

## Kapitel V

## Der lokale Grenzwertsatz

*Ich glaube nicht an Fügung und Schicksal, als Techniker bin ich gewohnt mit den Formeln der Wahrscheinlichkeit zu rechnen. [...]*

*„Hast du gewusst?“ frage ich, „dass die Mortalität*

*bei Schlangenbiss nur drei bis zehn Prozent beträgt?“ [...]*

*Ich fragte Hanna, wieso sie nicht an Statistik glaubt, statt dessen aber an Schicksal und Derartiges. „Du mit deiner Statistik!“ sagt sie. „Wenn ich hundert Töchter hätte, alle von einer Viper gebissen, dann ja! Dann würde ich nur drei bis zehn Töchter verlieren. Erstaunlich wenig!“*

Max Frisch (1911–1991), *Homo faber* (1957)

## Inhalt dieses Kapitels V

- 1 Kontinuierliche Wahrscheinlichkeitsverteilungen
  - Kontinuierliche Verteilungen
  - Wahrscheinlichkeitsräume
  - Eindimensionale Normalverteilung
- 2 Erwartung, Varianz, Streuung
  - Kumulative Verteilungsfunktion
  - Zufallsvariablen, Erwartung und Varianz
  - Verteilungen und ihre Kenngrößen
- 3 Der lokale Grenzwertsatz (LGS)
  - Von der Binomial- zur Normalverteilung
  - Der lokale Grenzwertsatz  $B(n, t) \approx N(\mu, \sigma^2)$
  - Erste Anwendungsbeispiele
- 4 Fazit: der lokale Grenzwertsatz
  - Zusammenfassung und Verständnisfragen
  - Weitere Aufgaben und Anwendungsbeispiele
  - Kleine Stichprobe und Intervallschätzung

Wir kennen bereits einen einfachen und nützlichen Grenzwertsatz: Das **Gesetz der kleinen Zahlen** besagt  $B(n, \lambda/n) \rightarrow P(\lambda)$  für  $n \rightarrow \infty$ . Bei seltenen Ereignissen nähert sich die Binomialverteilung  $B(n, t)$  der (viel einfacheren) Poisson-Verteilung  $P(nt)$ . Wir nutzen genauer:

$$B(n, t) \approx P(nt) \quad \text{mit Fehlerschranke} \quad \|B(n, t) - P(nt)\| \leq nt^2$$

Dies lohnt sich vor allem für kleine Wkten  $t$ , also seltene Ereignisse. Der **lokale Grenzwertsatz** (LGS) hingegen nützt für beliebige  $0 < t < 1$  und approximiert Binomialverteilungen durch die **Normalverteilung**: Es gilt  $B(n, t) \approx N(\mu, \sigma^2)$  mit  $\mu = nt$  und  $\sigma^2 = nt(1 - t)$ , genauer:

$$\sum_{k=a}^b \binom{n}{k} t^k (1-t)^{n-k} = \int_{\xi=\alpha}^{\beta} \frac{e^{-\xi^2/2}}{\sqrt{2\pi}} d\xi + \delta$$

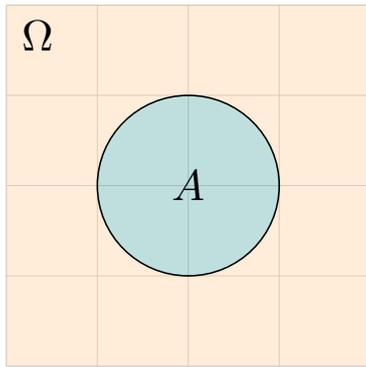
mit Integrationsgrenzen  $\alpha = (a - 1/2 - \mu)/\sigma$  und  $\beta = (b + 1/2 - \mu)/\sigma$ . Hierbei ist  $\mu = nt$  die **Erwartung** und  $\sigma = \sqrt{nt(1-t)}$  die **Streuung**. Der Approximationsfehler  $\delta$  ist beschränkt durch  $|\delta| < 1/(6\sigma)$  für  $\sigma \geq 5$ . Für  $n \rightarrow \infty$  gilt  $\sigma \rightarrow \infty$ , also  $\delta \rightarrow 0$ : Die Näherung wird beliebig gut.

## Vorgehensweise

Poissons Gesetz der kleinen Zahlen (Satz U3E) aus dem letzten Kapitel klärt präzise den Übergang von Binomial- zu Poisson-Verteilungen. Der lokale Grenzwertsatz V3A klärt nun den Übergang von Binomial- zu Normalverteilungen und erlaubt erfreulich genaue Fehlerschranken. Im nächsten Kapitel werden wir allgemeiner für Summen unabhängiger Zufallsvariablen den zentralen Grenzwertsatz (ZGS W1D) formulieren.

Diese Grenzwertsätze haben in der Wahrscheinlichkeitsrechnung eine zentrale Bedeutung: Im Kleinen, bei einzelnen Experimenten regiert der Zufall ungezügelt; im Großen, bei vielen Experimenten muss er sich den Gesetzen beugen. Als grundlegende Näherungen sind diese Werkzeuge deshalb in allen praktischen Anwendungen der Statistik unentbehrlich. Stochastische Werkzeuge helfen, mindern Risiken, optimieren Chancen!

Auch in diesem Kapitel formuliere ich Grenzwertsätze gleich mit möglichst expliziten und praktisch anwendbaren Fehlerschranken. Dies ermöglicht, die üblichen Approximationen nicht nur blind oder naiv nach Gefühl anzuwenden, sondern bequem ihre Güte abzuschätzen.



**Aufgabe:** Sie werfen zufällig und gleichverteilt Dartpfeile ins Quadrat  $\Omega = [-2, +2] \times [-2, +2]$ .

Mit welcher Wahrscheinlichkeit treffen Sie die Kreisscheibe  $A = \{ x \in \Omega \mid |x| \leq 1 \}$ ?

**Lösung:** Hier ist  $\text{vol}_2(\Omega) = 16$  und  $\text{vol}_2(A) = \pi$ , also gilt  $\mathbf{P}(A) = \text{vol}_2(A) / \text{vol}_2(\Omega) = \pi/16 \approx 0.196$ .

Die Wahrscheinlichkeit eines Ereignisses  $A \subseteq \Omega$  ist hier proportional zu seiner Fläche  $\text{vol}_2(A)$ . Die Wkt genau den Mittelpunkt  $(0, 0)$  zu treffen ist gleich Null! Gleiches gilt für jeden Punkt: Die Elementarwahrscheinlichkeiten  $\mathbf{P}(\{\omega\})$  sind alle gleich Null! Das Wahrscheinlichkeitsmaß  $\mathbf{P}$  entspricht hier nicht der Anzahl (Summation), sondern dem Volumen (Integration).

So können Sie experimentell  $\pi$  bestimmen: Werfen Sie 1000 Reiskörner auf das Quadrat  $\Omega$  und zählen Sie... Wenn Sie etwa 200 in  $A$  finden, so spräche dies für  $\pi \approx 0.2 \cdot 16 = 3.2$ . (Für  $\pi$  gibt es genauere und schnellere Methoden.) Diese Art von *Monte-Carlo Simulation* wird in der Praxis häufig vor allem für komplexe Modelle eingesetzt, die von vielen Parametern abhängen. Oft sind die Rechnungen in Einzelfällen (für fest vorgegebene Parameterwerte) leicht möglich, aber eine geschlossene Formulierung für alle Fälle nicht oder nur schwierig möglich. Durch zufällige Wahl der Parameter und wiederholte Simulation gewinnt man einen Eindruck des typischen Verhaltens. Die Methode geht zurück auf Stanisław Ulam und John von Neumann im Manhattan Projekt.

Kontinuierliche Gleichverteilung

Diskrete Erinnerung: Sei  $\Omega$  eine endliche Menge mit Elementezahl  $|\Omega|$ , kurz  $0 < |\Omega| < \infty$ . Hierauf haben wir die **diskrete Gleichverteilung** T1E

$$\mathbf{P} : \mathfrak{P}(\Omega) \rightarrow [0, 1] : A \mapsto \mathbf{P}(A) = \frac{|A|}{|\Omega|}.$$

Definition V1A: kontinuierliche Gleichverteilung

Sei  $\Omega \subseteq \mathbb{R}^n$  eine messbare Menge mit  $0 < \text{vol}_n(\Omega) < \infty$ .

Hierauf haben wir die **kontinuierliche Gleichverteilung**

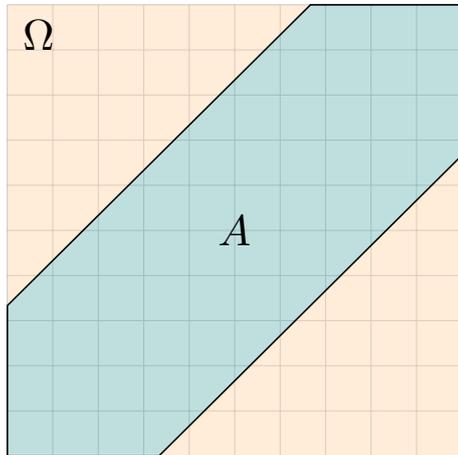
$$\mathbf{P} : \mathcal{B}(\Omega) \rightarrow [0, 1] : A \mapsto \mathbf{P}(A) = \frac{\text{vol}_n(A)}{\text{vol}_n(\Omega)}.$$

Im diskreten Fall ist  $\mathfrak{P}(\Omega)$  die Familie *aller* Teilmengen  $A \subseteq \Omega$ . [U207](#) Im kontinuierlichen Fall ist  $\mathcal{B}(\Omega)$  die Familie aller *messbaren* Teilmengen  $A \subseteq \Omega$ . [A102](#) Diese wird erzeugt von Quadern  $Q \subseteq \mathbb{R}^n$ ; zum Beispiel sind alle offenen und alle abgeschlossenen Mengen im  $\mathbb{R}^n$  messbar.

Extrem exotische Teilmengen  $A \subseteq \Omega$  lassen sich nicht messen. Aus diesem Grund *müssen* wir das WMaß  $\mathbf{P}$  einschränken auf die Familie  $\mathcal{B}(\Omega) \subsetneq \mathfrak{P}(\Omega)$  der messbaren Teilmengen  $A \subseteq \Omega$ . Diese Vorsichtsmaßnahme ist technischer Natur und spielt im Folgenden keine weitere Rolle.

Zwei Personen verabreden sich zwischen 12 und 13 Uhr zum Essen: Jede kommt zu einem zufälligen Zeitpunkt (unabhängig, gleichverteilt). Jede wartet bis zu 20 Minuten, spätestens aber bis 13 Uhr.

**Aufgabe:** Wie groß ist die Wkt, dass sich beide treffen?



**Lösung:** Ankunftszeiten  $(t_1, t_2) \in [0, 1]^2 = \Omega$ . Ereignis „beide Personen treffen sich“:

$$A = \left\{ (t_1, t_2) \in \Omega \mid |t_1 - t_2| \leq \frac{1}{3} \right\}$$

Bei kontinuierlicher Gleichverteilung ist die Wkt proportional zur Fläche:

$$\mathbf{P}(A) = \frac{\text{vol}_2(A)}{\text{vol}_2(\Omega)} = \frac{5}{9} \approx 56\%$$

Dies ist kein diskreter Wahrscheinlichkeitsraum! Wir können nicht jedem Punkt  $\omega \in \Omega$  eine feste positive Wahrscheinlichkeit  $\mathbf{P}(\{\omega\}) > 0$  zuordnen, denn dann wäre  $\mathbf{P}(\Omega) = \infty$ . Wir müssen deshalb jedem Punkt  $\omega \in \Omega$  die Wahrscheinlichkeit  $\mathbf{P}(\{\omega\}) = 0$  zuordnen. Das geht aber nicht mit den Rechenregeln für diskrete WRäume, denn dann wäre  $\mathbf{P}(\Omega) = 0$ . Näherung und Simulation mit diskreten Ankunftszeiten, etwa gleichverteilt in  $\{0, 1, \dots, 59\}$ .

## Was bedeutet hier Gleichverteilung?

Ausführlich: Der Zeitpunkt  $t_1$  liegt im Intervall  $[0, 1]$  der Länge 1. Die Wahrscheinlichkeit, dass  $t_1$  in  $[a, b]$  liegt, ist demnach

$$\mathbf{P}(t_1 \in [a, b]) = b - a.$$

Ebenso für  $t_2$ . Dank Unabhängigkeit ist die Wahrscheinlichkeit, dass  $(t_1, t_2)$  in  $[a_1, b_1] \times [a_2, b_2]$  liegt, gegeben durch das Produkt:

$$\begin{aligned} \mathbf{P}((t_1, t_2) \in [a_1, b_1] \times [a_2, b_2]) &= \mathbf{P}(t_1 \in [a_1, b_1], t_2 \in [a_2, b_2]) \\ &= \mathbf{P}(t_1 \in [a_1, b_1]) \cdot \mathbf{P}(t_2 \in [a_2, b_2]) = (b_1 - a_1)(b_2 - a_2). \end{aligned}$$

Das legt die Wkten für alle Rechtecke fest als Produktmaß.

Allgemein gilt: Bei Gleichverteilung ist die Wahrscheinlichkeit, dass  $(t_1, t_2)$  in  $A$  liegt, proportional zum Flächeninhalt  $\text{vol}_2(A)$ :

$$\mathbf{P}(A) = \frac{\text{vol}_2(A)}{\text{vol}_2(\Omega)}$$

Im Beispiel gilt  $\text{vol}_2(\Omega) = 1$  und  $\text{vol}_2(\Omega \setminus A) = 4/9$  und  $\text{vol}_2(A) = 5/9$ . Interpretation: In etwa 56% aller Fälle werden sich die beiden treffen.

Im diskreten Fall ist es oft nützlich oder meist auch nötig, neben der Gleichverteilung noch andere interessante Verteilungen zu betrachten.

Eine **diskrete Wahrscheinlichkeitsverteilung** auf einer Menge  $\Omega$  ist gegeben durch ihre Elementarwahrscheinlichkeiten, also eine Funktion

$$p : \Omega \rightarrow [0, 1] \quad \text{mit Gesamtmasse} \quad \sum_{x \in \Omega} p(x) = 1.$$

Diese definiert ein **diskretes Wahrscheinlichkeitsmaß**

$$\mathbf{P} : \mathfrak{P}(\Omega) \rightarrow [0, 1] : A \mapsto \mathbf{P}(A) := \sum_{x \in A} p(x).$$

Hierbei ist die Potenzmenge  $\mathfrak{P}(\Omega)$  die Familie aller Teilmengen  $A \subseteq \Omega$ .

**Beispiel:** Die diskrete Gleichverteilung auf  $\Omega$  mit  $0 < |\Omega| < \infty$  entsteht durch Summation über die konstante Verteilung  $p = 1/|\Omega|$ .

Weitere wichtige Beispiele kennen wir aus dem vorigen Kapitel U: Hypergeometrische, Binomial- und Poisson-Verteilungen auf  $\Omega = \mathbb{N}$ .

Wir betrachten nun neben diskreten auch kontinuierliche WVerteilungen. Die Summe der Elementarwkten wird dabei ersetzt durch ein Integral.

## Kontinuierliche Verteilungen

V106

### Definition V1B: kontinuierliche WDichte und WMaß

Eine **kontinuierliche Wahrscheinlichkeitsverteilung** oder **WDichte** auf einer Ergebnismenge  $\Omega \subseteq \mathbb{R}^n$  ist eine messbare Funktion

$$f : \Omega \rightarrow \mathbb{R}_{\geq 0} \quad \text{mit Gesamtmasse} \quad \int_{\Omega} f(x) dx = 1.$$

Diese definiert ein **kontinuierliches Wahrscheinlichkeitsmaß**

$$\mathbf{P} : \mathcal{B}(\Omega) \rightarrow [0, 1] : A \mapsto \mathbf{P}(A) := \int_A f(x) dx.$$

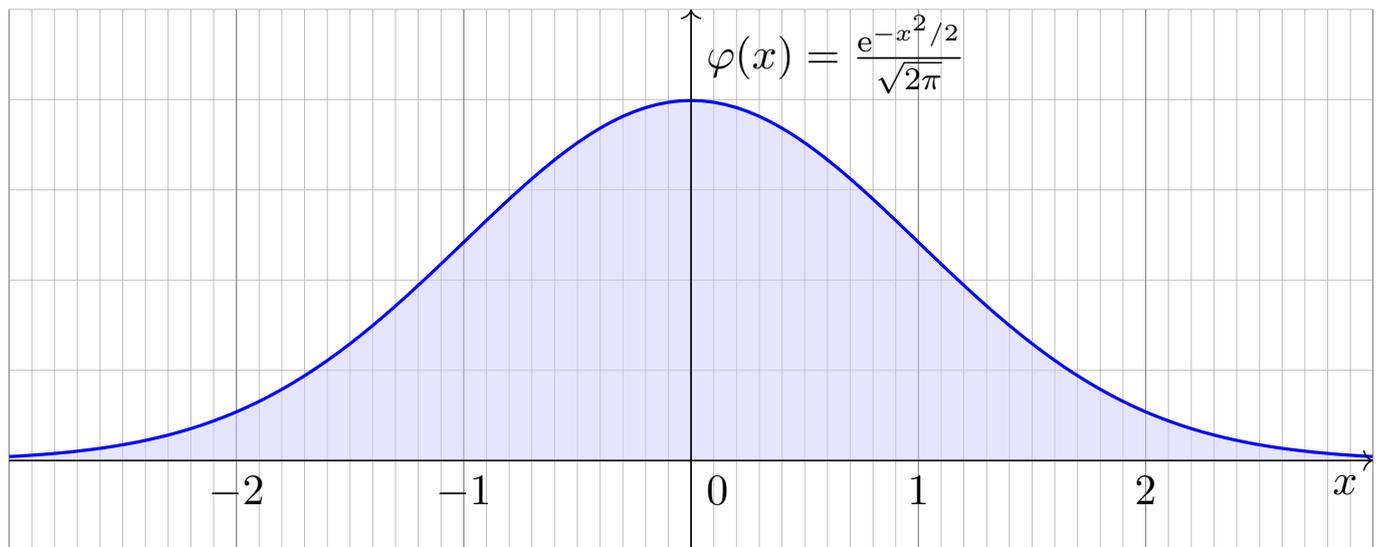
Hierbei ist  $\mathcal{B}(\Omega)$  die Familie aller messbaren Mengen  $A \subseteq \Omega$ .

**Beispiel:** Die Gaußsche Glockenkurve  $\varphi : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$  ist eine WDichte.

😊 Allgemein genügt auch hier  $0 < \int_{\Omega} \tilde{f}(x) dx < \infty$  mit Normierung.

**Beispiel:** Die kont. Gleichverteilung auf  $\Omega \subseteq \mathbb{R}^n$  mit  $0 < \text{vol}_n(\Omega) < \infty$  entsteht durch Integration über die konstante Dichte  $f = 1/\text{vol}_n(\Omega)$ .

In diesem Falle gilt einfach  $\mathbf{P}(A) = \text{vol}_n(A)/\text{vol}_n(\Omega)$  wie in V1A.



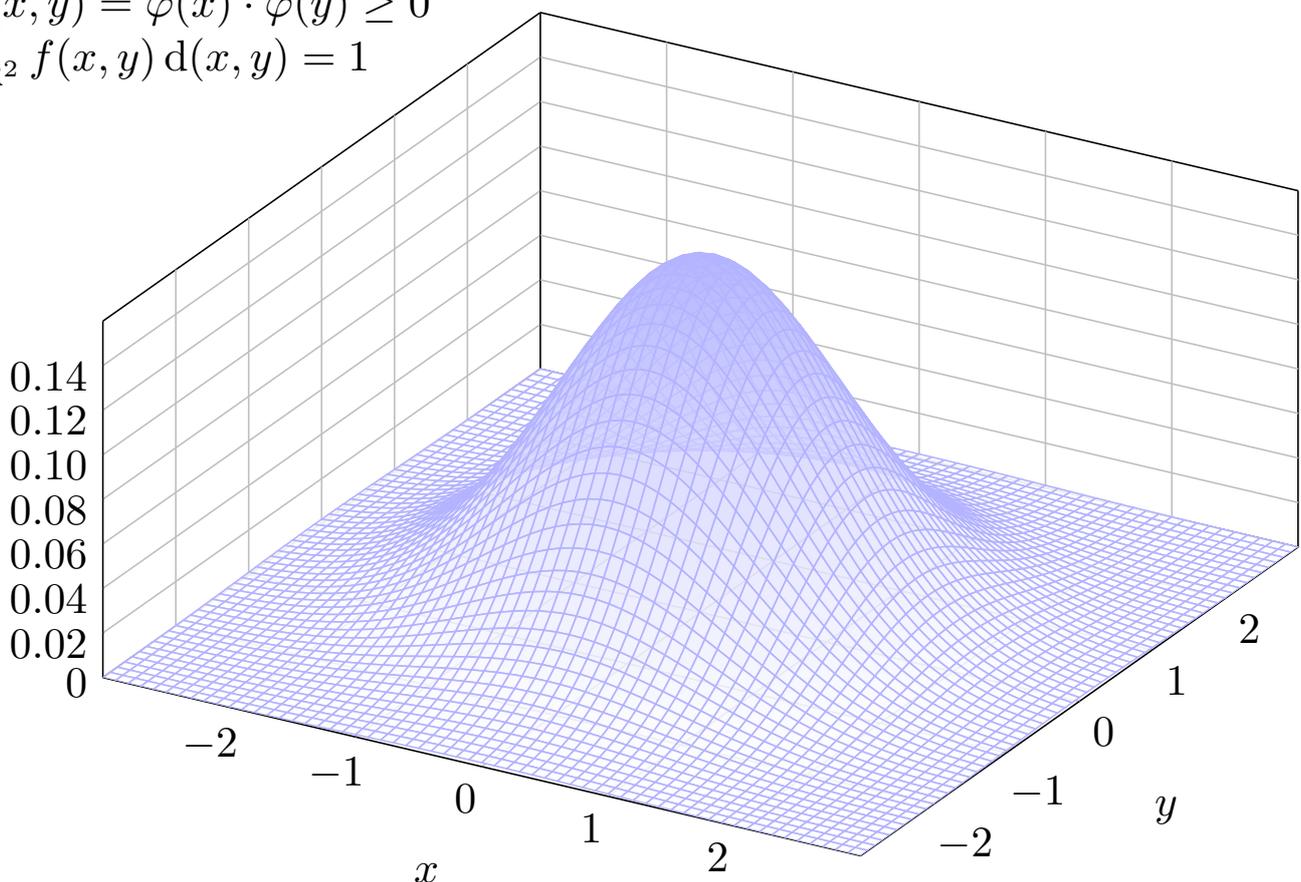
Zentrales Beispiel ist die **Gaußsche Glockenkurve**

$$\varphi : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Dies ist eine Wahrscheinlichkeitsdichte: Es gilt  $\varphi \geq 0$  und  $\int_{\mathbb{R}} \varphi(x) dx = 1$  dank C2G. Wir nennen  $\varphi$  die **Dichte der Standard-Normalverteilung**.

$$f(x, y) = \varphi(x) \cdot \varphi(y) \geq 0$$

$$\int_{\mathbb{R}^2} f(x, y) d(x, y) = 1$$



Bisher kennen wir zwei Typen von Wahrscheinlichkeitsräumen:

(1) Diskrete Wahrscheinlichkeitsräume  $(\Omega, \mathfrak{P}(\Omega), \mathbf{P})$ , wobei

$$\mathbf{P}(A) = \sum_{x \in A} p(x) \quad \text{mit Elementarwkt } p: \Omega \rightarrow [0, 1].$$

(2) Kontinuierliche Wahrscheinlichkeitsräume  $(\Omega, \mathcal{B}(\Omega), \mathbf{P})$ , wobei

$$\mathbf{P}(A) = \int_{x \in A} f(x) dx \quad \text{mit WDichte } f: \mathbb{R}^n \supseteq \Omega \rightarrow \mathbb{R}_{\geq 0}.$$

Im diskreten Fall ist die Wkt in einzelnen, höchstens abzählbar vielen Punkten konzentriert, im zweiten Fall ist die Wkt kontinuierlich verteilt. Auch Mischungen aus diskret und kontinuierlich sind ebenso möglich.

**Ziel:** Wir suchen eine einheitliche Formulierung für alle Fälle.

Hierzu fassen wir die wesentlichen Rechenregeln zusammen und erhalten Kolmogorovs Axiome der Wahrscheinlichkeitsrechnung (V1c).

*Historische Einordnung.* David Hilbert fragte 1900 im sechsten seiner berühmten 23 Probleme: *Wie kann die Physik axiomatisiert werden?* Ursprünglich ging es in erster Linie um WTheorie und Mechanik. Für die Physik ist die allgemeine Frage weiterhin offen, für die WTheorie fand Kolmogorov 1933 eine Lösung. Die mathematischen Grundlagen hierzu lieferte Lebesgue.

Ein Zufallsexperiment beschreiben wir durch seine **Ergebnismenge**  $\Omega$ .

Die beobachtbaren **Ereignisse**  $A \subseteq \Omega$  bilden eine Familie  $\mathcal{A} \subseteq \mathfrak{P}(\Omega)$ .

Das **Wahrscheinlichkeitsmaß** ist eine Abbildung  $\mathbf{P}: \mathcal{A} \rightarrow [0, 1]$ .

Es ordnet jedem Ereignis  $A \in \mathcal{A}$  sein WMaß  $\mathbf{P}(A) \in [0, 1]$  zu.

Um damit vernünftig rechnen zu können, benötigen wir folgendes:

**0** Unmögliches Ereignis: Es gilt  $\emptyset \in \mathcal{A}$  und

$$\mathbf{P}(\emptyset) = 0.$$

**1** Sicheres Ereignis: Es gilt  $\Omega \in \mathcal{A}$  und

$$\mathbf{P}(\Omega) = 1.$$

**2** Komplement: Aus  $A \in \mathcal{A}$  folgt  $(\Omega \setminus A) \in \mathcal{A}$  und

$$\mathbf{P}(\Omega \setminus A) = 1 - \mathbf{P}(A).$$

**3** Monotonie: Für alle  $A, B \in \mathcal{A}$  gilt

$$A \subseteq B \implies \mathbf{P}(A) \leq \mathbf{P}(B).$$

Mengenvereinigung und Maßaddition müssen wir sorgsam formulieren:

4 Vereinigung: Aus  $A, B \in \mathcal{A}$  folgt  $(A \cup B), (A \cap B) \in \mathcal{A}$  und

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B).$$

5 Paarweise Additivität: Für alle  $A, B \in \mathcal{A}$  mit  $A \cap B = \emptyset$  gilt

$$\mathbf{P}(A \sqcup B) = \mathbf{P}(A) + \mathbf{P}(B).$$

6 Endliche Additivität: Aus  $A_1, \dots, A_n \in \mathcal{A}$  folgt  $A_1 \cup \dots \cup A_n \in \mathcal{A}$ , und bei disjunkter Vereinigung addieren sich die Wkten:

$$\mathbf{P}(A_1 \sqcup \dots \sqcup A_n) = \mathbf{P}(A_1) + \dots + \mathbf{P}(A_n).$$

7 Abzählbare Additivität: Aus  $A_0, A_1, A_2, \dots \in \mathcal{A}$  folgt  $\bigcup_{k=0}^{\infty} A_k \in \mathcal{A}$ , und bei disjunkter Vereinigung addieren sich die Wkten:

$$\mathbf{P}\left(\bigcup_{k=0}^{\infty} A_k\right) = \sum_{k=0}^{\infty} \mathbf{P}(A_k) \quad \text{falls } A_i \cap A_j = \emptyset \text{ für } i \neq j.$$

Diese Wunschliste ist noch redundant: (1,2,7) implizieren alle anderen. Übung! Wir extrahieren daher nur diese grundlegenden Eigenschaften.

Das führt uns zu folgender Definition; sie ist genial-einfach und flexibel.

Definition V1c: Wahrscheinlichkeitsraum, Kolmogorov 1933

Ein **Wahrscheinlichkeitsraum**  $(\Omega, \mathcal{A}, \mathbf{P})$  besteht aus

- einer **Ergebnismenge**  $\Omega$ ,
- einer **Ereignismenge**  $\mathcal{A} \subseteq \mathfrak{P}(\Omega)$  und
- einem **Wahrscheinlichkeitsmaß**  $\mathbf{P} : \mathcal{A} \rightarrow [0, 1]$ ,

mit folgenden grundlegenden Eigenschaften, wie oben erklärt:

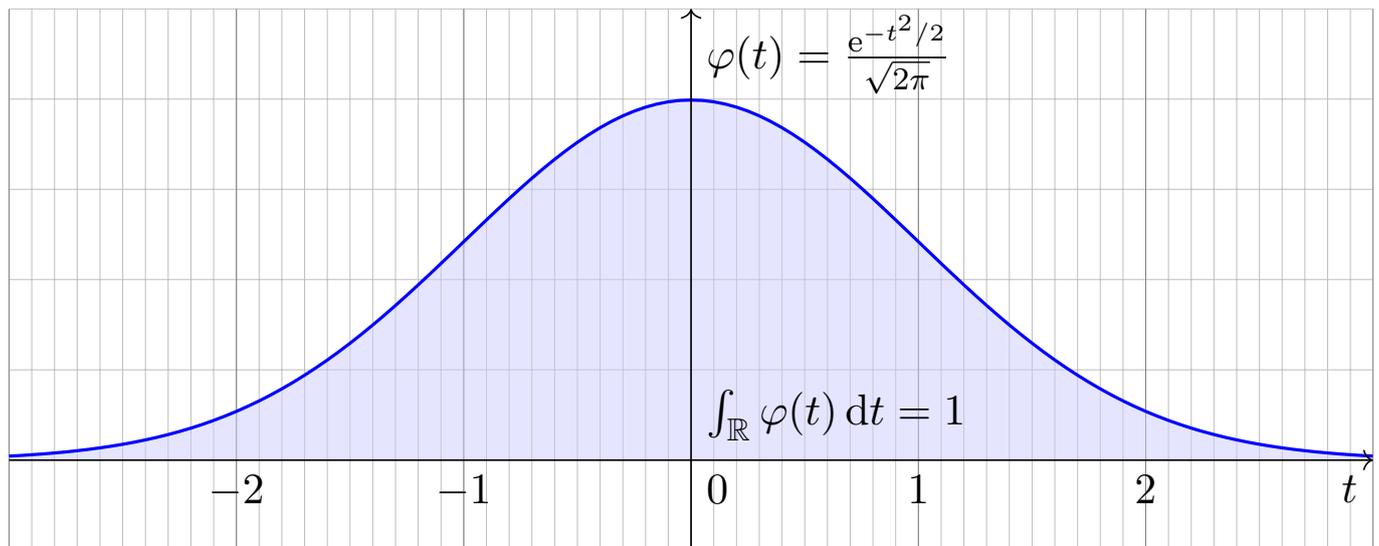
1 **Normierung:** Es gilt  $\Omega \in \mathcal{A}$  und  $\mathbf{P}(\Omega) = 1$ .

2 **Komplemente:** Aus  $A \in \mathcal{A}$  folgt  $(\Omega \setminus A) \in \mathcal{A}$ .

3  **$\sigma$ -Additivität:** Aus  $A_0, A_1, A_2, \dots \in \mathcal{A}$  folgt  $\bigcup_{k=0}^{\infty} A_k \in \mathcal{A}$ , sowie

$$\mathbf{P}\left(\bigcup_{k=0}^{\infty} A_k\right) = \sum_{k=0}^{\infty} \mathbf{P}(A_k) \quad \text{falls } A_i \cap A_j = \emptyset \text{ für } i \neq j.$$

☺ Ein WRaum ist ein Maßraum  $(\Omega, \mathcal{A}, \mathbf{P})$  im Sinne der Definition A1B mit der Normierung  $\mathbf{P}(\Omega) = 1$ . Diese Definition extrahiert und präzisiert die grundlegenden Rechenregeln, auf denen alle weiteren aufbauen.

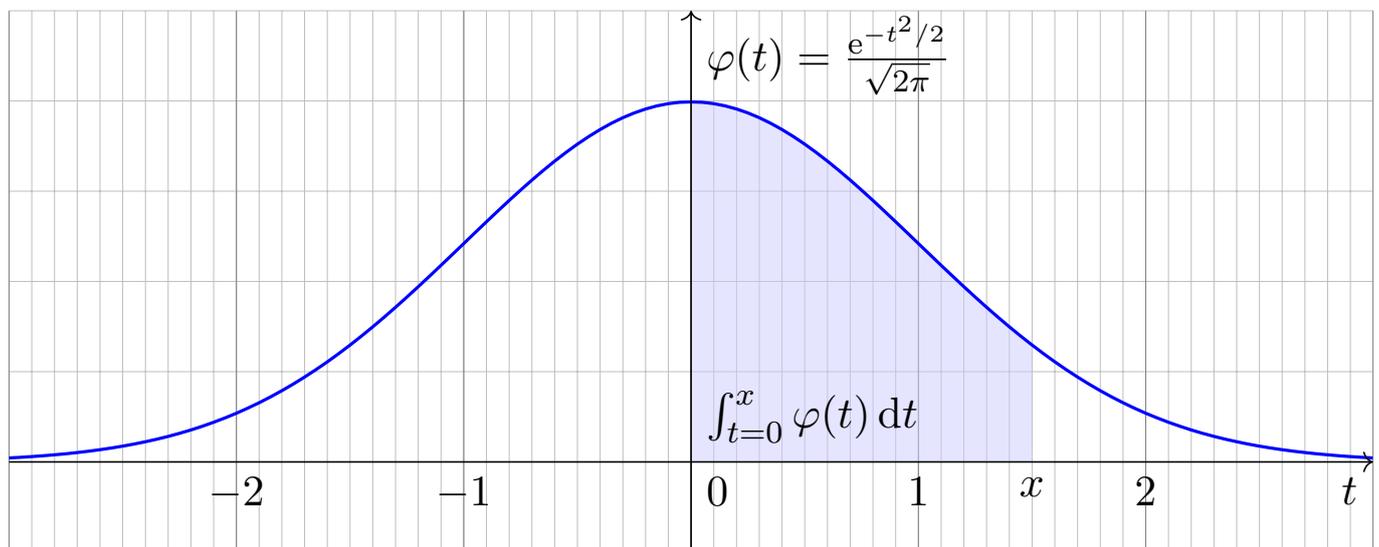


Die **Dichte der Standard-Normalverteilung** ist

$$\varphi : \mathbb{R} \rightarrow \mathbb{R} : t \mapsto \frac{1}{\sqrt{2\pi}} e^{-t^2/2} = \frac{1}{\sqrt{2\pi}} \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k}}{2^k k!}.$$

Dies ist eine WDichte, denn  $\varphi \geq 0$  und  $\int_{\mathbb{R}} \varphi(t) dt = 1$  dank Satz C2G, mit Erwartung  $\int_{\mathbb{R}} t \varphi(t) dt = 0$  [V239] und Varianz  $\int_{\mathbb{R}} t^2 \varphi(t) dt = 1$  [V240].

## Das Gaußsche Fehlerintegral



☹ Es gibt keine elementare Formel [B145] für die Integralfunktion

$$\int_{t=0}^x \varphi(t) dt = \frac{1}{\sqrt{2\pi}} \int_{t=0}^x e^{-t^2/2} dt = \frac{1}{\sqrt{2\pi}} \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k+1}}{2^k k! (2k+1)}.$$

😊 Die Potenzreihe hingegen ist wunderbar einfach und effizient! Die numerische Auswertung liefert die folgende hilfreiche Tabelle.

# Das Gaußsche Fehlerintegral: effiziente Näherung

**Aufgabe:** Approximieren Sie  $\varphi$  für  $x \in [-4, 4]$  auf 7 Nachkommastellen.

**Lösung:** Wir entwickeln die Funktion  $\varphi: \mathbb{R} \rightarrow \mathbb{R}$  in ihre Potenzreihe und integrieren termweise! Für  $k \geq 8$  ist dies eine Leibniz-Reihe (B147).

Zur Näherung durch eine endliche Summe schätzen wir den Fehler ab. Eine geeignete Approximation ist dann folgendes Polynom vom Grad 61:

$$\int_{t=0}^x \varphi(t) dt = \frac{1}{\sqrt{2\pi}} \sum_{k=0}^{30} \frac{(-1)^k x^{2k+1}}{2^k k! (2k+1)} + \delta \quad \left\{ \begin{array}{l} \text{mit } |\delta| < 0.4 \cdot 10^{-7} \\ \text{für alle } x \in [-4, 4] \end{array} \right.$$

Diese Näherung kann ein Computer leicht auswerten und tabellieren. In Python gelingt dies elegant wie folgt und liefert die ersehnte Tabelle.

```

1 def normal_distribution(x):
2     sum = term = x / sqrt(2*pi)      # Summe und Term für k=0
3     factor = (-1) * x*x / 2         # nützlicher gemeinsamer Faktor
4     for k in range(1,31):          # Grenze 30 aus dem Beweis
5         term *= factor / k         # Fortführung des Terms
6         sum += term / (2*k+1)     # Summe bis k
7     return sum

```

## Standard-Normalverteilung: Tabelle für $\int_0^x \varphi(t) dt$

	$x+0.00$	$x+0.01$	$x+0.02$	$x+0.03$	$x+0.04$	$x+0.05$	$x+0.06$	$x+0.07$	$x+0.08$	$x+0.09$
$x=0.0$	0.00000	0.00399	0.00798	0.01197	0.01595	0.01994	0.02392	0.02790	0.03188	0.03586
0.1	0.03983	0.04380	0.04776	0.05172	0.05567	0.05962	0.06356	0.06749	0.07142	0.07535
0.2	0.07926	0.08317	0.08706	0.09095	0.09483	0.09871	0.10257	0.10642	0.11026	0.11409
0.3	0.11791	0.12172	0.12552	0.12930	0.13307	0.13683	0.14058	0.14431	0.14803	0.15173
0.4	0.15542	0.15910	0.16276	0.16640	0.17003	0.17364	0.17724	0.18082	0.18439	0.18793
0.5	0.19146	0.19497	0.19847	0.20194	0.20540	0.20884	0.21226	0.21566	0.21904	0.22240
0.6	0.22575	0.22907	0.23237	0.23565	0.23891	0.24215	0.24537	0.24857	0.25175	0.25490
0.7	0.25804	0.26115	0.26424	0.26730	0.27035	0.27337	0.27637	0.27935	0.28230	0.28524
0.8	0.28814	0.29103	0.29389	0.29673	0.29955	0.30234	0.30511	0.30785	0.31057	0.31327
0.9	0.31594	0.31859	0.32121	0.32381	0.32639	0.32894	0.33147	0.33398	0.33646	0.33891
1.0	0.34134	0.34375	0.34614	0.34849	0.35083	0.35314	0.35543	0.35769	0.35993	0.36214
1.1	0.36433	0.36650	0.36864	0.37076	0.37286	0.37493	0.37698	0.37900	0.38100	0.38298
1.2	0.38493	0.38686	0.38877	0.39065	0.39251	0.39435	0.39617	0.39796	0.39973	0.40147
1.3	0.40320	0.40490	0.40658	0.40824	0.40988	0.41149	0.41309	0.41466	0.41621	0.41774
1.4	0.41924	0.42073	0.42220	0.42364	0.42507	0.42647	0.42785	0.42922	0.43056	0.43189
1.5	0.43319	0.43448	0.43574	0.43699	0.43822	0.43943	0.44062	0.44179	0.44295	0.44408
1.6	0.44520	0.44630	0.44738	0.44845	0.44950	0.45053	0.45154	0.45254	0.45352	0.45449
1.7	0.45543	0.45637	0.45728	0.45818	0.45907	0.45994	0.46080	0.46164	0.46246	0.46327
1.8	0.46407	0.46485	0.46562	0.46638	0.46712	0.46784	0.46856	0.46926	0.46995	0.47062
1.9	0.47128	0.47193	0.47257	0.47320	0.47381	0.47441	0.47500	0.47558	0.47615	0.47670
2.0	0.47725	0.47778	0.47831	0.47882	0.47932	0.47982	0.48030	0.48077	0.48124	0.48169
2.1	0.48214	0.48257	0.48300	0.48341	0.48382	0.48422	0.48461	0.48500	0.48537	0.48574
2.2	0.48610	0.48645	0.48679	0.48713	0.48745	0.48778	0.48809	0.48840	0.48870	0.48899
2.3	0.48928	0.48956	0.48983	0.49010	0.49036	0.49061	0.49086	0.49111	0.49134	0.49158
2.4	0.49180	0.49202	0.49224	0.49245	0.49266	0.49286	0.49305	0.49324	0.49343	0.49361
2.5	0.49379	0.49396	0.49413	0.49430	0.49446	0.49461	0.49477	0.49492	0.49506	0.49520
2.6	0.49534	0.49547	0.49560	0.49573	0.49585	0.49598	0.49609	0.49621	0.49632	0.49643
2.7	0.49653	0.49664	0.49674	0.49683	0.49693	0.49702	0.49711	0.49720	0.49728	0.49736
2.8	0.49744	0.49752	0.49760	0.49767	0.49774	0.49781	0.49788	0.49795	0.49801	0.49807
2.9	0.49813	0.49819	0.49825	0.49831	0.49836	0.49841	0.49846	0.49851	0.49856	0.49861
3.0	0.49865	0.49869	0.49874	0.49878	0.49882	0.49886	0.49889	0.49893	0.49896	0.49900
3.1	0.49903	0.49906	0.49910	0.49913	0.49916	0.49918	0.49921	0.49924	0.49926	0.49929
3.2	0.49931	0.49934	0.49936	0.49938	0.49940	0.49942	0.49944	0.49946	0.49948	0.49950

Diese Tabelle werden wir häufig nutzen! Ablesebeispiele:

$$\int_0^1 \varphi(t) dt = 0.34134$$

$$\int_0^{1.5} \varphi(t) dt = 0.43319$$

$$\int_0^{1.53} \varphi(t) dt = 0.43699$$

Ablezen eines Integrals der Form  $\int_{\alpha}^{\beta} \varphi(t) dt$  mit  $\alpha < 0 < \beta$ :

$$\begin{aligned} \int_{-1.23}^{2.34} \varphi(t) dt &= \int_{-1.23}^0 \varphi(t) dt + \int_0^{2.34} \varphi(t) dt \\ &= \int_0^{1.23} \varphi(t) dt + \int_0^{2.34} \varphi(t) dt \\ &= 0.39065 + 0.49036 = 0.88101 \end{aligned}$$

Ablezen eines Integrals der Form  $\int_{\alpha}^{\beta} \varphi(t) dt$  mit  $0 < \alpha < \beta$ :

$$\begin{aligned} \int_1^2 \varphi(t) dt &= \int_0^2 \varphi(t) dt - \int_0^1 \varphi(t) dt \\ &= 0.47725 - 0.34134 = 0.13591 \end{aligned}$$

Ebenso nutzen Sie auch die **kumulative Verteilungsfunktion**

$$\Phi(x) = \int_{-\infty}^x \varphi(t) dt = \frac{1}{2} + \int_0^x \varphi(t) dt$$

Ablesebeispiele:

$$\Phi(0.00) = 0.50000$$

$$\Phi(1.00) = 0.84134$$

$$\Phi(1.50) = 0.93319$$

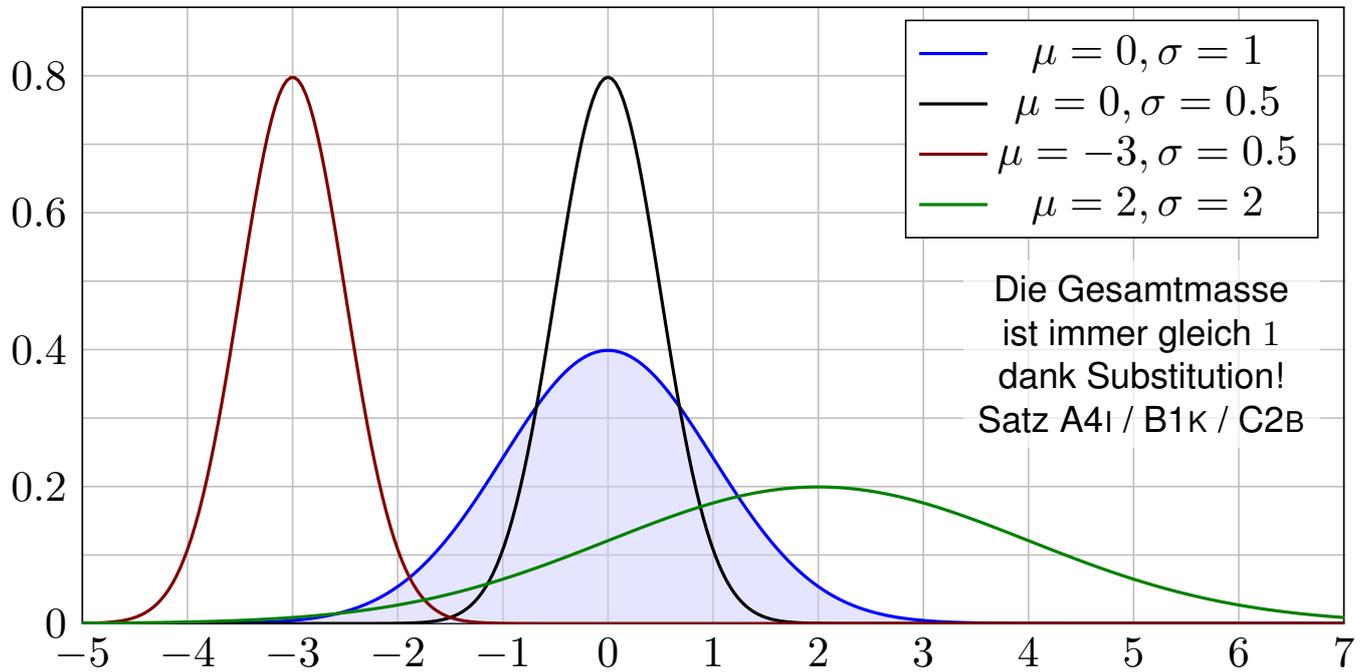
$$\Phi(1.53) = 0.93699$$

😊 Wahrscheinlichkeiten berechnen Sie leicht mit der nötigen Übung.

# Die Familie der Normalverteilungen $N(\mu, \sigma^2)$

Sei  $\mu, \sigma \in \mathbb{R}$  mit  $\sigma > 0$ . Die Dichte der **Normalverteilung**  $N(\mu, \sigma^2)$  ist

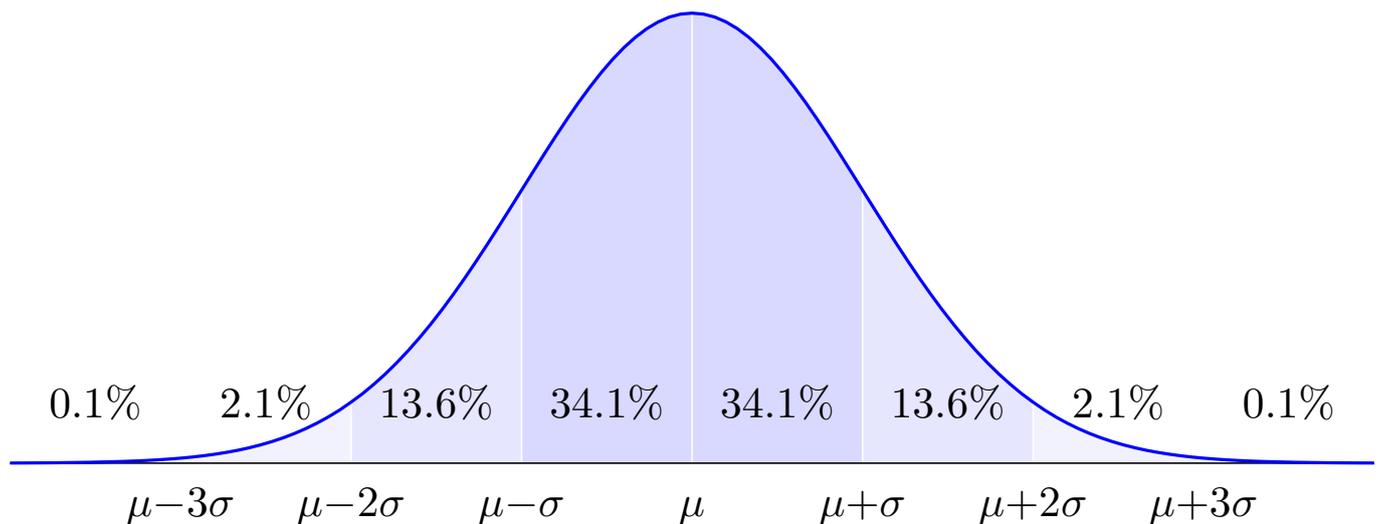
$$\varphi_{\mu, \sigma} : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto \frac{1}{\sigma} \varphi\left(\frac{x - \mu}{\sigma}\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$



Normalverteilungen: Kennst du eine, kennst du alle! Eine Tabelle genügt.

# Die 68–95–99–Regel

Die Normalverteilung  $N(\mu, \sigma^2)$  ist extrem konzentriert um  $\mu$ :



Jede Normalverteilung  $N(\mu, \sigma^2)$  ist symmetrisch um  $\mu$ . Gemäß unserer Tabelle gilt:

$$\int_{x=\mu}^{\mu+\sigma} \varphi_{\mu, \sigma}(x) dx \approx 0.34134, \quad \int_{x=\mu}^{\mu+2\sigma} \varphi_{\mu, \sigma}(x) dx \approx 0.47725, \quad \int_{x=\mu}^{\mu+3\sigma} \varphi_{\mu, \sigma}(x) dx \approx 0.49865$$

Das bedeutet ausformuliert: Die Masse liegt zu etwas über 68.2% in der  $1\sigma$ -Umgebung um  $\mu$ , zu über 95.4% in der  $2\sigma$ -Umgebung um  $\mu$ , und zu über 99.7% in der  $3\sigma$ -Umgebung um  $\mu$ . Chybechev T3F liefert universell nur die gröberen Ungleichungen  $\geq 0\%$ ,  $\geq 75\%$ ,  $\geq 88\%$ .

## Definition V2A: kumulative Verteilungsfunktion

Sei  $\mathbf{P} : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$  ein Wahrscheinlichkeitsmaß auf  $\mathbb{R}$ .

Hierzu definieren wir die **kumulative Verteilungsfunktion**

$$F : \mathbb{R} \rightarrow [0, 1] \quad \text{durch} \quad F(x) := \mathbf{P}([-\infty, x]).$$

Wichtige Spezialfälle: Für jede diskrete Verteilung  $p : \mathbb{R} \rightarrow [0, 1]$  gilt

$$F(x) = \sum_{t \leq x} p(t).$$

Für jede kontinuierliche Verteilung mit Dichte  $f : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$  gilt

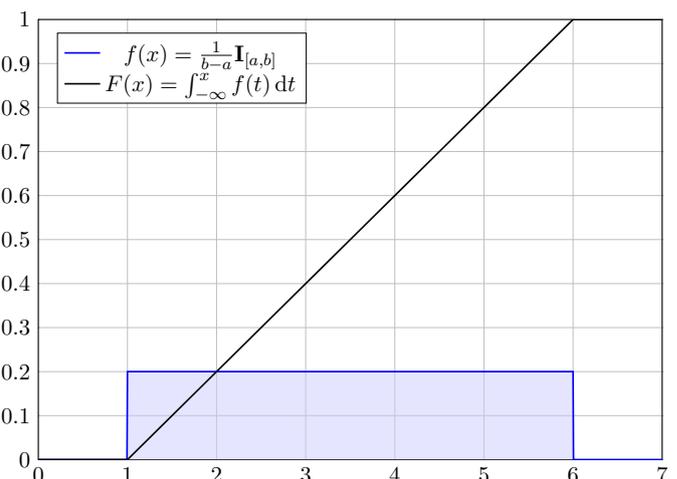
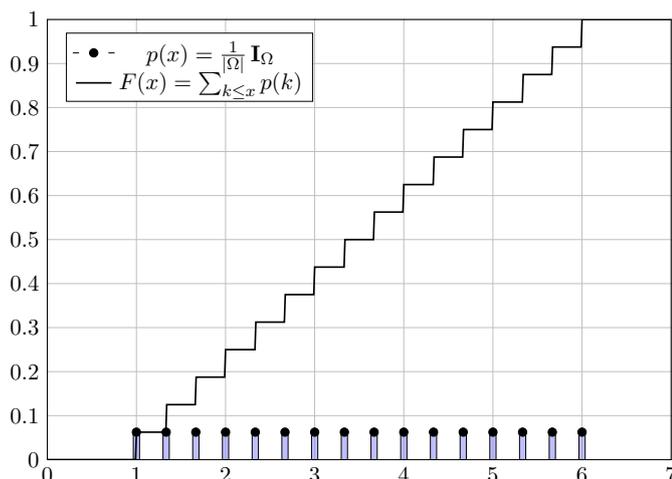
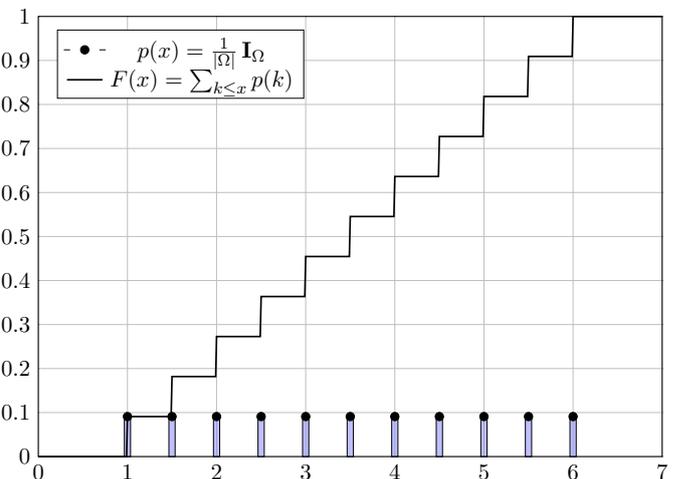
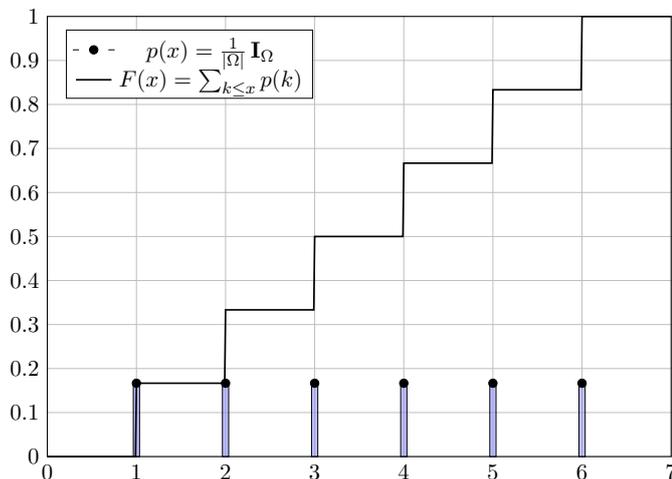
$$F(x) = \int_{-\infty}^x f(t) dt.$$

😊 Diese Beschreibung ist einheitlich, bequem und übersichtlich.

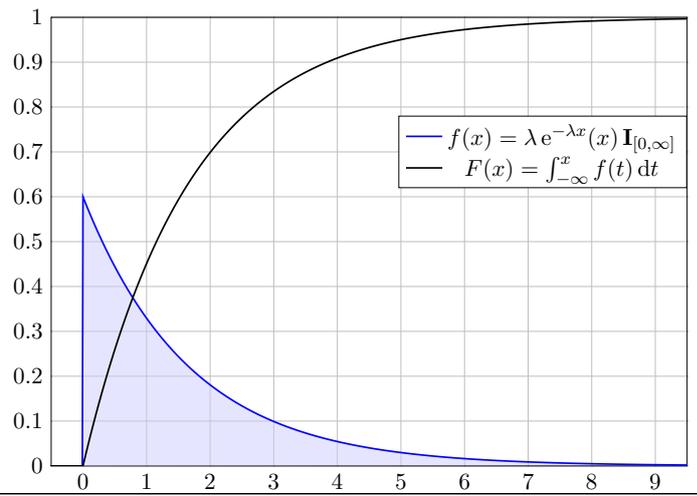
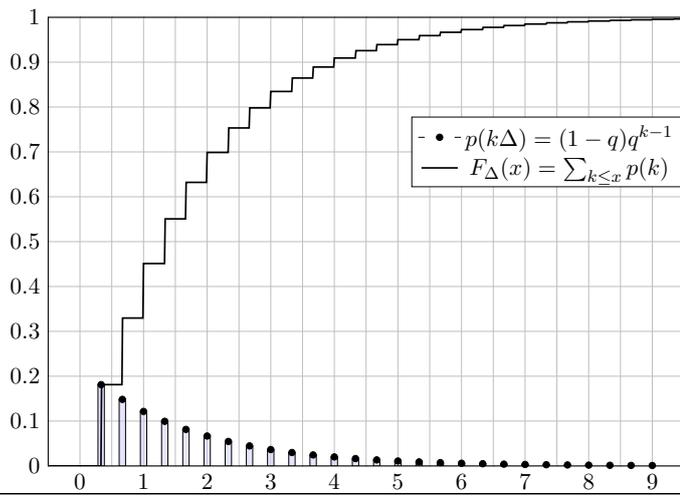
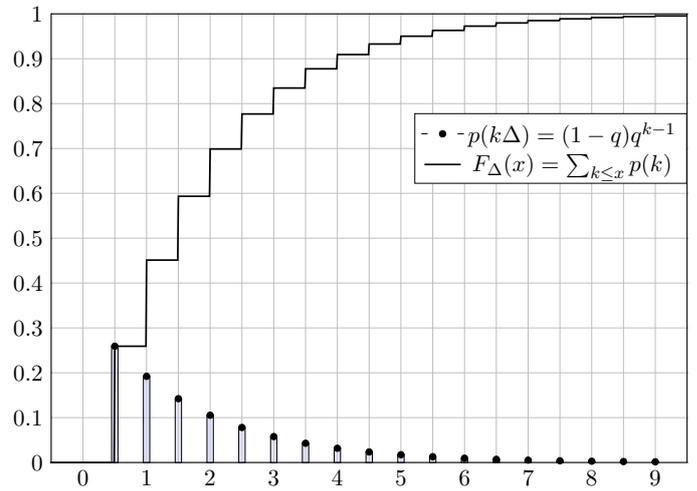
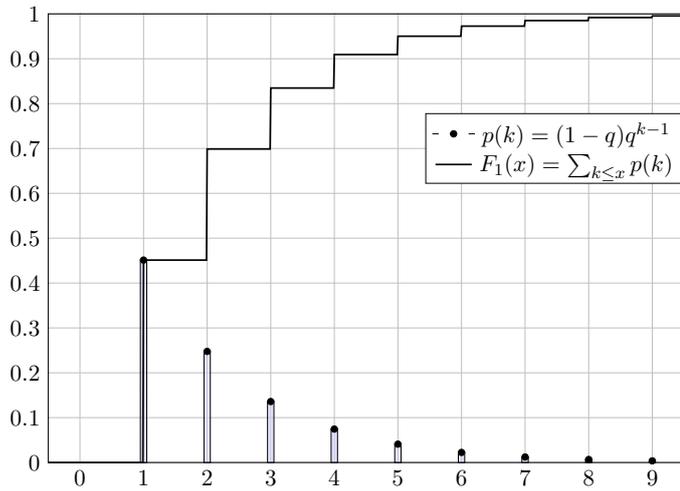
Sauber unterscheiden und benennen: Auf englisch heißt  $f$  *probability density function* (pdf) und  $F$  *cumulative distribution function* (cdf).

Aus  $F$  lässt sich  $\mathbf{P}$  rekonstruieren gemäß  $\mathbf{P}([a, b]) = F(b) - F(a)$  und  $\mathbf{P}(\{a\}) = F(a) - \lim_{x \nearrow a} F(x)$  im Falle einer Sprungstelle, siehe V2B.

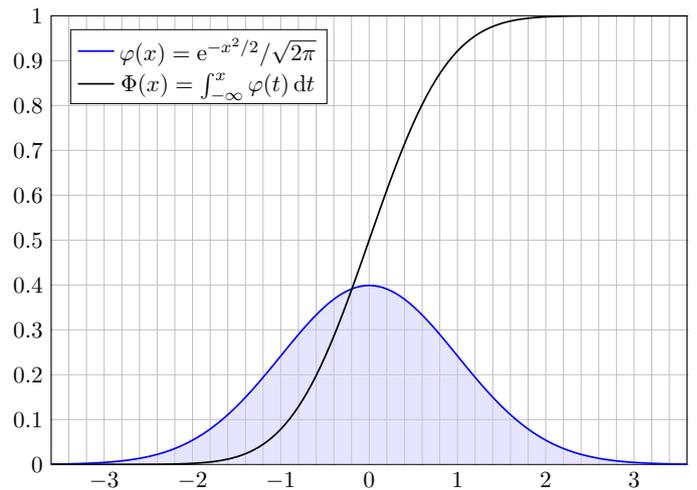
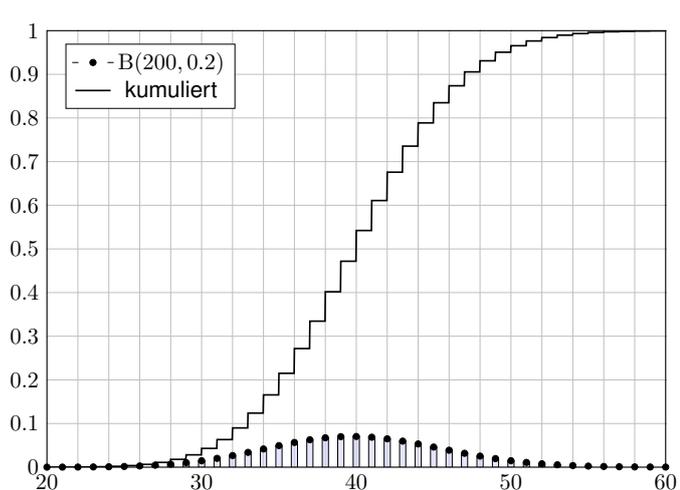
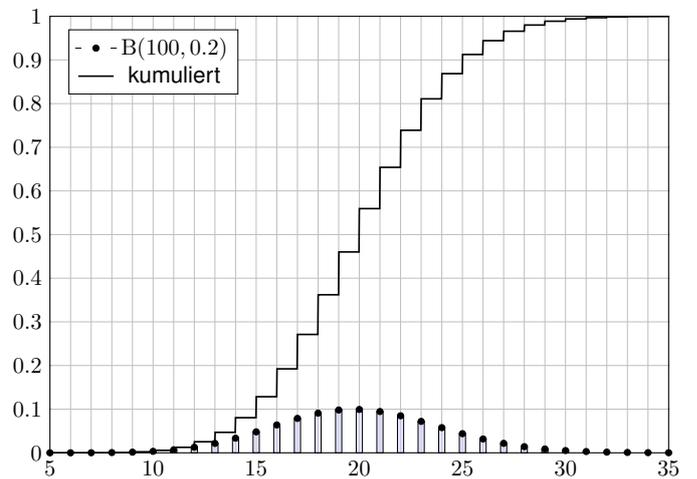
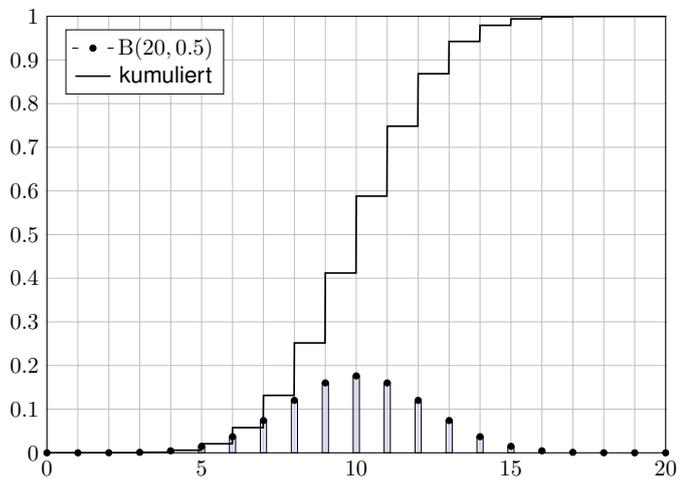
## Von diskreter zu kontinuierlicher Gleichverteilung



# Von geometrischer zu Exponentialverteilung



# Von Binomial- zu Normalverteilung



**Satz V2B: WMaße auf  $\mathbb{R}$  und ihre Verteilungsfunktionen**

Zu jedem WMaß  $\mathbf{P} : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$  haben wir seine Verteilungsfunktion

$$F : \mathbb{R} \rightarrow [0, 1] : x \mapsto F(x) := \mathbf{P}((-\infty, x]).$$

Sie erfreut sich folgender Eigenschaften:

- 1  $F$  ist monoton wachsend, für alle  $x \leq y$  gilt  $F(x) \leq F(y)$ .
- 2 Für  $x \searrow -\infty$  gilt  $F(x) \searrow 0$  und für  $x \nearrow +\infty$  gilt  $F(x) \nearrow 1$ .
- 3  $F$  ist rechtsseitig stetig, d.h. für  $x \searrow a$  gilt  $F(x) \searrow F(a)$ .

Umgekehrt definiert jede solche Funktion  $F : \mathbb{R} \rightarrow [0, 1]$  eindeutig ein Wahrscheinlichkeitsmaß  $\mathbf{P} : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$  mit  $\mathbf{P}((a, b]) = F(b) - F(a)$ .

In jedem Punkt  $a \in \mathbb{R}$  gilt  $\mathbf{P}(\{a\}) = F(a) - \lim_{x \nearrow a} F(x)$ .

Genau dann ist  $F$  stetig in  $a$ , wenn  $\mathbf{P}(\{a\}) = 0$  gilt.

Genau dann ist  $F$  differenzierbar mit stetiger Ableitung  $f = F'$ , wenn  $\mathbf{P}$  durch eine stetige Dichte  $f : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$  gegeben ist.

Sei  $(\Omega, \mathcal{A}, \mathbf{P})$  ein WRaum und hierin  $A_0, A_1, A_2, \dots, A \in \mathcal{A}$  Ereignisse.

Wir schreiben  $A_k \nearrow A$ , falls  $A_0 \subseteq A_1 \subseteq A_2 \subseteq \dots$  und  $A = \bigcup_{k=0}^{\infty} A_k$  gilt.

Wir schreiben  $A_k \searrow A$ , falls  $A_0 \supseteq A_1 \supseteq A_2 \supseteq \dots$  und  $A = \bigcap_{k=0}^{\infty} A_k$  gilt.

**Lemma V2c: Ausschöpfung und Eingrenzung**

Jedes Wahrscheinlichkeitsmaß  $\mathbf{P}$  ist in folgendem Sinne stetig:

Aus  $A_k \nearrow A$  folgt  $\mathbf{P}(A_k) \nearrow \mathbf{P}(A)$ ,

Aus  $A_k \searrow A$  folgt  $\mathbf{P}(A_k) \searrow \mathbf{P}(A)$ .

**Beweis:** Wir setzen  $B_0 := A_0$  und  $B_k = A_k \setminus A_{k-1}$  für  $k = 1, 2, 3, \dots$

Es gilt  $A_n = \bigsqcup_{k=0}^n B_k$  und  $A = \bigsqcup_{k=0}^{\infty} B_k$ . Dank abzählbarer Additivität

$$\mathbf{P}(A) = \mathbf{P}\left(\bigsqcup_{k=0}^{\infty} B_k\right) = \sum_{k=0}^{\infty} \mathbf{P}(B_k) = \lim_{n \rightarrow \infty} \sum_{k=0}^n \mathbf{P}(B_k) = \lim_{n \rightarrow \infty} \mathbf{P}(A_n).$$

Die zweite Aussage folgt aus der ersten durch Komplementbildung.

Für  $x < y$  gilt  $] -\infty, x] \subseteq ] -\infty, y]$  und daher

$$F(x) = \mathbf{P}(] -\infty, x]) \leq \mathbf{P}(] -\infty, y]) = F(y).$$

Für  $x \rightarrow \pm\infty$  gilt:

$$\lim_{n \rightarrow -\infty} F(n) = \lim_{n \rightarrow -\infty} \mathbf{P}(] -\infty, n]) = \mathbf{P}\left(\bigcap_{n=1}^{\infty} ] -\infty, n]\right) = \mathbf{P}(\emptyset) = 0$$

$$\lim_{n \rightarrow +\infty} F(n) = \lim_{n \rightarrow +\infty} \mathbf{P}(] -\infty, n]) = \mathbf{P}\left(\bigcup_{n=1}^{\infty} ] -\infty, n]\right) = \mathbf{P}(\mathbb{R}) = 1$$

Zur rechtsseitigen Stetigkeit:

$$\begin{aligned} \lim_{n \rightarrow \infty} F\left(x + \frac{1}{n}\right) &= \lim_{n \rightarrow \infty} \mathbf{P}\left(] -\infty, x + \frac{1}{n}]\right) \\ &= \mathbf{P}\left(\bigcap_{n=1}^{\infty} ] -\infty, x + \frac{1}{n}]\right) = \mathbf{P}(] -\infty, x]) = F(x) \end{aligned}$$

Umkehrung: Ist  $F$  gegeben, so setzen wir  $\mathbf{P}(]a, b]) = F(b) - F(a)$ .

Die Familie der messbaren Mengen wird von diesen Intervallen erzeugt.

Daraus folgt die Eindeutigkeit und (mit mehr Mühe) die Existenz von  $\mathbf{P}$ .

Zur linksseitigen Stetigkeit:

$$\begin{aligned} \lim_{n \rightarrow \infty} F\left(x - \frac{1}{n}\right) &= \lim_{n \rightarrow \infty} \mathbf{P}\left(] -\infty, x - \frac{1}{n}]\right) \\ &= \mathbf{P}\left(\bigcup_{n=1}^{\infty} ] -\infty, x - \frac{1}{n}]\right) = \mathbf{P}(] -\infty, x[) \end{aligned}$$

Dies ist genau dann gleich  $F(x) = \mathbf{P}(] -\infty, x])$ , wenn  $\mathbf{P}(\{x\}) = 0$  gilt.

Allgemein gilt in jedem Punkt  $\mathbf{P}(\{a\}) = F(a) - \lim_{x \nearrow a} F(x)$ .

Hat  $\mathbf{P}$  eine stetige Dichte  $f$ , so gilt  $F' = f$  nach dem HDI, denn

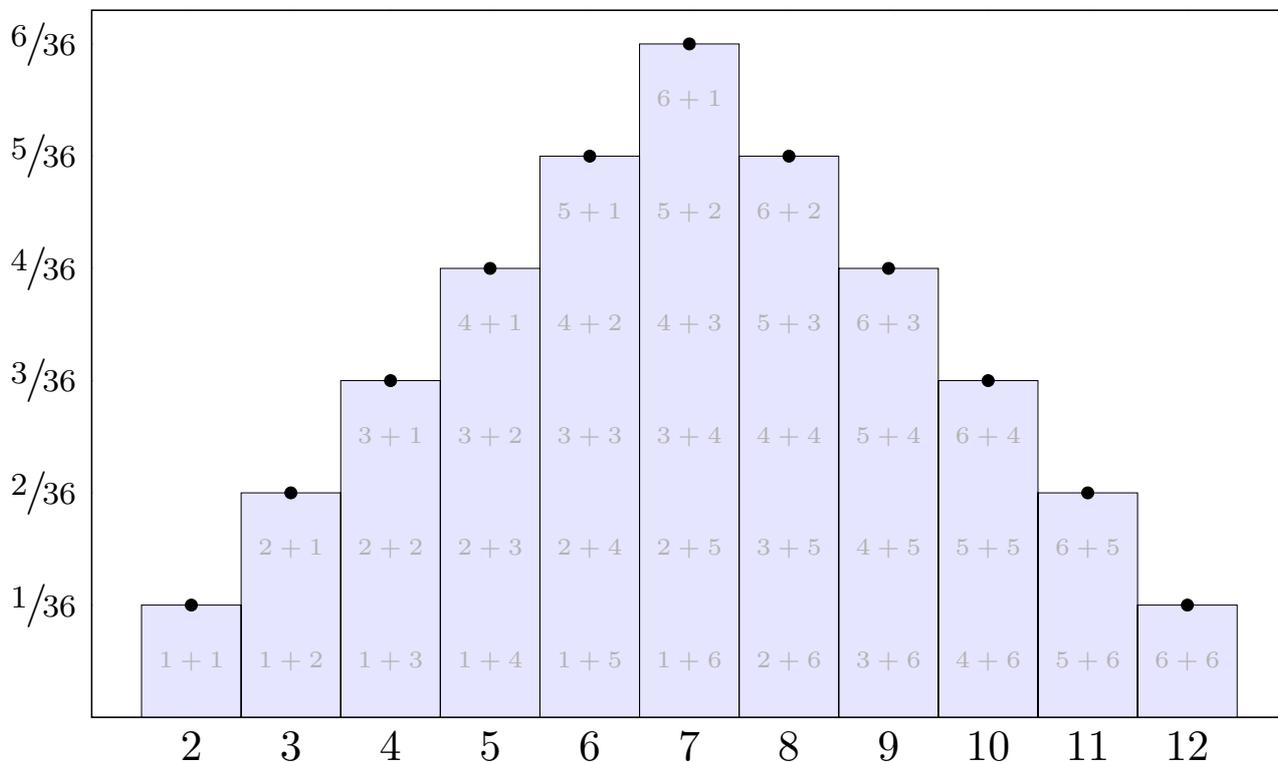
$$F(x) = \mathbf{P}(] -\infty, x]) = \int_{-\infty}^x f(t) dt.$$

Ist umgekehrt  $F$  stetig differenzierbar mit Ableitung  $f = F'$ , so gilt diese Gleichung. Für jede messbare Menge  $A \subseteq \mathbb{R}$  gilt dann

$$\mathbf{P}(A) = \int_A f(t) dt.$$

Somit ist das WMaß  $\mathbf{P}$  kontinuierlich mit stetiger Dichte  $f$ .

Ergebnisse  $\omega \in \Omega = \{1, 2, 3, 4, 5, 6\}^2$  mit Gleichverteilung  $\mathbf{P}(\{\omega\}) = \frac{1}{36}$ .  
 Augenzahlen  $X_1, X_2: \Omega \rightarrow \mathbb{R}$  mit  $X_1(\omega_1, \omega_2) = \omega_1$  und  $X_2(\omega_1, \omega_2) = \omega_2$ .  
 Augensumme  $X: \Omega \rightarrow \mathbb{R}$  mit  $X = X_1 + X_2$ . Die Verteilung von  $X$  ist:



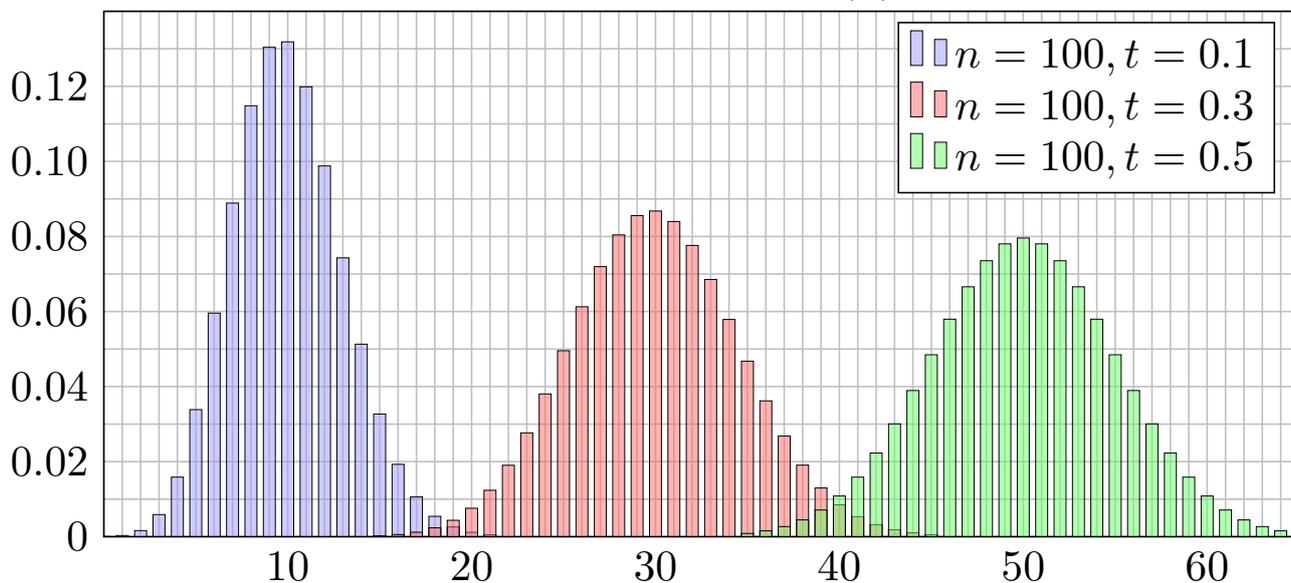
Ergebnisse sind  $n$ -Tupel  $\omega \in \Omega = \{0, 1\}^n$ . Die WVerteilung ist hier

$$\mathbf{P}(\{\omega\}) = t^k(1-t)^{n-k} \quad \text{bei Trefferzahl } k = \omega_1 + \dots + \omega_n.$$

Für  $X_i: \Omega \rightarrow \{0, 1\}$  mit  $X_i(\omega_1, \omega_2, \dots, \omega_n) = \omega_i$  gilt  $X_i \sim B(1, t)$ .

Für  $X = X_1 + \dots + X_n$  gilt  $X \sim B(n, t)$ , d.h.  $X$  hat die Verteilung

$$\mathbf{P}(X=k) = \mathbf{P}(\{\omega \mid X(\omega) = k\}) = \binom{n}{k} t^k (1-t)^{n-k}.$$



Oft interessiert bei Experimenten nicht das detaillierte Ergebnis  $\omega \in \Omega$ , sondern nur eine reelle Größe  $X(\omega) \in \mathbb{R}$ , die vom Ergebnis  $\omega$  abhängt.

### Definition V2D: Zufallsvariable

Wir betrachten einen Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, \mathbf{P})$ . V112

Eine reelle **Zufallsvariable** ist eine messbare Funktion  $X : \Omega \rightarrow \mathbb{R}$ .

Diese definiert auf  $\mathbb{R}$  eine WVerteilung  $\mathbf{P}_X : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$  durch

$$\mathbf{P}_X(B) = \mathbf{P}(X \in B) := \mathbf{P}(X^{-1}(B)) = \mathbf{P}(\{\omega \in \Omega \mid X(\omega) \in B\}).$$

Diese nennen wir die **Verteilung** der Zufallsvariablen  $X$  (Bildmaß).

Wir sagen hierzu  **$X$  ist  $\mathbf{P}_X$ -verteilt** und schreiben kurz  $X \sim \mathbf{P}_X$ .

Wir stellen uns  $\Omega$  als Lostopf vor; zu jedem Los  $\omega \in \Omega$  ist  $X(\omega) \in \mathbb{R}$  der Gewinn. Die Wkt  $\mathbf{P}$  ist auf die verschiedenen Ergebnisse  $\omega \in \Omega$  verteilt, entsprechend  $\mathbf{P}_X$  auf die möglichen Gewinne. Messbarkeit von  $X : \Omega \rightarrow \mathbb{R}$  ist eine technisch notwendige Bedingung. Sie bedeutet: Für jedes Intervall  $B \subseteq \mathbb{R}$  ist die Urbildmenge  $A = X^{-1}(B) = \{\omega \in \Omega \mid X(\omega) \in B\} = \{X \in B\}_1$  in  $\Omega$  messbar, das heißt  $A \in \mathcal{A}$ . Diese Vorsichtsmaßnahme ist mathematisch-technischer Natur: Sie stellt sicher, dass die Definition von  $\mathbf{P}_X$  und alle weiteren Rechnungen möglich sind.

### Zufallsvariablen: Anschauung und Bedeutung

Wir betrachten statt reeller auch vektorwertige Zufallsvariablen

$$X = (X_1, \dots, X_n) : (\Omega, \mathcal{A}, \mathbf{P}) \rightarrow (\mathbb{R}^n, \mathcal{B}, \mathbf{P}_X).$$

Allgemeiner: Jede Zufallsvariable ist eine messbare Abbildung

$$X : (\Omega, \mathcal{A}, \mathbf{P}) \rightarrow (\Omega', \mathcal{B}, \mathbf{P}_X)$$

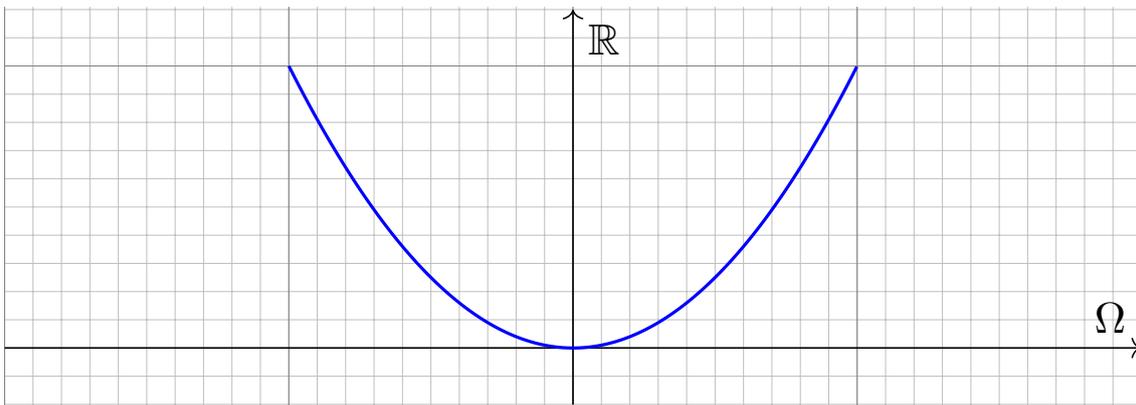
Anschaulich wird die Massenverteilung von  $\mathbf{P}$  auf  $\mathbf{P}_X$  transportiert.

Statt mit einer Zufallsvariable  $X : \Omega \rightarrow \Omega'$  auf dem WRaum  $(\Omega, \mathcal{A}, \mathbf{P})$  können wir ebenso auch direkt mit dem Bildraum  $(\Omega', \mathcal{B}, \mathbf{P}_X)$  arbeiten.

Mit der Zufallsvariable  $X$  können wir so einen *komplizierten* durch einen *einfachen* Wahrscheinlichkeitsraum ersetzen.

Die Einführung von Zufallsvariablen ist oft bequem und nützlich:

- Bequeme Notation und natürliche Sprechweise: ZVariablen sind z.B. Messungen, Stichproben, etc. Das ist für Anwendungen hilfreich.
- Wir können mehrere Experimente als Zufallsvariablen  $X_1, X_2, X_3, \dots$  auf dem gemeinsamen WRaum  $(\Omega, \mathcal{A}, \mathbf{P})$  betrachten. Das hilft.
- Wir können vom zugrundeliegenden WRaum  $(\Omega, \mathcal{A}, \mathbf{P})$  abstrahieren und müssen ihn häufig gar nicht explizit ausführen. Das vereinfacht.



**Aufgabe:** (1) Auf  $\Omega = [-1, 1]$  betrachten wir die Gleichverteilung  $\mathbf{P}$ . Zeichnen Sie hierzu die WDichte  $g$  und die kumul. Verteilung  $G$ .

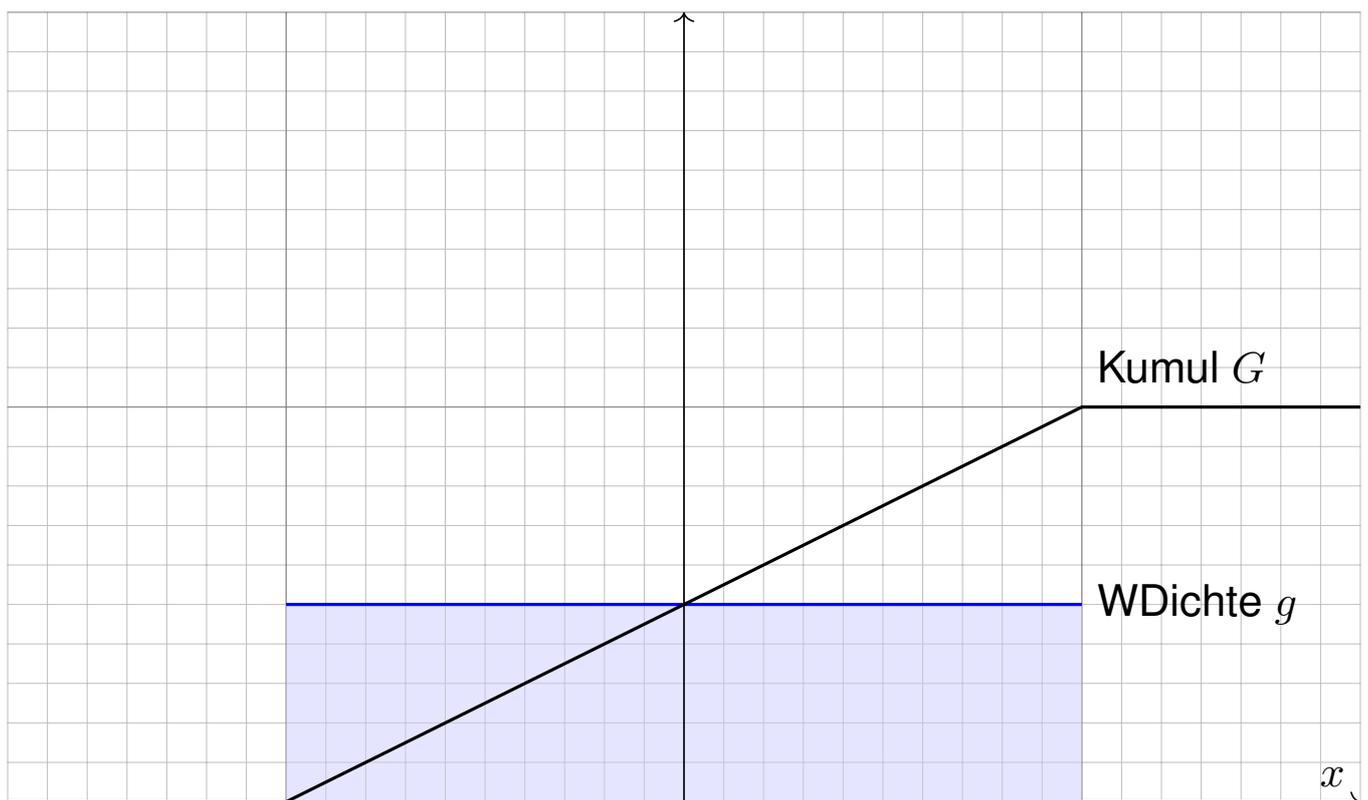
(2) Bestimmen Sie für die ZVariable  $X : \Omega \rightarrow \mathbb{R} : t \mapsto X(t) = t^2$  die kumulative Verteilungsfunktion  $F$  und ihre WDichte  $f$ .

**Lösung:** Die Bildmenge ist  $[0, 1]$ , aber die Wkt ist nicht gleichverteilt! Das Bildmaß  $\mathbf{P}_X$  hat für  $0 \leq x \leq 1$  die kumulative Verteilung

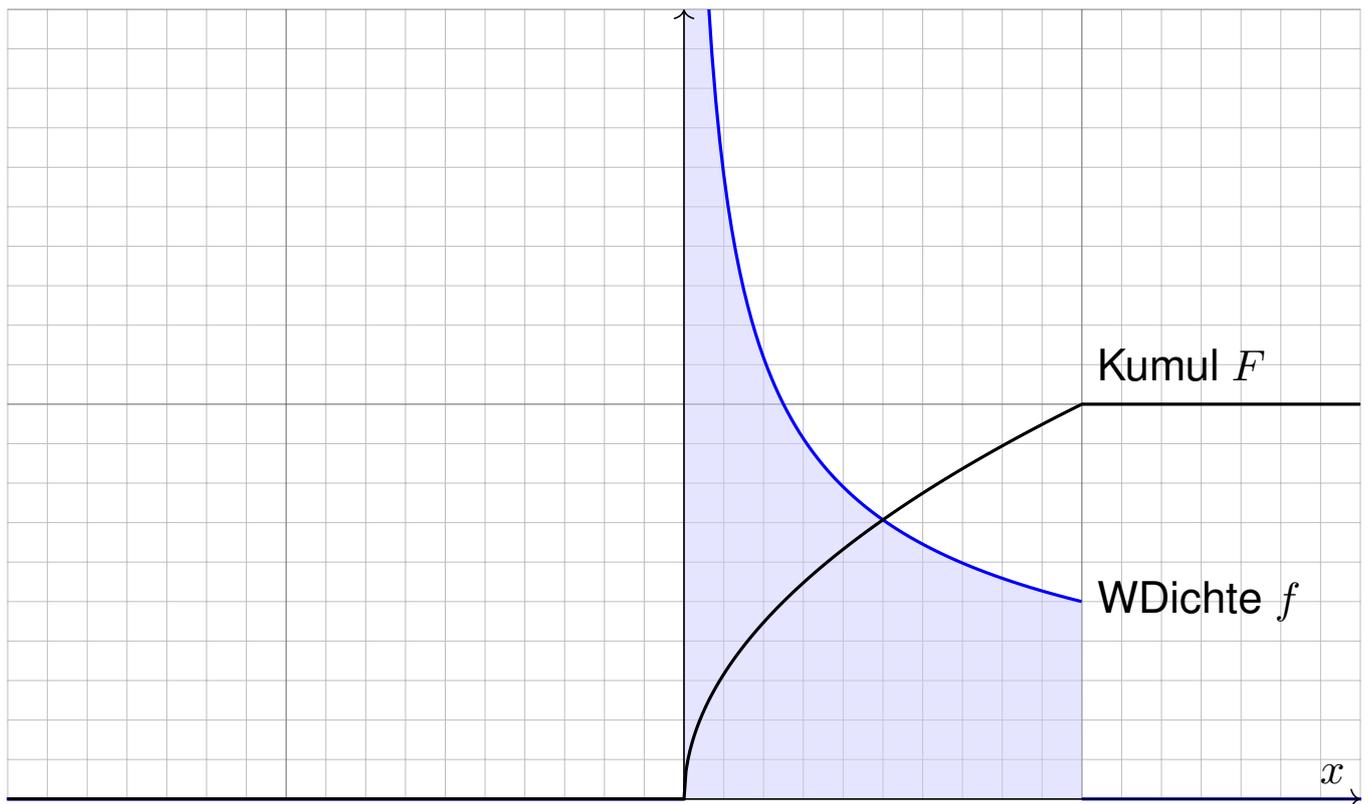
$$F(x) = \mathbf{P}(t^2 \leq x) = \mathbf{P}(|t| \leq \sqrt{x}) = \sqrt{x}.$$

Demnach hat  $\mathbf{P}_X$  die Dichte  $f(x) = F'(x) = \frac{1}{2\sqrt{x}}$ .

(1) Gleichverteilung  $\mathbf{P}$  auf dem Definitionsbereich  $\Omega = [-1, 1]$ :



(2) Verteilung  $\mathbf{P}_X$  der ZVaren  $X(t) = t^2$  im Bildbereich  $\Omega' = [0, 1]$ :



Die Massenverteilung  $(\Omega, \mathcal{A}, \mathbf{P})$  wird nach  $(\mathbb{R}, \mathcal{B}, \mathbf{P}_X)$  transportiert: Die Parameter  $t \in [-1, 1]$  sind gleichverteilt, nicht jedoch  $t^2 \in [0, 1]$ . Im Bildraum  $[0, 1]$  werden die niedrigen Werte um 0 wahrscheinlicher getroffen (hohe Dichte) als die hohen Werte um 1 (geringere Dichte).

Dies können Sie konkret simulieren: Generieren Sie gleichverteilte Zufallszahlen  $t \in [-1, 1]$  und zeichnen Sie ihre Bildpunkte  $X(t) = t^2$ . Es ergibt sich eine Punktwolke in  $[0, 1]$ , deren Dichte von 0 nach 1 monoton abnimmt. Nahe 0 wird diese Wolke tatsächlich extrem dicht.

Unser Ergebnis für die Zufallsvariable  $X(t) = t^2$  ist daher plausibel. Will man die genaue Verteilung, so muss man rechnen (wie gezeigt). Wir sehen insbesondere, dass WDichten nicht beschränkt sein müssen: Hier hat sie eine Polstelle, beim Nullpunkt 0 also unendlich große Dichte. Dennoch ist die Gesamtmasse weiterhin endlich, immer noch gleich 1.

Hinter diesen Beispielen steckt der Transformationssatz für Integrale: Ist die Zufallsvariable  $X : [a, b] \rightarrow \mathbb{R}$  stetig diff'bar, so hat das Bildmaß  $\mathbf{P}_X$  die kumulative Verteilung  $F(x) = \mathbf{P}(X^{-1}([a, x]))$  und WDichte  $f = F'$ .

Die Transformation von WMaßen können wir auch umgekehrt nutzen: Wir nehmen an, gleichverteilte unabhängige Zufallszahlen  $x \in [0, 1]$  werden uns kostengünstig zur Verfügung gestellt. Wie können Sie nun daraus Zufallszahlen erzeugen, die einer anderen Verteilung folgen, zum Beispiel einer Normalverteilung oder einer Exponentialverteilung?

Jedes WMaß auf  $\mathbb{R}$  können wir als Bildmaß aus der Gleichverteilung auf  $[0, 1]$  gewinnen! Diese Beobachtung können Sie zur Simulation verwenden. Das ist ein vielgenutztes Verfahren: Für manche Modelle ist es sinnvoll, stochastische Simulationen (am Computer) durchzuführen. Hierzu benötigen Sie einen Generator von gleichverteilten Zufallszahlen  $x \in [0, 1]$ . Das ist ein eigenes Kapitel der Informatik und birgt erstaunliche Schwierigkeiten und Überraschungen.

Wir betrachten dies als „black box“ und nehmen an, gleichverteilte unabhängige Zufallszahlen  $x \in [0, 1]$  werden uns kostengünstig zur Verfügung gestellt. Dann ist z.B.  $3x + 2$  gleichverteilt auf  $[2, 5]$ . Wie können Sie Zufallszahlen  $x \in \mathbb{R}$  erzeugen, die einer ganz anderen Verteilung folgen, zum Beispiel einer Normalverteilung auf  $\mathbb{R}$  oder einer Exponentialverteilung auf  $\mathbb{R}_{\geq 0}$ ?

Zur Erzeugung solcher Zufallszahlen nutzen wir, dass auch die kumulative Verteilungsfunktion einen Bildbereich zwischen Null und Eins aufweist! Wählen Sie in diesem Intervall zufällig eine Zahl  $y \in [0, 1]$ , so können Sie dies als Wert  $y = F(x)$  einer Verteilungsfunktion interpretieren. Der zugehörige Wert  $x$  ist dann die gewünschte Zufallszahl. Genau dies erklärt der folgende Satz.

### Satz V2E: Quantilfunktion

Vorgegeben sei eine kumulative Verteilungsfunktion  $F : \mathbb{R} \rightarrow [0, 1]$ .

Wir betrachten  $\Omega = [0, 1]$  mit Gleichverteilung  $\mathbf{P}$  und definieren

$$X : \Omega \rightarrow \bar{\mathbb{R}} \quad \text{durch} \quad X(t) = F^{-1}(t) = \inf \{ x \in \mathbb{R} \mid t \leq F(x) \}$$

Dann ist  $X$  messbar und  $\mathbf{P}_X$  hat die kumulative Verteilung  $F$ .

😊 Wenn  $F$  stetig und streng wachsend ist, so ist  $F : \mathbb{R} \rightarrow ]0, 1[$  bijektiv. Im allgemeinen Fall muss  $F$  nur schwach wachsend sein und darf auch Sprungstellen haben. Für diese Fälle ist die ausführlichere Formel als Infimum angegeben. Die Funktion  $X$  heißt **Quantilfunktion** von  $F$ .

**Nachrechnen:** Nach Konstruktion gilt  $X(t) \leq x \Leftrightarrow t \leq F(x)$ , also

$$\mathbf{P}_X(]-\infty, x]) = \mathbf{P}(X(t) \leq x) = \mathbf{P}(t \leq F(x)) = F(x)$$

😊 Anwendung: Simulation von WMaßen auf  $\mathbb{R}$ , z.B. am Computer. Aus der Gleichverteilung gewinnen Sie jede beliebige Verteilung.

**Aufgabe:** Was ist die Quantilfunktion der Exponentialverteilung  $E(\lambda)$ ? Wie können Sie diese Verteilung mit einer Gleichverteilung simulieren?

**Lösung:** Wir haben die Dichte  $f(x) = \lambda e^{-\lambda x}$  auf  $\mathbb{R}_{\geq 0}$ , kumuliert also

$$F(x) = \int_{-\infty}^x f(t) dt = 1 - e^{-\lambda x}.$$

(Siehe V203.) Die Quantilfunktion hierzu ist einfach die Umkehrfunktion

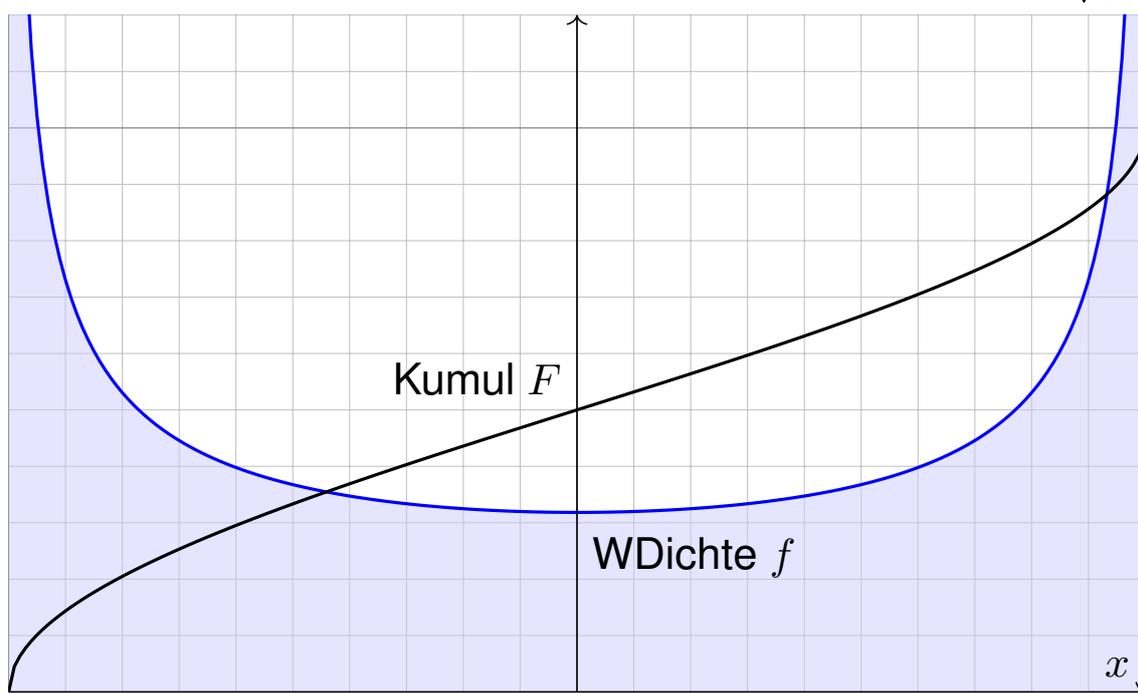
$$F^{-1}(y) = -\frac{\ln(1-y)}{\lambda}.$$

Die  $E(\lambda)$ -Verteilung simulieren wir demnach so: Wir ziehen zufällig Zahlen  $y \in [0, 1]$  und bilden diese ab auf  $x = F^{-1}(y) = -\ln(1-y)/\lambda$ . Auf  $\Omega = [0, 1]$  mit Gleichverteilung entspricht dies der Zufallsvariablen

$$X : \Omega \rightarrow \mathbb{R} : \omega \mapsto X(\omega) = -\frac{\ln(1-\omega)}{\lambda}.$$

Sie prüfen leicht nach, dass  $X$  die gewünschte Verteilungsfunktion hat: Abstrakt ist das der obige Beweis, hier ganz konkret und anschaulich.

**Aufgabe:** Wir betrachten das Intervall  $\Omega = [0, 1]$  mit Gleichverteilung  $\mathbf{P}$ . Finden Sie  $X : [0, 1] \rightarrow [-1, 1]$  sodass  $\mathbf{P}_X$  die Dichte  $f(x) = \frac{1}{\pi\sqrt{1-x^2}}$  hat.



**Lösung:** Kumul. Verteilung  $F(x) = \int_{-1}^x f(t) dt = \frac{1}{\pi} \arcsin(x) + \frac{1}{2}$ .

Wir wählen  $X : [0, 1] \rightarrow [-1, 1]$  gemäß  $X(t) = F^{-1}(t) = \sin(\pi t - \pi/2)$ .

## Definition V2F: einfache Zufallsvariablen und ihre Erwartung

Sei  $(\Omega, \mathcal{A}, \mathbf{P})$  ein WRaum. Eine Zufallsvariable  $X : \Omega \rightarrow \mathbb{R}$  heißt **einfach**, wenn sie nur endliche viele Werte  $a_1 < a_2 < \dots < a_n$  in  $\mathbb{R}$  annimmt.

Somit gilt  $X = \sum_{k=1}^n a_k \mathbf{I}_{A_k}$  mit  $A_1, \dots, A_n \in \mathcal{A}$  und  $\Omega = A_1 \sqcup \dots \sqcup A_n$ . In diesem Falle ist ihr Erwartungswert der gewichtete Mittelwert

$$\mathbf{E}(X) := \sum_{k=1}^n a_k \cdot \mathbf{P}(X=a_k) = \sum_{k=1}^n a_k \cdot \mathbf{P}(A_k)$$

**Aufgabe:** Berechnen Sie  $\mathbf{E}(X)$  für  $X \sim B(n, t)$  binomialverteilt.

**Lösung:** Die Berechnung des Erwartungswertes kennen wir [U421](#):

$$\mathbf{E}(X) = \sum_{k=0}^n k \cdot \mathbf{P}(X=k) = \sum_{k=0}^n k \cdot \binom{n}{k} t^k (1-t)^{n-k} = nt$$

Dies ist eine einfache Funktion, in diesem konkreten Beispiel gilt nämlich  $X = \sum_{k=0}^n k \mathbf{I}_{A_k}$  mit  $A_k = X^{-1}(\{k\}) = \{\omega \in \Omega \mid X(\omega) = k\}$ . Dies verallgemeinert Treppenfunktionen (A307): Hier nun sind  $A_1, \dots, A_n \in \mathcal{A}$  beliebige messbare Mengen in  $\Omega$ , nicht nur Quader im  $\mathbb{R}^n$ .

Zum WRaum  $(\Omega, \mathcal{A}, \mathbf{P})$  betrachten wir die Menge der Treppenfunktionen

$$T(\Omega, \mathcal{A}) = \left\{ \sum_{k=1}^n a_k \mathbf{I}_{A_k} \mid a_1, \dots, a_n \in \mathbb{R}, A_1, \dots, A_n \in \mathcal{A} \right\}$$

**Aufgabe:** (1) Die Menge  $T(\Omega, \mathcal{A}) \subseteq \text{Abb}(\Omega, \mathbb{R})$  ist ein  $\mathbb{R}$ -Vektorraum, erzeugt von den Indikatorfunktionen  $\mathbf{I}_A$  aller messbarer Mengen  $A \in \mathcal{A}$ .

(2) Hierauf ist die Erwartung  $\mathbf{E} : T(\Omega, \mathcal{A}) \rightarrow \mathbb{R}$  gegeben durch

$$\mathbf{E}\left(\sum_{k=1}^n a_k \mathbf{I}_{A_k}\right) = \sum_{k=1}^n a_k \mathbf{P}(A_k).$$

Dies ist die einzige lineare Abbildung mit der Normierung  $\mathbf{E}(\mathbf{I}_A) = \mathbf{P}(A)$ .

**Lösung:** Die Aussage (1) ist klar nach Konstruktion von  $T(\Omega, \mathcal{A})$ .

Aussage (2) folgt aus Definition V2F, falls die Mengen  $A_k \in \mathcal{A}$  disjunkt und die Werte  $a_k \in \mathbb{R}$  verschieden sind (wie in der vorigen Aufgabe). Jede Treppenfunktion können wir so darstellen (per Induktion über  $n$ ).

😊 Die Erwartung entspricht dem Integral! Wir nutzen unsere guten Erfahrungen und unsere solide Vorbereitung und führen die Konstruktion des Integrals aus Kapitel A hier erneut durch.

😊 Für die Erwartung gehen wir wie beim Integral vor (siehe Kapitel A): Für jede einfache Zufallsvariable (aka Treppenfunktion) können wir das Integral durch eine endliche Summe definieren. Anschließend gelangen wir per Grenzübergang von einfachen zu messbaren Funktionen:

### Lemma V2G: Ausschöpfung durch einfache Funktionen

Sei  $(\Omega, \mathcal{A}, \mathbf{P})$  ein Wahrscheinlichkeitsraum.

Für jede Funktion  $f : \Omega \rightarrow [0, \infty]$  sind äquivalent:

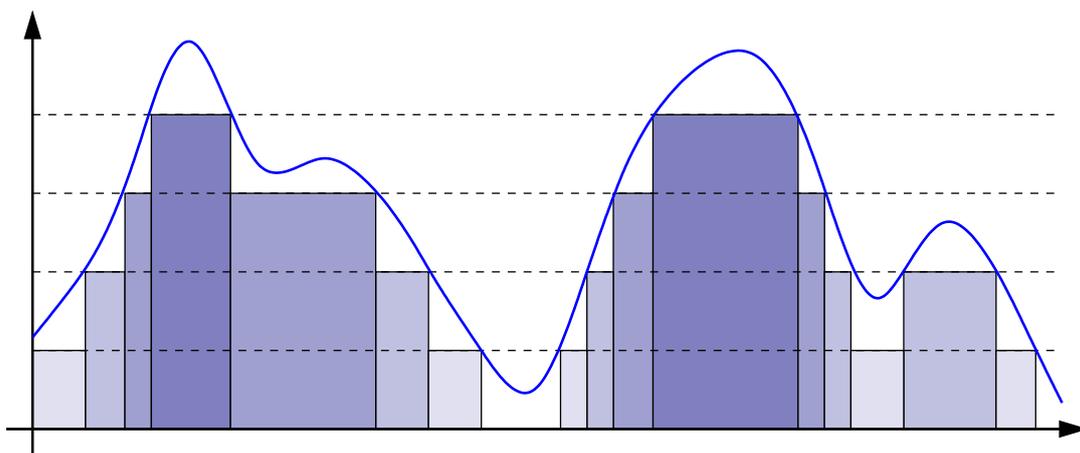
- (1) Für alle  $a \in \mathbb{R}$  ist  $f^{-1}([0, a])$  messbar:  $\{\omega \in \Omega \mid f(\omega) \leq a\} \in \mathcal{A}$ .
- (2) Es existiert eine wachsende Folge einfacher Funktionen

$$0 \leq f_0 \leq f_1 \leq f_2 \leq \dots : \Omega \rightarrow [0, \infty] \quad \text{mit} \quad f_k \nearrow f.$$

In diesem Fall gilt  $0 \leq \mathbf{E}(f_0) \leq \mathbf{E}(f_1) \leq \dots$  und wir definieren

$$\mathbf{E}(f) := \lim \mathbf{E}(f_k) \in [0, \infty].$$

Dies ist wohldefiniert, d.h. von der Wahl der Folge  $(f_k)_{k \in \mathbb{N}}$  unabhängig.



„(1)  $\Rightarrow$  (2)“: Wir konstruieren einfache Funktionen  $f_k : \Omega \rightarrow [0, \infty]$ .

Sei  $k \in \mathbb{N}$ . Für  $j \in \{0, \dots, N = k2^k\}$  setzen wir  $h_j := j/2^k$  und  $A_j := f^{-1}([h_j, h_{j+1}[)$ , wobei  $A_N := f^{-1}([h_N, \infty[)$ . Wir definieren  $f_k := \sum_{j=0}^N h_j \cdot \mathbf{I}_{A_j}$ . Für jeden Punkt  $x \in \Omega$  gilt dann  $f_k(x) \nearrow f(x)$ .

„(2)  $\Leftarrow$  (1)“: Seien  $f_k : \Omega \rightarrow [0, \infty]$  messbar und  $f_k \nearrow f$ .

Die Menge  $A_k = f_k^{-1}([0, a])$  ist messbar, also  $A_k \in \mathcal{A}$ .

Für  $A = f^{-1}([0, a])$  gilt dann  $A = \bigcap_{k=0}^{\infty} A_k$ , also  $A \in \mathcal{A}$ .

## Satz V2H: messbare und absolut integrierbare Funktionen

Genau dann ist  $f : \Omega \rightarrow \bar{\mathbb{R}}$  **messbar**, wenn  $f^+$  und  $f^-$  messbar sind.

(1) In diesem Fall ist auch  $|f| = f^+ + f^-$  messbar, und somit gilt

$$\mathbf{E}(|f|) = \mathbf{E}(f^+) + \mathbf{E}(f^-).$$

(2) Ist dieser Wert endlich, so nennen wir  $f$  **(absolut) integrierbar**.

In diesem Fall können wir den Erwartungswert von  $f$  definieren durch

$$\mathbf{E}(f) := \mathbf{E}(f^+) - \mathbf{E}(f^-).$$

(3) Der Erwartungswert verhält sich als Integral  $\mathbf{E}(f) = \int_{x \in \Omega} f(x) d\mathbf{P}$ :  
Die integrierbaren Funktionen  $f : \Omega \rightarrow \mathbb{R}$  bilden einen Vektorraum,

$$L^1(\Omega, \mathcal{A}, \mathbf{P}; \mathbb{R}) := \{ f : \Omega \rightarrow \mathbb{R} \mid \mathbf{E}(|f|) < \infty \}.$$

Hierauf ist die Erwartung  $\mathbf{E} : L^1(\Omega, \mathcal{A}, \mathbf{P}) \rightarrow \mathbb{R} : f \mapsto \mathbf{E}(f)$  eine  $\mathbb{R}$ -lineare Abbildung, erfüllt Normierung, Monotonie, und Ausschöpfung.

## Berechnung der Erwartung

## Satz V2I: Erwartungswert von Zufallsvariablen

Sei  $(\Omega, \mathcal{A}, \mathbf{P})$  ein WRaum und  $X : \Omega \rightarrow \bar{\mathbb{R}}$  eine Zufallsvariable.

Ihr **Erwartungswert** (oder kurz ihre **Erwartung**) ist gegeben durch

$$\mu(X) = \mathbf{E}(X) := \int_{\omega \in \Omega} X(\omega) d\mathbf{P} = \int_{x \in \mathbb{R}} x d\mathbf{P}_X.$$

Ist die Verteilung  $\mathbf{P}_X$  auf  $\mathbb{R}$  kontinuierlich mit Dichte  $f : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$ , so gilt

$$\mathbf{E}(X) = \int_{\mathbb{R}} x f(x) dx.$$

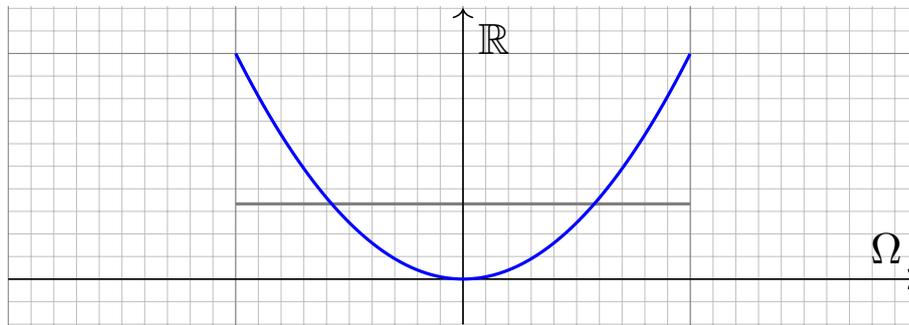
Ist  $\mathbf{P}_X$  diskret mit Wkten  $p(x) = \mathbf{P}(\{\omega \in \Omega \mid X(\omega) = x\})$ , so gilt

$$\mathbf{E}(X) = \sum_{x \in \mathbb{R}} x p(x).$$

Dies ist jeweils nur sinnvoll für  $\mathbf{E}(|X|) < \infty$ . In diesem Fall nennen wir  $X$  **integrierbar** oder sagen: **Der Erwartungswert von  $X$  existiert**.

Für jede Zufallsvariable  $X : \Omega \rightarrow \mathbb{R}$  ist der Erwartungswert das Integral:

$$\mathbf{E}(X) := \int_{\omega \in \Omega} X(\omega) d\mathbf{P} = \int_{x \in \mathbb{R}} x d\mathbf{P}_X$$



**Aufgabe:** Sei  $\mathbf{P}$  die kontinuierliche Gleichverteilung auf  $\Omega = [-1, 1]$ . Berechnen Sie den Erwartungswert von  $X : \Omega \rightarrow \mathbb{R}$  mit  $X(t) = t^2$ .

**Lösung:** Das gelingt durch Integration über  $(\Omega, \mathbf{P})$ :

$$\mathbf{E}(X) = \int_{\Omega} X(t) d\mathbf{P} = \int_{-1}^1 X(t) \cdot \frac{1}{2} dt = \int_{-1}^1 t^2 \cdot \frac{1}{2} dt = \left[ \frac{1}{6} t^3 \right]_{-1}^1 = \frac{1}{3}$$

😊 Die Verteilung  $\mathbf{P}_X$  dieser Zufallsvariable haben wir oben diskutiert. Der Erwartungswert ist anschaulich plausibel: Niedrige Werte nahe 0 werden wahrscheinlicher getroffen als hohe Werte nahe 1. Ein Wert unter  $1/2$  scheint daher plausibel. Den genauen Wert  $1/3$  hingegen muss man ausrechnen. Dafür haben wir alle Techniken zur Integration!

**Aufgabe:** Berechnen Sie den Erwartungswert alternativ mit dem Bildmaß  $\mathbf{P}_X$ . Stimmen beide Ergebnisse überein? Warum ist das so?

**Lösung:** Zu  $\mathbf{P}_X$  auf  $[0, 1]$  haben wir oben die Dichte  $f(x) = 1/(2\sqrt{x})$  ausgerechnet. Hieraus erhalten wir den Erwartungswert wie folgt:

$$\int_{\mathbb{R}} x d\mathbf{P}_X = \int_0^1 x f(x) dx = \int_0^1 \frac{x}{2\sqrt{x}} dx = \int_0^1 \frac{1}{2} x^{1/2} dx = \left[ \frac{1}{3} x^{3/2} \right]_0^1 = \frac{1}{3}$$

Beide Integrale sind gleich dank Transformationssatz / Substitution!

Die Erwartung lässt sich auf zwei Weisen berechnen: als Integral  $\int_{\Omega} X(\omega) d\mathbf{P}$  der messbaren Funktion  $X : \Omega \rightarrow \mathbb{R}^n$  bezüglich des Maßes  $\mathbf{P}$  auf  $\Omega$ , oder als Schwerpunkt  $\int_{\mathbb{R}^n} x d\mathbf{P}_X$  des Bildmaßes  $\mathbf{P}_X$  auf  $\mathbb{R}^n$ . Beide ergeben dasselbe Ergebnis. Daher darf sich jeder aussuchen, welcher Rechenweg im konkreten Beispiel jeweils bequemer ist. Ich werde im Folgenden meist Bildmaße auf  $\mathbb{R}^n$  vorziehen, da uns diese aus der  $n$ -dimensionalen Integration vertraut sind.

## Definition V2J: Erwartung, Varianz, Streuung

Sei  $p: \mathbb{R}^n \supseteq \Omega \rightarrow [0, 1]$  eine diskrete WVerteilung, also  $\sum_{x \in \Omega} p(x) = 1$ .  
Der Schwerpunkt von  $p$  heißt **Mittelwert** oder **Erwartung**:

$$\mu := \sum_{x \in \Omega} x p(x) \in \mathbb{R}^n.$$

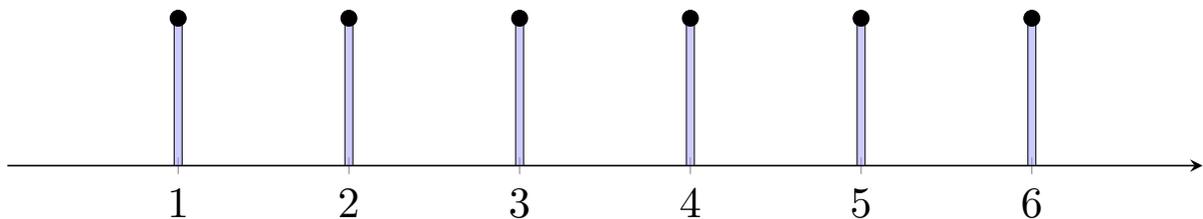
Hierbei setzen wir absolute Summierbarkeit voraus:  $\sum |x| p(x) < \infty$ .  
Die **Varianz** der Verteilung  $p$  ist ihr Trägheitsmoment bezüglich  $\mu$ :

$$\sigma^2 := \sum_{x \in \Omega} |x - \mu|^2 p(x) = \left[ \sum_{x \in \Omega} |x|^2 p(x) \right] - |\mu|^2 \geq 0.$$

Die Wurzel  $\sigma \geq 0$  dieses Wertes heißt die **Streuung** der Verteilung.

Anschaulich misst  $\sigma$ , wie weit die Werte um den Mittelwert  $\mu$  streuen.  
Wir nennen die Streuung  $\sigma$  daher auch die **Standardabweichung**.  
Genau dann gilt  $\sigma = 0$ , wenn  $p(\mu) = 1$  auf den Punkt  $\mu$  konzentriert ist.

## Beispiel: Erwartung und Varianz beim Würfeln



**Aufgabe:** Welchen **Mittelwert**  $\mu \in \mathbb{R}$  hat die Augenzahl beim Würfeln?

Was ist hier die Varianz  $\sigma^2 = \sum (x - \mu)^2 p(x)$ ? Und die Streuung  $\sigma$ ?

**Lösung:** Jede der Zahlen 1, 2, 3, 4, 5, 6 tritt in  $\frac{1}{6}$  der Fälle auf. Also:

$$\mu = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{21}{6} = 3.5$$

Anschaulich ist  $\mu$  der **Schwerpunkt** der Verteilung  $p$ .

Die **Varianz**  $\sigma^2$  misst die mittlere quadratische Abweichung von  $\mu$ :

$$\begin{aligned} \sigma^2 &= \frac{1}{6} \left[ \left(1 - \frac{7}{2}\right)^2 + \left(2 - \frac{7}{2}\right)^2 + \left(3 - \frac{7}{2}\right)^2 + \left(4 - \frac{7}{2}\right)^2 + \left(5 - \frac{7}{2}\right)^2 + \left(6 - \frac{7}{2}\right)^2 \right] \\ &= \frac{1}{3} \left[ \left(\frac{5}{2}\right)^2 + \left(\frac{3}{2}\right)^2 + \left(\frac{1}{2}\right)^2 \right] = \frac{1}{3} \left[ \frac{25}{4} + \frac{9}{4} + \frac{1}{4} \right] = \frac{35}{12} = 2.91666\dots \end{aligned}$$

Die **Streuung** ist demnach  $\sigma = \sqrt{35/12} \approx 1.70783$ .

**Definition V2k: Erwartung, Varianz, Streuung**

Sei  $f : \mathbb{R}^n \supseteq \Omega \rightarrow \mathbb{R}_{\geq 0}$  eine WDichte, also  $\int_{\Omega} f(x) dx = 1$ .

Der Schwerpunkt von  $f$  heißt **Mittelwert** oder **Erwartung**:

$$\mu := \int_{\Omega} x f(x) dx$$

Hierbei setzen wir absolute Integrierbarkeit voraus:  $\int |x| f(x) dx < \infty$ .

Die **Varianz** der Verteilung  $f$  ist ihr Trägheitsmoment bezüglich  $\mu$ :

$$\sigma^2 := \int_{\Omega} |x - \mu|^2 f(x) dx = \int_{\Omega} |x|^2 f(x) dx - |\mu|^2 \geq 0$$

Die Wurzel  $\sigma \geq 0$  dieses Wertes heißt die **Streuung** der Verteilung.

Wir nutzen hier euklidische Norm  $|\cdot|$  und Skalarprodukt  $\langle \cdot | \cdot \rangle$  des  $\mathbb{R}^n$ .

Anschaulich misst  $\sigma$ , wie weit die Werte um den Mittelwert  $\mu$  streuen.

Wir nennen die Streuung  $\sigma$  daher auch die **Standardabweichung**.

**Wichtige Verteilungen und ihre Kenngrößen**

<b>diskret</b>	$\Omega$	$p(k) = \mathbf{P}(\{k\})$	<b>E</b>	<b>V</b>
Gleichverteilung <small>V233</small>	$\{a, \dots, b\}$	$\frac{1}{b-a+1}$	$\frac{a+b}{2}$	$\frac{(b-a+1)^2-1}{12}$
Hyperg. $H(N, K, n)$ <small>U419</small>	$\{0, \dots, n\}$	$\binom{K}{k} \binom{N-K}{n-k} / \binom{N}{n}$	$n \frac{K}{N}$	$n \frac{K}{N} \frac{N-K}{N} \frac{N-n}{N-1}$
Binomial $B(n, t)$ <small>U421</small>	$\{0, \dots, n\}$	$\binom{n}{k} t^k (1-t)^{n-k}$	$nt$	$nt(1-t)$
Poisson $P(\lambda)$ <small>U425</small>	$\mathbb{N}$	$e^{-\lambda} \lambda^k / k!$	$\lambda$	$\lambda$
Geometrisch $G(q)$ <small>T421</small>	$\mathbb{N}_{\geq 1}$	$(1-q)q^{k-1}$	$\frac{1}{1-q}$	$\frac{q}{(1-q)^2}$
Zeta $Z(s), s = 2$ <small>T312</small>	$\mathbb{N}_{\geq 1}$	$\frac{6}{\pi^2} \frac{1}{k^2}$	— nicht definiert! — Diese Reihen divergieren.	
<b>kontinuierlich</b>	$\Omega$	Dichte $f(x)$	<b>E</b>	<b>V</b>
Gleichverteilung <small>V235</small>	$[a, b]$	$\frac{1}{b-a}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponential $E(\lambda)$ <small>V243</small>	$\mathbb{R}_{\geq 0}$	$\lambda e^{-\lambda x}$	$\lambda^{-1}$	$\lambda^{-2}$
Normal $N(\mu, \sigma^2)$ <small>V239</small>	$\mathbb{R}$	$\frac{e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}}{\sigma\sqrt{2\pi}}$	$\mu$	$\sigma^2$
Cauchy $C(m, a)$ <small>V241</small>	$\mathbb{R}$	$\frac{1}{\pi} \frac{a}{a^2+(x-m)^2}$	— nicht definiert! — Diese Integrale divergieren.	

**Aufgabe:** Seien  $a \leq b$  in  $\mathbb{Z}$  ganze Zahlen. Berechnen Sie Erwartung und Varianz der diskreten Gleichverteilung auf  $\Omega = \{a, a + 1, \dots, b\}$ .

**Lösung:** Wir setzen die Definition ein und rechnen's aus:

$$\mu = \frac{1}{n} \sum_{k=a}^b k = \frac{a+b}{2}$$

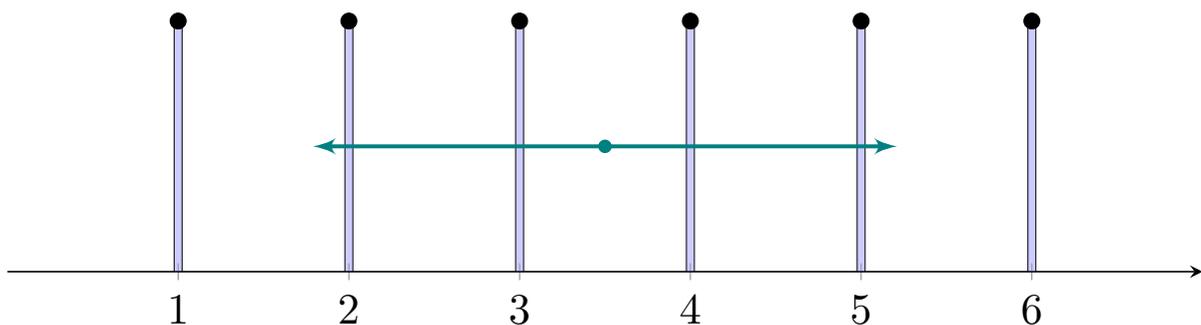
$$\sigma^2 = \left[ \frac{1}{n} \sum_{k=a}^b k^2 \right] - \mu^2 = \frac{n^2 - 1}{12}$$

Zur Abkürzung setzen wir  $n = |\Omega| = b - a + 1$  für die Elementezahl. Für große  $n$  ist die Streuung näherungsweise  $\sigma \approx n/\sqrt{12} \approx n/3.46$ . Zur Rechnung nutzen wir die bekannten Summenformeln (HDI B3B):

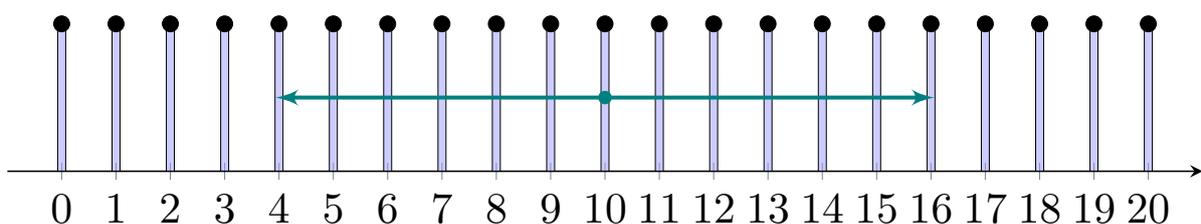
$$\sum_{k=a}^b k = (b - a + 1) \frac{a + b}{2},$$

$$\sum_{k=a}^b k^2 = (b - a + 1) \frac{2a^2 + 2ab + 2b^2 + b - a}{6}.$$

**Beispiel:** Für  $a = 1, b = 6$  gilt  $\mu = 3.5, \sigma^2 = 35/12, \sigma \approx 1.71$ .



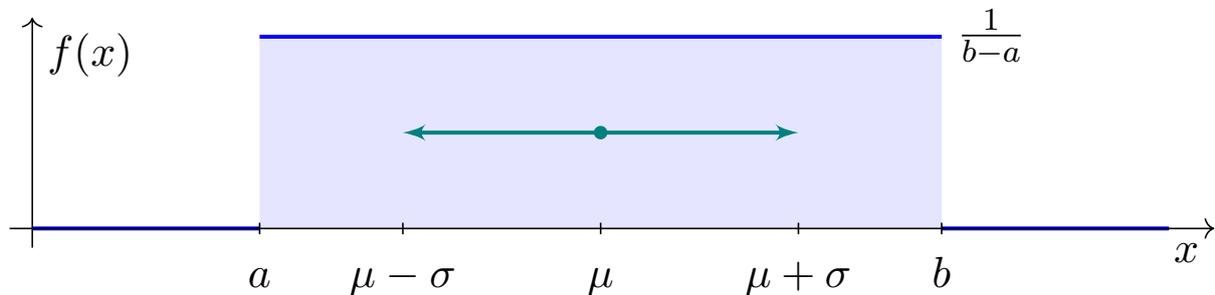
**Beispiel:** Für  $a = 0, b = 20$  gilt  $\mu = 10, \sigma^2 = 440/12, \sigma \approx 6.06$ .



😊 Anschaulich ist der Mittelwert  $\mu$  der Schwerpunkt der Verteilung. Die Standardabweichung  $\sigma$  misst die typische Breite der Verteilung.

Für  $a < b$  in  $\mathbb{R}$  und  $n \in \mathbb{N}_{\geq 2}$  betrachten wir die Schrittweite  $h = \frac{b-a}{n-1}$  und die diskrete Gleichverteilung auf  $\Omega := \{a + kh \mid k = 0, 1, \dots, n-1\}$ . Diese diskreten Verteilungen sind auf den vorigen Folien skizziert.

Im Grenzwert für  $n \rightarrow \infty$  erhalten wir als Idealisierung folgendes Bild:



Für  $a < b$  in  $\mathbb{R}$  betrachten wir die kontinuierliche Gleichverteilung auf dem Intervall  $[a, b]$ . Dies ist eine kontinuierliche Verteilung mit Dichte

$$f : [a, b] \rightarrow \mathbb{R}_{\geq 0} : x \mapsto f(x) = \frac{1}{b-a}.$$

**Aufgabe:** Berechnen Sie Erwartung und Varianz der kontinuierlichen Gleichverteilung auf  $[a, b]$ , mit Dichte  $f : [a, b] \rightarrow \mathbb{R}_{\geq 0} : x \mapsto f(x) = \frac{1}{b-a}$ .

**Lösung:** Wir setzen die Definition ein und rechnen's aus:

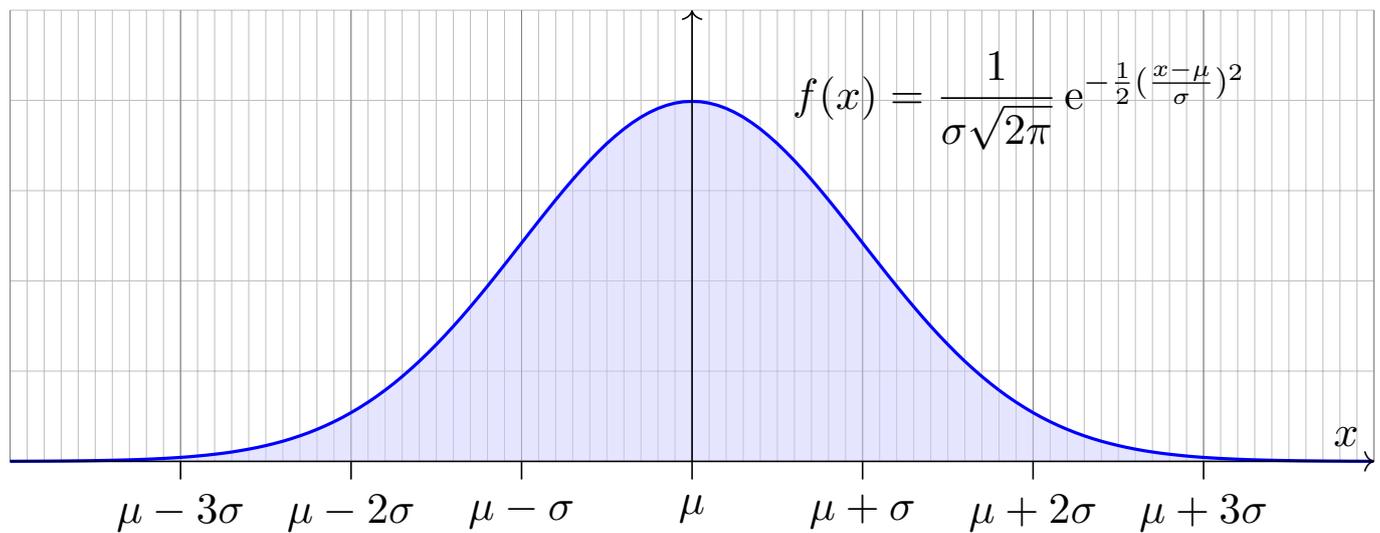
$$\mu = \frac{1}{b-a} \int_{x=a}^b x \, dx = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}$$

$$\sigma^2 = \frac{1}{b-a} \int_{x=a}^b x^2 \, dx - \mu^2 = \frac{b^3 - a^3}{3(b-a)} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}$$

$$\sigma = \sqrt{\sigma^2} = \frac{b-a}{2\sqrt{3}} \approx \frac{b-a}{3.46}$$

😊 Wie zuvor im diskreten Fall ist der Mittelwert  $\mu$  sehr anschaulich, die Varianz  $\sigma^2$  hingegen müssen wir als Integral explizit ausrechnen. Die ersten Beispiele hierzu sind nicht schwer und eine gute Übung, sowohl zur Schulung Ihrer Intuition als auch Ihrer Rechentechniken.

Für  $\mu, \sigma \in \mathbb{R}$  mit  $\sigma > 0$  ist die **Normalverteilung**  $N(\mu, \sigma^2)$  gegeben durch folgende Dichte, die allseits beliebte Gaußsche Glockenkurve:



- Aufgabe:** (1) Warum ist dies eine Wahrscheinlichkeitsverteilung?  
 (2) Skizzieren Sie  $x f(x)$  und berechnen Sie die Erwartung.  
 (3) Skizzieren Sie  $x^2 f(x)$  und berechnen Sie die Varianz.

## Normalverteilung: Gesamtmasse

**Lösung:** Offenbar gilt  $f \geq 0$ . Wir haben drei Integrale zu berechnen:

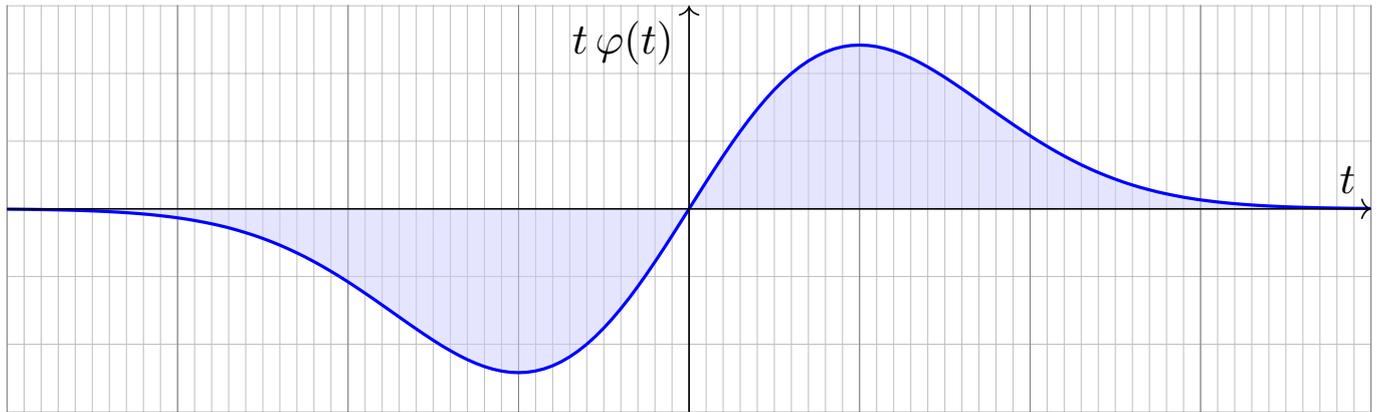
$$\int_{\mathbb{R}} f(x) dx = 1, \quad \int_{\mathbb{R}} x f(x) dx = \mu, \quad \int_{\mathbb{R}} (x - \mu)^2 f(x) dx = \sigma^2.$$

Vereinfachung: Nach Substitution  $t = (x - \mu)/\sigma$  erhalten wir  $\mu = 0$  und  $\sigma = 1$ , also die Dichte der Standard-Normalverteilung:

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

(1) Wir berechnen  $\int_{\mathbb{R}} \varphi(t) dt$  durch den Gaußschen Kunstgriff [C230](#):

$$\begin{aligned} \left( \int_{\mathbb{R}} e^{-t^2/2} dt \right)^2 &= \left( \int_{\mathbb{R}} e^{-x^2/2} dx \right) \cdot \left( \int_{\mathbb{R}} e^{-y^2/2} dy \right) \\ &\stackrel{\text{Fub}}{=} \int_{\mathbb{R} \times \mathbb{R}} e^{-(x^2+y^2)/2} d(x, y) \stackrel{\text{Trafo}}{=} \int_{\mathbb{R}_{\geq 0} \times [0, 2\pi]} e^{-r^2/2} r d(r, \alpha) \\ &\stackrel{\text{Fub}}{=} \int_{r=0}^{\infty} \int_{\alpha=0}^{2\pi} e^{-r^2/2} r d\alpha dr \stackrel{\text{HDI}}{=} 2\pi \left[ -e^{-r^2/2} \right]_{r=0}^{\infty} = 2\pi \end{aligned}$$



(2) Die Erwartung ist Null aufgrund der Symmetrie  $\varphi(-t) = \varphi(t)$ :

$$\mu = \int_{-\infty}^{\infty} t \varphi(t) dt = 0.$$

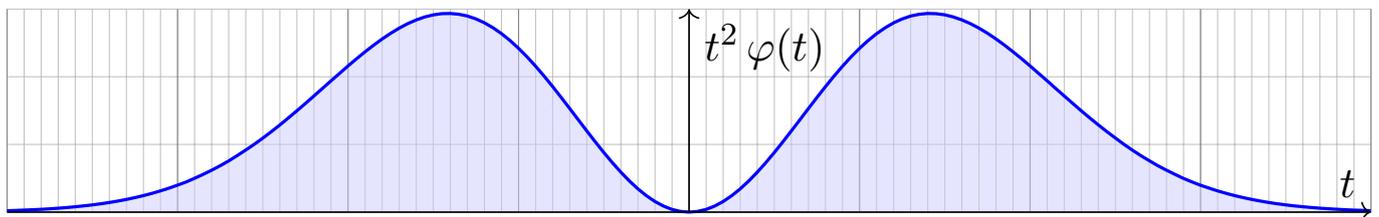
Alternativ können wir explizit eine Stammfunktion angeben:

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t e^{-t^2/2} dt \stackrel{\text{HDI}}{=} \frac{1}{\sqrt{2\pi}} \left[ -e^{-t^2/2} \right]_{-\infty}^{\infty} = 0$$

Der allgemeine Fall folgt mit Substitution  $t = (x - \mu)/\sigma$ :

Die Erwartung der Normalverteilung  $N(\mu, \sigma^2)$  ist  $\mu$ .

## Normalverteilung: Varianz



(3) Die Varianz berechnen wir mit partieller Integration:

$$\begin{aligned} \sigma^2 &\stackrel{\text{Def}}{=} \int_{-\infty}^{\infty} t^2 \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (-t) \cdot (-t) e^{-t^2/2} dt \\ &\stackrel{\text{part}}{=} \frac{1}{\sqrt{2\pi}} \left[ -t \cdot e^{-t^2/2} \right]_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-t^2/2} dt = 1 \end{aligned}$$

Der allgemeine Fall folgt mit Substitution  $t = (x - \mu)/\sigma$ :

Die Varianz der Normalverteilung  $N(\mu, \sigma^2)$  ist  $\sigma^2$ .

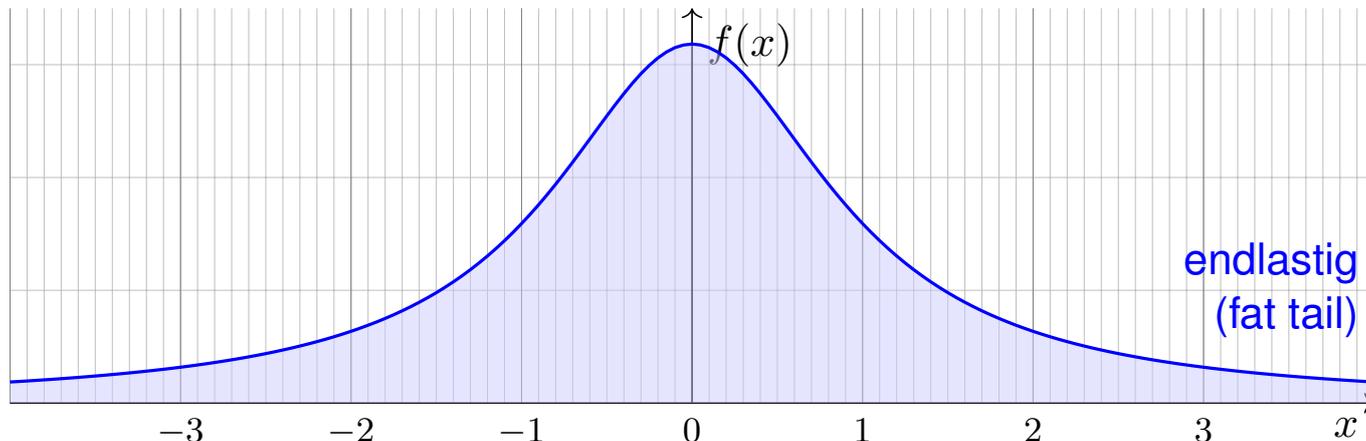
😊 Siehe D419 für die geraden Momente  $\rho^{2k}$  der Normalverteilung. Ungerade Momente  $\int_{\mathbb{R}} t^{2k+1} \varphi(t) dt$  verschwinden dank Symmetrie.

**Übung:** Legen Sie die Vorlage beiseite und versuchen Sie es selbst!

**Aufgabe:** Ist die folgende Funktion eine Wahrscheinlichkeitsdichte?

$$f : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto \frac{1/c}{1+x^2}$$

Für welche Konstanten  $c \in \mathbb{R}$ ? Existieren hierzu Erwartung und Varianz?



⚠ Der Graph erinnert grob an die Glockenkurve  $\varphi(x) = e^{-x^2/2}/\sqrt{2\pi}$ . Doch Vorsicht: Die Cauchy–Verteilung hat völlig andere Eigenschaften! Vor allem klingt  $f(x)$  nicht exponentiell ab, sondern nur wie  $x \mapsto 1/x^2$ . Das hat dramatische Konsequenzen: Die Erwartung existiert nicht!

**Lösung:** (1) Wir haben  $f \geq 0$  und für  $c = 1$  ist die Gesamtmasse

$$\int_{\mathbb{R}} f(x) dx = \int_{\mathbb{R}} \frac{1}{1+x^2} dx = \left[ \arctan(x) \right]_{-\infty}^{\infty} = \frac{\pi}{2} - \left( -\frac{\pi}{2} \right) = \pi$$

😊 Die Wahl  $c = \pi$  normiert die Gesamtmasse auf  $\int_{\mathbb{R}} f(x) dx = 1$ . Erst hierdurch wird  $f$  zu einer Wahrscheinlichkeitsdichte.

(2) Für Erwartung und Varianz benötigen wir folgende Integrale:

$$\mu \stackrel{?}{=} \frac{1}{\pi} \int_{\mathbb{R}} \frac{x}{1+x^2} dx \quad \text{und} \quad \sigma^2 \stackrel{?}{=} \frac{1}{\pi} \int_{\mathbb{R}} \frac{x^2}{1+x^2} dx = \infty.$$

Der erste Integrand ist nicht absolut integrierbar, siehe B417.

Der Median ist offensichtlich 0, aber die Erwartung existiert nicht!

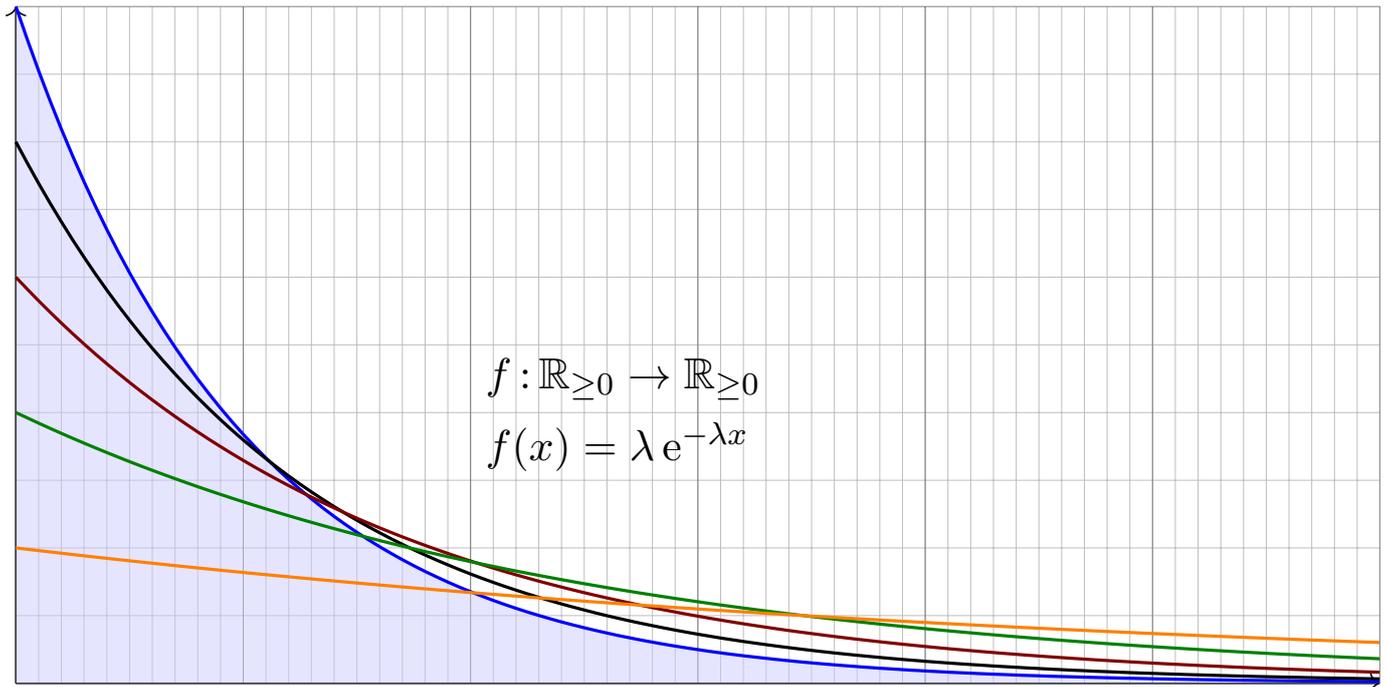
Der zweite Integrand ist nicht-negativ, sein Integral ist unendlich.

⚠ Hier haben Erwartung und Varianz demnach keinen Sinn.

Solche Schwierigkeiten vermeiden wir durch absolute Integrierbarkeit!

Dies garantiert für unsere Integrale alle gewohnten Rechenregeln.

Für  $\lambda > 0$  ist die **Exponentialverteilung**  $E(\lambda)$  gegeben durch die Dichte



**Aufgabe:** (1) Ist dies eine Wahrscheinlichkeitsverteilung?  
 (2) Berechnen Sie ihre Erwartung und ihre Varianz.

**Lösung:** (1) Ja: Auf  $\Omega = \mathbb{R}_{\geq 0}$  gilt  $f \geq 0$  und  $\int_{\Omega} f(x) dx = 1$ , denn

$$\int_{x=0}^{\infty} \lambda e^{-\lambda x} dx = \left[ -e^{-\lambda x} \right]_{x=0}^{\infty} = 1.$$

(2) Die Erwartung berechnen wir mit partieller Integration:

$$\begin{aligned} \mu &= \int_{x=0}^{\infty} x \lambda e^{-\lambda x} dx \\ &= \left[ -x e^{-\lambda x} \right]_{x=0}^{\infty} + \int_{x=0}^{\infty} e^{-\lambda x} dx = \left[ -\frac{1}{\lambda} e^{-\lambda x} \right]_{x=0}^{\infty} = \frac{1}{\lambda} \end{aligned}$$

Die Varianz berechnen wir ebenso:

$$\begin{aligned} \sigma^2 + \mu^2 &= \int_{x=0}^{\infty} x^2 \lambda e^{-\lambda x} dx = \left[ -x^2 e^{-\lambda x} \right]_{x=0}^{\infty} + \int_{x=0}^{\infty} 2x e^{-\lambda x} dx \\ &= \frac{2}{\lambda} \int_{x=0}^{\infty} x \lambda e^{-\lambda x} dx = \frac{2}{\lambda^2} \end{aligned}$$

Hieraus folgt  $\sigma^2 = 1/\lambda^2$ .

*Every day I get in the queue // To get on the bus that takes me to you.  
I'm so nervous, I just sit and smile. // Your house is only another mile.*

**Aufgabe:** Ihre Buslinie fährt streng nach Fahrplan alle 20 Minuten, also zum Beispiel um 6:00, 6:20, 6.40, etc. bis Dienstschluss um 24:00 Uhr. Sie kommen zufällig irgendwann im Laufe des Tages an die Haltestelle. Wie lang ist Ihre mittlere Wartezeit? (Gemeint ist der Erwartungswert, aber die Formulierung „erwartete Wartezeit“ klingt etwas seltsam.)

**Lösung:** Sei  $t$  Ihre Ankunftszeit und  $t_0$  die Abfahrtszeit des letzten Busses, den Sie verpasst haben. Dann gilt  $t_0 < t \leq t_0 + 20$  mit Wartezeit  $T = 20 - (t - t_0) \in [0, 20[$ . Daher ist Ihre mittlere Wartezeit:

$$\bar{T} = \frac{1}{20} \int_{t=t_0}^{t_0+20} \underbrace{20 - (t - t_0)}_{=\tau} dt = \frac{1}{20} \int_{\tau=0}^{20} \tau d\tau = \frac{1}{20} \left[ \frac{\tau^2}{2} \right]_{\tau=0}^{20} = 10$$

Sie müssen also im Durchschnitt 10min auf den nächsten Bus warten.

😊 Diese Rechnung ist völlig plausibel und wird Sie kaum überraschen. Sie dient zum Aufwärmen und als Kontrast für die folgenden Aufgaben.

**Aufgabe:** Die Exponentialverteilung hat kein Gedächtnis! Genauer, für jede exponentialverteilte Zufallsvariable  $X \sim E(\lambda)$  und  $s, t \in \mathbb{R}_{\geq 0}$  gilt:

$$\mathbf{P}(X \geq s + t \mid X \geq s) = \mathbf{P}(X \geq t) = e^{-\lambda t}$$

**Lösung:** Wir nutzen die Definitionen und rechnen es geduldig aus:

$$\mathbf{P}(X \geq s) = \int_{x=s}^{\infty} \lambda \cdot e^{-\lambda x} dx = - \left[ e^{-\lambda x} \right]_{x=s}^{\infty} = e^{-\lambda s}$$

$$\mathbf{P}(\{X \geq s + t\} \cap \{X \geq s\}) = \mathbf{P}(X \geq s + t) = e^{-\lambda(s+t)}$$

$$\mathbf{P}(X \geq s + t \mid X \geq s) = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = \mathbf{P}(X \geq t)$$

Die Exponentialverteilung modelliert kontinuierliche Wartezeiten. Sie ist ohne Gedächtnis: Wenn Sie schon bis zur Zeit  $s$  gewartet haben, so ist die weitere Wartezeit  $t$  davon unabhängig!

Diskrete Analogie ist die geometrische Verteilung [\[T421\]](#): „Ich kaufe jeden Tag ein Rubbellos mit Gewinnwkt 1:100. Im Mittel warte ich also 100 Tage auf den nächsten Gewinn. Seit 70 Tagen bin ich ohne Gewinn, also warte ich nur noch 30 Tage.“ Ist das plausibel? Was sagt die Rechnung?

**Aufgabe:** Ihr Bus fährt nun nicht mehr nach Fahrplan, sondern zufällig: Der Abstand aufeinanderfolgender Busse sei exponentialverteilt mit Dichte  $f(t) = e^{-t/20}/20$ . (Wir messen die Zeit  $t$  hier in Minuten.)

(1) Wie lang ist der mittlere Zeitabstand zwischen zwei Bussen?

(2) Sie kommen zufällig irgendwann am Tag an die Haltestelle.

Wie lang ist Ihre mittlere Wartezeit? Warum ist das erstaunlich?

Können Sie das Ergebnis dennoch möglichst anschaulich erklären?

(3) Diskutieren Sie das diskrete Analogon: An der Wendestelle lässt der Busfahrer zunächst den Vorgänger abfahren. Er würfelt jede Minute eine Zahl  $1, \dots, 20$  (mit einem regulären Ikosaeder), bei einer 1 fährt er los.

(4) Welcher Übergang führt vom diskreten zum kontinuierlichen Modell?

**Lösung:** (1) Diesen Erwartungswert haben wir bereits berechnet:

$$\mu = \int_{t=0}^{\infty} t \lambda e^{-\lambda t} dt = \left[ -t e^{-\lambda t} \right]_{t=0}^{\infty} + \int_{t=0}^{\infty} e^{-\lambda t} dt = \left[ -\frac{1}{\lambda} e^{-\lambda x} \right]_{t=0}^{\infty} = \frac{1}{\lambda}$$

In unserem Beispiel ist  $\lambda = 1/20$ , also der mittlere Abstand 20 Minuten.

(2) Nehmen wir zunächst an, Sie haben gerade einen Bus verpasst. Dann ist Ihre mittlere Wartezeit 20min wie zuvor in (1) berechnet.

Allgemeiner Fall: Sie haben den letzten Bus um  $s \geq 0$  Minuten verpasst. Nach der obigen Aufgabe hat die Exponentialverteilung kein Gedächtnis. Daher ist Ihre mittlere Wartezeit immer 20min! Vielleicht dachten Sie, das sei Ihr besonderes Pech. Jetzt kennen Sie den wahren Grund!

Naiv würde man hoffen, die Wartezeit wäre nur halb so groß, aber nein! Anschaulich erklärt: Mit großer Wkt treffen Sie große Intervalle, leider. Daher warten Sie im Mittel nicht nur 10min, sondern tatsächlich 20min.

(3) Die Wkt, nach einer Minute loszufahren, ist  $p = 1/20$ .

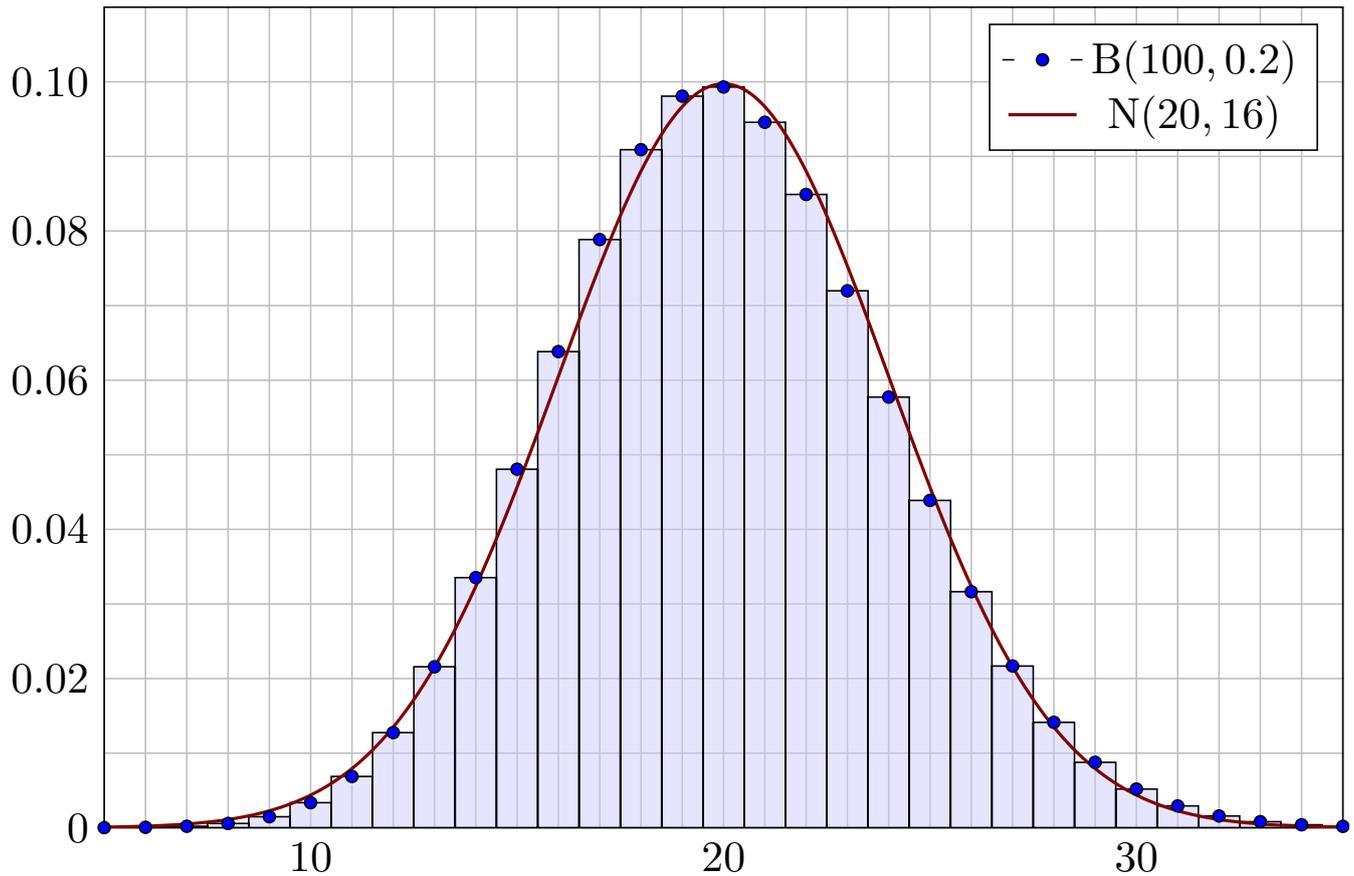
Die Wkt, nach zwei Minuten loszufahren, ist  $q \cdot p$  mit  $q := 1 - p$ .

Die Wkt, nach  $k$  Minuten loszufahren, ist dementsprechend  $q^{k-1} \cdot p$ .

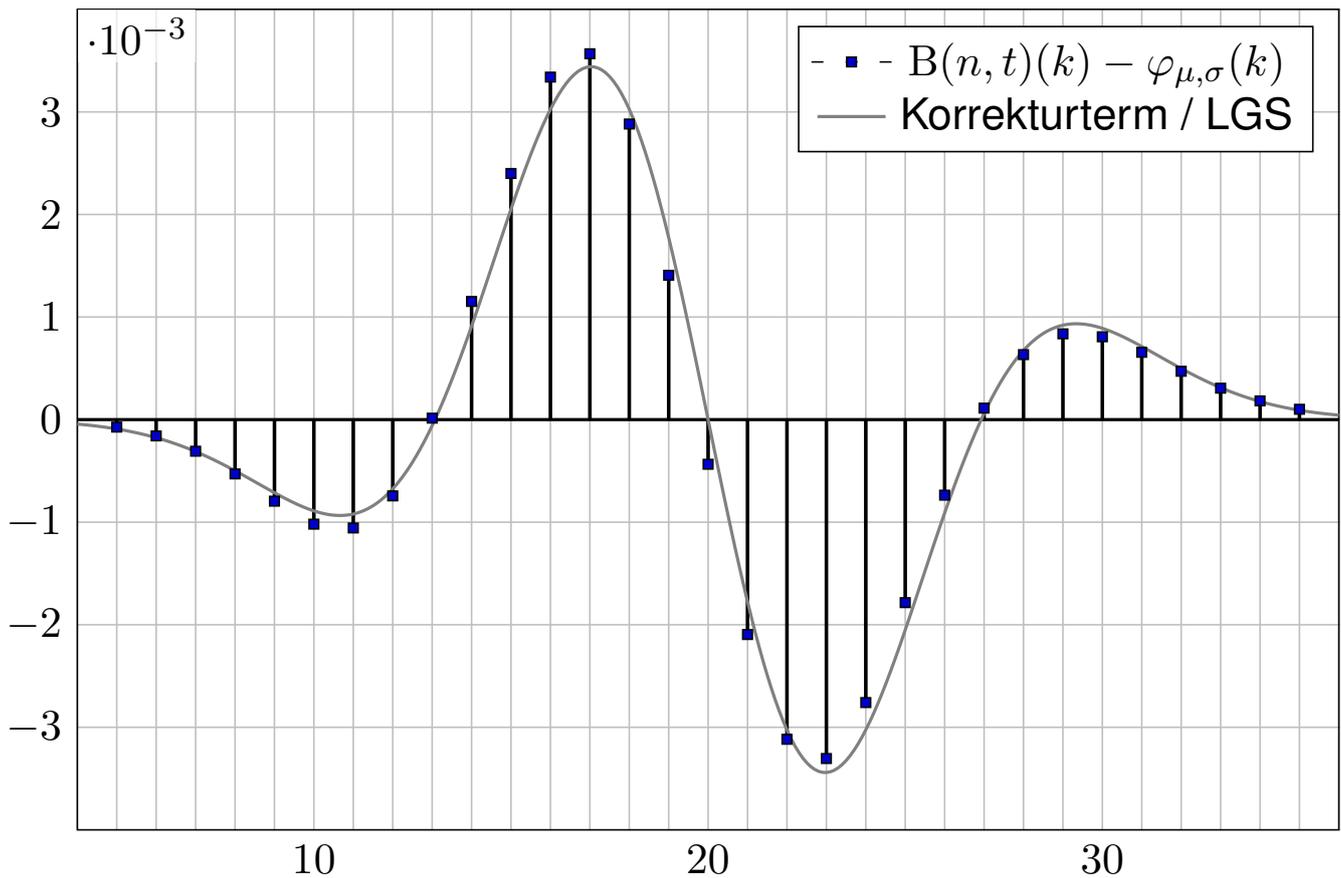
Dies ist die geometrische Verteilung  $G(q)$ . Der mittlere Abstand ist  $1/p$ , also 20 Minuten. In diesem diskreten Modell gilt dasselbe wie in (2).

(4) Lassen wir den Zeitschritt  $\Delta t$  gegen Null gehen, so erhalten wir aus der geometrischen Verteilung die Exponentialverteilung, siehe V203.

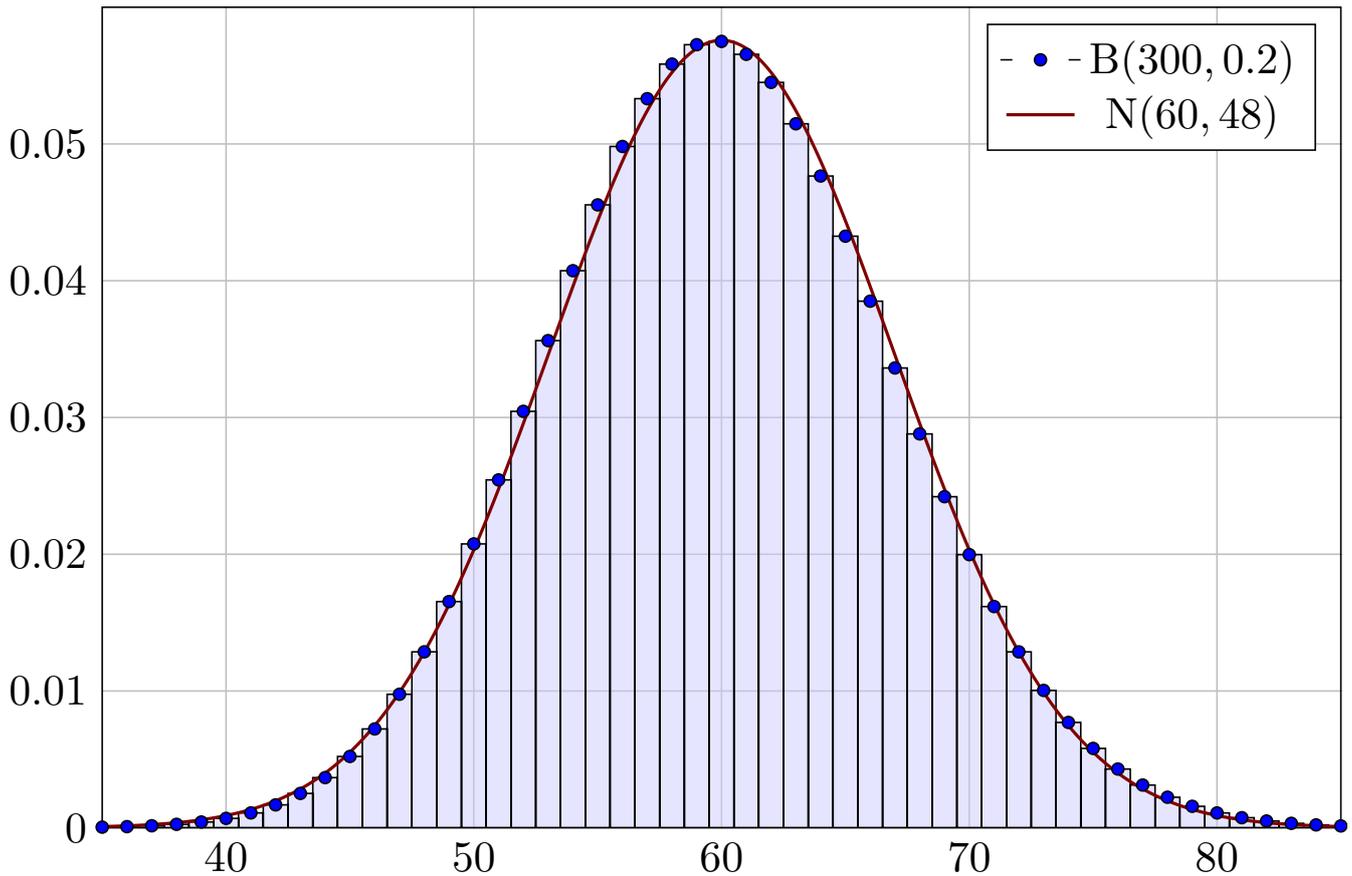
**Beispiel:**  $n = 100$  und  $t = 0.2$ , also  $\mu = nt = 20$  und  $\sigma^2 = nt(1-t) = 16$



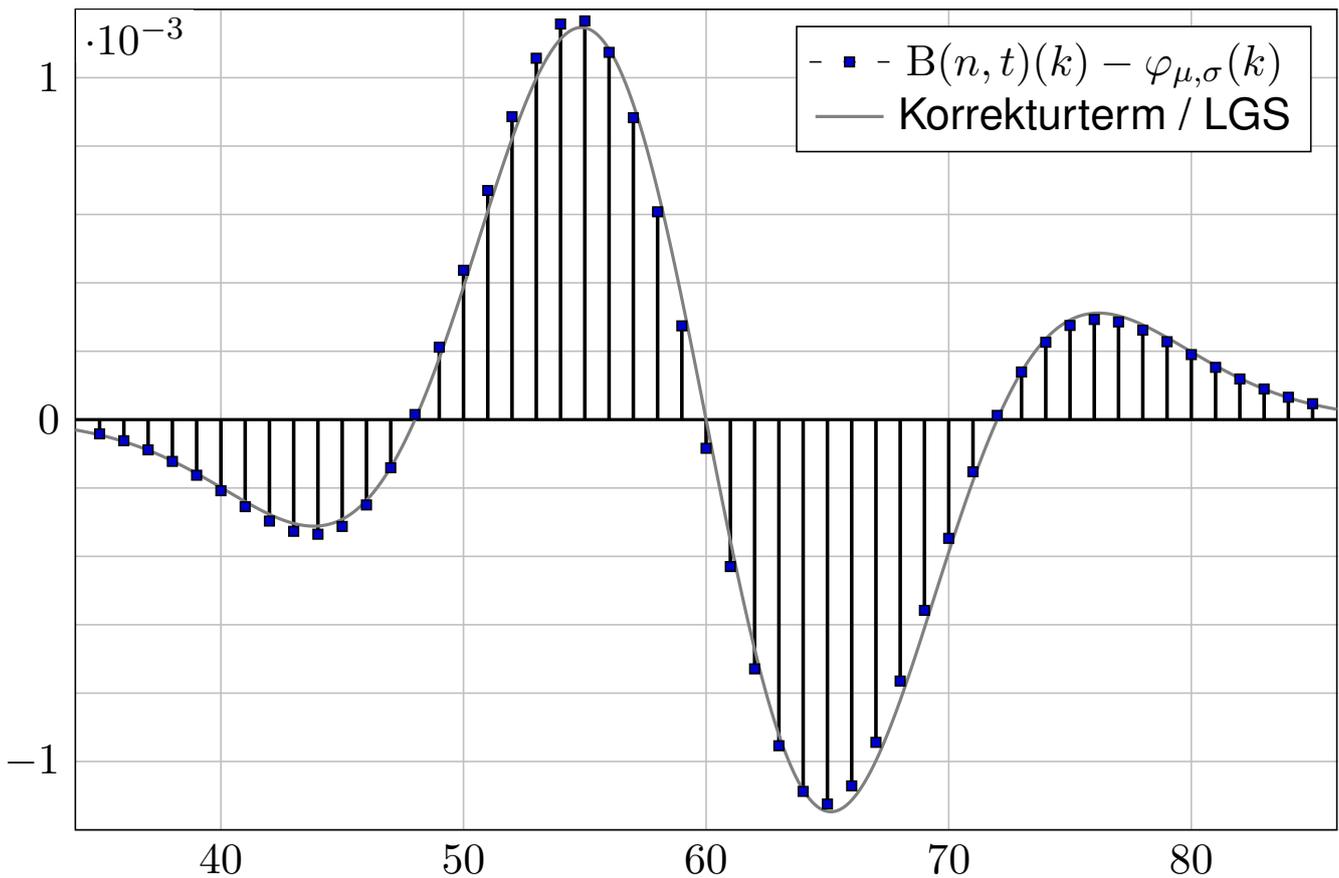
Wie gut ist die Näherung? Punktweise bis auf 0.004 genau!



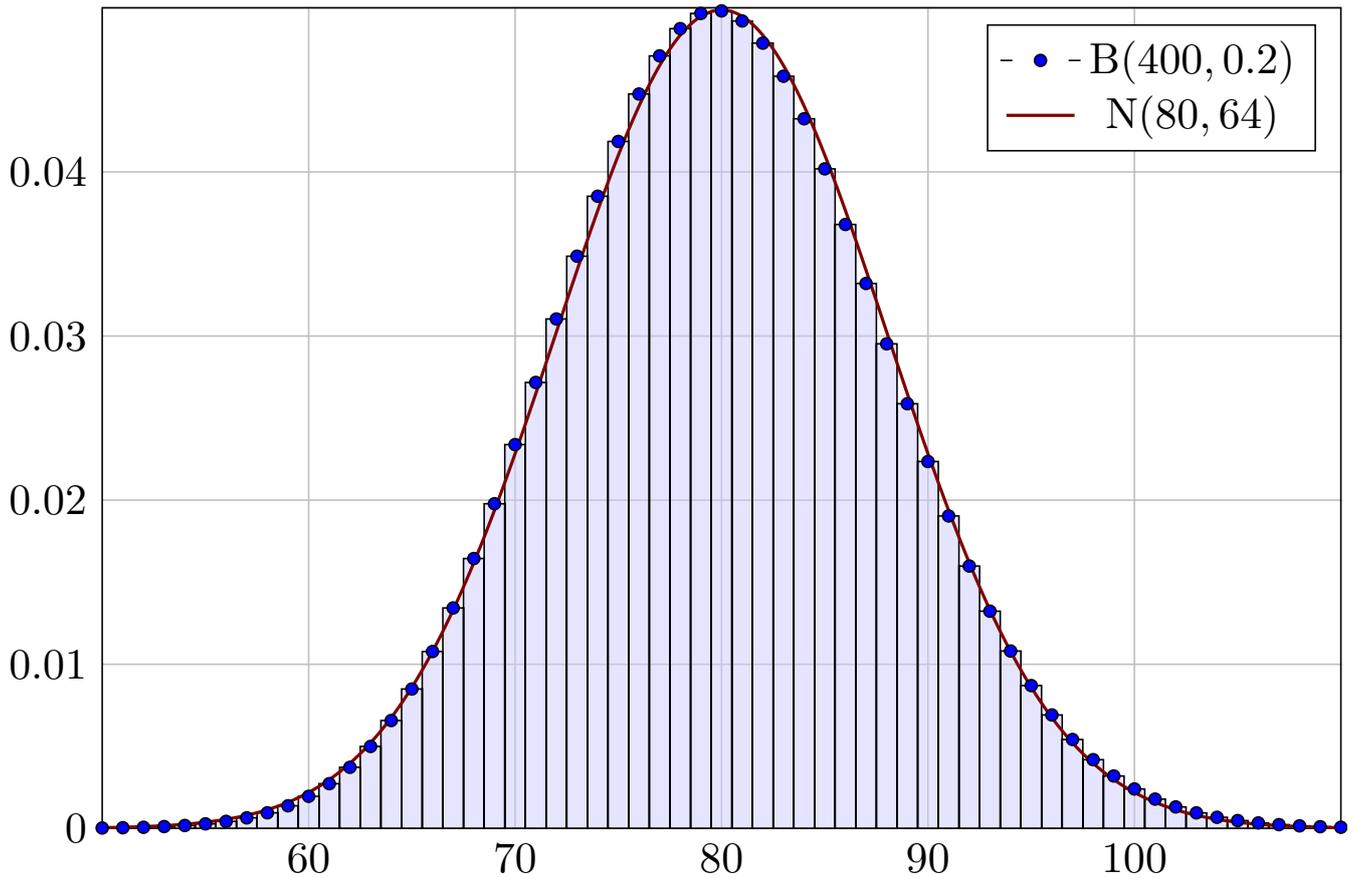
**Beispiel:**  $n = 300$  und  $t = 0.2$ , also  $\mu = nt = 60$  und  $\sigma^2 = nt(1-t) = 48$



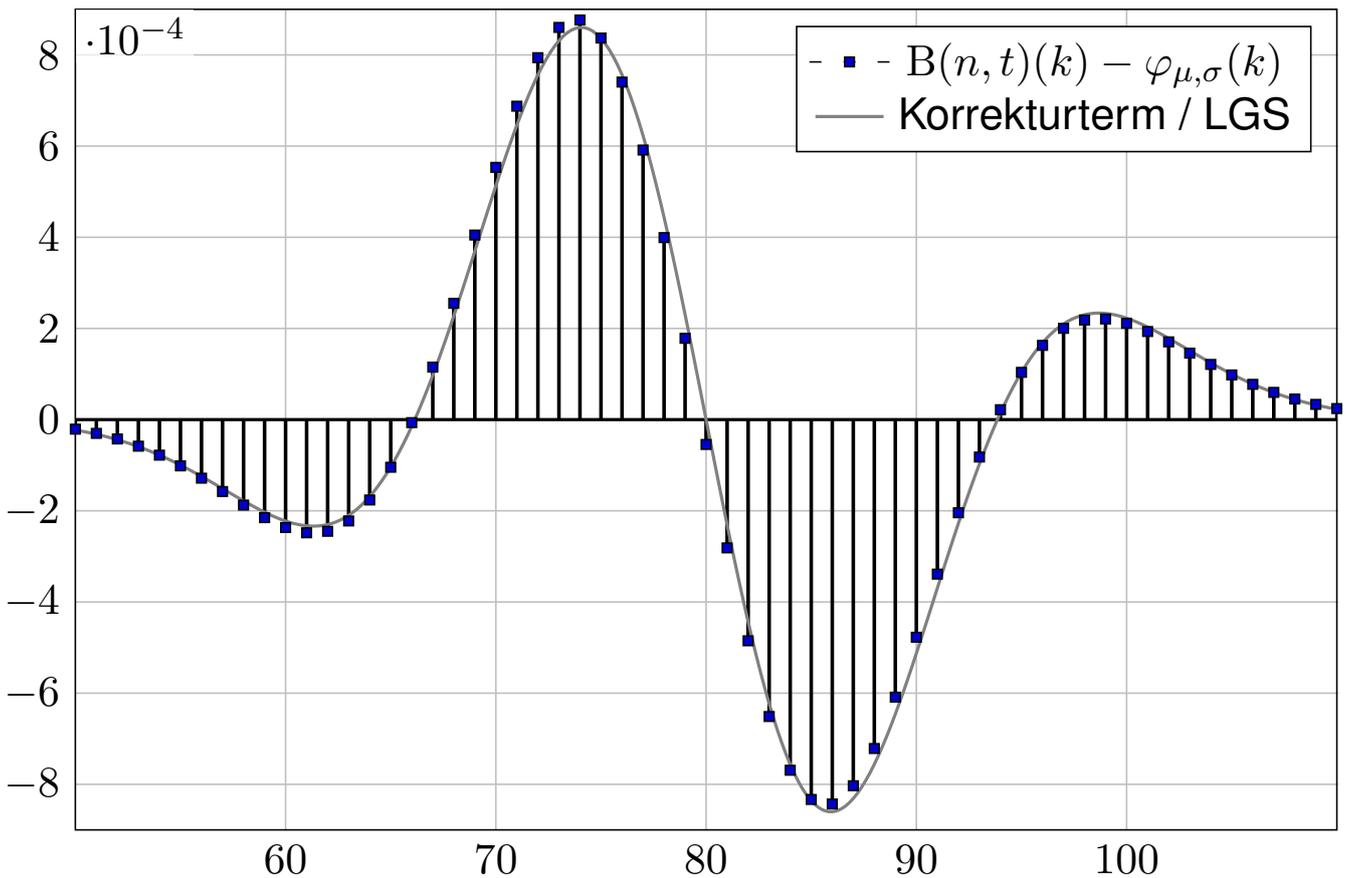
Wie gut ist die Näherung? Punktweise bis auf 0.0012 genau!



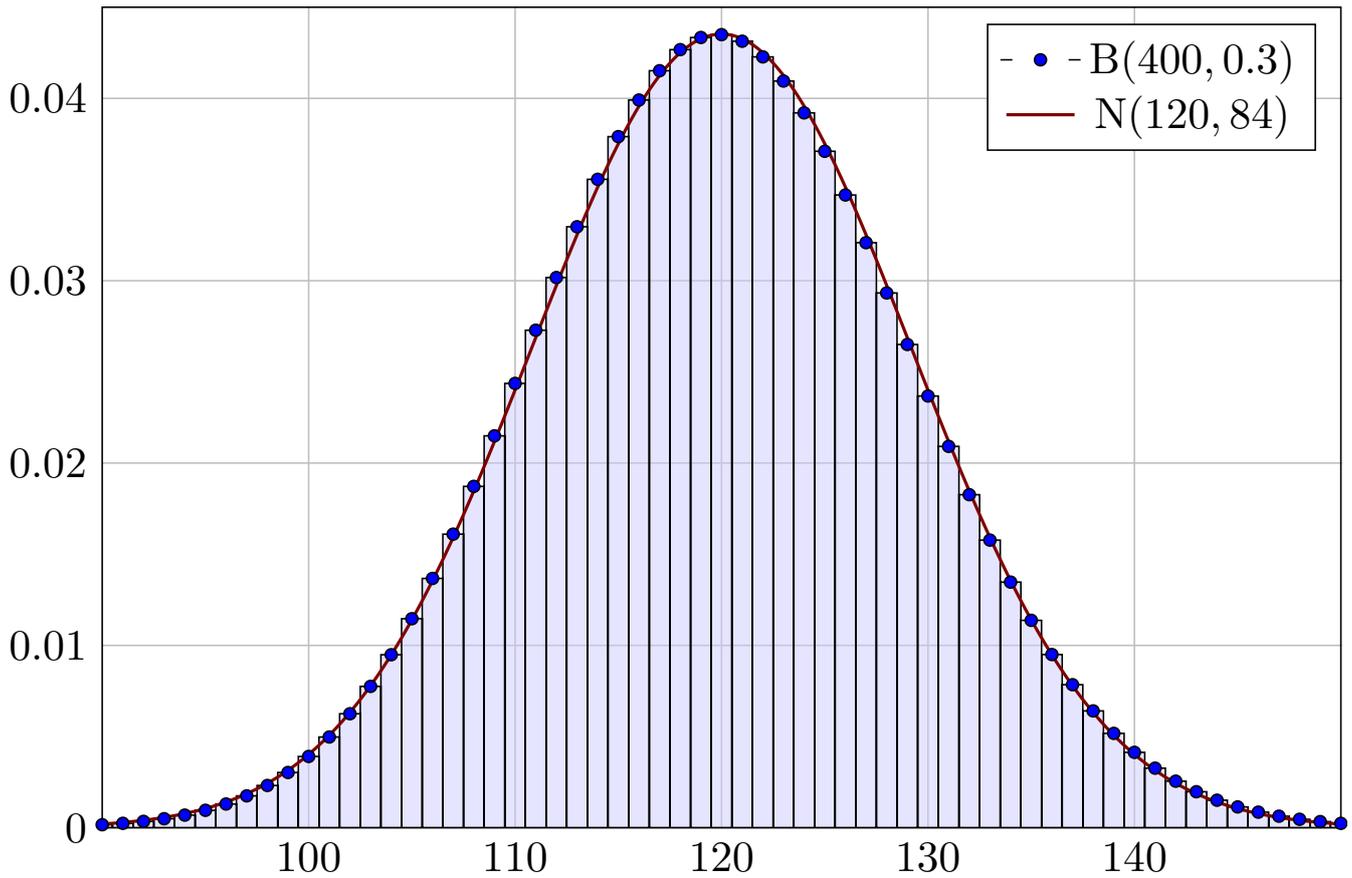
**Beispiel:**  $n = 400$  und  $t = 0.2$ , also  $\mu = nt = 80$  und  $\sigma^2 = nt(1-t) = 64$



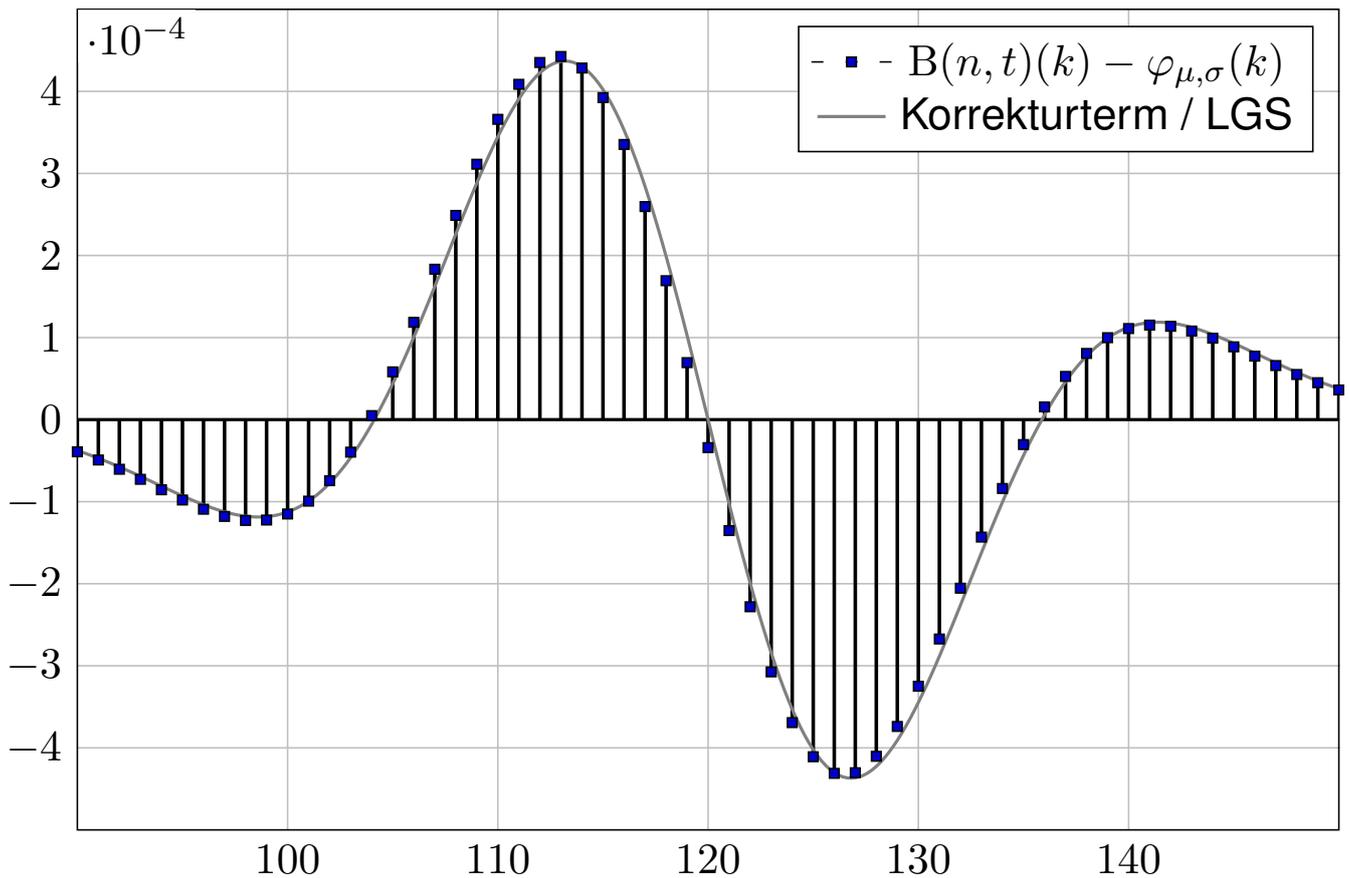
Wie gut ist die Näherung? Punktweise bis auf 0.0009 genau!



**Beispiel:**  $n = 400$  und  $t = 0.3$ , also  $\mu = nt = 120$  und  $\sigma^2 = nt(1-t) = 84$



Wie gut ist die Näherung? Punktweise bis auf 0.0005 genau!



😊 Obige Beispiele illustrieren die Annäherung der Binomial- an die Normalverteilung. Wir können sie nun als Grenzwert nachrechnen. Dank Stirling-Formel und Taylor-Entwicklung gilt (siehe unten)

$$\binom{n}{k} t^k (1-t)^{n-k} \approx \frac{e^{-\frac{(k-\mu)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}} \approx \int_{k-1/2}^{k+1/2} \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}} dx.$$

Die Stetigkeitskorrektur um  $-1/2$  bzw.  $+1/2$  ist hier sehr anschaulich. Mühsame Summen ersetzen wir so durch bequemere Integrale:

$$\sum_{k=a}^b \binom{n}{k} t^k (1-t)^{n-k} \approx \int_{x=a-1/2}^{b+1/2} \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}} dx = \int_{\xi=\alpha}^{\beta} \frac{e^{-\xi^2/2}}{\sqrt{2\pi}} d\xi$$

Die Substitution  $\xi = (x - \mu)/\sigma$  führt zur Standard-Normalverteilung.

😊 Wir müssen daher nur in einer einzigen Tabelle nachschlagen! Die Grenzen sind dabei  $\alpha = (a - 1/2 - \mu)/\sigma$  und  $\beta = (b + 1/2 - \mu)/\sigma$ . Die Stetigkeitskorrektur um  $-1/2$  bzw.  $+1/2$  ist auch hier anschaulich.

# Übergang von Binomial- zu Normalverteilung

Für  $n \in \mathbb{N}$  und  $0 < t < 1$  sei  $\mu = nt$  und  $\sigma = \sqrt{nt(1-t)}$ . Sei  $k = k_n$  ein Folge in  $\mathbb{N}$ . Für  $n \rightarrow \infty$  gelte  $(k - \mu)/\sigma \rightarrow \xi$ , somit insbesondere  $k/n \rightarrow t$ , jeweils gelesen als  $k = k_n$ . Dank Stirling-Formel  $n! \approx \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$  gilt dann:

$$\binom{n}{k} t^k (1-t)^{n-k} \approx \sqrt{\frac{n}{2\pi k(n-k)}} \cdot \left(\frac{nt}{k}\right)^k \left(\frac{n(1-t)}{n-k}\right)^{n-k} \rightarrow \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\xi^2/2}$$

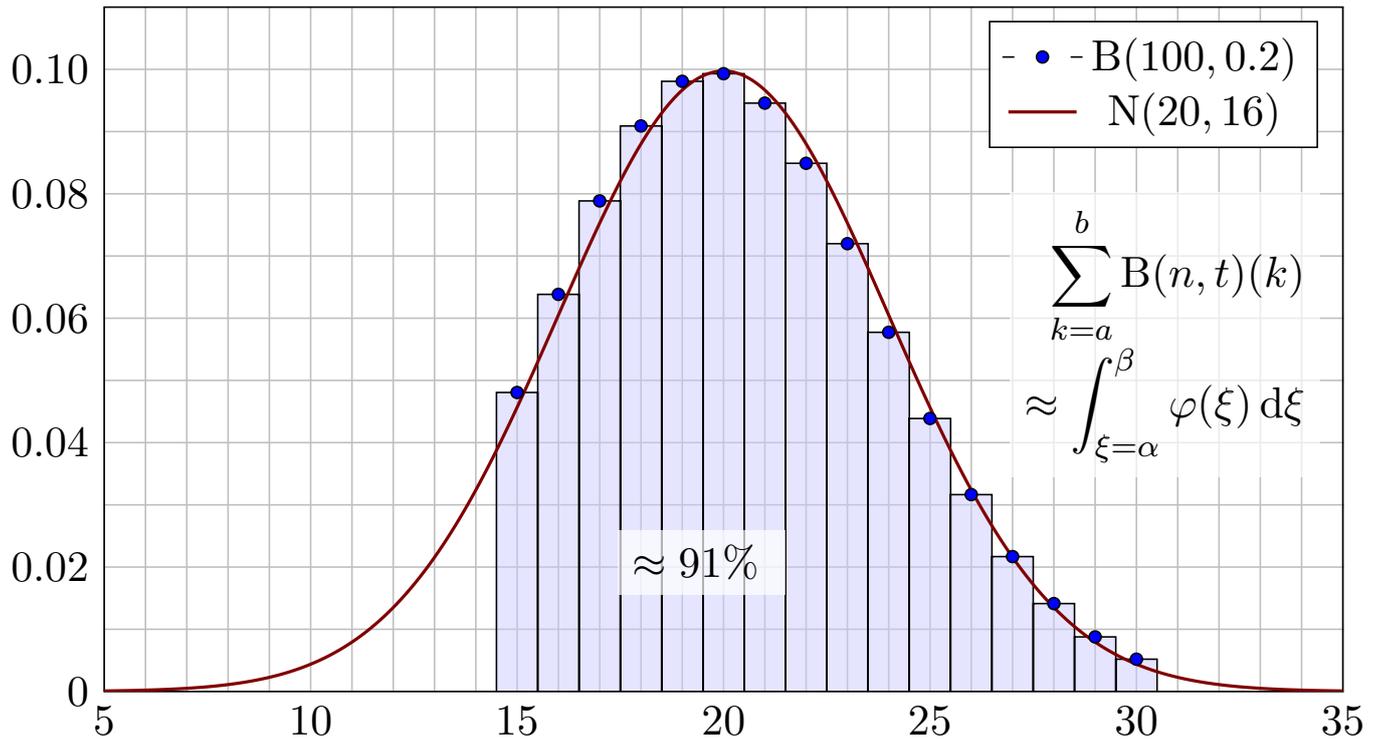
Zur Berechnung dieses Grenzwertes nutzen wir die Potenzreihe des natürlichen Logarithmus:

$$\ln(1+z) = z - \frac{z^2}{2} + \frac{z^3}{3} - \frac{z^4}{4} + \dots \quad \text{für } |z| < 1$$

Bis Grad 2 schreiben wir kurz  $\ln(1+z) \approx z - z^2/2$ . Hieraus erhalten wir

$$\begin{aligned} \left(\frac{nt}{k}\right)^k \left(\frac{n(1-t)}{n-k}\right)^{n-k} &= \left(1 + \frac{nt-k}{k}\right)^k \left(1 - \frac{nt-k}{n-k}\right)^{n-k} \\ &= \exp\left[k \ln\left(1 + \frac{nt-k}{k}\right) + (n-k) \ln\left(1 - \frac{nt-k}{n-k}\right)\right] \\ &\approx \exp\left[(nt-k) - \frac{(nt-k)^2}{2k} - (n-k) - \frac{(nt-k)^2}{2(n-k)}\right] \\ &= \exp\left[-\frac{(nt-k)^2 n}{2k(n-k)}\right] = \exp\left[-\frac{(k-nt)^2}{2n(k/n)(1-k/n)}\right] \rightarrow e^{-\xi^2/2} \end{aligned}$$

Wir **approximieren** mühsame Summen durch bequemere Integrale:



Beispiel für  $B(100, 0.2)$  mit Summationsgrenzen  $a = 15$  und  $b = 30$ , Integration von 14.5 bis 30.5, normiert also  $\alpha = -5.5/4 = -1.375$  und  $\beta = 10.5/4 = 2.625$ . Das Integral  $\int_{\alpha}^{\beta} \varphi(\xi) d\xi$  entnehmen wir der Tabelle.

## Übergang von Binomial- zu Normalverteilung

Wozu nützt die **Näherung** der Binomial- durch die Normalverteilung? Die Binomialverteilung  $B(n, t)$  ist für große  $n$  mühsam zu summieren. Die Normalverteilung hingegen ist leicht, und  $\Phi$  liegt tabelliert vor. V116

$$\sum_{k=a}^b \binom{n}{k} t^k (1-t)^{n-k} \approx \int_{\xi=\alpha}^{\beta} \frac{e^{-\xi^2/2}}{\sqrt{2\pi}} d\xi = [\Phi]_{\alpha}^{\beta}$$

Hierzu benötigen wir nur Erwartung  $\mu = nt$  und Varianz  $\sigma^2 = nt(1-t)$  und berechnen damit  $\alpha = (a - 1/2 - \mu)/\sigma$  und  $\beta = (b + 1/2 - \mu)/\sigma$ .

Der natürlich auftretende Summand  $\mp 1/2$  heißt **Stetigkeitskorrektur**. Er erklärt sich aus obiger Graphik. Denken Sie insbesondere an  $a = b!$

Die Binomialverteilung  $B(n, t)$  können wir somit durch die einfachere Normalverteilung approximieren. Wann ist diese Näherung gut genug? Die Faustregel ist „ $n$  soll groß sein“ oder „ $\sigma^2 = nt(1-t)$  soll groß sein“. Naja... Wie groß ist groß genug? Wie klein ist dann noch der Fehler?

Die genaue Analyse des Fehlers ist leider mühsam. Ich gebe hier ein praktisches Ergebnis von James Uspensky (1937) in vereinfachter Form.

Satz V3A: de Moivre 1733, Laplace 1812, Uspensky 1937

Für die Binomialverteilung  $B(n, t)$  gilt  $\mu = nt$  und  $\sigma = \sqrt{nt(1-t)}$ .  
 Sie ähnelt der Normalverteilung:  $B(n, t) \approx N(\mu, \sigma^2)$ . Genauer gilt:

$$\sum_{k=a}^b \binom{n}{k} t^k (1-t)^{n-k} = \int_{\xi=\alpha}^{\beta} \frac{e^{-\xi^2/2}}{\sqrt{2\pi}} d\xi + \delta$$

Die Grenzen sind hier  $\alpha = (a - 1/2 - \mu)/\sigma$  und  $\beta = (b + 1/2 - \mu)/\sigma$ .

- (1) Für  $\sigma \geq 5$  ist der Fehler  $\delta$  beschränkt durch  $|\delta| < \frac{|1-2t|}{10\sigma} + \frac{1}{3\sigma^2} \leq \frac{1}{6\sigma}$ .
- (2) Genauer gilt  $\delta = \kappa + \varepsilon$  mit Korrekturterm  $\kappa$  und restlichem Fehler  $\varepsilon$ :

$$\kappa = \frac{1-2t}{6\sigma\sqrt{2\pi}} \left[ (1-\xi^2) e^{-\xi^2/2} \right]_{\alpha}^{\beta}, \quad |\kappa| < \frac{|1-2t|}{10\sigma}, \quad |\varepsilon| < \frac{1}{3\sigma^2}.$$

$\sigma \geq 5$	$\sigma \geq 8$	$\sigma \geq 13$	$\sigma \geq 19$	$\sigma \geq 26$
$ \delta  < 0.034$	$ \delta  < 0.0178$	$ \delta  < 0.0097$	$ \delta  < 0.0062$	$ \delta  < 0.0043$
$ \varepsilon  < 0.014$	$ \varepsilon  < 0.0053$	$ \varepsilon  < 0.0020$	$ \varepsilon  < 0.0010$	$ \varepsilon  < 0.0005$

Der lokale Grenzwertsatz (LGS)

Für  $n \rightarrow \infty$  gilt  $\sigma = \sqrt{nt(1-t)} \rightarrow \infty$ , also  $\delta \rightarrow 0$ , und somit

$$\sum_{k=\lceil \mu+\alpha\sigma \rceil}^{\lfloor \mu+\beta\sigma \rfloor} \binom{n}{k} t^k (1-t)^{n-k} \rightarrow \int_{\xi=\alpha}^{\beta} \frac{e^{-\xi^2/2}}{\sqrt{2\pi}} d\xi.$$

Meist wird diese schwächere Aussage lokaler Grenzwertsatz genannt. Außer  $\delta \rightarrow 0$  macht sie keine quantitative Aussage über den Fehler.

In der Praxis rechnen wir nicht für  $n \rightarrow \infty$ , sondern mit festem  $n \in \mathbb{N}$ . Für praktische Rechnungen ist daher die Fehlerkontrolle wesentlich:

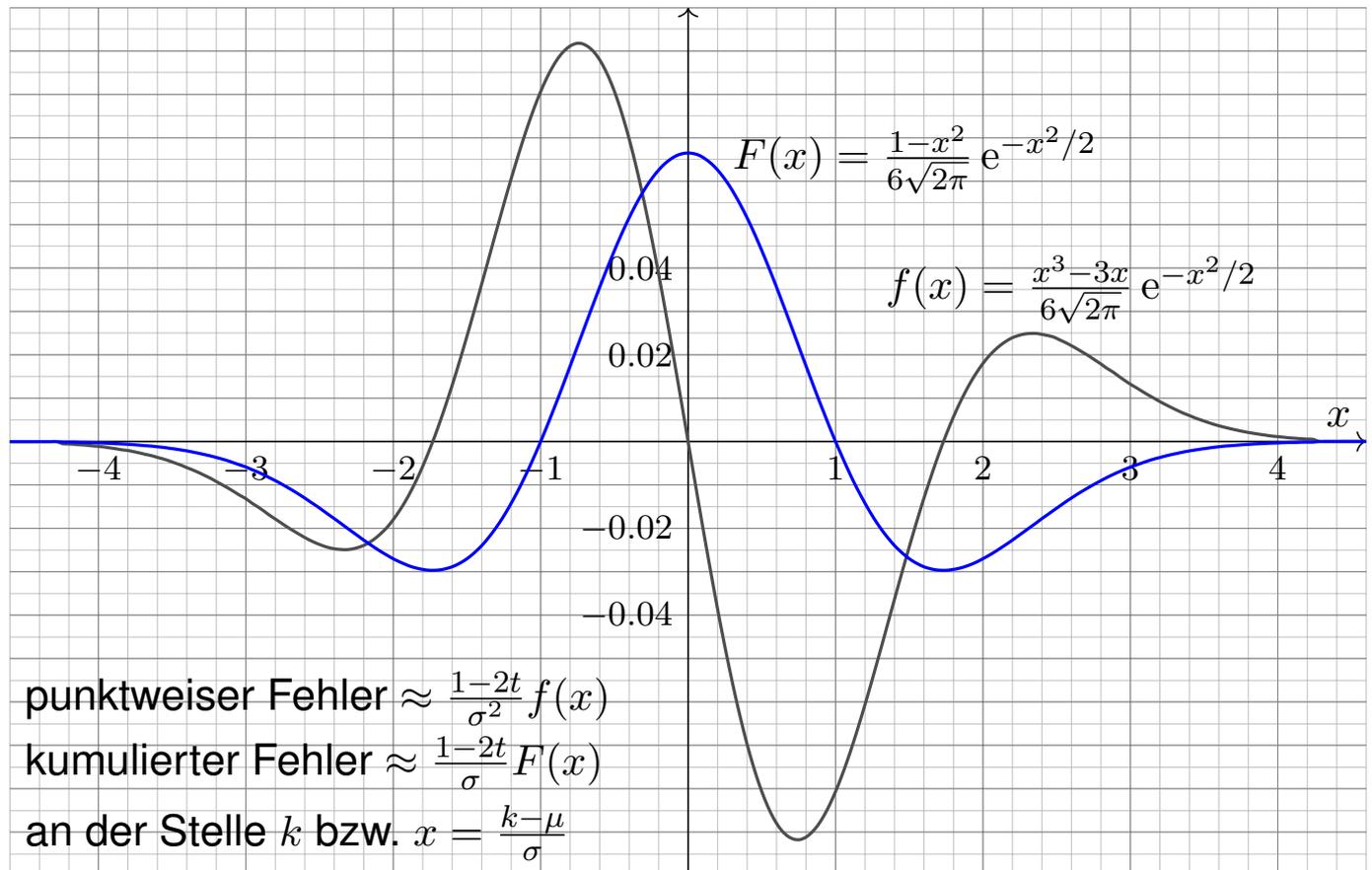
$$|\delta| < \frac{1}{6\sigma} = \frac{1}{6\sqrt{nt(1-t)}} \searrow 0$$

Wenn wir zum Integral den Korrekturterm  $\kappa$  hinzuaddieren, so wird der Approximationsfehler noch um eine Größenordnung kleiner, nämlich:

$$|\varepsilon| < \frac{0.15 + 0.18|1-2t|}{\sigma^2} < \frac{1}{3\sigma^2} = \frac{1}{3nt(1-t)}$$

😊 Für  $t = 1/2$  oder  $\alpha = -\beta$  gilt  $\kappa = 0$  und somit  $|\delta| < 1/(3\sigma^2)$ . In diesen Fällen ist die Näherung durch die Normalverteilung besonders genau.

Form des Korrekturterms  $\kappa(\alpha, \beta) = \frac{1-2t}{\sigma} [F(\beta) - F(\alpha)]$  wie illustriert:



😊 Gilt der Fehler als klein genug, so können wir die mühsame Summe der Binomialverteilung (links) durch das wesentlich bequemere Integral der Normalverteilung (rechts) ersetzen: Dieses liegt tabelliert vor. V116

Die Gleichung des Satzes (mit Fehler  $\delta$ ) können wir vergrößern zu:

$$\sum_{k=a}^b \binom{n}{k} t^k (1-t)^{n-k} \approx \int_{\xi=\alpha}^{\beta} \frac{e^{-\xi^2/2}}{\sqrt{2\pi}} d\xi = [\Phi]_{\alpha}^{\beta}$$

Das ist eine nützliche Faustformel, leider ohne jegliche Fehlerkontrolle. In der Praxis vernachlässigt man meist den Approximationsfehler  $\delta$  und beruft sich erfahrungsgemäß darauf, dass  $\sigma$  schon „groß genug“ sei.

😊 Möchte man ein möglichst genaues Ergebnis mit garantierter Fehler-schranke, so nimmt man die Grenzen mit der Stetigkeitskorrektur  $\mp 1/2$ . Statt der Grenzen  $\alpha = (a - 1/2 - \mu)/\sigma$  und  $\beta = (b + 1/2 - \mu)/\sigma$  nutzt man auch gerne  $(a - \mu)/\sigma$  und  $(b - \mu)/\sigma$ . Das ist bequem, aber ungenauer. Meist ist diese Vereinfachung gut genug, insbesondere bei großem  $\sigma$ . Die größeren Grenzen nutzen wir im zentralen Grenzwertsatzes W1D.

**Aufgabe:** Ein Experiment mit Trefferwkt 0.2 wird 400 mal unabhängig wiederholt. Mit welcher Wkt ergibt sich eine Trefferzahl von 60 bis 90? (Ergebnis in Prozent, gerundet zum nächsten Prozentpunkt)

**Lösung:** Die Trefferzahl ist  $B(n, t)$ -verteilt mit  $n = 400$  und  $t = 0.2$ . Dies ist die exakte Verteilung, die wir nun numerisch nähern wollen. Erwartung  $\mu = nt = 80$ , Varianz  $\sigma^2 = nt(1 - t) = 64$ , Streuung  $\sigma = 8$ .

**Grob:** Trefferzahl 60 bis 90 entspricht  $\mu - 2.5\sigma$  bis  $\mu + 1.25\sigma$ . Also

$$p = \sum_{k=60}^{90} \binom{400}{k} \left(\frac{1}{5}\right)^k \left(\frac{4}{5}\right)^{400-k} \approx \int_{-2.5}^{1.25} \varphi(t) dt.$$

Dank der Symmetrie  $\varphi(-t) = \varphi(t)$  gilt für das Integral

$$p \approx \int_{-2.5}^{1.25} \varphi(t) dt = \int_{t=0}^{2.5} \varphi(t) dt + \int_{t=0}^{1.25} \varphi(t) dt.$$

Die Tabelle liefert  $p \approx 0.49379 + 0.39435 = 0.88814$ , also  $p \approx 89\%$ .

Was ist hier schludrig? Schlechte Grenzen und keine Fehlerschranke!

**Fein:** Genauer ist die im Satz angegebene Näherung  $p \approx \int_{\alpha}^{\beta} \varphi(t) dt$  mit  $\alpha = (60 - 0.5 - 80)/8 \approx -2.56$  und  $\beta = (90 + 0.5 - 80)/8 \approx 1.31$ .

$$p \approx \int_{-2.56}^{1.31} \varphi(t) dt = \int_{t=0}^{2.56} \varphi(t) dt + \int_{t=0}^{1.31} \varphi(t) dt.$$

Die Tabelle liefert  $p \approx 0.49477 + 0.40490 = 0.89967$ , also  $p \approx 90\%$ .

**Güte:** Wie genau ist diese Approximation? Laut Satz gilt:

$$|\delta| < \frac{1}{6\sigma} = \frac{1}{48} \approx 0.02$$

Diese Schranke ist noch zu grob. Wir schauen daher genauer hin:

$$\kappa \approx 0.00047 \quad \text{und} \quad |\varepsilon| < \frac{1}{3\sigma^2} = \frac{1}{192} \approx 0.00521$$

Unsere Approximation  $p \approx 0.89967$  hat also einen Fehler  $|\delta| < 0.00568$ . Damit erfüllt der Wert  $p \approx 90\%$  die geforderte Genauigkeit 0.01.

**Exakt:** Mühsames Aufsummieren ergibt  $p = 0.89982\dots$ , also  $p \approx 90\%$ .

**Aufgabe:** Ein Experiment mit Trefferwkt 1% wird 3600 mal unabhängig wiederholt. Mit welcher Wkt ergibt sich eine Trefferzahl von 36 bis 46? (Ergebnis in Prozent, gerundet zum nächsten Prozentpunkt)

**Lösung:** Die Trefferzahl ist  $B(n, t)$ -verteilt mit  $n = 3600$  und  $t = 0.01$ . Dies ist die exakte Verteilung, die wir nun numerisch nähern wollen. Wir finden Erwartung  $\mu = nt = 36$ , Varianz  $\sigma^2 = nt(1 - t) = 35.64$ , Streuung  $\sigma \approx 5.97$ . Gesucht ist die Wahrscheinlichkeit

$$p = \sum_{k=36}^{46} \binom{3600}{k} \left(\frac{1}{100}\right)^k \left(\frac{99}{100}\right)^{3600-k} \approx \int_{t=\alpha}^{\beta} \varphi(t) dt$$

mit  $\alpha = -0.5/5.97 \approx -0.084$  und  $\beta = 10.5/5.97 \approx 1.759$ . Demnach gilt

$$p \approx \int_{-0.084}^{1.759} \varphi(t) dt = \int_{t=0}^{0.084} \varphi(t) dt + \int_{t=0}^{1.759} \varphi(t) dt.$$

Für die dritte Dezimalstelle interpolieren wir zwischen den Nachbarn. Interpolation der Tabelle liefert  $p \approx 0.03347 + 0.46071 = 0.49418$ .

**Güte:** Laut Satz ist der Approximationsfehler  $|\delta| < 1/(6\sigma) \lesssim 0.028$ . Damit können wir die gewünschte Genauigkeit noch nicht garantieren. Der Korrekturterm ist  $\kappa \approx -0.01567$ , also korrigieren wir zu

$$p \approx 0.49418 - 0.01567 = 0.47851$$

Dann bleibt nur noch  $|\varepsilon| < 1/(3\sigma^2) \lesssim 0.00936$ . Wir schließen  $p \approx 48\%$ .

In diesem Beispiel ist die Streuung  $\sigma \approx 6$  sehr klein, und die Näherungen des lokalen Grenzwertsatzes daher entsprechend grob. Auch sind sowohl die Trefferwahrscheinlichkeit  $t = 0.01$  als auch die Integrationsgrenzen  $\alpha$  und  $\beta$  besonders unsymmetrisch, und so ist der Korrekturterm  $\kappa$  hier recht groß. Die Addition von  $\kappa$  bringt eine spürbare Verbesserung:

Die erste Näherung  $p \approx 0.49418$  ist deutlich schlechter als die um  $\kappa$  korrigierte Näherung  $p \approx 0.47851$ . Nur letztere erfüllt die geforderte Genauigkeit von 0.01. Auch gerundet hätten wir zunächst 49% erhalten, aber die tatsächliche Wahrscheinlichkeit liegt knapp unter 48%:

**Exakt:** Mühsames Aufsummieren ergibt  $p = 0.47880\dots$ , also  $p \approx 48\%$ .

😊 Egal ob grob oder genau, der Satz stellt alle Werkzeuge bereit. Lesen Sie den Satz genau, er ist wunderbar explizit und hilfreich.

**Aufgabe:** Wie wahrscheinlich ist es, dass bei 600maligem Würfeln die Augenzahl 6 zwischen einschließlich 90 und 110 mal eintritt? Bestimmen Sie diese Wahrscheinlichkeit auf 0.005 genau.

**Lösung:** Die Anzahl ist  $B(n, t)$ -verteilt mit  $n = 600$  und  $t = 1/6$ . Dies ist die exakte Verteilung, die wir nun numerisch nähern wollen. Wir finden Erwartung  $\mu = nt = 100$ , Varianz  $\sigma^2 = nt(1 - t) = 500/6$ , Streuung  $\sigma = 10\sqrt{5/6} \approx 9.13$ . Gesucht ist die Wahrscheinlichkeit

$$p = \sum_{k=90}^{110} \binom{600}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{600-k} \approx \int_{t=\alpha}^{\beta} \varphi(t) dt$$

mit  $\alpha = (90 - 0.5 - 100)/9.13 \approx -1.15$  und  $\beta \approx 1.15$ . Demnach gilt

$$p \approx \int_{-1.15}^{1.15} \varphi(t) dt = 2 \int_{t=0}^{1.15} \varphi(t) dt \approx 2 \cdot 0.37493 = 0.74986.$$

Also  $p \approx 75\%$ . (Plausibel, denn etwas mehr als die  $1\sigma$ -Umgebung.)

**Güte:** Wie genau ist die Approximation in diesem Fall?

Dank der Symmetrie von  $\alpha$  und  $\beta$  verschwindet der Korrekturterm. Wegen  $\kappa = 0$  bleibt hier nur ein besonders kleiner Fehler von

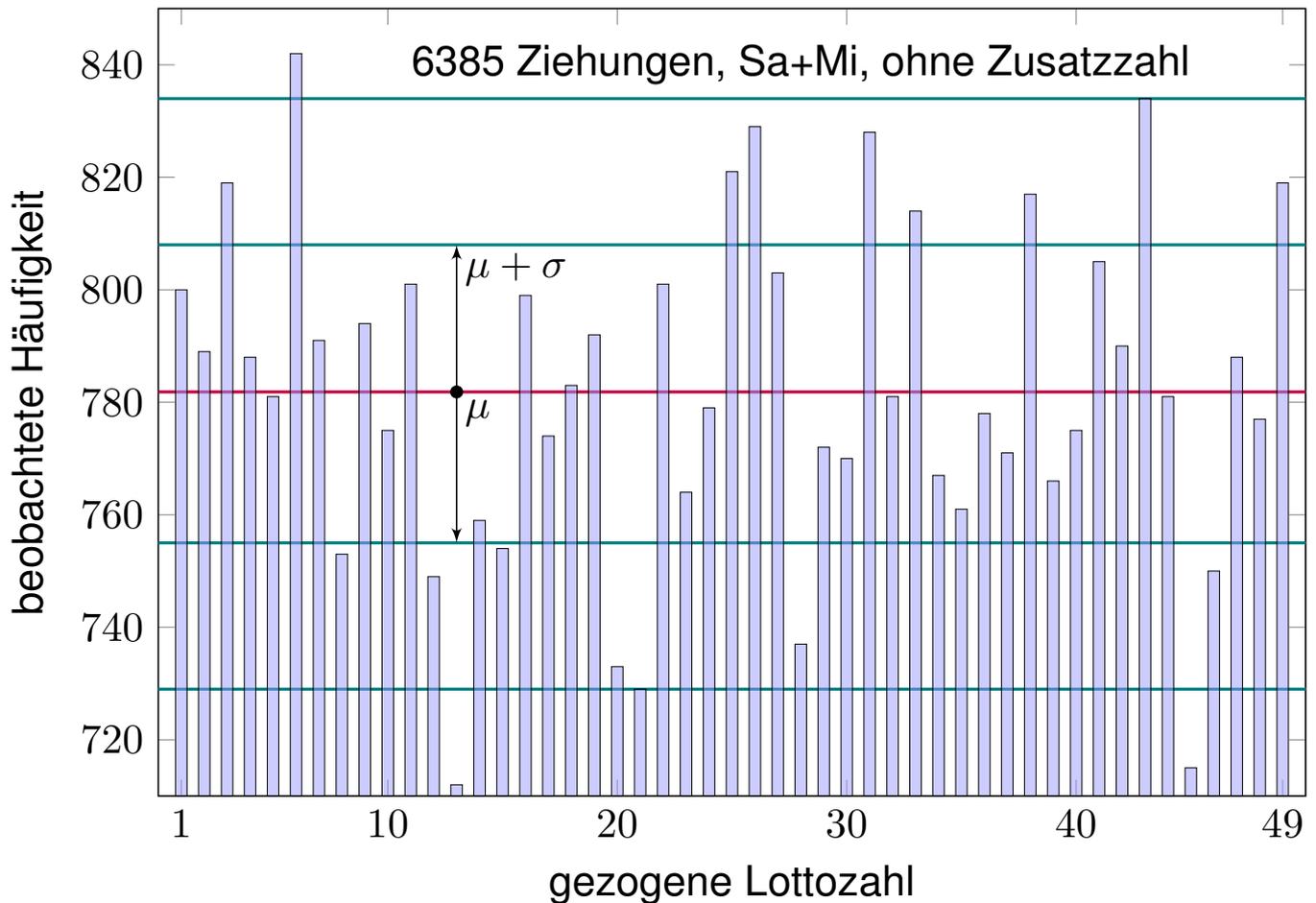
$$|\delta| = |\varepsilon| < \frac{1}{3\sigma^2} = \frac{2}{500} = 0.004.$$

Somit ist unsere Approximation  $p \approx 75\%$  auf 0.005 genau.

**Exakt:** Mühsames Aufsummieren ergibt  $p = 0.75012\dots$

Wir sehen in diesen Beispielen, dass die Approximation meist noch besser ist, als die allgemeine Fehlerabschätzung des Satzes vermuten lässt. Das ist auch verständlich: Es handelt sich nur um eine *obere Schranke* für den Fehler, aber sie ist doch recht nah an der wirklichen Abweichung. Die exakten Werte liefern immer einen Fehler unter dieser Schranke, manchmal deutlich darunter.

Der Nutzen des Satzes ist über die Konvergenz für  $n \rightarrow \infty$  hinaus vor allem die bequeme und aussagekräftige Fehlerschranke für endliches  $n$ . Damit können Sie schnell und sicher abschätzen, ob die gewünschte Genauigkeit gewährleistet ist. Ist eine höhere Genauigkeit nötig, so müssen Sie beim Fehler genauer hinschauen und gegebenenfalls eine präzisere Rechnung nutzen, notfalls aufwändig  $B(n, t)$  aufsummieren. Alle hierzu nötigen Informationen liegen explizit vor.



## Sind die Lotto-Zahlen verdächtig?

Die Graphik zeigt die Statistik über 6385 Lotto-Ziehungen 6 aus 49.

**Aufgabe:** Ist die geringe Ziehungshäufigkeit der Zahl 13 verdächtig?

**Lösung:** Wir untersuchen  $n = 6385$  Ziehungen von je 6 Zahlen.

Die Ziehungswahrscheinlichkeit jeder Zahl ist  $t = \binom{48}{5} / \binom{49}{6} = 6/49$ .

Bei ordnungsgemäßer Ziehung ist jede Anzahl  $B(n, t)$ -verteilt.

Erwartung  $\mu = nt \approx 781.84$

Varianz  $\sigma^2 = nt(1 - t) \approx 686.10$

Streuung  $\sigma \approx 26.19$

Die  $1\sigma$ -Umgebung  $[755, 808]$  enthält 31 der 49 Anzahlen, also 63%.

Die  $2\sigma$ -Umgebung  $[729, 834]$  enthält 46 der 49 Anzahlen, also 94%.

Die  $3\sigma$ -Umgebung  $[703, 860]$  enthält 49 der 49 Anzahlen, also 100%.

😊 Diese Verteilung entspricht der erwarteten 68–95–99–Regel.

Schwankungen dieser Größenordnung sind normal. Im Gegenteil, weniger wäre verdächtig. Also doch keine Lotto-Verschwörung.

Titanic (26.10.2011): „Experten warnen: Lottozahlen oft reiner Zufall!“

Mit Kolmogorovs Axiomen (V1c) fixieren wir die grundlegenden Forderungen für WMaße, woraus alle weiteren Rechenregeln folgen:

Ein **Wahrscheinlichkeitsraum**  $(\Omega, \mathcal{A}, \mathbf{P})$  besteht aus

- einer **Ergebnismenge**  $\Omega$ ,
- einer **Ereignismenge**  $\mathcal{A} \subseteq \mathfrak{P}(\Omega)$  und
- einem **Wahrscheinlichkeitsmaß**  $\mathbf{P} : \mathcal{A} \rightarrow [0, 1]$ ,

mit folgenden grundlegenden Eigenschaften:

- 1 Normierung:** Es gilt  $\Omega \in \mathcal{A}$  und  $\mathbf{P}(\Omega) = 1$ .
- 2 Komplemente:** Aus  $A \in \mathcal{A}$  folgt  $(\Omega \setminus A) \in \mathcal{A}$ .
- 3  $\sigma$ -Additivität:** Aus  $A_0, A_1, A_2, \dots \in \mathcal{A}$  folgt  $\bigcup_{k=0}^{\infty} A_k \in \mathcal{A}$ , sowie

$$\mathbf{P}\left(\bigsqcup_{k=0}^{\infty} A_k\right) = \sum_{k=0}^{\infty} \mathbf{P}(A_k) \quad \text{wenn } A_i \cap A_j = \emptyset \text{ für } i \neq j.$$

Die Bezeichnung „ $\sigma$ “ steht abkürzend für „abzählbar“. Wir nennen  $\mathcal{A}$  eine  **$\sigma$ -Algebra** und hierauf  $\mathbf{P} : \mathcal{A} \rightarrow [0, 1]$  ein  **$\sigma$ -additives Maß**.

Wichtigste Beispiele sind diskrete und kontinuierliche WRäume. Auch Mischungen aus diskret und kontinuierlich sind möglich.

## Diskrete und kontinuierliche WRäume

Eine **diskrete Wahrscheinlichkeitsverteilung** auf  $\Omega$  ist eine Funktion

$$p : \Omega \rightarrow [0, 1] \quad \text{mit Gesamtmasse} \quad \sum_{x \in \Omega} p(x) = 1.$$

Diese definiert ein **diskretes Wahrscheinlichkeitsmaß**

$$\mathbf{P} : \mathfrak{P}(\Omega) \rightarrow [0, 1] \quad \text{durch} \quad \mathbf{P}(A) := \sum_{x \in A} p(x).$$

Hierbei ist  $\mathfrak{P}(\Omega)$  die Familie aller Teilmengen  $A \subseteq \Omega$ .

Eine **kontinuierliche Wahrscheinlichkeitsverteilung** oder **WDichte** auf einer Ergebnismenge  $\Omega \subseteq \mathbb{R}^n$  ist eine messbare Funktion

$$f : \Omega \rightarrow \mathbb{R}_{\geq 0} \quad \text{mit Gesamtmasse} \quad \int_{\Omega} f(x) dx = 1.$$

Diese definiert ein **kontinuierliches Wahrscheinlichkeitsmaß**

$$\mathbf{P} : \mathcal{B}(\Omega) \rightarrow [0, 1] \quad \text{durch} \quad \mathbf{P}(A) := \int_A f(x) dx.$$

Hierbei ist  $\mathcal{B}(\Omega)$  die Familie aller messbaren Teilmengen  $A \subseteq \Omega$ .

Sei  $\mathbf{P}$  ein WMaß auf  $\Omega \subseteq \mathbb{R}^n$ , zum Beispiel diskret oder kontinuierlich. Der Schwerpunkt der Verteilung  $\mathbf{P}$  heißt **Mittelwert** oder **Erwartung**:

$$\mu := \sum_{x \in \Omega} x p(x) \quad \text{bzw.} \quad \mu := \int_{x \in \Omega} x f(x) dx.$$

Hierbei setzen wir absolute Summierbarkeit / Integrierbarkeit voraus. Die **Varianz** der Verteilung ist ihr Trägheitsmoment bezüglich  $\mu$ :

$$\sigma^2 := \sum_{x \in \Omega} (x - \mu)^2 p(x) = \left[ \sum_{x \in \Omega} x^2 p(x) \right] - \mu^2 \geq 0 \quad \text{bzw.}$$

$$\sigma^2 := \int_{x \in \Omega} (x - \mu)^2 f(x) dx = \int_{x \in \Omega} x^2 f(x) dx - \mu^2 \geq 0.$$

Die Wurzel  $\sigma \geq 0$  dieses Wertes heißt die **Streuung** der Verteilung. Anschaulich misst  $\sigma$ , wie weit die Werte um den Mittelwert  $\mu$  streuen. Man nennt die Streuung  $\sigma$  daher auch die **Standardabweichung**. Genau dann gilt  $\sigma = 0$ , wenn  $\mathbf{P}$  auf den Punkt  $\mu$  konzentriert ist, d.h. für  $A \subseteq \mathbb{R}$  gilt  $\mathbf{P}(A) = 1$  falls  $\mu \in A$  und  $\mathbf{P}(A) = 0$  falls  $\mu \notin A$ .

## Existenz von Erwartung und Varianz

**Aufgabe:** Welche der folgenden Funktionen  $f: \mathbb{R} \rightarrow \mathbb{R}$  sind WDichten?

$$f(x) = \frac{1/c}{1 + |x|},$$

$$f(x) = \frac{1/c}{1 + |x|^2},$$

$$f(x) = \frac{1/c}{1 + |x|^3},$$

$$f(x) = \frac{1/c}{1 + |x|^\alpha},$$

$$g(x) = \frac{1/c}{1 + |x|^\alpha \ln(e + |x|)},$$

$$h(x) = \frac{1/c}{1 + |x|^\alpha \ln(e + |x|)^2},$$

mit geeignetem  $c \in \mathbb{R}$ . Für welche existieren Erwartung und Varianz?

**Lösung:** Wir haben jeweils  $f \geq 0$ . Die erste lässt sich nicht normieren:

$$\int_{\mathbb{R}} \frac{1}{1 + |x|^\alpha} dx = \begin{cases} +\infty & \text{für } \alpha \leq 1, \\ < \infty & \text{für } \alpha > 1. \end{cases}$$

Für  $\alpha = 2$  erhalten wir die Cauchy-Verteilung  $f(x) = \frac{1/\pi}{1+x^2}$ . V241

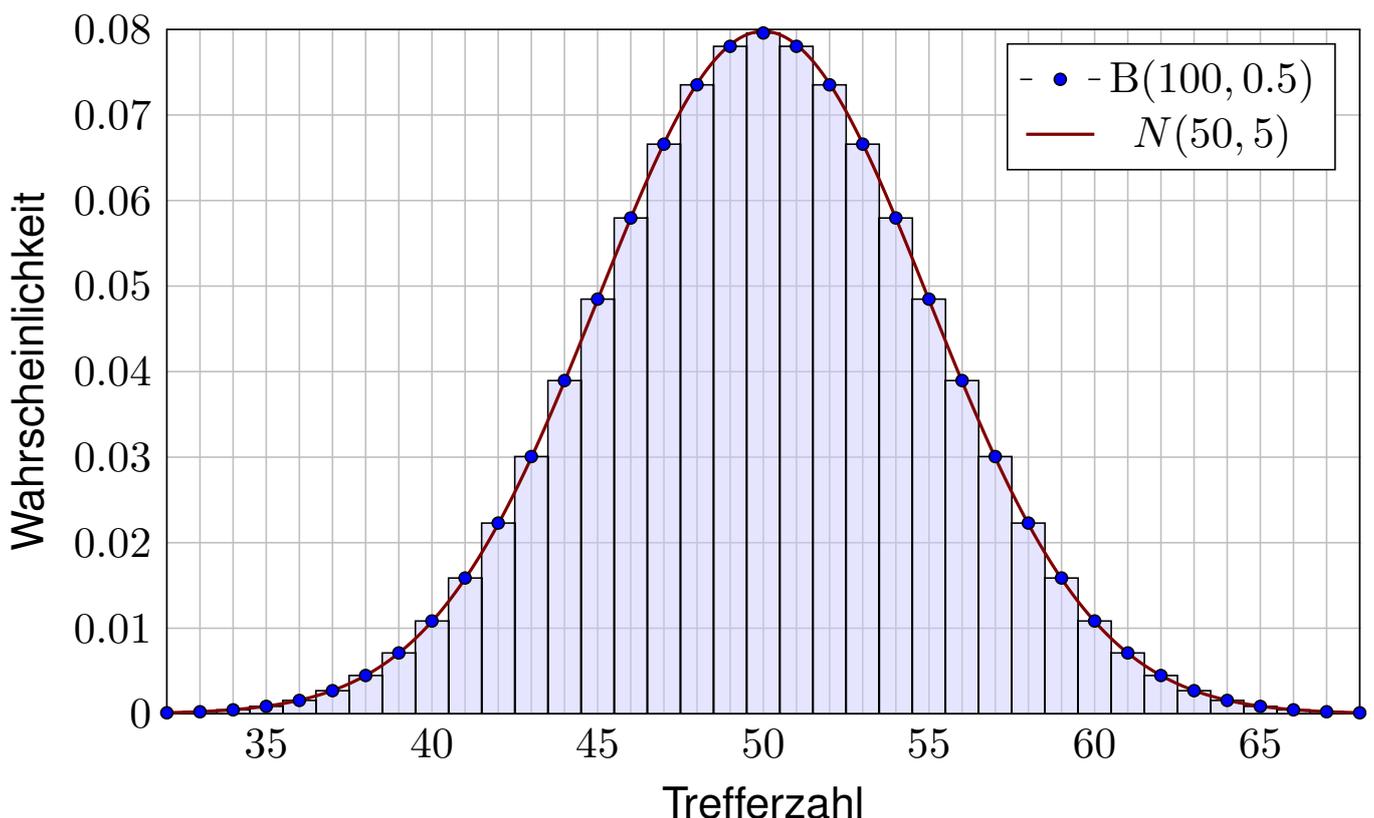
Das Integral konvergiert für  $\alpha > 1$ . Für die Erwartung braucht's  $\alpha > 2$ .

Für die Varianz braucht's  $\alpha > 3$ . Für  $h$  genügt jeweils schon  $\alpha \geq 1, 2, 3$ .

<b>diskret</b>		$\Omega$	$p(k) = \mathbf{P}(\{k\})$	<b>E</b>	<b>V</b>
Gleichverteilung	V233	$\{a, \dots, b\}$	$\frac{1}{b-a+1}$	$\frac{a+b}{2}$	$\frac{(b-a+1)^2-1}{12}$
Hyperg. $H(N, K, n)$	U419	$\{0, \dots, n\}$	$\binom{K}{k} \binom{N-K}{n-k} / \binom{N}{n}$	$n \frac{K}{N}$	$n \frac{K}{N} \frac{N-K}{N} \frac{N-n}{N-1}$
Binomial $B(n, t)$	U421	$\{0, \dots, n\}$	$\binom{n}{k} t^k (1-t)^{n-k}$	$nt$	$nt(1-t)$
Poisson $P(\lambda)$	U425	$\mathbb{N}$	$e^{-\lambda} \lambda^k / k!$	$\lambda$	$\lambda$
Geometrisch $G(q)$	T421	$\mathbb{N}_{\geq 1}$	$(1-q)q^{k-1}$	$\frac{1}{1-q}$	$\frac{q}{(1-q)^2}$
Zeta $Z(s), s = 2$	T312	$\mathbb{N}_{\geq 1}$	$\frac{6}{\pi^2} \frac{1}{k^2}$	— nicht definiert! — Diese Reihen divergieren.	
<b>kontinuierlich</b>		$\Omega$	Dichte $f(x)$	<b>E</b>	<b>V</b>
Gleichverteilung	V235	$[a, b]$	$\frac{1}{b-a}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponential $E(\lambda)$	V243	$\mathbb{R}_{\geq 0}$	$\lambda e^{-\lambda x}$	$\lambda^{-1}$	$\lambda^{-2}$
Normal $N(\mu, \sigma^2)$	V239	$\mathbb{R}$	$\frac{e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}}{\sigma\sqrt{2\pi}}$	$\mu$	$\sigma^2$
Cauchy $C(m, a)$	V241	$\mathbb{R}$	$\frac{1}{\pi} \frac{a}{a^2+(x-m)^2}$	— nicht definiert! — Diese Integrale divergieren.	

## Von Binomial- zu Normalverteilungen

Grundbeobachtung: Binomialverteilungen ähneln Normalverteilungen. Der lokale Grenzwertsatz präzisiert dies und sichert Fehlerschranken.



Für die **Binomialverteilung**  $B(n, t)$  gilt  $\mu = nt$  und  $\sigma^2 = nt(1 - t)$ .  
 Sie ähnelt der **Normalverteilung**:  $B(n, t) \approx N(\mu, \sigma^2)$ , das heißt:

$$\binom{n}{k} t^k (1 - t)^{n-k} \approx \frac{e^{-\frac{(k-\mu)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}} \approx \int_{k-1/2}^{k+1/2} \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}} dx.$$

Dies nennt man auch Satz von de Moivre (1733) und Laplace (1812).  
 Mühsame Summen ersetzen wir so durch bequemere Integrale:

$$\sum_{k=a}^b \binom{n}{k} t^k (1 - t)^{n-k} = \int_{\xi=\alpha}^{\beta} \frac{e^{-\xi^2/2}}{\sqrt{2\pi}} d\xi + \delta$$

mit den Grenzen  $\alpha = (a - 1/2 - \mu)/\sigma$  und  $\beta = (b + 1/2 - \mu)/\sigma$ .

(1) Für  $\sigma \geq 5$  ist der Approximationsfehler  $\delta$  beschränkt durch

$$|\delta| < \frac{|1 - 2t|}{10\sigma} + \frac{1}{3\sigma^2} \leq \frac{1}{6\sigma} = \frac{1}{6\sqrt{nt(1 - t)}}$$

😊 Gilt der Fehler als klein genug, so können wir die mühsame Summe der Binomialverteilung (links) durch das wesentlich bequemere Integral der Normalverteilung (rechts) ersetzen: Dieses liegt tabelliert vor. V116

(2) Ist eine genauere Approximation nötig, so addiert man zur obigen Näherung durch das Integral noch folgenden Korrekturterm  $\kappa$ :

$$\sum_{k=a}^b \binom{n}{k} t^k (1 - t)^{n-k} = \int_{\xi=\alpha}^{\beta} \frac{e^{-\xi^2/2}}{\sqrt{2\pi}} d\xi + \kappa + \varepsilon$$

mit Korrektur  $\kappa = \frac{1 - 2t}{6\sigma\sqrt{2\pi}} \left[ (1 - \xi^2) e^{-\xi^2/2} \right]_{\alpha}^{\beta}$

Der verbleibende Fehler  $\varepsilon$  ist noch eine Größenordnung kleiner, nämlich

$$|\kappa| < \frac{|1 - 2t|}{10\sigma}, \quad |\varepsilon| < \frac{0.15 + 0.18|1 - 2t|}{\sigma^2} < \frac{1}{3\sigma^2} = \frac{1}{3nt(1 - t)}.$$

Speziell für  $t = 1/2$  oder  $\alpha = \beta$  verschwindet der Korrekturterm  $\kappa$ .

Für  $n \rightarrow \infty$  geht der Fehler gegen Null mit  $1/\sqrt{n} \rightarrow 0$  bzw.  $1/n \rightarrow 0$ .

$\sigma \geq 5$	$\sigma \geq 8$	$\sigma \geq 13$	$\sigma \geq 19$	$\sigma \geq 26$
$ \delta  < 0.034$	$ \delta  < 0.0178$	$ \delta  < 0.0097$	$ \delta  < 0.0062$	$ \delta  < 0.0043$
$ \varepsilon  < 0.014$	$ \varepsilon  < 0.0053$	$ \varepsilon  < 0.0020$	$ \varepsilon  < 0.0010$	$ \varepsilon  < 0.0005$

Erklären Sie in eigenen Worten den lokalen Grenzwertsatz:

$$B(n, t) \approx N(\mu, \sigma^2)$$

In welchem Sinne nähert sich die Binomial- der Normalverteilung?  
Nennen / erfinden Sie mindestens zehn konkrete Anwendungsbeispiele, in denen diese Näherung praktisch nützlich ist, sinnvoll und hilfreich.

Wie lautet und was besagt der lokale Grenzwertsatz? Wie bestimmt man hierzu die nötigen Kenngrößen  $\mu$  und  $\sigma$ ? In welchen Situationen kann man den LGS anwenden und wie? Welche Voraussetzungen braucht man und welche Schlussfolgerung gewinnt man?

Woher kommen im LGS die Integrationsgrenzen  $\alpha = (a - 1/2 - \mu)/\sigma$  und  $\beta = (b + 1/2 - \mu)/\sigma$ ? Woher kommt die Stetigkeitskorrektur  $\pm 1/2$ ? Wann darf man sie ignorieren? Was passiert im Extremfall  $a = b$ ?

Wie lauten die Fehlerschranken im LGS? Was passiert für  $n \rightarrow \infty$ ?

Wie verbessert man bei Bedarf die Näherung um eine Größenordnung?

Wozu nützen die Schranken in konkreten Anwendungen, bei festem  $n$ ?

Wann darf man sie getrost ignorieren? meistens? manchmal? nie?

Vergleichen Sie den LGS mit Poissons Gesetz der kleinen Zahlen:

$$B(n, t) \approx P(\lambda)$$

In welchem Sinne nähert sich die Binomial- der Poisson-Verteilung?  
Nennen / erfinden Sie mindestens zehn konkrete Anwendungsbeispiele, in denen diese Näherung praktisch nützlich ist, sinnvoll und hilfreich.

Wie lautet und was besagt Poissons Gesetz der kleinen Zahlen? Wie bestimmt man hierzu die nötige Kenngröße  $\lambda$ ? In welchen Situationen kann man diese Näherung anwenden? Welche Voraussetzungen braucht man und welche Schlussfolgerung gewinnt man?

Wie lautet die Fehlerschranke beim Gesetz der kleinen Zahlen?

Wie verhält sie sich für  $n \rightarrow \infty$ ? Wie muss  $t$  angepasst werden?

Wozu nützt diese Schranke in konkreten Anwendungen, bei festem  $n$ ?

Wann darf man sie getrost ignorieren? meistens? manchmal? nie?

*Die Klärung dieser Fragen ist Gegenstand von Kapitel U und V.*

*Prüfen Sie sich selbst und lesen Sie alles nötige gründlich nach.*

Im Experiment  $(\Omega, \mathcal{A}, \mathbf{P})$  wollen Sie die Wkt  $p = \mathbf{P}(A)$  messen. Sie führen  $n$  unabhängige Versuche durch und erhalten  $k$  Treffer.

Diese Stichprobe liefert für  $p$  Ihre **Punktschätzung**  $\hat{p} = k/n \in [0, 1]$ . Diese Zahl sagt leider rein gar nichts über die Güte Ihrer Schätzung!

**Aufgabe:** Bestimmen Sie zu  $k \in \{0, \dots, n\}$  ein **Intervall**  $I_k = [a_k, b_k]$ , das in  $\alpha = 95\%$  aller Stichproben die wahre Wkt  $p \in [0, 1]$  überdeckt.

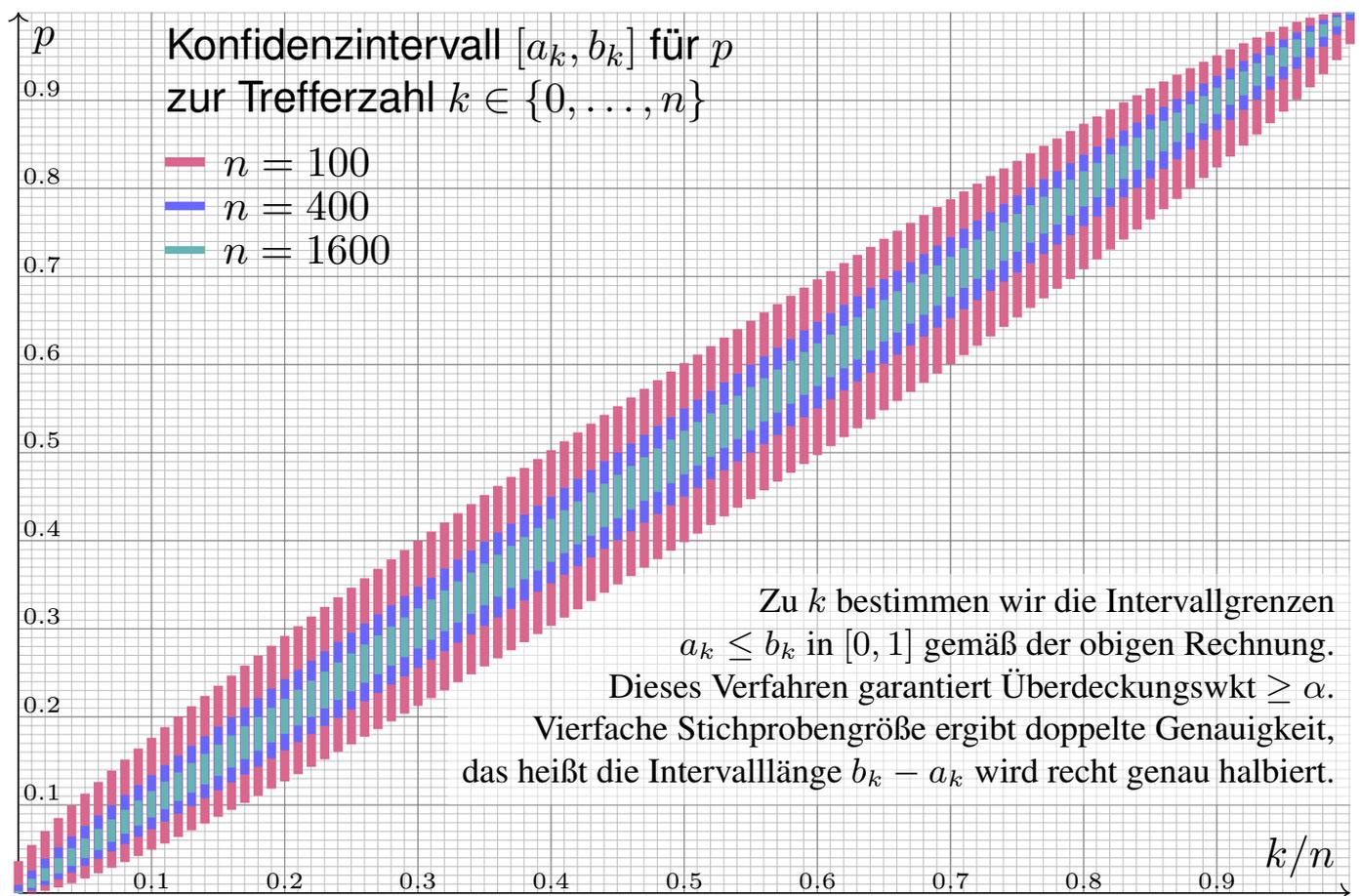
**Lösung:** Wir setzen  $c := (1 - \alpha)/2$  und bestimmen  $a_k, b_k \in [0, 1]$  durch

$$f_k(a) := \sum_{j < k} B(n, a)(j) \stackrel{!}{=} c \quad \text{und} \quad g_k(b) := \sum_{j > k} B(n, b)(j) \stackrel{!}{=} c.$$

Dabei ist  $f_k$  ein Polynom und für  $k > 0$  streng fallend von  $f_k(0) = 1$  nach  $f_k(1) = 0$ . Daher existiert genau ein  $a_k \in [0, 1]$  mit  $f_k(a_k) = c$ . Für  $k = 0$  setzen wir  $a_0 := 0$ . Ebenso verfahren wir für  $b_k$  und setzen  $b_n := 1$ .

**Begründung:** Nicht-Überdeckung  $p \notin I_k$  bedeutet  $p < a_k$  oder  $b_k < p$ . Wir bestimmen  $\ell \in \{0, \dots, n\}$  durch  $a_\ell \leq p \leq b_\ell$  und  $p < a_k$  für  $k > \ell$ . Damit finden wir  $\mathbf{P}(a_X > p) = \sum_{k > \ell} B(n, p)(k) = g_\ell(p) \leq g_\ell(b_\ell) \leq c$ . Ebenso  $\mathbf{P}(b_X < p) \leq c$ . Somit gilt  $\mathbf{P}(I_X \ni p) \geq \alpha$ .

## Konfidenzintervall für die Wkt $p$ (nach Clopper–Pearson)



Das Meinungsforschungsinstitut *Infratest dimap* schreibt über ihre Befragungsmethode auf [www.infratest-dimap.de/service/faqs](http://www.infratest-dimap.de/service/faqs) (aufgerufen am 10.03.2016) folgendes:

*Aus finanziellen und organisatorischen Gründen ist es nicht möglich, Millionen oder Hunderttausende Wahlberechtigte zu befragen. Die Wahlforschung arbeitet daher generell mit Stichproben. Befragt wird lediglich eine kleine Auswahl von Personen, zum Beispiel 1000, die für viele Millionen wahlberechtigte Bundesbürger stehen. [...]*

*Das Ziel, ein kleines Abbild des Ganzen zu bekommen, ist am besten mit einer Zufallsstichprobe zu erreichen. Sie stellt einerseits sicher, dass jeder einzelne Wahlberechtigte oder jeder einzelne Wähler prinzipiell die gleiche Chance besitzt, befragt zu werden, andererseits garantiert sie, dass Interviewer keinen subjektiven Einfluss auf die Auswahl der Befragten haben. [...]*

*Die Genauigkeit einer Befragung oder die Fehlergröße hängt ab von der Anzahl der befragten Personen. Faustregel: Eine Vervierfachung der Befragten halbiert den Zufallsfehler.*

*Wahltagsbefragungen sind meist exakter als Erhebungen in der Zwischen- und Vorwahlzeit. Bei Vorwählerhebungen befragt Infratest dimap in der Regel 1 000 Wahlberechtigte, am Wahlsonntag bei Landtags-, Bundestags- und Europawahlen aber bis zu 50 000 Wähler. [...] Vorwählerhebungen finden in der Regel als computergestützte Telefonerhebung statt.*

**Aufgabe:** Erklären Sie möglichst genau die genannte Faustregel: Eine Vervierfachung der Befragten halbiert den Zufallsfehler.

Der Stimmanteil  $p \in [0.07, 0.93]$  einer Partei soll prognostiziert werden. Hierzu werden  $n = 1000$  zufällige Personen befragt. Den gesuchten Anteil  $p = K/N$  schätzen wir durch den Stichprobenanteil  $\hat{p} = k/n$ .

**Aufgabe:** Mit welcher Wkt  $q$  ist die Schätzung auf 0.01 genau?

**Lösung:** Die Anzahl  $k$  ist  $B(n, p)$ -verteilt. Erwartung  $\mu = np \in [70, 930]$ , Varianz  $\sigma^2 = np(1 - p) \in [65, 250]$ , Streuung  $\sigma \in [8, 16]$  (ausgerundet). Abweichung  $|k/n - p| \leq 0.01$  bedeutet  $|k - np| \leq 0.01n = 10$ . Also

$$q = \sum_{k=\lceil np-10 \rceil}^{\lfloor np+10 \rfloor} \binom{n}{k} p^k (1-p)^{n-k} \approx \int_{-\alpha}^{\alpha} \varphi(t) dt = 2 \int_{t=0}^{\alpha} \varphi(t) dt.$$

Hier ist  $\alpha = 10.5/\sigma \in [0.65, 1.32]$ . Tabelle ablesen:  $q \in [0.47, 0.82]$ .

Der Näherungsfehler  $|\varepsilon| < 1/(3\sigma^2) < 0.006$  wurde hier mitgerechnet.

😊 Mit Wahrscheinlichkeit  $q \geq 47\%$  ist die Schätzung auf 0.01 genau.

$ k/n - p $	$\leq 0.01$	$\leq 0.02$	$\leq 0.03$	$\leq 0.04$	$\leq 0.05$
$\alpha$	[0.65, 1.32]	[1.28, 2.57]	[1.90, 3.82]	[2.53, 5.07]	[3.15, 6.32]
$q$	[0.47, 0.82]	[0.79, 0.99]	[0.94, 1.00]	[0.98, 1.00]	[0.99, 1.00]

Eine Wahl mit über 5 000 000 Stimmberechtigten steht bevor. Der Stimmanteil  $p \in [0.02, 0.98]$  einer Partei soll geschätzt werden. Hierzu werden  $n$  zufällige Personen befragt. Die Schätzung soll mit Wahrscheinlichkeit  $\geq 95\%$  bis auf einen Fehler von  $\leq 0.01$  genau sein.

**Aufgabe:** Wie viele Personen sollten hierfür befragt werden? Welche Näherungen nutzen Sie hierzu? mit welcher Güte?

**Lösung:** Von allen  $N$  Stimmberechtigten stimmen  $K$  für die Partei. Sei  $k$  die Anzahl der  $n$  befragten Personen, die sich für die Partei aussprechen. Diese Anzahl ist  $H(N, K, n)$ -verteilt (Stichprobe oZ). Für große  $N$  ist dies annähernd  $B(n, p)$ -verteilt (Stichprobe mZ).

Diese Näherung werden wir nachträglich rechtfertigen: Der totale Abstand von  $H(N, K, n)$  zu  $B(n, p)$  ist kleiner als  $n/N$ , für  $n \leq 10\,000$  und  $N \geq 5\,000\,000$  also kleiner als 0.002. Alternativ könnte man die Befragung auch so gestalten, dass statt der hypergeometrischen gleich die Binomialverteilung entsteht, nämlich durch Auslosung mit Zurücklegen.

Den unbekanntem Anteil  $p = K/N$  schätzen wir durch den Wert  $k/n$ . Fehler  $\leq 0.01$  bedeutet  $|k/n - p| \leq 0.01$ , also  $|k - np| \leq 0.01n$ . Das heißt  $a \leq k \leq b$  mit  $a = \lceil np - 0.01n \rceil$  und  $b = \lfloor np + 0.01n \rfloor$ .

Diese noch tolerierbare Abweichung tritt auf mit Wahrscheinlichkeit

$$\sum_{k=a}^b \binom{n}{k} p^k (1-p)^{n-k} \approx \int_{-\alpha}^{\alpha} \varphi(t) dt = 2 \int_0^{\alpha} \varphi(t) dt \stackrel{!}{\geq} 0.95$$

wobei  $\alpha = 0.01n / \sqrt{np(1-p)}$ . Aus der Tabelle lesen wir ab:  $\alpha \stackrel{!}{\geq} 2$ . Hieraus bestimmen wir eine geeignete Stichprobengröße  $n$ :

$$\alpha = \frac{0.01\sqrt{n}}{\sqrt{p(1-p)}} \geq 2 \iff \sqrt{n} \geq 200\sqrt{p(1-p)}$$

$$\iff n \geq 40\,000 \cdot p(1-p) \iff n \geq 10\,000$$

Wegen  $p(1-p) \leq 1/4$  genügt es in jedem Falle,  $n \geq 200^2 \cdot 1/4 = 10\,000$  zu wählen. Genauer: Für  $p \in [0.2, 0.8]$  gilt  $\sigma^2 = np(1-p) \in [1600, 2500]$ , also  $\sigma \in [40, 50]$ . Mit Rundung und Stetigkeitskorrektur gilt demnach  $\alpha' = (0.01n - 0.5)/\sigma \in [1.99, 2.49]$  und Wkt  $\geq 95.34\%$ . Für  $p \in [0.02, 0.2] \cup [0.8, 0.98]$  gilt  $\sigma^2 = np(1-p) \in [196, 1600]$ , also  $\sigma \in [14, 40]$ . Wir erhalten demnach  $\alpha' = (0.01n - 0.5)/\sigma \in [2.48, 7.11]$  und sogar Wkt  $\geq 98.68\%$ . Der Approximationsfehler des LGS ist höchstens  $|\varepsilon| < 1/(3\sigma^2) < 0.00021$  bzw.  $< 0.00171$ .

😊 Bei Stichprobengröße  $n \geq 10\,000$  ist die Schätzung  $k/n$  für  $K/N$  mit Wahrscheinlichkeit  $\geq 95\%$  bis auf einen Fehler von  $\leq 0.01$  genau.

Jede Messung unterliegt zufälligen Störungen, wir betrachten sie daher als eine Zufallsvariable  $X : \Omega \rightarrow \mathbb{R}$ . Wir nehmen nun  $X \sim N(\mu, \sigma^2)$  an, das heißt, die Messwerte sind normalverteilt um den Mittelwert  $\mu \in \mathbb{R}$  mit Streuung  $\sigma \in \mathbb{R}_{>0}$ . Ziel der Messung ist die Schätzung des Wertes  $\mu$ .

Wir untersuchen das Motto: „Die Wahrheit liegt irgendwo in der Mitte.“

Sie betrachten folgende besonders simple Methode: Sie wiederholen die Messung  $n$  mal unabhängig und erhalten Messwerte  $X_1, \dots, X_n$ . Hieraus bilden Sie das Minimum  $A = \min\{X_1, \dots, X_n\}$  und Maximum  $B = \max\{X_1, \dots, X_n\}$  und nehmen als Schätzung das Intervall  $[A, B]$ .

- Aufgabe:** (1) Berechnen Sie die Wahrscheinlichkeit  $\mathbf{P}([A, B] \ni \mu)$ , dass Ihr Intervall tatsächlich den gesuchten Mittelwert  $\mu$  überdeckt. Wie verhält sich diese Wkt mit wachsender Fallzahl  $n = 1, 2, 3, \dots$ ?
- (2) Berechnen Sie die Wkten  $\mathbf{P}(B \leq \mu + \sigma)$  und  $\mathbf{P}(B \leq \mu + 2\sigma)$  etc., dass Ihre Schätzung  $[A, B]$  keine allzu große Unsicherheit offen lässt. Wie verhält sich diese Wkt mit wachsender Fallzahl  $n = 1, 2, 3, \dots$ ?
- (3) Warum ist diese Min-Max-Schätzung noch nicht zufriedenstellend?

**Lösung:** (1) Nach Konstruktion unseres Verfahrens gilt immer  $A \leq B$ . Zu hohe Schätzung  $\mu < A \leq B$  bedeutet, dass  $n$  mal ausschließlich Werte  $X_1, \dots, X_n > \mu$  gemessen wurde. Wegen  $\mathbf{P}(X_k > \mu) = 1/2$  und Unabhängigkeit der Messungen geschieht dies mit Wkt  $\mathbf{P}(A > \mu) = 2^{-n}$ . Zu niedrige Schätzung  $A \leq B < \mu$  bedeutet, dass  $n$  mal ausschließlich Werte  $X_1, \dots, X_n < \mu$  gemessen wurde. Wegen  $\mathbf{P}(X_k < \mu) = 1/2$  und Unabhängigkeit der Messungen geschieht dies mit Wkt  $\mathbf{P}(B < \mu) = 2^{-n}$ . Das Ereignis  $A \leq \mu \leq B$  hat also die Wkt  $\mathbf{P}(A \leq \mu \leq B) = 1 - 2^{1-n}$ .

Für  $n = 1$  erhalten wir nur die Punktschätzung  $[A, B] = \{X_1\}$ .

Ihre Trefferwahrscheinlichkeit  $\mathbf{P}(X_1 = \mu)$  ist tatsächlich gleich Null.

Für  $n = 2$  ist die Trefferwkt  $\mathbf{P}(A \leq \mu \leq B) = 1/2$ : immerhin etwas.

Für  $n = 3$  ist die Trefferwkt  $\mathbf{P}(A \leq \mu \leq B) = 3/4$ : etwas besser.

Für  $n = 4$  ist die Trefferwkt  $\mathbf{P}(A \leq \mu \leq B) = 7/8$ : noch besser.

Für  $n = 11$  ist die Trefferwkt  $\mathbf{P}(A \leq \mu \leq B) = 1 - 2^{-10} \approx 99.9\%$ .

Für  $n \rightarrow \infty$  erhalten wir den Grenzwert  $\mathbf{P}(A \leq \mu \leq B) = 1 - 2^{1-n} \nearrow 1$ .

(2) Laut Tabelle der Normalverteilung gilt  $\mathbf{P}(X_k \leq \mu + \sigma) \approx 0.84134$ .  
Demnach gilt  $\mathbf{P}(B \leq \mu + \sigma) \approx 0.84134^n \searrow 0$  für  $n \rightarrow \infty$ . **Schade!**  
Für  $n = 11$  finden wir  $\mathbf{P}(B \leq \mu + \sigma) \approx 0.84134^{11} \approx 0.15$ .

Laut Tabelle der Normalverteilung gilt  $\mathbf{P}(X_k \leq \mu + 2\sigma) \approx 0.97725$ .  
Demnach gilt  $\mathbf{P}(B \leq \mu + 2\sigma) \approx 0.97725^n \searrow 0$  für  $n \rightarrow \infty$ . **Schade!**  
Für  $n = 11$  finden wir  $\mathbf{P}(B \leq \mu + 2\sigma) \approx 0.97725^{11} \approx 0.78$ .

Laut Tabelle der Normalverteilung gilt  $\mathbf{P}(X_k \leq \mu + 3\sigma) \approx 0.99865$ .  
Demnach gilt  $\mathbf{P}(B \leq \mu + 3\sigma) \approx 0.99865^n \searrow 0$  für  $n \rightarrow \infty$ . **Schade!**  
Für  $n = 11$  finden wir  $\mathbf{P}(B \leq \mu + 3\sigma) \approx 0.99865^{11} \approx 0.98$ .

(3) Das hier genutzte Verfahren ist einfach und sicher, aber ungenau:

😊 Die Berechnung von  $A = \min$  und  $B = \max$  ist denkbar einfach.  
Die Intervallschätzung  $[A, B]$  wird mit größerem  $n$  immer sicherer.

😞 Leider wird sie gleichzeitig immer ungenauer, sogar sehr schnell,  
obwohl mehr Messwerte doch größere Genauigkeit bringen sollten!

⚠️ Aus diesen triftigen Gründen nutzt man für die Berechnung von **Konfidenzintervallen** genauere Methoden (nach Neyman 1937).

Als Stichprobe für  $X$  führen wir  $n$  **unabhängige Messungen** aus.  
Aus den so gewonnenen Messwerten  $x_1, \dots, x_n \in \mathbb{R}$  berechnen wir

den **Stichprobenmittelwert**  $\hat{x} := \frac{1}{n} \sum_{k=1}^n x_k,$

die **Stichprobenvarianz**  $\hat{\sigma}^2 := \frac{1}{n-1} \sum_{k=1}^n (x_k - \hat{x})^2,$

das **Konfidenzintervall**  $I_k := \left[ \hat{x} - k \frac{\hat{\sigma}}{\sqrt{n}}, \hat{x} + k \frac{\hat{\sigma}}{\sqrt{n}} \right].$

**Aufgabe:** Für  $X \sim N(\mu, \sigma)$  machen wir  $n = 2$  Messungen.

(1) Berechnen Sie die Wahrscheinlichkeit  $\mathbf{P}([\hat{x} \pm \hat{\sigma}/\sqrt{n}] \ni \mu)$ ,  
dass das Intervall  $I_1$  tatsächlich den gesuchten Wert  $\mu$  überdeckt.

(2) Berechnen Sie ebenso  $\mathbf{P}([\hat{x} \pm k\hat{\sigma}/\sqrt{n}] \ni \mu)$  für das Intervall  $I_k$ .

😊 Größeres  $k$  bringt Sicherheit, mehr Messwerte bringen Genauigkeit!  
Für  $n = 2, 3, 4, \dots$  nutzen wir Gosssets Student- $t$ -Verteilung (Satz V4A).  
Für große  $n$  nähert sie sich rasch der Normalverteilung, das vereinfacht.

**Lösung:** Wir rechnen den einfachsten Fall  $n = 2$  sorgfältig durch.

$$\begin{aligned}\hat{x} &= \frac{1}{n} \sum_{k=1}^n x_k && \stackrel{n=2}{=} \frac{x_1 + x_2}{2} \\ \hat{\sigma}^2 &= \frac{1}{n-1} \sum_{k=1}^n (x_k - \hat{x})^2 && = \frac{(x_1 - x_2)^2}{2} \\ \hat{\sigma} &= \sqrt{\hat{\sigma}^2} && = \frac{|x_1 - x_2|}{\sqrt{2}} \\ I_k &= \left[ \hat{x} \pm k \frac{\hat{\sigma}}{\sqrt{n}} \right] && = \left[ \frac{x_1 + x_2}{2} \pm k \frac{|x_1 - x_2|}{2} \right]\end{aligned}$$

😊 Speziell für  $k = 1$  erhalten wir erneut  $I_1 = [\min(x_1, x_2), \max(x_1, x_2)]$ .

Die beiden Messungen  $X_1, X_2: \Omega \rightarrow \mathbb{R}$  setzen wir als unabhängig und  $N(\mu, \sigma)$ -verteilt voraus. Das Intervall  $I_k \subseteq \mathbb{R}$  hängt nur von den beiden Messwerten  $x_1 = X_1(\omega)$  und  $x_2 = X_2(\omega)$  ab, und ist wie diese ebenso dem Zufall unterworfen, also dem zufälligen Ergebnis  $\Omega \ni \omega \mapsto (x_1, x_2)$ .

Wir berechnen nun die Wahrscheinlichkeit  $\mathbf{P}(I_k \ni \mu)$ , dass unser zufälliges Intervall  $I_k \subseteq \mathbb{R}$  den tatsächlichen Wert  $\mu \in \mathbb{R}$  überdeckt. Dank Unabhängigkeit ist die Verteilung von  $(X_1, X_2): \Omega \rightarrow \mathbb{R}^2$  eine zweidimensionale Normalverteilung  $N(\mu, \sigma^2) \otimes N(\mu, \sigma^2)$  mit Dichte

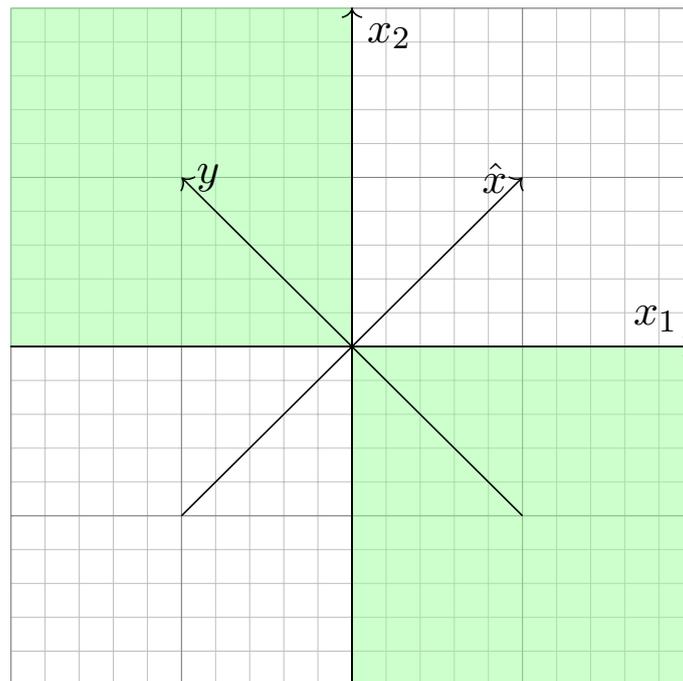
$$\begin{aligned}f(x_1, x_2) &= \frac{1}{\sqrt{2\pi}} e^{-(x_1 - \mu)^2 / (2\sigma^2)} \cdot \frac{1}{\sqrt{2\pi}} e^{-(x_2 - \mu)^2 / (2\sigma^2)} \\ &= \frac{1}{2\pi} e^{-|(x_1, x_2) - (\mu, \mu)|^2 / (2\sigma^2)}.\end{aligned}$$

Sie ist rotationssymmetrisch um ihren Schwerpunkt  $(\mu, \mu)$ , siehe V108. Unsere Formeln für  $\hat{x}$  und  $\hat{\sigma}$  legen die folgende Transformation nahe:

$$\left. \begin{aligned} \hat{X} &= \frac{1}{2}(X_1 + X_2) \\ Y &= \frac{1}{2}(X_2 - X_1) \end{aligned} \right\} \iff \begin{pmatrix} \hat{X} \\ Y \end{pmatrix} = \begin{pmatrix} 1/2 & 1/2 \\ -1/2 & 1/2 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

Dies ist eine Drehstreckung: Drehung um  $45^\circ$  und Stauchung um  $\sqrt{2}$ . Demnach ist auch die Verteilung von  $(\hat{X}, Y): \Omega \rightarrow \mathbb{R}^2$  normalverteilt und ebenso rotationssymmetrisch um ihren Schwerpunkt  $(\mu, 0)$ .

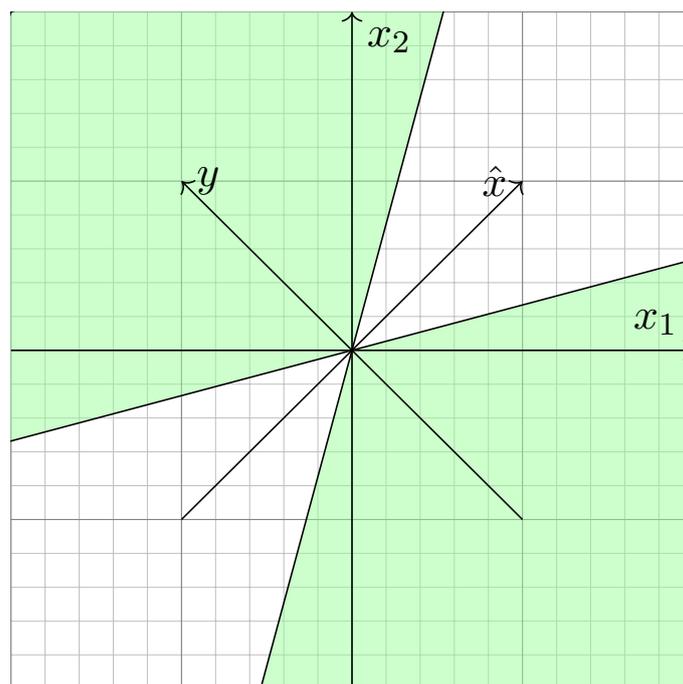
(1) Das Ereignis  $I_1 \ni \mu$  entspricht  $|\hat{X} - \mu| \leq |Y|$ . Skizze für  $\mu = 0$ :



Seine Wkt entspricht dem grünen Winkel, somit gilt  $\mathbf{P}(I_1 \ni \mu) = 1/2$ .

😊 Dies entspricht der Min-Max-Schätzung aus der vorigen Aufgabe.

(2) Das Ereignis  $I_2 \ni \mu$  entspricht  $|\hat{X} - \mu| \leq 2|Y|$ . Skizze für  $\mu = 0$ :

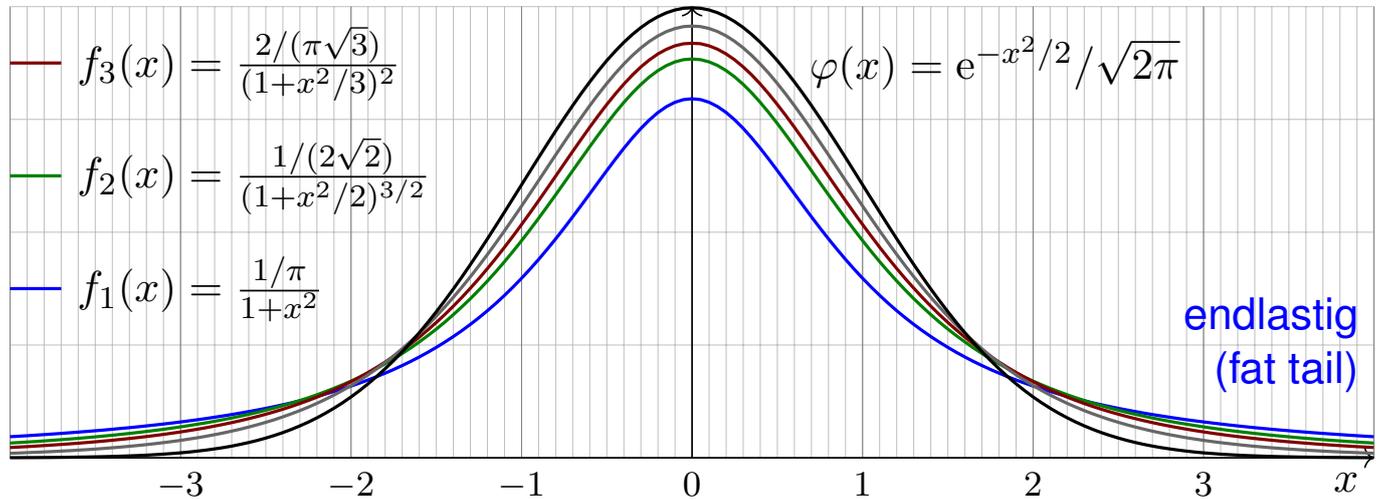


Seine Wkt ist  $\mathbf{P}(I_2 \ni \mu) \approx 70\%$ , allgemein  $\mathbf{P}(I_k \ni \mu) = \frac{2}{\pi} \arctan(k)$ .

😊 Für  $k \rightarrow \infty$  gilt demnach  $\mathbf{P}(I_k \ni \mu) = \frac{2}{\pi} \arctan(k) \nearrow 1$ .

Zur Anzahl  $\nu \in \mathbb{N}_{\geq 1}$  von „Freiheitsgraden“ definieren wir die Wktsdichte

$$f_\nu : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0} : x \mapsto \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\nu\pi}} (1 + x^2/\nu)^{-\frac{\nu+1}{2}}.$$



Normierung  $\int_{\mathbb{R}} f_\nu(x) dx = 1$  gelingt dank Gamma-Funktion. B316 C232

Für  $\nu = 1$  erhalten wir die obige Cauchy-Verteilung  $f_1(x) = \frac{1/\pi}{1+x^2}$ . V241

Für  $\nu \rightarrow \infty$  erhalten wir die Standard-Normalverteilung  $\varphi$ . (Übung!)

Diese Verteilung hat folgende zentrale Bedeutung in der Statistik:

**Satz V4A: Gosset 1908, alias „Student“**

Seien  $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}$  unabhängig, normalverteilt,  $X_k \sim N(\mu, \sigma^2)$ .  
Wir betrachten Stichprobenmittelwert, -Varianz und -Abweichung:

$$\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k, \quad S^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2 \quad \text{und} \quad T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

Dann sind die ZVaren  $\bar{X}$  und  $S^2$  stochastisch unabhängig, der Stichprobenmittelwert  $\bar{X} \sim N(\mu, \sigma^2/n)$  ist normalverteilt, die Stichprobenvarianz  $S^2$  ist  $\chi^2$ -verteilt, und die Abweichung  $T$  ist Student- $t$ -verteilt mit der oben angegebenen Dichte  $f_{n-1}$ .

Das bedeutet: Der Abstand des gemessenen Mittelwertes  $\bar{X}$  zum wahren (aber uns unbekanntem) Mittelwert  $\mu$  ist verteilt wie  $T S/\sqrt{n}$ . Den Fall  $n = 2$  haben wir oben ausgeführt; der allgemeine Fall gelingt ähnlich. Diese Rechnung und die  $\chi^2$ -Verteilung führe ich hier nicht aus.

Zum Konfidenzniveau  $\alpha$  nutzen wir das **Konfidenzintervall**

$$I_k = \left[ \bar{X} - k S / \sqrt{n}, \bar{X} + k S / \sqrt{n} \right].$$

Den Wert  $k$  bestimmen wir dabei so, dass  $\int_{-k}^k f_\nu(x) dx = 1 - \alpha$  gilt.

$\alpha =$	0.500	0.750	0.800	0.900	0.950	0.980	0.990	0.998
$\nu = 1$	1.000	2.414	3.078	6.314	12.706	31.821	63.657	318.309
2	0.816	1.604	1.886	2.920	4.303	6.965	9.925	22.327
3	0.765	1.423	1.638	2.353	3.182	4.541	5.841	10.215
4	0.741	1.344	1.533	2.132	2.776	3.747	4.604	7.173
5	0.727	1.301	1.476	2.015	2.571	3.365	4.032	5.893
10	0.700	1.221	1.372	1.812	2.228	2.764	3.169	4.144
20	0.687	1.185	1.325	1.725	2.086	2.528	2.845	3.552
30	0.683	1.173	1.310	1.697	2.042	2.457	2.750	3.385
40	0.681	1.167	1.303	1.684	2.021	2.423	2.704	3.307
50	0.679	1.164	1.299	1.676	2.009	2.403	2.678	3.261
100	0.677	1.157	1.290	1.660	1.984	2.364	2.626	3.174
200	0.676	1.154	1.286	1.653	1.972	2.345	2.601	3.131
300	0.675	1.153	1.284	1.650	1.968	2.339	2.592	3.118
400	0.675	1.152	1.284	1.649	1.966	2.336	2.588	3.111
500	0.675	1.152	1.283	1.648	1.965	2.334	2.586	3.107
$\infty$	0.674	1.150	1.282	1.645	1.960	2.326	2.576	3.090

# Kleine Stichproben und Gossets Student- $t$ -Verteilung

(0) Durch Stichprobe wollen wir den Mittelwert  $\mu$  schätzen. Aus Messwerten  $X_1, X_2, \dots, X_n$  bestimmen wir dazu ein geeignetes Konfidenzintervall  $I = [A, B]$ . Dieses überdeckt den wahren (aber uns unbekannt) Wert  $\mu$  mit der Wkt  $\mathbf{P}(I \ni \mu)$ . Wir geben uns ein Konfidenzniveau  $\alpha$  vor und verlangen  $\mathbf{P}(I \ni \mu) \geq \alpha$ . Typische Werte sind  $\alpha = 95\%$  oder  $\alpha = 99\%$ .

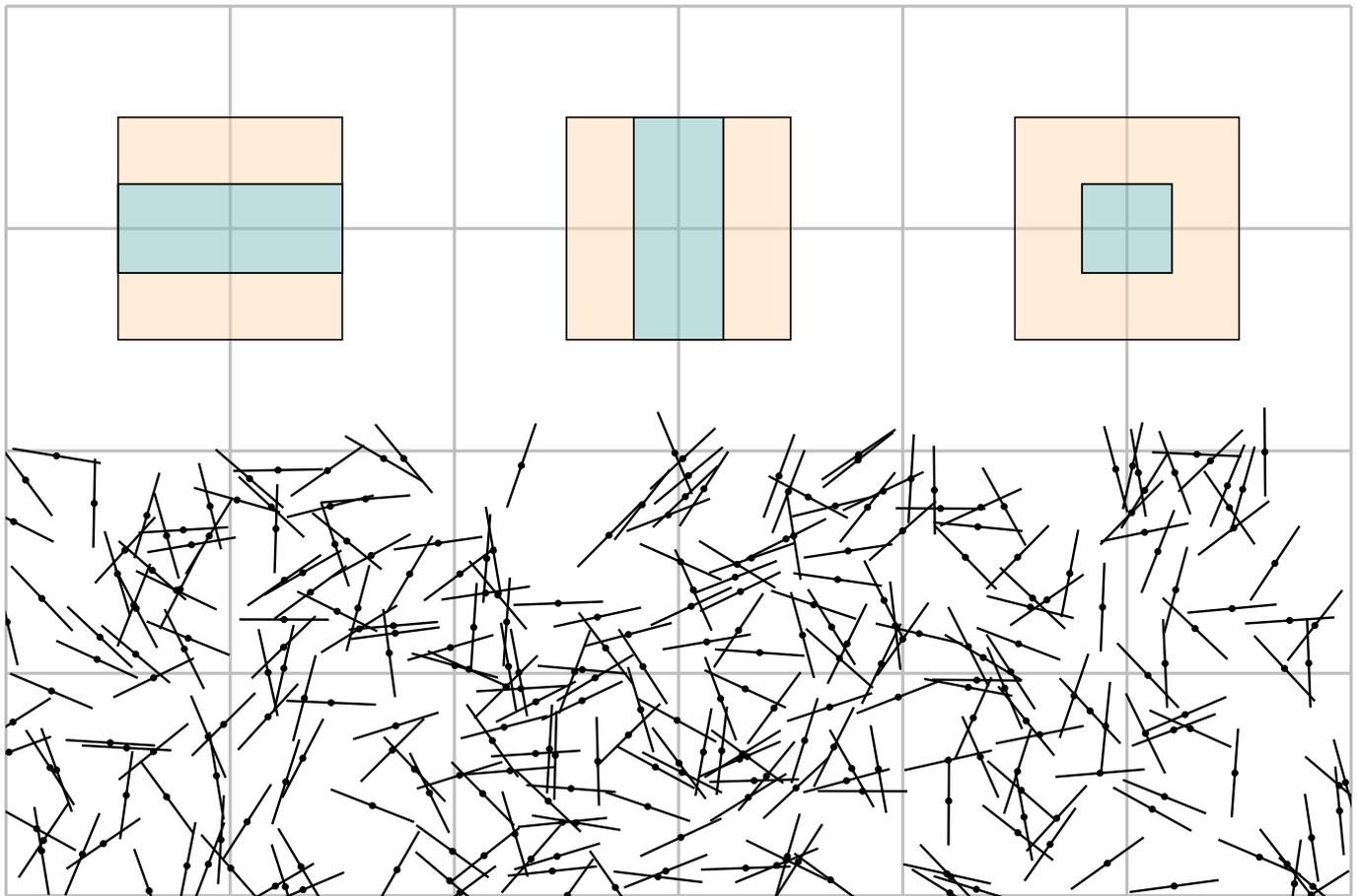
(1) Die Zufallsvariable  $\bar{X}$  ist  $N(\mu, \sigma^2/n)$ -verteilt. Ist die Varianz  $\sigma^2$  bekannt, so betrachten wir das Intervall  $I = [\bar{X} - \tau\sigma/\sqrt{n}, \bar{X} + \tau\sigma/\sqrt{n}]$  und bestimmen  $\tau$  so, dass  $\int_{-\tau}^{\tau} \varphi(x) dx = 1 - \alpha$ .

(2) In realistischen Anwendungen ist die Varianz  $\sigma^2$  unbekannt und muss ebenso wie  $\mu$  aus der Stichprobe geschätzt werden. Dies gelingt durch die (Bessel-korrigierte) Stichprobenvarianz  $S^2$ . Wir betrachten  $T = (\bar{X} - \mu)/(S/\sqrt{n})$  als Abweichung vom Mittelwert. Bei großen Stichproben ( $n \rightarrow \infty$ ) nimmt man näherungsweise an, dass  $T$  normalverteilt ist, wie in (1), das vereinfacht.

(3) Bei kleinen Stichprobengrößen  $n$  ist dies jedoch allzu optimistisch. William Gosset stellte 1908 fest, dass  $T$  nicht normalverteilt ist, sondern der oben angegebenen Verteilung  $f_\nu$  folgt, wobei  $\nu = n - 1$ , genannt *Student- $t$ -Verteilung*, kurz  *$t$ -Verteilung*. Das ist das richtige Modell! Die Abweichung von  $f_\nu$  zu  $\varphi$  ist für kleine  $n$  deutlich spürbar; für  $n \rightarrow \infty$  nähert sich  $f_\nu$  rasch der Standard-Normalverteilung  $\varphi$ . (Lehrreiche Übung: Rechnen Sie dies sorgfältig nach!)

Zur Mittelwertschätzung nutzen wir das Konfidenzintervall  $I_k = [\bar{X} - kS/\sqrt{n}, \bar{X} + kS/\sqrt{n}]$  und bestimmen den Wert  $k$  so, dass  $\int_{-k}^k f_\nu(x) dx = 1 - \alpha$  gilt. Die obige Tabelle zeigt hierzu Quantile der Student- $t$ -Verteilungen. Das Ergebnis  $k$  ist etwas größer als der Wert  $\tau$  aus (1).

*Fun fact:* Gosset arbeitete in der Dubliner Guinness-Brauerei. Da sie ihm die Veröffentlichung seines Artikels untersagte, veröffentlichte er kurzerhand unter dem Pseudonym „Student“.



**Aufgabe:** In der Ebene  $\mathbb{R}^2$  liegen die Geradenscharen  $X = \mathbb{R} \times \mathbb{Z}$  und  $Y = \mathbb{Z} \times \mathbb{R}$ . Sie werfen zufällig eine Münze vom Durchmesser  $\ell \in [0, 1]$  und beobachten, ob sie die Geraden  $X$  oder die Geraden  $Y$  schneidet.

- (1) Warum ist das Quadrat  $\Omega = [-1/2, 1/2]^2 \subseteq \mathbb{R}^2$  mit kontinuierlicher Gleichverteilung ein geeignetes Modell für dieses Zufallsexperiment?  
 (2) Wie wahrscheinlich ist es, dass die Münze die Geraden  $X$  schneidet?  $Y$  schneidet? Beide gleichzeitig? Sind beide stochastisch unabhängig?

**Lösung:** (1) Die Lage der Münze wird durch die Position  $(x, y) \in \mathbb{R}^2$  ihres Mittelpunktes festgelegt. (Alle Drehungen sind hier irrelevant.) Schnitte sind invariant unter ganzzahligen Verschiebungen um  $v \in \mathbb{Z}^2$ . Wir notieren nur die Position  $(x, y) \in \Omega$  zur nächstgelegenen Geraden. Die Koordinaten  $x, y \in [-1/2, 1/2]$  seien gleichverteilt und unabhängig. Dann ist  $\mathbf{P}(A) = \text{vol}_2(A)$  auf  $\Omega$  das richtige stochastische Modell.

(2) Die gefragten Wahrscheinlichkeiten lesen wir aus der Skizze ab: Das Ereignis  $A := \{\text{Münze trifft } X\}$  entspricht  $|y| \leq \ell/2$ , also  $\mathbf{P}(A) = \ell$ . Das Ereignis  $B := \{\text{Münze trifft } Y\}$  entspricht  $|x| \leq \ell/2$ , also  $\mathbf{P}(B) = \ell$ . Entsprechend gilt  $\mathbf{P}(A \cap B) = \ell^2 = \mathbf{P}(A) \cdot \mathbf{P}(B)$ . Beide sind unabhängig!

Der junge Graf Georges-Louis de Buffon (1707–1788) warf Stöcke über die Schulter auf einen gekachelten Fußboden. Anschließend zählte er, wie oft sie die Fugen trafen. Die Berechnung dieser geometrischen Wahrscheinlichkeiten bezeichnet man als **Buffonsches Nadelproblem**. Als wunderbares Video zu mathematischer Eleganz:  Weitz, [youtu.be/YIkqZsz4faA](https://youtu.be/YIkqZsz4faA)

**Aufgabe:** Sie werfen zufällig eine Nadel der Länge  $\ell \in [0, 1]$ .

(1) Warum ist  $\Omega = [-1/2, 1/2] \times [-1/2, 1/2] \times [0, 2\pi]$  mit kontinuierlicher Gleichverteilung ein geeignetes Modell für dieses Zufallsexperiment?

(2) Wie wahrscheinlich ist es, dass die Nadel die Geraden  $X$  schneidet?  $Y$  schneidet? Beide gleichzeitig? Sind beide stochastisch unabhängig?

**Lösung:** (1) Die Lage der Nadel wird durch die Position  $(x, y) \in \mathbb{R}^2$  ihres Mittelpunktes und den Winkel  $\varphi \in [0, 2\pi]$  festgelegt, gemessen zwischen der  $x$ -Achse und dem Vektor vom Mittelpunkt zur Spitze.

Schnitte sind invariant unter ganzzahligen Verschiebungen um  $v \in \mathbb{Z}^2$ . Wir notieren daher nur die Position  $(x, y)$  zur nächstgelegenen Geraden. Wir erhalten so  $(x, y, \varphi) \in \Omega = [-1/2, 1/2] \times [-1/2, 1/2] \times [0, 2\pi]$ . Annahme: Die drei Koordinaten  $x, y, \varphi$  sind jeweils gleichverteilt und unabhängig. Daher ist  $\mathbf{P}(A) = \frac{1}{2\pi} \text{vol}_3(A)$  auf  $\Omega$  das richtige stochastische Modell.

(2) Das Ereignis  $A := \{\text{Nadel trifft } X\}$  entspricht  $|y| \leq \ell/2 \cdot |\cos \varphi|$ . Also:

$$\begin{aligned} \mathbf{P}(A) &= \frac{1}{2\pi} \text{vol}_3(A) = \frac{1}{2\pi} \int_{\varphi=0}^{2\pi} \int_{x=-1/2}^{1/2} \int_{y=-\ell/2 \cdot |\cos \varphi|}^{\ell/2 \cdot |\cos \varphi|} dy dx d\varphi \\ &= \frac{1}{2\pi} \int_{\varphi=0}^{2\pi} \ell |\cos \varphi| d\varphi = \frac{\ell}{\pi} \int_{\varphi=-\pi/2}^{\pi/2} \cos \varphi d\varphi = \frac{\ell}{\pi} \left[ \sin(\varphi) \right]_{-\pi/2}^{\pi/2} = \frac{2\ell}{\pi} \end{aligned}$$

Das Ereignis  $B := \{\text{Nadel trifft } Y\}$  entspricht dann  $|x| \leq \ell/2 \cdot |\sin \varphi|$ .

Dank Symmetrie gilt  $\mathbf{P}(B) = \mathbf{P}(A) \dots$  oder nachrechnen! Schließlich:

$$\begin{aligned} \mathbf{P}(A \cap B) &= \frac{1}{2\pi} \int_{\varphi=0}^{2\pi} \int_{x=-\ell/2 \cdot |\sin \varphi|}^{\ell/2 \cdot |\sin \varphi|} \int_{y=-\ell/2 \cdot |\cos \varphi|}^{\ell/2 \cdot |\cos \varphi|} dy dx d\varphi \\ &= \frac{\ell^2}{2\pi} \int_{\varphi=0}^{2\pi} |\sin \varphi| \cdot |\cos \varphi| d\varphi = \frac{2\ell^2}{\pi} \int_{\varphi=0}^{\pi/2} \sin \varphi \cos \varphi d\varphi \\ &= \frac{\ell^2}{\pi} \int_{\varphi=0}^{\pi/2} \sin(2\varphi) d\varphi = \frac{\ell^2}{\pi} \left[ -\frac{\cos(2\varphi)}{2} \right]_{\varphi=0}^{\pi/2} = \frac{\ell^2}{\pi} < \frac{4\ell^2}{\pi^2} = \mathbf{P}(A) \cdot \mathbf{P}(B) \end{aligned}$$

 Die Ereignisse  $A$  und  $B$  sind nicht unabhängig! Das ist plausibel:  $A$  ist besonders wahrscheinlich für  $\varphi \approx \pm\pi/2$ , aber  $B$  eher für  $\varphi \approx 0, \pi$ .