SS 2007

Diskrete Wahrscheinlichkeitstheorie

Javier Esparza

Fakultät für Informatik
TU München

http://www7.in.tum.de/um/courses/DS2/ss2007

Sommersemester 2007

Inhaltsverzeichnis

Kapitel 0 Organisatorisches

- Vorlesungen:
  - Di 14:30–16:00 (MI HSI), Do 8:15–9:45 (MI HS1)
    Pflichtvorlesung Grundstudium (Bachelor IN, Bioinformatik)
- Übung:
  - 1SWS Tutorübung
- Umfang:
  - 3V+1TÜ, 5 ECTS-Punkte
  - In den ersten Wochen: 4V
- Sprechstunde:
  - Nach Vereinbarung

- Vorkenntnisse:
  - Einführung in die Informatik I/II/III
  - Diskrete Strukturen I
- Weiterführende Vorlesungen:
  - Effiziente Algorithmen und Datenstrukturen
  - Randomisierte Algorithmen
  - Internetalgorithmik
  - ...
- Webseite:

  http://www7.in.tum.de/um/courses/DS2/ss2007/

- Ubungsleitung:
  - Dr. Stefan Schwoon
    Michael Luttenberger
- Sekretariat:
  - Frau Leber, MI 03.11.052 (leber@in.tum.de)

- Übungsaufgaben und Klausur:
  - Ausgabe jeweils am Donnerstag in der Vorlesung bzw. auf der Webseite der Vorlesung
  - Besprechung in der Tutorübung
- Klausur:
  - Zwei Tests (jeweils 10 Punkte) am 22. Mai 2007, 14:30-16:00 und am 26. Juni 2007, 14:30-16:00
  - Endklausur (45 Punkte)
  - Bestanden mit 30 Punkten
  - Wiederholungsklausur
  - bei den Klausuren sind keine Hilfsmittel außer einem handbeschriebenen DIN-A4-Blatt zugelassen

1. Vorlesungsinhalt

- Endliche Wahrscheinlichkeitsräume
  - Wahrscheinlichkeitsraum, Ereignis, Zufallsvariable
  - Spezielle Verteilungen
  - Ungleichungen von Markov und Chebyshev
- Unendliche Wahrscheinlichkeitsräume
  - Normalverteilung, Exponentialverteilung
  - Zentraler Grenzwertsatz
- Stochastische Prozesse
  - Markovketten
  - Warteschlangen
- Statistik
  - Schätzvariablen
  - Konfidenzintervalle
  - Testen von Hypothesen

2. Literatur

T. Schickinger, A. Steger:
Diskrete Strukturen - Band 2,
Springer Verlag 2001

N. Henze:
Stochastik für Einsteiger, 5. Auflage
Vieweg, 2004

M. Greiner, G. Tinhofer:
Stochastik fÃ¼r Informatiker,
Carl Hanser Verlag, 1996

H. Gordon:
Discrete Probability,
Springer-Verlag, 1997

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

3. Einleitung

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.*

- Pragmatischer 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.

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

Thema der Vorlesung: Zufallsexperimente

- Das Experiment wird unter genau festgelegten Versuchsbedingungen durchgeführt.
- Die Menge der moglichen 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. Was ist die W'keit, dass insgesamt über 12 Augenzahlen angezeigt werden?

- 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!

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

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

Kapitel I Diskrete Wahrscheinlichkeitsräume

1. Grundlagen

Definition 1

1. Ein diskreter Wahrscheinlichkeitsraum ist durch eine Ergebnismenge $\Omega = \{\omega_1, \omega_2, \ldots\}$ von Elementarereignissen gegeben.
2. Jedem Elementarereignis ω_i ist eine (Elementar-)Wahrscheinlichkeit $\Pr[\omega_i]$ zugeordnet, wobei wir fordern, dass $0 \leq \Pr[\omega_i] \leq 1$ und

$$\sum_{\omega \in \Omega} \Pr[\omega] = 1.$$

3. Eine Menge $E \subseteq \Omega$ heißt Ereignis. Die Wahrscheinlichkeit $\Pr[E]$ eines Ereignisses ist durch

$$\Pr[E] := \sum_{\omega \in E} \Pr[\omega]$$

definiert.

Beispiel 2

Zwei faire Würfel werden geworfen. Mit welcher W'keit beträgt die Gesamtzahl der Augen 10?

$$
\begin{aligned}
\Omega \quad = \{ \quad & (1,1),(1,2),(1,3),(1,4),(1,5),(1,6),\\
& (2,1),(2,2),(2,3),(2,4),(2,5),(2,6),\\
& (3,1),(3,2),(3,3),(3,4),(3,5),(3,6),\\
& (4,1),(4,2),(4,3),(4,4),(4,5),(4,6),\\
& (5,1),(5,2),(5,3),(5,4),(5,5),(5,6),\\
& (6,1),(6,2),(6,3),(6,4),(6,5),(6,6) \quad \}
\end{aligned}
$$

- Die Wahrscheinlichkeit $\Pr[(i,j)]$ eines jeden Elementarereignisses (i,j) ist $\frac{1}{36}$.

- Die Wahrscheinlichkeit $\Pr[E]$ des Ereignisses

$$E = \{\text{Die Gesamtzahl der Augen ist } 10\}$$

ist $\frac{1}{12}$.

Wir hätten aber auch folgendermaßen modellieren können:

$$\Omega = \{2, 3, 4, \dots, 10, 11, 12\}$$

Die Wahrscheinlichkeit der Elementarereignisse muss nun sorgfältig modelliert werden:

- $\Pr[2] = \frac{1}{36}$
- $\Pr[4] = \frac{1}{12}$
- $\Pr[7] = \frac{1}{6}$

Achtung: $\Pr(n) = \frac{1}{11}$ für alle $n \in \{2, \dots 12\}$ ist aus mathematischer Sicht zulässig, jedoch ein falsches Modell dieses Problems. Es entspricht nicht dem, was man unter "faire Würfel" versteht!

Mehrstufige Experimente:

- Experimente bestehen oft aus Teilexperimente, die der Reihe nach durchgeführt werden.
- Welches Teilexperiment als nächstes durchgeführt wird, kann vom Ausgang des vorigen Teils abhängen.
- Elementarereignisse = mögliche Sequenzen von "Teilereignissen".
- Daumenregel: die W'keit eines Elementarereignisses ist das Produkt der W'keiten der Teilereignisse.

Beispiel 3

Eine faire Münze wird geworfen. Zeigt sie Kopf, werden zwei faire Würfel geworfen. Zeigt sie Zahl, wird nur eine faire Wurfel geworfen. Was ist die W'keit, höchstens 4 Augen zu bekommen?

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

W'keiten der Elementarereignisse:

- $\Pr(\mathsf{k}\,(1,1)) = \frac{1}{2} \cdot \frac{1}{36}$
- $\Pr(\mathsf{k}\,(6,6)) = \frac{1}{2} \cdot \frac{1}{36}$
- $\Pr(\mathsf{z}\,3) = \frac{1}{2} \cdot \frac{1}{6}$

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

Beispiel 4

Eine faire Münze wird so lange geworfen, bis die gleiche Seite zweimal hintereinander fällt. Mit welcher W'keit ist die Anzahl der Würfe gerade?

Dann ist

$$\Omega = \{\mathsf{kk}, \mathsf{zz}, \mathsf{kzz}, \mathsf{zkk}, \mathsf{kzkk}, \mathsf{zkzz}\ldots\}$$

Frage: Was sind die W'keiten der einzelnen Elementarereignisse?

Es gilt:

$$\Pr(\mathsf{gerade}) = \Pr(\mathsf{kk}) + \Pr(\mathsf{zz}) + \Pr(\mathsf{kzkk}) + \ldots$$

Beispiel 5 (Forts.)

Ein zweites Modell: die Elementarereignisse sind die Anzahlen der Würfe:

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

Frage: Was sind die W'keiten der einzelnen Elementarereignisse?

Beispiel 5 (Forts.)

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

Beispiel 6 (Das Ziegenproblem)

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, sagen wir Nummer eins. 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, zum Beispiel Nummer drei, 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.

Beispiel 6 (Forts.: Modellierung des Ziegenproblems I)

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

Annahmen: 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 kann, dann wählt er eine Tür mit W'keit $1/2$.

Elementarereignis: (Auto-wird-versteckt
Erste-Wahl
Moderator-macht-Tür-auf
Zweite-Wahl
Ergebnis)

Beispiel 6 (Forts.: Modellierung des Ziegenproblems I)

- Experiment "Ich bleibe bei meiner ersten Wahl"

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

  $\Pr((1\,1\,2\,1\,A)) = \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{2} \quad \Pr((1\,2\,3\,1\,A)) = \frac{1}{3} \cdot \frac{1}{3}$
  $\Pr(\text{Auto}) = \frac{1}{3}$

- Experiment "Ich wechsle"

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

  $\Pr((1\,1\,2\,3\,Z)) = \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{2} \quad \Pr((1\,2\,3\,1\,A)) = \frac{1}{3} \cdot \frac{1}{3}$
  $\Pr(\text{Auto}) = \frac{2}{3}$

Beispiel 6 (Forts.: Modellierung des Ziegenproblems II)

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

Neue Annahmen: Das Auto wird mit W'keit $1/3$ hinter Tür 1, 2, oder 3 gestellt. Sie wählen Tür 1. Wenn der Moderator zwei Türen aufmachen kann, dann wählt er die Tür mit der kleinsten Nummer.

Beispiel 6 (Forts.: Modellierung des Ziegenproblems II)

- Modell “Ich bleibe bei meiner ersten Wahl”

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

$\Pr((1\,1\,2\,1\,A)) = \Pr((2\,1\,3\,1\,Z)) = \Pr((3\,1\,2\,1\,Z)) = \frac{1}{3}$
$\Pr(\text{Auto}) = \frac{1}{3}$

- Modell “Ich wechsle”

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

$\Pr((1\,1\,2\,3\,Z)) = \Pr((2\,1\,3\,2\,A)) = \Pr((3\,1\,2\,3\,A)) = \frac{1}{3}$
$\Pr(\text{Auto}) = \frac{2}{3}$

$\bar{E}$ heißt komplementäres Ereignis zu E.

Allgemein verwenden wir bei der Definition von Ereignissen alle bekannten Operatoren aus der Mengenlehre. Wenn also A und B Ereignisse sind, dann sind auch $A \cup B$, $A \cap B$, $A \setminus B$ etc. Ereignisse.

Zwei Ereignisse A und B heißen disjunkt oder auch unvereinbar, wenn $A \cap B = \emptyset$ gilt.

Beispiel 7

Wir beobachten die an einer Straße vorbeifahrenden Autos. Dabei gelte:

1. Es fahren doppelt so viele Autos von links nach rechts wie von rechts nach links.
2. Von zehn Autos sind acht silbergrau und zwei beige.

- Frage:: Welche W'keit hat das Ereignis "*Wir beobachten ein von links nach rechts fahrendes Auto*"?
- Frage: Welche W'keit hat das Ereignis "*Das nächste Auto ist beige und kommt von rechts*"?

Definition 8

Ein Wahrscheinlichkeitsraum mit $\Omega = \{\omega_1, \ldots, \omega_n\}$ heißt endlicher Wahrscheinlichkeitsraum.

Bei unendlichen Wahrscheinlichkeitsräumen werden wir gewöhnlich nur den Fall $\Omega = \mathbb{N}_0$ betrachten. Dies stellt keine große Einschränkung dar, da wir statt einer Ergebnismenge $\Omega = \{\omega_1, \omega_2, \ldots\}$ auch $\mathbb{N}_0$ als Ergebnismenge verwenden können, indem wir ω_i mit $i - 1$ identifizieren. Wir sagen, dass durch die Angabe der Elementarwahrscheinlichkeiten ein Wahrscheinlichkeitsraum auf Ω definiert ist.

Lemma 9

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

1. $\Pr[\emptyset] = 0$, $\Pr[\Omega] = 1$.
2. $0 \leq \Pr[A] \leq 1$.
3. $\Pr[\bar{A}] = 1 - \Pr[A]$.
4. *Wenn* $A \subseteq B$, *so folgt* $\Pr[A] \leq \Pr[B]$.

Lemma 9 (Forts.)

5. *(Additionssatz) Wenn die Ereignisse* $A_1, \ldots, A_n$ *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 disjunkte Ereignisse A, B *erhalten wir insbesondere*

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

Für eine unendliche Menge von disjunkten Ereignissen $A_1, A_2, \ldots$ *gilt analog*

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

Beweis:

Die Aussagen folgen unmittelbar aus Definition 1, den Eigenschaften der Addition und der Definition der Summe. $\square$

Eigenschaft 5 in Lemma 9 gilt nur für disjunkte Ereignisse. Für den allgemeinen Fall erhalten wir folgenden

Satz 10 (Siebformel, Prinzip der Inklusion/Exklusion)

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

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

Satz 10 (Forts.)

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

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

Beweis:

Wir betrachten zunächst den Fall $n = 2$. Dazu setzen wir $C := A \setminus B = A \setminus (A \cap B)$. Gemäß dieser Definition gilt, dass C und $A \cap B$ sowie C und B disjunkt sind. Deshalb können wir Eigenschaft 5 von Lemma 9 anwenden:

$$\Pr[A] = \Pr[C \cup (A \cap B)] = \Pr[C] + \Pr[A \cap B]\ .$$

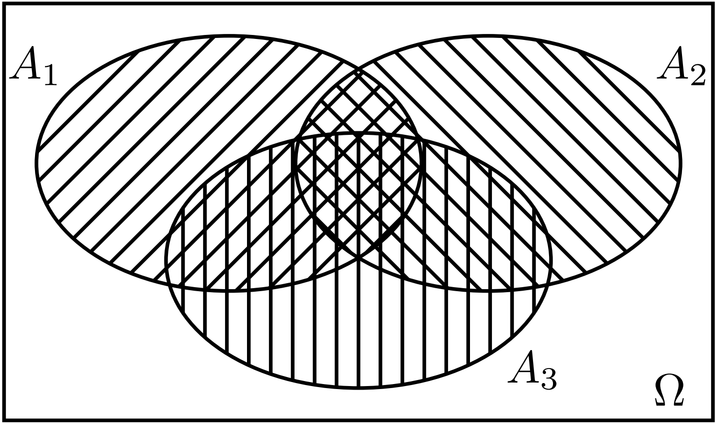
Wegen $A \cup B = C \cup B$ folgt daraus

$$\begin{aligned} \Pr[A \cup B] = \Pr[C \cup B] = \Pr[C] + \Pr[B] = \\ \Pr[A] - \Pr[A \cap B] + \Pr[B] \end{aligned}$$

und wir haben die Behauptung für $n = 2$ gezeigt.

Beweis (Forts.):

Der Fall $n = 3$:



Man beachte, dass durch die im Satz angegebene Summe jedes Flächenstück insgesamt genau einmal gezählt wird.

Beweis (Forts.):

Der allgemeine Fall kann nun durch Induktion über n gezeigt werden (was wir aber hier nicht ausführen!).

Satz 10 findet man manchmal auch unter der Bezeichung *Satz von Poincaré-Sylvester*, nach dem Franzosen

Henri PoincarÃ© (1854–1912)

und dem Engländer

James Joseph Sylvester (1814–1897)

benannt.

Boolesche Ungleichung:

Die folgende Abschätzung ist nach George Boole (1815–1864) benannt:

Korollar 11

Für Ereignisse $A_1, \ldots, A_n$ *gilt*

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

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

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

Beweis:

Zunächst betrachten wir die linke Seite der Ungleichung für den endlichen Fall und erhalten

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

Für die rechte Seite gilt

$$\sum_{i=1}^{n} \Pr[A_i] = \sum_{i=1}^{n} \sum_{\omega \in A_i} \Pr[\omega]\ .$$

Jedes Elementarereignis kommt links also genau einmal und rechts mindestens einmal vor. $\square$

1.1 Wahl der Wahrscheinlichkeiten

Frage: Wie können Wahrscheinlichkeiten sinnvoll festgelegt werden?

Prinzip von Laplace (Pierre Simon Laplace (1749–1827)):
Wenn nichts dagegen spricht, gehen wir davon aus, dass alle Elementarereignisse gleich wahrscheinlich sind.

Also:

$$\Pr[E] = \frac{|E|}{|\Omega|}$$

Was bedeutet "wenn nichts dagegen spricht"?
So was wie: Nach unserem Kenntnis der Gesetze (physikalische, psychologische, soziologische ...), die das Ergebnis des Experiments bestimmen, wird kein Elementarereignis "bevorzugt".

Und wenn etwas dagegen spricht?

Definition 12

$$\text{relative Häufigkeit von } E \;:=\; \frac{\text{absolute Häufigkeit von } E}{\text{Anzahl aller Beobachtungen}} \;=\; \frac{\text{Anzahl Eintreten von } E}{\text{Anzahl aller Beobachtungen}}.$$

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

Basiert auf das Prinzip:

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

1.2 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.

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. Bedingte Wahrscheinlichkeiten

Beispiel 13

A und B spielen Poker (52 Karten, 5 Karten pro Spieler, keine getauschten Karten).

A hält vier Asse und eine Herz Zwei in der Hand. B kann dieses Blatt nur überbieten, wenn er einen Straight Flush (fünf Karten *einer* Farbe in aufsteigender Reihenfolge) hat. Die Wahrscheinlichkeit für das Ereignis
$F :=$ „B hat einen Straight Flush" beträgt

$$\Pr[F] = \frac{|F|}{|\Omega|} = \frac{3 \cdot 8 + 7}{\binom{52-5}{5}} = \frac{31}{1533939} = 2{,}02.. \cdot 10^{-5}.$$

Beispiel 13 (Forts.)

A hat die Karten allerdings gezinkt und weiß, dass B nur Kreuz in der Hand hält. Bezeichne nun Ω' den Wahrscheinlichkeitsraum aller Möglichkeiten für B und F' das Ereignis, dass B einen Straight Flush der Farbe Kreuz hat:

$$\Pr[F'] = \frac{|F'|}{|\Omega'|} = \frac{8}{\binom{12}{5}} = \frac{8}{792} \approx 0{,}01 \text{ !!}$$

Wir sagen: Die W'keit, dass B einen Straight Flush hat, unter der Bedingung, dass B nur Kreuz in der Hand hält, beträgt $\frac{8}{792}$ und schreiben

$$\Pr[B \text{ hat SF} \mid B \text{ hat nur Kreuz}] = \frac{8}{792}$$

Definition 14

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

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

Eigenschaften:

1. $\Pr[B|B] = 1$;
2. $\Pr[A|\Omega] = \Pr[A]$;
3. für festes B ist $\Pr[A|B]$ proportional zu $\Pr[A \cap B]$.

Die bedingten Wahrscheinlichkeiten $\Pr[\cdot|B]$ bilden für ein beliebiges Ereignis $B \subseteq \Omega$ mit $\Pr[B] > 0$ einen neuen Wahrscheinlichkeitsraum über Ω.

Es ist leicht nachzurechnen, dass dadurch die Definition eines *diskreten Wahrscheinlichkeitsraums* erfüllt ist:

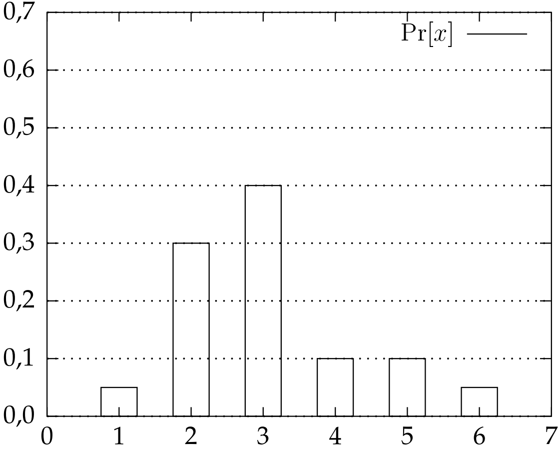
$$\sum_{\omega \in \Omega} \Pr[\omega|B] = \sum_{\omega \in \Omega} \frac{\Pr[\omega \cap B]}{\Pr[B]} = \sum_{\omega \in B} \frac{\Pr[\omega]}{\Pr[B]} = \frac{\Pr[B]}{\Pr[B]} = 1.$$

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

$$\Pr[\emptyset|B] = 0 \text{ sowie } \Pr[\bar{A}|B] = 1 - \Pr[A|B]\ .$$

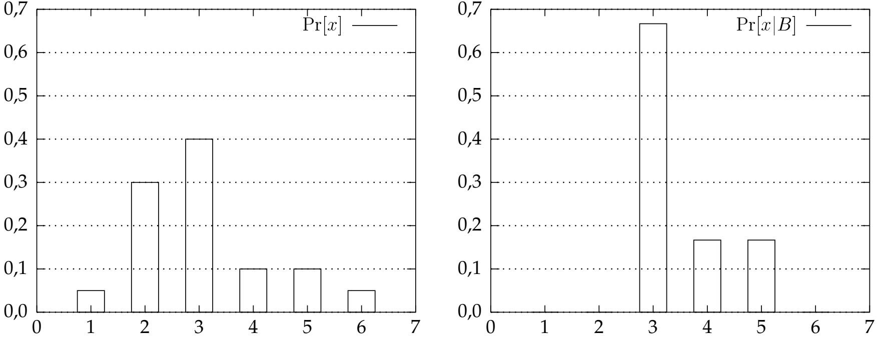
Beispiel 15 (Reskalierung bei bedingten Wahrscheinlichkeiten)

Betrachte folgenden gezinkten Würfel:



Beispiel 15 (Forts.)

Wir betrachten nun den durch $B := \{3, 4, 5\}$ gegebenen bedingten Wahrscheinlichkeitsraum:



Was genau war die Bedingung?

Beispiel 16 (Zweikinderproblem)

Wir nehmen an, dass bei der Geburt eines Kindes beide Geschlechter gleich wahrscheinlich sind. Wir wissen, dass eine bestimmte Familie zwei Kinder hat und eines davon ein Mädchen ist. Wie groß ist die Wahrscheinlichkeit, dass beide Kinder der Familie Mädchen sind?

Natürlich $\frac{1}{2}$.

Wirklich?

Beispiel 16 (Forts.)

Eigentlich gilt:

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

und

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

Wir bedingen auf M, und damit gilt für $A := \{mm\}$:

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

Häufig verwendet man die Definition der bedingten Wahrscheinlichkeit in der Form

$$\Pr[A \cap B] = \Pr[B|A] \cdot \Pr[A] = \Pr[A|B] \cdot \Pr[B]\,. \tag{1}$$

Damit:

Satz 17 (Multiplikationssatz)

Seien die Ereignisse $A_1, \ldots, A_n$ *gegeben. Falls* $\Pr[A_1 \cap \ldots \cap A_n] > 0$ *ist, gilt*

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

Beweis:

Zunächst halten wir fest, dass alle bedingten Wahrscheinlichkeiten wohldefiniert sind, 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 im Satz 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 18 (Geburtstagsproblem)

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

Umformulierung:
Man werfe m Bälle zufällig und gleich wahrscheinlich in n Körbe. Wie groß ist die Wahrscheinlichkeit, dass nach dem Experiment jeder Ball allein in seinem Korb liegt?

Für das Geburtstagsproblem: $n = 365$

Offensichtlich muss $m \leq n$ sein, damit überhaupt jeder Ball allein in einem Korb liegen kann.

Wir nehmen an, dass die Bälle nacheinander geworfen werden. A_i bezeichne das Ereignis „Ball i landet in einem noch leeren Korb". Das gesuchte Ereignis „Alle Bälle liegen allein in einem Korb" bezeichnen wir mit A. Nach Satz 17 können wir $\Pr[A]$ berechnen durch

$$\begin{aligned}
\Pr[A] &= \Pr\left[\cap_{i=1}^{m} A_i\right] \\
&= \Pr[A_1] \cdot \Pr[A_2|A_1] \cdot \ldots \cdot \Pr[A_m| \cap_{i=1}^{m-1} A_i].
\end{aligned}$$

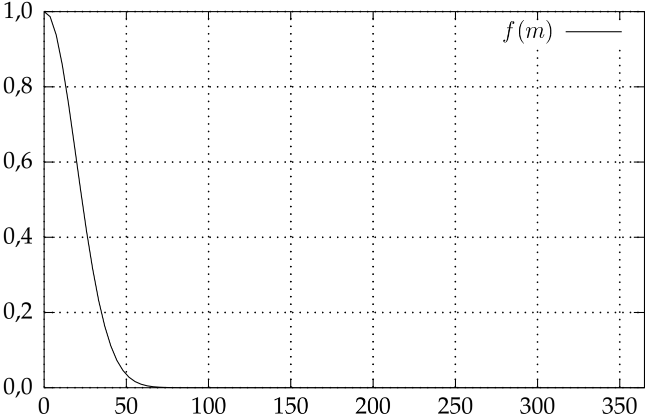
Unter der Bedingung, dass die ersten $j-1$ Bälle jeweils in einer leeren Urne gelandet sind, bedeutet A_j, dass der j-te Ball in eine der $n-(j-1)$ leeren Urnen fallen muss, die aus Symmetriegründen jeweils mit derselben Wahrscheinlichkeit gewählt werden.

Daraus folgt

$$\Pr[A_j | \cap_{i=1}^{j-1} A_i] = \frac{n-(j-1)}{n} = 1 - \frac{j-1}{n}.$$

Mit der Abschätzung $1 - x \leq e^{-x}$ und wegen $\Pr[A_1] = 1$ erhalten wir

$$\begin{aligned}
\Pr[A] &= \prod_{j=1}^{m} \left(1 - \frac{j-1}{n}\right) \\
&\leq \prod_{j=2}^{m} e^{-(j-1)/n} = e^{-(1/n)\cdot\sum_{j=1}^{m-1} j} \\
&= e^{-m(m-1)/(2n)} =: f(m)\ .
\end{aligned}$$



Verlauf von $f(m)$ für $n = 365$

Beispiel 19 (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 40jährigen 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.

Frage: Was ist die Wahrscheinlichkeit, daß die erste Wiederholung einer Gewinnreihe (13,983,816 verschiedene Sechserreihen) spätestens bei der 3016te Ausspielung auftritt?

Antwort: Ca. $1 - e^{-3016 \cdot 3015/27967632} \approx 1 - e^{-0.325} \approx 0,278$

Ausgehend von der Darstellung der bedingten Wahrscheinlichkeit in Gleichung 1 zeigen wir:

Satz 20 (Satz von der totalen Wahrscheinlichkeit)

Die Ereignisse $A_1, \ldots, A_n$ *seien paarweise disjunkt und es gelte* $B \subseteq A_1 \cup \ldots \cup A_n$*. Dann folgt*

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

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

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

Beweis:

Wir zeigen zunächst den endlichen Fall. Wir halten fest, dass

$$B = (B \cap A_1) \cup \ldots \cup (B \cap A_n)\ .$$

Da für beliebige i, j mit $i \neq j$ gilt, dass $A_i \cap A_j = \emptyset$, sind auch die Ereignisse $B \cap A_i$ und $B \cap A_j$ disjunkt. Wegen (1) folgt $\Pr[B \cap A_i] = \Pr[B|A_i] \cdot \Pr[A_i]$ (auch für den Fall, dass $\Pr[A_i] = 0$!). Wir wenden nun den Additionssatz (Lemma 5) an

$$\begin{aligned} \Pr[B] &= \Pr[B \cap A_1] + \ldots + \Pr[B \cap A_n] = \\ &\Pr[B|A_1] \cdot \Pr[A_1] + \ldots + \Pr[B|A_n] \cdot \Pr[A_n] \end{aligned}$$

und haben damit die Behauptung gezeigt. Da der Additionssatz auch für unendlich viele Ereignisse $A_1, A_2, \ldots$ gilt, kann dieser Beweis direkt auf den unendlichen Fall übertragen werden. $\square$

3. Unabhängigkeit

Bei einer bedingten Wahrscheinlichkeit $\Pr[A|B]$ kann der Fall auftreten, dass die Bedingung auf B, also das Vorwissen, dass B eintritt, keinen Einfluss auf die Wahrscheinlichkeit hat, mit der wir das Eintreten von A erwarten. Es gilt also $\Pr[A|B] = \Pr[A]$, und wir nennen dann die Ereignisse A und B unabhängig.

Beispiel 21 (Zweimaliges Würfeln)

$$\Omega := \{(i,j) \mid 1 \le i,j \le 6\} .$$

Alle Elementarereignisse erhalten die Wahrscheinlichkeit $\frac{1}{36}$ (faire Würfel) Wir definieren die Ereignisse

$$A := \text{Augenzahl im ersten Wurf ist gerade,}$$
$$B := \text{Augenzahl im zweiten Wurf ist gerade,}$$
$$C := \text{Summe der Augenzahlen beider Würfe beträgt 7.}$$

Es gilt $\Pr[A] = \Pr[B] = \frac{1}{2}$ und $\Pr[C] = \frac{1}{6}$. Wie groß ist $\Pr[B|A]$?

Beispiel 21 (Forts.)

Nach unserer Intuition beeinflusst der Ausgang des ersten Wurfs den zweiten Wurf nicht. Daher gewinnen wir durch das Eintreten von A keine Information in Bezug auf das Ereignis B hinzu:

$$B \cap A = \{(2,2),(2,4),(2,6),(4,2),(4,4),(4,6),(6,2),(6,4),(6,6)\}.$$

Daraus folgt

$$\Pr[B|A] = \frac{\Pr[B \cap A]}{\Pr[A]} = \frac{\frac{9}{36}}{\frac{1}{2}} = \frac{1}{2} = \Pr[B] \,.$$

Das Eintreffen des Ereignisses B hat mit dem Ereignis A „nichts zu tun“.

Definition 22

Die Ereignisse A und B heißen unabhängig, wenn gilt

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

Falls $\Pr[B] \neq 0$, so können wir diese Definition zu

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

umschreiben.

Beispiel 21 (Zweimaliges Würfeln, Forts.)

Zur Erinnerung:

$$A := \text{Augenzahl im ersten Wurf ist gerade,}$$
$$B := \text{Augenzahl im zweiten Wurf ist gerade,}$$
$$C := \text{Summe der Augenzahlen beider Würfe beträgt 7.}$$

Bei den Ereignissen A und B ist die Unabhängigkeit klar, da offensichtlich kein kausaler Zusammenhang zwischen den Ereignissen besteht. Wie steht es mit A und C?

$$A \cap C = \{(2,5),(4,3),(6,1)\}$$

und damit

$$\Pr[A \cap C] = \frac{3}{36} = \frac{1}{2} \cdot \frac{1}{6} = \Pr[A] \cdot \Pr[C] \text{ bzw. } \Pr[C|A] = \Pr[C]\,.$$

Beispiel 21 (Forts.)

Also sind auch A und C (und analog B und C) unabhängig.

Bemerkung: Im Beispiel ist $A \cap C \neq \emptyset$.
Es gilt sogar allgemein für zwei unabhängige Ereignisse A und B mit $\Pr[A], \Pr[B] > 0$, dass sie gar nicht disjunkt sein können, da ansonsten

$$0 = \Pr[\emptyset] = \Pr[A \cap B] \neq \Pr[A] \cdot \Pr[B] \, .$$

Beispiel 21 (Zweimaliges Würfeln (Forts.))

Zur Erinnerung:

$$A := \text{Augenzahl im ersten Wurf ist gerade,}$$
$$B := \text{Augenzahl im zweiten Wurf ist gerade,}$$
$$C := \text{Summe der Augenzahlen beider Würfe beträgt 7.}$$

Wir betrachten das Ereignis $A \cap B \cap C$. Wenn $A \cap B$ eintritt, so sind beide gewürfelten Augenzahlen gerade und somit ergibt auch die Summe davon eine gerade Zahl. Daraus folgt $\Pr[A \cap B \cap C] = 0$ bzw. $\Pr[C|A \cap B] = 0 \neq \Pr[C]$. Das Ereignis $A \cap B$ liefert uns also Information über das Ereignis C.

Definition 23

Die Ereignisse $A_1, \ldots, A_n$ heißen unabhängig, wenn für alle Teilmengen $I \subseteq \{1, \ldots, n\}$ mit $I = \{i_1, \ldots, i_k\}$ gilt, dass

$$\Pr[A_{i_1} \cap \ldots \cap A_{i_k}] = \Pr[A_{i_1}] \cdot \ldots \cdot \Pr[A_{i_k}]. \tag{2}$$

Eine unendliche Familie von Ereignissen A_i mit $i \in \mathbb{N}$ heißt unabhängig, wenn (2) für jede endliche Teilmenge $I \subseteq \mathbb{N}$ erfüllt ist.

Lemma 24

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*

$$\Pr[A_1^{s_1} \cap \ldots \cap A_n^{s_n}] = \Pr[A_1^{s_1}] \cdot \ldots \cdot \Pr[A_n^{s_n}], \tag{3}$$

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

Beweis:

Zunächst zeigen wir, dass aus (2) die Bedingung (3) folgt. Wir beweisen dies durch 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. Andernfalls gelte ohne Einschränkung $s_1 = 0$. Aus dem Additionssatz folgt dann

$$\begin{aligned}\Pr[\bar{A}_1 \cap A_2^{s_2} \cap \ldots \cap A_n^{s_n}] &= \Pr[A_2^{s_2} \cap \ldots \cap A_n^{s_n}] \\ &\quad -\Pr[A_1 \cap A_2^{s_2} \cap \ldots \cap A_n^{s_n}].\end{aligned}$$

Darauf können wir die Induktionsannahme anwenden und erhalten

$$\begin{aligned}&\Pr[\bar{A}_1 \cap A_2^{s_2} \cap \ldots \cap A_n^{s_n}] \\ &= \Pr[A_2^{s_2}] \cdot \ldots \cdot \Pr[A_n^{s_n}] - \Pr[A_1] \cdot \Pr[A_2^{s_2}] \cdot \ldots \cdot \Pr[A_n^{s_n}] \\ &= (1 - \Pr[A_1]) \cdot \Pr[A_2^{s_2}] \cdot \ldots \cdot \Pr[A_n^{s_n}],\end{aligned}$$

woraus die Behauptung wegen $1 - \Pr[A_1] = \Pr[\bar{A}_1]$ folgt.

Beweis (Forts.):

Für die Gegenrichtung zeigen wir nur, dass aus (3) $\Pr[A_1 \cap A_2] = \Pr[A_1] \cdot \Pr[A_2]$ folgt. Es gilt wegen des Satzes von der totalen Wahrscheinlichkeit, dass

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

und es folgt die Behauptung. $\square$

Aus der Darstellung in Lemma 24 folgt die wichtige Beobachtung, dass für zwei unabhängige Ereignisse A und B auch die Ereignisse $\bar{A}$ und B (und analog auch A und $\bar{B}$ bzw. $\bar{A}$ und $\bar{B}$) unabhängig sind!

Ebenso folgt:

Lemma 25

Seien A, B und C unabhängige Ereignisse. Dann sind auch $A \cap B$ und C bzw. $A \cup B$ und C unabhängig.

Beweis:

Zur Unabhängigkeit von $A \cap B$ und C siehe das vorangehende Beispiel.

Aus

$$\begin{aligned}
\Pr[(A \cup B) \cap C] &= \Pr[(A \cap C) \cup (B \cap C)] \\
&= \Pr[A \cap C] + \Pr[B \cap C] - \Pr[A \cap B \cap C] \\
&= \Pr[C] \cdot (\Pr[A] + \Pr[B] - \Pr[A \cap B]) \\
&= \Pr[A \cup B] \cdot \Pr[C]
\end{aligned}$$

folgt die Unabhängigkeit von $A \cup B$ und C. $\square$

4. Satz von Bayes

Satz 26 (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 paarweis disjunkte Ereignisse $A_1, A_2, \ldots$ *mit* $B \subseteq \bigcup_{i=1}^{\infty} A_i$*, dass*

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

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

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

Wir definieren: $B =$ 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.

Im Spezialfall, in dem die Ereignisse $S_1 \cap K_i, \ldots, S_m \cap K_i$ paarweise unabhängig sind für alle $1 \leq i \leq n$ erhalten wir

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

Durch Einsetzen im Satz von Bayes:

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

mit

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

Beispiel 27 (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.

Beispiel 0 (Forts.: *A priori* Wahrscheinlichkeiten)

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

Beispiel 0 (Forts.: 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

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.

5. Zufallsvariablen

5.1 Grundlagen

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

Definition 1

Sei ein Wahrscheinlichkeitsraum auf der Ergebnismenge Ω gegeben. Eine Abbildung

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

heißt (numerische) Zufallsvariable.

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

Bei diskreten Zufallsvariablen ist der Wertebereich

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

ebenfalls wieder endlich (bzw. abzählbar unendlich).

Beispiel 2

Wir werfen eine ideale Münze drei Mal. Als Ergebnismenge erhalten wir $\Omega := \{H, T\}^3$. Die Zufallsvariable Y bezeichne die Gesamtanzahl der Würfe mit Ergebnis „Head“.

Beispielsweise gilt also $Y(HTH) = 2$ und $Y(HHH) = 3$. Y hat den Wertebereich $W_Y = \{0, 1, 2, 3\}$.

Für $W_X = \{x_1, \ldots, x_n\}$ bzw. $W_X = \{x_1, x_2, \ldots\}$ betrachten wir (für ein beliebiges $1 \le i \le n$ bzw. $x_i \in \mathbb{N}$) das Ereignis

$$A_i := \{\omega \in \Omega;\ X(\omega) = x_i\} = X^{-1}(x_i).$$

Bemerkung: Anstelle von $\Pr[X^{-1}(x_i)]$ verwendet man häufig auch die Schreibweise $\Pr[„X = x_i“]$. Analog setzt man

$$\begin{aligned} \Pr[„X \le x_i“] &= \sum_{x \in W_X\,:\, x \le x_i} \Pr[„X = x“] \\ &= \Pr[\{\omega \in \Omega;\ X(\omega) \le x_i\}]\,. \end{aligned}$$

Oft lässt man auch die Anführungszeichen weg.

Definition 3

- Die Funktion

$$f_X : \mathbb{R} \ni x \mapsto \Pr[X = x] \in [0,1] \tag{4}$$

  nennt man (diskrete) Dichte(funktion) der Zufallsvariablen X.

- Die Funktion

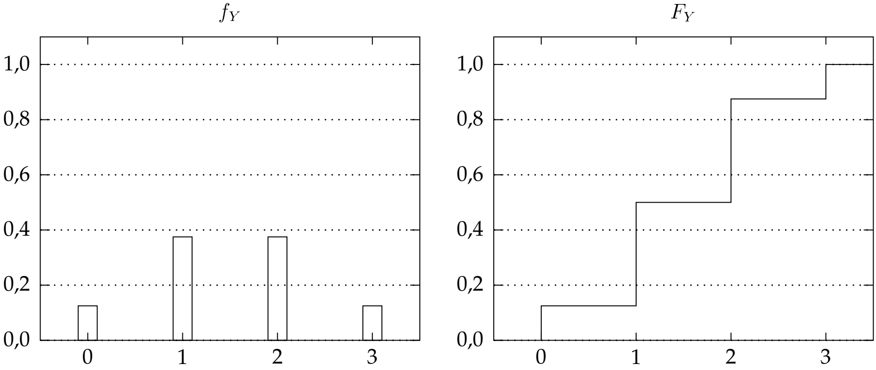
$$F_X : \mathbb{R} \ni x \mapsto \Pr[X \leq x] = \sum_{x \in W_X \,:\, x' \leq x} \Pr[X = x'] \in [0,1] \tag{5}$$

  heißt Verteilung(sfunktion) der Zufallsvariablen X.

Beispiel 4

Für die Zufallsvariable Y erhalten wir

$$\begin{aligned}
\Pr[Y=0] &= \Pr[TTT] = \frac{1}{8}, \\
\Pr[Y=1] &= \Pr[HTT] + \Pr[THT] + \Pr[TTH] = \frac{3}{8}, \\
\Pr[Y=2] &= \Pr[HHT] + \Pr[HTH] + \Pr[THH] = \frac{3}{8}, \\
\Pr[Y=3] &= \Pr[HHH] = \frac{1}{8}.
\end{aligned}$$



Dichte und Verteilung von Y

Bemerkung: Man kann statt Ω auch den zugrunde liegenden Wahrscheinlichkeitsraum über W_X betrachten.

5.2 Erwartungswert und Varianz

Definition 5

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.

Beispiel 6

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

Beispiel 7

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.

$$G := \begin{cases} k & \text{falls } k \text{ ungerade,} \\ -k & \text{falls } k \text{ gerade.} \end{cases}$$

Wie schon gesehen, gilt dann

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

Damit erhalten wir

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

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

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

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

$$G' := \begin{cases} 2^k & \text{falls } k \text{ ungerade,} \\ -2^k & \text{falls } k \text{ gerade ,} \end{cases}$$

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

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

Berechnung des Erwartungswerts:

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

Bei unendlichen Wahrscheinlichkeitsräumen ist dabei analog zur Definition des Erwartungswerts erforderlich, dass $\sum_{\omega \in \Omega} |X(\omega)| \cdot \Pr[\omega]$ konvergiert (absolute Konvergenz).

Satz 8 (Monotonie des Erwartungswerts)

Seien X und Y Zufallsvariablen über dem Wahrscheinlichkeitsraum Ω mit $X(\omega) \leq Y(\omega)$ für alle $\omega \in \Omega$. Dann gilt $\mathbb{E}[X] \leq \mathbb{E}[Y]$.

Beweis:

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

$\square$

Aus Satz 8 folgt insbesondere, dass $a \leq \mathbb{E}[X] \leq b$ gilt, wenn für die Zufallsvariable X die Eigenschaft $a \leq X(\omega) \leq b$ für alle $\omega \in \Omega$ erfüllt ist.

5.2.1 Rechenregeln für den Erwartungswert

Oft betrachtet man eine Zufallsvariable X nicht direkt, sondern wendet noch eine Funktion darauf an:

$$Y := f(X) = f \circ X\,,$$

wobei $f : \mathcal{D} \to \mathbb{R}$ eine beliebige Funktion sei mit $W_X \subseteq \mathcal{D} \subseteq \mathbb{R}$.

Beobachtung: $f(X)$ ist wieder eine Zufallsvariable.

Aus

$$\Pr[Y = y] = \Pr[\{\omega \mid f(X(\omega)) = y\}] \;\; = \sum_{x\,:\,f(x)=y} \Pr[X = x]$$

folgt

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

Satz 9 (Linearität des Erwartungswerts, einfache Version)

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

$$\mathbb{E}[a \cdot X + b] = a \cdot \mathbb{E}[X] + b\,.$$

Beweis:

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

$\square$

Satz 10

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:

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

$\square$

Definition 11

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

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

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

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

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

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

Satz 12

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.*

Beweis:

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

Der Beweis für den unendlichen Fall verläuft analog. $\square$

Beispiel 13

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

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

Beispiel 13

Andere Berechnungsmethode: (gestützt auf Satz 12)
Definiere das Ereignis

$$K_1 := \text{„Im ersten Wurf fällt Kopf"}\,.$$

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.

Beispiel 13

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 mit Satz 20:

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

Daraus ergibt sich wiederum $\mathbb{E}[X] = 1/p$.

Beispiel 13

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

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

5.2.2 Varianz

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.

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

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

Eine nahe liegende Lösung wäre,

$$\mathbb{E}[|X - \mu|]$$

zu berechnen, wobei $\mu = \mathbb{E}[X]$ sei. Dies scheitert jedoch meist an der „unhandlichen" Betragsfunktion. Aus diesem Grund betrachtet man stattdessen $\mathbb{E}[(X - \mu)^2]$, also die quadratische Abweichung vom Erwartungswert.

Definition 14

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.

Satz 15

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

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

$\square$

Beispiel 16

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

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

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

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

Satz 17

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 9 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 15:

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

und daraus zusammen die Behauptung. $\square$

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

Definition 18

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.

5.3 Mehrere Zufallsvariablen

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. Die Werte von X und Y hängen offensichtlich stark voneinander ab. Beispielsweise muss $Y = 0$ sein, wenn $X = 4$ gilt.

Wie kann man mit mehreren Zufallsvariablen über demselben Wahrscheinlichkeitsraum rechnen, auch wenn sie, wie im obigen Beispiel, sehr voneinander abhängig sind?

Wir untersuchen Wahrscheinlichkeiten der Art

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

Beispiel 20

Wenn wir nur die Zufallsvariable X betrachten, so gilt für $0 \leq x \leq 4$

$$\Pr[X = x] = \frac{\binom{4}{x}\binom{28}{10-x}}{\binom{32}{10}}\,.$$

Allgemein nennt man Zufallsvariablen mit der Dichte

$$\Pr[X = x] = \frac{\binom{b}{x}\binom{a}{r-x}}{\binom{a+b}{r}}$$

hypergeometrisch verteilt. Durch diese Dichte wird ein Experiment modelliert, bei dem r Elemente ohne Zurücklegen aus einer Grundmenge der Mächtigkeit $a + b$ mit b besonders ausgezeichneten Elementen gezogen werden.

Beispiel 20 (Forts.)

Die Zufallsvariable Y ist für sich gesehen ebenfalls hypergeometrisch verteilt mit $b = 4$, $a = 28$ und $r = 2$.

Für X und Y zusammen gilt jedoch z.B.

$$\Pr[X = 4, Y = 1] = 0,$$

und allgemein

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

Bemerkung: Die Schreibweise $\Pr[X = x, Y = y]$ stellt eine Abkürzung von $\Pr[„X = x \wedge Y = y“]$ dar. Ein anderes Beispiel ist

$$\Pr[X \le x, Y \le y_1, \sqrt{Y} = y_2]\,.$$

Die Funktion

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

heißt gemeinsame Dichte der Zufallsvariablen X und Y.

Aus der gemeinsamen Dichte $f_{X,Y}$ kann man ableiten

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

Die Funktionen f_X und f_Y nennt man 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.

Für zwei Zufallsvariablen definiert man die gemeinsame Verteilung

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

Die Randverteilung 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

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

5.3.1 Unabhängigkeit von Zufallsvariablen

Definition 21

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

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

Alternativ:

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

Bei unabhängigen Zufallsvariablen ist also die gemeinsame Dichte gleich dem Produkt der Randdichten. Ebenso gilt

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

Satz 22

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:

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

$\square$

Satz 23

$f_1, \ldots, f_n$ 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 $S_i = \{x;\ f(x) = z_i\}$.

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

$\square$

5.3.2 Zusammengesetzte Zufallsvariablen

Beispiel 24

Ein Würfel werde zweimal geworfen. X bzw. Y bezeichne die Augenzahl im ersten bzw. zweiten Wurf. Sei $Z := X + Y$ die Summe der gewürfelten Augenzahlen.

Für Z gilt z.B.:
$\Pr[Z = 1] = \Pr[\emptyset] = 0$,
$\Pr[Z = 4] = \Pr[\{(1,3),(2,2),(3,1)\}] = \frac{3}{36}$.

Für die Verteilung der Summe zweier unabhängiger Zufallsvariablen gilt der folgende Satz:

Satz 25

Für zwei unabhängige Zufallsvariablen X und Y sei $Z := X + Y$. 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 folgt, dass

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

$\square$

Den Ausdruck $\sum_{x \in W_X} f_X(x) \cdot f_Y(z - x)$ aus Satz 25 nennt man in Analogie zu den entsprechenden Begriffen bei Potenzreihen auch Faltung oder Konvolution der Dichten f_X und f_Y.

Beispiel (Forts.)

Berechne 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

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

Und für $7 < z \leq 12$:

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

5.3.3 Momente zusammengesetzter Zufallsvariablen

Satz 26 (Linearität des Erwartungswerts)

Für Zufallsvariablen $X_1, \ldots, X_n$ *und* $X := a_1 X_1 + \cdots + a_n X_n$ *mit* $a_1, \ldots, a_n \in \mathbb{R}$ *gilt*

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

Beweis:

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

$\square$

Beispiel 27

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). 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 $X := X_1 + \cdots + X_n$.

Beispiel 27

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 hat also nur ein Seemann sein eigenes Bett aufgesucht.

Satz 28 (Multiplikativität des Erwartungswerts)

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

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

Beweis:

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

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

□

Dass für die Gültigkeit von Satz 28 die Unabhängigkeit der Zufallsvariablen wirklich notwendig ist, sieht man beispielsweise am Fall $Y = -X$ für eine Zufallsvariable mit einer von Null verschiedenen Varianz. Dann gilt

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

Satz 29 (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:

Wir betrachten nur den Fall $n = 2$ mit den 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 15 folgt die Behauptung. $\square$

Für abhängige Zufallsvariablen $X_1, \ldots, X_n$ gilt Satz 29 im Allgemeinen nicht. Als Beispiel funktioniert wiederum der Fall $X = -Y$:

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

Definition 30

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.

Beobachtung:
Für die Indikatorvariable I_A gilt nach Definition

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

Ebenso gilt

$$\mathbb{E}[I_{A_1} \cdot \ldots \cdot I_{A_n}] = \Pr[A_1 \cap \ldots \cap A_n],$$

da das Produkt von Indikatorvariablen genau dann gleich 1 ist, wenn alle entsprechenden Ereignisse eintreten.

Beispiel (Forts.)

Wir betrachten wieder das Beispiel der total betrunkenen Matrosen.

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 für beliebige* $i, j \in \{1, \ldots, n\}$, $i \neq j$*:*

$$\mathbb{E}[X_i X_j] = \mathbb{E}[I_{A_i} I_{A_j}] = \Pr[A_i \cap A_j] = \frac{1}{n(n-1)},$$

sowie

$$\mathbb{E}[X_i^2] = 0^2 \cdot \Pr[\bar{A}_i] + 1^2 \cdot \Pr[A_i] = \Pr[A_i] = 1/n.$$

Beispiel (Forts.)

Daraus folgt wegen der Linearität des Erwartungswerts für $X = X_1 + \cdots + X_n$:

$$\begin{aligned}\mathbb{E}[X^2] &= \mathbb{E}\left[\sum_{i=1}^{n} X_i^2 + \sum_{i=1}^{n} \sum_{j \neq i} X_i X_j\right] \\ &= n \cdot \frac{1}{n} + n(n-1) \cdot \frac{1}{n(n-1)} = 2\,.\end{aligned}$$

Für die Varianz erhalten wir somit den Wert

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

Einfacher Beweis für Satz 10 mit Hilfe von Indikatorvariablen:

Zur Erinnerung:

Satz 10 (Siebformel, Prinzip der Inklusion/Exklusion)

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

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

Beweis:

Zur Erinnerung: Zu Ereignissen $A_1, \ldots, A_n$ wollen wir die Wahrscheinlichkeit $\Pr[B]$ des Ereignisses $B := A_1 \cup \ldots \cup A_n$ ermitteln.

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

Das Produkt $\prod_{i=1}^n (1 - I_i)$ ist genau dann gleich 1, wenn $I_1 = \ldots = I_n = 0$, d.h. wenn B nicht eintritt. Somit gilt $I_{\bar{B}} = \prod_{i=1}^n (1 - I_i)$ und wir erhalten:

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

Beweis:

Wegen der Eigenschaften von Indikatorvariablen gilt

$$\Pr[B] = 1 - \Pr[\bar{B}] = 1 - \mathbb{E}[I_{\bar{B}}].$$

Mit Hilfe von Satz 26 und Satz 28 „verteilen“ wir den Erwartungswert auf die einzelnen Produkte von Indikatorvariablen. Wenn wir nun $\mathbb{E}[I_i]$ durch $\Pr[A_i]$ und allgemein $\mathbb{E}[I_{i_1} \cdot \ldots \cdot I_{i_k}]$ durch $\Pr[A_{i_1} \cap \ldots \cap A_{i_k}]$ ersetzen, haben wir Satz 10 (noch einmal) bewiesen. $\square$

6. Formelsammlung I

6.1 Gesetze zum Rechnen mit Ereignissen

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.

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

$\forall i \neq j : A_i \cap A_j = \emptyset \implies$ $\Pr\left[\bigcup_{i=1}^{n} A_i\right] = \sum_{i=1}^{n} \Pr[A_i]$	Additionssatz
$\Pr[A \cup B] = \Pr[A] + \Pr[B] - \Pr[A \cap B]$ allgemeine Form: siehe Satz 10	Inklusion/Exklusion, Siebformel
$\Pr\left[\bigcup_{i=1}^{n} A_i\right] \leq \sum_{i=1}^{n} \Pr[A_i]$	Boolesche Ungleichung
$\Pr[A\|B] = \frac{\Pr[A \cap B]}{\Pr[B]}$ für $\Pr[B] > 0$	Def. bedingte Ws.

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

6.2 Erwartungswert und Varianz diskreter Zufallsvariablen

Sei X eine diskrete Zufallsvariable. Für Erwartungswert und Varianz gelten die folgenden Formeln (sofern $\mathbb{E}[X]$ und $\mathrm{Var}[X]$ existieren).

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

Erwartungswert

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

Varianz

6.3 Gesetze zum Rechnen mit Zufallsvariablen

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	
$\mathbb{E}[a \cdot X + b] = a \cdot \mathbb{E}[X] + b$	

$X(\omega) \leq Y(\omega)$ für alle $\omega \in \Omega \implies \mathbb{E}[X] \leq \mathbb{E}[Y]$	Monotonie des Erwartungswerts
$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X \mid A_i] \cdot \Pr[A_i]$	
$\operatorname{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$	
$\operatorname{Var}[a \cdot X + b] = a^2 \cdot \operatorname{Var}[X]$	

$\mathbb{E}[a_1 X_1 + \ldots + a_n X_n]$ $= a_1 \mathbb{E}[X_1] + \ldots + a_n \mathbb{E}[X_n]$	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

6.4 Wichtige Reihen

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

7. Wichtige diskrete Verteilungen

Wir diskutieren nun einige wichtige *diskrete* Verteilungen. Bei diesen Verteilungen handelt es sich um Funktionen, die von gewissen *Parametern* abhängen. Eigentlich betrachten wir also immer eine ganze Familie von ähnlichen Verteilungen.

7.1 Bernoulli-Verteilung

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

$$f_X(x) = \begin{cases} p & \text{für } x = 1, \\ 1 - p & \text{für } x = 0. \end{cases}$$

heißt Bernoulli-verteilt. Den Parameter p nennen wir Erfolgswahrscheinlichkeit.

Eine solche Verteilung erhält man z.B. bei einer einzelnen Indikatorvariablen. Es gilt mit $q := 1 - p$

$$\mathbb{E}[X] = p \text{ und } \mathrm{Var}[X] = pq,$$

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

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.

7.2 Binomialverteilung

Eine Bernoulli-verteilte Zufallsvariable entspricht der Verteilung *einer* Indikatorvariablen. Häufig betrachtet man jedoch Summen von Indikatorvariablen.

Definition 31

Sei $X := X_1 + \ldots + X_n$ als Summe von n unabhängigen, Bernoulli-verteilten Zufallsvariablen mit gleicher Erfolgswahrscheinlichkeit p definiert. Dann heißt X binomialverteilt mit den Parametern n und p. In Zeichen schreiben wir

$$X \sim \mathrm{Bin}(n, p)\,.$$

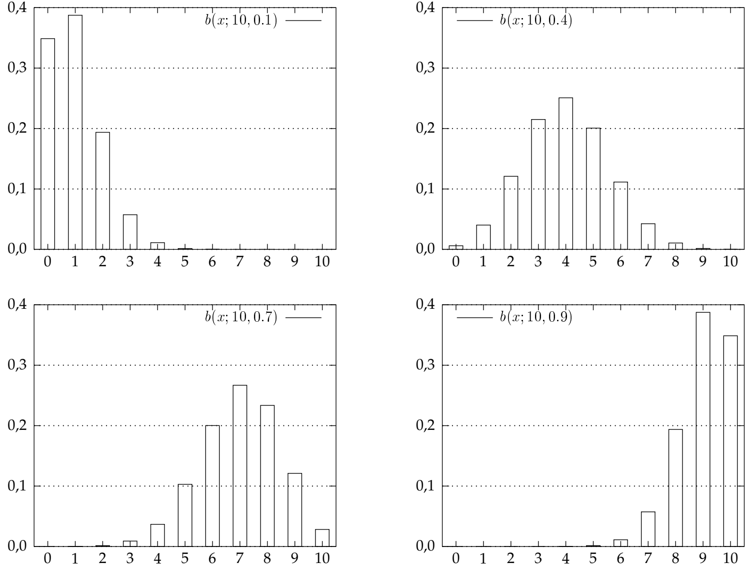
Es gilt $W_X = \{0, \ldots, n\}$. Die Binomialverteilung besitzt die Dichte

$$f_X(x) := b(x; n, p) = \binom{n}{x} p^x q^{n-x}$$

mit $q := 1 - p$. Da die Binomialverteilung eine sehr wichtige Rolle spielt, führen wir für die Dichtefunktion die Abkürzung $b(x; n, p)$ ein.

Mit den Sätzen über Erwartungswert und Varianz von Summen unabhängiger Zufallsvariablen erhalten wir sofort

$$\mathbb{E}[X] = np \quad \text{und} \quad \mathrm{Var}[X] = npq\,.$$



Dichte der Binomialverteilung

Satz 32

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

Beweis:

Die Aussage folgt sofort, wenn man gemäß der Definition der Binomialverteilung X und Y als Summen von Indikatorvariablen darstellt. Z ist dann offensichtlich wieder eine Summe von unabhängigen Indikatorvariablen. $\square$

7.3 Geometrische Verteilung

Definition 33

Die Dichte der geometrischen Verteilung mit Parameter/Erfolgswahrscheinlichkeit $p \in [0,1]$ und $q := 1 - p$ ist gegeben durch

$$f_X(i) = pq^{i-1} \quad \text{für } i \in \mathbb{N}\,.$$

Für Erwartungswert und Varianz geometrisch verteilter Zufallsvariablen gilt

$$\mathbb{E}[X] = \frac{1}{p} \quad \text{und} \quad \mathrm{Var}[X] = \frac{q}{p^2},$$

denn es gilt:

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

$\mathbb{E}[X^2]$ ergibt sich gemäß der Formel (siehe DS I)

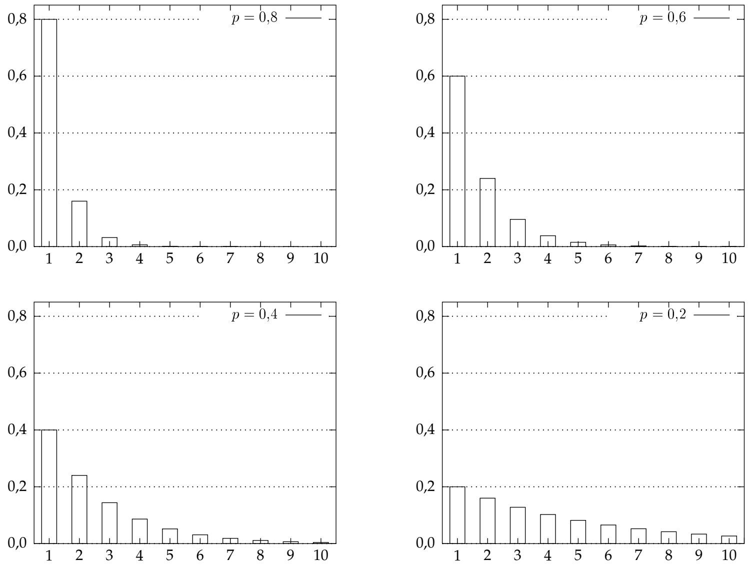
$$\sum_{n\geq 0}\binom{c+n-1}{n}z^n = \frac{1}{(1-z)^c} = (1-z)^{-c}$$

zu

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

und damit

$$\mathrm{Var}[X] = \frac{q}{p^2}\,.$$



Dichte der geometrischen Verteilung

Sei X geometrisch verteilt mit Erfolgswahrscheinlichkeit p. Dann ist $\Pr[X = k]$ die Wahrscheinlichkeit, dass wir bei einem binären Experiment mit Erfolgswahrscheinlichkeit p genau in der k-ten unabhängigen Wiederholung das erste Mal erfolgreich sind.

Wie groß ist die Wahrscheinlichkeit $\Pr[X > y + x \mid X > x]$?

Da bei den ersten x Versuchen kein Erfolg eintrat, stellen wir uns vor, dass das „eigentliche" Experiment erst ab dem $(x+1)$-ten Versuch beginnt. Die Zeit bis zum ersten Erfolg bei diesem neuen Experiment nennen wir X'. Damit $X > y + x$ gilt, muss $X' > y$ gelten. Es ist intuitiv, dass X' wieder geometrisch verteilt ist mit Erfolgswahrscheinlichkeit p, dass also für $x, y \in \mathbb{N}$ gilt:

$$\Pr[X > y + x \mid X > x] = \Pr[X' > y]. \tag{6}$$

Formal gilt

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

sowie

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

Diese Eigenschaft nennt man *Gedächtnislosigkeit*, da eine geometrisch verteilte Zufallsvariable gewissermaßen vergisst, dass sie schon x Misserfolge hinter sich hat und sich deshalb zum Zeitpunkt $y + x$ genauso verhält wie ursprünglich zur Zeit y.

Warten auf den n-ten Erfolg.
Wir betrachten n unabhängige Zufallsvariablen $X_1, \ldots, X_n$, die jeweils geometrisch verteilt sind mit Parameter p, und bestimmen die Dichte der Zufallsvariablen $Z := X_1 + \cdots + X_n$. Damit bezeichnet Z also die Anzahl der Versuche bis zum n-ten erfolgreichen Experiment (einschließlich).

Falls $Z = z$ ist, so werden also genau n erfolgreiche und $z - n$ nicht erfolgreiche Experimente durchgeführt. Dafür gibt es genau $\binom{z-1}{n-1}$ Möglichkeiten, von denen jede mit Wahrscheinlichkeit $p^n(1-p)^{z-n}$ eintritt. Es gilt also

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

Die Zufallsvariable Z nennt man negativ binomialverteilt mit Ordnung n.

Das Coupon-Collector-Problem

In manchen Branchen legen Firmen den Verpackungen ihrer Produkte oft kleine Bilder oder andere Gegenstände bei, um den Käufer zum Sammeln anzuregen. Wenn es insgesamt n verschiedene solche Beilagen gibt, wie viele Packungen muss man im Mittel erwerben, bis man eine vollständige Sammlung besitzt? Hierbei nehmen wir an, dass bei jedem Kauf jede Beilage mit gleicher Wahrscheinlichkeit auftritt.

Sei

- X die Anzahl der zu tätigenden Käufe, und
- bezeichne Phase i die Schritte vom Erwerb der $(i-1)$-ten Beilage (ausschließlich) bis zum Erwerb der i-ten Beilage (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.

Somit ist X_i geometrisch verteilt mit Parameter $p = \frac{n-i+1}{n}$ und es gilt $\mathbb{E}[X_i] = \frac{n}{n-i+1}$.

Damit folgt aber sofort

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

wobei $H_n := \sum_{i=1}^{n} \frac{1}{i}$ die n-te harmonische Zahl bezeichnet. Da $H_n = \ln n + O(1)$, folgt $\mathbb{E}[X] = n \ln n + O(n)$.

7.4 Poisson-Verteilung

Die Poisson-Verteilung mit dem Parameter λ hat den Wertebereich $W_X = \mathbb{N}_0$ und besitzt die Dichte

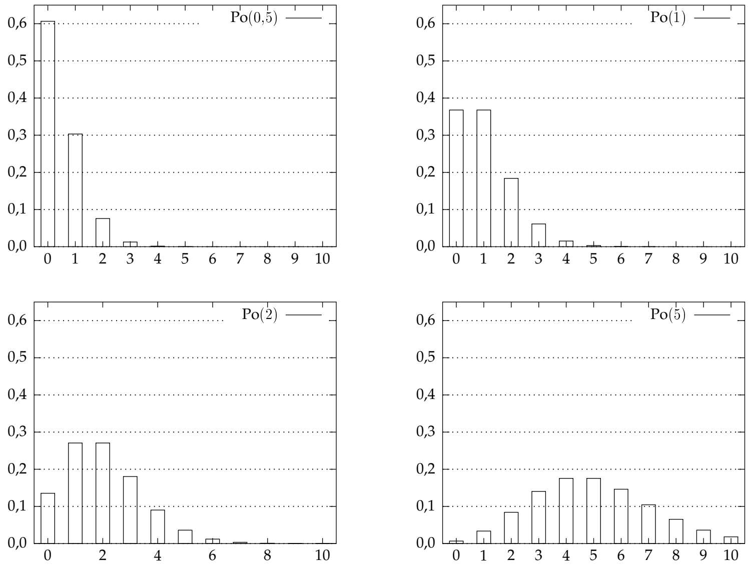
$$f_X(i) = \frac{e^{-\lambda}\lambda^i}{i!} \quad \text{für } i \in \mathbb{N}_0.$$

f_X ist eine zulässige Dichte, da

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

Dafür, dass eine Zufallsvariable X Poisson-verteilt mit Parameter λ ist, schreiben wir auch

$$X \sim \mathrm{Po}(\lambda).$$



Dichte der Poisson-Verteilung

Poisson-Verteilung als Grenzwert der Binomialverteilung

Wir betrachten eine Folge von binomialverteilten Zufallsvariablen X_n mit $X_n \sim \mathrm{Bin}(n, p_n)$, wobei $p_n = \lambda/n$. Für ein beliebiges k mit $0 \le k \le n$ ist die Wahrscheinlichkeit, dass X_n den Wert k annimmt, gleich

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

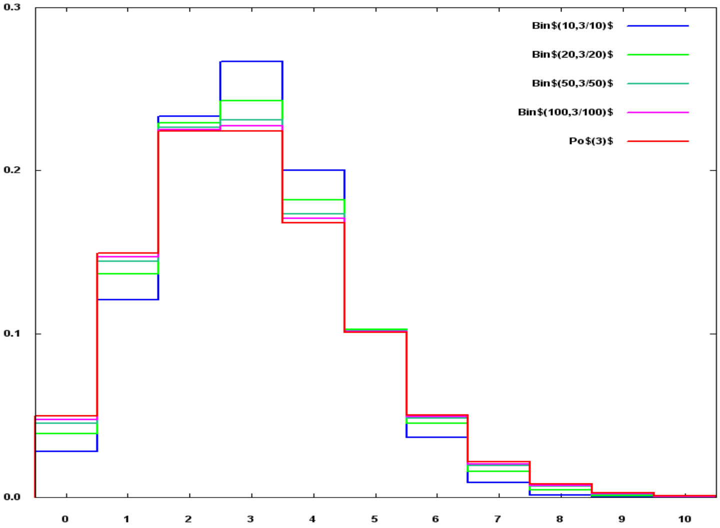
Wir betrachten nun $n \to \infty$ und erinnern uns, dass

$$\lim_{n\to\infty} \frac{n^{\underline{k}}}{n^k} = 1,$$
$$\lim_{n\to\infty} (1 - \frac{\lambda}{n})^{-k} = 1, \text{ und}$$
$$\lim_{n\to\infty} (1 - \frac{\lambda}{n})^{n} = e^{-\lambda}\,.$$

Damit folgt

$$\lim_{n\to\infty} b(k; n, p_n) = \lim_{n\to\infty} \binom{n}{k} \cdot p_n^k \cdot (1 - p_n)^{n-k} = e^{-\lambda} \cdot \frac{\lambda^k}{k!}.$$

Die Wahrscheinlichkeit $b(k; n, p_n)$ konvergiert also für $n \to \infty$ gegen die Wahrscheinlichkeit, dass eine Poisson-verteilte Zufallsvariable mit Parameter λ den Wert k annimmt. Insgesamt folgt somit, dass die Verteilung einer Zufallsvariablen $X \sim \mathrm{Bin}(n, \lambda/n)$ sich für $n \to \infty$ der Poisson-Verteilung $\mathrm{Po}(\lambda)$ annähert.



Vergleich von Binomial- und Poisson-Verteilung

Ist also n im Vergleich zu λ hinreichend groß, so kann man die Poisson-Verteilung als Approximation der Binomialverteilung verwenden.

Diese Tatsache wird manchmal auch als Gesetz seltener Ereignisse bezeichnet, da die Wahrscheinlichkeit eines einzelnen Treffers $p_n = \lambda/n$ relativ klein sein muss, wenn die Approximation gute Ergebnisse liefern soll.

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 34

Beim Kundenservice einer Firma rufen im Durschnitt k Kunden pro Tag an. Wir nehmen an, dass die Kunden zu jedem Zeitpunkt mit der selben W'keit anrufen.

Wir betrachten ein diskretes Modell, in dem der Tag in n gleich lange Zeitintervallen unterteilt wird (jeweils $24/n$ Stunden lang). Unter der Annahme, dass höchstens ein Kunde pro Zeitintervall anruft, ergibt sich für die Anzahl X der Anrufe an einem Tag:

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

Beispiel 34 (Forts.)

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

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

Für eine Poisson-verteilte Variable X mit $\lambda = 3$ erhalten wir $\Pr[X \leq 5] = 0.9161$

Beispiel 35

Wir wollen wissen, wie oft eine bestimmte Gegend im Durchschnitt von einer Naturkatastrophe (z.B. Vulkanausbruch) getroffen wird. Aus Statistiken entnehmen wir, dass so ein Ereignis im Mittel 10^{-4}-mal pro Jahr auftritt. Wir interessieren uns nun für die Wahrscheinlichkeit, dass die Region in einem Jahr mehr als einmal von einem solchen Unglück heimgesucht wird.

Die Voraussetzungen scheinen erfüllt zu sein, die Anzahl X der Katastrophen durch eine Poisson-Verteilung mit Parameter $\lambda = 10^{-4}$ zu modellieren.

Damit gilt

$$\begin{aligned}\Pr[X \geq 2] &= 1 - \Pr[X = 0] - \Pr[X = 1] = 1 - e^{-\lambda} - \lambda e^{-\lambda} \\ &\approx 1 - 0{,}999900005 - 0{,}000099990 = 5 \cdot 10^{-9}.\end{aligned}$$

Erwartungswert und Varianz

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

Da

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

und

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

folgt

$$\mathrm{Var}[X] = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - \mathbb{E}[X]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda. \quad (7)$$

Summe von Poisson-verteilten Zufallsvariablen

Satz 36

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

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

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 $p := \frac{\lambda}{\lambda+\mu}$.

Da die Summe gleich 1 ist, folgt

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

$\square$

8. Abschätzen von Wahrscheinlichkeiten

8.1 Die Ungleichungen von Markov und Chebyshev

Satz 37 (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:

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

Beweis:

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

$\square$

Alternativer Beweis:

Es gilt

$$E[X] = E[X|X < t]Pr[X < t] + E[X|X \geq t]Pr[X \geq t]\,.$$

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

$$E[X] \geq t * Pr[X \geq t]\,.$$

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.).

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

Satz 38 (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:

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

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:

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

$\square$

Beispiel 39

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

X ist binomialverteilt mit $X \sim \mathrm{Bin}(1000, p = \frac{1}{2})$,also gilt

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

Wie groß ist die Wahrscheinlichkeit, dass mehr als 550-mal „Kopf“ fällt?

Beispiel 39

Chebyshev-Ungleichung:

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

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

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

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

8.2 Gesetz der großen Zahlen

Wir haben diskutiert, wie Wahrscheinlichkeiten als Grenzwerte von relativen Häufigkeiten aufgefasst werden können.

Satz 40 (Gesetz der großen Zahlen)

Gegeben sei eine Zufallsvariable X*. Ferner seien* $\varepsilon, \delta > 0$ *beliebig aber fest. Dann gilt für alle* $n \geq \frac{\mathrm{Var}[X]}{\varepsilon\delta^2}$*:*

Sind $X_1, \ldots, X_n$ *unabhängige Zufallsvariablen mit derselben Verteilung wie* X *und setzt man*

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

so gilt

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

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

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

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. $\square$

Wahrscheinlichkeit und relative Häufigkeit.

Sei X eine Indikatorvariable für ein Ereignis A, $\Pr[A] = p$. 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 beliebig nahe an die „wahre“ Wahrscheinlichkeit p an.

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

$$\text{relative Abweichung } |\tfrac{1}{n}\textstyle\sum_i X_i - p|$$

und nicht die

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

abschätzt!

8.3 Chernoff-Schranken

8.3.1 Chernoff-Schranken für Summen von 0–1–Zufallsvariablen

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 41

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 gilt für* $X := \sum_{i=1}^n X_i$ *und* $\mu := \mathbb{E}[X] = \sum_{i=1}^n p_i$*, sowie jedes* $\delta > 0$*, dass*

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

Beweis:

Für $t > 0$ gilt

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

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

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

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

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

Beweis (Forts.):

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

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

Damit wird

$$f(t) = \frac{e^{(e^t-1)\mu}}{e^{t(1+\delta)\mu}} = \frac{e^{\delta\mu}}{(1+\delta)^{(1+\delta)\mu}}\,.$$

$\square$

Beispiel 42

Wir betrachten wieder das Beispiel, dass wir eine faire Münze n-mal werfen und abschätzen wollen, mit welcher Wahrscheinlichkeit „Kopf“

$$\frac{n}{2}(1 + 10\%)$$

oder öfter fällt.

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

Satz 43

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 gilt für $X := \sum_{i=1}^n X_i$ und $\mu := \mathbb{E}[X] = \sum_{i=1}^n p_i$, sowie jedes $0 < \delta < 1$, dass

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

Beweis:

Analog zum Beweis von Satz 41. $\square$

Bemerkung: Abschätzungen, wie sie in Satz 41 und Satz 43 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 (vergleiche Satz 41) und vom lower tail (vergleiche Satz 43).

Die Chernoff-Schranken hängen exponentiell von μ ab!

Lemma 44

Für $0 \le \delta < 1$ *gilt*

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

Beweis:

Wir betrachten

$$f(x) = (1-x)\ln(1-x) \text{ und } g(x) = -x + \frac{1}{2}x^2\,.$$

Es gilt für $0 \le x < 1$:

$$g'(x) = x - 1 \le -\ln(1-x) - 1 = f'(x)$$

sowie

$$f(0) = 0 = g(0)\,,$$

also im angegebenen Intervall $f(x) \ge g(x)$.

Die Ableitung der zweiten Ungleichung erfolgt analog. □

Korollar 45

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* $t \geq 2e\mu$.

Beweis:

1 und 2 folgen direkt aus Satz 41 bzw. 43 und Lemma 44.

Aus 1 und 2 zusammen folgt 3.

Die Abschätzung 4 erhalten wir direkt aus Satz 41, da für den Zähler gilt

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

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

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

□

Beispiel 46

Wir betrachten wieder balls into bins und werfen n Bälle unabhängig und gleichverteilt in n Körbe. Sei

$$X_i := \text{Anzahl der Bälle im } i\text{-ten Korb}$$

für $i = 1, \ldots, n$, sowie $X := \max_{1 \le i \le n} X_i$.

Für die Analyse von X_i ($i \in \{1, \ldots, n\}$ beliebig) verwenden wir Aussage 5 von Korollar 45, mit $p_1 = \ldots = p_n = \frac{1}{n}$, $\mu = 1$ und $t = 2 \log n$. Es folgt

$$\Pr[X_i \ge 2 \log n] \le 1/n^2.$$

Daraus ergibt sich

$$\Pr[X \ge 2 \log n] = \Pr[X_1 \ge 2 \log n \vee \ldots \vee X_n \ge 2 \log n] \le n \cdot \frac{1}{n^2} = \frac{1}{n}.$$

Es gilt also mit Wahrscheinlichkeit $1 - 1/n$, dass $X < 2 \log n$ ist.

Literatur:

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

9. Erzeugende Funktionen

9.1 Einführung

Definition 47

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

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

Eine wahrscheinlichkeitserzeugende Funktion ist also die (gewöhnliche) erzeugende Funktion der Folge $(f_i)_{i \in \mathbb{N}_0}$ mit $f_i := \Pr[X = i]$.

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

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

Beobachtung:
Sei $Y := X + t$ mit $t \in \mathbb{N}_0$. Dann gilt

$$G_Y(s) = \mathbb{E}[s^Y] = \mathbb{E}[s^{X+t}] = \mathbb{E}[s^t \cdot s^X] = s^t \cdot \mathbb{E}[s^X] = s^t \cdot G_X(s)\,.$$

Ebenso lässt sich leicht nachrechnen, dass

$$\begin{aligned} G'_X(s) &= \sum_{k=1}^{\infty} k \cdot \Pr[X = k] \cdot s^{k-1}, \text{ also} \\ G'_X(0) &= \Pr[X = 1], \text{ sowie} \\ G_X^{(i)}(0) &= \Pr[X = i] \cdot i!, \text{ also} \\ G_X^{(i)}(0)/i! &= \Pr[X = i]\,. \end{aligned}$$

Satz 48 (Eindeutigkeit der w.e. Funktion)

Die Dichte und die Verteilung einer Zufallsvariablen X *mit* $W_X \subseteq \mathbb{N}$ *sind durch ihre wahrscheinlichkeitserzeugende Funktion eindeutig bestimmt.*

Beweis:

Folgt aus der Eindeutigkeit der Potenzreihendarstellung. $\square$

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

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

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

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

Poisson-Verteilung

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

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

Beispiel 49

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 wahrscheinlichkeitserzeugenden Funktion die Konvergenz der Verteilung folgt.

9.1.1 Zusammenhang zwischen der w.e. Funktion und den Momenten

Da

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

gilt

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

Beispiel 50

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

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

Beispiel 50

Ebenso ergibt sich

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

also etwa

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

Andere Momente von X kann man auf ähnliche Art und Weise berechnen.