Stochastik und Statistik

Vorlesung WT 2 Zufallsvariablen und Verteilungen

K.Gerald van den Boogaart

http://www.stat.boogaart.de

Kombinatorik

Motivation

Bei einer Stichprobe zur Qualitätskontrolle werden 4 von 10 Bauteilen überprüft.
Wie wahrscheinlich ist es, dass keine defektes Bauteil untersucht wird, wenn 2 Bauteile defekt waren.

$$P(\text{Defekt gefunden}) = \frac{\text{Anzahl Auswahlmöglichkeiten mit defektem Teil}}{\text{Anzahl Auswahlmöglichkeiten}}$$

Möglichkeiten zählen

Wieviel Möglichkeiten gibt es

- aus n_1 weiblichen und n_2 männlichen Zwergen eine gemischtes Doppel zusammenzustellen?
- sich aus n_1 weiblichen und n_2 männlichen Zwergen einen Zwerg herauszusuchen?
- k mal nacheinander einen der n Zwerge auszusuchen?
- n Zwerge in eine Reihenfolge zu bringen?
- k Zwerge aus n Zwergen auszusuchen?
- k Zwerge nacheinander aus n Zwergen auszusuchen?
- Zwerge aus n Zwergen auszuwählen?

Von denen und von jenen

Wieviel Möglichkeiten gibt es

- **aus n_1 weiblichen und n_2 männlichen Zwergen eine gemischtes Doppel zusammenzustellen?**

$$n_{\text{ges}} = n_1 n_2$$

$$n_1 \left\{ \underbrace{\begin{matrix} (1,1) & \cdots & (1,n_2) \\ \vdots & \ddots & \vdots \\ (n_1,1) & \cdots & (n_1,n_2) \end{matrix}}_{n_2} \right.$$

Von denen oder von jenen

Wieviel Möglichkeiten gibt es

- **sich aus n_1 weiblichen und n_2 männlichen Zwergen einen Zwerg herauszusuchen?**

$$n_{\text{ges}} = |\{\underbrace{w_1, \ldots, w_{n_1}}_{n_1}, \underbrace{m_1, \ldots, m_{n_2}}_{n_2}\}| = n_1 + n_2$$

Variation mit Wiederholung

Wieviel Möglichkeiten gibt es

- **k mal nacheinander einen der n Zwerge auszusuchen?**

$$n_{\text{ges}} = \underbrace{n \cdot n \cdot \cdots \cdot n}_{k-\text{mal}} = n^k$$

Permutation ohne Wiederholung

Wieviel Möglichkeiten gibt es

- **n Zwerge in eine Reihenfolge zu bringen?**

$$n_{\text{ges}} = n(n-1)(n-2)\cdots 1 =: n!$$

n	Möglichkeiten für den 1. Zwerg
$(n-1)$	Möglichkeiten für den 2. Zwerg
$(n-2)$	Möglichkeiten für den 3. Zwerg
$\vdots$	$\vdots$
$(n-(n-1))$	Möglichkeiten für den n-ten Zwerg

Fakultät

Def: Das Rechensymbol “!”

$$n! := n \cdot (n-1) \cdot (n-2) \cdot \cdots \cdot 3 \cdot 2 \cdot 1$$

bezeichnet die **Fakultät** einer Zahl.
Spezialfälle:

$$0! = 1, \quad 1! = 1, \quad 2! = 2, \quad 3! = 6, \quad 4! = 24 \ \ldots$$

Anwendung: Anzahl der Anordungsmöglichkeiten von n Dingen.

Variation ohne Wiederholung

Wieviel Möglichkeiten gibt es

- **k Zwerge nacheinander aus n Zwergen auszusuchen?**

$$n_{\text{ges}} = n(n-1)(n-2)\cdots(n-k+1) =: \frac{n!}{k!}$$

Kombination ohne Wiederholung

Wieviel Möglichkeiten gibt es

- k **Zwerge aus** n **Zwergen auszusuchen?**
  Es gibt $k!$ Möglichkeiten die k ausgesuchten Zwerge in eine Reihenfolge zu bringen.
  Also wird jede Möglichkeit die k-Zwerge auszusuchen von der Formel $\frac{n!}{(n-k)!}$ dann $k!$-mal gezählt:

$$n_{\text{ges}} = \frac{n!}{k!(n-k)!} =: \begin{pmatrix} n \\ k \end{pmatrix}$$

Binomialkoeffizient

Def: Das Rechensymbol

$$\begin{pmatrix} n \\ k \end{pmatrix} := \frac{n!}{k!(n-k)!}$$

wir als **Binomialkoeffizient** bezeichnet und als

$$k \text{ aus } n$$

gelesen.

Spezialfälle:

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = 1, \quad \begin{pmatrix} n \\ 0 \end{pmatrix} = \begin{pmatrix} n \\ n \end{pmatrix} = 1, \quad \begin{pmatrix} n \\ 1 \end{pmatrix} = n, \ldots$$

Anwendung: Anzahl Auswahlmöglichkeiten von k aus n

Teilmengen

Wieviel Möglichkeiten gibt es

- **Zwerge aus n Zwergen auszusuchen?** Für jeden der n Zwerge sucht man sich "dabei" oder "nicht dabei" aus.
  Also nach "Variation mit Wiederholung":

$$n_{\text{ges}} = 2^n$$

  da wir n mal aus $\{dabei, nichtdabei\}$ ziehen.

Wie lernt man das?

- diese Woche noch!!!
- mit einer Packung Smarties oder bunten Büroklammern.
- 10. Klasse aber schon vergessen?
- in der Vorlesung trivial, im Testat vergessen!!!

Wozu benutzt man das?

z.B. bei der Binomialverteilung:
Wieviele Möglichkeiten gibt es um zwei (kaputte) Triebwerke aus 4 Triebwerken auszusuchen?

$$\text{Antwort: } \binom{4}{2}$$

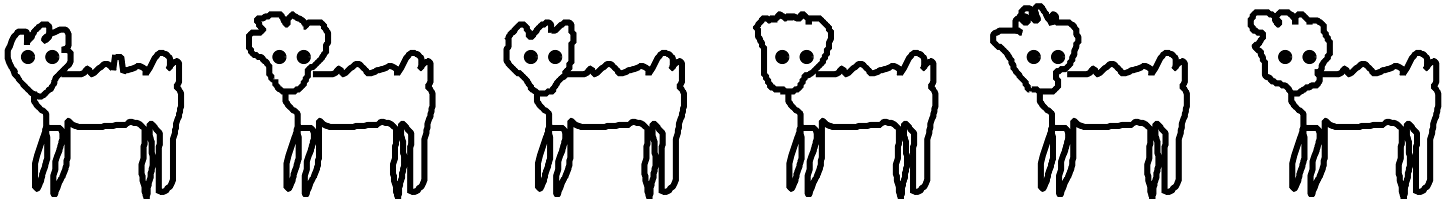
Wie wahrscheinlich ist es das genau diese zwei kaputt sind:

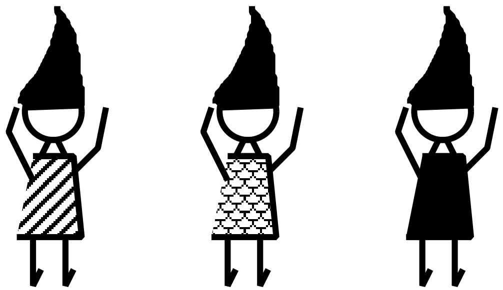
$$p \cdot p \cdot (1-p) \cdot (1-p)$$

Also insgesamt:

$$P(\text{genau zwei Triebwerke kaputt}) = \binom{4}{2} p^2 (1-p)^{4-2}$$

Jetzt wollen wir Schafe aufteilen





Schafe

Wieviele Möglichkeiten gibt es:

- n-Schafe auf k Zwerge aufzuteilen?
- n-Schafe auf k Zwerge so aufzuteilen, dass Zwerg i genau n_i Schafe bekommt.
- n-Eier auf k Zwerge so aufzuteilen, dass Zwerg i genau n_i Eier bekommt.
- n-Eier auf k Zwerge aufzuteilen?

Zerlegung

Wieviele Möglichkeiten gibt es:

- n-Schafe auf k Zwerge aufzuteilen (wenn ich die Schafe kenne)?

$$n_{\text{ges}} = k^n$$

Permutation mit Wiederholung

Wieviele Möglichkeiten gibt es:

- n-Schafe auf k Zwerge so aufzuteilen, dass Zwerg i genau n_i Schafe bekommt.

$$n_{\text{ges}} = \frac{n!}{n_1!n_2!\cdots n_k!} =: \left(\begin{array}{c} n \\ n_1, n_2, \ldots, n_k \end{array} \right)$$

Wie ein Ei dem anderen.

Wieviele Möglichkeiten gibt es:

- n-Eier auf k Zwerge so aufzuteilen, dass Zwerg i genau n_i Eier bekommt.

$$n_{\text{ges}} = 1$$

Kombination mit Wiederholung

Wieviele Möglichkeiten gibt es:

- n-Eier auf k Zwerge aufzuteilen?

$$n_{\text{ges}} = \binom{n+k-1}{n}$$

Beispiel I Kombinatorik

- Bei einer Stichprobe zur Qualitätskontrolle werden 4 von 10 Bauteilen überprüft. (2 defekt)
- Wieviele Möglichkeiten gibt es, dass 2 der 10 Bauteile kaputt sind?
- Wieviele Möglichkeiten gibt es, dass wir die zwei kaputten nicht ziehen?

$$\binom{4}{0}\binom{6}{2} = 1 \cdot \frac{6 \cdot 5}{2 \cdot 1} = 15$$

- Wie hoch ist also die Wahrscheinlichkeit kein defektes Teil zu sehen?

$$P(\text{kein defektes Teil entdeckt}) = \frac{15}{45} \approx \frac{1}{3}$$

Zufallsvariablen

Zufallsvariablen

- **Zufallsvariable** X
  Eine Wert den es noch gar nicht gibt.
- **Wertebereich** Ω_X
  Menge der möglichen Werte von X.
- **Realisierung** x
  Eine Wert der es geworden ist.
- **Verteilung** P^X
  Eine Beschreibung aller Wahrscheinlichkeiten, von Ereignissen, die sich als Aussagen über X formulieren lassen.

Zufallsvariablen mit diskrete Verteilungen

Beispiel I:Hauptantrieb der Simplon

Der Hauptantrieb der Simplon besteht aus 4 Raketentriebwerken.
Mindestens 3 müssen funktionieren, um die Simplon in die Umlaufbahn zu bringen.
Zufallsvariable:

$$X = \text{Anzahl der defekten Triebwerke}$$

Wertebereich:

$$\Omega_X = \{0, 1, 2, 3, 4\}$$

Skala: Anzahl
Realisierung: z.B.

$$x = 0$$

Messbare Ereignisse

Def (vereinfacht): Ein Ereignis heißt X-messbar, falls es genügt die Realisierung x von X zu kennen, um zu entscheiden, ob das Ereignis eingetreten ist.
Solch ein Ereignisse läßt sich durch eine Menge $A \subset \Omega_X$ der Werte beschreiben, bei dem es eintritt.
Beispiele X-messbarer Ereignisse:

- Hauptantrieb betriebsbereit$= \{0, 1\}$
- Hauptantrieb nicht betriebsbereit$= \{2, 3, 4\}$
- Alle Triebwerke arbeiten $= \{0\}$
- Weniger als drei Triebwerke defekt $= \{0, 1, 2\}$

Verteilung

Die Verteilung P^X von X ist eine Abbildung, die X-messbare Ereignisse auf ihre Wahrscheinlichkeit abbildet:

$$P^X(A) = \text{Wahrscheinlichkeit von } A$$

Probleme:

- Woher bekomme ich P^X?
- Wie beschreibe ich P^X?
- Wie rechne ich P^X aus?

Wie beschreibe ich P^X?

Allgemein: Bei diskreten Verteilungen (d.h. Ω_X ist aufzählbar) genügt es die Wahrscheinlichkeit für einzelne Werte $k \in \Omega_X$ zu kennen:

$$p_k = P(X = k)$$

Diese Werte p_k heißen **Elementarwahrscheinlichkeit.** z.B.

$$\begin{aligned}
p_0 &= 0.7435 = P(\text{0 Triebwerke defekt}) \\
p_1 &= 0.2288 = P(\text{1 Triebwerk defekt}) \\
p_2 &= 0.0264 = P(\text{2 Triebwerke defekt}) \\
p_3 &= 0.00135 = P(\text{3 Triebwerke defekt}) \\
p_4 &= 2.6e-05 = P(\text{4 Triebwerke defekt})
\end{aligned}$$

Wie rechne ich $P^X(A)$ aus?

1. Idee:
Summiere die einzelnen Wahrscheinlichkeiten für die möglichen Werte von X auf:

$$P^X(A) = \sum_{k \in A} p_k$$

Kennen wir also die **Elementarwahrscheinlichkeiten** für X, so auch die Wahrscheinlichkeit für jedes X-messbare Ereignis.

$$P(H) = p_0 + p_1 = 0.7435 + 0.2288 = 0.9722$$

Woher bekomme ich P^X bzw. p_k

Allgemein:

- Literatur
- Deligieren
- Standardverteilungsmodelle für verschiedene Situationen, mit statistisch geschätzten Parametern.
- Statistische Schätzung
- Stochastische Modellbildung

Modellbildung

- Jedes Triebwerk fällt mit Wahrscheinlichkeit p aus.
- Annahme: Die Triebwerke fallen unabhängig voneinander aus.
- Also z.B.
  $P(Nur Triebwerk Nr.1 fällt aus) = p\,\underbrace{(1-p)}_{q}\,\underbrace{(1-p)}_{q}\,\underbrace{(1-p)}_{q}$
- $q = 1 - p = P($Triebwerk A fällt nicht aus.$)$

Binomialverteilung

k aus n

Es gibt

$$\begin{pmatrix} n \\ k \end{pmatrix} = \frac{n!}{k!(n-k)!} = \frac{1 \cdot 2 \cdot \ldots \cdot n}{1 \cdot 2 \cdot \ldots \cdot k \; 1 \cdot 2 \cdot \ldots \cdot (n-k)}$$

Möglichkeiten k Objekte aus n Stück auszusuchen.
Dabei ist:

$$k! = 1 \cdot 2 \cdot \cdots \cdot k = \prod_{i=1}^{k} i$$

Binomialverteilung

Def: Die diskrete Verteilung auf $\Omega_X = \{0, \ldots, n\}$ mit

$$p_k = \binom{n}{k} p^k (1-p)^{n-k}$$

heißt Binomialverteilung für n Wiederholungen zur "Erfolgswahrscheinlichkeit" p.
In Zeichen: $P^X = Bi(n, p)$

Einsatz: Die Anzahl der "Erfolge" bei n unabhängigen Versuchen mit "Erfolgswahrscheinlichkeit" p ist $Bi(n, p)$ verteilt.

Beispiel für Modellierung

Wir würden also annehmen, dass Zufallsvariable

$$X = \text{“Anzahl der ausgefallenen Raktentriebwerke”}$$

binomialverteilt ist mit $n = 4$ Triebwerken und einer Einzelausfallwahrscheinlichkeit von

$$p = \text{“Wahrscheinlichkeit für ein Triebwerk auszufallen”}$$

und schreiben

$$X \sim Bi(4, p)$$

Wie bestimmen wir p?

Bestimmung von p

Es gibt verschiedene Wege p zu bestimmen:

- Weitere Modellierung des Triebwerks.
- “Erfahrung”/ “Einschätzung”
- Statistische Schätzung:

$$\hat{p} = \frac{\text{Anzahl Triebwerksausfälle}}{\text{Anzahl Versuche}}$$

Geht nur, wenn wir schon “vergleichbare” Tests gemacht haben.
Problem: Ungenau

Beispiel

Der Triebwerkshersteller hat das Triebwerk getestet und gibt die Ausfallwahrscheinlichkeit mit $p \leq \frac{1}{4}$ an.
Wir modellieren also unseren “worst case” mit:

$$p = P(\text{Triebwerk fällt aus}) = \frac{1}{14}$$

Dann sind wegen $p_i = \left(\begin{array}{c} 4 \\ i \end{array}\right) p^i (1-p)^{4-i}$

$$p_0 = 1 \cdot \frac{13^4}{14^4}, p_1 = 4 \cdot \frac{13^3}{14^4}, p_2 = 6 \cdot \frac{13^2}{14^4}, p_3 = 4 \cdot \frac{13^1}{14^4}, p_4 = 1 \cdot \frac{13^0}{14^4}$$

und also wegen $P(A) = \sum_{k \in A} p_k$

$$P(H) = P(\{0,1\}) = p_0 + p_1 = 0.97222511453561 \approx 0.97$$

Rückblick

- Wir haben die Anzahl Triebwerksausfälle als **Zufallsvariable** X modelliert.
- Wir haben die Verteilung durch das Standardmodell **“Binomialverteilung” modelliert.**
- Der Hersteller hat einen unbekannten Modellparameter p **geschätzt.**
- Wir haben das **Ereignis** $H =$ “Hauptantrieb funktioniert” als Menge $\{0, 1\}$ der Werte von X geschrieben, bei dem das Triebwerk funktionert.
- Wir haben die **Wahrscheinlichkeit** von H mit den Formeln für die Binomialverteilung berechnen.

Diskrete Verteilungen angeben

Diskrete Verteilungen können angegeben werden durch die Wahrscheinlichkeiten:

$$p_x = P(X = x), x \in \Omega_X$$

z.B. $p_x = \begin{pmatrix} 4 \\ x \end{pmatrix} \left(\frac{1}{14}\right)^x \left(\frac{13}{14}\right)^{4-x}$

Diese heißen Elementarwahrscheinlichkeiten

oder

durch eine eine Verteilungsnotation: z.B.

$$X \sim Bi(4, \frac{1}{14})$$

Reellwertige und stetige Verteilungen

Bsp: Wirkungsgrad des Hauptantriebs

Zufallsvariable:

$$W = \text{“Spezifische Wirkung des Hauptantriebs”}$$

Wertebereich: $\Omega_W = \mathbb{R}_+$
Skala: Positiv Reel
Realisierungen: z.B.

$$w = \frac{10.2345kN\ 523.2s}{14.21t} = 37682.55\frac{Ns}{kg}$$

Interessierende Ereignisse

Interessierendes Ereignis:

$$W \geq W_{\mathrm{krit}}$$

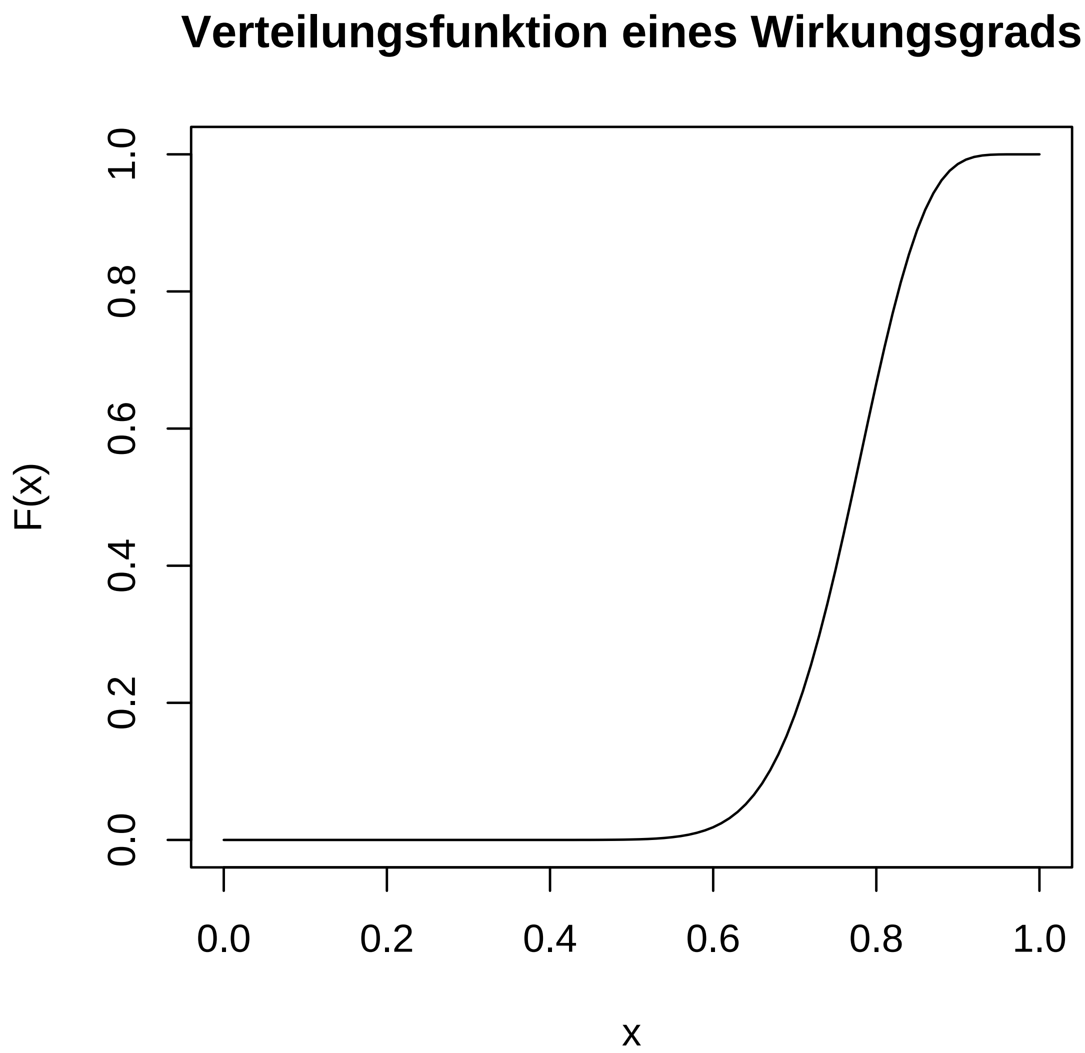
Wobei W_{krit} die minimale spezifische Wirkung ist, um mit der vorhanden Treibstoffmenge die Umlaufbahn zu erreichen.

Verteilungsfunktion

Die Verteilung reller Zufallsvariablen X kann über die Verteilungsfunktion F_X beschrieben werden:

$$F_X(x) := P(X \leq x) = P^X((-\infty, x])$$

Beispiel Verteilungsfunktion I



Verteilungsfunktion

Die Verteilung reller Zufallsvariablen X kann über die Verteilungsfunktion F_X beschrieben werden:

$$F_X(x) := P(X \leq x) = P^X((-\infty, x])$$

Berechnung von Wahrscheinlichkeiten:

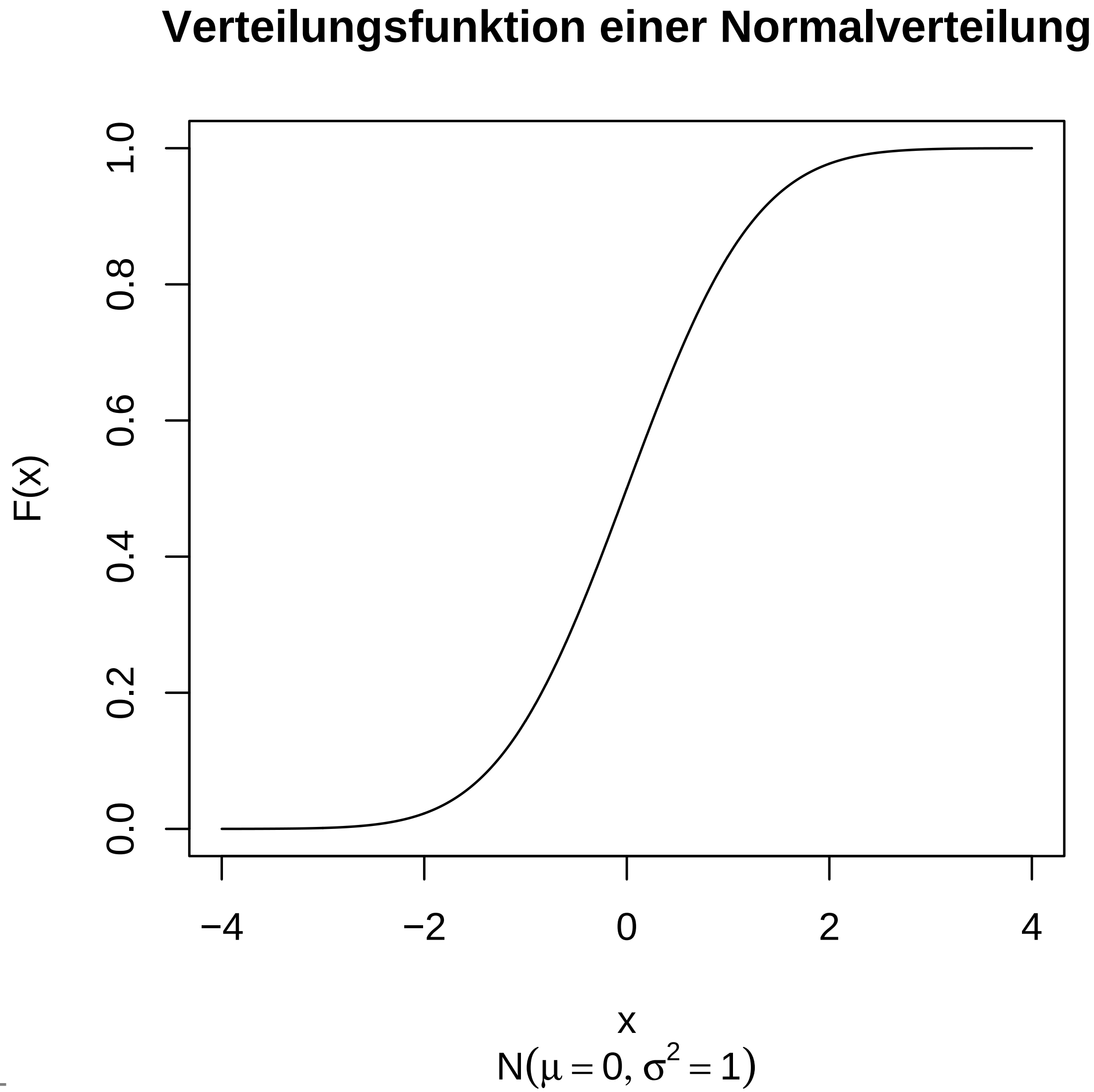
$$P^X((a,b]) = \underbrace{F_X(b)}_{P((-\infty,b])} - \underbrace{F_X(b)}_{P((-\infty,a])}$$

$$P^X((a,b)) = \underbrace{\lim_{\beta \uparrow b} F_X(b)}_{P((-\infty,b))} - \underbrace{F_X(b)}_{P((-\infty,a])}$$

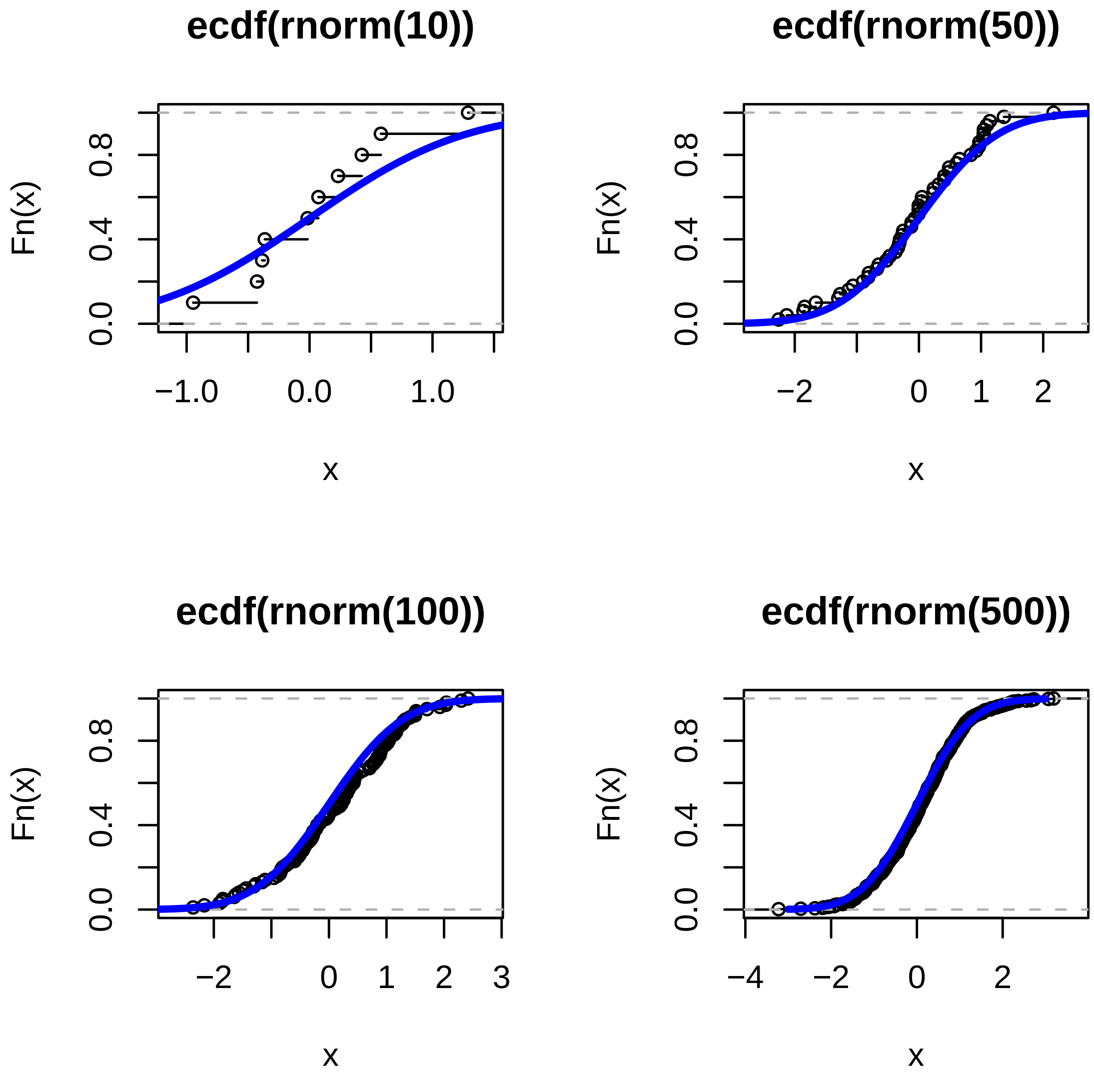
$$P^X([a,b]) = \underbrace{F_X(b)}_{P((-\infty,b])} - \underbrace{\lim_{\alpha \uparrow a} F_X(b)}_{P((-\infty,a))}$$

Beispiel Verteilungsfunktion II

Beispiel Verteilungsfunktion III



Empirische Verteilungsfunktion



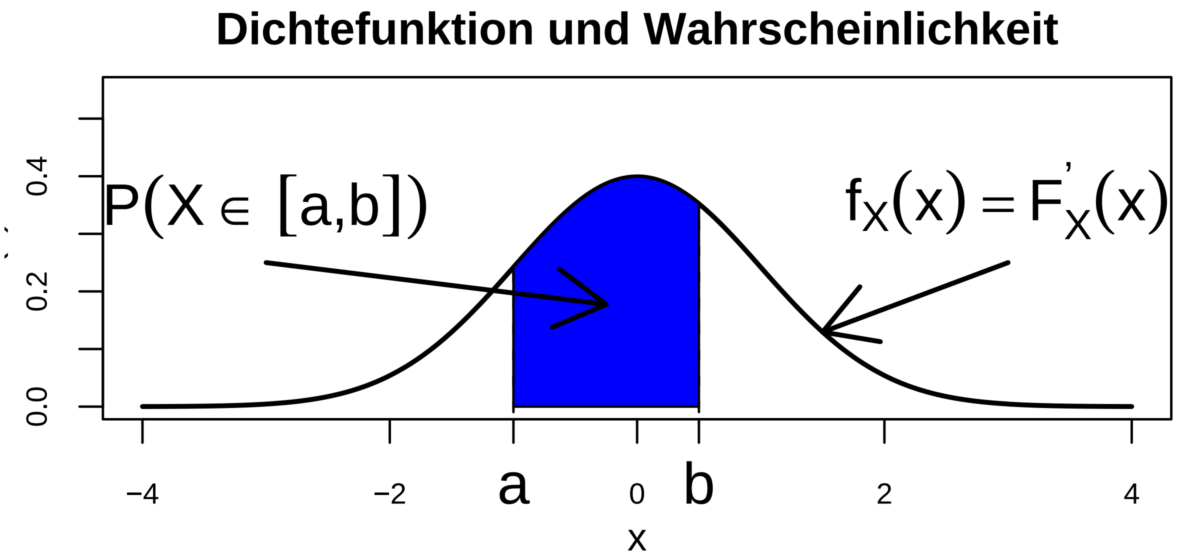
Wahrscheinlichkeitsdichte

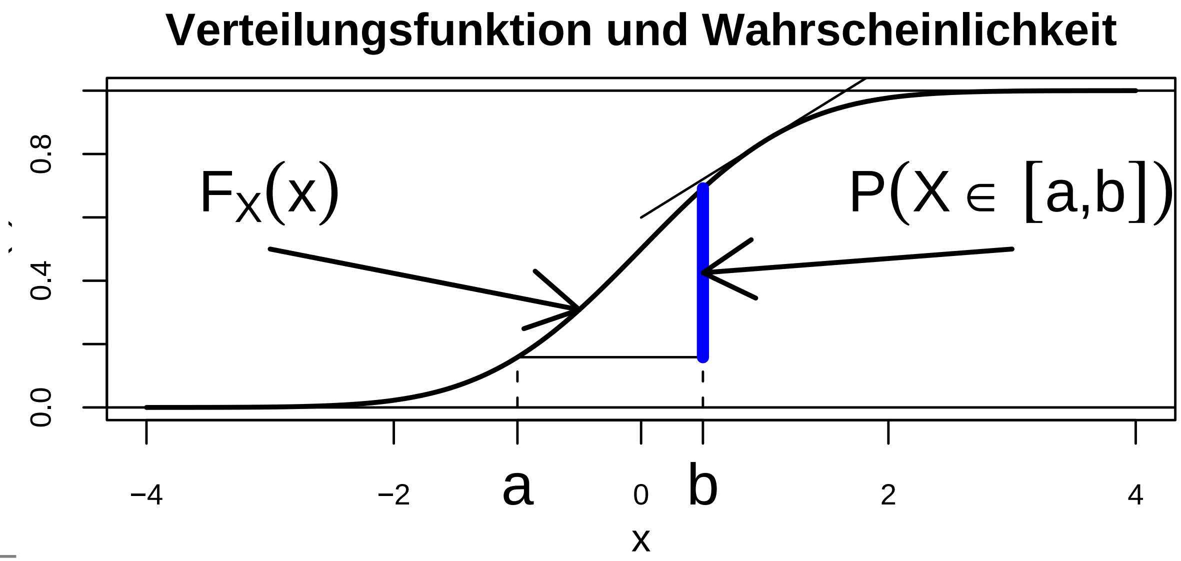
Wahrscheinlichkeitsdichte

Für stetige Verteilungsfunktionen definieren wir:
Def: Die Ableitung $f_X(x) = F'_X(x)$ der Verteilungsfunktion $F_X(x)$ heißt die **Wahrscheinlichkeitsdichte** oder kurz **Dichte** von X.

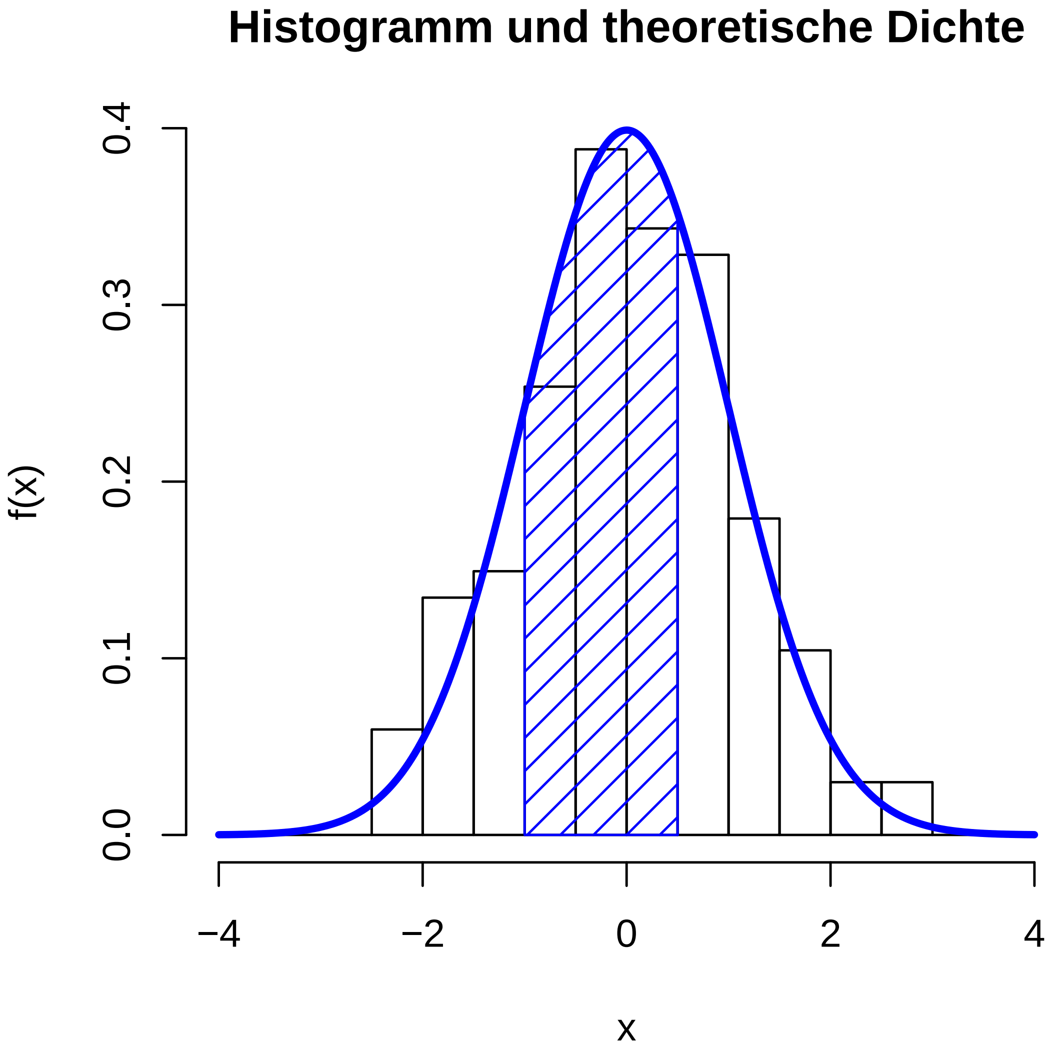
$$f_X(x) = F'_X(x), \quad F_X(x) = \int_{-\infty}^{x} f_X(x)dx$$

Fläche unter der Dichte





Dichte und Histogramm



Kerndichteschätzung

Kerndichteschätzung

$$\hat{f}_X(x) := \sum_{i=1}^{n} \frac{1}{bn} \phi\left(\frac{X_i - x}{b}\right)$$

z.B. mit der Kernfunktion

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

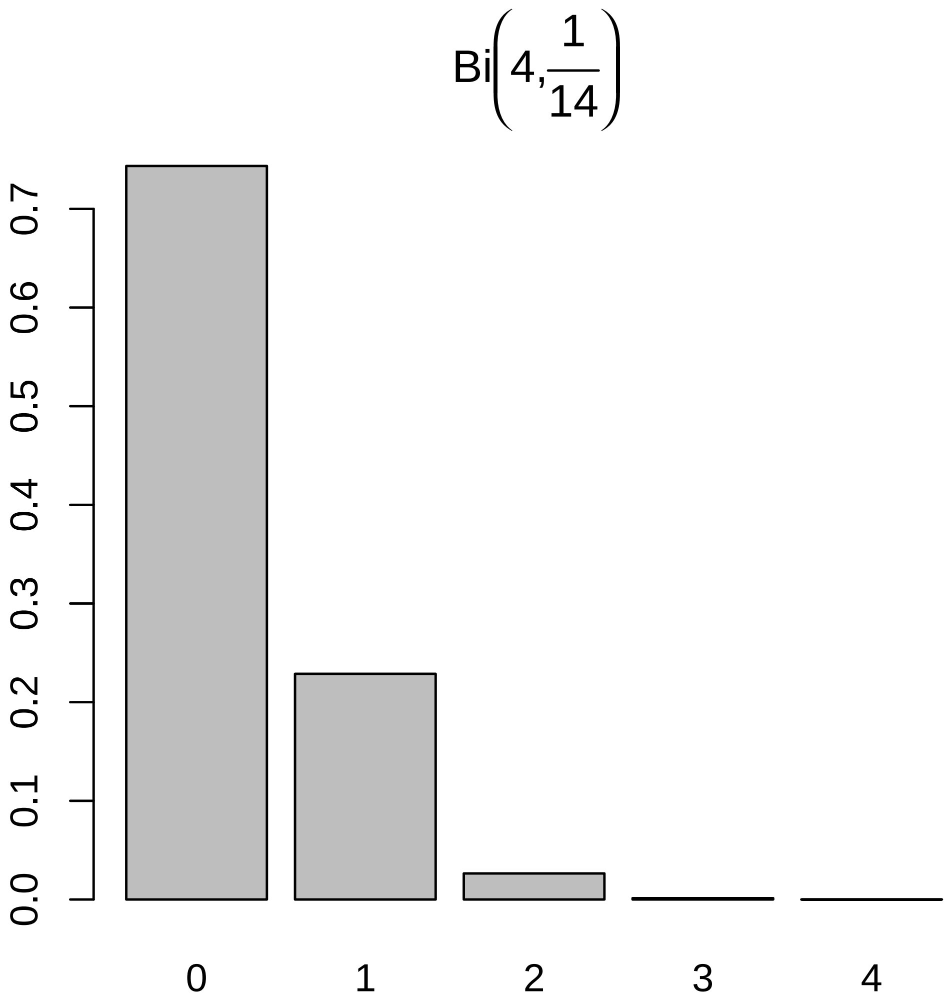
- Die Breite b des Kerns muß richtig gewählt sein.
- Zu schmale Kerne machen “Gewackele”.
- Zu breite Kerne machen “alles platt”.
- Die geschätzten Dichten sind tententiell zu flach.
- Die geschätzten Dichten weisen unsinnige “Strukturen” auf.

Verteilungen beschreiben und zusammenfassen

Beschreiben und zusammenfassen

- Verteilungen lassen sich so ähnlich wie Stichproben graphisch darstellen und durch Parameter zusammenfassen.
- Wieder unterscheidet man nach der Skala der Zufallsvariable.
- Manchmal hat die analoge Technik für Verteilungen einen etwas anderen Namen.
- Will man explizit sagen, dass sich die Größe auf die Verteilung bezieht, so spricht man von der **theoretischen** Größe.
- Will man explizit sagen, dass sich die Größe auf die Stichprobe bezieht, so spricht man von der **empirischen** Größe.

Balkendiagramm f. diskr. Verteilungen



Wahrscheinlichkeiten

```
           0            1            2            3            4
7.434663e-01 2.287589e-01 2.639525e-02 1.353603e-03 2.603082e-05
```

Für dichotome Größen

Der Odd:

$$odd(p_{\text{j}a}) = \frac{P(\text{ja})}{P(\text{nein})} = \frac{p_{\text{j}a}}{1 - p_{\text{j}a}}$$

Wieviel wahrscheinlicher ist es zu gewinnen als zu verlieren.
z.B. für den Funktionieren des einzelnen Triebwerks $p = \frac{13}{14}$:

$$odd\left(\frac{13}{14}\right) = \frac{\frac{13}{14}}{\frac{1}{14}} = 13$$

Es ist dreizehn mal so wahrscheinlich, dass das Triebwerk funktioniert, als das es nicht funktioniert.

Momente:
Erwartungswert, Varianz

Motivation Erwartungswert

Erwartungswert als Mittelwert der Grundgesamtheit:

$$\begin{aligned} E[X] &:= \sum_{x \in \Omega_X} x p_x \\ &= \sum_{x \in \Omega_X} x \frac{\text{“Anzahl x in Grundgesamtheit”}}{n} \\ &= \frac{1}{n} \sum_{\text{Grundgesamtheit}} X \end{aligned}$$

Definition des Erwartungswerts

Def:
Erwartungswert (für diskrete Größen)

$$E[X] = \sum_{x \in \Omega_X} x p_x$$

Erwartungswert (für stetige Größen)

$$\begin{aligned} E[X] &:= \int_{\Omega_X} x f(x) dx \\ &= \lim_{\Delta \downarrow 0} \sum_{x \in \Delta_x \mathbb{N}} x f(x) \Delta_x \\ &= \lim_{\Delta_x \downarrow 0} \sum_{x \in \Delta_x \mathbb{N}} x P([x, x + \Delta_x]) \end{aligned}$$

Beispiel

$X \sim Bi\left(4, \frac{1}{14}\right)$:

$$
\begin{aligned}
E[X] &= 0 \cdot 1 \left(\frac{13}{14}\right)^4 \\
&\quad +4 \left(\frac{13}{14}\right)^3 \left(\frac{1}{14}\right)^1 \\
&\quad +6 \left(\frac{13}{14}\right)^2 \left(\frac{1}{14}\right)^2 \\
&\quad +4 \left(\frac{13}{14}\right)^1 \left(\frac{1}{14}\right)^1 \\
&\quad +1 \left(\frac{1}{14}\right)^4 \\
&= \frac{0 \cdot 1 \cdot 13^4 + 1 \cdot 4 \cdot 13^3 + 2 \cdot 6 \cdot 13^2 + 3 \cdot 4 \cdot 13^1 + 1}{14^4} \\
&= \frac{4}{14}
\end{aligned}
$$

Mittelwert und Erwartungswert

Satz: Starkes Gesetz der großen Zahlen
Sind die X_i stochastisch unabhängig und wie X verteilt, so gilt:

$$P\left(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}X_i = E[X]\right) = 1$$

ist die Wahrscheinlichkeit dafür, dass der Mittelwert einer immer größer werden Stichprobe letztlich gegen den Erwartungswert konvergiert gleich 1.

Transformierter Erwartungswert

Sei T eine Funktion von X so ist $Y = T(X)$ eine neue Zufallsvariable und es gilt:

$$E[T(X)] = \sum_{x \in \Omega_X} T(x) p_x$$

bzw.

$$E[T(X)] = \int_{\Omega_X} T(x) f(x) dx$$

z.B. $T(x) = (x-3)^2$

z.B. $T(x) =$ Treistoffverbrauch pro Sekunde bei x kaputten Triebwer

Rechenregel für $E[X]$

Seien X, Y Zufallsvariablen und α eine relle Zahl:

- $E[\alpha X] = \alpha E[X]$,
- $E[X + Y] = E[X] + E[Y]$

Beispiel

Seien X_i, $i = 1, \ldots, n$ unabhängig und genauso wie X verteilt:

$$\begin{aligned} E[\bar{X}] &= E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \\ &= \frac{1}{n}E\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E\left[X_i\right] \\ &= \frac{1}{n}nE\left[X\right] = E\left[X\right] \end{aligned}$$

Beispiel: Binomialverteilung

Für ein $X_i \sim Bi(1, p)$ gilt:

$$E[X_i] = 0(1-p) + 1p = p$$

Also gilt für $X_1 + X_2 + \ldots + X_n \sim Bi(n, p)$ sofort:

$$\begin{aligned} E[X_1 + X_2 + \ldots + X_n] &= E[X_1] + \ldots + E[X_n] \\ &= p + \ldots + p \\ &= np \end{aligned}$$

Formel für Binomialverteilung

Für $X \sim Bi(n, p)$ gilt:

$$E[X] = np$$

Weil man ja die Anzahl der Ereignisse zählt, die mit Wahrscheinlichkeit p zustande gekommen sind.

Was tut man mit dem Erwartungswert?

- Man beschreibt die Lage der Verteilung (wie Mittelwert)
- Markovungleichung (für positive Zufallsgrößen):

$$P(X \geq g) \leq \frac{E[X]}{g}$$

- Erwartungswerte kann man durch Mittelwerte leicht schätzen.
- Viele Parameter kann man aus geschätzen Erwartungswerten ausrechnen.
- Mit Erwartungswerten kann man leicht rechnen.

Theoretische Quantile

Def: Ein Wert q_p mit der Eigenschaft:

$$\lim_{b \to q} F_X(b) \leq p \leq F_X(q)$$

heißt p-Quantil der Verteilung von X.
Im Klartext:

$$P(X < q_p) = p$$

Theoretische Quantile

Andere Lageparameter

Theoretischer Median

$$\mathrm{med}(X) := q_{0.5}$$

Streuungsparameter

Def: Die (theoretische) **Varianz**:

$$\mathrm{var}(X) = E[(X - E[X])^2]$$

Die (theoretische) **Standardabweichung**:

$$\mathrm{sd}(X) = \sqrt{\mathrm{var}(X)}$$

Die **Mittlere Absolute Abweichung** (mean absolute deviation).

$$mad(X) = E[|X - \mathrm{med}(X)|]$$

Anwendung

Streuparameter, insbesondere die Varianz werden eingesetzt in:

- Wahrscheinlichkeitsabschätzungen
- Parameterschätzung
- Fehlerrechnung
- Für statistische Nachweise in t-Tests und Varianzanalyse

Tschebyschewsche-Ungleichung + ...

Satz:

$$
\begin{aligned}
P(|X - E[X]|^2 \geq \epsilon^2) &\leq \frac{\mathrm{var}(X)}{\epsilon^2} \\
P\left(|X - E[X]| \geq \epsilon\right) &\leq \frac{\mathrm{var}(X)}{\epsilon^2} \\
P(|X - \mathrm{med}(X)| \geq \epsilon) &\leq \frac{\mathrm{mad}(X)}{\epsilon}
\end{aligned}
$$

Markov zum Vergleich: Für $X > 0$

$$P(X \geq \epsilon) \leq \frac{E[X]}{\epsilon}$$

Schätzung der Varianz

Aus einer repräsentativen Stichprobe $X_1, \ldots, X_n$ wird die Varianz üblicherweise geschätzt durch:

$$\hat{\mathrm{var}}(X) = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

Grund: Mit $\frac{1}{n}$ wird es im Mittel zu klein.

$$E[\hat{\mathrm{var}}(X)] = \mathrm{var}(X)$$

$$E\left[\frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2\right] = \frac{n-1}{n} \mathrm{var}(X)$$

Schätzung der anderen Streuparamter

$$\hat{\mathrm{sd}}(X) = \sqrt{\hat{\mathrm{var}}(X)}$$

$$\hat{\mathrm{mad}}(X) = \sum_{i=1}^{n} |X_i - \hat{\mathrm{med}}(X)|$$

Aber

$$E[\hat{\mathrm{sd}}(X)] \neq \mathrm{sd}(X)$$

$$E[\hat{\mathrm{mad}}(X)] \neq \mathrm{mad}(X)$$

Grundlagen zur Schätzung von Parametern

Motivation Schätzung

- Ein konzeptionelles Modell der Vorgänge liefert meist ein Verteilungsmodell.
- Darin sind meist Parameter enthalten deren Wert man nicht kennt.
- Diese Parameter müssen wir aus Daten schätzen.
- Ich will Ihnen mitgeben wie das geht.
- Sie müssen dazu grundsätzlich verstehen, wie statistische Schätzung funktioniert.

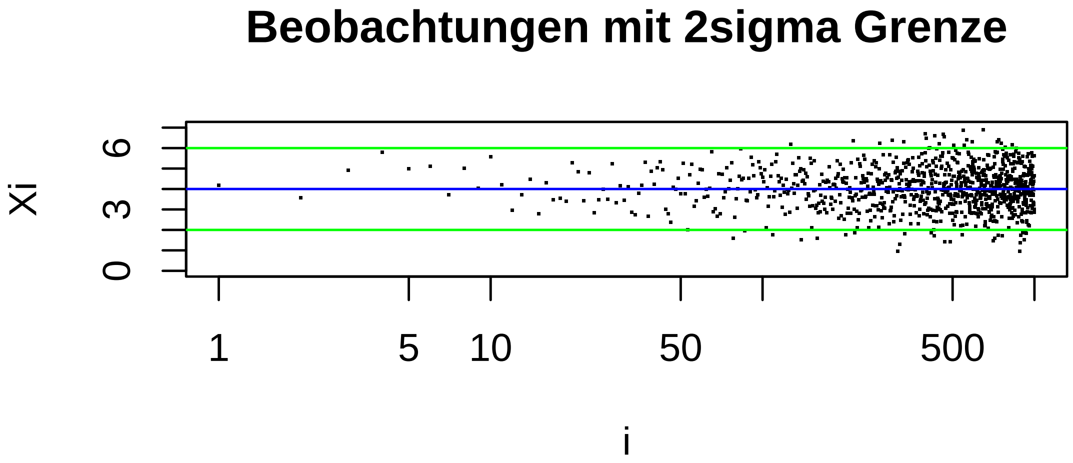
Beispiel: Schätzer

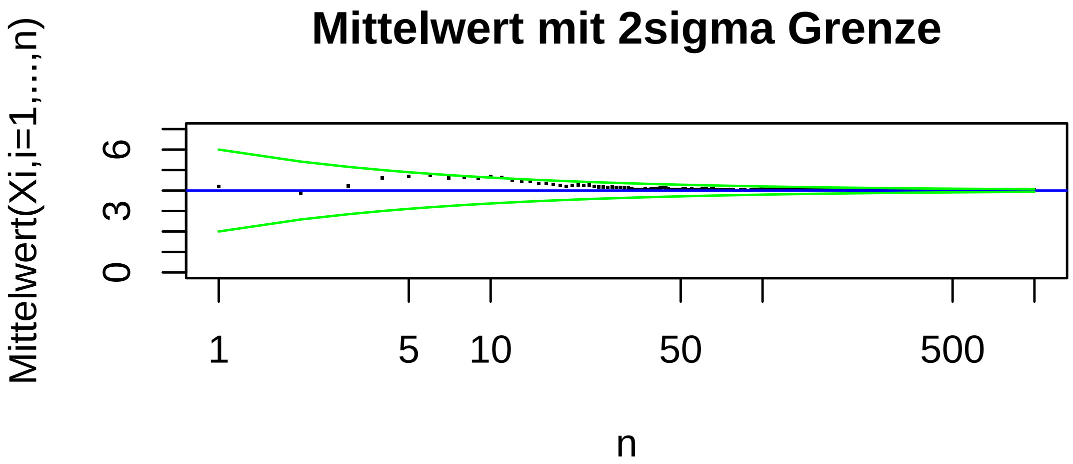
- Der Mittelwert $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ wird als Schätzer für den Erwartungswert $E[X]$ verwendet.
- Der Erwartungswert ist die theoretische Größe, die wir nicht kennen.
- Der Mittelwert ist eine Größe, die wir aus den Daten ausrechnen können.
- Der Mittelwert ist nicht gleich dem Erwartungswert.
- Der Mittelwert konvergiert nach dem starken Gesetz der großen Zahlen gegen den Erwartungswert. (konsistent)
- Der Mittelwert hat den Erwartungswert als Erwartungswert:

$$E[\bar{X}] = E[X]$$

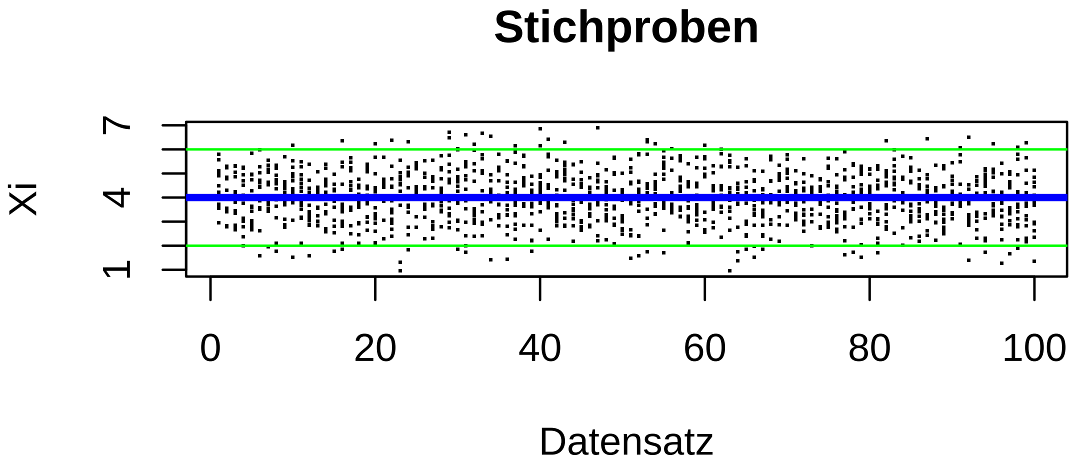
Man schätzt also im Mittel richtig. (erwartungstreu)

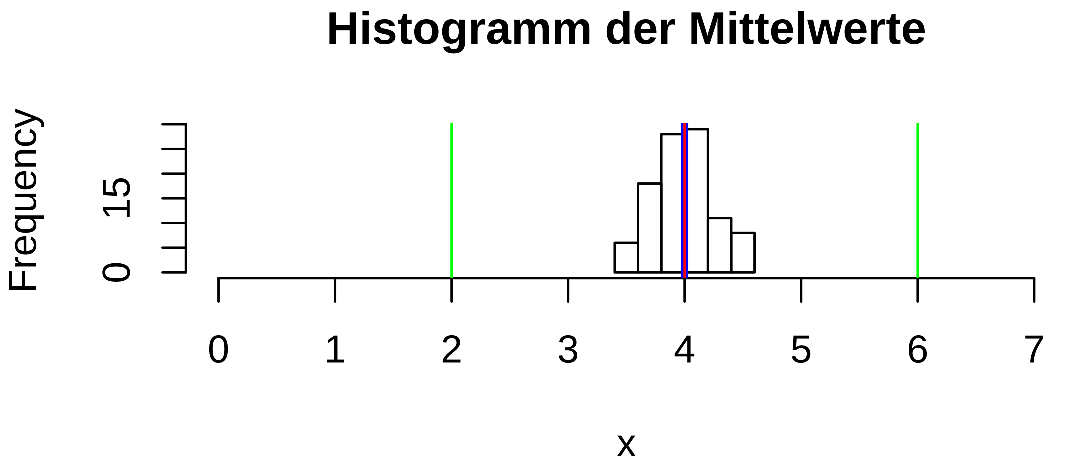
Konsistenz des Mittelwertes





Erwartungstreue des Mittelwertes





Schätzer

- **Def:** Eine Funktion der Daten, welche den Wert eines Parameters ungefähr ermitteln soll heißt Schätzer.
- Ein Schätzer wird mit dem Namen des Parameters und einem Dach darüber notiert: z.B. $\hat{\mu}$, $\widehat{\text{var}}(X)$, $\hat{\sigma}$, $\hat{p}, \hat{n}$
- Der Schätzer ist selbst Zufallsvariable.
- Sein Wert ist zufällig.

Erwartungstreue

Def:
Ein Schätzer $\hat{S}(X_1, \ldots, X_n)$ heißt "**erwartungstreu**" für einen Parameter S wenn gilt:

$$E[\hat{S}(X_1, \ldots, X_n)] = S$$

Konsistenz

Def:
Ein Folge von Schätzern $\hat{S}_n(X_1, \ldots, X_n),\ n = n_0, \ldots, \infty$ heißt "**stark konsistent**" für einen Parameter S wenn gilt:

$$P(\lim_{n\to\infty} E[\hat{S}_n(X_1, \ldots, X_n)] = S) = 1$$

Starkes Gesetz der großen Zahlen

Satz: Für den Mittelwert

$$\bar{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i$$

gilt:

$$E[\bar{X}] = E[X]$$

$$P(\lim_{n\to\infty} \bar{X}_n = E[X]) = 1$$

Voraussetzung: $X_i, i = 1, \ldots, \infty$ sind i.i.d. (repräsentativ) und der Erwartungswert existiert.

Varianzschätzung

Satz: Für den die empirische Varianz

$$\hat{\text{var}} X_n := \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$$

gilt:

$$E[\hat{\text{var}} X] = \text{var}(X)$$

$$P(\lim_{n \to \infty} \hat{\text{var}}_n(X) = \text{var}(X)) = 1$$

Voraussetzung: $X_i, i = 1, \ldots, \infty$ sind i.i.d. (repräsentativ) und Erwartungswert und Varianz existieren.

Konsistenzsatz für $\hat{F}$

Satz: Für die empirische Verteilungsfunktion gilt:

$$E[\hat{F}_n(x)] = F_X(x)$$

$$P(\text{Für alle } x \text{ gilt } \lim_{n \to \infty} \hat{F}_n(x) = F_X(x)) = 1$$

Voraussetzung: $X_i, i = 1, \ldots, \infty$ sind i.i.d. (repräsentativ)

Zusammenfassung

Die folgenden Schätzer sind erwartungstreu und (stark) konsistent:

- Mittelwert für den Erwartungswert
- empirische Varianz für die theoretische Varianz
- empirische Verteilungsfunktion für die Verteilungsfunktion

Rezepte für konsistente Schätzer

Wenn Sie einen Parameter schätzen wollen, können Sie folgendes versuchen:

- Stellen Sie fest, ob sich der Parameter stetige Funktion einer oder mehrer konsistent schätzbarer Größen schreiben läßt.
- Schätzen sie diese Größen konsistent.
- Wenden Sie die Funktion an.
- Sie erhalten eine konsistente Schätzung, da ja

$$P(\lim_{n\to\infty} G(\hat{p}_n) = G(\lim_{n\to\infty} \hat{p}_n) = p) = 1$$

Beispiel

Binomalverteilung $Bi(4, n)$:

$$E[X] = 4p$$

also

$$p = \frac{1}{4}E[X]$$

$$\hat{p} = \frac{1}{4}\bar{X}$$

Ist konsistenter Schätzer.

Wie geht es weiter?

- Welche Standardmodelle gibt es?
- Welches Modell gehört zu welcher Situtation?
- Wie schätzt man die Parameter?
- Wie kann man mit den Modellen weiterrechnen?