SS 2013

Diskrete Wahrscheinlichkeitstheorie

Javier Esparza

Fakultät für Informatik
TU München

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

Sommersemester 2013

Teil I

Einleitung

Vorlesungsinhalt

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

Literatur I

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

Literatur II

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

Teil II

Kontinuierliche Wahrscheinlichkeitsräume

1. Kontinuierliche Wahrscheinlichkeitsräume

Informelle Einführung I

Wir suchen nach einem mathematischen Modell für dieses Zufallsexperiment:

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

Wir erwarten z.B.:

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

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

Lösung: Physikalische Analogie

Informelle Einführung II: Physikalische Analogie

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

Ausgedehnter (eindimensionaler) Körper:

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

Informelle Einführung III: Physikalische Analogie

In einem kontinuierlichen W'keitsraum:

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

Informelle Einführung IV: Beispiele

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

$f(x) = 1/3$ für alle $x \in [0,3]$.

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

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

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

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

Informelle Einführung V: Beispiele

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

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

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

Informelle Einführung VI: Probleme

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

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

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

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

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

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

W'keitsräume I

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

Einige Anforderungen an $\Pr$:

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

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

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

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

W'keitsräume II: Satz von Vitali

Satz 1 (Vitali)

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

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

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

W'keitsräume III

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

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

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

Was bedeutet "gut"?

Antwort (Kolmogorov):

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

$\Pr$ soll die Kolmogorov-Axiome erfüllen.

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

W'keitsräume IV: σ-Algebren

Definition 2

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

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

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

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

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

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

W'keitsräume V: Kolmogorov-Axiome und Definition

Definition 3 (Kolmogorov-Axiome, Wahrscheinlichkeitsraum)

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

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

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

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

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

W'keitsräume VI: Einfache Eigenschaften

Aus der Definition folgt:

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

W'keitsräume VII: Ausblick auf die nächsten Folien

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

Satz 4

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

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

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

In den kommenden Folien führen wir diese Begriffe ein.

Intervalle

Definition 5

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

Offene und halboffene Intervalle werden analog definiert.

Borel'sche Mengen I

Definition 6

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

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

Borel'sche Mengen II

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

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

Lemma 7

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

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

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

Definition 8

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

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

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i).$$

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

Satz 9 (Caratheodory)

Sei Ω ein Intervall. Es gibt eine und nur eine Maßfunktion $\mu\colon \mathcal{B}(\Omega) \to \mathbb{R}^n$ *mit*

$$\mu\left(\prod_{i=1}^{n}[a_i, b_i]\right) = \prod_{i=1}^{n}(b_i - a_i)\ .$$

Intuition: Jede Borel'sche Menge hat eine wohldefinierte Länge, Fläche, Volumen, ...

- $\mu(\{x\}) = 0$ für jedes $x \in \Omega$.
- $\mu(\mathbb{Q}) = 0$, denn $\mathbb{Q}$ ist abzählbar.
- $\mu([0,1] \setminus \mathbb{Q}) = 1$.

Wir suchen nach Dichtefunktionen mit der Eigenschaft

> Für jeden Dichtwert ist das Gesamtgewicht der Punkte mit dieser Dichte wohldefiniert.

Noch allgemeiner:

> Für jede Borel'sche Menge M von Dichtwerten ist das Gesamtgewicht der Punkte mit Dichte in M wohldefiniert.

Lebesgue Integral II: Messbare Funktionen

Definition 10

Sei Ω ein Intervall. Eine Funktion $f : \Omega \to \mathbb{R}$ heißt messbar, falls das Urbild jeder Borel'schen Menge ebenfalls eine Borel'sche Menge ist.

- Indikatorfunktionen von Borel'schen Mengen sind messbar.
- Stetige Funktionen sind messbar.
- Summen und Produkte von messbaren Funktionen sind wiederum messbar.
- Die Verkettung messbarer Funktionen ist messbar.
- Der punktweise Grenzwert einer Folge von messbaren Funktionen ist wiederum messbar.

Lebesgue-Integral III: Informelle Definition

Jeder messbaren Funktion $f\colon \Omega \to \mathbb{R}_0^+$ kann ein Integral, das so genannte Lebesgue-Integral, geschrieben $\int f \, \mathrm{d}x$, zugeordnet werden.

$\int f \, \mathrm{d}x$ ist eine Funktion $\mathcal{B}(\Omega) \to \mathbb{R}$, die jedem Borel'schen Menge A eine reelle Zahl $\displaystyle\int_A f \, \mathrm{d}x$ zuordnet.

Informelle Idee für die Definition von $\displaystyle\int_A f \, \mathrm{d}x$ (nicht korrekt!):

- Partitioniere $\mathbb{R}_0^+$ in Subintervallen $[0, a_1], [a_1, a_2] \ldots [a_n, \infty)$.
- Für jedes $[a_i, a_{i+1}]$, berechne $a_i \cdot \mu(f^{-1}(\,[a_i, a_{i+1}]\,) \cap A)$

  ($a_i \times$ Maß der Punkte x mit $f(x) \in [a_i, a_{i+1}]$)
- Addiere die Ergebnisse.
- Nehme den Grenzwert wenn $n \to \infty$.

Lebesgue-Integral IV: Bezug zur Riemann-Integral

Satz 11

Sei $A = [a, b] \subseteq \Omega$ *mit* $a, b \in \mathbb{R}$ *und sei* $f\colon A \to \mathbb{R}$ *beschränkt (d.h., es gibt* $c \in \mathbb{R}$ *mit* $f(x) \leq c$ *für alle* $x \in A$*).*
Wenn $\int_a^b f(x)dx$ *existiert, dann gilt*

$$\int_A f \, \mathsf{d}x = \int_a^b f(x)dx$$

Der Satz gilt auch für Intervalle $(-\infty, b]$, $[a, \infty)$, oder $(-\infty, \infty)$ und kann auch auf mehrere Dimensionen erweitert werden.

W'keitsdichte

Definition 12

Sei $\Omega \subseteq \mathbb{R}^n$ ein Intervall. Eine W'keitsdichte über Ω ist eine messbare Funktion $f : \Omega \to \mathbb{R}$ mit $\int_\Omega f \,\mathrm{d}x = 1$.

Physikalische Analogie: jede Borel'sche Menge hat ein wohldefiniertes Gewicht und Ω wiegt 1 Kg.

Satz 13

Sei $\Omega \subseteq \mathbb{R}^n$ ein Intervall und sei $f : \Omega \to \mathbb{R}$ eine W'keitsdichte. Die Abbildung $\Pr\colon \mathcal{B}(\Omega) \to \mathbb{R}_0^+$ mit $\Pr[A] = \int_A f \cdot dx$ erfüllt die Kolmogorov-Axiome und die Triple $(\Omega, \mathcal{B}(\Omega), \Pr)$ bildet einen W'keitsraum.

W'keitsräume: Zusammenfassung

Der Ansatz der informellen Einführung ist ausreichend für fast alle Beispiele dieser Vorlesung .

Es hat jedoch Probleme, insbesondere für Fragen, die nicht-terminierenden Zufallsexperimenten betreffen.

Die formale Definition von W'keitsraum erlaubt eine fundierte Betrachtung dieser Fragen.

Wir haben die formale Definition nur für den Fall betrachtet, in dem Ω ein Intervall ist. Sie kann jedoch viel allgemeiner formuliert werden.

Bertrand'sches Paradoxon I

In diskreten W'keisträume gibt es nur eine mögliche Modellierung von „alle Ausgänge sind gleichwahrscheinlich".

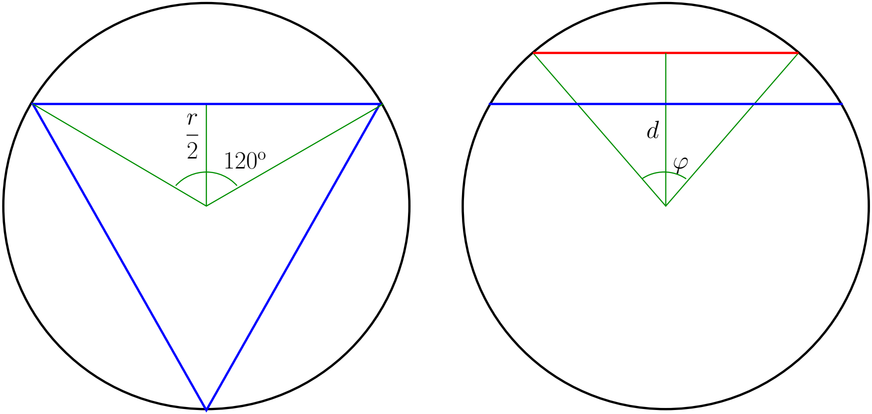
Im kontinuierlichen Fall kann es verschiedene Möglichkeiten geben, „gleichwahrscheinliche Ausgänge" zu modellieren, die zu verschiedenen Ergbenissen führen!

Beispiel 14 (Bertrand'sches Paradoxon)

Wir betrachten einen Kreis von Radius r mit einem eingeschriebenen gleichseitigen Dreieck (Spitze nach unten).
Wir wählen eine waagerechte Sehne des oberen Halbkreises zufällig.
Alle diese Sehnen seien "gleichwahrscheinlich".

Frage: Was ist die Wahrscheinlichkeit, mit der die Länge der Sehne die Seitenlänge des Dreiecks übersteigt?

Bertrand'sches Paradoxon II



Bertrand'sches Paradoxon II

Die Seiten des Dreiecks haben Abstand $\frac{r}{2}$ vom Mittelpunkt.

Die Lage der Sehne ist

- durch den Abstand d zum Kreismittelpunkt, oder
- durch den Winkel φ mit dem Kreismittelpunkt

eindeutig festgelegt.

Modell 1: $\Omega = [0, r]$, W'keitsdichte von d: $f(x) = 1/r$.

A tritt ein, wenn $d < \frac{r}{2}$, und es folgt $\Pr[A] = \frac{1}{2}$.

Modell 2: $\Omega = [0, 180]$, W'keitsdichte von φ: $f(x) = 1/180$.

A tritt ein, wenn $\varphi \in]120°, 180°]$, und es folgt somit $\Pr[A] = \frac{1}{3}$.

2. (Kontinuierliche) Zufallsvariablen

Beispiel 15

Wir vergleichen zwei Probleme:

Problem I: Eine Zahl $x \in \{1, \ldots, 10\}$ wird zufällig gewählt.
Alle Zahlen haben die gleiche W'keit, gewählt zu werden.
Wenn x gewählt wird, dann gewinnt man x^2 Euro.

Frage: Mit welcher W'keit gewinnt man mindestens 9 Euro?

Problem II: Eine reelle Zahl $x \in [1, 10]$ wird zufällig gewählt.
Alle Zahlen haben die gleiche W'keit, gewählt zu werden.
Wenn x gewählt wird, dann gewinnt man x^2 Euro.

Frage: Mit welcher W'keit gewinnt man mindestens 9 Euro?

Zufallsvariablen II: Informelle Einführung

Modell I: $\Omega = \{1, \dots, 10\}$ $\Pr[i] = 1/10$.

Zufallsvariable $G \colon \Omega \to \mathbb{R}$ mit $G(\omega) = \omega^2$.

$\Pr[G \leq 9] = \Pr[\{1, 2, 3\}] = 3/10$.

Modell II: $\Omega = [1, 10]$ W'keitsdichte $f(x) = 1/9$.

Zufallsvariable $G \colon \Omega \to \mathbb{R}$ mit $G(\omega) = \omega^2$.

$$\Pr[G \leq 9] = \Pr[\ [1, 3]\ ] = \int_1^3 \frac{1}{9}\,\mathsf{d}x = 2/9$$

Und wenn „$G \leq 9$“ keine Borel'sche Menge wäre ?

Zufallsvariablen III: Definition, Verteilung, Dichte

Definition 16

Sei $(\Omega, \mathcal{A}, \Pr)$ ein W'keitsraum. Eine kontinuierliche Zufallsvariable ist eine messbare Abbildung $X : \Omega \to \mathbb{R}$.

Eine Zufallsvariable $X : \Omega \to \mathbb{R}$ ist stetig wenn es eine W'keitsdichtefunktion $f_X : \mathbb{R} \to \mathbb{R}_0^+$ (die Dichte von X) gibt mit

$$\Pr[X \in A] = \int_A f_X(x)\,\mathrm{d}x$$

Die Verteilung oder Verteilungsfunktion von X ist die Funktion:

$$F_X(x) := \Pr[X \le x] = \int_{-\infty}^{x} f_X(t)\,\mathrm{d}t.$$

In den kommenden Folien: Zufallsvariable = stetige Zufallsvariable.

Zufallsvariablen IV: Eigenschaften

Sei X eine Zufallsvariable:

- F_X ist monoton steigend und stetig.
- $\Pr[a \le X \le b] = F_X(b) - F_X(a)$.
- $\lim\limits_{x\to-\infty} F_X(x) = 0$ und $\lim\limits_{x\to\infty} F_X(x) = 1$.
- Zwischen den Ereignissen „$a < X \le b$", „$a \le X \le b$", „$a \le X < b$" und „$a < X < b$" besteht kein wesentlicher Unterschied, da

$$\int_{[a,b]} f(t)\,\mathrm{d}t = \int_{(a,b]} f(t)\,\mathrm{d}t = \int_{[a,b[} f(t)\,\mathrm{d}t = \int_{]a,b[} f(t)\,\mathrm{d}t.$$

- $f_X(x) = \dfrac{\mathrm{d}\,F_X(x)}{\mathrm{d}\,x}$

Zufallsvariablen V: Gleichverteilung

Definition 17 (Gleichverteilung)

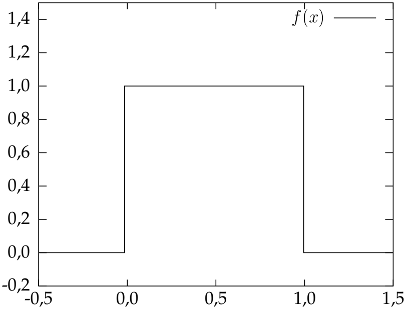
Eine kontinuierliche Zufalsvariable X ist gleichverteilt auf dem Intervall $[a, b]$ wenn

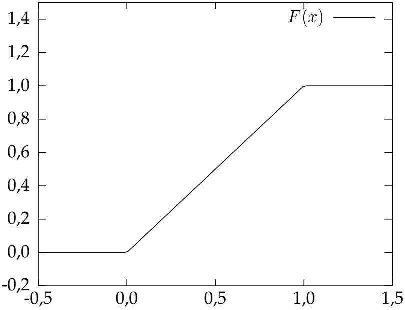
$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & \text{für } x \in [a, b], \\ 0 & \text{sonst.} \end{cases}$$

Verteilungsfunktion der Gleichverteilung:

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t = \begin{cases} 0 & \text{für } x < a, \\ \dfrac{x-a}{b-a} & \text{für } a \le x \le b, \\ 1 & \text{für } x > b. \end{cases}$$

Zufallsvariablen VI: Gleichverteilung





Gleichverteilung über dem Intervall $[0, 1]$

Zufallsvariablen VII: Exponentialverteilung

Definition 18

Eine kontinuierliche Zufalsvariable X ist exponentialverteilt wenn

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

Verteilungsfunktion der Exponentialverteilung:

$$\begin{aligned} F_X(x) &= \int_0^x e^{-\lambda t}\,\mathrm{d}t = \\ &= \left[-e^{-\lambda t}\right]_0^x \\ &= \begin{cases} 1 - e^{-\lambda y} & \text{für } y \geq 0, \\ 0 & \text{sonst.} \end{cases} \end{aligned}$$

W'keitsräume als Zufallsvariablen

- Sei $R = (\Omega, \mathcal{B}(\mathbb{R}), \Pr)$ ein W'keitsraum mit W'keitsdichte $f : \Omega \to \mathbb{R}_0^+$.
  Sei $X : \Omega \to \mathbb{R}$ eine Zufallsvariable.
- $R' = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \Pr')$ mit $\Pr'[A] = \int_A f_X \,\mathrm{d}x$ bildet ebenfalls einen W'keitsraum.
- Wenn $\Omega = \mathbb{R}$ und $X(\omega) = \omega$, dann sind die zwei Räume identisch.
- W'keitsräume mit $\Omega = \mathbb{R}$ werden daher oft direkt als Zufallsvariablen modelliert.

Beispiel: Experiment: Eine Zahl $x \in [0, 1]$ wird zufällig gewählt.

Modell: Sei X gleichverteilt auf $[0, 1]$.

Zusammengesetzte Zufallsvariablen I

Beispiel 19

Die Zeit X, die ein Server für eine Anfrage braucht sei exponentialverteilt mit Dichte $f_X(t) = \lambda e^{-\lambda t}$.
Die Gesamtkosten einer Anfrage seien $at + b$, wenn die Anfrage in t Sekunden beantwortet wird.
Frage: Mit welcher W'keit Kostet eine Anfrage mindestens k Euro ($k \geq b$)?

Die Kosten werden von der Zufallsvariable $Y := aX + b$ modelliert.
Wir müssen $\Pr[Y \leq k]$ berechnen.

Zusammengesetzte Zufallsvariablen II

- Allgemein gilt für $Y = g(X)$:

$$F_Y(y) = \Pr[Y \leq y] = \Pr[g(X) \leq y] = \int_C f_X(t)\,\mathrm{d}t.$$

  mit $C := \{t \in \mathbb{R} \mid g(t) \leq y\}$.

- Wenn g messbar, dann ist C eine Borel'sche Menge und das Integral ist definiert.
- Wenn $g(X) = aX + b$ dann

$$\Pr[g(X) \leq y] = \Pr\left[X \leq \frac{y-b}{a}\right] = F_X\left(\frac{y-b}{a}\right)$$

In unserem Beispiel:

$$\Pr[Y \leq k] = F_X\left(\frac{k-b}{a}\right) = 1 - e^{\lambda(k-b)/a}$$

Simulation von Zufallsvariablen I: Die Inversionsmethode

Die Simulation einer Zufallsvariablen X mit Dichte f_X ist die algorithmische Erzeugung von Zufallswerten, deren Verteilung der Verteilung von X entspricht.

Zufallsgeneratoren simulieren eine Zufallsvariable U gleichverteilt über $[0, 1]$.

Wie können Variablen mit anderen Verteilungen simuliert werden?

Simulation von Zufallsvariablen II: Die Inversionsmethode

Satz 20

Sei X eine Zufallsvariable mit einer streng monoton wachsenden Verteilungsfunktion F_X. Dann hat F_X eine (eindeutige) inverse Funktion F_X^{-1} mit $F_X(F_X^{-1}(x)) = x$ für alle $x \in (0,1)$.

Sei $\tilde{X} = F_X^{-1}(U)$. Es gilt $F_{\tilde{X}}(t) = F_X(t)$ für alle $t \in (0,1)$.

Beweis:

$$\begin{aligned} F_{\tilde{X}}(t) &= \Pr[\tilde{X} \le t] \\ &= \Pr[F_X^{-1}(U) \le t] \\ &= \Pr[U \le F_X(t)] \\ &= F_U(F_X(t)) \\ &= F_X(t)\,. \end{aligned}$$

□

Simulation von Zufallsvariablen II: Exponentialverteilung

Beispiel 21

Wir simulieren die Variable X mit Exponentialverteilung

$$F_X(x) = 1 - e^{-\lambda t} \text{ für } t \geq 0 \ .$$

Wir erhalten auf $(0, 1)$ die Umkehrfunktion

$$F_X^{-1}(t) = -\frac{1}{\lambda} \ln(1 - t) \ .$$

und damit

$$\tilde{X} = -\frac{1}{\lambda} \ln(1 - U) \ .$$

Kontinuierliche ZV als Grenzwerte diskreter ZV I

Sei X eine kontinuierliche Zufallsvariable.

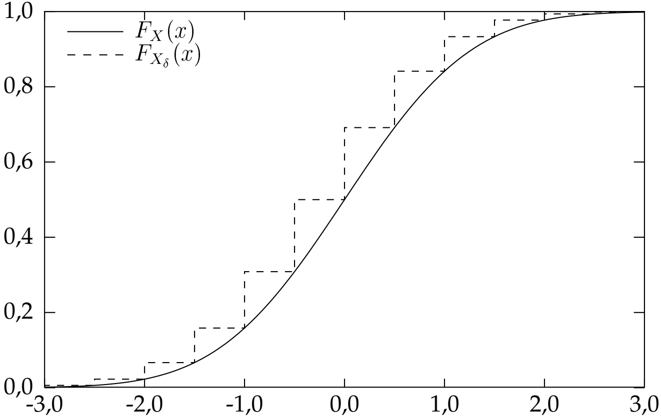
Wir konstruieren eine diskrete Zufallsvariable X_δ, indem wir für ein festes $\delta > 0$ definieren

$$X_\delta = n\delta \iff X \in [n\delta, (n+1)\delta[\text{ für } n \in \mathbb{Z}.$$

Es gilt

$$\Pr[X_\delta = n\delta] = F_X((n+1)\delta) - F_X(n\delta)\,.$$

Kontinuierliche ZV als Grenzwerte diskreter ZV II



Für $\delta \to 0$ nähert sich die Verteilung von X_δ der Verteilung von X immer mehr an.

Erwartungswert und Varianz I

Definition 22

Für eine kontinuierliche Zufallsvariable X ist der Erwartungswert definiert durch

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} t \cdot f_X(t)\,\mathrm{d}t,$$

sofern das Integral $\displaystyle\int_{-\infty}^{\infty} |t| \cdot f_X(t)\,\mathrm{d}t$ endlich ist.

Für die Varianz gilt entsprechend

$$\mathrm{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \int_{-\infty}^{\infty} (t - \mathbb{E}[X])^2 \cdot f_X(t)\,\mathrm{d}t,$$

wenn $\mathbb{E}[(X - \mathbb{E}[X])^2]$ existiert.

Erwartungswert und Varianz II

Beispiel 23

Für Erwartungswert und Varianz der Gleichverteilung ergibt sich

$$\begin{aligned}
\mathbb{E}[X] &= \int_a^b t \cdot \frac{1}{b-a}\,\mathrm{d}t = \frac{1}{b-a} \cdot \int_a^b t \cdot \mathrm{d}t \\
&= \frac{1}{2(b-a)} \cdot [t^2]_a^b \\
&= \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}, \\
\mathbb{E}[X^2] &= \frac{1}{b-a} \cdot \int_a^b t^2 \cdot \mathrm{d}t = \frac{b^2+ba+a^2}{3}, \\
\mathrm{Var}[X] &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \ldots = \frac{(a-b)^2}{12}\,.
\end{aligned}$$

Erwartungswert einer zusammengesetzten Variable I

Lemma 24

Sei X eine kontinuierliche Zufallsvariable, und sei $Y := g(X)$. Dann gilt

$$\mathbb{E}[Y] = \int_{-\infty}^{\infty} g(t) \cdot f_X(t)\, \mathrm{d}t\,.$$

Beweis:

Nur für $Y := a \cdot X + b$ mit $a, b \in \mathbb{R}$ und $a > 0$. Es gilt

$$\mathbb{E}[a \cdot X + b] = \int_{-\infty}^{\infty} t \cdot f_Y(t)\, \mathrm{d}t = \int_{-\infty}^{\infty} t \cdot f_X\left(\frac{t-b}{a}\right) \cdot \frac{1}{a}\, \mathrm{d}t.$$

Durch die Substitution $u := (t-b)/a$ mit $\mathrm{d}\, u = (1/a)\, \mathrm{d}\, t$ erhalten wir

$$\mathbb{E}[a \cdot X + b] = \int_{-\infty}^{\infty} (au + b) f_X(u)\, \mathrm{d}u.$$

$\square$

Definition 25

Die gemeinsame Dichte zweier kontinuierlichen Zufallsvariablen X, Y ist eine integrierbare Dichtefunktion $f_{X,Y} : \mathbb{R}^2 \to \mathbb{R}_0^+$ mit

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y) \, \mathrm{d}x \, \mathrm{d}y = 1$$

Für ein Ereignis $A \in \mathcal{B}(\mathbb{R}^2)$ gilt

$$\Pr[A] = \int_A f_{X,Y}(x,y) \, \mathrm{d}x \, \mathrm{d}y.$$

Die (gemeinsame) Verteilung $F_{X,Y} : \mathbb{R}^2 \to [0,1]$ von X und Y ist definiert durch:

$$F_{X,Y}(x,y) = \Pr[X \le x, Y \le y] = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(u,v) \, \mathrm{d}u \, \mathrm{d}v.$$

Randverteilungen

Definition 26

Sei $f_{X,Y}$ die gemeinsame Dichte der Zufallsvariablen X und Y. Die Randverteilung der Variablen X ist gegeben durch

$$F_X(x) = \Pr[X \le x] = \int_{-\infty}^{x} \left[\int_{-\infty}^{\infty} f_{X,Y}(u,v) \,\mathrm{d}v \right] \mathrm{d}u.$$

Analog nennen wir

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,v) \,\mathrm{d}v$$

die Randdichte von X. Entsprechende Definitionen gelten symmetrisch für Y.

Unabhängigkeit I

Definition 27

Zwei kontinuierliche Zufallsvariablen X und Y heißen unabhängig, wenn

$$\Pr[X \leq x, Y \leq y] = \Pr[X \leq x] \cdot \Pr[Y \leq y]$$

für alle $x, y \in \mathbb{R}$ gilt.

Dies ist gleichbedeutend mit

$$F_{X,Y}(x, y) = F_X(x) \cdot F_Y(y)\,.$$

Differentiation ergibt

$$f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y)\,.$$

Unabhängigkeit II

Definition 28

Die kontinuierlichen Zufallsvariablen $X_1, \ldots, X_n$ sind unabhängig, wenn

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

bzw.

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

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

Summen von unabhängigen Zufallsvariablen I

Satz 29

Seien X und Y unabhängige kontinuierliche Zufallsvariablen. Für die Dichte von $Z := X + Y$ gilt

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) \cdot f_Y(z-x)\,\mathrm{d}x\,.$$

Beweis:

Nach Definition der Verteilungsfunktion gilt

$$F_Z(t) = \Pr[Z \le t] = \Pr[X + Y \le t] = \int_{A(t)} f_{X,Y}(x,y)\,\mathrm{d}x\mathrm{d}y$$

wobei $A(t) = \{(x,y) \in \mathbb{R}^2 \mid x + y \le t\}$.

Summen von unabhängigen Zufallsvariablen II

Beweis (Forts.):

Aus der Unabhängigkeit von X und Y folgt

$$\begin{aligned} F_Z(t) &= \int_{A(t)} f_X(x) \cdot f_Y(y)\,\mathrm{d}x\mathrm{d}y \\ &= \int_{-\infty}^{\infty} f_X(x) \cdot \left(\int_{-\infty}^{t-x} f_Y(y)\,\mathrm{d}y \right) \mathrm{d}x. \end{aligned}$$

Mittels der Substitution $z := x + y$, $\mathrm{d}\,z = \mathrm{d}\,y$ ergibt sich

$$\int_{-\infty}^{t-x} f_Y(y)\,\mathrm{d}y = \int_{-\infty}^{t} f_Y(z - x)\,\mathrm{d}z$$

und somit

$$F_Z(t) = \int_{-\infty}^{t} \left(\int_{-\infty}^{\infty} f_X(x) f_Y(z - x)\,\mathrm{d}x \right) \mathrm{d}z\,.$$

$\square$

3. Wichtige stetige Verteilungen

Die Normalverteilung I

Die Normalverteilung ist in gewisser Weise das kontinuierliche Analogon zur Binomialverteilung wenn $n \to \infty$ bei konstanter Erfolgswahrscheinlichkeit p.

Definition 30

Eine Zufallsvariable X mit Wertebereich $W_X = \mathbb{R}$ heißt normalverteilt mit den Parametern $\mu \in \mathbb{R}$ und $\sigma \in \mathbb{R}^+$, wenn sie die Dichte

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) =: \varphi(x; \mu, \sigma)$$

besitzt.

Die Normalverteilung II

In Zeichen schreiben wir $X \sim \mathcal{N}(\mu, \sigma^2)$.

$\mathcal{N}(0,1)$ heißt Standardnormalverteilung.

Die Dichte $\varphi(x; 0, 1)$ kürzen wir durch $\varphi(x)$ ab.

Die Verteilungsfunktion zu $\mathcal{N}(\mu, \sigma^2)$ ist

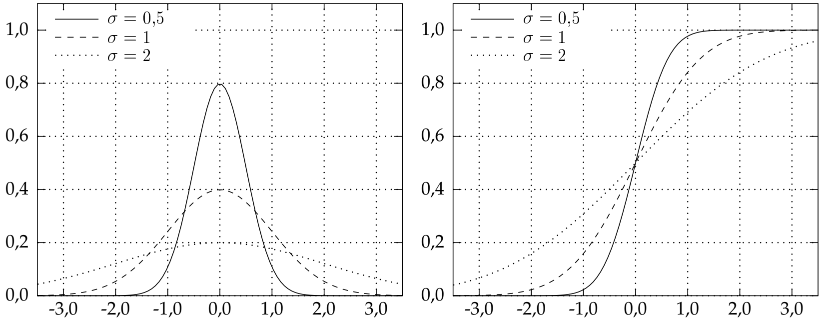
$$F(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \int_{-\infty}^{x} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) \mathrm{d}t =: \Phi(x; \mu, \sigma)\,.$$

Diese Funktion heißt Gauß'sche Φ-Funktion.

Die Verteilung $\varphi(x; 0, 1)$ kürzen wir durch $\Phi(x)$ ab.

Es gibt keinen geschlossenen Ausdruck für $\Phi(x; \mu, \sigma)$.

Die Normalverteilung III



Dichte und Verteilung von $\mathcal{N}(0, \sigma^2)$

Die Normalverteilung IV: Erwartungswert und Varianz

Wir berechnen zunächst Erwartungswert und Varianz einer Zufallsvariable $X \sim \mathcal{N}(0,1)$. (Satz 32).

(D.h., wir betrachten den Speziallfall $\mu = 0, \sigma = 1$.)

Dann berechnen wir Erwartungswert und Varianz für beliebige Werte der Parameter μ und σ. (Satz 34).

Erwartungswert und Varianz von $\mathcal{N}(0,1)$ I

Lemma 31

$$I := \int_{-\infty}^{\infty} e^{-x^2/2}\,\mathrm{d}x = \sqrt{2\pi}.$$

Beweis:

Wir berechnen zunächst I^2:

$$\begin{aligned} I^2 &= \left(\int_{-\infty}^{\infty} e^{-x^2/2}\,\mathrm{d}x\right)\left(\int_{-\infty}^{\infty} e^{-y^2/2}\,\mathrm{d}y\right) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)/2}\,\mathrm{d}x\,\mathrm{d}y\,. \end{aligned}$$

Erwartungswert und Varianz von $\mathcal{N}(0,1)$ II

Beweis (Forts.):

Wir gehen nun zu Polarkoordinaten über durch die Substitution:

$$x := r\cos\phi \qquad\qquad y := r\sin\phi$$

Die Derterminante der Jacobi-Matrix der Substitution ist

$$\begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial y}{\partial r} \\ \frac{\partial x}{\partial \phi} & \frac{\partial y}{\partial \phi} \end{vmatrix} = \begin{vmatrix} \cos\phi & \sin\phi \\ -r\sin\phi & r\cos\phi \end{vmatrix} = r(\cos^2\phi + \sin^2\phi) = r$$

und wir erhalten

$$\begin{aligned} I^2 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)/2}\,\mathrm{d}x\,\mathrm{d}y \;=\; \int_0^{2\pi}\int_0^{\infty} e^{-r^2/2}\,r\;\mathrm{d}r\,\mathrm{d}\phi \\ &= \int_0^{2\pi}\left[-e^{-r^2/2}\right]_0^{\infty}\mathrm{d}\phi \;=\; \int_0^{2\pi} 1\,\mathrm{d}\phi \;=\; 2\pi. \end{aligned}$$

Erwartungswert und Varianz von $\mathcal{N}(0,1)$ III

Satz 32

X *sei* $\mathcal{N}(0,1)$*-verteilt. Dann gilt*

$$\mathbb{E}[X] = 0 \textit{ und } \quad \mathrm{Var}[X] = 1.$$

Beweis:

Es gilt

$$\mathbb{E}[X] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x \cdot e^{-x^2/2} \, \mathrm{d}x.$$

Sei $f(x) = x \cdot e^{-x^2/2}$.

Es gilt $f(-x) = -f(x)$ und damit $\mathbb{E}[X] = 0$.

Erwartungswert und Varianz von $\mathcal{N}(0,1)$ IV

Für die Varianz berechnen wir $\mathbb{E}[X^2]$.

Mit Lemma 31 gilt

$$\int_{-\infty}^{\infty} e^{-x^2/2}\,\mathrm{d}x = \sqrt{2\pi}.$$

Durch partielle Integration erhalten wir

$$\begin{aligned}\sqrt{2\pi} &= \int_{-\infty}^{\infty} e^{-x^2/2}\,\mathrm{d}x \\ &= xe^{-x^2/2}\Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} x^2 \cdot e^{-x^2/2}\,\mathrm{d}x \\ &= 0 \quad + \quad \mathbb{E}[X^2]\end{aligned}$$

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

Erwartungswert und Varianz von $\mathcal{N}(\mu, \sigma)$ I

Lemma 33 (Lineare Transformation der Normalverteilung)

Wenn $X \sim \mathcal{N}(\mu, \sigma^2)$ *und* $Y = aX + b$ *dann*

$$Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)\ .$$

Beweis: Fall „$a > 0$“ („$a < 0$“ analog):

$$\Pr[Y \le y] = \Pr[aX + b \le y] = \Pr\left[X \le \frac{y-b}{a}\right]$$

$$= \frac{1}{\sqrt{2\pi}\sigma} \cdot \int_{-\infty}^{(y-b)/a} \exp\left(-\frac{(u-\mu)^2}{2\sigma^2}\right) \mathrm{d}u.$$

Mit $u = (v - b)/a$ und $\mathrm{d}u = (1/a)\ \mathrm{d}v$:

$$\Pr[Y \le y] = \frac{1}{\sqrt{2\pi}a\sigma} \cdot \int_{-\infty}^{y} \exp\left(-\frac{(v-(a\mu+b))^2}{2(a\sigma)^2}\right) \mathrm{d}v\ .$$

Also $Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.

Erwartungswert und Varianz von $\mathcal{N}(\mu, \sigma)$ II

Satz 34

Sei $X \sim \mathcal{N}(\mu, \sigma^2)$. *Dann gilt*

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

Beweis:

Sei $Y := \frac{1}{\sigma}(X - \mu)$ (und so $X = \sigma Y + \mu$).

Mit Lemma 33 gilt $Y \sim \mathcal{N}\left(\left(\frac{1}{\sigma}\right)\mu - \frac{\mu}{\sigma}, \left(\frac{1}{\sigma}\right)^2 \sigma^2\right) \sim \mathcal{N}(0,1)$

und so

$$\mathbb{E}[X] = \mathbb{E}[\sigma Y + \mu] = \sigma \cdot \mathbb{E}[Y] + \mu = \mu$$

und

$$\mathrm{Var}[X] = \mathrm{Var}[\sigma Y + \mu] = \sigma^2 \cdot \mathrm{Var}[Y] = \sigma^2.$$

$\square$

Normalverteilung als Grenzwert der Binomialverteilung I

Seien $X_1, \ldots, X_n$ unabhängige Bernoulli-verteilte Zufallsvariablen mit gleicher Erfolgswahrscheinlichkeit p.

Sei $H_n := X_1 + \ldots + X_n$. Es gilt $H_n \sim \mathrm{Bin}(n; p)$ mit

$$\mathbb{E}[H_n] = np \qquad \text{und} \qquad \mathrm{Var}[H_n] = npq\ .$$

Die standardisierte Variable $H_n^* := \dfrac{H_n - np}{\sqrt{npq}}$ erfüllt

$$\mathbb{E}[H_n^*] = \frac{\mathbb{E}[H_n] - np}{\sqrt{npq}} = 0 \quad \text{und} \quad \mathrm{Var}[H_n] = \left(\frac{1}{\sqrt{npq}}\right)^2 \mathrm{Var}[H_n] = 1\ .$$

Wir betrachten die Sequenz $H_1^*, H_2^*, H_3^* \ldots$ und untersuchen den Grenzwert von $F_{H_1^*}, F_{H_2^*}, F_{H_3^*} \ldots$.

Normalverteilung als Grenzwert der Binomialverteilung II

Satz 35 (Grenzwertsatz von de Moivre)

Seien $X_1, \ldots, X_n$ unabhängige Bernoulli-verteilte Zufallsvariablen mit gleicher Erfolgswahrscheinlichkeit p.

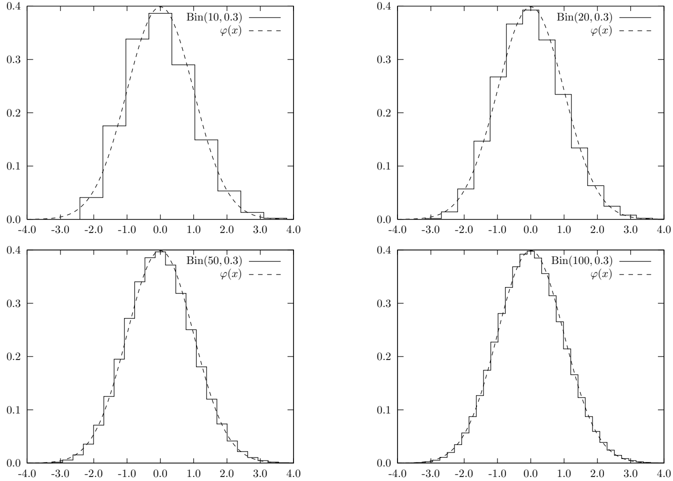
Sei $H_n := X_1 + \ldots + X_n$. Die Verteilung $F_{H_n^*}$ der Zufallsvariable

$$H_n^* := \frac{H_n - np}{\sqrt{npq}}$$

konvergiert gegen die Standardnormalverteilung $\mathcal{N}(0, 1)$ für $n \to \infty$.

Der Satz wird später als Spezialfall des zentralen Grenzwertsatzes bewiesen. (Satz 40 und Korollar 14).

Normalverteilung als Grenzwert der Binomialverteilung III



Vergleich von Binomial- und Normalverteilung

$\mathrm{Bin}(n, 0.3)$ bei $0.3n$ zentriert, mit $\sqrt{0.3 \cdot 0.7n}$ horizontal gestaucht und vertikal gestreckt

Normalverteilung als Grenzwert der Binomialverteilung III

H_n beschreibt die Anzahl der Erfolge bei n Versuchen.

$\frac{H_n}{n}$ beschreibt die Frequenz der Erfolge.

Die folgende Aussage ist eine Konsequenz von Satz 35:

Korollar 36

Sei $H_n \sim \mathrm{Bin}(n, p)$ eine binomialverteilte Zufallsvariable.

Die Verteilung von $\frac{H_n}{n}$ konvergiert gegen

$$\mathcal{N}\left(p, \frac{pq}{n}\right)$$

für $n \to \infty$. Insbesondere gilt $\lim_{n\to\infty} \mathrm{Var}\left[\frac{H_n}{n}\right] = 0$.

Beispiel 37

25.05.2012, ZDF: Die Umfragen zu diesem Politbarometer wurden wie immer von der Mannheimer Forschungsgruppe Wahlen durchgeführt. Die Interviews wurden in der Zeit vom 22. bis 24. Mai 2012 bei 1312 zufällig ausgewählten Wahlberechtigten telefonisch erhoben. Die Befragung ist repräsentativ für die wahlberechtigte Bevölkerung in Deutschland. Der Fehlerbereich beträgt bei einem Parteianteil von 40 Prozent rund +/- drei Prozentpunkte und bei einem Parteianteil von zehn Prozent rund +/- zwei Prozentpunkte. Daten zur politischen Stimmung: CDU/CSU: 38 Prozent, SPD: 34 Prozent, FDP: zwei Prozent, Linke: vier Prozent, Grüne: 14 Prozent, Piraten: sechs Prozent.

Wie werden diese Fehlerbereiche berechnet?
Warum sind sie vom Parteilanteil abhängig?

Normalverteilung und Stichproben II

Jede Umfrage muss zwei Parameter angeben:

- Konfidenz- oder Vertrauensintervall.
  Intervall um den geschätzten Wert s.
  Für die CDU/DSU: $s = 0.38$, Intervall $[s - 0.03, s + 0.03]$.
  („Fehlerbereich“ ist nocht der übliche Fachbegriff.)
- Konfidenzniveau.
  Untere Schranke für die W'keit, dass der unbekannte wahre Wert p im Konfidenzintervall liegt.
  Das übliche Konfidenzniveau bei diesen Studien beträgt 95%.

Ohne die Angabe beider Parameter ist eine Umfrage Wertlos!

Normalverteilung und Stichproben III

Umfrage als mehrstufiges Experiment:

- Eine Urne enthält weisse und schwarze Bälle.
- Das Experiment besteht aus n Stufen.
- In jeder Stufe wird ein zufälliger Ball aus der Urne extrahiert und zurück in die Urne gestellt.
- Sei p die (unbekannte!) W'keit, dass der extrahierter Ball schwarz ist.
- Sei k die Anzahl der gezogenen Bällen, die Schwarz sind $(0 \leq k \leq n)$.
- Ergebnis s des Experiments: $s = k/n$.
- Die Umfrage hat ein Konfidenzniveau von 95%:
  Die W'keit, dass $|s - p| \leq 0.03$ beträgt (mindestens) 95%.

Normalverteilung und Stichproben IV

Zu berechnen ist die Zahl δ, für die die W'keit, dass bei 1312 Befragten das Ergebnis der Umfrage im Intervall $[p - \delta, p + \delta]$ liegt, 0.95 beträgt.

Dazu modellieren wir die Befragung durch folgendes n-stufiges Experiment ($n = 1312$).

Normalverteilung und Stichproben V

Es gilt $\Pr[-2 \leq \mathcal{N}(0,1) \leq 2] \approx 0.9554$ (Tabelle, Internet))

Wir haben

$$
\begin{aligned}
0.95 &\approx \Pr[-2 \leq H_n^* \leq 2] \\
&= \Pr[np - 2\sqrt{np(1-p)} \leq H_n \leq np + 2\sqrt{np(1-p)}] \\
&= \Pr\left[p - 2\sqrt{\tfrac{p(1-p)}{n}} \leq \tfrac{H_n}{n} \leq p + 2\sqrt{\tfrac{p(1-p)}{n}}\right]
\end{aligned}
$$

Damit gilt: bei n befragten und 95% Konfidenz erhählt man einen Fehlerbereich von

$$\delta = 2\sqrt{\frac{p(1-p)}{n}}$$

.

Normalverteilung und Stichproben VI

Mit $p(1-p) \leq 1/4$ für alle $0 \leq p \leq 1$ erhalten wir

$$0.95 \approx \Pr[-2 \leq H_n^* \leq 2] \leq \Pr\left[p - \sqrt{\frac{1}{n}} \leq \frac{H_n}{n} \leq p + \sqrt{\frac{1}{n}}\right]$$

Damit gilt $\delta \leq \dfrac{1}{\sqrt{n}}$ für alle p. In anderen Worten:

> *unabhängig vom Wert von p erhält man bei n befragten und 95% Konfidenz einen Fehlerbereich von ca. $\dfrac{1}{\sqrt{n}}$.*

Bei $n = 1312$ beträgt der Fehlerbereich $\frac{1}{\sqrt{1312}} \approx 0.0276$ oder 2,8 Prozentpunkte für alle p.

Für $p \leq 0.4$: $2\sqrt{\frac{0.24}{1312}} \approx 0.0271$ oder 2,7 Prozentpunkte.

Für $p \leq 0.1$: $2\sqrt{\frac{0.09}{1312}} \approx 0.0166$ oder 1,7 Prozentpunkte.

4. Momenterzeugende Funktionen

Momenterzeugende Funktionen I

Definition 38

Zu einer Zufallsvariablen X ist die momenterzeugende Funktion gemäß

$$M_X(s) = \mathbb{E}[e^{Xs}] = \int_{-\infty}^{\infty} e^{ts} \cdot f_X(t)\,\mathrm{d}t$$

definiert.

Momenterzeugende Funktionen II: Gleichverteilung

Für eine auf $[a, b]$ gleichverteilte Zufallsvariable U gilt

$$
\begin{aligned}
M_U(t) = \mathbb{E}[e^{tX}] &= \int_a^b e^{tx} \cdot \frac{1}{b-a}\,\mathsf{d}x \\
&= \left[\frac{e^{tx}}{t(b-a)}\right]_a^b \\
&= \frac{e^{tb} - e^{ta}}{t(b-a)}\,.
\end{aligned}
$$

Momenterzeugende Funktionen III: Normalverteilung

Für eine standardnormalverteilte Zufallsvariable $N \sim \mathcal{N}(0,1)$ gilt

$$\begin{aligned} M_N(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{t\xi} e^{-\xi^2/2}\, \mathsf{d}\xi \\ &= e^{t^2/2} \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-(t-\xi)^2/2}\, \mathsf{d}\xi = e^{t^2/2}\,. \end{aligned}$$

Daraus ergibt sich für $Y \sim \mathcal{N}(\mu, \sigma^2)$

$$\begin{aligned} M_Y(t) &= \mathbb{E}[e^{tY}] \\ &= e^{t\mu} \cdot \mathbb{E}[e^{(t\sigma)\cdot \frac{Y-\mu}{\sigma}}] \\ &= e^{t\mu} \cdot M_N(t\sigma) \qquad \text{(wegen } \tfrac{Y-\mu}{\sigma} \sim \mathcal{N}(0,1)\text{)} \\ &= e^{t\mu + (t\sigma)^2/2}\,. \end{aligned}$$

Summe von Normalverteilungen

Satz 39 (Additivität der Normalverteilung)

Seien $X_1, \ldots, X_n$ unabhängige Zufallsvariablen mit $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ für alle $1 \leq i \leq n$.

Sei $Z := a_1X_1 + \ldots + a_nX_n$. Es gilt $Z \sim \mathcal{N}(\mu, \sigma^2)$ mit

$$\mu = a_1\mu_1 + \ldots + a_n\mu_n \quad \textit{und} \quad \sigma^2 = a_1^2\sigma_1^2 + \ldots + a_n^2\sigma_n^2\,.$$

Beweis: Mit $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ gilt $M_{X_i}(t) = e^{t\mu_i + (t\sigma_i)^2/2}$.
Aus der Unabhängigkeit der X_i folgt

$$\begin{aligned} M_Z(t) = \mathbb{E}[e^{t(a_1X_1+\cdots+a_nX_n)}] &= \prod_{i=1}^{n} \mathbb{E}[e^{(a_it)X_i}] \\ &= \prod_{i=1}^{n} M_{X_i}(a_it) = \prod_{i=1}^{n} e^{a_it\mu_i + (a_it\sigma_i)^2/2} = e^{t\mu + (t\sigma)^2/2}\,. \end{aligned}$$

5. Der zentraler Grenzwertsatz

Der zentraler Grenzwertsatz I

Dieser Satz ist von großer Bedeutung für die Anwendung der Normalverteilung in der Statistik.

Informell lautet die Aussage des Satzes:

Die Verteilung einer Summe unabhängiger
identisch ABER BELIEBIG verteilter Zufallsvariablen
nähert der Normalverteilung umso mehr an,
je mehr Zufallsvariablen an der Summe beteiligt sind.

(OK, nicht ganz: Erwartungswert und Varianz der Variablen müssen endlich sein.)

Zentraler Grenzwertsatz II

Satz 40 (Zentraler Grenzwertsatz)

Die Zufallsvariablen $X_1, \ldots, X_n$ *besitzen jeweils dieselbe Verteilung und seien unabhängig.*

Erwartungswert und Varianz von X_i *existieren für* $i = 1, \ldots, n$ *und seien mit* μ *bzw.* σ^2 *bezeichnet* ($\sigma^2 > 0$).

Die Zufallsvariablen Y_n *seien definiert durch* $Y_n := X_1 + \ldots + X_n$ *für* $n \geq 1$. *Dann folgt, dass die Zufallsvariablen*

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

asymptotisch standardnormalverteilt sind, also $Z_n \sim \mathcal{N}(0, 1)$ *für* $n \to \infty$.

Zentraler Grenzwertsatz II

Etwas formaler ausgedrückt gilt: Die Folge der zu H_n^* gehörenden Verteilungsfunktionen F_n hat die Eigenschaft

$$\lim_{n \to \infty} F_n(x) = \Phi(x) \text{ für alle } x \in \mathbb{R}.$$

Wir sagen dazu auch: Die Verteilung von H_n^* konvergiert gegen die Standardnormalverteilung für $n \to \infty$.

Zentraler Grenzwertsatz III

Beweis:

Sei $X_i^* := (X_i - \mu)/\sigma$ für $i = 1, \ldots, n$.

Es gilt $\mathbb{E}[X_i^*] = 0$ und $\mathrm{Var}[X_i^*] = 1$.

Damit haben wir

$$\begin{aligned}
M_Z(t) \quad &= \mathbb{E}[e^{tZ}] = \mathbb{E}[e^{t(X_1^* + \ldots + X_n^*)/\sqrt{n}}] \\
&= \mathbb{E}\left[e^{tX_1^*/\sqrt{n}} \cdot \ldots \cdot e^{tX_1^*/\sqrt{n}}\right] \\
&= \mathbb{E}\left[e^{tX_1^*/\sqrt{n}}\right] \cdot \ldots \cdot \mathbb{E}\left[e^{tX_1^*/\sqrt{n}}\right] && \text{(Unabh.)} \\
&= M_{X_1^*}(t/\sqrt{n}) \cdot \ldots \cdot M_{X_n^*}(t/\sqrt{n})\,. \\
&= \left(M_{X_1^*}(t/\sqrt{n})\right)^n && \text{(X_i's gleich verteilt)}
\end{aligned}$$

Zentraler Grenzwertsatz IV

Wir betrachten die Taylorentwicklung von $M_{X_i^*}(t) =: h(t)$ mit Entwicklungsstelle 0:

$$h(t) = h(0) + h'(0) \cdot t + \frac{h''(0)}{2} \cdot t^2 + \mathcal{O}(t^3).$$

Es gilt

$$\begin{aligned}
h'(t) &= \frac{\mathsf{d}}{\mathsf{d}\,t}\mathbb{E}[e^{tX_i^*}] = \frac{\mathsf{d}}{\mathsf{d}\,t}\mathbb{E}\left[\sum_{i=0}^{\infty} \frac{(tX_i^*)^n}{n!}\right] \\
&= \sum_{i=0}^{\infty} n \cdot t^{n-1}\mathbb{E}\left[\frac{(X_i^*)^n}{n!}\right] \qquad \text{(Lin. und Arith)} \\
&= \sum_{i=1}^{\infty} \mathbb{E}\left[X_i^* \frac{(tX_i^*)^{n-1}}{(n-1)!}\right] = \mathbb{E}\left[X_i^* \sum_{i=0}^{\infty} \frac{(tX_i^*)^n}{n!}\right] \\
&= \mathbb{E}[e^{tX_i^*} \cdot X_i^*]
\end{aligned}$$

Zentraler Grenzwertsatz V

Analog erhalten wir $h''(t) = \mathbb{E}[e^{tX_i^*} \cdot (X_i^*)^2]$.

Damit gilt

$$h'(0) = \mathbb{E}[X_i^*] = 0 \qquad h''(0) = \mathbb{E}[(X_i^*)^2] = \mathrm{Var}[X] = 1 \; .$$

Durch Einsetzen in die Taylorreihe folgt

$$h(t) = 1 + t^2/2 + \mathcal{O}(t^3)$$

und damit

$$\begin{aligned} M_Z(t) \quad &= \quad \left(M_{X_1^*}(t/\sqrt{n})\right)^n = \left(1 + \frac{t^2}{2n} + \mathcal{O}\left(\frac{t^3}{n^{3/2}}\right)\right)^n \\ &\to \quad e^{t^2/2} \text{ für } n \to \infty. \end{aligned}$$

Aus der Konvergenz der momenterzeugenden Funktion folgt auch die Konvergenz der Verteilung. Damit ist Z asymptotisch normalverteilt.

Zentraler Grenzwertsatz VI

Die momenterzeugende Funktion existiert nicht bei allen Zufallsvariablen.

Der Beweis ist deshalb unvollständig.

Man umgeht dieses Problem, indem man statt der momenterzeugenden Funktion die charakteristische Funktion

$$\tilde{M}_X(t) = \mathbb{E}[e^{\mathrm{i}tX}]$$

betrachtet.

Beispiel 41

Physical constants, National Institute of Standards and Technology (http://physics.nist.gov/constants) (June 2012)

Newtonian constant of gravitation G

Value	6.673 84 $\times 10^{-11}$ m^3 kg^{-1} s^{-2}
Standard uncertainty	0.000 80 $\times 10^{-11}$ m^3 kg^{-1} s^{-2}
Concise form	6.673 84(80) $\times 10^{-11}$ m^3 kg^{-1} s^{-2}

Was ist die "standard uncertainty" ?

Der ZGWS und experimentelle Messungen II

Beispiel 42

New Scientist Physics & Math, 09.02.2012: The Higgs boson is the missing piece of the standard model of physics ... In December, the LHC's two main particle detectors, CMS and ATLAS, each reported excesses of events, such as the appearance of a pair of photons in the shrapnel from particle collisions. These excess events could be due to a Higgs with a mass of around 125 gigaelectron volts ... By convention, researchers only declare a discovery when an anomaly reaches a statistical significance known as 5 sigma, which means there is less than a 1-in-a-million chance it is just a fluke. The size of the anomalies reported in December was 1.9 sigma for CMS and 2.5 sigma for ATLAS, which indicate a probability of a fluke of roughly 1 per cent.

Wieso bedeutet „5 sigma“, dass die W'keit eines Zufallstreffers unter 10^{-6} liegt?

Der ZGWS und experimentelle Messungen III

Messfehler: Abweichung eines aus Messungen gewonnenen Wertes vom wahren Wert der Messgröße.

Systematische Fehler: Messfehler, die sich bei wiederholter Messung nicht im Mittel aufheben.
(Z.B. Fehler durch falsche Kalibrierung von Messgeräten.

Systematische Fehler werden durch sorgfältige Analyse und unabhängige Reproduktion des Experimentes beseitigt.

Zufallsfehler: Messfehler, die sich bei wiederholter Messung im Mittel aufheben.
(Z.B. Fehler durch nicht beherrschbare Einflüsse der Messgeräte oder nicht beherrschbare Änderungen des Wertes der Messgröße.)

Der ZGWS und experimentelle Messungen IV

Annahme: systematische Fehler sind beseitigt worden.

Zufallsfehler werden durch Wiederholung mit Hilfe des Zentralen Grenzwertsatzes beherrscht:

- Ergebnis des Experiments $\rightarrow$ Zufallsvariable X.
- $\mu := \mathbb{E}[X]$ ist der wahre Wert der Messgröße. (Folgt aus der Annahme).
- Experiment wird N-Mal wiederholt $\rightarrow$ Variablen $X_1, \ldots, X_N$, identisch verteilt.
- Die Verteilung von X ist nicht bekannt, aber aus dem ZGWS folgt: $Z_n \approx \mathcal{N}(0,1)$ wenn $n \rightarrow \infty$.

Annahme: $Z_N \approx \mathcal{N}(0,1)$.

Der ZGWS und experimentelle Messungen V

Sei σ_N^2 die Varianz von Y_N (σ_N die Standardabweichung).

$$
\begin{aligned}
\Phi(k)-\Phi(-k) &= \Pr[-k \le Z_N \le k] \\
&= \Pr\left[\mu - k\frac{\sigma}{\sqrt{N}} \le \frac{Y_N}{N} \le \mu + k\frac{\sigma}{\sqrt{N}}\right] \\
&= \Pr\left[\mu - k\sigma_N \le \frac{Y_N}{N} \le \mu + k\sigma_N\right]
\end{aligned}
$$

Die W'keit, dass Y_N/N mehr als k Standardabweichungen vom wahren Wert liegt, beträgt höchstens $\Phi(k)-\Phi(-k)$.

Der ZGWS und experimentelle Messungen VI

Die W'keit, dass Y_N/N mehr als k Standardabweichungen vom wahren Wert liegt, beträgt höchstens $\Phi(k) - \Phi(-k)$.

k	$\Phi(k) - \Phi(-k)$
1	0.6826895
2	0.9544997
3	0.9973003
4	0.9999367
5	0.9999995

Der ZGWS und experimentelle Messungen I

Beispiel 12

Physical constants, National Institute of Standards and Technology (http://physics.nist.gov/constants) (June 2012)

Newtonian constant of gravitation G

Value	6.673 84 $\times$ 10^{-11} m^3 kg^{-1} s^{-2}
Standard uncertainty	0.000 80 $\times$ 10^{-11} m^3 kg^{-1} s^{-2}
Concise form	6.673 84(80) $\times$ 10^{-11} m^3 kg^{-1} s^{-2}

Was ist die “standard uncertainty” ?

Die standard uncertainty ist (im Wesentlichen) σ_N.

Normalverteilung und experimentelle Messungen

Beispiel 13

New Scientist Physics & Math, 09.02.2012: In December, the LHC's two main particle detectors, CMS and ATLAS, each reported excesses of events, such as the appearance of a pair of photons in the shrapnel from particle collisions. ... By convention, researchers only declare a discovery when an anomaly reaches a statistical significance known as 5 sigma, which means there is less than a 1-in-a-million chance it is just a fluke. The size of the anomalies reported in December was 1.9 sigma for CMS and 2.5 sigma for ATLAS, which indicate a probability of a fluke of roughly 1 per cent.

Wieso bedeutet „5 sigma", dass die W'keit eines Zufallstreffers unter 10^{-6} liegt?

Weil $\Pr[-5 \leq Z_N \leq 5] \approx 10^{-6}$ wenn $Z_N \sim \mathcal{N}(0,1)$.

Grenzwertsatz von de Moivre I

Korollar 14 (Grenzwertsatz von de Moivre)

$X_1, \ldots, X_n$ seien unabhängige Bernoulli-verteilte Zufallsvariablen mit gleicher Erfolgsw'keit p. Dann gilt für $H_n := X_1 + \ldots + X_n$ und $n \geq 1$, dass die Verteilung der Zufallsvariablen

$$H_n^* := \frac{H_n - np}{\sqrt{np(1-p)}}$$

für $n \to \infty$ gegen die Standardnormalverteilung konvergiert.

Beweis:

Wenn $X_i \sim Ber(p)$ dann $\mu = \mathbb{E}[X_i] = p$, $\sigma^2 = \mathrm{Var}[X_i] = pq$.
Einsetzen im ZGWS ergibt $Z_n = H_n^*$ und so $H_n^* \sim \mathcal{N}(0,1)$. $\square$

Grenzwertsatz von de Moivre II

Historisch gesehen entstand Korollar 14 vor Satz 40.

Für den Fall $p = 1/2$ wurde Korollar 14 bereits von Abraham de Moivre (1667–1754) bewiesen. De Moivre war gebürtiger Franzose, musste jedoch aufgrund seines protestantischen Glaubens nach England fliehen. Dort wurde er unter anderem Mitglied der Royal Society, erhielt jedoch niemals eine eigene Professur.

Die allgemeine Formulierung von Korollar 14 geht auf Pierre Simon Laplace (1749–1827) zurück. Allerdings vermutet man, dass die Lösung des allgemeinen Falls $p \neq 1/2$ bereits de Moivre bekannt war.

Elementarer Beweis des GWSdM für $p = 1/2$ I

Wir schreiben $f(n) \sim_\infty g(n)$ für $\lim_{n\to\infty} f(n)/g(n) = 1$.

Zu zeigen: $\Pr[a \le H^*_{2n} \le b] \sim_\infty \int_a^b \varphi(t)\,\mathrm{d}t$.

Mit $\mathbb{E}[H_{2n}] = n$ und $\mathrm{Var}[H_{2n}] = n/2$ ($p = 1/2$!) erhalten wir

$$H^*_{2n} = \frac{H_{2n} - n}{\sqrt{n/2}}$$

Es folgt

$$\begin{aligned}
& \Pr[a \le H^*_{2n} \le b] \\
= \; & \sum_{i=\alpha}^{\beta} \Pr[H_{2n} = n + i] \quad \alpha = \left\lceil a\sqrt{n/2} \right\rceil, \beta = \left\lfloor b\sqrt{n/2} \right\rfloor \\
= \; & \sum_{i=\alpha}^{\beta} \binom{2n}{n+i} \cdot \left(\frac{1}{2}\right)^{2n} =: \sum_{i=\alpha}^{\beta} p_{n,i}
\end{aligned}$$

Elementarer Beweis des GWSdM für $p = 1/2$ II

Es gilt

$$\max_i p_{n,i} \leq p_n^* := \binom{2n}{n} \cdot \left(\frac{1}{2}\right)^{2n} = \frac{(2n)!}{(n!)^2} \cdot \left(\frac{1}{2}\right)^{2n}$$

Wir setzen

$$q_{n,i} := \frac{p_{n,i}}{p_n^*}$$

d.h.

$$\Pr[a \leq H_{2n}^* \leq b] = \sum_{i=\alpha}^{\beta} p_n^* \cdot q_{n,i}$$

Wir schätzen erst p_n^*, dann $q_{n,i}$ und zuletzt $\Pr[a \leq H_{2n}^* \leq b]$ ab.

Elementarer Beweis des GWSdM für $p = 1/2$ III

Abschätzung von p_n^*:

Mit der Stirling'schen Approximation für $n!$ gilt

$$p_n^* \sim_\infty \frac{(2n)^{2n} \cdot e^{-2n} \cdot \sqrt{2\pi \cdot 2n}}{(n^n \cdot e^{-n} \cdot \sqrt{2\pi n})^2} \cdot \left(\frac{1}{2}\right)^{2n} = \frac{1}{\sqrt{\pi n}}.$$

Elementarer Beweis des GWSdM für $p = 1/2$ IV

Abschätzung von $q_{n,i}$:

Für $i > 0$ gilt

$$\begin{aligned} q_{n,i} = \frac{\binom{2n}{n+i} \cdot \left(\frac{1}{2}\right)^{2n}}{\binom{2n}{n} \cdot \left(\frac{1}{2}\right)^{2n}} &= \frac{(2n)! \cdot n! \cdot n!}{(n+i)! \cdot (n-i)! \cdot (2n)!} \\ = \frac{\prod_{j=0}^{i-1}(n-j)}{\prod_{j=1}^{i}(n+j)} &= \prod_{j=1}^{i} \frac{n-j+1}{n+j} = \prod_{j=1}^{i} \left(1 - \frac{2j-1}{n+j}\right). \end{aligned}$$

Für $i < 0$ gilt $q_{n,-i} = q_{n,i}$.

Wir schätzen $\ln \prod_{j=1}^{i} \left(1 - \frac{2j-1}{n+j}\right)$ ab.

Elementarer Beweis des GWSdM für $p = 1/2$ V

Mit $1 - 1/x \le \ln x \le x - 1$ für $x > 0$ gilt

$$\begin{aligned}
\ln\left(\prod_{j=1}^{i}\left(1 - \frac{2j-1}{n+j}\right)\right) &= \sum_{j=1}^{i} \ln\left(1 - \frac{2j-1}{n+j}\right) \\
&\le -\sum_{j=1}^{i} \frac{2j-1}{n+j} \ \le \ -\sum_{j=1}^{i} \frac{2j-1}{n+i} \\
&= -\frac{i(i+1)-i}{n+i} \ = \ -\frac{i^2}{n} + \frac{i^3}{n(n+i)} \\
&= -\frac{i^2}{n} + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right),
\end{aligned}$$

da $i = \mathcal{O}(\sqrt{n})$ für $\alpha \le i \le \beta$.

Elementarer Beweis des GWSdM für $p = 1/2$ VI

Ebenso erhalten wir

$$\ln\left(\prod_{j=1}^{i}\left(1-\frac{2j-1}{n+j}\right)\right) \geq \sum_{j=1}^{i}\left(1-\left(1-\frac{2j-1}{n+j}\right)^{-1}\right)$$

$$= \sum_{j=1}^{i}\frac{-2j+1}{n-j+1} \geq -\sum_{j=1}^{i}\frac{2j-1}{n-i}$$

$$= -\frac{i^2}{n-i} \;=\; -\frac{i^2}{n}-\mathcal{O}\left(\frac{1}{\sqrt{n}}\right).$$

Zusammen gilt

$$e^{-\frac{i^2}{n-i} = -\frac{i^2}{n}-\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)} \leq q_{n,i} \leq e^{-\frac{i^2}{n}+\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)}$$

Wegen $e^{\pm\mathcal{O}(1/\sqrt{n})} = 1 \pm o(1)$ folgt daraus

$$q_{n,i} \;\sim_\infty\; e^{-i^2/n}$$

Elementarer Beweis des GWSdM für $p = 1/2$ VII

Abschätzung von $\Pr[a \le H_{2n}^* \le b] = \sum_{i=\alpha}^{\beta} p_n^* \cdot q_{n,i}$:

$$\Pr[a \le H_{2n}^* \le b] \;\sim_\infty\; \frac{1}{\sqrt{\pi n}} \cdot \sum_{i=\alpha}^{\beta} e^{-i^2/n}.$$

Elementarer Beweis des GWSdM für $p = 1/2$ VIII

Letzter Schritt:

Mit $\delta := \sqrt{2/n}$ erhalten wir

$$\Pr[a \leq H_{2n}^* \leq b] \ \sim_\infty \ \frac{1}{\sqrt{2\pi}} \cdot \sum_{i=\alpha}^{\beta} \delta e^{-(i\delta)^2 \cdot \frac{1}{2}}.$$

Die rechte Seite entspricht einer Näherung für

$$\int_a^b \varphi(t)\,\mathrm{d}t = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-t^2/2}\,\mathrm{d}t$$

durch Aufteilung der integrierten Fläche in Balken der Breite δ.

Für $n \to \infty$ konvergiert die Fläche der Balken gegen das Integral und wir erhalten

$$\Pr[a \leq H_{2n}^* \leq b] \ \sim_\infty \ \int_a^b \varphi(t)\,\mathrm{d}t$$

q. e. d.

6. Die Exponentialverteilung

Exponentialverteilung I

Beispiel 15

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

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

Damit ruft ein Kunde in ein Intervall mit W'keit $p_n = \dfrac{k}{n}$ an. Die Zeit X bis zum ersten Anruf ist geometrisch verteilt:

$$\Pr[X \leq a] = \sum_{i=1}^{a} (1-p_n)^{(i-1)} \cdot p_n = 1 - (1-p_n)^a$$

Exponentialverteilung II

Wir berechnen die W'keit, dass für $k = 3$ der erste Anruf in der ersten Hälfte des Tages stattfindet.

Bei einer Einteilung in n Intervallen ist diese W'keit gleich $\Pr[X \leq n/2]$.

Die folgende Tabelle zeigt $\Pr[X \leq n/2]$ für $k = 3$ und verschiedene Werte von n:

n	$\Pr[X \leq 5]$
6	0.8750
12	0.8221
24	0.7986
48	0.7875
$24 * 60$	0.7772

Frage: Zu welchem Wert konvergiert diese Folge?

Die Exponentialverteilung I

Definition 16

Eine Zufallsvariable X heißt exponentialverteilt mit dem Parameter $\lambda > 0$, wenn sie die Dichte

$$f(x) = \begin{cases} \lambda \cdot e^{-\lambda x} & \text{falls } x \geq 0, \\ 0 & \text{sonst} \end{cases}$$

besitzt.

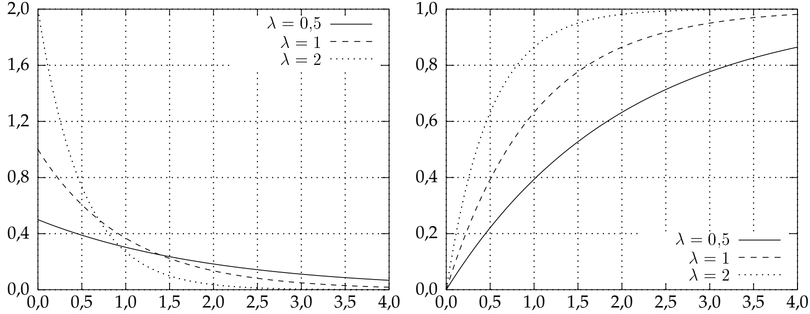
Wir schreiben auch $X \sim Exp(\lambda)$.

Für die Verteilungsfunktion gilt für $x \geq 0$

$$F(x) = \int_0^x \lambda \cdot e^{-\lambda t}\,\mathrm{d}t = \left[-e^{-\lambda t}\right]_0^x = 1 - e^{-\lambda x}.$$

sowie $F(x) = 0$ für $x < 0$.

Die Exponentialverteilung II



Dichte und Verteilung der Exponentialverteilung

Die Exponentialverteilung III

Satz 17

Sei $\lambda > 0$ *und sei* $X_n \sim Geo(\lambda/n)$ *eine Folge von Zufallsvariablen. Die Folge* $Y_n = X_n/n$ *ist asymptotisch exponentialverteilt mit Parameter* λ*, d.h.* $Y_n \sim Exp(\lambda)$ *für* $n \to \infty$*.*

Vergleiche mit: $\mathrm{Po}(\lambda) = \lim_{n\to\infty} \mathrm{Bin}(n, \lambda/n)$.

In unserem Beispiel: $\lambda = 3$ und

$$\Pr[X \le 1/2] = F(1/2) = 1 - e^{-3/2} \approx 0.7768 \; .$$

Die Exponentialverteilung IV

Beweis: Für ein beliebiges $k \in \mathbb{N}$ gilt

$$\begin{aligned}\Pr[X_n \le kn] &= \sum_{i=1}^{kn} (1-p_n)^{i-1} \cdot p_n \;=\; p_n \cdot \sum_{i=0}^{kn-1} (1-p_n)^i \\ &= p_n \cdot \frac{1-(1-p_n)^{kn}}{p_n} \;=\; 1 - \left(1 - \frac{\lambda}{n}\right)^{kn}.\end{aligned}$$

und damit

$$\begin{aligned}\lim_{n\to\infty} \Pr[Y_n \le t] = \lim_{n\to\infty} \Pr[X_n \le t \cdot n] &= \lim_{n\to\infty} \left[1 - \left(1 - \frac{\lambda}{n}\right)^{tn}\right] \\ &= 1 - e^{-\lambda t}.\end{aligned}$$

Die Folge Y_n geht also für $n \to \infty$ in eine exponentialverteilte Zufallsvariable mit Parameter λ über.

Exponentialverteilung: Erwartungswert

$$\begin{aligned}\mathbb{E}[X] &= \int_0^\infty t \cdot \lambda \cdot e^{-\lambda t}\,\mathrm{d}t \\ &= \Big[t \cdot (-e^{-\lambda t})\Big]_0^\infty + \int_0^\infty e^{-\lambda t}\,\mathrm{d}t \\ &= 0 + \left[-\frac{1}{\lambda} \cdot e^{-\lambda t}\right]_0^\infty = \frac{1}{\lambda}\,.\end{aligned}$$

Exponentialverteilung: Varianz

Analog erhalten wir

$$\begin{aligned}\mathbb{E}[X^2] &= \int_0^\infty t^2 \cdot \lambda \cdot e^{-\lambda t}\,\mathsf{d}t \\ &= \left[t^2 \cdot (-e^{-\lambda t})\right]_0^\infty + \int_0^\infty 2t \cdot e^{-\lambda t}\,\mathsf{d}t \\ &= 0 + \frac{2}{\lambda} \cdot \mathbb{E}[X] = \frac{2}{\lambda^2}\end{aligned}$$

und somit

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

Lemma 18 (Skalierung exponentialverteilter Variablen)

Wenn $X \sim Exp(\lambda)$ *und* $Y := aX$ *für* $a > 0$ *dann* $Y \sim Exp(\lambda/a)$.

Beweis:

$$\begin{aligned} F_Y(x) &= \Pr[Y \le x] = \Pr[aX \le x] \\ &= \Pr\left[X \le \frac{x}{a}\right] = F_X\left(\frac{x}{a}\right) \\ &= 1 - e^{-\lambda x/a}\,. \end{aligned}$$

$\square$

Gedächtnislosigkeit der Exponentialverteilung II

Satz 19 (Gedächtnislosigkeit)

Eine kontinuierliche Zufallsvariable X *mit Wertebereich* $\mathbb{R}^+$ *ist* *genau dann* *exponentialverteilt, wenn für alle* $x, y > 0$ *gilt, dass*

$$\Pr[X > x + y \mid X > y] = \Pr[X > x]\,. \qquad (*)$$

Beweis:

Sei X exponentialverteilt mit Parameter λ. Dann gilt

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

Gedächtnislosigkeit der Exponentialverteilung III

Beweis (Forts.):

Sei umgekehrt X eine kontinuierliche Zufallsvariable, die die Gleichung $(*)$ erfüllt. Definiere $g(x) := \Pr[X > x]$.

Wir zeigen $g(x) = e^{-\lambda x}$ für alle $x \in \mathbb{Q}^+$.

Aufgrund der Stetigkeit folgt daraus $g(x) = e^{-\lambda x}$ für alle $x \in \mathbb{R}^+$.

Aus (*) folgt

$$
\begin{aligned}
g(x+y) &= \Pr[X > x+y] \\
&= \Pr[X > x+y \mid X > y] \cdot \Pr[X > y] \\
&= \Pr[X > x] \cdot \Pr[X > y] = g(x)g(y)\,.
\end{aligned}
$$

Gedächtnislosigkeit der Exponentialverteilung IV

Beweis (Forts.):

Durch wiederholte Anwendung erhalten wir

$$g(1) = g\Big(\underbrace{\frac{1}{n} + \cdots + \frac{1}{n}}_{n\text{-mal}}\Big) = \left(g\Big(\frac{1}{n}\Big)\right)^n \text{ für alle } n \in \mathbb{N}$$

und somit insbesondere auch $g(1/n) = (g(1))^{1/n}$.

Da X positiv, gibt es $n \in \mathbb{N}$ mit $g(1/n) > 0$.

Wegen $0 < g(1) \leq 1$ gibt es auch ein $\lambda \geq 0$ mit $g(1) = e^{-\lambda}$.

Nun gilt für beliebige $p, q \in \mathbb{N}$

$$g(p/q) = g(1/q)^p = g(1)^{p/q}$$

$\square$

Gedächtnislosigkeit der Exponentialverteilung V

Beispiel 20

Das Cäsium-Isotop $^{134}_{55}\mathrm{Cs}$ besitzt eine mittlere Lebensdauer von ungefähr 3,03 Jahren oder $1{,}55 \cdot 10^6$ Minuten.

Die Zufallsvariable X messe die Lebenszeit eines bestimmten $^{134}_{55}\mathrm{Cs}$-Atoms.

X ist exponentialverteilt mit dem Parameter

$$\lambda = \frac{1}{\mathbb{E}[X]} = \frac{1}{1{,}55 \cdot 10^6} \approx 0{,}645 \cdot 10^{-6} \ \left[\frac{1}{\text{min}}\right]$$

Da λ den Kehrwert einer Zeit als Einheit besitzt, spricht man von der Zerfallsrate.

Auch bei anderen Anwendungen ist es üblich, λ als Rate einzuführen.

Minimum von Exponentialverteilungen I

Satz 21

Die Zufallsvariablen $X_1, \ldots, X_n$ *seien unabhängig und exponentialverteilt mit den Parametern* $\lambda_1, \ldots, \lambda_n$. *Dann ist auch* $X := \min\{X_1, \ldots, X_n\}$ *exponentialverteilt mit dem Parameter* $\lambda_1 + \ldots + \lambda_n$.

Beweis:

Der allgemeine Fall folgt mittels Induktion aus dem Fall $n = 2$. Für die Verteilungsfunktion F_X gilt:

$$\begin{aligned}
1 - F_X(t) &= \Pr[X > t] = \Pr[\min\{X_1, X_2\} > t] \\
&= \Pr[X_1 > t, X_2 > t] \\
&= \Pr[X_1 > t] \cdot \Pr[X_2 > t] \\
&= e^{-\lambda_1 t} \cdot e^{-\lambda_2 t} = e^{-(\lambda_1 + \lambda_2) t}.
\end{aligned}$$

Minimum von Exponentialverteilungen II

Anschaulich besagt Satz 21:

> *Wenn man auf das erste Eintreten eines Ereignisses aus mehreren unabhängigen Ereignissen wartet, dann addieren sich die Raten.*

Wenn beispielsweise ein Atom die Zerfallsrate λ besitzt, so erhalten wir bei n Atomen die Zerfallsrate $n\lambda$.

Poisson-Prozess I

Wir wissen:

Wenn der zeitliche Abstand der Treffer geometrisch verteilt ist, so ist ihre Anzahl in einer festen Zeitspanne binomialverteilt.

Im Grenzwert $n \to \infty$ mit Trefferw'keit $p_n = \lambda/n$ konvergiert

- die Binomialverteilung gegen die Poisson-Verteilung und
- die geometrische Verteilung gegen die Exponentialverteilung.

Im Grenzwert $n \to \infty$ erwarten wir deshalb:

Wenn man Ereignisse zählt, deren zeitlicher Abstand exponentialverteilt ist, so ist die Anzahl dieser Ereignisse in einer festen Zeitspanne Poisson-verteilt.

Poisson-Prozess II

Seien $T_1, T_2 \ldots$ unabhängige Zufallsvariablen.

T_i modelliert die Zeit, die zwischen Treffer $i-1$ und i vergeht.

Für den Zeitpunkt $t > 0$ definieren wir

$$X(t) := \max\{n \in \mathbb{N} \mid T_1 + \ldots + T_n \leq t\}.$$

$X(t)$ modelliert die Anzahl der Treffer von Zeit 0 bis t.

Satz 22 (ohne Beweis)

$X(t) \sim \mathrm{Po}(t\lambda)$ *genau dann, wenn* $T_1, T_2, \ldots \sim Exp(\lambda)$*.*

Eine Familie $(X(t))_{t>0}$ von Zufallsvariablen nennt man allgemein einen stochastischen Prozess.

Der Prozess, bei dem $T_1, T_2, \ldots$ unabhängige, exponentialverteilte Zufallsvariablen sind, heißt Poisson-Prozess.

Beispiel 23

Eine Menge von Jobs werden auf einem Prozessor sequentiell abgearbeitet.

Die Laufzeiten der Jobs seien unabhängig und exponentialverteilt mit Parameter $\lambda = 1/30[1/s]$.

Jeder Job benötigt also im Mittel $30s$.

Gemäß Satz 22 ist die Anzahl von Jobs, die in einer Minute vollständig ausgeführt werden, Poisson-verteilt mit Parameter $t\lambda = 60 \cdot (1/30) = 2$.

Die Wahrscheinlichkeit, dass in einer Minute höchstens ein Job abgearbeitet wird, beträgt für $t\lambda = 2$

$$e^{-t\lambda} + t\lambda e^{-t\lambda} \approx 0{,}406\ .$$

Approximationen der Binomialverteilung I

Sei $H_n \sim \mathrm{Bin}(n, p)$ mit der Verteilungsfunktion F_n.

Für $n \to \infty$ gilt (Korollar 14)

$$F_n(t) = \Pr[H_n/n \le t/n] \longrightarrow \Phi\left(\frac{t/n - p}{\sqrt{p(1-p)/n}}\right) = \Phi\left(\frac{t - np}{\sqrt{p(1-p)n}}\right).$$

F_n kann somit für große n durch Φ approximieren.

Faustregel: Gute Approximation wenn $np \ge 5$ und $n(1-p) \ge 5$.

Für die Berechnung von Φ liegen effiziente numerische Methoden vor.

Approximationen der Binomialverteilung II

Beispiel 24

Die W'keit, mit der bei 10^6 Würfen mit einem idealen Würfel mehr als 500500-mal eine gerade Augenzahl fällt beträgt

$$T := \sum_{i=5{,}005\cdot 10^5}^{10^6} \binom{10^6}{i} \left(\frac{1}{2}\right)^{10^6} .$$

Die Approximation durch die Normalverteilung ergibt

$$\begin{aligned} T &\approx 1 - \Phi\left(\frac{5{,}005\cdot 10^5 - 5\cdot 10^5}{\sqrt{2{,}5\cdot 10^5}}\right) \\ &= 1 - \Phi\left(\frac{5\cdot 10^2}{5\cdot 10^2}\right) = 1 - \Phi(1) \approx 0{,}1573 . \end{aligned}$$

Approximationen der Binomialverteilung III

Oft führt man eine Stetigkeitskorrektur durch.

Zur Berechnung von $\Pr[X \le x]$ für $X \sim \mathrm{Bin}(n, p)$ setzt man

$$\Pr[X \le x] \approx \Phi\left(\frac{x + 0{,}5 - np}{\sqrt{np(1-p)}}\right)$$

statt

$$\Pr[X \le x] \approx \Phi\left(\frac{x - np}{\sqrt{np(1-p)}}\right)$$

an.

Approximationen der Binomialverteilung IV

Der Korrekturterm läßt sich in der Histogramm-Darstellung der Binomialverteilung veranschaulichen.

Die Binomialverteilung wird durch Balken angegeben, deren Fläche in etwa der Fläche unterhalb der Dichte φ von $\mathcal{N}(0,1)$ entspricht.

Wenn man die Fläche der Balken mit „$X \leq x$“ durch das Integral von φ approximieren will, soll man bis zum Ende des Balkens für „$X = x$“ integrieren (nicht nur bis zur Mitte).

Dafür sorgt der Korrekturterm $0{,}5$.

Approximationen der Binomialverteilung V

Zusammenfassung:

- Approximation durch die Normalverteilung.

  $\mathrm{Bin}(n,p)$ wird approximiert durch $\Phi((t-np)/\sqrt{p(1-p)n})$.

  Gut für grosses n.

  Faustregel: Gut wenn $np \geq 5$ und $n(1-p) \geq 5$.

- Approximation durch die Poisson-Verteilung.

  $\mathrm{Bin}(n,p)$ wird approximiert durch $\mathrm{Po}(np)$.

  Gut für seltene Ereignisse, d. h. wenn $np << n$.

  Faustregel: gut wenn $n \geq 30$ und $p \leq 0{,}05$.