12. Parameterschätzung

Yichen Han

25. Januar 2024

Inhalt

Wiederholungsquiz

Abbildung: Quiz 11

Gute Nachricht

Alle von mir erstellen Kursmaterialien stehen ab jetzt auf einer inoffiziellen(!) Art und Weise in einer GitHub Repository unter urheberrechtlichen Bedingungen zur Verfügung.

Abbildung: Link zur GitHub Repository

Schwaches Gesetz der großen Zahlen

Seien $X_1, ..., X_n$ iid. verteilt mit $\mathbb{E}(X_i) = \mu$ und $Var(X_i) = \sigma^2 < \infty$, $\forall\ i = 1, 2, ..., n$. Dann gilt:

$$\bar{X}_n := \frac{1}{n}\Sigma_{i=1}^{n} X_n \xrightarrow{\mathbb{P}} \mu \tag{1}$$

das ist äquivalent zu:

$$\mathbb{P}(|\bar{X}_n - \mu| > \epsilon) \overset{n\to\infty}{\to} 0 \quad \forall\ \epsilon > 0 \tag{2}$$

Zentraler Grenzwertsatz (Lindenberg-Lévy)

Seien $X_1, ..., X_n$ iid. verteilt mit $\mathbb{E}(X_i) = \mu$ und $Var(X_i) = \sigma^2 < \infty$, $\forall\ i = 1, 2, ..., n$. Dann gilt:

$$\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \tag{3}$$

Wir definieren $Z := \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$. Das ist äquivalent zu:

$$\lim_{n \to \infty} F_n(z) = F_{\mathcal{N}(0,1)}(z) =: \Phi(z) \ \ \forall\ z \in T_Z \tag{4}$$

Zentraler Grenzwertsatz (de Moivre-Laplace)

Sei X binomialverteilt, i.e. $X \sim Bin(n, \pi)$. Dann gilt:

$$\frac{X - n\pi}{\sqrt{n\pi(1-\pi)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \quad \text{bzw.} \quad X \xrightarrow{\mathcal{D}} \mathcal{N}(n\pi, n\pi(1-\pi)) \tag{5}$$

Wir definieren $Z := \frac{X-n\pi}{\sqrt{n\pi(1-\pi)}}$. Das ist äquivalent zu:

$$\lim_{n\to\infty} F_n(z) = F_{\mathcal{N}(0,1)}(z) =: \Phi(z) \;\; \forall\, z \in T_Z \tag{6}$$

Erweiterung 2

Satz von Gliwenko-Cantelli (Hauptsatz der Statistik)

Seien $X_1, ..., X_n$ iid. verteilt mit gemeinsamer Verteilungsfunktion $F : \mathbb{R} \to \mathbb{R}$ und der zu Stichprobenumfang n gehörigen ECDF F_n.
Wir definieren $D_n : T_X \to \mathbb{R}$ mit $D_n(x) := \sup_x |F_n(x) - F(x)|$. Es gilt:

$$D_n(x) \overset{f.s.}{\to} 0 \;\text{ bzw. } \mathbb{P}(x \in T_X | D_n(x) \to 0 \text{ für } n \to \infty) = 1 \qquad (7)$$

Inhalt

Aufgabe 1.1

Bei der Lufthansa ist aus Erfahrung bekannt, dass etwa 18% der Fluggäste ihre gebuchte Reise nicht antreten. Um die Auslastung der Flugzeuge möglichst hoch zu halten, werden mehr als die verfügbaren 150 Plätze in einem Airbus A320 angeboten.
Berechnen Sie mit Hilfe des zentralen Grenzwertsatzes die Wahrscheinlichkeit dafür, dass aufgrund von Überbelegung nicht alle Passagiere, die den Flug tatsächlich antreten, mitgenommen werden können, wenn 171 Plätze gebucht wurden.

Aufgabe 1.1

Bei der Lufthansa ist aus Erfahrung bekannt, dass etwa 18% der Fluggäste ihre gebuchte Reise nicht antreten. Um die Auslastung der Flugzeuge möglichst hoch zu halten, werden mehr als die verfügbaren 150 Plätze in einem Airbus A320 angeboten.
Berechnen Sie mit Hilfe des zentralen Grenzwertsatzes die Wahrscheinlichkeit dafür, dass aufgrund von Überbelegung nicht alle Passagiere, die den Flug tatsächlich antreten, mitgenommen werden können, wenn 171 Plätze gebucht wurden.

X: Anzahl von Personen, die die Reise antreten.
$X \sim \mathcal{B}(n, \pi), \quad \pi = 0.82, \quad n = 171.$

Aufgabe 1.1

Bei der Lufthansa ist aus Erfahrung bekannt, dass etwa 18% der Fluggäste ihre gebuchte Reise nicht antreten. Um die Auslastung der Flugzeuge möglichst hoch zu halten, werden mehr als die verfügbaren 150 Plätze in einem Airbus A320 angeboten.
Berechnen Sie mit Hilfe des zentralen Grenzwertsatzes die Wahrscheinlichkeit dafür, dass aufgrund von Überbelegung nicht alle Passagiere, die den Flug tatsächlich antreten, mitgenommen werden können, wenn 171 Plätze gebucht wurden.

X: Anzahl von Personen, die die Reise antreten.
$X \sim \mathcal{B}(n, \pi), \quad \pi = 0.82, \quad n = 171.$

Bekannt: $E(X) = n\pi$, $Var(X) = n\pi(1-\pi)$.

$$\text{ZGWS: } X \overset{a}{\sim} \mathcal{N}(n\pi, n\pi(1-\pi)) \quad \text{bzw.} \quad \frac{X - n\pi}{\sqrt{n\pi(1-\pi)}} \overset{a}{\sim} \mathcal{N}(0,1)$$

Aufgabe 1.2

$$P(X \geq 151) \quad = \quad 1 - P(X < 151)$$

Aufgabe 1.2

$$
\begin{aligned}
P(X \geq 151) &= 1 - P(X < 151) \\
&= 1 - P\left(\frac{X - n\pi}{\sqrt{n\pi(1-\pi)}} \leq \frac{151 - n\pi}{\sqrt{n\pi(1-\pi)}}\right) \\
&= 1 - F_{\mathcal{N}(0,1)}\left(\frac{151 - n\pi}{\sqrt{n\pi(1-\pi)}}\right) \\
&= 1 - \Phi\left(\frac{151 - n\pi}{\sqrt{n\pi(1-\pi)}}\right)
\end{aligned}
$$

$$
\begin{aligned}
P(X \geq 151) &= 1 - P(X < 151) \\
&= 1 - P\left(\frac{X - n\pi}{\sqrt{n\pi(1-\pi)}} \leq \frac{151 - n\pi}{\sqrt{n\pi(1-\pi)}}\right) \\
&= 1 - F_{\mathcal{N}(0,1)}\left(\frac{151 - n\pi}{\sqrt{n\pi(1-\pi)}}\right) \\
&= 1 - \Phi\left(\frac{151 - n\pi}{\sqrt{n\pi(1-\pi)}}\right) \\
&= 1 - \Phi\left(\frac{151 - 171 \cdot 0.82}{\sqrt{171 \cdot 0.82 \cdot 0.18}}\right) \\
&= 1 - \Phi(2.15) \\
&\approx 1 - 0.984 = 0.016
\end{aligned}
$$

Aufgabe 2.1

Wie oft muss mit einer idealen Münze mindestens geworfen werden, so dass die relative Häufigkeit von Wappen, mit einer Wahrscheinlichkeit von 0.95, um höchstens 0.01 respektive 0.001 von $\pi = 0.5$ abweicht?

Aufgabe 2.1

Wie oft muss mit einer idealen Münze mindestens geworfen werden, so dass die relative Häufigkeit von Wappen, mit einer Wahrscheinlichkeit von 0.95, um höchstens 0.01 respektive 0.001 von $\pi = 0.5$ abweicht?

$X_i \sim \mathcal{B}\left(\pi = \frac{1}{2}\right)$ mit π: Wahrscheinlichkeit für Wappen

$\bar{X} = \frac{1}{n} \sum x_i$: relative Häufigkeit von Wappen

Wie oft muss mit einer idealen Münze mindestens geworfen werden, so dass die relative Häufigkeit von Wappen, mit einer Wahrscheinlichkeit von 0.95, um höchstens 0.01 respektive 0.001 von $\pi = 0.5$ abweicht?

$X_i \sim \mathcal{B}\left(\pi = \frac{1}{2}\right)$ mit π: Wahrscheinlichkeit für Wappen

$\bar{X} = \frac{1}{n}\sum x_i$: relative Häufigkeit von Wappen

$$
\begin{aligned}
\text{ZGWS: } \bar{X} &\overset{a}{\sim} \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) \\
&= \mathcal{N}\left(\pi, \frac{\pi(1-\pi)}{n}\right) = \mathcal{N}\left(\frac{1}{2}, \frac{\frac{1}{2}(1-\frac{1}{2})}{n}\right) \\
&= \mathcal{N}\left(\frac{1}{2}, \frac{1}{4n}\right) \text{ bzw. } \frac{\bar{X}-\frac{1}{2}}{\sqrt{\frac{1}{4n}}} \overset{a}{\sim} \mathcal{N}(0,1)
\end{aligned}
$$

Aufgabe 2.2

Gesucht ist die Wahrscheinlichkeit, dass $\bar{X}$ um maximal ϵ von $\frac{1}{2}$ abweicht:

$$P\left(\frac{1}{2} - \epsilon \leq \bar{X} \leq \frac{1}{2} + \epsilon\right)$$

Aufgabe 2.2

Gesucht ist die Wahrscheinlichkeit, dass $\bar{X}$ um maximal ϵ von $\frac{1}{2}$ abweicht:

$$
\begin{aligned}
& P\left(\frac{1}{2} - \epsilon \leq \bar{X} \leq \frac{1}{2} + \epsilon\right) \\
= \; & P\left(\bar{X} \leq \frac{1}{2} + \epsilon\right) - P\left(\bar{X} \leq \frac{1}{2} - \epsilon\right)
\end{aligned}
$$

Gesucht ist die Wahrscheinlichkeit, dass $\bar{X}$ um maximal ϵ von $\frac{1}{2}$ abweicht:

$$
\begin{aligned}
& P\left(\frac{1}{2}-\epsilon \leq \bar{X} \leq \frac{1}{2}+\epsilon\right) \\
= \quad & P\left(\bar{X} \leq \frac{1}{2}+\epsilon\right) - P\left(\bar{X} \leq \frac{1}{2}-\epsilon\right) \\
= \quad & P\left(\frac{\bar{X}-\frac{1}{2}}{\sqrt{\frac{1}{4n}}} \leq \frac{\frac{1}{2}+\epsilon-\frac{1}{2}}{\sqrt{\frac{1}{4n}}}\right) - P\left(\frac{\bar{X}-\frac{1}{2}}{\sqrt{\frac{1}{4n}}} \leq \frac{\frac{1}{2}-\epsilon-\frac{1}{2}}{\sqrt{\frac{1}{4n}}}\right)
\end{aligned}
$$

Aufgabe 2.2

Gesucht ist die Wahrscheinlichkeit, dass $\bar{X}$ um maximal ϵ von $\frac{1}{2}$ abweicht:

$$
\begin{aligned}
& P\left(\frac{1}{2} - \epsilon \leq \bar{X} \leq \frac{1}{2} + \epsilon\right) \\
= \quad & P\left(\bar{X} \leq \frac{1}{2} + \epsilon\right) - P\left(\bar{X} \leq \frac{1}{2} - \epsilon\right) \\
= \quad & P\left(\frac{\bar{X} - \frac{1}{2}}{\sqrt{\frac{1}{4n}}} \leq \frac{\frac{1}{2} + \epsilon - \frac{1}{2}}{\sqrt{\frac{1}{4n}}}\right) - P\left(\frac{\bar{X} - \frac{1}{2}}{\sqrt{\frac{1}{4n}}} \leq \frac{\frac{1}{2} - \epsilon - \frac{1}{2}}{\sqrt{\frac{1}{4n}}}\right) \\
= \quad & \Phi\left(\frac{\frac{1}{2} + \epsilon - \frac{1}{2}}{\sqrt{\frac{1}{4n}}}\right) - \Phi\left(\frac{\frac{1}{2} - \epsilon - \frac{1}{2}}{\sqrt{\frac{1}{4n}}}\right)
\end{aligned}
$$

Gesucht ist die Wahrscheinlichkeit, dass $\bar{X}$ um maximal ϵ von $\frac{1}{2}$ abweicht:

$$\begin{aligned}
& P\left(\frac{1}{2}-\epsilon \leq \bar{X} \leq \frac{1}{2}+\epsilon\right) \\
= \;& P\left(\bar{X} \leq \frac{1}{2}+\epsilon\right) - P\left(\bar{X} \leq \frac{1}{2}-\epsilon\right) \\
= \;& P\left(\frac{\bar{X}-\frac{1}{2}}{\sqrt{\frac{1}{4n}}} \leq \frac{\frac{1}{2}+\epsilon-\frac{1}{2}}{\sqrt{\frac{1}{4n}}}\right) - P\left(\frac{\bar{X}-\frac{1}{2}}{\sqrt{\frac{1}{4n}}} \leq \frac{\frac{1}{2}-\epsilon-\frac{1}{2}}{\sqrt{\frac{1}{4n}}}\right) \\
= \;& \Phi\left(\frac{\frac{1}{2}+\epsilon-\frac{1}{2}}{\sqrt{\frac{1}{4n}}}\right) - \Phi\left(\frac{\frac{1}{2}-\epsilon-\frac{1}{2}}{\sqrt{\frac{1}{4n}}}\right) \\
\overset{(*)}{=} \;& \Phi(2\epsilon\sqrt{n}) - \left(1-\Phi(2\epsilon\sqrt{n})\right) \quad \text{mit} \quad (*):\; \Phi(n) = 1-\Phi(-n)
\end{aligned}$$

Gesucht ist die Wahrscheinlichkeit, dass $\bar{X}$ um maximal ϵ von $\frac{1}{2}$ abweicht:

$$
\begin{aligned}
& P\left(\frac{1}{2}-\epsilon \leq \bar{X} \leq \frac{1}{2}+\epsilon\right) \\
= \quad & P\left(\bar{X} \leq \frac{1}{2}+\epsilon\right) - P\left(\bar{X} \leq \frac{1}{2}-\epsilon\right) \\
= \quad & P\left(\frac{\bar{X}-\frac{1}{2}}{\sqrt{\frac{1}{4n}}} \leq \frac{\frac{1}{2}+\epsilon-\frac{1}{2}}{\sqrt{\frac{1}{4n}}}\right) - P\left(\frac{\bar{X}-\frac{1}{2}}{\sqrt{\frac{1}{4n}}} \leq \frac{\frac{1}{2}-\epsilon-\frac{1}{2}}{\sqrt{\frac{1}{4n}}}\right) \\
= \quad & \Phi\left(\frac{\frac{1}{2}+\epsilon-\frac{1}{2}}{\sqrt{\frac{1}{4n}}}\right) - \Phi\left(\frac{\frac{1}{2}-\epsilon-\frac{1}{2}}{\sqrt{\frac{1}{4n}}}\right) \\
\overset{(*)}{=} \quad & \Phi(2\epsilon\sqrt{n}) - \left(1-\Phi(2\epsilon\sqrt{n})\right) \quad \text{mit} \quad (*):\ \Phi(n) = 1-\Phi(-n) \\
= \quad & 2\Phi(2\epsilon\sqrt{n}) - 1 \overset{!}{>} 0.95
\end{aligned}
$$

$$
\begin{aligned}
2\Phi(2\epsilon\sqrt{n}) - 1 &\overset{!}{>} 0.95 \\
\Phi(2\epsilon\sqrt{n}) &> 0.975 \\
2\epsilon\sqrt{n} &> \Phi^{-1}(0.975) = 1.96 \\
\Rightarrow n &> \left(\frac{1.96}{2\epsilon}\right)^2 = \frac{0.9604}{\epsilon^2}
\end{aligned}
$$

$$\begin{aligned} 2\Phi(2\epsilon\sqrt{n}) - 1 &\overset{!}{>} 0.95 \\ \Phi(2\epsilon\sqrt{n}) &> 0.975 \\ 2\epsilon\sqrt{n} &> \Phi^{-1}(0.975) = 1.96 \\ \Rightarrow n &> \left(\frac{1.96}{2\epsilon}\right)^2 = \frac{0.9604}{\epsilon^2} \end{aligned}$$

a) $\epsilon = 0.01$: $n > 9604$

$\Longrightarrow$ mind. 9604 mal werfen

b) $\epsilon = 0.001$: $n > 960400$

$\Longrightarrow$ mind. 960365 mal werfen

Aufgabe 3.1

Ein Beamter verlässt an den 225 Arbeitstagen eines Jahres sein Büro immer erst kurz nach Dienstschluss. Die Dauern der täglichen zusätzlichen Arbeitszeiten lassen sich jeweils durch exponentialverteilte Zufallsvariablen mit Erwartungswert $\frac{1}{\lambda} = 5$ Minuten angemessen beschreiben und seien als unabhängig vorausgesetzt.

a) Leiten Sie die approximative Verteilung der gesamten zusätzlichen Arbeitszeit eines Jahres für den Beamten her.

b) Berechnen Sie approximativ die Wahrscheinlichkeit, dass der Beamte in einem Jahr mehr als 16 Stunden zusätzlich arbeitet.

Aufgabe 3.1

Ein Beamter verlässt an den 225 Arbeitstagen eines Jahres sein Büro immer erst kurz nach Dienstschluss. Die Dauern der täglichen zusätzlichen Arbeitszeiten lassen sich jeweils durch exponentialverteilte Zufallsvariablen mit Erwartungswert $\frac{1}{\lambda} = 5$ Minuten angemessen beschreiben und seien als unabhängig vorausgesetzt.

a) Leiten Sie die approximative Verteilung der gesamten zusätzlichen Arbeitszeit eines Jahres für den Beamten her.

b) Berechnen Sie approximativ die Wahrscheinlichkeit, dass der Beamte in einem Jahr mehr als 16 Stunden zusätzlich arbeitet.

X_i: Zusätzliche Arbeitszeit am i-ten Arbeitstag, $i = 1, \ldots, 225$, mit:

$$\begin{aligned} X_i &\overset{\text{iid}}{\sim} \mathcal{E}(\lambda) \\ \mathrm{E}(X_i) &= \frac{1}{\lambda} = 5[min], \\ \mathrm{Var}(X_i) &= \frac{1}{\lambda^2} = \frac{1}{0.04} = 25. \end{aligned}$$

Aufgabe 3.2

a) Jährliche zusätzliche Arbeitszeit: $Y_{225} := \sum_{i=1}^{225} X_i$.

a) Jährliche zusätzliche Arbeitszeit: $Y_{225} := \sum_{i=1}^{225} X_i$.

Nach ZGWS (Lemma), es gilt für Y_n, wobei:

$$Y_n := \sum_{i=1}^{n} X_i \text{ mit } \mathrm{E}(X_i) = \mu \text{ und } \mathrm{Var}(X_i) = \sigma^2$$

eine Summe aus n iid ZV ist, das Folgende:

$$Y_n \overset{a}{\sim} \mathcal{N}(n \cdot \mu, n \cdot \sigma^2)$$

a) Jährliche zusätzliche Arbeitszeit: $Y_{225} := \sum_{i=1}^{225} X_i$.

Nach ZGWS (Lemma), es gilt für Y_n, wobei:

$$Y_n := \sum_{i=1}^{n} X_i \text{ mit } \mathrm{E}(X_i) = \mu \text{ und } \mathrm{Var}(X_i) = \sigma^2$$

eine Summe aus n iid ZV ist, das Folgende:

$$Y_n \overset{a}{\sim} \mathcal{N}(n \cdot \mu, n \cdot \sigma^2)$$

Also:

$$Y_{225} = \sum_{i=1}^{225} X_i \overset{a}{\sim} \mathcal{N}(\underbrace{225 \cdot 5}_{1125}, \underbrace{225 \cdot 25}_{5625})$$

Alternativ:

$$Z_{225} = \frac{Y_{225} - 225 \cdot 5}{\sqrt{225 \cdot 25}} = \frac{Y_{225} - 1125}{75} \overset{a}{\sim} \mathcal{N}(0, 1)$$

b) 16 Stunden entsprechen 960 Minuten. Gesucht ist $P(Y_{225} > 960)$.

$$\begin{aligned}
P(Y_{225} > 960) &= 1 - P(Y_{225} \leq 960) \\
&= 1 - P\left(Z_{225} \leq \frac{960 - 1125}{75}\right) \\
&= 1 - P\,(Z_{225} \leq -2.2) \\
&= 1 - \Phi(-2.2) \\
&= \Phi(2.2) \approx 0.9861
\end{aligned}$$

Die Wahrscheinlichkeit, dass der Beamte in einem Jahr mehr als 16 Stunden zusätzlich arbeitet, beträgt approximativ 0.9861.

Aufgabe 4

Veranschaulichen Sie das Gesetz der großen Zahlen mit Hilfe von R. Nutzen Sie dabei graphische und numerische Möglichkeiten. Demonstrieren Sie das GGZ (auch) für Zufallszahlen die nicht aus einer Normalverteilung stammen und dokumentieren Sie ihren Code ausreichend.

Bonus: Diese Aufgabe wäre auch eine exzellente Gelegenheit Ihre erste interaktive App mit dem `shiny`-Paket zu programmieren!