Statistik I für Betriebswirte
Vorlesung 12

Dr. Andreas Wünsche

TU Bergakademie Freiberg
Institut für Stochastik

1. Juli 2019

d) Klumpenstichproben

- Bisher wurde die Grundgesamtheit in Schichten unterteilt, die im Allgemeinen verschieden in Erwartungswerten und Varianzen sind.
- Jetzt betrachten wir eine Einteilung der Grundgesamtheit in M Klumpen, die jeder für sich die Grundgesamtheit schon recht gut repräsentieren.
- **Beispiel 3.4:** Alter der Personen in Vierpersonenhaushalten

Nr.	Vater	Mutter	1. Kind	2. Kind	
1	41	37	15	11	1. Klumpen
2	38	38	18	16	2. Klumpen
3	43	33	10	7	3. Klumpen
4	39	40	18	8	4. Klumpen
	Schicht 1		Schicht 2		

- Weitere Beispiele für Klumpen sind: Bevölkerung in Kreisen für Bevölkerung in BRD, Studenten einer Uni für Studenten in BRD, usw.

Auswertungen im Beispiel 3.4

- Gleichgültig, ob Schichten oder Klumpen: beides sind Zerlegungen der Grundgesamtheit, für die die Zerlegungsformeln für Erwartungswerte und Varianzen gelten, d.h.

$$\mu = \sum_{i=1}^{M} p_i \mu_i \,; \qquad \sigma^2 = \sum_{i=1}^{M} p_i \sigma_i^2 + \sum_{i=1}^{M} p_i (\mu_i - \mu)^2 \ .$$

-

i	Vater	Mutter	1. Kind	2. Kind	μ_i	σ_i^2
1	41	37	15	11	26	173.0
2	38	38	18	16	27.5	110.8
3	43	33	10	7	23.25	231.2
4	39	40	18	8	26.25	188.2
	$\mu_I = 38.625$		$\mu_{II} = 12.875$		$\mu = 25.75$	
	$\sigma_I^2 = 7.7$		$\sigma_{II}^2 = 17.1$			$\sigma^2 = 178.2$

Zufällige Auswahl der Klumpen

- Vorhanden seien M Klumpen.
- N_i sei die Anzahl von Merkmalsträgern im Klumpen i,

  $N = \sum_{i=1}^{M} N_i$ sei der Umfang der kompletten Grundgesamtheit.
- Bei einer Klumpenstichprobe sollen zufällig m der M Klumpen so ausgewählt werden, dass die Wahrscheinlichkeit, dass Klumpen i ausgewählt wird, gleich $p_i = \frac{N_i}{N}$ $(i = 1, \ldots, M)$ ist.
- Dann erfolgt eine Totalerhebung in den ausgewählten m Klumpen, d.h. die exakte Ermittlung des dortigen Mittelwertes μ_i. Damit liegt auch hier wieder eine eingeschränkte Zufallsauswahl vor.
- Der Stichprobenumfang ist hier im Allgemeinen zufällig.
- Die Auswahl der m Klumpen kann zum Beispiel mit Hilfe von Zufallszahlen (ZZ) erfolgen.

Schätzung von μ bei einer Klumpenstichprobe I

- Wir betrachten ein Untersuchungsmerkmal X und identifizieren dieses wieder mit einer Zufallsgröße (der Wert der Zufallsgröße entspricht dem Merkmalswert eines zufällig ausgewählten Merkmalsträgers). Zu schätzen ist der Erwartungswert $\mathbf{E}X = \mu$.
- Innerhalb der m ausgewählten Klumpen erfolgt eine Totalerhebung

  $\Rightarrow$ der Erwartungswert $\mu_i = \dfrac{1}{N_i} \displaystyle\sum_{j=1}^{N_i} x_{ij}$ in einem ausgewählten Klumpen mit Index i kann exakt bestimmt werden.
- Es seien $i_1, \ldots, i_m$ die Indizes der m ausgewählten Klumpen.
- Jeder ausgewählte Index i_k kann als Realisierung einer Zufallsgröße I_k betrachtet werden, die die Werte $i = 1, ..., M$ jeweils mit den Wahrscheinlichkeiten p_i annimmt.
- Die Zufallsgrößen I_k werden als unabhängig und identisch verteilt angenommen (gilt eigentlich nur bei Ziehen mit Zurücklegen).

Schätzung von μ bei einer Klumpenstichprobe II

- Der Mittelwert $\mu = \mathbf{E}X$ wird geschätzt durch

$$\hat{\mu} = \frac{1}{m} \sum_{k=1}^{m} \mu_{I_k} \,.$$

- Diese Schätzung ist unverzerrt (erwartungstreu), denn es gilt

$$\mathbf{E}\hat{\mu} = \frac{1}{m} \sum_{k=1}^{m} \mathbf{E}\mu_{I_k} = \frac{1}{m} \sum_{k=1}^{m} \sum_{i=1}^{M} p_i \mu_i = \frac{1}{m} \sum_{k=1}^{m} \mu = \mu \,.$$

- Für die Varianz der Schätzung gilt

$$\mathbf{Var}\hat{\mu} = \frac{1}{m} \sum_{i=1}^{M} p_i (\mu_i - \mu)^2 \,.$$

- Die Formel ist korrekt für eine Klumpenziehung mit Zurücklegen, eine Auswahl ohne Zurücklegen führt zu einer noch geringeren Varianz.

Statistische Vorteile von Klumpenstichproben

- Klumpenstichproben sind häufig kostengünstiger als reine Zufallsstichproben vom gleichen Umfang (Reisekosten o.ä.).
- Bei statistischen Untersuchungen muss man berücksichtigen, dass eine Klumpenstichprobe im Allgemeinen einen zufälligen Stichprobenumfang besitzt: $n' = \sum_{k=1}^{m} N_{I_k}$.
- Daher beschränken wir uns beim Vergleich mit dem Fall einer reinen Zufallsauswahl auf den Spezialfall $N_i = \frac{N}{M} = \text{const}$, d.h. $p_i = \frac{1}{M}$, $i = 1, \ldots, M$, und einen Stichprobenumfang $n = m\frac{N}{M}$.
- Für die reine Zufallsauswahl gilt wegen $p_i = \frac{1}{M}$

$$\mathbf{Var}\overline{X} = \frac{\sigma^2}{n} = \frac{M\sigma^2}{Nm} = \frac{1}{Nm}\left(\sum_{i=1}^{M}\sigma_i^2 + \sum_{i=1}^{M}(\mu_i - \mu)^2\right) .$$

Vergleich Klumpenstichprobe und reine Zufallsauswahl I

- Für die Klumpenstichprobe (beim Ziehen mit Zurücklegen) gilt

$$\mathbf{Var}\hat{\mu} = \frac{1}{Mm} \sum_{i=1}^{M} (\mu_i - \mu)^2 \,.$$

- Zum Vergleich: exakte Formel beim Ziehen ohne Zurücklegen

$$\mathbf{Var}\hat{\mu} = \frac{M-m}{M-1} \cdot \frac{1}{Mm} \sum_{i=1}^{M} (\mu_i - \mu)^2 \,.$$

- Absoluter Klumpeneffekt (beim Ziehen mit Zurücklegen)

$$\begin{aligned} \mathbf{Var}\overline{X} - \mathbf{Var}\hat{\mu} &= \frac{1}{mN} \sum_{i=1}^{M} \sigma_i^2 + \left(\frac{1}{mN} - \frac{1}{mM} \right) \sum_{i=1}^{M} (\mu_i - \mu)^2 \\ &= \frac{1}{mN} \left[\sum_{i=1}^{M} \sigma_i^2 - \left(\frac{N}{M} - 1 \right) \sum_{i=1}^{M} (\mu_i - \mu)^2 \right] . \end{aligned}$$

Vergleich Klumpenstichprobe und reine Zufallsauswahl II

$\Rightarrow$ Klumpenstichproben sind statistisch vorteilhafter als reine Zufallsstichproben, falls

$$\left(\frac{N}{M} - 1\right) \sum_{i=1}^{M} (\mu_i - \mu)^2 < \sum_{i=1}^{M} \sigma_i^2 \,.$$

- Die Bedingung für einen positiven Klumpeneffekt ist auf jeden Fall erfüllt, falls alle Klumpenmittelwerte μ_i gleich sind.
- Der Klumpeneffekt ist umso größer, je weniger die Mittelwerte μ_i schwanken und je größer die Varianz innerhalb der Klumpen ist.
- Im **Beispiel 3.4** vom Anfang gelten $M = 4, m = 1, N = 16$ und

$$\mathbf{Var}\overline{X} = \frac{178.2}{4} = 44.55\,; \quad \mathbf{Var}\hat{\mu} = \frac{1}{4}\sum_{i=1}^{4}(\mu_i - \mu)^2 = 2.41\,;$$

$$\mathbf{Var}\hat{\mu}^* = \frac{1}{2}\left(\sigma_I^2 + \sigma_{II}^2\right) = 12.42 \quad \text{(geschichtete Stichprobe zum Vgl.)}$$

3.2 Parameterschätzungen: Punktschätzungen

- Für alle theoretischen Aussagen wird im Weiteren eine mathematische Stichprobe $X_1, \ldots, X_n$ vorausgesetzt.
- Eine der Grundaufgaben der schließenden Statistik besteht in der Schätzung unbekannter Verteilungsparameter, z.B. μ bzw. σ^2 bei einer Normalverteilung oder p bei einer Binomialverteilung.
- Im Weiteren bezeichnet ϑ den unbekannten Parameter und $\hat{\vartheta}$ eine Schätzung für diesen Parameter.
- Ein Punktschätzer ist eine Stichprobenfunktion:

$$\hat{\vartheta}(X_1, .., X_n) \, .$$

- Für die beobachtete Stichprobe kann man mit Hilfe dieser Stichprobenfunktion einen konkreten Zahlenwert $\hat{\vartheta}(x_1, .., x_n)$, die Punktschätzung (den geschätzten Wert für ϑ,) errechnen.

3.2.1 Punktschätzungen für allgemeine Parameter

a) Schätzung des Erwartungswertes

- Der arithmetische Mittelwert $\overline{X}$ ist ein Punktschätzer für den Erwartungswert μ:

$$\hat{\mu}(X_1, \ldots, X_n) = \overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \,.$$

- Der arithmetische Mittelwert $\overline{X}$ ist ein erwartungstreuer Schätzer für den Erwartungswert μ:

$$\mathbf{E}(\hat{\mu}(X_1, \ldots, X_n)) = \mathbf{E}\overline{X} = \mu \,.$$

- Die Varianz des Schätzers ist:

$$\mathbf{Var}(\hat{\mu}(X_1, \ldots, X_n)) = \mathbf{Var}\overline{X} = \frac{\sigma^2}{n} \,.$$

Eigenschaften des arithmetischen Mittelwertes

- Damit liefert der arithmetische Mittelwert $\overline{X}$ eine gute Schätzvorschrift für μ, da
  - der Erwartungswert von $\overline{X}$ gleich μ ist (sogenannte „Erwartungstreue des Schätzers $\overline{X}$“) und
  - die Varianz von $\overline{X}$ mit wachsendem Stichprobenunfang n immer kleiner wird und für $n \to \infty$ gegen Null konvergiert.
- Man kann sogar zeigen, dass unter obigen Bedingungen

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

gilt („Gesetz der großen Zahlen“, „Konsistenz des Schätzers $\overline{X}$“).

Wünschenswerte Eigenschaften von Punktschätzungen

i) Erwartungstreue

Die Schätzung $\hat{\vartheta}$ heißt erwartungstreu oder unverzerrt, wenn der Erwartungswert des Schätzers gleich dem zu schätzenden Wert ist, d.h. wenn man im Mittel den Parameter richtig schätzt, $\mathbf{E}\hat{\vartheta} = \vartheta$. Gilt jedoch $\mathbf{E}\hat{\vartheta} \neq \vartheta$, dann heißt der Schätzer $\hat{\vartheta}$ verzerrt und die Differenz $\mathbf{E}\hat{\vartheta} - \vartheta$ heißt systematischer Fehler, Verzerrung oder Bias.

- Wird die Forderung $\mathbf{E}\hat{\vartheta} = \vartheta$ dahingehend abgeschwächt, dass man nur fordert

$$\lim_{n \to \infty} \mathbf{E}\hat{\vartheta} = \vartheta,$$

kommt man zu dem Begriff der asymptotisch erwartungstreuen Schätzung.

Konsistenz einer Punktschätzung

ii) Konsistenz

Die Schätzung $\hat{\vartheta}$ heißt konsistent, wenn sie mit wachsendem Stichprobenumfang n dem zu schätzenden Parameter immer näher kommt. Insbesondere heißt sie schwach konsistent, wenn für beliebige $\varepsilon > 0$ gilt

$$\lim_{n\to\infty} P(|\hat{\vartheta} - \vartheta| \leq \varepsilon) = 1\,.$$

- Hinreichend dafür sind zum Beispiel die Bedingungen
  - $\hat{\vartheta}$ ist (asymptotisch) erwartungstreu und
  - $\lim\limits_{n\to\infty} \mathbf{Var}\,\hat{\vartheta} = 0\,.$
- **Beispiel:** Wegen $\mathbf{Var}\overline{X} = \dfrac{\sigma^2}{n}$ ist für eine Zufallsgröße mit endlicher Varianz σ^2 das Stichprobenmittel $\overline{X}$ ein schwach konsistenter Schätzer für den Erwartungswert $\mathbf{E}X$.

Optimalität

iii) Optimalität

Eine weitere wünschenswerte Eigenschaft einer Schätzung ist ihre Optimalität. Dies besagt, dass die Schätzung in einer Menge von möglichen Schätzungen auf bestimmte Art und Weise optimal ist, zum Beispiel die geringste Varianz besitzt.

- **Beispiel:** Für eine Zufallsgröße X mit endlicher Varianz ist das Stichprobenmittel $\overline{X}$ derjenige lineare erwartungstreue Schätzer für den Erwartungswert $\mathbf{E}X$, der die kleinste Varianz besitzt ("BLUE", "best linear unbiased estimator").

Schätzung einer (unbekannten) Wahrscheinlichkeit p

- In n unabhängigen Versuchen tritt das Ereignis A ein oder nicht. Die Eintrittswahrscheinlichkeit ist $P(A) = p$.
- Damit ist X_i

$$X_i := \begin{cases} 1 & : \text{falls } A \text{ im } i\text{-ten Versuch eintritt} \\ 0 & : \text{sonst} \end{cases} \qquad i = 1, \ldots, n\,,$$

  Bernoulli-verteilt mit Parameter $p = P(A) = P(X_i = 1)$.
- Der Erwartungswert ist: $\mu = \mathbf{E}X_i = p$.
- $H(A) = \sum_{i=1}^{n} X_i$ - absolute Häufigkeit des Ereignisses A.
- Die relative Häufigkeit des Ereignisse A ist eine Punktschätzung für die (unbekannte) Wahrscheinlichkeit $p = P(A)$:

$$\hat{p}(X_1, \ldots, X_n) = \hat{\mu}(X_1, \ldots, X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i = \frac{H(A)}{n}\,.$$

b) Schätzung der Varianz

- Die empirische Varianz S^2 ist ein Punktschätzer für die Varianz σ^2:

$$\hat{\sigma}^2(X_1, \ldots, X_n) = S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2 \,.$$

- Die empirische Varianz S^2 ist ein erwartungstreuer Schätzer für die Varianz σ^2:

$$\mathbf{E} S^2 = \sigma^2 \,.$$

- Die Schätzfunktion

$$\frac{1}{n} \sum_{i=1}^{n} \left(X_n - \overline{X}\right)^2$$

  ist im Gegensatz dazu nur eine verzerrte Schätzung der Varianz. Sie ist jedoch eine asymptotisch erwartungstreue Schätzung der Varianz σ^2.

c) Schätzung des Medians

- Ausgehend von der geordneten mathematischen Stichprobe

$$X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$$

kann der Median geschätzt werden.

- Der Stichprobenmedian ist eine Punktschätzung für den Median

$$\widehat{\text{Median}}(X) = \widetilde{X} = \begin{cases} X_{\left(\frac{n+1}{2}\right)}, & \text{falls } n \text{ ungerade,} \\ \frac{1}{2}\left(X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}\right), & \text{falls } n \text{ gerade.} \end{cases}$$

3.2.2 Schätzung von Parametern spezieller Verteilungen

a) Normalverteilung

Beispiel: Es seien X_i iid. mit $X_i \sim \mathrm{N}(\mu, \sigma^2), \quad i = 1, \ldots, n$.

- $\overline{X}$ ist ein Schätzer für μ, d.h. $\hat{\mu} = \overline{X}$ mit Schätzwert

$$\hat{\mu}(x_1, \ldots, x_n) = \overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \,;$$

- die Stichprobenvarianz S^2 ist ein Schätzer für σ^2, d.h. $\hat{\sigma^2} = S^2$ mit Schätzwert

$$\hat{\sigma^2}(x_1, \ldots, x_n) = s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2 \,.$$

b) weitere Verteilungen

Es seien X_i iid., $i = 1, \dots, n$, mit:

- Exponentialverteilung: $X_i \sim \mathrm{Exp}(\lambda)$

$$\hat{\lambda} = \frac{1}{\overline{x}}\,.$$

- Stetige Gleichverteilung: $X_i \sim \mathrm{U}[0, a]$

$$\hat{a} = \frac{n+1}{n} \cdot \max_{i=1}^{n} x_i = \frac{n+1}{n} \cdot x_{(n)}\,.$$

- Bernoulliverteilung: $X_i \sim \mathrm{B}(p)$

$$\hat{p} = \overline{x} = \frac{\sum\limits_{i=1}^{n} x_i}{n}\,.$$

- Poissonverteilung: $X_i \sim \mathrm{Poi}(\lambda)$

$$\hat{\lambda} = \overline{x}\,.$$

3.2.3 Weitere Schätzungen

a) Schätzung der Quantile

- Ausgehend von der geordneten mathematischen Stichprobe

$$X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$$

kann das p-Quantil geschätzt werden.

- Die Quantilsfunktion an der Stelle p $(0 < p < 1)$ wird geschätzt durch das Stichprobenquantil der Ordnung p

$$\widetilde{X}_p = \begin{cases} X_{(k)}, & \text{falls } np \text{ nicht ganzzahlig:} \\ & k \text{ ist die auf } np \text{ folgende ganze Zahl,} \\ \frac{1}{2}\left(X_{(k)} + X_{(k+1)}\right), & \text{falls } np \text{ ganzzahlig: } k = np. \end{cases}$$

b) Die empirische Verteilungsfunktion

- Für eine konkrete Stichprobe $x_1, \ldots, x_n$ kann man die zugehörige empirische Verteilungsfunktion durch

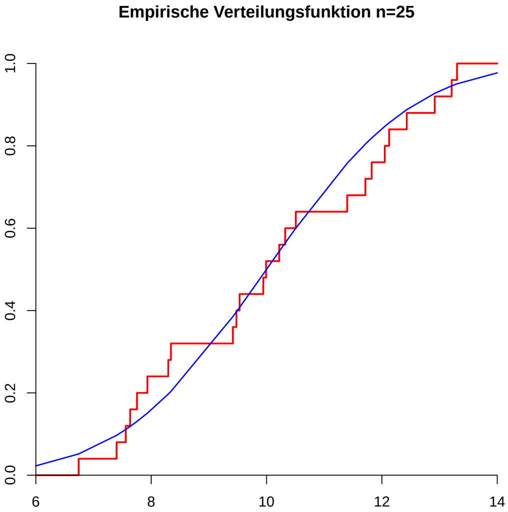
$$\hat{F}_n(x) = \frac{\text{Anzahl der Stichprobenwerte } < x}{n}, \qquad x \in \mathbb{R},$$

definieren.

- Dann kann man unter der klassischen Voraussetzung der Stichprobenentnahme zeigen, dass mit Wahrscheinlichkeit 1 die Folge dieser empirischen Verteilungsfunktionen gegen die Verteilungsfunktion der Merkmalszufallsgröße X konvergiert.
- Somit kann jede mögliche Wahrscheinlichkeitsverteilung eines beobachteten Merkmals mit gewünschter Genauigkeit erkannt werden, wenn man eine genügend große Stichprobe mit unabhängigen und identisch verteilten Stichprobenwerten zieht.

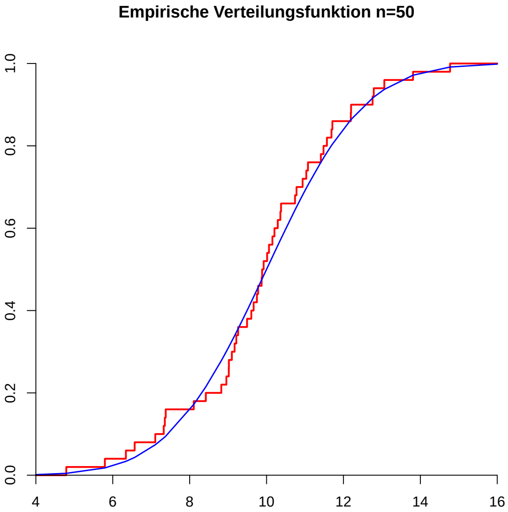
Beispiel für empirische Verteilungsfunktion

$X_i \sim \mathrm{N}(10, 4), \quad i = 1, \ldots n$, F_{X_i} - VF und $\hat{F}_n$ - emp. VF .



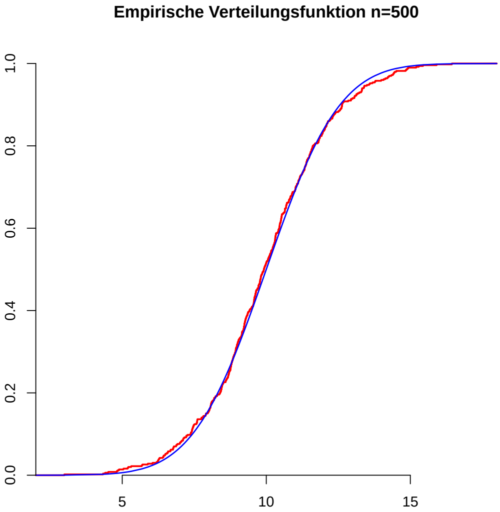
Beispiel für empirische Verteilungsfunktion II

$X_i \sim \mathrm{N}(10, 4), \quad i = 1, \ldots n$, F_{X_i} - VF und $\hat{F}_n$ - emp. VF .



Beispiel für empirische Verteilungsfunktion III

$X_i \sim \mathrm{N}(10, 4), \quad i = 1, \dots n$, F_{X_i} - VF und $\hat{F}_n$ - emp. VF .

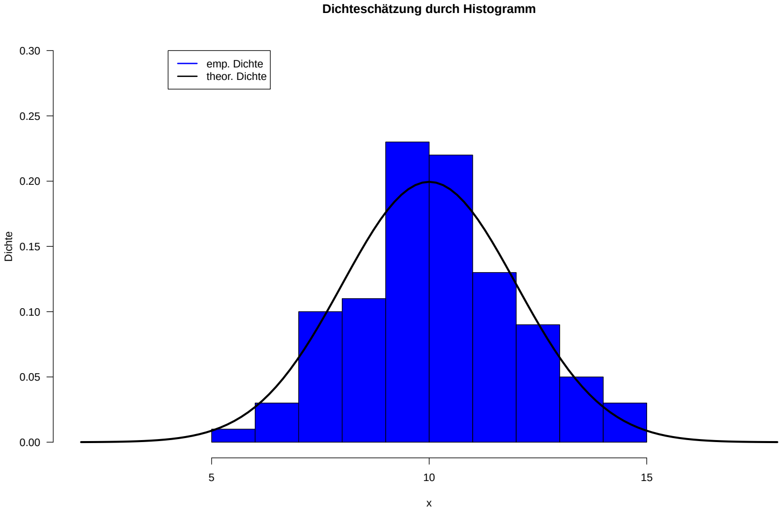


c) Schätzung der Dichtefunktion

- Eine einfache Schätzung der Dichtefunktion erhält man durch ein Histogramm . Dabei müssen die Höhen der Rechtecke so normiert werden, dass der Gesamtflächeninhalt unter allen Rechtecken gleich 1 ist.
- Ist von der Stichprobe $X_1, \ldots, X_n$ die Verteilung bis auf die Parameter bekannt und schätzt man diese aus der Stichprobe, so kann man damit die Dichtefunktion bestimmen. Man spricht dabei auch von einer parametrischen Dichteschätzung .
- Eine nichtparametrische Dichteschätzung erhält man z.B. durch Kerndichteschätzer.

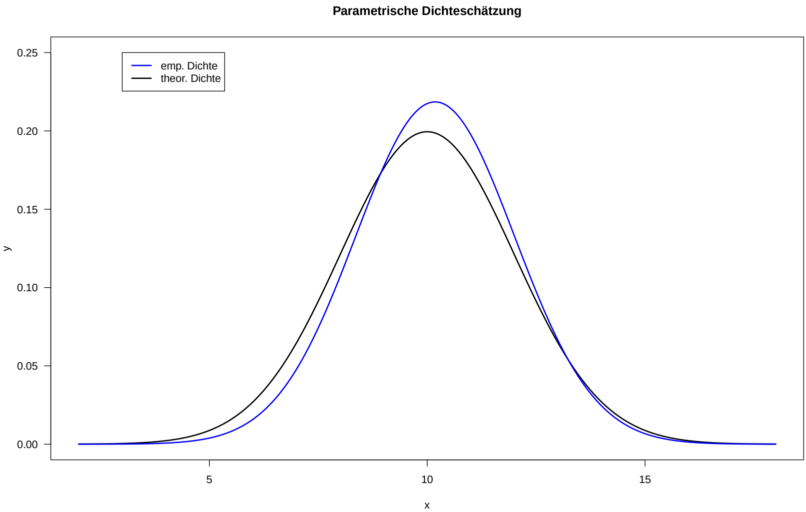
Dichteschätzung durch Histogramm

$$X_i \sim \mathrm{N}(10, 4), \quad i = 1, \ldots 100.$$



Parametrische Dichteschätzung

$X_i \sim \mathrm{N}(10, 4), \quad i = 1, \ldots 100, \qquad \hat{\mu} = \overline{x} = 10.18$ und $\hat{\sigma^2} = s^2 = 3.33$.



Nichtparametrische Dichteschätzung

$$X_i \sim \mathrm{N}(10, 4), \quad i = 1, \ldots 100\,.$$