Berner Fachhochschule
Haute école spécialisée bernoise
Bern University of Applied Sciences

Wochen 6 und 7: Vertrauensintervalle und Parameterschätzung

WBL 15/17, 01.06.2015

Alain Hauser <alain.hauser@bfh.ch>

Teil V
Zentraler Grenzwertsatz und Vertrauensintervalle

Lernziele

Sie können...

- ...den zentralen Grenzwertsatz erläutern
- ...den Standardfehler des arithmetischen Mittels berechnen
- ...den Unterschied zwischen Standardfehler und Standardabweichung erläutern
- ...ein approximatives Vertrauensintervall für einen Erwartungswert berechnen
- ...eine Binomialverteilung durch eine Normalverteilung approximieren

Vorlesung basiert auf Kapitel 4.6 im Skript

Beispiel: Latenz messen

- Auftrag: durchschnittliche Latenz einer Datenverbindung bestimmen
- Problem: *durchschnittliche* Latenz kann nicht gemessen werden, dafür Latenz einer einzelner Datenübertragung
- Idee: Datenübertragung mehrmals wiederholen, jeweils Latenz messen; Mittelwert aus Messwerten berechnen
- Intuitiv hofft man, dass der berechnete Mittelwert (Stichprobenmittel) bei mehr und mehr Messungen näher an die „wahre“ durchschnittliche Latenz herankommt.

Mathematisches Modell

- Wiederholte Messungen $X_1, X_2, \ldots, X_n$
- $X_1, X_2, \ldots, X_n$ sind **unabhängige und identisch verteilte** Zufallsvariablen („**i.i.d.**“)
- Insbesondere haben alle Messungen denselben Erwartungswert und dieselbe Varianz:

$$\mathcal{E}(X_1) = \ldots = \mathcal{E}(X_n) = \mu, \quad \mathrm{Var}(X_1) = \ldots = \mathrm{Var}(X_n) = \sigma^2$$

- Stichprobenmittel $\overline{X} = \frac{1}{n}(X_1 + X_2 + \ldots X_n)$ ist auch eine Zufallsvariable
- Aus den Rechenregeln für Erwartungswert und Varianz wissen wir bereits:

$$\mathcal{E}(\overline{X}) = \mu, \quad \mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n}$$

- Vgl. Aussage aus Kapitel „Deskriptive Statistik“: $\overline{X}$ ist ein konsistenter Schätzer für μ.

Standardfehler des arithmetischen Mittels

- Das arithmetische Mittel aus wiederholten Messungen $\overline{X}$ ist ein Schätzer für den Erwartungswert μ einer Zufallsvariablen
- Je kleiner die Varianz von $\overline{X}$ ist, desto präziser wird die Schätzung. Mit wachsender Stichprobengrösse n wird die Varianz von $\overline{X}$ kleiner.

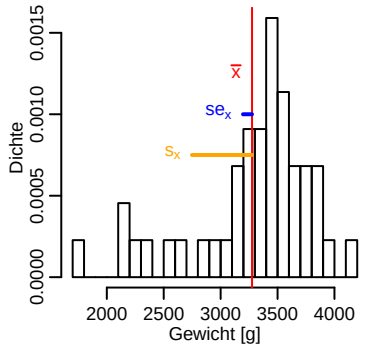
Definition (Standardfehler des arithmetischen Mittels)

Die Grösse $\mathrm{se}_x = \frac{s_x}{\sqrt{n}}$ heisst **Standardfehler des arithmetischen Mittels** (s_x bezeichnet die empirische Standardabweichung der Stichprobe). se_x ist ein Schätzer für die Standardabweichung des arithmetischen Mittels $\overline{X}$.

Abkürzung für den Standardfehler des arithmetischen Mittels: **SEM** (“standard error of the mean”)

Beispiel: Standardfehler eines durchschnittlichen Geburtsgewichts

Datensatz mit Geburtsgewichten von 44 Neugeborenen (in g):



Mittelwert $\overline{x} = 3276\,\mathrm{g}$

emp. St.abw. $s_x = 528\,\mathrm{g}$

SEM $\mathrm{se}_x = 80\,\mathrm{g}$

Wenn die Stichprobengrösse n wächst, konvergiert ...

- ... $\overline{x} \to \mathcal{E}(X)$,
- ... $s_x \to \sigma(X)$,
- ... $\mathrm{se}_x \to 0$.

Der zentale Grenzwertsatz

- Wir können sogar mehr als bloss Erwartungswert und Varianz des Stichprobenmittels $\overline{X}$ bestimmen.
- Der **zentrale Grenzwertsatz** besagt, dass $\overline{X}$ für grosse Stichprobengrössen n näherungsweise normalverteilt ist (mit Erwartungswert μ und Varianz $\frac{\sigma^2}{n}$)
- Gleichwertige Aussage:

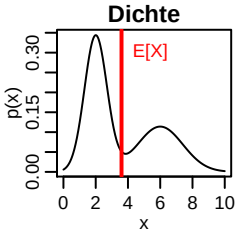
Theorem (Zentraler Grenzwertsatz)

X sei eine Zufallsvariable mit Erwartungswert μ und Varianz σ^2, und $X_1, \ldots, X_n$ sind i.i.d. Kopien davon. Dann ist

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \approx \mathcal{N}(0,1) \text{ für grosse } n\,.$$

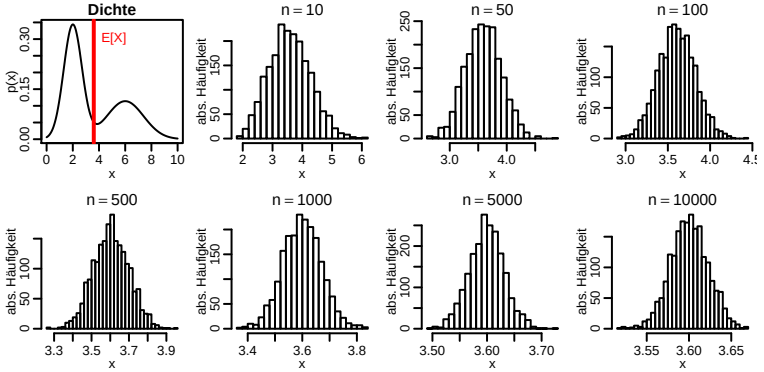
Simulation zum zentralen Grenzwertsatz

- n unabhängige Messungen simulieren; Verteilung einer Einzelmessung durch Dichte rechts gegeben (klar nicht normalverteilt!)
- Mittelwert der n Messwerte berechnen
- Simulation jeweils 2000 mal wiederholen, Histogramm der 2000 berechneten Stichprobenmittel zeichnen



Zentraler Grenzwertsatz: Illustration

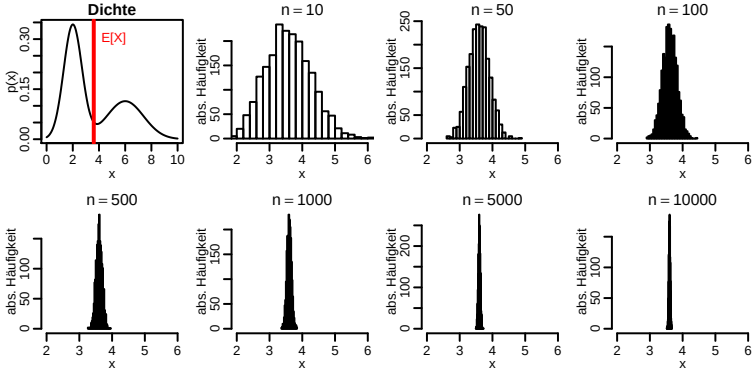
Die Verteilung des Stichprobenmittels gleicht mehr und mehr einer Normalverteilung:



Histogramme: empirische Verteilung der Stichprobenmittel für unterschiedliches n

Zentraler Grenzwertsatz: Illustration

Die Verteilung des Stichprobenmittels gleicht mehr und mehr einer Normalverteilung:



Histogramme: empirische Verteilung der Stichprobenmittel für unterschiedliches n

ZGS: was heisst „näherungsweise normalverteilt“? (I)

- Z_n bezeichne das *standardisierte* Stichprobenmittel

$$Z_n = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

- Mathematische präzise Formulierung des zentralen Grenzwertsatzes: für alle reellen Zahlen z gilt

$$\lim_{n\to\infty} P(Z_n \leq z) = \lim_{n\to\infty} F_{Z_n}(z) = \Phi(z) \; ;$$

  $\Phi(z)$: kumulative Verteilungsfunktion von $\mathcal{N}(0,1)$
- In Worten: die kumulative Verteilungsfunktion des standardisierten Stichprobenmittels nähert sich mit wachsender Stichprobengrösse mehr und mehr der kumulativen Verteilungsfunktion der Standard-Normalverteilung an.
- Für die mathematisch Interessierten: man nennt das „schwache Konvergenz“

ZGS: was heisst „näherungsweise normalverteilt“? (II)

- Die kumulative Verteilungsfunktion des standardisierten Stichprobenmittels nähert sich mit wachsender Stichprobengrösse mehr und mehr der kumulativen Verteilungsfunktion der Standard-Normalverteilung an.
- Konvergenz ist nur für kumulative Verteilungsfunktion wahr, nicht für Dichte: u.U. ist $\overline{X}$ diskret und hat keine Dichte!
- Der ZGS ist tatsächlich auch für wiederholte Messungen diskreter Zufallsvariablen richtig! (Anwendung: Normalapproximation einer Binomialverteilung)

Anwendungen des zentralen Grenzwertsatzes

- (Approximative) Vertrauensintervalle für den Erwartungswert
- Normalapproximation einer Binomialverteilung

Beispiel (Forts.): Latenz messen

- Bestimmung der durchschnittlichen Latenz einer Datenverbindung: wiederholte Messungen durchführen, Stichprobenmittel Schätzwert für Durchschnitt verwenden
- Das ist ein sogenannter **Punktschätzer**: ein einzelner Schätzwert ohne Genauigkeitsangabe
- Unsere Intuition wird mathematisch bestätigt: je mehr Messungen wir durchführen, desto näher liegt der Punktschätzer am wahren Durchschnitt
- Die Genauigkeit einer Schätzung können wir mit einem **Vertrauensintervall** angeben: das ist ein Bereich, in dem der wahre Durchschnitt mit einer gewissen Sicherheit liegt.

Vertrauensintervalle konstruieren I

- Modell für Messungen: $X_1, X_2, \ldots, X_n$: i.i.d. Messungen (z.B. der Latenz) mit $E(X_i) = \mu$ und $\mathrm{Var}(X_i) = \sigma^2$.
- Zentraler Grenzwertsatz besagt, dass $\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}$ näherungsweise standard-normalverteilt ist
- Das heisst: $\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}$ liegt mit 95% Wahrscheinlichkeit zwischen dem 2.5%- und dem 97.5%-Quantil der Standard-Normalverteilung (also zwischen $\Phi^{-1}(0.025) = -1.96$ und $\Phi^{-1}(0.975) = 1.96$)
- Konsequenz: μ liegt mit 95% Wahrscheinlichkeit zwischen $\overline{X} - 1.96\frac{\sigma}{\sqrt{n}}$ und $\overline{X} + 1.96\frac{\sigma}{\sqrt{n}}$

Vertrauensintervalle konstruieren II

- μ liegt mit 95% Wahrscheinlichkeit zwischen $\overline{X} - 1.96\frac{\sigma}{\sqrt{n}}$ und $\overline{X} + 1.96\frac{\sigma}{\sqrt{n}}$
- Problem: wahre Standardabweichung σ ist nicht bekannt.
- Wir können aber Unter- und Obergrenze dieses Bereichs (des Vertrauensintervalls!) abschätzen, indem wir σ durch die *empirische* Standardabweichung s_x der Stichprobe ersetzen.

Vertrauensintervall für den Erwartungswert

- Der Erwartungswert μ liegt mit (ungefähr) 95% Wahrscheinlichkeit im Intervall $I = \left[\overline{X} - 1.96\frac{s}{\sqrt{n}}, \overline{X} + 1.96\frac{s}{\sqrt{n}}\right]$.
- I heisst **Vertrauensintervall** für μ zum **Konfidenzniveau** 95%.

Definition (Vertrauensintervall für den Erwartungswert)

Ein **Vertrauensintervall** zu einem **Konfidenzniveau** $1 - \alpha$ ist ein Intervall $I \subset \mathbb{R}$ mit der Eigenschaft, dass $P(\mu \in I) = 1 - \alpha$.

Ein Vertrauensintervall für μ zu einem beliebigen Konfidenzniveau $1 - \alpha$ ist gegeben durch

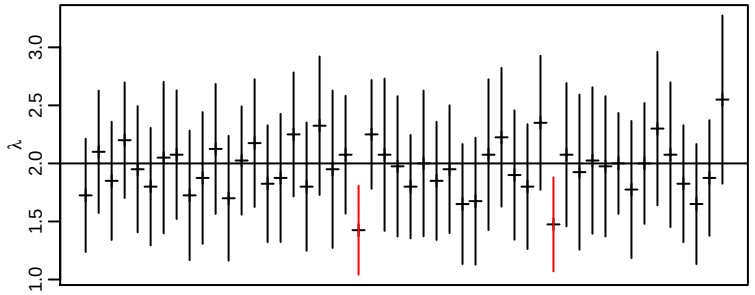
$$\left[\overline{x} - \Phi^{-1}(1 - \alpha/2)\frac{s}{\sqrt{n}}, \overline{x} + \Phi^{-1}(1 - \alpha/2)\frac{s}{\sqrt{n}}\right]$$

Bemerkungen zum Vertrauensintervall

- Wahl des Konfidenzniveaus hängt vom Kontext ab. Für nicht-sicherheitsrelevante Anwendungen verwendet man häufig $1 - \alpha = 0.95$.
- Häufiges Missverständnis: die Aussage $P(\mu \in I) = 1 - \alpha$ bedeutet nicht, dass der Erwartungswert μ eine Zufallsvariable ist!! Zufällig sind die Grenzen des Vertrauensintervalls, da sie mit Hilfe von zufälligen Messwerten berechnet werden.
- *Wenn ich die Stichprobe vergrössere: wird das Vertrauensintervall grösser oder kleiner? Und wenn ich das Konfidenzniveau erhöhe?*

Vertrauensintervalle sind zufällig

Simulation: 50 mal wird eine Poisson-verteilte Stichprobe (mit Parameter $\lambda = 2$) der Grösse 40 gezogen und jedesmal ein 95%-Vertrauensintervall für den Erwartungswert berechnet.



Berechnete Vertrauensintervalle für die 50 Stichproben

Normalapproximation einer Binomialverteilung

- $X \sim \text{Bin}(n, \pi)$ sei eine binomialverteilte Zufallsvariable
- Falls n gross genug ist, können wir deren kumulative Verteilungsfunktion durch eine kumulative Verteilungsfunktion einer Normalverteilung approximieren:

$$X \approx \mathcal{N}(n\pi, n\pi(1-\pi))$$

- Faustregel: Approximation darf verwendet werden, falls $n\pi > 5$ und $n(1-\pi) > 5$
- Beachte: Approximation macht nur Sinn, wenn wir uns für die kumulative Verteilungsfunktion interessieren. Die W'keitsverteilung von X kann natürlich nicht durch eine Dichte approximiert werden!

Teil VI
Verteilungen an Daten anpassen („fitten"): Maximum-Likelihood-Schätzung

Lernziele

Sie können...

- ...die Definition der Likelihood angeben
- ...analytisch einen Maximum-Likelihood-Schätzer für eine einfache Dichte berechnen
- ...die Maximum-Likelihood-Schätzer für die Verteilungen aus Teil II der Vorlesung angeben
- ...einen Q-Q-Plot erzeugen, lesen und interpretieren
- ...für einen gegebene Datensatz eine geeignete Verteilung finden und anpassen

Vorlesung basiert auf Kapitel 3.2 im Skript

Verteilungen an Daten anpassen („fitten")

- Ziel: eine parametrische Verteilung finden, die einen gegebenen Datensatz „gut erklärt"
- Konkreter: eine Verteilungsfamilie (Binomial-, Poisson-, Normalverteilung, etc.) samt zugehören Parametern finden, die sich gut auf einen vorhandenen Datensatz anpassen lässt
- Bisher sind wir immer davon ausgegangen, eine passende Verteilung auf Grund theoretischer Überlegungen zu kennen. Nun wollen wir passende Verteilung *basierend auf Daten* suchen.

Beispiel: Wahrscheinlichkeit schätzen

- Beispiel: wir vermuten, dass im neuen SBB-Fahrplan eine bestimmte Verspätung recht instabil sein könnte. Wir wollen daher die Wahrscheinlichkeit schätzen, dass der betroffene Zug verspätet ist.
- Formales Ziel: Wahrscheinlicht π schätzen, dass ein bestimmter Zug verspätet ist.
- Vorgehen: an zufällig ausgewählten Tagen im Jahr messen wir, ob der betreffende Zug verspätet eintrifft oder nicht.
- Beispiel: von 20 Tagen ist der Zug an 4 Tagen verspätet.
- Intuitiver Schluss: Wahrscheinlichkeit für Verspätung beträgt ungefähr 20%.
- Mathematische Begründung für diesen Schluss: Maximum-Likelihood-Schätzung

Verspäteter Zug: probabilistisches Modell I

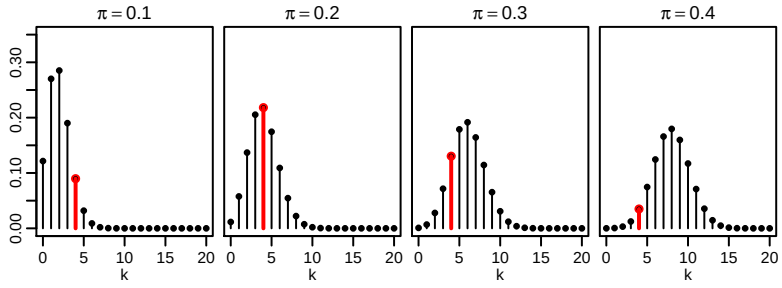
- Binäre Zufallsvariable gibt an, ob Zug verspätet ist:

$$X = \begin{cases} 1, & \text{Zug ist verspätet,} \\ 0, & \text{sonst.} \end{cases} \qquad P(X = 1) = \pi$$

  X ist Bernoulli-verteilt.
- Wir sammeln i.i.d. Daten der Pünktlichkeit an 20 Tagen, stellen z.B. an $k = 4$ Tagen Verspätung fest.
- Wie wahrscheinlich ist es, dass der Zug an k Tagen verspätet ist?

Verspäteter Zug: probabilistisches Modell II

- Wie wahrscheinlich ist es, dass der Zug an k Tagen verspätet ist?
- Die Zufallsvariable $K =$ „Anzahl Tage mit Verspätung“ ist binomialverteilt: $K \sim \text{Bin}(20, \pi)$ (π: W'keit einer einzelnen Verspätung)
- Mögliche Verteilungen abhängig vom (unbekannten) Parameter π:



Verspäteter Zug: Likelihood I

- Bernoulli-Variable gibt an, ob Zug verspätet ist:

$$X = \begin{cases} 1, & \text{Zug ist verspätet,} \\ 0, & \text{sonst.} \end{cases} \qquad P(X = 1) = \pi$$

- Wir messen Verspätung an 20 Tagen $\rightsquigarrow$ „Messwerte“ $x_1, \ldots, x_n$ aus i.i.d. Messungen $X_1, \ldots, X_n$
- Bernoulli-Verteilung kann geschrieben werden als

$$P(X_1 = x_1) = \pi^{x_1}(1 - \pi)^{1 - x_1}$$

- Daher ist **gemeinsame Verteilung** von $X_1, \ldots, X_n$

$$P(X_1 = x_1, \ldots, X_n = x_n) = \pi^k(1 - \pi)^{n - k}$$

(k: Anzahl Tage mit Verspätung)

Verspäteter Zug: Likelihood II

- Die **Likelihood** der Stichprobe ist ihre Wahrscheinlichkeit unter der gemeinsamen Verteilung, *aufgefasst als Funktion des Parameters* π:

$$L(\pi; x_1, \ldots, x_n) = \pi^k (1-\pi)^{n-k}$$

- Die Likelihood ist kein neues Konzept, keine neue Definition. Wir sprechen von der Likelihood (statt von W'keit) um auszudrücken, dass wir die Stichprobe als *fest, gegeben* und den Parameter als *variabel* oder *unbekannt* betrachten. Bisweilen lässt man die Stichprobe aus der Notation weg:

$$L(\pi) = L(\pi; x_1, \ldots, x_n)$$

- **Maximum-Likelihood-Schätzer** für π: $\hat{\pi}$ = Wert des Parameters, der die Likelihood für eine gegebene Stichprobe maximiert.

Maximum-Likelihood-Schätzer für diskrete Verteilung

- Messungen $X_1, X_2, \ldots, X_n$: i.i.d. Kopien einer diskreten Zufallvariablen X mit Wahrscheinlichkeitsverteilung $P(X = x; \theta)$: **parametrisiert** durch θ
- **Likelihood**

$$L(\theta) := P(X_1 = x_1; \theta) \cdot \ldots \cdot P(X_n = x_n; \theta) = \prod_{i=1}^{n} P(X = x_i; \theta)$$

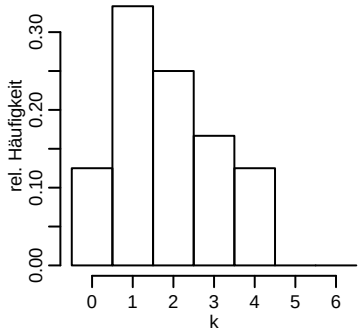
- **Log-Likelihood** $\ell(\theta) := \log(L(\theta)) = \sum_{i=1}^{n} \log(P(X = x_i; \theta))$
- **Maximum-Likelihood-Schätzer** für θ: $\hat{\theta}$ = Wert von θ, der $L(\theta)$ (und damit $\ell(\theta)$) maximiert
- Berechnung: aus praktischen Gründen ist es i.d.R. einfacher, $\ell(\theta)$ zu maximieren statt $L(\theta)$. Funktion $\ell(\theta)$ nach θ ableiten, Ableitung 0 setzen, nach θ auflösen.

Maximum-Likelihood-Schätzer für stetige Verteilung

- Messungen $X_1, X_2, \ldots, X_n$: i.i.d. Kopien einer stetigen Zufallsvariablen X mit Dichte $f(x; \theta)$: **parametrisiert** durch θ
- **Likelihood** $L(\theta) := f(x_1; \theta) \cdot \ldots \cdot f(x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$
- **Log-Likelihood** $\ell(\theta) := \log(L(\theta)) = \sum_{i=1}^{n} \log(f(x_i; \theta))$
- **Maximum-Likelihood-Schätzer** für θ: $\hat{\theta}$ = Wert von θ, der $L(\theta)$ (und damit $\ell(\theta)$) maximiert
- Berechnung: Funktion $\ell(\theta)$ nach θ ableiten, Ableitung 0 setzen, nach θ auflösen.

Beispiel: Geburten-Statistik

- Datensatz: Geburtszeit, Geschlecht und Gewicht von 44 Neugeborenen (Aufzeichnung aller Neugeborener während 24 h)
- Histogramm: Anzahl Geburten pro Stunde



- Geeignete Verteilung, um Daten zu beschreiben?

ML-Schätzer für Poisson-Verteilung

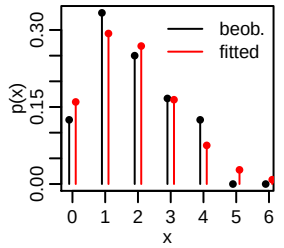
- Stichprobe von Poisson-Verteilung: $x_1, x_2, \ldots, x_n$
- Likelihood der Stichprobe: $L(\lambda) = \prod_{i=1}^{n} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!}$
- Log-Likelihood: $\ell(\lambda) = \sum_{i=1}^{n} \left[x_i \log(\lambda) - \lambda - \log(x_i!) \right]$
- **Maximum-Likelihood-Schätzer**:
  $\hat{\lambda} = \frac{1}{n}(x_1 + x_2 + \ldots x_n) = \frac{1}{n} \sum_{i=1}^{n} x_i = \overline{x}$

Beispiel: Geburten-Statistik

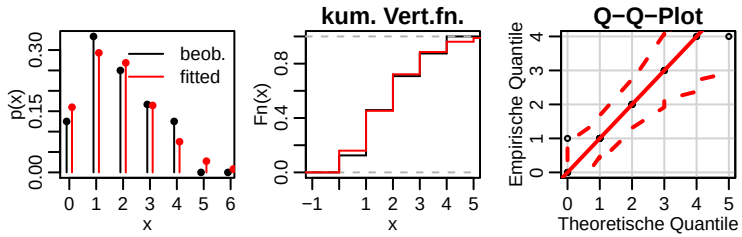
Datensatz:

Stunde i:	0	1	2	3	4	5	6	7	8	9	$\cdots$
Anz. Geburten x_i:	1	3	1	0	4	0	0	2	2	1	$\cdots$

- ML-Schätzer für Parameter λ:
  $\hat{\lambda} = \overline{x} = 1.833$
- $\rightsquigarrow$ Modell: $X \sim \text{Pois}(\hat{\lambda})$ mit $\hat{\lambda} = 1.833$
- *Erklärt die Verteilung unsere Daten gut?*



Q-Q (Quantil-Quantil)-Plot



Je näher die Punkte an der Diagonale des Q-Q-Plots sind, desto besser passen die Daten zur gewählten Verteilung.

Nächste Frage: wie präzise ist das geschätzte $\hat{\lambda}$?

Vertrauensintervall für geschätzten Parameter

- $\hat{\lambda}$ wird als arithmetisches Mittel der Stichprobe $x_1, \ldots, x_n$ berechnet, daher können wir das (approximative) 95%-Vertrauensintervall aus dem letzten Kapitel verwenden:

$$\left[\hat{\lambda} - \Phi^{-1}(0.975) \frac{s_x}{\sqrt{n}}, \hat{\lambda} + \Phi^{-1}(0.975) \frac{s_x}{\sqrt{n}} \right]$$

- Mit (ungefähr) 95% W'keit enthält dieses Intervall den wahren Parameter λ
- Beachte: das Intervall ist zufällig, nicht der Parameter λ!

Beispiel: Geburten-Statistik

- (Approximatives) 95%-Vertrauensintervall für λ:

$$\left[\hat{\lambda} - \Phi^{-1}(0.975)\frac{s_x}{\sqrt{n}}, \hat{\lambda} + \Phi^{-1}(0.975)\frac{s_x}{\sqrt{n}}\right]$$

- In unserem Beispiel:

```
> mean(x)
[1] 1.833333
> sd(x)
[1] 1.239448
> qnorm(0.975)
[1] 1.959964
```

- 95%-Vertrauensintervall für λ: $[1.337, 2.329]$

Publizieren Sie geschätzte Parameter immer zusammen mit Konfidenzintervallen!

Normalverteilung anpassen

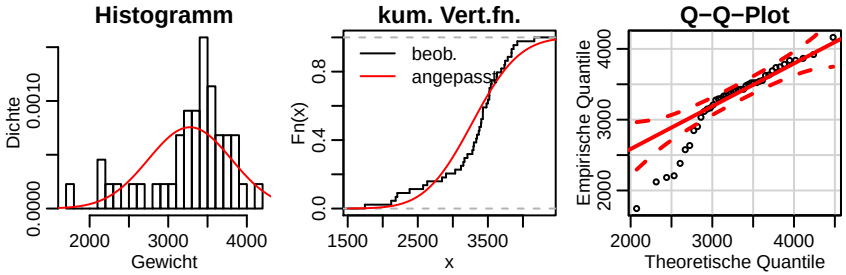
- Normalverteilung $\mathcal{N}(\mu, \sigma^2)$: parametrisiert durch Erwartungswert (μ) und Varianz (σ^2)
- Wir betrachten eine i.i.d. Stichprobe $x_1, x_2, \ldots, x_n$ einer Normalverteilung $\mathcal{N}(\mu, \sigma^2)$
- Wir kennen bereits erwartungstreue Schätzer: $\overline{x}$ für μ, s_x^2 für σ^2
- Maximum-Likelihood-Schätzer

$$\hat{\mu} = \overline{x}, \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

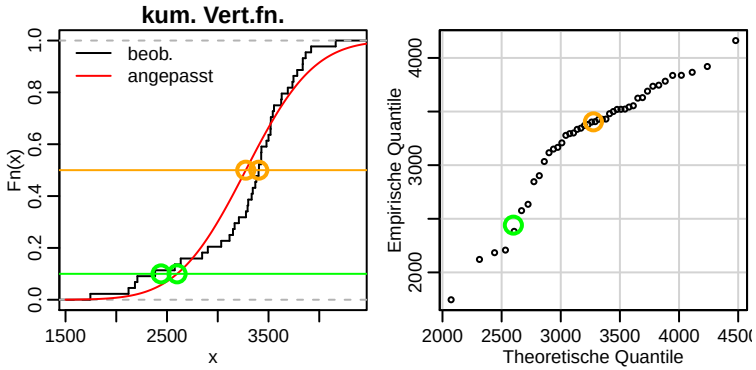
(beachte: $\hat{\sigma}^2$ ist nicht ganz erwartungstreu [leicht "biased"])

Beispiel: Geburten-Statistik

- Zurück zum Beispiel-Datensatz: betrachte X = Geburtsgewicht
- Ist das Gewicht normalverteilt?



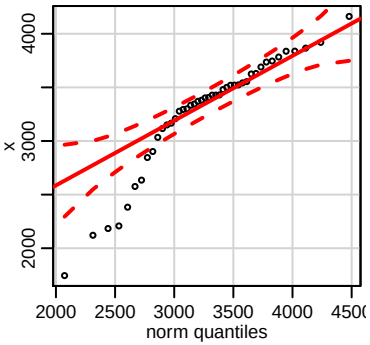
Q-Q-Plots im Detail



Q-Q-Plot: R-Funktionen

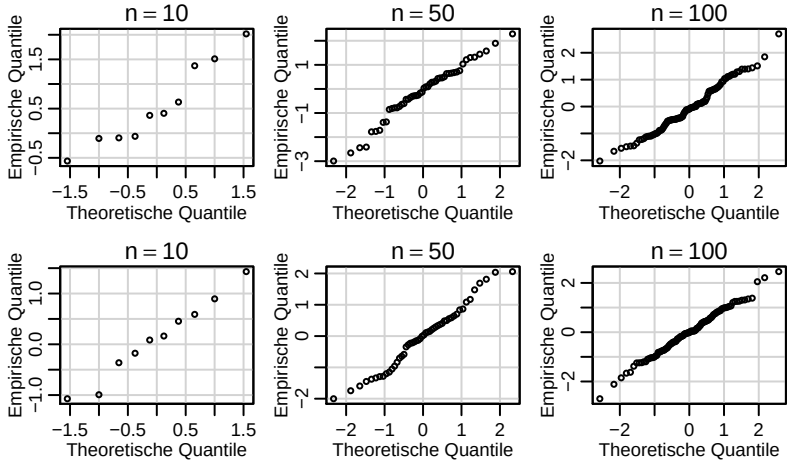
Mit Hilfe des R-Pakets `car` ("Companion to Applied Regression") können Q-Q-Plots erstellt werden:

```
> library(car)
> (est.mean <- mean(x))

[1] 3275.955

> (est.sd <- sd(x))

[1] 528.0325

> qqPlot(x, dist = "norm",
    mean = est.mean, sd = est.sd)
```

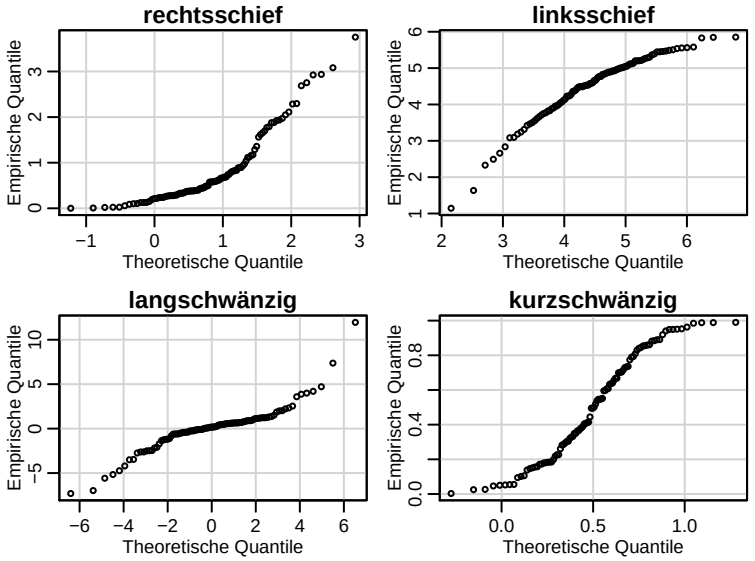


Q-Q-Plots normalverteilter Stichproben

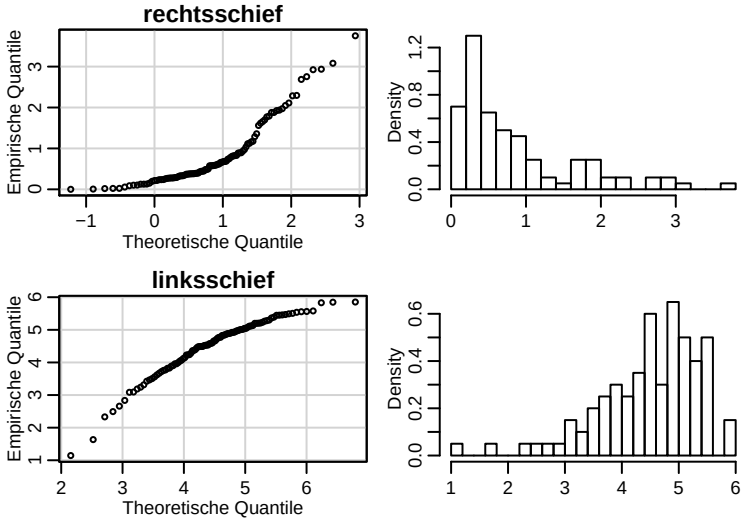
Q-Q-Plots von Stichproben, die aus einer Normalverteilung simuliert wurden:



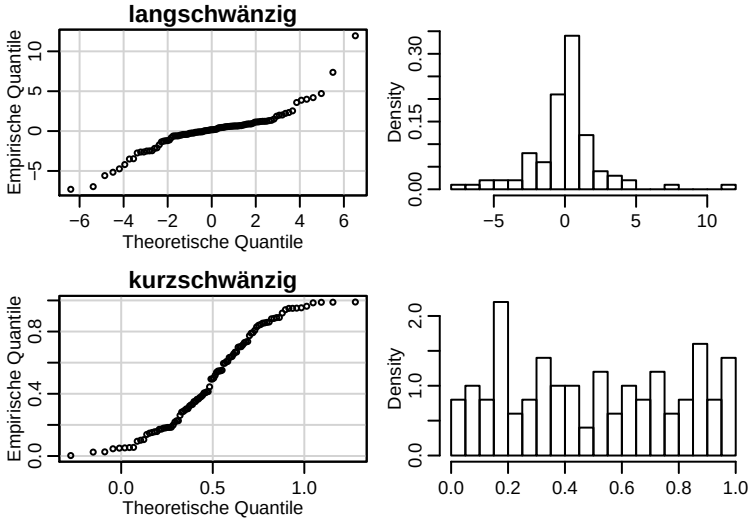
Q-Q-Plots nicht normalverteilter Stichproben I



Q-Q-Plots nicht normalverteilter Stichproben II

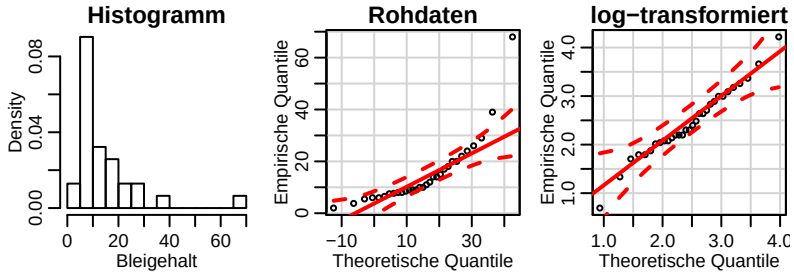


Q-Q-Plots nicht normalverteilter Stichproben III



Rechtsschiefe und Transformationen

- Rechtsschiefe Daten können teilweise durch eine Transformation in eine Normalverteilung überführt werden
- Häufig benutzte Transformationen: log-Transformation, Wurzel-Transformation
- Beispiel: Messungen des Bleigehalts in Granit



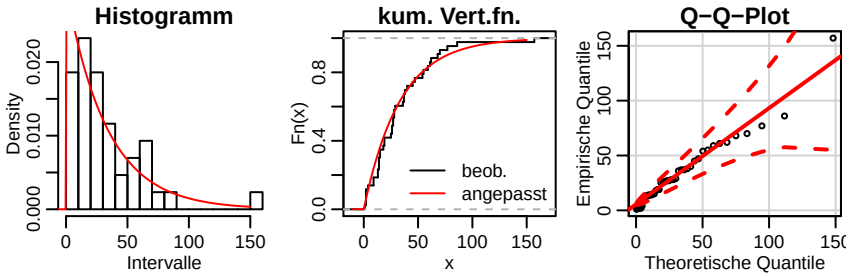
Exponentialverteilung anpassen

- Exponentialverteilung $\text{Exp}(\lambda)$: parametrisiert durch Parameter λ („Rate“)
- Wir betrachten eine i.i.d. Stichprobe $x_1, x_2, \ldots, x_n$ von $\text{Exp}(\lambda)$
- Maximum-Likelihood-Schätzer: $\hat{\lambda} = \dfrac{1}{\bar{x}}$
- (Approximatives) 95%-Vertrauensintervall:

$$\left[\hat{\lambda}\left(1 - \frac{\Phi^{-1}(0.975)}{\sqrt{n}}\right), \hat{\lambda}\left(1 + \frac{\Phi^{-1}(0.975)}{\sqrt{n}}\right)\right]$$

Beispiel: Geburten-Statistik

- Datensatz über Geburten: betrachte T = Zeitintervall zwischen 2 aufeinanderfolgenden Geburten
- Wenn Geburtenzahl in festem Zeitintervall Poisson-verteilt ist, ist Zeitintervall zwischen 2 Geburten exponentialverteilt (ohne Beweis)



Beispiel: Geburten-Statistik

R-Befehle:

```
> (rate <- 1/mean(x))
[1] 0.03006993
> qqPlot(x, dist = "exp", rate = rate)
```