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 $\widehat{\theta}$ für θ: Untersuchung des **zufälligen** Schätzfehlers $\widehat{\theta} - \theta$ (bzw. dessen Verteilung)
- Naheliegende Forderung für „gute“ Schätzfunktionen:

  Verteilung des Schätzfehler sollte möglichst „dicht“ um 0 konzentriert sein (d.h. Verteilung von $\widehat{\theta}$ sollte möglichst „dicht“ um θ konzentriert sein)
- Aber:
  - Was bedeutet das?
  - Wie vergleicht man zwei Schätzfunktionen $\widehat{\theta}$ und $\widetilde{\theta}$? Wann ist Schätzfunktion $\widehat{\theta}$ „besser“ als $\widetilde{\theta}$ (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 Θ, $\widehat{\theta}$ eine Schätzfunktion für θ. Dann heißt

1. der erwartete Schätzfehler

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

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

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

Bemerkungen

- Obwohl Definition 3.4 auch für mehrdimensionale Parameterräume Θ 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 $\widehat{p} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ bei Vorliegen einer Zufallsstichprobe $X_1, \ldots, X_n$ zu Y erwartungstreu für p, denn es gilt:

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

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

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

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

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

- Der nach der Methode der Momente erhaltene Schätzer

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

für den Parameter σ^2 einer normalverteilten Zufallsvariable ist **nicht** erwartungstreu für σ^2.
Bezeichnet $\sigma^2 := \text{Var}(Y)$ nämlich die (unbekannte) Varianz der Zufallsvariablen Y, so kann gezeigt werden, dass für $\widehat{\sigma^2}$ generell

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

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

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

denn es gilt offensichtlich

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

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 μ, denn es gilt dann offensichtlich

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

- 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 W eine parametrische Verteilungsannahme mit Parameterraum Θ.

1. Seien $\widehat{\theta}$ und $\widetilde{\theta}$ erwartungstreue Schätzfunktionen für θ. Dann heißt $\widehat{\theta}$ **mindestens so wirksam** wie $\widetilde{\theta}$, wenn
$$\text{Var}(\widehat{\theta}) \leq \text{Var}(\widetilde{\theta}) \text{ für alle } \theta \in \Theta$$
gilt. $\widehat{\theta}$ heißt **wirksamer** als $\widetilde{\theta}$, wenn *außerdem* $\text{Var}(\widehat{\theta}) < \text{Var}(\widetilde{\theta})$ für mindestens ein $\theta \in \Theta$ gilt.
2. Ist $\widehat{\theta}$ mindestens so wirksam wie alle (anderen) Schätzfunktionen einer Klasse mit erwartungstreuen Schätzfunktionen für θ, so nennt man $\widehat{\theta}$ **effizient** in dieser Klasse erwartungstreuer Schätzfunktionen.

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

Beispiel: Effizienz

- Betrachte Klasse der (linearen) erwartungstreuen Schätzfunktionen

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

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

- Für welche $w_1, \ldots, w_n$ erhält man (bei Vorliegen einer einfachen Stichprobe) die in dieser Klasse **effiziente** Schätzfunktion $\widehat{\mu}_{w_1,\ldots,w_n}$?

$\rightsquigarrow$ Suche nach den Gewichten $w_1, \ldots, w_n$ (mit $\sum_{i=1}^{n} w_i = 1$), für die $\text{Var}(\widehat{\mu}_{w_1,\ldots,w_n})$ möglichst klein wird.

- Man kann zeigen, dass $\text{Var}(\widehat{\mu}_{w_1,\ldots,w_n})$ minimal wird, wenn

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

gewählt wird.

- Damit ist $\overline{X}$ also effizient in der Klasse der linearen erwartungstreuen Schätzfunktionen für den Erwartungswert μ 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 Θ, $\widehat{\theta}$ eine Schätzfunktion für $\theta \in \Theta$. Dann heißt $\mathrm{MSE}(\widehat{\theta}) := \mathrm{E}\left[(\widehat{\theta} - \theta)^2\right]$ der **mittlere quadratische Fehler (mean square error, MSE)** von $\widehat{\theta}$.

- Mit dem (umgestellten) Varianzzerlegungssatz erhält man direkt

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

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 $\widehat{\theta}_n$ für „variable" Stichprobengrößen $n \in \mathbb{N}$ betrachtet werden.

Definition 3.7 (Konsistenz im quadratischen Mittel)

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

$$\lim_{n\to\infty} \mathrm{MSE}(\widehat{\theta}_n) = \lim_{n\to\infty} \mathrm{E}\left[(\widehat{\theta}_n - \theta)^2\right] = 0$$

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 Θ, $\widehat{\theta}_n$ eine Schätzfunktion für $\theta \in \Theta$ zum Stichprobenumfang $n \in \mathbb{N}$. Dann ist die Familie $\widehat{\theta}_n$ von Schätzfunktionen genau dann konsistent im quadratischen Mittel für θ, wenn sowohl

1. $\lim_{n\to\infty} \mathrm{E}(\widehat{\theta}_n - \theta) = 0$ *bzw.* $\lim_{n\to\infty} \mathrm{E}(\widehat{\theta}_n) = \theta$ *als auch*
2. $\lim_{n\to\infty} \mathrm{Var}(\widehat{\theta}_n) = 0$

für alle $\theta \in \Theta$ gilt.

- Eigenschaft 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 θ zwar konsistent im quadratischen Mittel sind, aber nicht erwartungstreu.

Beispiel: Konsistenz im quadratischen Mittel

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

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

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

Verteilung des Stichprobenmittels $\overline{X}$

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

$$\mathrm{E}(\overline{X}) = \mu \qquad \text{sowie} \qquad \text{Var}(\overline{X}) = \frac{\sigma^2}{n} \ .$$

- Damit Aussagen über Erwartungstreue, Wirksamkeit, Konsistenz möglich.
- **Jetzt:** Interesse an ganzer **Verteilung** von Schätzfunktionen, insbesondere $\overline{X}$.
- 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 $\overline{X}$.
  - Y nicht normalverteilt $\rightsquigarrow$ Verwendung der Näherung der Verteilung von $\overline{X}$ aus dem zentralen Grenzwertsatz.

Aus „Deskriptive Statistik und Wahrscheinlichkeitsrechnung“:

1. Gilt $Y \sim N(\mu, \sigma^2)$, so ist $\overline{X}$ **exakt** normalverteilt mit Erwartungswert μ und Varianz $\frac{\sigma^2}{n}$, es gilt also
$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \; .$$
2. Ist Y beliebig verteilt mit $\mathrm{E}(Y) =: \mu$ und $\mathrm{Var}(Y) =: \sigma^2$, so rechtfertigt der zentrale Grenzwertsatz **für ausreichend große Stichprobenumfänge** n die Näherung der tatsächlichen Verteilung von $\overline{X}$ durch eine Normalverteilung mit Erwartungswert μ und Varianz $\frac{\sigma^2}{n}$ (wie oben!), man schreibt dann auch
$$\overline{X} \stackrel{\bullet}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$$
und sagt „$\overline{X}$ ist approximativ (näherungsweise) $N\left(\mu, \frac{\sigma^2}{n}\right)$-verteilt“.

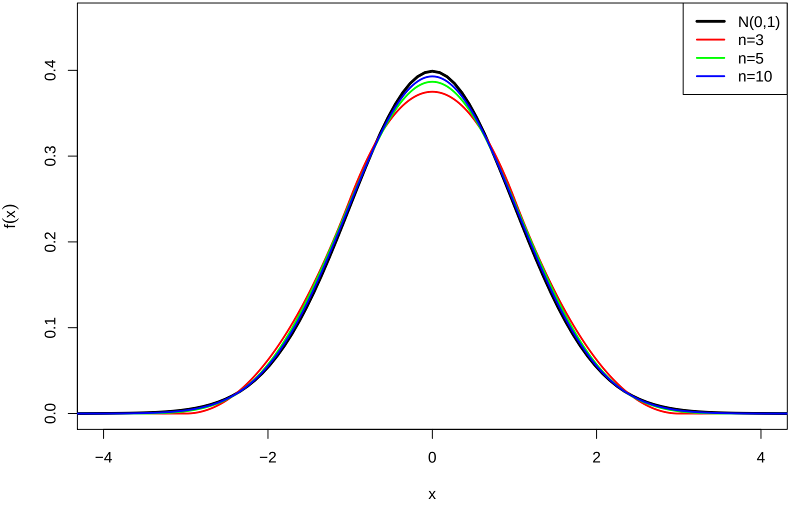
Der Standardabweichung $\mathrm{Sd}(\overline{X}) = \sqrt{\mathrm{Var}(\overline{X})}$ von $\overline{X}$ (also der Standardfehler der Schätzfunktion $\overline{X}$ für μ) wird häufig mit $\sigma_{\overline{X}} := \frac{\sigma}{\sqrt{n}}$ abgekürzt.

- 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 $\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ bzw. $\overline{X} \overset{\bullet}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$ wird meistens (äquivalent!) in der (auch aus dem zentralen Grenzwertsatz bekannten) Gestalt

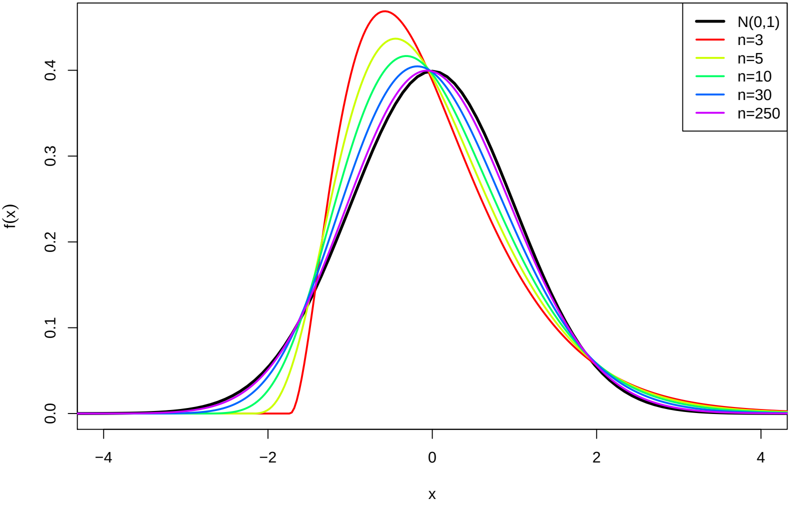
$$\frac{\overline{X} - \mu}{\sigma}\sqrt{n} \sim N(0,1) \qquad \text{bzw.} \qquad \frac{\overline{X} - \mu}{\sigma}\sqrt{n} \overset{\bullet}{\sim} N(0,1)$$

  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 $\frac{\overline{X}-\mu}{\sigma}\sqrt{n}$ für unterschiedliche Stichprobenumfänge n.

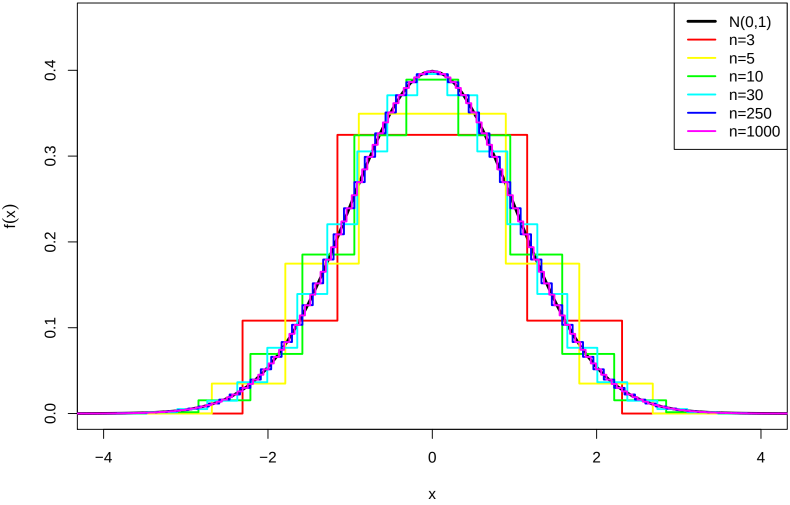
Beispiel: Näherung, falls $Y \sim \text{Unif}(20, 50)$



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



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



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