ &= \Pr[X > x] \cdot \Pr[X > y] = g(x)g(y)\,. \end{aligned}$$

Die Exponentialverteilung: Gedächtnislosigkeit IV

Beweis (Forts.):

Durch wiederholte Anwendung erhalten wir

$$g(1) = g\Big(\underbrace{\frac{1}{n} + \cdots + \frac{1}{n}}_{n\text{-mal}}\Big) = \left(g\Big(\frac{1}{n}\Big)\right)^n \text{ für alle } n \in \mathbb{N}$$

und somit insbesondere auch $g(1/n) = (g(1))^{1/n}$.

Da X positiv, gibt es $n \in \mathbb{N}$ mit $g(1/n) > 0$.

Wegen $0 < g(1) \leq 1$ gibt es auch ein $\lambda \geq 0$ mit $g(1) = e^{-\lambda}$.

Nun gilt für beliebige $p, q \in \mathbb{N}$

$$g(p/q) = g(1/q)^p = g(1)^{p/q}$$

□

Beispiel 145

Das Cäsium-Isotop ${}^{134}_{55}\mathrm{Cs}$ besitzt eine mittlere Lebensdauer von ungefähr 3,03 Jahren oder $1{,}55 \cdot 10^6$ Minuten.

Die Zufallsvariable X messe die Lebenszeit eines bestimmten ${}^{134}_{55}\mathrm{Cs}$-Atoms.

X ist exponentialverteilt mit dem Parameter

$$\lambda = \frac{1}{\mathbb{E}[X]} = \frac{1}{1{,}55 \cdot 10^6} \approx 0{,}645 \cdot 10^{-6} \ \left[\frac{1}{\text{min}}\right]$$

Da λ den Kehrwert einer Zeit als Einheit besitzt, spricht man von der Zerfallsrate.

Auch bei anderen Anwendungen ist es üblich, λ als Rate einzuführen.

Minimum von Exponentialverteilungen I

Satz 146

Die Zufallsvariablen $X_1, \ldots, X_n$ *seien unabhängig und exponentialverteilt mit den Parametern* $\lambda_1, \ldots, \lambda_n$. *Dann ist auch* $X := \min\{X_1, \ldots, X_n\}$ *exponentialverteilt mit dem Parameter* $\lambda_1 + \ldots + \lambda_n$.

Beweis:

Der allgemeine Fall folgt mittels Induktion aus dem Fall $n = 2$. Für die Verteilungsfunktion F_X gilt:

$$\begin{aligned} 1 - F_X(t) &= \Pr[X > t] = \Pr[\min\{X_1, X_2\} > t] \\ &= \Pr[X_1 > t, X_2 > t] \\ &= \Pr[X_1 > t] \cdot \Pr[X_2 > t] \\ &= e^{-\lambda_1 t} \cdot e^{-\lambda_2 t} = e^{-(\lambda_1 + \lambda_2) t}. \end{aligned}$$

Minimum von Exponentialverteilungen II

Anschaulich besagt Satz 146:

> *Wenn man auf das erste Eintreten eines Ereignisses aus mehreren unabhängigen Ereignissen wartet, dann addieren sich die Raten.*

Wenn beispielsweise ein Atom die Zerfallsrate λ besitzt, so erhalten wir bei n Atomen die Zerfallsrate $n\lambda$.

Poisson-Prozess I

Wir wissen:

> *Wenn der zeitliche Abstand der Treffer geometrisch verteilt ist, so ist ihre Anzahl in einer festen Zeitspanne binomialverteilt.*

Im Grenzwert $n \to \infty$ mit Trefferw'keit $p_n = \lambda/n$ konvergiert

- die Binomialverteilung gegen die Poisson-Verteilung und
- die geometrische Verteilung gegen die Exponentialverteilung.

Im Grenzwert $n \to \infty$ erwarten wir deshalb:

> *Wenn man Ereignisse zählt, deren zeitlicher Abstand exponentialverteilt ist, so ist die Anzahl dieser Ereignisse in einer festen Zeitspanne Poisson-verteilt.*

Poisson-Prozess II

Seien $T_1, T_2 \ldots$ unabhängige Zufallsvariablen.

T_i modelliert die Zeit, die zwischen Treffer $i-1$ und i vergeht.

Für den Zeitpunkt $t > 0$ definieren wir

$$X(t) := \max\{n \in \mathbb{N} \mid T_1 + \ldots + T_n \leq t\}.$$

$X(t)$ modelliert die Anzahl der Treffer von Zeit 0 bis t.

Satz 147 (ohne Beweis)

$X(t) \sim \mathrm{Po}(t\lambda)$ *genau dann, wenn* $T_1, T_2, \ldots \sim Exp(\lambda)$.

Eine Familie $(X(t))_{t>0}$ von Zufallsvariablen nennt man allgemein einen stochastischen Prozess.

Der Prozess, bei dem $T_1, T_2, \ldots$ unabhängige, exponentialverteilte Zufallsvariablen sind, heißt Poisson-Prozess.

Poisson-Prozess III

Beispiel 148

Eine Menge von Jobs werden auf einem Prozessor sequentiell abgearbeitet.

Die Laufzeiten der Jobs seien unabhängig und exponentialverteilt mit Parameter $\lambda = 1/30[1/s]$.

Jeder Job benötigt also im Mittel $30s$.

Gemäß Satz 147 ist die Anzahl von Jobs, die in einer Minute vollständig ausgeführt werden, Poisson-verteilt mit Parameter $t\lambda = 60 \cdot (1/30) = 2$.

Die Wahrscheinlichkeit, dass in einer Minute höchstens ein Job abgearbeitet wird, beträgt für $t\lambda = 2$

$$e^{-t\lambda} + t\lambda e^{-t\lambda} \approx 0{,}406 \; .$$

Approximationen der Binomialverteilung I

Sei $H_n \sim \mathrm{Bin}(n, p)$ mit der Verteilungsfunktion F_n.

Für $n \to \infty$ gilt (Korollar 139)

$$F_n(t) = \Pr[H_n/n \le t/n] \longrightarrow \Phi\left(\frac{t/n - p}{\sqrt{p(1-p)/n}}\right) = \Phi\left(\frac{t - np}{\sqrt{p(1-p)n}}\right).$$

F_n kann somit für große n durch Φ approximieren.

Faustregel: Gute Approximation wenn $np \ge 5$ und $n(1-p) \ge 5$.

Für die Berechnung von Φ liegen effiziente numerische Methoden vor.

Beispiel 149

Die W'keit, mit der bei 10^6 Würfen mit einem idealen Würfel mehr als 500500-mal eine gerade Augenzahl fällt beträgt

$$T := \sum_{i=5{,}005\cdot 10^5}^{10^6} \binom{10^6}{i} \left(\frac{1}{2}\right)^{10^6} .$$

Die Approximation durch die Normalverteilung ergibt

$$\begin{aligned} T &\approx 1 - \Phi\left(\frac{5{,}005\cdot 10^5 - 5\cdot 10^5}{\sqrt{2{,}5\cdot 10^5}}\right) \\ &= 1 - \Phi\left(\frac{5\cdot 10^2}{5\cdot 10^2}\right) = 1 - \Phi(1) \approx 0{,}1573 . \end{aligned}$$

Approximationen der Binomialverteilung III

Oft führt man eine Stetigkeitskorrektur durch.

Zur Berechnung von $\Pr[X \leq x]$ für $X \sim \mathrm{Bin}(n, p)$ setzt man

$$\Pr[X \leq x] \approx \Phi\left(\frac{x + 0{,}5 - np}{\sqrt{np(1-p)}}\right)$$

statt

$$\Pr[X \leq x] \approx \Phi\left(\frac{x - np}{\sqrt{np(1-p)}}\right)$$

an.

Approximationen der Binomialverteilung IV

Der Korrekturterm läßt sich in der Histogramm-Darstellung der Binomialverteilung veranschaulichen.

Die Binomialverteilung wird durch Balken angegeben, deren Fläche in etwa der Fläche unterhalb der Dichte φ von $\mathcal{N}(0,1)$ entspricht.

Wenn man die Fläche der Balken mit „$X \leq x$“ durch das Integral von φ approximieren will, soll man bis zum Ende des Balkens für „$X = x$“ integrieren (nicht nur bis zur Mitte).

Dafür sorgt der Korrekturterm $0{,}5$.

Approximationen der Binomialverteilung V

Zusammenfassung:

- Approximation durch die Normalverteilung.

  $\mathrm{Bin}(n,p)$ wird approximiert durch $\Phi((t-np)/\sqrt{p(1-p)n})$.

  Gut für grosses n.

  Faustregel: Gut wenn $np \geq 5$ und $n(1-p) \geq 5$.

- Approximation durch die Poisson-Verteilung.

  $\mathrm{Bin}(n,p)$ wird approximiert durch $\mathrm{Po}(np)$.

  Gut für seltene Ereignisse, d. h. wenn $np << n$.

  Faustregel: gut wenn $n \geq 30$ und $p \leq 0{,}05$.

Teil V

Induktive Statistik

Induktive Statistik

Das Ziel der induktiven Statistik besteht darin,

- aus gemessenen Zufallsgrößen (z.B. Häufigkeiten)
- auf die zugrunde liegenden Gesetzmäßigkeiten (z.B. W'keiten)

zu schließen.

20. Schätzvariablen

Schätzvariablen I: Einführendes Beispiel

Sei X die Anzahl von Lesezugriffen auf eine Festplatte bis zum ersten Lesefehler.

Wir nehmen $X \sim Geo(p)$ an mit unbekannter Erfolgsw'keit p. Die W'keit p soll empirisch geschätzt werden.

Wegen $p = \dfrac{1}{\mathbb{E}[X]}$ schätzen wir $\mathbb{E}[X]$.

Wir führen n Messungen der Anzahl von Zugriffen bis zum ersten Lesefehler. Sei X_i die Zufallsvariable für das Ergebnis der i-ten Messung.

$X_1, \ldots, X_n$ sind unabhängig und besitzen jeweils dieselbe Verteilung wie X.

Die Variablen $X_1, \ldots, X_n$ nennt man Stichprobenvariablen.

Schätzvariablen II: Einführendes Beispiel

Sei $\overline{X}$ das arithmetische Mittel der X_i, d.h. $\overline{X} := \frac{1}{n} \sum_{i=1}^{n} X_i$.

Wir schätzen $\mathbb{E}[X]$ durch $\overline{X}$.

Wir nennen $\overline{X}$ einen Schätzer für den Erwartungswert $\mathbb{E}[X]$.

Definition 150

Sei X eine Zufallsvariable mit Dichte $f_X(x) = f(x;\theta)$.

Eine Schätzvariable oder kurz Schätzer für θ ist eine Zufallsvariable U der Gestalt $U = f(x_1, \ldots, X_n)$, wobei die $X_1, \ldots, X_n$ unabhängige Stichprobenvariablen sind mit derselben Verteilung wie X.

Frage: Welche Eigenschaften soll eine Schätzvariable erfüllen?

Definition 151

Ein Schätzer U eines Parameters θ heißt erwartungstreu, wenn $\mathbb{E}[U] = \theta$ gilt.

Die Größe $\mathbb{E}[U - \theta]$ nennt man Bias der Schätzvariablen U.

Satz 152

Sei X eine beliebige Zufallsvariable. $\overline{X}$ is ein erwartungstreuer Schätzer für $\mathbb{E}[X]$.

Beweis:

$$\mathbb{E}[\overline{X}] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X] = \mathbb{E}[X].$$

$\square$

Schätzvariablen IV: Mittlere quadratische Abweichung

Definition 153

Sei U ein Schätzer eines Parameters θ. Die mittlere quadratische Abweichung von U ist

$$MSE := \mathbb{E}[(U - \theta)^2] \, .$$

Ist U erwartungstreu, dann $MSE = \mathbb{E}[(U - \mathbb{E}[U])^2] = \mathrm{Var}[U]$.

Eine Schätzvariable heißt konsistent im quadratischen Mittel, wenn $MSE \to 0$ für $n \to \infty$ gilt (mit n Umfang der Stichprobe).

Schätzvariablen V: Mittlere quadratische Abweichung

Satz 154

Sei X eine Zufallsvariable mit endlicher Varianz. Der Schätzer $\overline{X}$ für $\mathbb{E}[X]$ ist konsistent im quadratischen Mittel.

Beweis:

Wegen der Unabhängigkeit von $X_1, \ldots, X_n$ gilt

$$MSE = \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}[X_i] = \frac{1}{n}\mathrm{Var}[X].$$

und damit $MSE \to 0$ für $n \to \infty$. $\square$

Schätzvariablen VI: Schwache Konsistenz

Definition 155

Ein Schätzer U eines Parameters θ ist schwach konsistent wenn

$$\lim_{n\to\infty} \Pr[|U-\theta| \geq \varepsilon] = 0$$

Satz 156

Ist ein Schätzer konsistent im quadratischen Mittel, dann ist er auch schwach konsistent. Insbesondere ist $\overline{X}$ ein schwach konsistenter Schätzer für X.

Beweis:

Mit der Ungleichung von Chebyshev gilt für ein beliebiges, festes $\varepsilon > 0$:

$$\Pr[|\overline{X}-\theta| \geq \varepsilon] = \Pr[|\overline{X}-\mathbb{E}[X]| \geq \varepsilon] \leq \frac{\operatorname{Var}[\overline{X}]}{\varepsilon^2} \to 0 \text{ für } n \to \infty$$

$\square$

Schätzvariablen VII: Schätzung der Varianz

Das Ergebnis einer wiederholten Messung wird präsentiert als

$$\overline{X} \pm S$$

wobei S^2 ein Schätzer für die Varianz von X darstellt.

Es liegt nah,

$$S_n^2 := \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

als Schätzer zu verwenden.

S_n^2 ist jedoch keinen erwartungstreuen Schätzer für die Varianz!

Schätzvariablen VIII: Schätzung der Varianz

Satz 157

Sei X eine Zufallsvariable und sei

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

Die Variable S_{n-1}^2 ist ein erwartungstreuer Schätzer für $\mathrm{Var}[X]$.

Schätzvariablen IX: Schätzung der Varianz

Beweis: Sei $\mu := \mathbb{E}[X] = \mathbb{E}[X_i] = \mathbb{E}[\overline{X}]$.

$$
\begin{aligned}
(X_i - \overline{X})^2 &= (X_i - \mu + \mu - \overline{X})^2 \\
&= (X_i - \mu)^2 + (\mu - \overline{X})^2 + 2(X_i - \mu)(\mu - \overline{X}) \\
&= (X_i - \mu)^2 + (\mu - \overline{X})^2 - \frac{2}{n}\sum_{j=1}^{n}(X_i - \mu)(X_j - \mu) \\
&= \frac{n-2}{n}(X_i - \mu)^2 + (\mu - \overline{X})^2 - \frac{2}{n}\sum_{j \neq i}(X_i - \mu)(X_j - \mu).
\end{aligned}
$$

Für je zwei unabhängige Zufallsvariablen X_i, X_j mit $i \neq j$ gilt

$$
\mathbb{E}[(X_i - \mu)(X_j - \mu)] = \mathbb{E}[X_i - \mu] \cdot \mathbb{E}[X_j - \mu] \;=\; 0 \cdot 0 \;=\; 0.
$$

Schätzvariablen X: Schätzung der Varianz

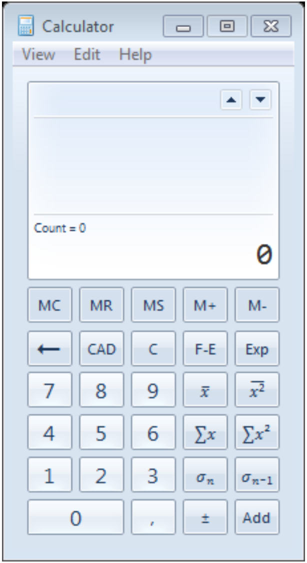
Daraus folgt

$$\begin{aligned}
\mathbb{E}[(X_i - \overline{X})^2] &= \frac{n-2}{n} \cdot \mathbb{E}[(X_i - \mu)^2] + \mathbb{E}[(\mu - \overline{X})^2] \\
&= \frac{n-2}{n} \cdot \mathrm{Var}[X_i] + \mathrm{Var}[\overline{X}] \\
&= \frac{n-2}{n} \cdot \frac{1}{n} \cdot \mathrm{Var}[X] + \mathrm{Var}[\overline{X}] = \frac{n-1}{n} \cdot \mathrm{Var}[X]
\end{aligned}$$

und somit gilt für S_{n-1}^2

$$\begin{aligned}
\mathbb{E}[S_{n-1}^2] &= \frac{1}{n-1} \sum_{i=1}^{n} \mathbb{E}[(X_i - \overline{X})^2] \\
&= \frac{1}{n-1} \cdot n \cdot \frac{n-1}{n} \cdot \mathrm{Var}[X] = \mathrm{Var}[X].
\end{aligned}$$

Schätzvariablen XI: Schätzung der Varianz



Definition 158

Die Zufallsvariablen

$$\overline{X} := \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{und} \quad S_{n-1}^2 := \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

heißen Stichprobenmittel bzw. Stichprobenvarianz der Stichprobe $X_1, \ldots, X_n$.

$\overline{X}$ und S_{n-1}^2 sind erwartungstreue Schätzer für den Erwartungswert bzw. die Varianz von X.

Oft wird S^2 statt S_{n-1}^2 geschrieben.

Schätzvariablen XIII

Beispiel 159

Ein Program läuft 10 Mal mit demselben Input mit Laufzeiten (ms)

2582	2580	2568	2576	2582
2581	2623	2616	2622	2617

Damit gilt für (die Realisierungen von) $\overline{X}$ und S_{n-1}

$$\overline{X} \approx 2594,7 \qquad S_{n-1} \approx 21,8$$

und wir schreiben als Ergebnis: $2594.7 \pm 21.8.$

Unter den Annahmen der Normalverteilung beträgt die W'keit, dass die Laufzeit einer Ausführung ausserhalb des Intervalls

$$[2594.7 - 2 \cdot 21.8\ ,\ 2594.7 + 2 \cdot 21.8] = [2551.1\ ,\ 2638.3]$$

liegt, höchstens 5% .

21. Das Maximum-Likelihood-Prinzip

Maximum-Likelihood-Prinzip I

Das Maximum-Likelihood-Prinzip ist ein allgemeines Prinzip zur Bestimmung des Parameters einer W'eitsverteilung.

> *Wähle den Wert des Parameters, für den die beobachtete Stichprobe maximale W'keit hat.*

Beispiel 160

Folgendes Experiment wird 10 mal wiederholt: Eine Münze wird bis zum ersten „Kopf“ geworfen.

Die Münze wird jeweils $4, 4, 7, 2, 4, 6, 5, 3, 5, 2$ Mal geworfen.

Frage: Welche ist die Erfolgsw'keit der Münze ?

Sei X die Anzahl der Würfe. Es gilt $X \sim Geo(p)$.

Nach dem Prinzip sollen wir p so wählen, das die W'keit einer Beobachtung $4, 4, 7, 2, 4, 6, 5, 3, 5, 2$ maximiert wird.

Maximum-Likelihood-Prinzip II: Formalisierung

Sei X eine Zufallsvariable mit $\Pr[X = x] = f(x;\theta)$ für eine bekannte Funktion f mit einem unbekannten Parameter θ.

Sei $X_1, \ldots, X_n$ Stichprobenvariablen mit Dichte $f(x;\theta)$.

Eine Stichprobe liefert für jede X_i einen Wert x_i und wir schreiben $\vec{x} = (x_1, \ldots, x_n)$.

Für eine feste Stichprobe $\vec{x}$ gibt die Likelihood-Funktion

$$\begin{aligned} L(\vec{x};\theta) := \prod_{i=1}^{n} f(x_i;\theta) &= \prod_{i=1}^{n} \Pr_\theta[X_i = x_i] \\ &\stackrel{\text{unabh.}}{=} \Pr_\theta[X_1 = x_1, \ldots, X_n = x_n] \end{aligned}$$

die Wahrscheinlichkeit an, dass man die Stichprobe $\vec{x}$ erhält, wenn der Parameter den Wert θ hat.

Maximum-Likelihood-Prinzip IV: Formalisierung

Definition 161

Ein Schätzwert $\widehat{\theta}$ für den Parameter einer Verteilung $f(x;\theta)$ heißt Maximum-Likelihood-Schätzwert (ML-Schätzwert) für eine Stichprobe $\vec{x}$, wenn gilt

$$L(\vec{x};\theta) \leq L(\vec{x};\widehat{\theta}) \text{ für alle } \theta.$$

Wenn $L(\vec{x};\theta)$ differenzierbar ist, dann kann ein Maximum von $L(\vec{x};\theta)$ mit Hilfe der üblichen Methode berechnet werden:

- Berechne $L'(\vec{x};\theta) := \frac{\mathrm{d}\, L(\vec{x};\theta)}{\mathrm{d}\,\theta}$.
- Finde eine Lösung θ_0 der Gleichung $L'(\vec{x};\theta) = 0$ mit $L''(\vec{x};\theta_0) > 0$.

Maximum-Likelihood-Prinzip V: Beispiele

Beispiel 162

Wir konstruieren mit der ML-Methode einen Schätzer für den Parameter p der Bernoulli-Verteilung.

Mit $\Pr_p[X_i = 1] = p$ und $\Pr_p[X_i = 0] = 1 - p$ gilt

$$\Pr_p[X_i = x_i] = p^{x_i}(1-p)^{1-x_i} \ .$$

Die Likelihood-Funktion ist

$$L(\vec{x}; p) = \prod_{i=1}^{n} p^{x_i} \cdot (1-p)^{1-x_i}$$

Maximum-Likelihood-Prinzip VI: Beispiele

Statt L maximieren wir $\ln L$ (einfachere Berechnung).

$$\begin{aligned}\ln L(\vec{x};p) &= \sum_{i=1}^{n}(x_i \cdot \ln p + (1-x_i)\cdot \ln(1-p)) \\ &= n\bar{x}\cdot \ln p + (n-n\bar{x})\cdot \ln(1-p).\end{aligned}$$

mit $\bar{x} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i$.

Nullsetzen der Ableitung ergibt:

$$\frac{\mathsf{d}\ln L(\vec{x};p)}{\mathsf{d}\,p} = \frac{n\bar{x}}{p} - \frac{n-n\bar{x}}{1-p} = 0.$$

mit Lösung $p = \bar{x}$, ein Maximum.

Maximum-Likelihood-Prinzip VII: Beispiele

Sei $X \sim \mathcal{N}(\mu, \sigma^2)$. Wir bestimmen Schätzer für μ und σ^2. Es gilt

$$L(\vec{x}; \mu, \sigma^2) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \cdot \prod_{i=1}^{n} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right).$$

Durch Logarithmieren erhalten wir

$$\ln L(\vec{x}; \mu, \sigma^2) = -n(\ln\sqrt{2\pi} + \ln\sigma) + \sum_{i=1}^{n} \left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right).$$

Maximum-Likelihood-Prinzip VII: Beispiele

Nullsetzen der Ableitungen ergibt

$$\frac{\partial \ln L}{\partial \mu} = \sum_{i=1}^{n} \frac{x_i - \mu}{\sigma^2} \stackrel{!}{=} 0$$

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{\sigma^3} \stackrel{!}{=} 0$$

und wir erhalten

$$\mu = \bar{x} \quad \text{und} \quad \sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2.$$

Der zweiter Wert ist fast die Stichprobenvarianz S_{n-1}^2 aber mit Vorfaktor $\frac{1}{n}$ statt $\frac{1}{n-1}$.

Die ML-Schätzvariable für die Varianz ist somit nicht erwartungstreu.

22. Konfidenzintervalle

Konfidenzintervalle I

Problem: Wie gut kann man sich auf einen Schätzwert verlassen?

Lösung: Berechne statt einem Schätzer U zwei Schätzer U_1 und U_2 mit

$$\Pr[U_1 \leq \theta \leq U_2] \geq 1 - \alpha.$$

Die W'keit $1-\alpha$ heißt Konfidenzniveau und kann dem „Sicherheitsbedürfnis“ angepasst werden.

Informalle Bedeutung:

> *Wenn wir für eine Stichprobe die Schätzer U_1 und U_2 berechnen und $\theta \in [U_1, U_2]$ schliessen, dann irren wir höchstens mit W'keit α.*

$[U_1, U_2]$ heißt Konfidenzintervall.

Oft setzt man $U_1 := U - \delta$ und $U_2 := U + \delta$ für einen Schätzer U (symmetrisches Konfidenzintervall).

Konfidenzintervalle II: Normalverteilung

Beispiel 163

Sei $X \sim \mathcal{N}(\mu, \sigma^2)$ und seien $X_1, \ldots, X_n$ zugehörige Stichprobenvariablen.

Wir Schätzen μ durch $\overline{X}$.

Wir suchen ein symmetrisches Konfidenzintervall für $\overline{X}$, d.h. einen Wert δ mit

$$\Pr[\overline{X} - \delta \leq \theta \leq \overline{X} + \delta] \geq 1 - \alpha.$$

Es gilt (Satz 133) $\overline{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$. Nach Lemma 127 ist

$$Z := \sqrt{n} \cdot \frac{\overline{X} - \mu}{\sigma}$$

standardnormalverteilt.

Konfidenzintervalle III: Normalverteilung

Für Z suchen wir nach einem Konfidenzintervall $[-c, c]$ mit

$$\Pr[-c \leq Z \leq c] \stackrel{!}{=} 1 - \alpha.$$

Auflösen nach μ ergibt

$$\Pr\left[\overline{X} - \frac{c\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + \frac{c\sigma}{\sqrt{n}}\right] \stackrel{!}{=} 1 - \alpha\,.$$

Für c muss also gelten:

$$\Pr[-c \leq Z \leq c] = \Phi(c) - \Phi(-c) \stackrel{!}{=} 1 - \alpha.$$

Mit $\Phi(-c) = 1 - \Phi(c)$ erhalten wir

$$2 \cdot \Phi(c) - 1 \stackrel{!}{=} 1 - \alpha$$

also

$$c = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right).$$

Konfidenzintervalle IV: Normalverteilung

Definition 164

X sei eine stetige Zufallsvariable mit Verteilung F_X.
Die Zahl x_γ mit

$$F_X(x_\gamma) = \gamma$$

heißt γ-Quantil der Verteilung F_X.

Für die Standardnormalverteilung bezeichnet z_γ das γ-Quantil.

Damit kann das gesuchte Konfidenzintervall durch

$$\left[\, \overline{X} - \frac{z_{(1-\frac{\alpha}{2})}\sigma}{\sqrt{n}} \,,\, \overline{X} + \frac{z_{(1-\frac{\alpha}{2})}\sigma}{\sqrt{n}} \,\right]$$

angegeben werden.

23. Testen von Hypothesen

Testen von Hypothesen I: Einführung

Bislang haben wir versucht, Parameter von Verteilungen zu schätzen, z.B. die Erfolgsw'keit p einer Bernoulli-verteilte Variable. X.

In der Praxis ist man jedoch nur an einer Eigenschaft des Parameters interessiert, z.B. ob $p \geq 1/3$, $p \leq 0.8$, oder $p = 0.5$ gilt.

Statistische Tests werden verwendet, um mit einer gewissen Fehlerw'keit zu entscheiden, ob die Eigenschaft gilt oder nicht.

Testen von Hypothesen II: Terminologie

Definition 165

Die zu überprüfende Eigenschaft bezeichnet man mit H_0 und wird Nullhypothese genannt.

Man nimmt an, dass entweder die Nullhypothese oder die mit H_1 bezeichnete Alternative gilt.

Bei den meisten Tests gilt $H_1 := \neg H_0$ (triviale Alternative).

Der Test entscheidet, welche von den beiden Eigenschaften abgelehnt wird.

Testen von Hypothesen III: Terminologie

Ein Beispiel, in dem eine nicht triviale Alternative Sinn macht:

Beispiel 166

Von einer Festplatte ist bekannt, dass sie zu einer von zwei Baureihen gehört. Die mittleren Zugriffszeiten dieser Baureihen betragen 9ms bzw. 12ms.

Es soll mit einem statistischen Test entschieden werden, zu welchem Typ die betrachtete Festplatte gehört.

In diesem Test $H_0 : \mu \leq 9$ und $H_1 : \mu \geq 12$ mit μ die mittlere Zugriffszeit.

Testen von Hypothesen IV: Terminologie

Ein Test wiederholt n-Mal das zur Variablen X zugehörigen Zufallsexperiment. Seien $X_1, \ldots, X_n$ unabhängige Stichprobenvariablen mit derselben Verteilung wie X.

Definition 167

Die Testgröße eines Tests ist eine Zufallsvariable der Gestalt $T := f(X_1, \ldots, X_n)$.

Die Ablehnungsbedingung ist eine Bedingung der Gestalt

$$
\begin{array}{ll}
T < c \ \text{ oder } \ T > c & \text{(einseitiger Test)} \\
T < c_1 \ \vee \ T > c_2 & \text{(zweiseitiger Test)}
\end{array}
$$

Die Nullhypothese wird abgelehnt, wenn der Stichprobenvektor $\vec{x} = (x_1, \ldots, x_n)$ die Ablehnungsbedingung erfüllt (d.h., $f(\vec{x}) < c$, $f(\vec{x}) > c$, oder $f(\vec{x}) < c_1 \vee f(\vec{x}) > c_2$ gilt).

Der Ablehnungsbereich ist die Menge der Stichprobenvektoren $\vec{x} = (x_1, \ldots, x_n)$, für die die Nullhypothese abgelehnt wird.

Testen von Hypothesen V: Terminologie

Definition 168

Ein Fehler 1. Art tritt ein, wenn H_0 irrtümlich abgelehnt wird.
(D.h., H_0 gilt aber die Stichprobe $\vec{x}$ liegt im Ablehnungsbereich.)

Ein Fehler 2. Art tritt ein, wenn H_0 irrtümlich angenommen wird.
(D.h., H_0 gilt nicht aber die Stichprobe $\vec{x}$ liegt nicht im Ablehnungsbereich.)

Sei α eine obere Schranke der W'keit für den Fehler 1. Art. Wir sagen, dass der Test α-Fehler oder Signifikanzniveau von α hat.

In der Praxis gibt man sich einen Wert für α vor und bestimmt man den Ablehnungsbereich so, dass der Test Signifikanzniveu α hat.

Übliche Werte für α sind $0{,}05$, $0{,}01$ oder $0{,}001$.

Testen von Hypothesen VI: Terminologie

Die Minimierung des Fehlers 1. Art und des Fehlers 2. Art sind gegenläufige Ziele, z.B.:

- Lehnt man die Nullhypothese nie ab, hat ein Fehler 1. Art W'keit 0. Allerdings tritt ein Fehler 2. Art immer ein, wenn H_0 nicht gilt.
- Lehnt man die Nullhypothese immer ab, hat ein Fehler 2. Art W'keit 0. Allerdings tritt ein Fehler 1. Art immer ein, wenn H_0 gilt.

Ziel der meisten statistischen Tests ist eine kleine W'keit für den Fehler 1. Art zu garantieren. Die W'keit des Fehlers 2.Art kann groß sein!

Bei der Wahl der Nullhypothese (setzt man $H_0 = H$ oder $H_0 = \neg H$?) muss das berücksichtigt werden.

Testen von Hypothesen VII: Entwurf eines Tests

Wir entwerfen einen Test für den Parameter p einer Bernoulli-verteilten Zufallsvariablen X.

Wir setzen $H_0 : p \geq p_0$ und $H_1 : p < p_0$.

Als Testgröße verwenden wir $T := X_1 + \ldots + X_n$.
(Anzahl der Köpfe.)

Wir lehnen H_0 ab wenn der Wert von T "zu klein" ist.
Als Gestalt des Ablehnungsbereichs wählen wir also $T < k$.
(Einseitiger Test.)
Für eine Nullhypothese $H_0 : p = p_0$ würden wir $T < k_1 \ \vee \ T > k_2$ wählen.

Der Wert $k \in \mathbb{R}$ muss in Abhängigkeit des Signifikanzniveaus α festgelegt werden.

Testen von Hypothesen VIII: Entwurf eines Tests

Wir bestimmen k.

Es gilt $T \sim Bin(n,p)$. Bei großer Stichprobenumfang n ist die Variable

$$\tilde{T} := \frac{T - np}{\sqrt{np(1-p)}}.$$

annähernd standardnormalverteilt (siehe Korollar 139).

Wir berechnen für jeden Wert von k die maximale Fehlerw'keit über die möglichen Werten von p:

$$\begin{aligned}
\text{Fehlerw'keit 1. Art} &\leq \sup_{p \text{ erfüllt } H_0} \Pr_p[\text{„}H_0 \text{ wird abgelehnt“}] \\
&= \max_{p \geq p_0} \Pr_p[T < k] = \Pr_{p=p_0}[T < k] \\
\text{Fehlerw'keit 2. Art} &\leq \sup_{p \text{ erfüllt } H_1} \Pr_p[\text{„}H_1 \text{ wird abgelehnt“}] \\
&= \sup_{p < p_0} \Pr_p[T \geq k] = \Pr_{p=p_0}[T \geq k]
\end{aligned}$$

Testen von Hypothesen IX: Entwurf eines Tests

Für das gewünschte Signifikanzniveau α erhalten wir

$$\begin{aligned}
\alpha &= \Pr_{p=p_0}[T < k] \\
&= \Pr_{p=p_0}\left[\tilde{T} < \frac{k - np}{\sqrt{np(1-p)}}\right] \\
&= \Pr\left[\tilde{T} < \frac{k - np_0}{\sqrt{np_0(1-p_0)}}\right] \approx \Phi\left(\frac{k - np_0}{\sqrt{np_0(1-p_0)}}\right).
\end{aligned}$$

Testen von Hypothesen X: Entwurf eines Tests

Damit gilt: Für jedes

$$k \leq z_\alpha \sqrt{np_0(1-p_0)} + np_0$$

kann die Nullhypothese $p \geq p_0$ abgelehnt werden mit Signifikanzniveau α.

Das größte solche k minimiert den Fehler 2. Art.

Für diesen Wert gilt:

$$\text{Fehlerw'keit 1. Art} = \Pr_{p=p_0}[T < k]$$
$$\text{Fehlerw'keit 2. Art} = \Pr_{p=p_0}[T \geq k]$$

und damit $\text{Fehlerw'keit 2. Art} = 1 - \text{Fehlerw'keit 1. Art}$

Es ist also nicht möglich, beide W'keiten gleichzeitig zu reduzieren.

Beispiel 169

Ein Test soll mit Signifikanz $\alpha = 0.05$ bestimmen, ob die W'keit von Kopf bei einer Münze mindestens $1/3$ beträgt.

Wenn die Münze $n = 450$ Mal geworfen wird, dann wird die Hypothese $p \geq 1/3$ abgelehnt werde, wenn die Anzahl der Köpfe unter

$$\begin{aligned} k &= z_{0.05} \cdot \sqrt{450 \cdot \frac{1}{3} \cdot \frac{2}{3}} + 450 \cdot \frac{1}{3} \\ &= z_{0.05} \cdot 10 + 150 \\ &\approx -1.64 \cdot 10 + 150 = 133,6 \end{aligned}$$

liegt.

Testen von Hypothesen XI: Entwurf eines Tests

Bei echten Alternativtests werden für hinreichend große Stichproben beide Testfehler klein.

Beispiel 170

Bei zwei Sorten von gezinkten Münzen ist die Erfolgs'wkeit jeweils $p \geq 3/5$ und $p \leq 2/5$. Mit Hilfe eines Tests soll die Sorte einer gegebenen Münze bestimmt werden.

Unter dieser Annahme setzt man:

$$H_0 : p \geq 3/5 \qquad H_1 : p \leq 2/5.$$

Testen von Hypothesen XII: Entwurf eines Tests

$$\begin{aligned}
\text{Fehlerw'keit 1. Art} &\leq \sup_{p \text{ erfüllt } H_0} \Pr_p[\text{„}H_0 \text{ wird abgelehnt“}] \\
&= \max_{p \geq 3/5} \Pr_p[T < k] = \Pr_{p=3/5}[T < k] \\
&\approx \Phi\left(\frac{k - n \cdot 3/5}{\sqrt{n \cdot 6/25}}\right)
\end{aligned}$$

$$\begin{aligned}
\text{Fehlerw'keit 2. Art} &\leq \sup_{p \text{ erfüllt } H_1} \Pr_p[\text{„}H_1 \text{ wird abgelehnt“}] \\
&= \sup_{p \leq 2/5} \Pr_p[T \geq k] = \Pr_{p=2/5}[T \geq k] \\
&\approx 1 - \Phi\left(\frac{k - n \cdot 2/5}{\sqrt{n \cdot 6/25}}\right)
\end{aligned}$$

Testen von Hypothesen XIII: Entwurf eines Tests

Mit z.B. $n = 100$ und $\alpha = 0.05$ ergibt sich $k \approx 51.94$. Für diesen Wert erhalten wir

$$\begin{aligned}\text{Fehlerw'keit 2. Art} &\approx 1 - \Phi\left(\frac{51.94 - 40}{2\sqrt{6}}\right) \\ &\approx 1 - \Phi(2.437) \approx 1 - 0.9926 \\ &= 0.0074\end{aligned}$$

24. Praktische Anwendung statistischer Tests

Praktische Anwendung statistischer Tests

Das im vorhergehenden Abschnitt konstruierte Testverfahren taucht in der Literatur unter dem Namen approximativer Binomialtest auf.

Die folgende Tabelle 1 gibt einen Überblick über die Eckdaten dieses Tests.

Tabelle: **Approximativer Binomialtest**

Annahmen:

$X_1, \ldots, X_n$ seien unabhängig und identisch verteilt mit $\Pr[X_i = 1] = p$ und $\Pr[X_i = 0] = 1 - p$, wobei p unbekannt sei.
n sei hinreichend groß, so dass die Approximation aus Korollar 139 brauchbare Ergebnisse liefert.

Hypothesen:

$$\begin{array}{lllll} a) & H_0 : p = p_0 & \text{gegen} & H_1 : p \neq p_0, \\ b) & H_0 : p \geq p_0 & \text{gegen} & H_1 : p < p_0, \\ c) & H_0 : p \leq p_0 & \text{gegen} & H_1 : p > p_0. \end{array}$$

Testgröße:

$$Z := \frac{h - np_0}{\sqrt{np_0(1 - p_0)}},$$

wobei $h := X_1 + \ldots + X_n$ die Häufigkeit bezeichnet, mit der die Ereignisse $X_i = 1$ aufgetreten sind.

Ablehnungskriterium für H_0 bei Signifikanzniveau α:

$$\begin{array}{ll} a) & |Z| > z_{1-\alpha/2}, \\ b) & Z < z_\alpha, \\ c) & Z > z_{1-\alpha}. \end{array}$$

Allgemeines Vorgehen bei statistischen Tests

1. Schritt: Formulierung von Annahmen.
   Ganz ohne Annahmen kommt man meist nicht aus.
   Übliche Annahmen betreffen die Verteilung der Stichprobenvariablen und deren Unabhängigkeit.
2. Schritt: Formulierung der Nullhypothese.
3. Schritt: Auswahl des Testverfahrens.
4. Schritt: Durchführung des Tests und Entscheidung.

Wie findet man das richtige Testverfahren? I

Statistische Tests kann man nach mehreren Kriterien in Klassen einteilen.

- **Anzahl der beteiligten Zufallsgrößen**
  - Ein-Stichproben-Test.
    Es wird nur eine Zufallsgröße untersucht, für die eine Stichprobe erzeugt wird.
    Beispiel: Beträgt die mittlere Zugriffszeit auf einen Datenbankserver im Mittel höchstens 10ms?
  - Zwei-Stichproben-Test.
    Zwei Zufallsgrößen, für die jeweils eine Stichprobe erzeugt wird, werden verglichen.
    Beispiel: Hat Datenbankserver A eine kürzere mittlere Zugriffszeit als Datenbankserver B?

Wie findet man das richtige Testverfahren? II

- **Art der Nullhypothese**
  - Tests auf Lageparameter.
    Die Nullhypothese ist eine Aussage über Lageparameter der Verteilung wie den Erwartungswert oder die Varianz.
  - Tests auf eine vorgegebene Verteilung.
    Die Nullhypotese ist eine Aussage über den Verteilungstyp, z.B. dass die Zufallsgröße Normalverteilt ist.
  - Tests auf ein Maß für die Abhängigkeit verschiedener Zufallsgröße.
    Z.B. sagt die Nullhypothese, dass zwei Zufallsvariablen unabhängig sind.

Wie findet man das richtige Testverfahren? III

- **Annahmen über die Zufallsgrößen**
  - Bekannter/unbekannter Verteilungstyp.
  - Bekannter/unbekannter Erwartungswert.
  - Bekannte/unbekannte Varianz.

Ein-Stichproben-Tests für Lageparameter I: Gaußtest

Tabelle: **Gaußtest**

Annahmen:

$X_1, \dots, X_n$ seien unabhängig und identisch verteilt mit $X_i \sim \mathcal{N}(\mu, \sigma^2)$, wobei σ^2 bekannt ist.
Alternativ gelte $\mathbb{E}[X_i] = \mu$ und $\mathrm{Var}[X_i] = \sigma^2$, und n sei groß genug.

Hypothesen:

$$\begin{array}{lllll} a) & H_0 : \mu = \mu_0 & \text{gegen} & H_1 : \mu \neq \mu_0, \\ b) & H_0 : \mu \geq \mu_0 & \text{gegen} & H_1 : \mu < \mu_0, \\ c) & H_0 : \mu \leq \mu_0 & \text{gegen} & H_1 : \mu > \mu_0. \end{array}$$

Testgröße:

$$Z := \frac{\overline{X} - \mu_0}{\sigma} \sqrt{n} \, .$$

Ablehnungskriterium für H_0 bei Signifikanzniveau α:

$$\begin{array}{ll} a) & |Z| > z_{1-\alpha/2}, \\ b) & Z < z_{\alpha}, \\ c) & Z > z_{1-\alpha}. \end{array}$$

Ein-Stichproben-Tests für Lageparameter II: Gaußtest

Der Gaußtest hat den Nachteil, dass man die Varianz σ^2 der Variablen $X_1, \ldots, X_n$ kennen muss.

Wenn diese unbekannt ist kann die Varianz durch die Stichprobenvarianz S^2 (siehe Definition 158) ersetzt werden.

Dies führt auf den so genannten t-Test.

Ein-Stichproben-Tests für Lageparameter III: t-Test

Tabelle: ***t*-Test**

Annahmen:

$X_1, \ldots, X_n$ seien unabhängig und identisch verteilt mit $X_i \sim \mathcal{N}(\mu, \sigma^2)$.
Alternativ gelte $\mathbb{E}[X_i] = \mu$ und $\mathrm{Var}[X_i] = \sigma^2$, und n sei groß genug.

Hypothesen:

a) $H_0 : \mu = \mu_0$ *gegen* $H_1 : \mu \neq \mu_0$,
b) $H_0 : \mu \geq \mu_0$ *gegen* $H_1 : \mu < \mu_0$,
c) $H_0 : \mu \leq \mu_0$ *gegen* $H_1 : \mu > \mu_0$.

Testgröße:

$$T := \frac{\overline{X} - \mu_0}{S} \sqrt{n}.$$

Ablehnungskriterium für H_0 bei Signifikanzniveau α:

a) $|T| > t_{n-1,1-\alpha/2}$,
b) $T < t_{n-1,\alpha}$,
c) $T > t_{n-1,1-\alpha}$.

Ein-Stichproben-Tests für Lageparameter IV: t-Test

Hierbei bezeichnet $t_{n-1,1-\alpha}$ das

$(1-\alpha)$-*Quantil der t-Verteilung mit* $n-1$ *Freiheitsgraden*

an.

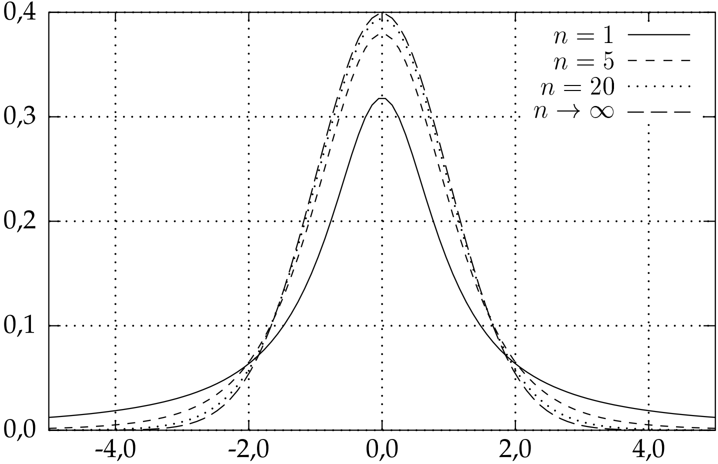
Die t-Verteilung nennt man auch Student-Verteilung.

Es handelt sich eigentlich um eine Familie von Verteilungen, eine für jede Anzahl von Freihaitsgraden.

Die Dichte ähnelt der Dichte der Normalverteilung.

Für große n (Faustregel: $n \geq 30$) wird in der Praxis die t-Verteilung durch die Normalverteilung angenähert.

Ein-Stichproben-Tests für Lageparameter V: t-Test



Dichte der t-Verteilung mit n Freiheitsgraden

Zwei-Stichproben-Tests für Lageparameter I: t-Test

Tabelle: **Zwei-Stichproben-t-Test**

Annahmen:

$X_1, \ldots, X_m$ und $Y_1, \ldots, Y_n$ seien unabhängig und jeweils identisch verteilt, wobei $X_i \sim \mathcal{N}(\mu_X, \sigma_X^2)$ und $Y_i \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ gelte.
Die Varianzen seien identisch, also $\sigma_X^2 = \sigma_Y^2$.

Hypothesen:

a) $H_0 : \mu_X = \mu_Y$ *gegen* $H_1 : \mu_X \neq \mu_Y$,
b) $H_0 : \mu_X \geq \mu_Y$ *gegen* $H_1 : \mu_X < \mu_Y$,
c) $H_0 : \mu_X \leq \mu_Y$ *gegen* $H_1 : \mu_X > \mu_Y$.

Testgröße:

$$T := \sqrt{\frac{n+m-2}{\frac{1}{n}+\frac{1}{m}}} \cdot \frac{\overline{X}-\overline{Y}}{\sqrt{(m-1)\cdot S_X^2 + (n-1)\cdot S_Y^2}}.$$

Ablehnungskriterium für H_0 bei Signifikanzniveau α:

a) $|T| > t_{m+n-2,1-\alpha/2}$,
b) $T < t_{m+n-2,\alpha}$,
c) $T > t_{m+n-2,1-\alpha}$.

Zwei-Stichproben-Tests für Lageparameter II

Beispiel 171

Im SS 2012 nahmen 48 Studierende an die Evaluierung der DWT-Vorlesung teil. Sie beurteilten die Vorlesung 21x mit der Note 1, 21x mit der Note 2 und 6x mit der Note 3.

Damit ergibt sich für die Zufallsvariable X, die die Note darstellt, $\overline{X} \approx 1.69$ und $\overline{S}_X \approx 0.68$.

Im SS 2013 nahmen 61 Studierende teil. Sie beurteilten die Vorlesung 20x mit der Note 1, 33x mit der Note 2, 7x Mal mit der Note 3 und 1x mit der Note 4.

Damit ergibt sich für die Zufallsvariable Y, die die Note darstellt, $\overline{Y} \approx 1.82$ und $\overline{S}_Y \approx 0.69$.

Prof. Evilsparza ist vom schlechteren Ergebniss im 2013 schwer beleidigt. Er erwägt, alle Teilnehmer der Klausur eine 5 zu verpassen. Dr. Luttenberger redet ihn jedoch ein, dass er erst die statistische Signifikanz der Ergebnisse prüfen soll.

Frage: Kann Prof. Evilsparza mit Signifikanzniveau von $\alpha = 0.05$ behaupten, dass die Studierenden vom SS 2013 eine schlechtere Meinung haben?

Zwei-Stichproben-Tests für Lageparameter III

Seien $X_1, \ldots, X_{48}$ und $Y_1, \ldots, Y_{61}$ jeweils unabhängig und identisch verteilt. Wir definieren

$$\begin{array}{rclcrcl} X_1' & = & (X_1 + \cdots + X_6)/6 & \cdots & X_8' & = & (X_{43} + \cdots + X_{48})/6 \\ Y_1' & = & (Y_1 + \cdots + Y_6)/6 & \cdots & Y_{10}' & = & (Y_{55} + \cdots + X_{60})/6 \end{array}$$

und nehmen an, dass

$$X_i' \sim \mathcal{N}(\mu_{X'}, \sigma^2) \qquad \text{und} \qquad \mu_{Y'} \sim \mathcal{N}(\mu_{Y'}, \sigma^2)$$

für unbekannte $\mu_{X'}, \mu_{Y'}, \sigma^2$ gilt.

Wir haben

$$\overline{X'} = \overline{X} \qquad \overline{Y'} = \overline{Y}$$

und sowei

$$S_{X'} = \frac{1}{\sqrt{6}}\, S_X \qquad S_{Y'} = \frac{1}{\sqrt{6}}\, S_Y$$

Zwei-Stichproben-Tests für Lageparameter IV

Wir wählen als Nullhypothese

$$H_0 : \mu_{X'} = \mu_{Y'}$$

(identischen W'keitsverteilungen in beiden Semestern).

Für die Testgröße ergibt sich

$$T = \sqrt{\frac{8+10-2}{1/8+1/10}} \cdot \frac{1.69-1.82}{\sqrt{7 \cdot 0.68^2/6 + 9 \cdot 0.69^2/6}} \approx -0.861$$

Aus einer Tabelle entnehmen wir

$$t_{m+n-2,1-\alpha/2} = t_{16,0.975} = 2.120 > 0.861 = |T|$$

Damit kann H_0 nicht abgelehnt werden.

Für $\alpha = 0.4$ gilt

$$t_{m+n-2,1-\alpha/2} = t_{16,0.40} = 0.865 > 0.861 = |T|$$

und so kann H_0 wieder nicht abgelehnt werden (aber fast).

Tests auf Verteilungen I: χ^2-Anpassungstest

Tabelle: **χ^2-Anpassungstest**

Annahmen:

$X_1, \ldots, X_n$ seien unabhängig und identisch verteilt mit $W_{X_i} = \{1, \ldots, k\}$.

Hypothesen:

$$\begin{aligned} H_0 &: \Pr[X = i] = p_i \quad \text{für } i = 1, \ldots, k, \\ H_1 &: \Pr[X = i] \neq p_i \quad \text{für mindestens ein } i \in \{1, \ldots, k\}, \end{aligned}$$

Testgröße:

$$T = \sum_{i=1}^{k} \frac{(h_i - np_i)^2}{np_i},$$

wobei h_i die Häufigkeit angibt, mit der $X_1, \ldots, X_n$ den Wert i angenommen haben.

Ablehnungskriterium für H_0 bei Signifikanzniveau α:

$$T > \chi^2_{k-1,1-\alpha};$$

dabei sollte gelten, dass $np_i \geq 1$ für alle i und $np_i \geq 5$ für mindestens 80% der Werte $i = 1, \ldots, k$.

Tests auf Verteilungen II: χ^2-Anpassungstest

Für die Testgröße T wird näherungsweise eine

χ^2-Verteilung mit $k-1$ Freiheitsgraden

angenommen.

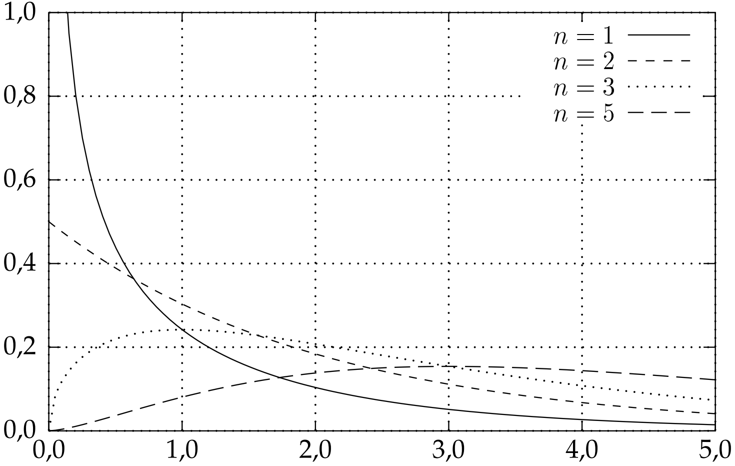
Das γ-Quantil einer χ^2-Verteilung mit k Freiheitsgraden bezeichen wir mit $\chi^2_{k,\gamma}$.

Die Werte dieser Verteilung finden sich in entsprechenden Tabellen in der Literatur.

Damit die Approximation gerechtfertigt ist, sollte gelten (Faustregel)

- $np_i \geq 1$ für alle i und
- $np_i \geq 5$ für mindestens 80% der Werte $i = 1, \dots, k$.

Tests auf Verteilungen III: χ^2-Anpassungstest



Dichte der χ^2-Verteilung mit n Freiheitsgraden

Beispiel 172

Wir wollen überprüfen, ob der Zufallszahlengenerator von Maple eine gute Approximation der Gleichverteilung liefert.

Dazu lassen wir Maple $n = 100000$ Zufallszahlen aus der Menge $\{1, \ldots, 10\}$ generieren.

Die Nullhypothese lautet $p_1 = \ldots = p_{10} = 1/10$, wobei p_i die W'keit von i bezeichnet.

Die Nullhypothese soll mit Signifikanzniveau von $\alpha = 0{,}05$ getestet werden.

Tests auf Verteilungen V: χ^2-Anpassungstest

Beispiel:

i	1	2	3	4	5	6	7	8	9	10
h_i	10102	10070	9972	9803	10002	10065	10133	9943	10009	9901

Es gilt $T = 8{,}9946$ und $\chi^2_{9,0{,}95} \approx 16{,}919$.

Die Nullhypothese wird nicht abgelehnt.

Was bedeutet das?

Nicht viel! Der Test garantiert nur, dass die W'keit, die Nullhypothese irrtümlich abzulehnen, höchstens 5% beträgt. Die Nullhypothese ist jedoch nicht abgelehnt worden!

Teil VI

Markov-Ketten

Markov-Ketten

Markov-Ketten modellieren mehrstufige Experimente mit unendlich vielen Stufen.

Der Ausgang einer Stufe bestimmt welches Experiment in der nächsten Stufe ausgeführt wird.

Erinnerung: Markov-Diagramme

Definition 173

Ein (verallgemeinertes) Markov-Diagramm $D = (Q, T, \delta)$ besteht aus

- einer (nicht notwendigerweise endlichen) Menge Q von Zuständen,
- einer Menge $T \subseteq Q \times Q$ von Transitionen, und
- einer W'keitsfunktion $\delta\colon T \to (0, 1]$, die Folgendes erfüllt für jeden Zustand q:

$$\sum_{(q,q')\in T} \delta(q, q') = 1 \ .$$

Definition einer Markov-Kette

Definition 174

Eine W'keitsverteilung oder Verteilung für ein Markov-Diagramm mit Zustandsmange Q ist eine Funktion $v\colon Q \to [0,1]$ mit der Eigenschaft

$$\sum_{q \in Q} v(q) = 1 \, .$$

Definition 175

Eine Markov-Kette ist ein Tuple $M = (Q, T, \delta, \pi_0)$, wobei:

- (Q, T, δ) ist ein Markov-Diagramm und
- $\pi_0\colon Q \to [0,1]$ ist die Anfangsverteilung.

Eine Markov-Kette ist endlich bzw. abzählbar, wenn Q endlich bzw. abzählbar ist.

Die ménàge a trois von Armand, Bertrand und Cécile I

Fragen von Armand an Denis, der sich in W'keitstheorie auskennt:

- Heute Morgen (Donnerstag) ist Cécile zu Bertrand gegangen. Mit welcher W'keit wird sie den Sonntag mit mir verbringen?
- Wenn Cécile mich verlässt, wie lange dauert es im Schnitt, bis Sie zurückkommt?
- Wenn diese Situation für immer so weiter geht, wieviel Prozent der Tage wird Cécile mit mir verbringen?

Wir untersuchen Methoden, um diese Fragen zu beantworten.

W'keitsraum einer Markov-Kette I

Definition 176

Ein Pfad einer Markov-Kette $M = (Q, T, \delta, \mathcal{Q}_0)$ ist eine endliche oder unendliche Sequenz $\sigma = q_0\, q_1 \ldots q_k \ldots$ von Zuständen mit $k \geq 0$ und $(q_i, q_{i+1}) \in T$ für alle $q_i\, q_{i+1}$ in σ.

Π bzw. Π_ω bezeichnen die Menge aller endlichen bzw. unendlichen Pfaden von M.

$\sigma(k)$ bezeichnet den Zustand q_k, d.h. $\sigma = \sigma(0)\, \sigma(1) \ldots \sigma(k) \ldots$.

Die Konkatenation von $\sigma \in \Pi$ und $\sigma' \in \Pi \cup \Pi_\omega$ wird mit $\sigma \cdot \sigma'$ oder $\sigma\, \sigma'$ bezeichnet.

Sei $\sigma \in \Pi$ ein endlicher Pfad. Die von σ generierte Zylindermenge $Cyl(\sigma)$ ist die Menge aller unendlichen Pfaden $\sigma' \in \Pi_\omega$ mit σ als Präfix.

W'keitsraum einer Markov-Kette II

Definition 177

Der W'keitsraum einer abzählbaren Markov-Kette M mit Anfangsverteilung $\mathcal{Q}_0$ ist die Triple $(\Omega, \mathcal{A}, \Pr)$ mit

- $\Omega = \Pi_\omega$.
- $\mathcal{A}$ enthält die von den Zylindermengen generierten Borel'sche Mengen, d.h.:
  - $Cyl(\sigma) \in \mathcal{A}$ für jedes $\sigma \in \Pi$.
  - Wenn $A \in \mathcal{A}$, dann $\Omega \setminus A \in \mathcal{A}$.
  - Wenn $A_1, A_2, \ldots \in \mathcal{A}$, dann $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$.
- Die W'keitsfunktion $\Pr$ ist die einzige Funktion, die

  $$\Pr\left[Cyl(q_0 q_1 \ldots q_n)\right] = \mathcal{Q}_0(q_0) \cdot \prod_{i=0}^{n-1} \delta(q_i, q_{i+1})$$

  und die Kolmogorov-Axiome erfüllt.

Zufallsvariablen einer Markov-Kette

Definition 178

Sei $M = (Q, T, \delta, \mathcal{Q}_0)$ eine abzählbare Markov-Kette.

Für jedes $t \in \mathbb{N}_0$ bezeichnet X_t die Zufallsvariable $X_t \colon \Omega \to Q$ mit

$$X_t(\sigma) = \sigma(t) \ .$$

- X_t gibt den Zustand der Kette zum Zeitpunkt t.
- X_t ist wohldefiniert: Man kann leicht zeigen, dass für jeden Zustand $q \in Q$ die Menge „$X_t = q$“ Borel ist.
- Für alle $t \geq 0$ gilt:

$$\begin{aligned} \Pr[X_{t+1} = q' \mid X_t = q] &= \delta(q, q') \\ \Pr[X_{t+1} = q_{t+1} \mid X_t = q_t, \ldots, X_0 = q_0] &= \delta(q, q') \ . \end{aligned}$$

25. Übergangswahrscheinlichkeiten

Übergangsw'keiten I: Übergangsmatrix

Definition 179

Sei $M = (Q, T, \delta, \mathcal{Q}_0)$ eine endliche Markov-Kette mit $Q = \{q_1, \ldots, q_n\}$.

Die $n \times n$ Matrix $P = (p_{ij})_{0 \leq i,j < n}$ mit

$$p_{ij} = \delta(q_i, q_j) = \Pr[X_{t+1} = q_j \mid X_t = q_i]$$

ist die Übergangsmatrix von M.

Beispiel 180

Die Matrix der Armand-Bertrand-Cécile-Kette (mit $q_1 :=$ Armand und $q_2 :=$ Bertrand) ist:

$$P = \begin{pmatrix} 0.8 & 0.2 \\ 0.1 & 0.9 \end{pmatrix}$$

Übergangsw'keiten II: Berechnung

Sei $\mathcal{Q}_t = (\ \Pr[X_t = q_1], \ldots \Pr[X_t = q_n]\ )$

Es gilt

$$\Pr[X_0 = q_k] = \mathcal{Q}_0(q_k)$$

$$\begin{aligned}\Pr[X_{t+1} = q_k] &= \sum_{i=1}^{n} \Pr[X_{t+1} = q_k \mid X_t = q_i] \cdot \Pr[X_t = q_i] \\ &= \sum_{i=1}^{n} p_{ik} \cdot \Pr[X_t = q_i]\end{aligned}$$

also

$$(\mathcal{Q}_{t+1})_k = \sum_{i=1}^{n} p_{ik} \cdot (\mathcal{Q}_t)_i$$

und in Matrixschreibweise

$$\mathcal{Q}_{t+1} = \mathcal{Q}_t \cdot P$$

Übergangsw'keiten III: Berechnung

Mit der Matrixschreibweise erhalten wir für alle $t, k \in \mathbb{N}$

$$\mathcal{Q}_t = \mathcal{Q}_0 \cdot P^t \qquad \mathcal{Q}_{t+k} = \mathcal{Q}_t \cdot P^k$$

Beispiel 181 (Erste Frage von Armand)

Heute Morgen (Donnerstag) ist Cécile zu Bertrand gegangen. Mit welcher W'keit wird sie den Sonntag mit mir verbringen?

Modellierung: Sei $\mathcal{Q}_0 = (0, 1)$.
Gesucht wird $\mathcal{Q}_3(q_1) = \Pr[X_3 = q_1]$.

$$\mathcal{Q}_3 = \mathcal{Q}_0 \cdot P^3 = (0, 1) \cdot \begin{pmatrix} 0.8 & 0.2 \\ 0.1 & 0.9 \end{pmatrix}^3 = (0.219, 0.781)$$

Die W'keit beträgt somit 0.219.

Übergangsw'keiten IV: Exponentiation von Matrizen

Wenn P diagonalisierbar ist, so existiert eine Diagonalmatrix D und eine invertierbare Matrix B mit

$$P = B \cdot D \cdot B^{-1}$$

und somit

$$P^k = B \cdot D^k \cdot B^{-1}$$

wobei D^k sehr leicht zu berechnen ist. Die Diagonale von D enthält die Eigenwerte von P, d.h., die λ-Lösungen der Gleichung

$$P \cdot v = \lambda v$$

Die Spalten von B sind die Eigenvektoren von P, d.h., die v-Lösungen derselben Gleichung.

Übergangsw'keiten IV: Berechnung von D und B

Beispiel 182

$$P = \begin{pmatrix} 0.8 & 0.2 \\ 0.1 & 0.9 \end{pmatrix}$$

Eigenwerte: Nullstellen des charakteristischen Polynoms

$$|P - \lambda \cdot I| = \begin{vmatrix} 0.8 - \lambda & 0.2 \\ 0.1 & 0.9 - \lambda \end{vmatrix} = \lambda^2 - 1.7\lambda + 0.7$$

Wir erhalten: $\lambda_1 = 0.7$ und $\lambda_2 = 1$.

Dazugehörige Eigenvektoren:

$$\nu_1 = \begin{pmatrix} -2 \\ 1 \end{pmatrix} \text{ und } \nu_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Übergangsw'keiten V: Berechnung von D und B

Damit gilt

$$D = \begin{pmatrix} 0.7 & 0 \\ 0 & 1 \end{pmatrix} \qquad B = \begin{pmatrix} -2 & 1 \\ 1 & 1 \end{pmatrix}$$

$$B^{-1} = \begin{pmatrix} -\frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}$$

Es ergibt sich z.B.

$$P^{10} = \begin{pmatrix} -2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 0.7^{10} & 0 \\ 0 & 1^{10} \end{pmatrix} \begin{pmatrix} -\frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix} \approx \begin{pmatrix} 0.352 & 0.648 \\ 0.324 & 0.676 \end{pmatrix}$$

Wir berechnen die W'keit, dass Cécile am 10. Juli bei Armand bzw. Bertrand ist, wenn sie den 1. Juli bei Bertrand verbringt:

$$\mathcal{Q}_{10} = (0, 1) \cdot \begin{pmatrix} 0.352 & 0.648 \\ 0.324 & 0.676 \end{pmatrix} = (0.324, 0.676)$$

26. Ankunftsw'keiten und Übergangszeiten

Ankunftsw'keiten und Übergangszeiten

Wir untersuchen Fragestellungen auf, die sich auf zwei bestimmte Zustände q_i und q_j beziehen:

- Wie wahrscheinlich ist es, von q_i irgendwann nach q_j zu kommen?
- Wie viele Schritte benötigt die Kette im Mittel, um von q_i nach q_j zu gelangen?

Bemerkung: Die zweite Frage wurde schon im Wesentlichen im Abschnitt „Markov-Diagramme“ betrachtet.

Übergangszeiten

Definition 183

Sei T_j die Zufallsvariable

$$T_j(\sigma) := \begin{cases} \min\{n \geq 0 \mid X_n(\sigma) = q_j\} & \text{wenn Menge nichtleer} \\ \infty & \text{sonst} \end{cases}$$

Die bedingte Zufallsvariable

$$T_{ij} := T_j \mid \text{„}X_0 = q_i\text{“}$$

nennen wir die Übergangszeit (engl. hitting time) von q_i nach q_j.

T_{ij} zählt die Anzahl der Schritte, die für den Weg von q_i nach q_j benötigt werden.

Notation: $h_{ij} := \mathbb{E}[T_{ij}]$ (falls der bedingte Erwartungswert existiert).

Rückkehrzeiten

Im Fall $q_i = q_j$ gilt $T_{ii} = 0$ weil “die Kette schon in q_i ist“.

Wir untersuchen auch, wie lange es dauert, bis Zustand q_i zu einem späteren Zeitpunkt wieder besucht wird.

Definition 184

Sei T'_j die Zufallsvariable

$$T'_j(\sigma) := \begin{cases} \min\{n \geq 1 \mid X_n(\sigma) = q_j\} & \text{wenn Menge nichtleer} \\ \infty & \text{sonst} \end{cases}$$

Die bedingte Zufallsvariable

$$T_i := T'_i \mid \text{„}X_0 = q_i\text{“}$$

ist die Rückkehrzeit (engl. recurrence time) von q_i.

Notation: $h_i := \mathbb{E}[T_i]$ (falls der bedingte Erwartungswert existiert).

Definition 185

Die Ankunftsw'keit f_{ij}, vom Zustand q_i nach beliebig vielen Schritten in den Zustand q_j zu gelangen ist definiert durch

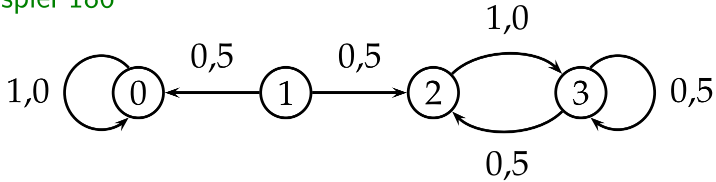
$$f_{ij} := \Pr[T_j < \infty \mid X_0 = q_i].$$

Die Rückkehrw'keit f_i, vom Zustand q_i nach beliebig vielen Schritten (mindestens 1) zurück zum Zustand q_i zu kehren ist definiert durch

$$f_i := \Pr[T_i < \infty \mid X_0 = q_i].$$

Ein Beispiel

Beispiel 186



Für alle $\sigma \in \Pi_\omega$ gilt

$$T_0(\sigma) = 1 \qquad T_{01}(\sigma) = T_{02}(\sigma) = T_{03}(\sigma) = \infty$$

$$T_{10}(\sigma) = \begin{cases} 1 & \text{falls } X_1(\sigma) = 0 \\ \infty & \text{falls } X_1(\sigma) = 2 \end{cases}$$

Es gilt auch

$$f_{10} = 0.5\ ,\ f_{32} = 1\ ,\ f_2 = 1 \quad \text{und} \quad h_{10} = \infty\ ,\ h_{32} = 2$$

Berechnung von f_{ij} und h_{ij} I

Lemma 187

Für die erwarteten Übergangs-/Rückkehrzeiten gilt

$$h_{ij} = 1 + \sum_{k \neq j} p_{ik} h_{kj} \qquad \text{für alle } q_i, q_j \in Q, q_i \neq q_j$$

$$h_j = 1 + \sum_{k \neq j} p_{jk} h_{kj} \qquad \text{für alle } q_j \in Q$$

sofern die Erwartungswerte h_{ij} und h_{kj} existieren.

Für die Ankunfts-/Rückkehrwahrscheinlichkeiten gilt analog

$$f_{ij} = p_{ij} + \sum_{k \neq j} p_{ik} f_{kj} \qquad \text{für alle } q_i, q_j \in Q, q_i \neq q_j$$

$$f_j = p_{jj} + \sum_{k \neq j} p_{jk} f_{kj} \qquad \text{für alle } q_j \in Q$$

Berechnung von f_{ij} und h_{ij} II

Beweis für f_{ij}:

Sei $q_i \neq q_j$. Es gilt

$$\begin{aligned}
\Pr[T_{ij} < \infty \mid X_1 = q_k] &= \Pr[T_{kj} < \infty] \quad \text{für} \quad q_k \neq q_j \\
\Pr[T_{ij} < \infty \mid X_1 = q_j] &= 1
\end{aligned}$$

und damit

$$\begin{aligned}
f_{ij} = \Pr[T_{ij} < \infty] &= \sum_{q_k \in Q} \Pr[T_{ij} < \infty \mid X_1 = q_k] \cdot p_{ik} \\
&= p_{ij} + \sum_{k \neq j} \Pr[T_{kj} < \infty] \cdot p_{ik} = p_{ij} + \sum_{k \neq j} p_{ik} f_{kj}
\end{aligned}$$

Die Ableitung für f_j ist analog.

Berechnung von f_{ij} und h_{ij} III

Beweis für h_{ij}:

Sei $q_i \neq q_j$. Es gilt

$$\begin{array}{rcl} \mathbb{E}[T_{ij} \mid X_1 = q_k] & = & 1 + \mathbb{E}[T_{kj}] \quad \text{für } q_k \neq q_j \\ \mathbb{E}[T_{ij} \mid X_1 = q_j] & = & 1 \end{array}$$

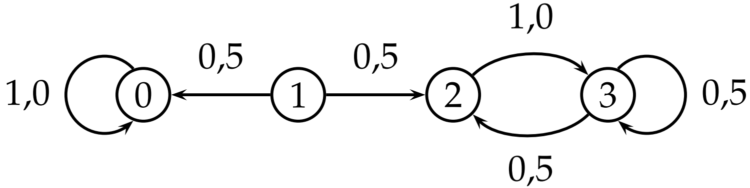
und damit

$$\begin{aligned} h_{ij} = \mathbb{E}[T_{ij}] \; &= \sum_{q_k \in Q} \mathbb{E}[T_{ij} \mid X_1 = q_k] \cdot p_{ik} \\ &= p_{ij} + \sum_{k \neq j} (1 + \mathbb{E}[T_{kj}]) \cdot p_{ik} = 1 + \sum_{k \neq j} h_{kj} \cdot p_{ik} \end{aligned}$$

Die Ableitung für h_j analog.

Berechnung von f_{ij} und h_{ij} IV

Beispiel 188



Für die Übergangszeiten für die Zustände 2 und 3 erhalten wir die Gleichungen

$$\begin{aligned} h_{23} &= 1 & \qquad h_2 &= 1 + h_{32} \\ h_{32} &= 1 + \tfrac{1}{2} h_{32} & \qquad h_3 &= 1 + \tfrac{1}{2} h_{23} \end{aligned}$$

mit Lösung

$$h_{23} = 1 \qquad h_{32} = 2 \qquad h_2 = 3 \qquad h_3 = 1.5$$

Berechnung von f_{ij} und h_{ij} V

Beispiel 189 (Zweite Frage von Armand)

Wenn Cécile mich verlässt, wie lange dauert es im Schnitt, bis sie zurückkommt?

Armand interessiert sich für h_{21} für die Kette mit

$$P = \begin{pmatrix} 0.8 & 0.2 \\ 0.1 & 0.9 \end{pmatrix}$$

Wir erhalten das Gleichungssystem

$$\begin{aligned} h_{12} &= 1 + p_{11}\, h_{12} = 1 + 0.8\, h_{12} & \qquad h_1 &= 1 + p_{12}\, h_{21} = 1 + 0.2\, h_{21} \\ h_{21} &= 1 + p_{22}\, h_{21} = 1 + 0.9\, h_{21} & \qquad h_2 &= 1 + p_{21}\, h_{12} = 1 + 0.1\, h_{12} \end{aligned}$$

mit Lösung

$$h_{12} = 5 \qquad h_{21} = 10 \qquad h_1 = 3 \qquad h_2 = 1.5$$

Beispiel 190

Cécile entscheidet, Armand und Bertrand sollen Poker spielen, bis einer von ihnen bankrott ist. Sie zieht dann endgültig beim Gewinner ein.

Armand und Bertrand verfügen jeweils über Kapital a und $m - a$.

In jeder Pokerrunde setzen beide jeweils eine Geldeinheit.

Armand gewinnt jedes Spiel mit W'keit p und Bertrand mit W'keit $q := 1 - p$.

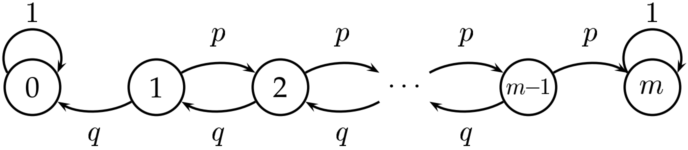
Frage 1: Mit welcher W'keit zieht Cécile bei Armand ein?

Frage 2: Wieviele Runden müssen im Schnitt gespielt werden?

Das Gamblers Ruin Problem II

Frage 1: Mit welcher W'keit zieht Cécile bei Armand ein?

Wir modellieren das Spiel durch die Markov-Kette



Die Zustände modellieren das aktuelle Kapital von Armand.

Die W'keit, mit der Armand Bertrand in den Ruin treibt is $f_{a,m}$.

Wir erhalten

$$
\begin{aligned}
f_{0,m} &= 0 \\
f_{i,m} &= p \cdot f_{i+1,m} + q \cdot f_{i-1,m} \qquad \text{für } 1 \leq i < m-1 \\
f_{m-1,m} &= p + q \cdot f_{m-2,m} \\
f_{m,m} &= 1
\end{aligned}
$$

Das Gamblers Ruin Problem III

Die Gleichungen können umgeschrieben werden zu

$$
\begin{aligned}
f_{0,m} &= 0 \\
f_{1,m} &= \xi \\
f_{i+1,m} &= (1/p) \cdot f_{i,m} - (q/p) \cdot f_{i-1,m} \qquad \text{für } 1 \leq i < m
\end{aligned}
$$

mit ξ so gewählt, dass $f_{m,m} = 1$ erfüllt ist.

Es ergibt sich für $p \neq 1/2$ (Fall $p = 1/2$ analog):

$$
f_{i,m} = \frac{p \cdot \xi}{2p-1} \cdot \left(1 - \left(\frac{1-p}{p}\right)^i\right) \qquad \xi = \frac{2p-1}{p \cdot \left(1 - \left(\frac{1-p}{p}\right)^m\right)}
$$

und insgesamt

$$
f_{a,m} = \frac{1 - \left(\frac{1-p}{p}\right)^a}{1 - \left(\frac{1-p}{p}\right)^m}
$$

Das Gamblers Ruin Problem IV

Frage 2: Wieviele Runden müssen im Schnitt gespielt werden?

Wir betrachten die Zufallsvariable

$$U_i := \text{„Anzahl der Schritte von } q_i \text{ nach } q_0 \text{ oder } q_m\text{“}$$

Mit $d_i := \mathbb{E}[U_i]$ gilt

$$\begin{aligned} d_0 &= 0 \\ d_i &= q \cdot d_{i-1} + p \cdot d_{i+1} + 1 \qquad \text{für } 1 \le i < m \\ d_m &= 0 \end{aligned}$$

Der Spezialfall $p = q = 1/2$ ist besonders einfach:

$$d_i = i \cdot (m - i) \qquad \text{für alle } i = 0, \ldots, m$$

womit unabhängig vom Startzustand das Spiel im Mittel nach höchstens m^2 Schritten beendet ist.

27. Stationäre Verteilung

Stationäre Verteilung I: Motivation

Die Übergangsmatrix der ABC-Kette erfüllt für alle $t \in \mathbb{N}$:

$$P^t = B \cdot D^t \cdot B^{-1} = \begin{pmatrix} -2 & 1 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0.7^t & 0 \\ 0 & 1^t \end{pmatrix} \cdot \begin{pmatrix} -1/3 & 1/3 \\ 1/3 & 2/3 \end{pmatrix}$$

Für $t \to \infty$ erhalten wir

$$\lim_{t\to\infty} P^t = \begin{pmatrix} -2 & 1 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} -1/3 & 1/3 \\ 1/3 & 2/3 \end{pmatrix} = \begin{pmatrix} 1/3 & 2/3 \\ 1/3 & 2/3 \end{pmatrix}$$

und so gilt für eine beliebige Anfangsverteilung $\mathcal{Q}_0 = (a, 1-a)$

$$\lim_{t\to\infty} \mathcal{Q}_t = \lim_{t\to\infty} \mathcal{Q}_0 \cdot P^t = (a, 1-a) \cdot \begin{pmatrix} 1/3 & 2/3 \\ 1/3 & 2/3 \end{pmatrix} = \left(\frac{1}{3}, \frac{2}{3}\right)$$

Stationäre Verteilung II: Motivation

Das System konvergiert also unabhängig von der Anfangsverteilung in die feste Verteilung $\pi = (1/3\,,\,2/3)$.

Intuitive Antwort auf Armands dritte Frage:

> *Die Verteilung der Zeit, die Cécile bei Armand und Bertrand verbringt, konvergiert gegen: 1/3 der Zeit bei Armand, 2/3 bei Bertrand.*

Offene Punkte:

(1) Konvergiert jede Kette in eine feste Verteilung unabhängig von der Anfangsverteilung?

(2) Wenn so, wie kann diese Verteilung berechnet werden?

(3) Stimmt die intuitive Antwort auf Armands dritte Frage? Wie kann die Frage überhaupt formalisiert werden?

Stationäre Verteilung III: Motivation

Wenn eine Kette immer in eine Verteilung π konvergiert, dann muss sie mit π als Anfangsverteilung „in π bleiben'".

Wir erwarten also

$$\pi \cdot P = \pi$$

d.h., π soll Eigenvektor von P zum Eigenwert 1 sein (bezüglich Multiplikation von links).

In der Tat gilt:

$$\pi \cdot P = (1/3,\ 2/3) \cdot \begin{pmatrix} 0.8 & 0.2 \\ 0.1 & 0.9 \end{pmatrix} = (1/3,\ 2/3) = \pi.$$

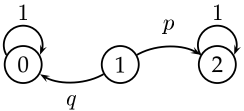
Stationäre Verteilung IV: Definition

Definition 191

Sei P die Übergangsmatrix einer Markov-Kette. Einen Zustandsvektor π mit $\pi = \pi \cdot P$ nennen wir stationäre Verteilung der Markov-Kette.

Umformulierung der Frage (a): Besitzen alle Markov-Ketten die Eigenschaft, dass sie unabhängig von der Anfangsverteilung in eine bestimmte stationäre Verteilung konvergieren?

Nein!



Diese Kette hat unendlich viele zwei stationäre Verteilungen:

$$(a, 0, 1-a) \quad \text{für alle } a \in [0, 1]$$

Stationäre Verteilung V: Irreduzible Ketten

Definition 192

Eine Markov-Kette heißt irreduzibel, wenn es für alle Zustandspaare q_i, q_j eine Zahl $n \in \mathbb{N}$ gibt, so dass $(P^n)_{ij} > 0$.

Wir bezeichen $p_{ij}^{(n)} := (P^n)_{ij}$.

Äquivalente Definition: Eine Markov-Kette heißt irreduzibel, wenn ihr Markov-Diagramm stark zusammenhängend ist.

Lemma 193

Für irreduzible endliche Markov-Ketten gilt für alle Zustände $q_i, q_j \in Q$*:*

(a) $f_{ij} = \Pr[T_{ij} < \infty] = 1$, *und*

(b) *der Erwartungswert* $h_{ij} = \mathbb{E}[T_{ij}]$ *existiert.*

Stationäre Verteilung VI: Irreduzible Ketten

Beweis:

Zu (b): Sei $q_k \in Q$ beliebig. Es gibt n_k mit $p_{kj}^{(n_k)} > 0$.

Sei $n := \max_k\{n_k\}$ und $p := \min_k\{p_{kj}^{(n_k)}\}$.

Wir unterteilen die Zeit in Phasen zu n Schritten.

Wir nennen eine Phase erfolgreich, wenn während dieser Phase ein Besuch bei q_j stattfindet.

Die Anzahl von Phasen bis zur ersten erfolgreichen Phase können wir durch eine geometrische Verteilung mit Parameter p abschätzen.

Die erwartete Anzahl von Phasen ist somit höchstens $1/p$.

Es folgt $h_{ij} \leq (1/p)\, n$ und $f_{ij} = \Pr[T_{ij} < \infty] = 1$. $\square$

Stationäre Verteilung VII: Irreduzible Ketten

Satz 194

Eine irreduzible endliche Markov-Kette besitzt eine eindeutige stationäre Verteilung π *und es gilt* $\pi(q_j) = 1/h_j$ *für alle* $q_j \in Q$.

Beweis:

(a) Z.z.: Es gibt einen Vektor $\pi \neq 0$ mit $\pi = \pi \cdot P$.

Sei $e := (1, \ldots, 1)^T$ und sei I die Einheitsmatrix.

Es gilt $P\,e = e$. (Einträge einer Zeile von P addieren sich zu 1).

Daraus folgt $0 = P\,e - e = (P - I)e$. Damit ist die Matrix $P - I$ singulär.

Es gibt also $\pi \neq 0$ mit $(P^T - I) \cdot \pi = 0$.

Stationäre Verteilung VIII: Irreduzible Ketten

(b) Z.z.: Wenn $\pi = \pi \cdot P$ für $\pi \neq 0$, dann $\pi(q_j) = 1/h_j$ für alle $q_j \in Q$.

Wir betrachten zwei Fälle:

Fall 1. $\displaystyle\sum_{q_i \in Q} \pi(q_i) \neq 0$.

O.B.d.A. sei $\sum_{q_i \in Q} \pi(q_i) = 1$.

Für jeden Zustand $q_j \in Q$ gilt (Lemma 193 und 187)

$$\begin{array}{rcll} \pi(q_i)h_{ij} & = & \pi(q_i)\left(1 + \displaystyle\sum_{k \neq j} p_{ik}h_{kj}\right) & \text{für } q_i \neq q_j \\ \pi(q_j)h_j & = & \pi(q_j)\left(1 + \displaystyle\sum_{k \neq j} p_{jk}h_{kj}\right) & \end{array}$$

Stationäre Verteilung IX: Irreduzible Ketten

Addition der Gleichungen ergibt

$$\pi(q_j)h_j + \sum_{q_i \neq q_j} \pi(q_i)h_{ij} = \sum_{q_i \in Q} \pi(q_i) + \sum_{q_i \in Q} \sum_{q_k \neq q_j} \pi(q_i)p_{ik}h_{kj}$$

Mit $\sum_{q_i \in Q} \pi(q_i) = 1$ folgt

$$\begin{aligned}
\pi(q_j)h_j + \sum_{q_i \neq q_j} \pi(q_i)h_{ij} &= 1 + \sum_{q_i \in Q} \sum_{q_k \neq q_j} \pi(q_i)p_{ik}h_{kj} \\
&= 1 + \sum_{q_k \neq q_j} h_{kj} \sum_{q_i \in Q} \pi(q_i)p_{ik} \\
&= 1 + \sum_{q_k \neq q_j} h_{kj}\pi(q_k)
\end{aligned}$$

und so $\pi(q_j)h_j = 1$.

Stationäre Verteilung X: Irreduzible Ketten

Fall 2. $\sum_{q_i \in Q} \pi(q_i) = 0$.

Dieselbe Rechnung ergibt nun

$$\begin{aligned} \pi(q_j)h_j + \sum_{q_i \neq q_j} \pi(q_i)h_{ij} &= 0 + \sum_{q_i \in Q} \sum_{q_k \neq q_j} \pi(q_i)p_{ik}h_{kj} \\ &= \sum_{q_k \neq q_j} h_{kj}\pi(q_k) \end{aligned}$$

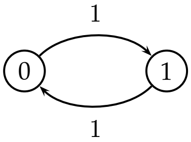
Es folgt $\pi(q_j) = 0$ für alle $q_j \in Q$, im Widerspruch zu $\pi \neq 0$.

Dieser Fall ist also eigentlich nicht möglich.

Stationäre Verteilung XI: Aperiodische Ketten

Auch wenn eine Markov-Kette eine eindeutige stationäre Verteilung besitzt, so muss sie nicht für jede Anfangsverteilung in diese Verteilung konvergieren.

Beispiel:



Als Anfangsverteilung nehmen wir $\pi_0 = (1,0)$ an. Es gilt:

$$\pi_t = \begin{cases} (1,0) & \text{falls } t \text{ gerade,} \\ (0,1) & \text{sonst.} \end{cases}$$

Die Kette pendelt also zwischen den beiden Vektoren $(1,0)$ und $(0,1)$ hin und her.

Die eindeutige stationäre Verteilung ist $(1/2, 1/2)$.

Stationäre Verteilung XII: Aperiodische Ketten

Definition 195

Die Periode eines Zustands q_j ist die größte Zahl $\xi \in \mathbb{N}$, so dass gilt:

$$\{n \in \mathbb{N}_0 \mid p_{jj}^{(n)} > 0\} \subseteq \{i \cdot \xi \mid i \in \mathbb{N}_0\}$$

Ein Zustand mit Periode $\xi = 1$ heißt aperiodisch.

Eine Markov-Kette ist aperiodisch, wenn alle Zustände aperiodisch sind.

Stationäre Verteilung XIII: Aperiodische Ketten

Lemma 196

Ein Zustand $q_i \in Q$ ist genau dann aperiodisch, wenn es ein $n_0 \in \mathbb{N}$ gibt mit $p_{ii}^{(n)} > 0$ für alle $n \geq n_0$.

Beweis:

($\Rightarrow$) Aus $p_{ii}^{(n_0)} > 0$ und $p_{ii}^{(n_0+1)} > 0$ folgt $\xi = 1$.

($\Leftarrow$) Wenn q_i aperiodisch ist, dann gibt es teilerfremde $a, b \in \mathbb{N}$ mit $p_{ii}^{(a)} > 0$ und $p_{ii}^{(b)} > 0$.

Ein bekannter Fakt der elementaren Zahlentheorie besagt: Da $a, b \in \mathbb{N}$ teilerfremd gibt es $n_0 \in \mathbb{N}$, so dass für alle $n \geq n_0$ es nichtnegative Zahlen $x, y \in \mathbb{N}_0$ gibt mit $n = xa + yb$.

Es folgt $p_{ii}^{(n)} > 0$ für alle $n \geq n_0$ und wir sind fertig.

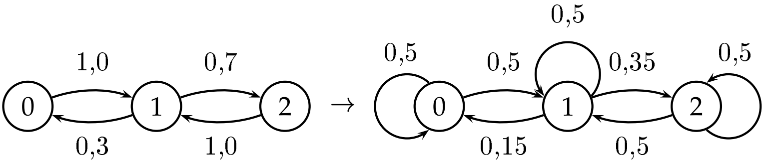
Stationäre Verteilung XIV: Aperiodische Ketten

$p_{ii}^{(n)} > 0$ gilt genau dann, wenn das Markov-Diagramm einen Pfad von q_i nach q_i der Länge n hat.

Es folgt: Wenn q_i eine Schleife hat (d.h. $p_{ii} > 0$ gilt), dann ist q_i aperiodisch.

Damit kann eine Kette folgendermaßen durch eine aperiodische Kette „simuliert“ werden:

- Füge an jedem Zustand eine Schleife an mit W'keit $1/2$.
- Halbiere die W'keiten an allen übrigen Kanten.



Bei irreduziblen Ketten genügt es, eine einzige Schleife einzuführen.

Stationäre Verteilung XV: Ergodische Ketten

Definition 197

Irreduzible, aperiodische Markov-Ketten nennt man ergodisch.

Satz 198 (Fundamentalsatz für ergodische Markov-Ketten)

Für jede ergodische endliche Markov-Kette $M = (Q, T, \delta, \mathcal{Q}_0)$ *gilt*

$$\lim_{n\to\infty} \mathcal{Q}_n = \pi,$$

wobei π *die eindeutige stationäre Verteilung von* M *bezeichnet.*

Bemerkung: π ist unabhängig von der Anfangsverteilung!

Stationäre Verteilung XVI: Ergodische Ketten

Beweis:

(Skizze.) Wir zeigen, dass für beliebige q_i, q_k gilt

$$\lim_{n\to\infty} p_{ik}^{(n)} = \pi_k.$$

Daraus folgt die Behauptung, da

$$\pi_n(q_k) = \sum_{q_i \in Q} \mathcal{Q}_0(q_i) \cdot p_{ik}^{(n)} \to \pi(q_k) \cdot \sum_{q_i \in Q} \mathcal{Q}_0(q_i) = \pi(q_k).$$

Wir betrachten das „Produkt“ zweier Kopien der Kette mit Zuständen (q_i, q_j) und Übergangsw'keiten

$$\delta((q_i, q_j), (q_{i'}, q_{j'})) = p_{ii'} \cdot p_{jj'}$$

Diese Produktkette ist ebenfalls ergodisch.

Stationäre Verteilung XVII: Ergodische Ketten

Sei H die Zufallsvariable, die die kleinste Zeit angibt, an die sich die Kette in einen Zustand der Gestalt (q, q) befindet.

Aus Lemma 193 und der Endlichkeit der Markov-Kette folgt

$$\Pr[H < \infty] = 1 \qquad \text{und} \qquad \mathbb{E}[H] < \infty$$

unabhängig von der Anfangsverteilung.

Seien X_t, Y_t Zufallsvariablen, die den Zustand der ersten bzw. der zweiten Komponente angeben.

Für ein festes t gilt $\Pr[X_t = q \mid t \geq H] = \Pr[Y_t = q \mid t \geq H]$ und somit auch

$$\Pr[X_t = q, t \geq H] = \Pr[Y_t = q, t \geq H].$$

Stationäre Verteilung XVIII: Ergodische Ketten

Wir wählen ein $q_i \in Q$ und setzen für die Anfangsverteilung $\mathcal{Q}_0$ der Produktkette

$$\mathcal{Q}_0(q, q') = \begin{cases} \pi(q') & \text{wenn } q = q_i \\ 0 & \text{sonst} \end{cases}$$

Intuition: die erste Komponente startet im Zustand q_i, die zweite startet (und bleibt) in der stationären Verteilung π.

Wir erhalten für alle $q_k \in Q$ und $n \geq 1$

$$\begin{aligned} |p_{ik}^{(n)} - \pi(q_k)| &= |\Pr[X_n = q_k] - \Pr[Y_n = q_k]| \\ &= |\Pr[X_n = q_k, n \geq H] + \Pr[X_n = q_k, n < H] \\ &\quad - \Pr[Y_n = q_k, n \geq H] - \Pr[Y_n = q_k, n < H]| \end{aligned}$$

Stationäre Verteilung XIX: Ergodische Ketten

Mit $\Pr[X_t = q, t \geq H] = \Pr[Y_t = q, t \geq H]$ gilt

$$|p_{ik}^{(n)} - \pi(q_k)| = |\Pr[X_n = q_k, n < H] - \Pr[Y_n = q_k, n < H]|$$

und wegen $|\Pr[A \cap B] - \Pr[A \cap C]| \leq \Pr[A]$ folgt

$$|p_{ik}^{(n)} - \pi(q_k)| \leq \Pr[n < H]$$

Da $\Pr[H < \infty] = 1$, gilt $\lim_{n \to \infty} \Pr[n < H] = 0$, d.h.

$$\lim_{n \to \infty} p_{ik}^{(n)} = \pi(q_k)$$

für alle $q_i, q_k \in Q$.

Stationäre Verteilung XX: Ergodische Ketten

Sei $q \in Q$ und $k \geq 0$. Seien X_q^k und B_q die Zufallsvariablen mit

$$X_q^k(\sigma) = \begin{cases} 1 & \text{falls } \sigma(k) = q \\ 0 & \text{sonst} \end{cases}$$

$$B_q(\sigma) = \begin{cases} \lim\limits_{n\to\infty} \dfrac{X_q^0 + \cdots + X_q^n}{n} & \text{wenn der Grenzwert existiert} \\ \perp & \text{sonst} \end{cases}$$

Satz 199 (Ergodischer Satz (ohne Beweis))

Für jeden Zustand q einer ergodischen endlichen Kette mit stationärer Verteilung π gilt

$$\Pr[B_q = \pi(q)] = 1 \ .$$

Beispiel 200 (Armands dritte Frage)

Wenn unsere ménàge a trois für immer so weiter geht, wieviel Prozent der Tage wird Cécile mit mir verbringen?

Armand fragt nach der Verteilung von B_{q_1}.

Der ergodische Satz zeigt, dass B_{q_1} den Wert $1/3$ mit W'keit 1 nimmt.

Cécile wird „auf langer Sicht“ mit W'keit 1 ein drittel der Tage mit Armand verbringen.