Mathe III

Ralf Herbrich

Konfidenzintervalle

Überblick

1. Punktschätzer und Intervallschätzer
2. Konfidenzintervalle unter Normalverteilungsannahme
   - Konfidenzintervalle für den Erwartungswert
   - Konfidenzintervalle für die Varianz
3. Approximative Konfidenzintervalle
   - Approximation über Normalverteilung
   - Approximation über Ungleichungen

Überblick

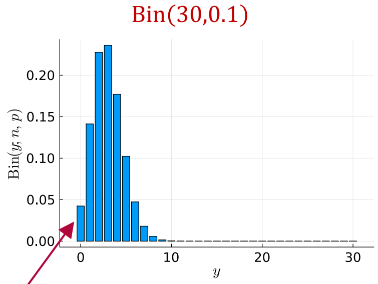
1. Punktschätzer und Intervallschätzer
2. Konfidenzintervalle unter Normalverteilungsannahme
   - Konfidenzintervalle für den Erwartungswert
   - Konfidenzintervalle für die Varianz
3. **Approximative Konfidenzintervalle**
   - Approximation über Normalverteilung
   - Approximation über Ungleichungen

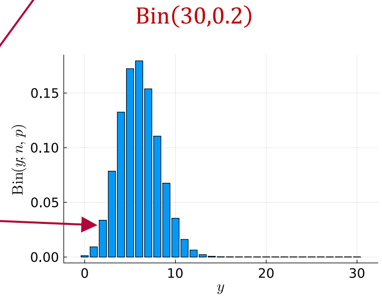
Motivation

- **Situation**: Bisher haben wir Konfidenzintervalle $C_\alpha : \Omega \to \mathcal{B}(\mathbb{R})$ konstruiert, bei denen wir die Verteilung der Stichprobe $x \in \Omega$ für jeden Wert von $\theta \in \Theta$ kennen und formal leicht nach θ umstellen können.
- **Frage**: Was machen wir, wenn wir die Verteilung der Stichprobe $x \in \Omega$ für jedes $\theta \in \Theta$ kennen aber nicht leicht nach θ umstellen können?
- **Beispiel (Club)**. Wir beobachten 30 Wartende vor einem Club, bei dem die Türsteherin mit einer unbekannten Wahrscheinlichkeit einen Gast einlässt; nur 3 Wartende dürfen in den Club. Was ist ein Konfidenzintervall für die Wahrscheinlichkeit auf einem Niveau von 95%?
  - **Ansatz**: Sei X_i eine Bernoulli-verteilte Zufallsvariable, $X_i \sim \mathrm{Bern}(p)$ mit der Bedeutung

$$X_i = \begin{cases} 0 & \text{Gast } i \text{ wird nicht reingelassen} \\ 1 & \text{Gast } i \text{ wird reingelassen} \end{cases}$$

    Alle X_i sind paarweise unabhängig und $Y = \sum_{i=1}^{30} X_i \sim \mathrm{Bin}(30, p)$.
  - **Problem**: Nicht für jeden Wert von p gibt es ein Intervall in $\{0, \ldots, 30\}$, was genau 95% Wahrscheinlichkeitsmasse hat!





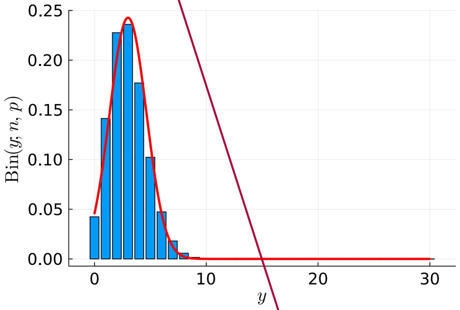
Überblick

1. Punktschätzer und Intervallschätzer
2. Konfidenzintervalle unter Normalverteilungsannahme
   - Konfidenzintervalle für den Erwartungswert
   - Konfidenzintervalle für die Varianz
3. Approximative Konfidenzintervalle
   - **Approximation über Normalverteilung**
   - Approximation über Ungleichungen

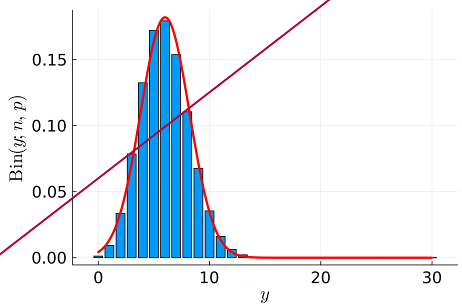
Approximatives Konfidenzintervall

- **Idee**: Wenn n groß ist ($n \geq 30$), dann gilt nach dem zentralen Grenzwertsatz, dass die Variable Y approximativ normalverteilt ist mit den Parametern $E[Y] = n \cdot p$ und $V[Y] = n \cdot p \cdot (1 - p)$

$\mathrm{Bin}(30{,}0.1)$

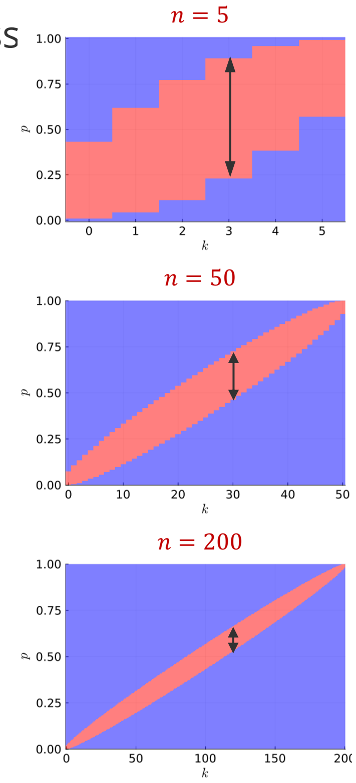


$\mathrm{Bin}(30{,}0.2)$



- **Konstruktion von approximativen Konfidenzintervallen**. Für $p \in [0{,}1]$, wähle via $z_{\frac{\alpha}{2}}$ ein symmetrisches Intervall um $\frac{Y - E[Y]}{\sqrt{V[Y]}}$ mit Wahrscheinlichkeit $1 - \alpha$

$$P\left(z_{\frac{\alpha}{2}} < \frac{Y - n \cdot p}{\sqrt{n \cdot p \cdot (1 - p)}} < -z_{\frac{\alpha}{2}}\right) \approx 1 - \alpha$$



Approximatives Konfidenzintervall

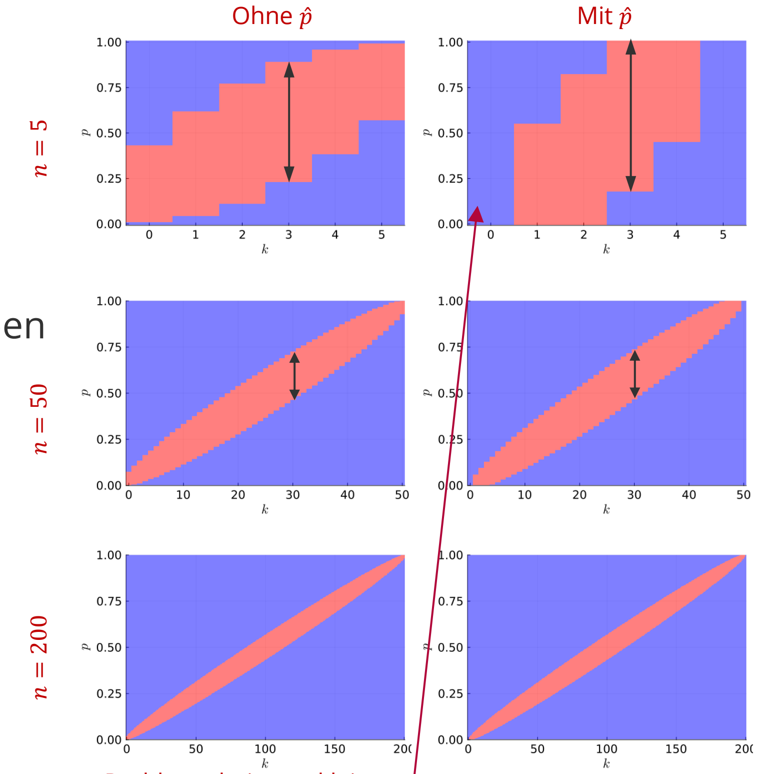
- **Problem**: Wenn wir die Lesart ändern, müssen wir die Ungleichungen nach p lösen. Das ergibt quadratische Ungleichungen (kompliziert!)

$$P\left(z_{\frac{\alpha}{2}} < \frac{Y - n \cdot p}{\sqrt{n \cdot p \cdot (1-p)}} < -z_{\frac{\alpha}{2}}\right) \approx 1 - \alpha$$

- **Idee**: Für den Nenner benutzen wir den MLE von p gegeben durch $\hat{p} = \frac{Y}{n}$

$$P\left(z_{\frac{\alpha}{2}} < \frac{Y - n \cdot p}{\sqrt{n \cdot \hat{p} \cdot (1-\hat{p})}} < -z_{\frac{\alpha}{2}}\right) \approx 1 - \alpha$$

$$P\left(\hat{p} + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}} < p < \hat{p} - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}}\right) \approx 1 - \alpha$$



Probleme bei ganz kleinen und großen Beobachtungen

Approximatives Konfidenzintervall: Beispiel

- **Beispiel (Wahlen)**. Im August 2013 berichtete die *New York Times*, dass eine Wählerumfrage ergeben hätte, dass 52% der US-Bevölkerung mit der Arbeit von Präsident Obama zufrieden sind mit einer Konfidenz von $\pm 4\%$ bei einem Konfidenzniveau von 95%. Wie viele Leute wurden befragt?
- **Lösung**: Für ein Konfidenzniveau von $1 - \alpha = 95\%$ gilt, dass $\alpha = 0.05$ und damit $z_{\frac{\alpha}{2}} = -1.96$. Daher gilt, dass

$$P\left(0.52 - 1.96 \cdot \sqrt{\frac{0.52 \cdot 0.48}{n}} < p < 0.52 + \boxed{1.96 \cdot \sqrt{\frac{0.52 \cdot 0.48}{n}}}\right) \approx 0.95$$

$$1.96 \cdot \sqrt{\frac{0.52 \cdot 0.48}{n}} = 0.04 \quad \Leftrightarrow \quad n = 1.96^2 \cdot \frac{0.52 \cdot 0.48}{0.04^2} \approx 599.29$$

Wenn man die Unabhängigkeit der Befragten annimmt, waren es 599 Leute.

Approximative Konfidenzintervall: Stichprobengröße

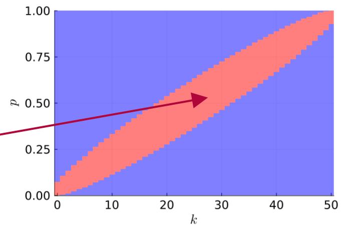
- **Frage**: Wie groß muss (im gleichen Beispiel) eine Stichprobe mindestens sein, damit das Konfidenzintervall auf Konfidenzniveau $1-\alpha$ höchstens 1% ist?
  - **Lösung**: Wir benutzen die Definition des approximativen Konfidenzintervalls

$$P\left(\hat{p} + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}} < p < \hat{p} - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}}\right) \approx 1-\alpha$$

$$-2z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}} = \frac{1}{100} \quad \Leftrightarrow 4z_{\frac{\alpha}{2}}^2 \cdot \frac{\hat{p}\cdot(1-\hat{p})}{n} = \frac{1}{100^2} \quad \Leftrightarrow n = 4z_{\frac{\alpha}{2}}^2 \cdot \hat{p}\cdot(1-\hat{p})\cdot 100^2$$

Die Funktion $\hat{p}\cdot(1-\hat{p})$ nimmt das Maximum bei $\hat{p} = \frac{1}{2}$ an. Daher gilt

$$n \geq 100^2 \cdot z_{\frac{\alpha}{2}}^2$$



- ## Bemerkungen (Stichprobengröße)
  - Die Stichprobengröße wächst mit der inversen **quadratischen** Genauigkeit.
  - Die Schranke wird besser, wenn der empirische Schätzer näher an 0% oder 100% ist.

Überblick

1. Punktschätzer und Intervallschätzer
2. Konfidenzintervalle unter Normalverteilungsannahme
   - Konfidenzintervalle für den Erwartungswert
   - Konfidenzintervalle für die Varianz
3. Approximative Konfidenzintervalle
   - Approximation über Normalverteilung
   - **Approximation über Ungleichungen**

Approximatives Konfidenzintervall für den Erwartungswert

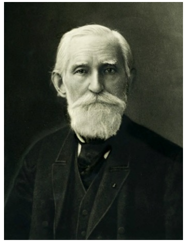
- **Frage**: Können wir ein Konfidenzintervall für den Erwartungswert finden, wenn wir keine konkrete Verteilungsannahme machen können?
- **Antwort**: Ja, denn per Definition verlangen wir nur, dass Stichproben mit **mindestens** $1-\alpha$ Wahrscheinlichkeit ins Konfidenzintervall fallen!
- **Satz (Tschebyscheff-Ungleichung)**. Sei X eine reelle Zufallsvariable mit Erwartungswert und Varianz. Dann gilt für alle $\lambda > 0$

$$P\left(|X - E[X]| \geq \lambda \cdot \sqrt{V[X]}\right) \leq \frac{1}{\lambda^2} \quad \Leftrightarrow \quad P\left(|X - E[X]| \leq \lambda \cdot \sqrt{V[X]}\right) \geq 1 - \frac{1}{\lambda^2}$$

- **Satz (Approximatives Konfidenzintervall für den Erwartungswert)**. Sei $(\Omega, \mathcal{F}, \{P_\theta \mid \theta \in \Theta\})$ ein parametrisches Modell mit $V_{P_\theta}[X] = v$ für alle $\theta \in \Theta$. Dann ist das folgende Intervall ein Konfidenzintervall zum Niveau $\alpha \in (0,1)$

$$C_\alpha(x) = \left[x - \sqrt{v/\alpha}, x + \sqrt{v/\alpha}\right]$$

- **Bemerkung (Approximatives Konfidenzintervall)**
  - Es gibt noch andere Ungleichungen, die man zum Konstruieren von approximativen Konfidenzintervallen nutzen kann (z.B. Markov Ungleichung).

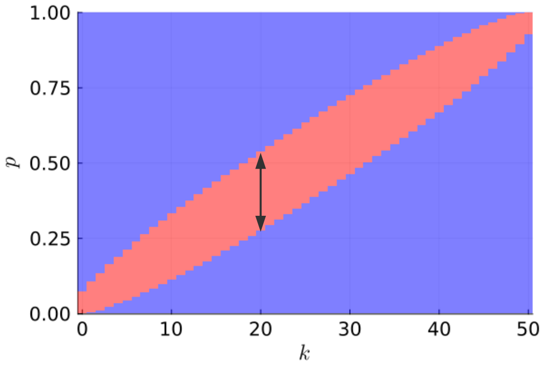


Pafnuty Chebyshev
(1821 – 1894)

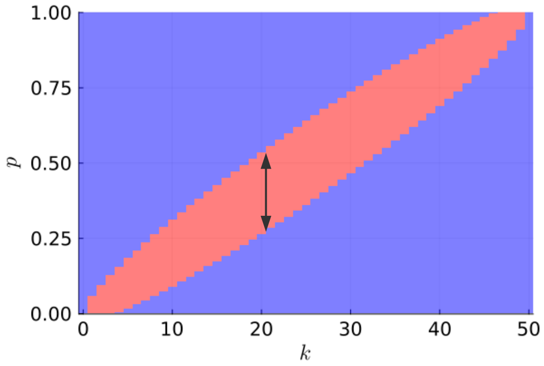
Approximative Konfidenzintervalle im Vergleich

- **Beispiel (Binomialverteilung).** Wir beobachten $X \sim \mathrm{Bin}(50, p)$ mit unbekannter Erfolgswahrscheinlichkeit p und betrachten ein 95% Konfidenzniveau!

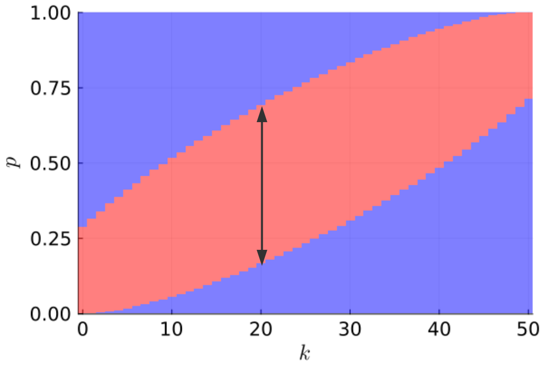
Normalverteilungsapproximation von $\mathrm{Bin}(50, p)$



Normalverteilungsapproximation von $\mathrm{Bin}(50, p)$ mit Varianzapproximation



Tschebyscheffapproximation von $\mathrm{Bin}(50, p)$



- **Bemerkung (Approximatives Konfidenzintervall)**
  - Weniger Annahmen führen zu unsichereren Aussagen über Parameter (d.h., größeren Konfidenzintervallen).
  - Approximative Konfidenzintervalle sind unter Umständen einfacher/schneller zu berechnen.

Viel Spaß bis zur nächsten Vorlesung!