

# Beurteilung von Schätzfunktionen

- *Bisher*: Zwei Methoden zur Konstruktion von Schätzfunktionen bekannt.
- *Problem*:  
Wie kann Güte/Qualität dieser Methoden bzw. der resultierenden Schätzfunktionen beurteilt werden?
- *Lösung*:  
Zu gegebener Schätzfunktion  $\hat{\theta}$  für  $\theta$ : Untersuchung des **zufälligen** Schätzfehlers  $\hat{\theta} - \theta$  (bzw. dessen Verteilung)
- Naheliegende Forderung für „gute“ Schätzfunktionen:  
Verteilung des Schätzfehler sollte möglichst „dicht“ um 0 konzentriert sein (d.h. Verteilung von  $\hat{\theta}$  sollte möglichst „dicht“ um  $\theta$  konzentriert sein)
- Aber:
  - ▶ Was bedeutet das?
  - ▶ Wie vergleicht man zwei Schätzfunktionen  $\hat{\theta}$  und  $\tilde{\theta}$ ? Wann ist Schätzfunktion  $\hat{\theta}$  „besser“ als  $\tilde{\theta}$  (und was bedeutet „besser“)?
  - ▶ Was ist zu beachten, wenn Verteilung des Schätz**fehlers** noch vom zu schätzenden Parameter abhängt?

# Bias, Erwartungstreue

- Eine offensichtlich gute Eigenschaft von Schätzfunktionen ist, wenn der zu schätzende (wahre) Parameter zumindest *im Mittel* getroffen wird, d.h. der *erwartete* Schätzfehler gleich Null ist:

## Definition 3.4 (Bias, Erwartungstreue)

Seien  $W$  eine parametrische Verteilungsannahme mit Parameterraum  $\Theta$ ,  $\hat{\theta}$  eine Schätzfunktion für  $\theta$ . Dann heißt

- 1 der erwartete Schätzfehler

$$\text{Bias}(\hat{\theta}) := E(\hat{\theta} - \theta) = E(\hat{\theta}) - \theta$$

die **Verzerrung** oder der **Bias** von  $\hat{\theta}$ ,

- 2 die Schätzfunktion  $\hat{\theta}$  **erwartungstreu für**  $\theta$  oder auch **unverzerrt für**  $\theta$ , falls  $\text{Bias}(\hat{\theta}) = 0$  bzw.  $E(\hat{\theta}) = \theta$  für alle  $\theta \in \Theta$  gilt.
- 3 Ist allgemeiner  $g : \Theta \rightarrow \mathbb{R}$  eine (messbare) Abbildung, so betrachtet man auch Schätzfunktionen  $\widehat{g(\theta)}$  für  $g(\theta)$  und nennt diese **erwartungstreu für**  $g(\theta)$ , wenn  $E(\widehat{g(\theta)} - g(\theta)) = 0$  bzw.  $E(\widehat{g(\theta)}) = g(\theta)$  für alle  $\theta \in \Theta$  gilt.

## Bemerkungen

- Obwohl Definition 3.4 auch für mehrdimensionale Parameterräume  $\Theta$  geeignet ist („0“ entspricht dann ggf. dem Nullvektor), betrachten wir zur Vereinfachung im Folgenden meist nur noch **eindimensionale** Parameterräume  $\Theta \subseteq \mathbb{R}$ .
- Ist beispielsweise  $W$  als Verteilungsannahme für  $Y$  die Menge aller Alternativverteilungen  $B(1, p)$  mit Parameter  $p \in \Theta = [0, 1]$ , so ist der ML-Schätzer  $\hat{p} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  bei Vorliegen einer Zufallsstichprobe  $X_1, \dots, X_n$  zu  $Y$  erwartungstreu für  $p$ , denn es gilt:

$$\begin{aligned} E(\hat{p}) &= E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \stackrel{E \text{ linear}}{=} \frac{1}{n} \sum_{i=1}^n E(X_i) \\ &\stackrel{F_{X_i}=F_Y}{=} \frac{1}{n} \sum_{i=1}^n E(Y) \\ &= \frac{1}{n} \cdot n \cdot p = p \text{ für alle } p \in [0, 1] \end{aligned}$$

- Allgemeiner gilt, dass  $\bar{X}$  bei Vorliegen einer Zufallsstichprobe stets erwartungstreu für  $E(Y)$  ist, denn es gilt analog zu oben:

$$\begin{aligned} E(\bar{X}) &= E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \stackrel{E \text{ linear}}{=} \frac{1}{n} \sum_{i=1}^n E(X_i) \\ &\stackrel{F_{X_i}=F_Y}{=} \frac{1}{n} \sum_{i=1}^n E(Y) \\ &= \frac{1}{n} \cdot n \cdot E(Y) = E(Y) \end{aligned}$$

- Genauso ist klar, dass man für beliebiges  $k$  mit dem  $k$ -ten empirischen Moment  $\bar{X}^k$  bei Vorliegen einer Zufallsstichprobe stets erwartungstreu Schätzer für das  $k$ -te theoretische Moment  $E(Y^k)$  erhält, denn es gilt:

$$E(\bar{X}^k) = E\left(\frac{1}{n} \sum_{i=1}^n X_i^k\right) = \frac{1}{n} \sum_{i=1}^n E(X_i^k) = \frac{1}{n} \sum_{i=1}^n E(Y^k) = E(Y^k)$$

- Der nach der Methode der Momente erhaltene Schätzer

$$\widehat{\sigma}^2 = \overline{X^2} - \bar{X}^2 \stackrel{\text{Verschiebungssatz}}{=} \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

für den Parameter  $\sigma^2$  einer normalverteilten Zufallsvariable ist **nicht** erwartungstreu für  $\sigma^2$ .

Bezeichnet  $\sigma^2 := \text{Var}(Y)$  nämlich die (unbekannte) Varianz der Zufallsvariablen  $Y$ , so kann gezeigt werden, dass für  $\widehat{\sigma}^2$  generell

$$E(\widehat{\sigma}^2) = \frac{n-1}{n} \sigma^2$$

gilt. Einen erwartungstreuen Schätzer für  $\sigma^2$  erhält man folglich mit der sogenannten **Stichprobenvarianz**

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{n}{n-1} \widehat{\sigma}^2,$$

denn es gilt offensichtlich

$$E(S^2) = E\left(\frac{n}{n-1} \widehat{\sigma}^2\right) = \frac{n}{n-1} E(\widehat{\sigma}^2) = \frac{n}{n-1} \cdot \frac{n-1}{n} \cdot \sigma^2 = \sigma^2.$$

## Vergleich von Schätzfunktionen

- Beim Vergleich von Schätzfunktionen: **oft** Beschränkung auf erwartungstreue Schätzfunktionen
- In der Regel: viele erwartungstreue Schätzfunktionen denkbar.
- Für die Schätzung von  $\mu := E(Y)$  beispielsweise alle *gewichteten* Mittel

$$\widehat{\mu}_{w_1, \dots, w_n} := \sum_{i=1}^n w_i \cdot X_i$$

mit der Eigenschaft  $\sum_{i=1}^n w_i = 1$  erwartungstreu für  $\mu$ , denn es gilt dann offensichtlich

$$E(\widehat{\mu}_{w_1, \dots, w_n}) = E\left(\sum_{i=1}^n w_i \cdot X_i\right) = \sum_{i=1}^n w_i E(X_i) = E(Y) \cdot \sum_{i=1}^n w_i = E(Y) = \mu.$$

- Problem: Welche Schätzfunktion ist „die beste“?
- Übliche Auswahl (bei Beschränkung auf erwartungstreue Schätzfunktionen!): Schätzfunktionen mit geringerer **Streuung (Varianz)** bevorzugen.

# Wirksamkeit, Effizienz

## Definition 3.5 (Wirksamkeit, Effizienz)

Sei  $\mathcal{W}$  eine parametrische Verteilungsannahme mit Parameterraum  $\Theta$ .

- ① Seien  $\hat{\theta}$  und  $\tilde{\theta}$  erwartungstreue Schätzfunktionen für  $\theta$ . Dann heißt  $\hat{\theta}$  **mindestens so wirksam** wie  $\tilde{\theta}$ , wenn

$$\text{Var}(\hat{\theta}) \leq \text{Var}(\tilde{\theta}) \text{ für alle } \theta \in \Theta$$

gilt.  $\hat{\theta}$  heißt **wirksamer** als  $\tilde{\theta}$ , wenn *außerdem*  $\text{Var}(\hat{\theta}) < \text{Var}(\tilde{\theta})$  für mindestens ein  $\theta \in \Theta$  gilt.

- ② Ist  $\hat{\theta}$  mindestens so wirksam wie alle (anderen) Schätzfunktionen einer Klasse mit erwartungstreuen Schätzfunktionen für  $\theta$ , so nennt man  $\hat{\theta}$  **effizient** in dieser Klasse erwartungstreuer Schätzfunktionen.

- Die Begriffe „Wirksamkeit“ und „Effizienz“ betrachtet man analog zu Definition 3.5 ebenfalls, wenn Funktionen  $g(\theta)$  von  $\theta$  geschätzt werden.
- $\text{Sd}(\hat{\theta}) = \sqrt{\text{Var}(\hat{\theta})}$  wird auch **Standardfehler** oder **Stichprobenfehler** von  $\hat{\theta}$  genannt.

## Beispiel: Effizienz

- Betrachte Klasse der (linearen) erwartungstreuen Schätzfunktionen

$$\hat{\mu}_{w_1, \dots, w_n} := \sum_{i=1}^n w_i \cdot X_i$$

mit  $\sum_{i=1}^n w_i = 1$  für den Erwartungswert  $\mu := E(Y)$  aus Folie 57.

- Für welche  $w_1, \dots, w_n$  erhält man (bei Vorliegen einer einfachen Stichprobe) die in dieser Klasse **effiziente** Schätzfunktion  $\hat{\mu}_{w_1, \dots, w_n}$ ?
- ↪ Suche nach den Gewichten  $w_1, \dots, w_n$  (mit  $\sum_{i=1}^n w_i = 1$ ), für die  $\text{Var}(\hat{\mu}_{w_1, \dots, w_n})$  möglichst klein wird.
- Man kann zeigen, dass  $\text{Var}(\hat{\mu}_{w_1, \dots, w_n})$  minimal wird, wenn

$$w_i = \frac{1}{n} \text{ für alle } i \in \{1, \dots, n\}$$

gewählt wird.

- Damit ist  $\bar{X}$  also effizient in der Klasse der linearen erwartungstreuen Schätzfunktionen für den Erwartungswert  $\mu$  einer Verteilung!

## Mittlerer quadratischer Fehler (MSE)

- Wenn Erwartungstreue im Vordergrund steht, ist Auswahl nach minimaler Varianz der Schätzfunktion sinnvoll.
- Ist Erwartungstreue nicht das „übergeordnete“ Ziel, verwendet man zur Beurteilung der Qualität von Schätzfunktionen häufig auch den sogenannten mittleren quadratischen Fehler (mean square error, MSE).

### Definition 3.6 (Mittlerer quadratischer Fehler (MSE))

Sei  $W$  eine parametrische Verteilungsannahme mit Parameterraum  $\Theta$ ,  $\hat{\theta}$  eine Schätzfunktion für  $\theta \in \Theta$ . Dann heißt  $\text{MSE}(\hat{\theta}) := E[(\hat{\theta} - \theta)^2]$  der **mittlere quadratische Fehler (mean square error, MSE)** von  $\hat{\theta}$ .

- Mit dem (umgestellten) Varianzzerlegungssatz erhält man direkt

$$E[(\hat{\theta} - \theta)^2] = \underbrace{\text{Var}(\hat{\theta} - \theta)}_{=\text{Var}(\hat{\theta})} + \underbrace{\left[E(\hat{\theta} - \theta)\right]^2}_{=(\text{Bias}(\hat{\theta}))^2},$$

für erwartungstreue Schätzfunktionen stimmt der MSE einer Schätzfunktion also gerade mit der Varianz überein!

## Konsistenz im quadratischen Mittel

- Basierend auf dem MSE ist ein „minimales“ Qualitätskriterium für Schätzfunktionen etabliert.
- Das Kriterium fordert (im Prinzip), dass man den MSE durch Vergrößerung des Stichprobenumfangs beliebig klein bekommen muss.
- Zur Formulierung des Kriteriums müssen Schätzfunktionen  $\hat{\theta}_n$  für „variable“ Stichprobengrößen  $n \in \mathbb{N}$  betrachtet werden.

### Definition 3.7 (Konsistenz im quadratischen Mittel)

Seien  $W$  eine parametrische Verteilungsannahme mit Parameterraum  $\Theta$ ,  $\hat{\theta}_n$  eine Schätzfunktion für  $\theta \in \Theta$  zum Stichprobenumfang  $n \in \mathbb{N}$ . Dann heißt die (Familie von) Schätzfunktion(en)  $\hat{\theta}_n$  **konsistent im quadratischen Mittel für  $\theta$** , falls

$$\lim_{n \rightarrow \infty} \text{MSE}(\hat{\theta}_n) = \lim_{n \rightarrow \infty} E[(\hat{\theta}_n - \theta)^2] = 0$$

für alle  $\theta \in \Theta$  gilt.

- Mit der (additiven) Zerlegung des MSE in Varianz und quadrierten Bias aus Folie 60 erhält man sofort:

### Satz 3.8

Seien  $W$  eine parametrische Verteilungsannahme mit Parameterraum  $\Theta$ ,  $\hat{\theta}_n$  eine Schätzfunktion für  $\theta \in \Theta$  zum Stichprobenumfang  $n \in \mathbb{N}$ . Dann ist die Familie  $\hat{\theta}_n$  von Schätzfunktionen genau dann konsistent im quadratischen Mittel für  $\theta$ , wenn sowohl

$$\textcircled{1} \lim_{n \rightarrow \infty} E(\hat{\theta}_n - \theta) = 0 \quad \text{bzw.} \quad \lim_{n \rightarrow \infty} E(\hat{\theta}_n) = \theta \quad \text{als auch}$$

$$\textcircled{2} \lim_{n \rightarrow \infty} \text{Var}(\hat{\theta}_n) = 0$$

für alle  $\theta \in \Theta$  gilt.

- Eigenschaft  $\textcircled{1}$  aus Satz 3.8 wird auch **asymptotische Erwartungstreue** genannt; asymptotische Erwartungstreue ist offensichtlich schwächer als Erwartungstreue.
- Es gibt also auch (Familien von) Schätzfunktionen, die für einen Parameter  $\theta$  zwar konsistent im quadratischen Mittel sind, aber nicht erwartungstreu.

## Beispiel: Konsistenz im quadratischen Mittel

- Voraussetzung (wie üblich):  $X_1, \dots, X_n$  einfache Stichprobe zu  $Y$ .
- Bekannt: Ist  $\mu := E(Y)$  der unbekannte Erwartungswert der interessierenden Zufallsvariable  $Y$ , so ist  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$  für alle  $n \in \mathbb{N}$  erwartungstreu.
- Ist  $\sigma^2 := \text{Var}(Y)$  die Varianz von  $Y$ , so erhält man für die Varianz von  $\bar{X}_n$  (vgl. Beweis der Effizienz von  $\bar{X}$  unter allen linearen erwartungstreuen Schätzfunktionen für  $\mu$ ):

$$\text{Var}(\bar{X}_n) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \underbrace{\text{Var}(X_i)}_{=\sigma^2} = \frac{\sigma^2}{n}$$

- Es gilt also  $\lim_{n \rightarrow \infty} \text{Var}(\bar{X}_n) = \lim_{n \rightarrow \infty} \frac{\sigma^2}{n} = 0$ , damit folgt zusammen mit der Erwartungstreue, dass  $\bar{X}_n$  konsistent im quadratischen Mittel für  $\mu$  ist.

## Verteilung des Stichprobenmittels $\bar{X}$

- **Bisher:** Interesse meist an einigen *Momenten* (Erwartungswert und Varianz) von Schätzfunktionen, insbesondere des Stichprobenmittels  $\bar{X}$ .
- Bereits bekannt: Ist  $\mu := E(Y)$ ,  $\sigma^2 := \text{Var}(Y)$  und  $X_1, \dots, X_n$  eine einfache Stichprobe zu  $Y$ , so gilt

$$E(\bar{X}) = \mu \quad \text{sowie} \quad \text{Var}(\bar{X}) = \frac{\sigma^2}{n} .$$

- Damit Aussagen über Erwartungstreue, Wirksamkeit, Konsistenz möglich.
- **Jetzt:** Interesse an ganzer **Verteilung** von Schätzfunktionen, insbesondere  $\bar{X}$ .
- Verteilungsaussagen entweder
  - ▶ auf Grundlage des Verteilungstyps von  $Y$  aus der Verteilungsannahme in speziellen Situationen **exakt** möglich oder
  - ▶ auf Grundlage des zentralen Grenzwertsatzes (bei genügend großem Stichprobenumfang!) allgemeiner **näherungsweise (approximativ)** möglich.
- Wir unterscheiden im Folgenden nur zwischen:
  - ▶  $Y$  normalverteilt  $\rightsquigarrow$  Verwendung der exakten Verteilung von  $\bar{X}$ .
  - ▶  $Y$  nicht normalverteilt  $\rightsquigarrow$  Verwendung der Näherung der Verteilung von  $\bar{X}$  aus dem zentralen Grenzwertsatz.

### Aus „Deskriptive Statistik und Wahrscheinlichkeitsrechnung“:

- 1 Gilt  $Y \sim N(\mu, \sigma^2)$ , so ist  $\bar{X}$  **exakt** normalverteilt mit Erwartungswert  $\mu$  und Varianz  $\frac{\sigma^2}{n}$ , es gilt also

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) .$$

- 2 Ist  $Y$  beliebig verteilt mit  $E(Y) =: \mu$  und  $\text{Var}(Y) =: \sigma^2$ , so rechtfertigt der zentrale Grenzwertsatz **für ausreichend große Stichprobenumfänge**  $n$  die Näherung der tatsächlichen Verteilung von  $\bar{X}$  durch eine Normalverteilung mit Erwartungswert  $\mu$  und Varianz  $\frac{\sigma^2}{n}$  (wie oben!), man schreibt dann auch

$$\bar{X} \dot{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$$

und sagt „ $\bar{X}$  ist approximativ (näherungsweise)  $N\left(\mu, \frac{\sigma^2}{n}\right)$ -verteilt“.

Der Standardabweichung  $\text{Sd}(\bar{X}) = \sqrt{\text{Var}(\bar{X})}$  von  $\bar{X}$  (also der Standardfehler der Schätzfunktion  $\bar{X}$  für  $\mu$ ) wird häufig mit  $\sigma_{\bar{X}} := \frac{\sigma}{\sqrt{n}}$  abgekürzt.

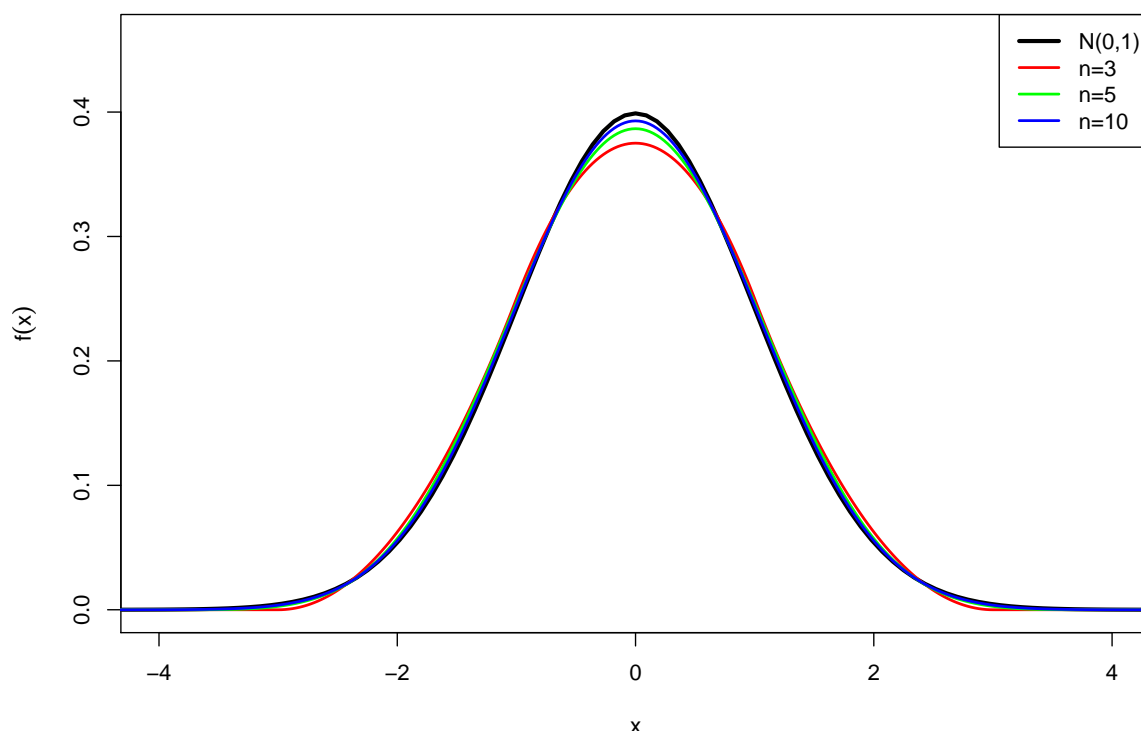
- Die Qualität der Näherung der Verteilung im Fall ② wird mit zunehmendem Stichprobenumfang höher, hängt aber **ganz entscheidend** vom Verteilungstyp (und sogar der konkreten Verteilung) von  $Y$  ab!
- Pauschale Kriterien an den Stichprobenumfang  $n$  („Daumenregeln“, z.B.  $n \geq 30$ ) finden sich häufig in der Literatur, sind aber nicht ganz unkritisch.
- Verteilungseigenschaft  $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$  bzw.  $\bar{X} \dot{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$  wird meistens (äquivalent!) in der (auch aus dem zentralen Grenzwertsatz bekannten) Gestalt

$$\frac{\bar{X} - \mu}{\sigma} \sqrt{n} \sim N(0, 1) \quad \text{bzw.} \quad \frac{\bar{X} - \mu}{\sigma} \sqrt{n} \dot{\sim} N(0, 1)$$

verwendet, da dann Verwendung von Tabellen zur Standardnormalverteilung möglich.

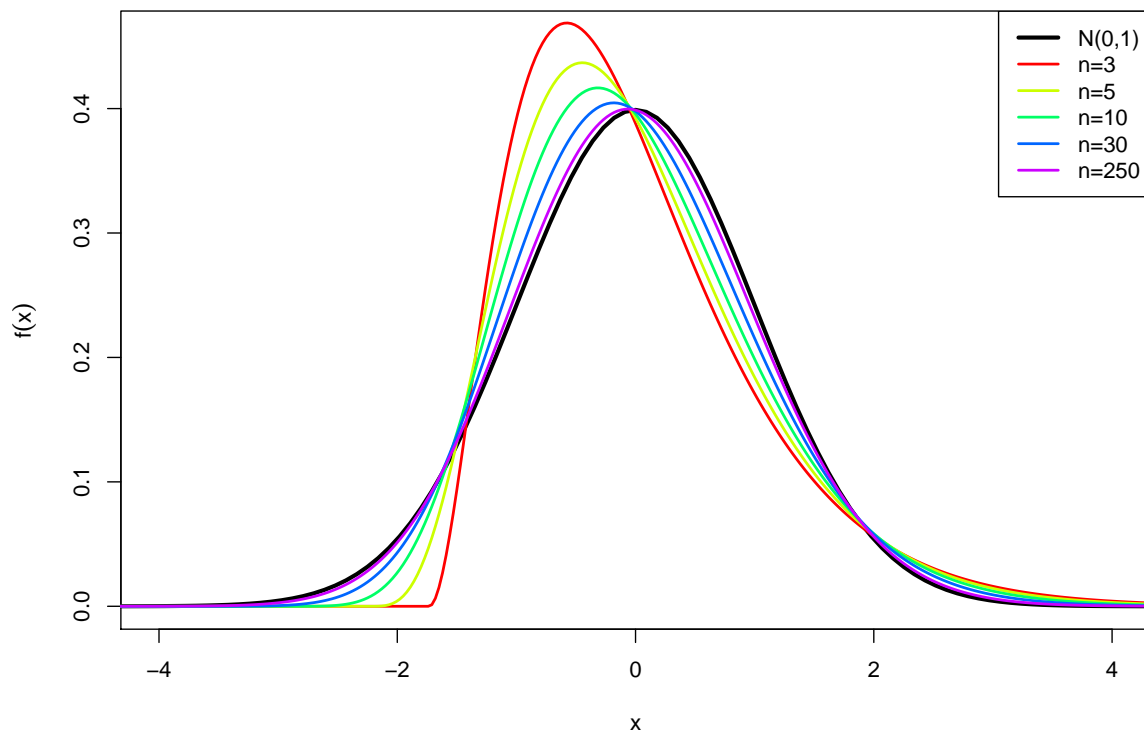
- Im Folgenden: Einige Beispiele für Qualität von Näherungen durch Vergleich der Dichtefunktion der Standardnormalverteilungsapproximation mit der tatsächlichen Verteilung von  $\frac{\bar{X} - \mu}{\sigma} \sqrt{n}$  für unterschiedliche Stichprobenumfänge  $n$ .

## Beispiel: Näherung, falls $Y \sim \text{Unif}(20, 50)$

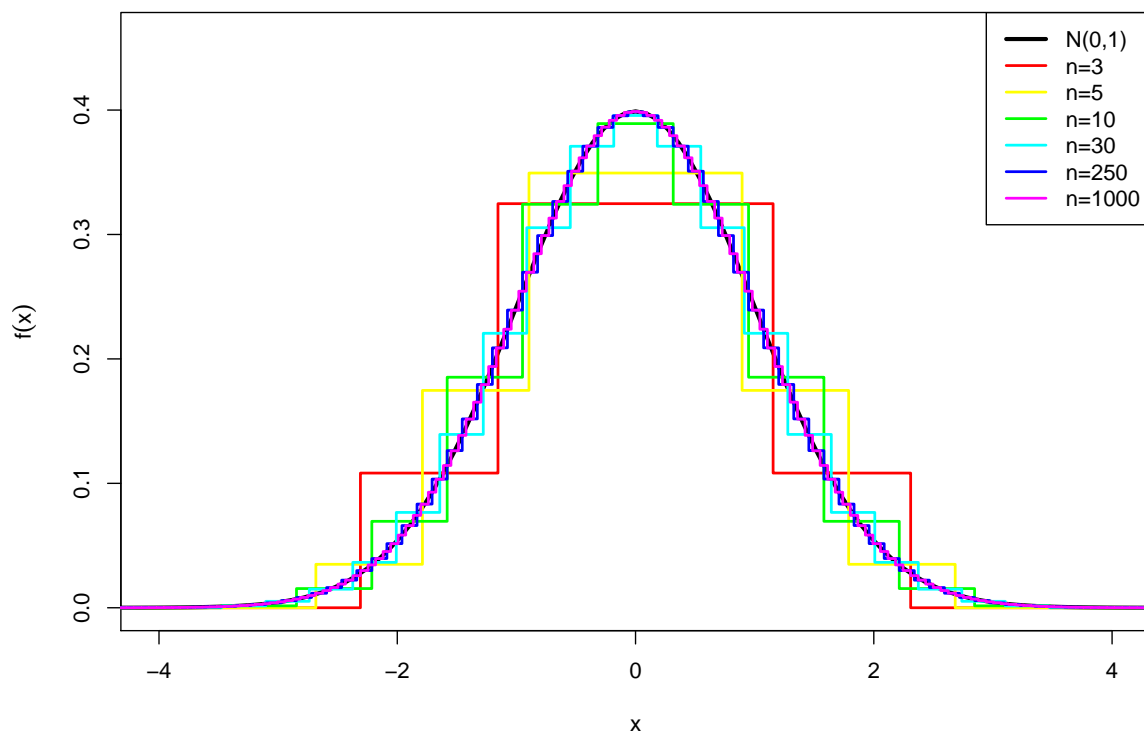




## Beispiel: Näherung, falls $Y \sim \text{Exp}(2)$



## Beispiel: Näherung, falls $Y \sim B(1, 0.5)$



# Beispiel: Näherung, falls $Y \sim B(1, 0.05)$

