2 Induktive Statistik

2.1 Grundprinzipien der induktiven Statistik

Ziel: Inferenzschluss, Repräsentationsschluss: Schluss von einer Stichprobe auf Eigenschaften der Grundgesamtheit, aus der sie stammt.

Genauer: Von Interesse sei ein Merkmal $\widetilde{X}$ in Grundgesamtheit $\widetilde{\Omega}$. Ziehe Stichprobe $(\omega_1, \dots \omega_n)$ von Elementen aus $\widetilde{\Omega}$ und werte $\widetilde{X}$ jeweils aus. Man erhält Werte $x_1, \dots x_n$; sie sind Realisationen der i.i.d Zufallsvariablen oder Zufallselemente $X_1, \dots X_n$ wobei die Wahrscheinlichkeitsverteilung der $X_1, \dots X_n$ genau die Häufigkeitsverhältnisse in der Grundgesamtheit widerspiegelt.

Die Frage lautet also: wie kommt man von Realisationen $x_1, \dots x_n$ von i.i.d. Zufallsvariablen $X_1, \dots X_n$ auf die Verteilung der X_i?

- Dazu nimmt man häufig an, man kenne den Grundtyp der Verteilung der $X_1 \dots X_n$, unbekannt seinen nur einzelne Parameter davon. z.B.

Annahme, X_i sei normalverteilt (bekannter Typ!), unbekannt seien nur μ, σ^2.
$\Longrightarrow$ *„parametrische Verteilungsannahme“* (meist im Folgenden)

- Alternativ: *nichtparametrische Modelle*
  Verteilungstyp nicht oder nur schwach festgelegt (z.B. symmetrische Verteilung)

- Klarerweise gilt im Allgemeinen (generelles Problem bei der Modellierung):
  Parametrische Modelle liefern schärfere Aussagen – wenn ihre Annahmen zutreffen. Wenn ihre Annahmen nicht zutreffen, dann existiert die große Gefahr von Fehlschlüssen.

Wichtige Fragestellungen der induktiven Statistik:

2.2 Punktschätzung

Jann (2002, Kap 5.5.1)5.5.1
Fahrmeir et al. (2004, 9.1, 9.2, teilw. 9.3)
Ziel: Finde einen möglichst guten Schätzwert für eine bestimmte Kenngröße ϑ (Parameter) der Grundgesamtheit (GG), z.B. den wahren Anteil der rot/grün-Wähler, den wahren Mittelwert, die wahre Varianz, aber auch z.B. das Maximum (Windgeschwindigkeit...)

2.2.1 Schätzfunktion, typische Beispiele

Gegeben sei die in Kapitel 2.1 beschriebene Situation:
$X_1, \ldots, X_n$ i.i.d. Stichprobe eines Merkmales $\widetilde{X}$

Def 2.1 Sei $X_1, \ldots, X_n$ i.i.d. Stichprobe. „Jede" Funktion

$$T = g(X_1, \ldots, X_n)$$

heißt *Schätzer* oder *Schätzfunktion*.

Bsp 2.2: *Typische Beispiele für Schätzfunktionen*

i) $$\overline{X} = g(X_1, \ldots, X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{2.2.1}$$
arithmetisches Mittel der Stichprobe.
Falls $X_i = \begin{cases} 1 \\ 0 \end{cases} \Longrightarrow \overline{X}$ rel. Häufigkeit des Auftretens von „1" in der Stichprobe

ii) $$\tilde{S}^2 = g(X_1, \ldots, X_n) = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2 \tag{2.2.2}$$
Stichprobenvarianz

iii) $$S^2 = g(X_1, \ldots, X_n) = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2 = \frac{1}{n-1} \left(\sum_{i=1}^{n} X_i^2 - n \cdot (\overline{X})^2 \right) \tag{2.2.3}$$
korrigierte Stichprobenvarianz

iv) $$X_{(n)} = g(X_1, \ldots, X_n) = \max_{i=1,\ldots,n} X_i \tag{2.2.4}$$
größter Stichprobenwert

v) $$X_{(1)} = g(X_1, \ldots, X_n)) = \min_{i=1,\ldots,n} X_i \tag{2.2.5}$$
kleinster Stichprobenwert

Bem 2.3:

(Zur Berechnung von $\widetilde{S}_X^2$ und S^2:)

Häufig ist es einfacher, den Verschiebungssatz heranzuziehen (Siehe Stat I, vgl. auch (1.4.3)). Es ist

$$\tilde{S}_X^2 = \underbrace{\left(\frac{1}{n}\sum_{i=1}^n X_i^2\right)}_{\triangleq \mathbb{E}X_i^2} - \underbrace{\left(\frac{1}{n}\sum_{i=1}^n X_i\right)^2}_{\triangleq (\mathbb{E}X_i)^2} = \left(\frac{1}{n}\sum_{i=1}^n X_i^2\right) - (\overline{X})^2$$

und damit

$$S_X^2 = \frac{n}{n-1}\tilde{S}_X^2 = \frac{1}{n-1}\left(n \cdot \frac{1}{n}\sum_{i=1}^n X_i^2 - n \cdot (\overline{X})^2\right),$$

also

$$S_X^2 = \frac{1}{n-1}\left(\sum_{i=1}^n X_i^2 - n \cdot (\overline{X})^2\right). \tag{2.2.6}$$

Da $X_1, \ldots, X_n$ zufällig sind, ist auch die Schätzfunktion $T = g(X_1, \ldots, X_n)$ **zufällig.** (Zieht man mehrere Stichproben, so erhält man jeweils andere Realisationen von $X_1, \ldots, X_n$, und damit auch von T.)

Die Realisation t (konkreter Wert) der Zufallsvariable T (Variable) heißt *Schätzwert.*

Man hat in der Praxis immer nur eine konkrete Stichprobe und damit auch nur einen konkreten Wert t von T. Zur Beurteilung der mathematischen Eigenschaften werden aber alle denkbaren Stichproben und die zugehörigen Realisationen der Schätzfunktion T herangezogen.

D.h. beurteilt wird **nicht** der einzelne Schätz**wert** als solcher, sondern die Schätz**funktion**, also sozusagen die **Methode** insgesamt.

Bsp: Ausgang der Wahl

Schätzung für Anteil von Rot/Grün: 48.5%

Diese Zahl kann auf verschiedene Art zustandegekommen sein

i) sauber geplante Umfrage: Durchschnittswert unter 1000 Befragten

ii) Zufallszahl aus dem Computer

Mehr Vertrauen in i) als in ii)!

Andere Notation in der Literatur:
$\hat{\vartheta}$ Schätzer und Schätzwert von ϑ, aber nicht klar an Notation erkennbar, wann Zufallsvariable und wann nicht.
$\Longrightarrow$ Schreibe $\hat{\vartheta}(X_1, \ldots X_n)$ bzw. $\hat{\vartheta}(x_1, \ldots x_n)$

Bsp 2.4: (zurechtgestutzt, um mit der Hand rechnen zu können)
Durchschnittliche Anzahl der Statistikbücher in der Grundgesamtheit schätzen
Grundgesamtheit $\widetilde{\Omega} = \{\widetilde{\omega_1}, \widetilde{\omega_2}, \widetilde{\omega_3}\}$: 3 Personen
Stichprobenumfang 2: Stichprobe X_1, X_2 ohne Zurücklegen,
ω_1 erste gezogene Person, ω_2 zweite gezogene Person.

Person	$\widetilde{\omega_1}$	$\widetilde{\omega_2}$	$\widetilde{\omega_3}$	
$\widetilde{X}$:	$\widetilde{X}(\omega_1) = 3$	$\widetilde{X}(\omega_2) = 1$	$\widetilde{X}(\omega_3) = 2$	(durchschnittliche Zahl: $\mu = 2$)

$$X_1 = \widetilde{X}(\omega_1) \qquad X_2 = \widetilde{X}(\omega_2)$$

Betrachte Schätzer:

$$\begin{aligned} T_1 &= g_1(X_1, X_2) = \bar{X} = \frac{X_1 + X_2}{2} \\ T_2 &= X_1 \\ T_3 &= g(X_1, X_2) = \frac{1}{2} X_{(2)} = \frac{\max(X_1, X_2)}{2} \end{aligned}$$

2.2.2 Gütekriterien

Wie gesagt: Beurteile die Schätz**funktionen**, also das Verfahren **an sich**, nicht den einzelnen Schätzwert. Besonders bei komplexeren Schätzproblemen sind klar festgelegten Güteeigenschaften wichtig.

Natürlich auch festzulegen: Was soll geschätzt werden?
Hier stets: Parameter ϑ, eine eindimensionale Kenngröße der Grundgesamtheit (z.B. Mittelwert, Varianz, Maximum)
Da T zufällig ist, kann man nicht erwarten, dass man immer den richtigen Wert trifft, aber den Erwartungswert von T berechnen:

Erwartungstreue, Bias

Def 2.5 Gegeben sei eine i.i.d. Stichprobe $X_1, \ldots, X_n$ und eine Schätzfunktion $T = g(X_1, \ldots, X_n)$ (mit existierendem Erwartungswert).

a) T heißt *erwartungstreu für den Parameter* ϑ, falls gilt

$$\mathbb{E}_\vartheta(T) = \vartheta \tag{2.2.7}$$

für alle ϑ.

b) Die Größe

$$\mathrm{Bias}_\vartheta(T) := \mathbb{E}_\vartheta(T) - \vartheta \tag{2.2.8}$$

heißt Bias (oder Verzerrung) der Schätzfunktion.

(Man schreibt $\mathbb{E}_\vartheta(T)$, $\mathrm{Bias}_\vartheta(T)$, um deutlich zu machen, dass die Größen von dem wahren ϑ abhängen.)

nochmals: Erwartungstreue bedeutet anschaulich:

Erwartungstreue Schätzfunktionen haben per definitionem einen Bias von 0.

Bsp 2.6: (Fortsetzung von Bsp 2.4)
Stichprobenziehung sei so, dass jede Stichprobe dieselbe Wahrscheinlichkeit hat, gezogen zu werden, nämlich $\frac{1}{6}$.

Bsp 2.7: Bias/Erwartungstreue bei einigen typischen Schätzfunktionen

i) $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ ist erwartungstreu für den Mittelwert μ einer Grundgesamtheit:

Da $X_1, \dots X_n$ i.i.d. und $\mathbb{E}_\mu(X_1) = \mathbb{E}_\mu(X_2) = \dots = \mu$ gilt:

$$\begin{aligned}
\mathbb{E}_\mu\left(\frac{1}{n}\sum_{i=1}^n X_i\right) &= \frac{1}{n}\,\mathbb{E}_\mu\left(\sum_{i=1}^n X_i\right) = \\
= \frac{1}{n}\sum_{i=1}^n \mathbb{E}(X_i) &= \frac{1}{n}\sum_{i=1}^n \mu = \frac{1}{n}\cdot n\cdot\mu = \mu
\end{aligned}$$

(Argumentation gilt für jedes μ)

ii) Sei σ^2 die Varianz in der Grundgesamtheit:
Es gilt (ohne Beweis)

$$\mathbb{E}_{\sigma^2}(\tilde{S}^2) = \frac{n-1}{n}\sigma^2,$$

also ist $\tilde{S}^2$ **nicht erwartungstreu** für σ^2:

$$Bias_{\sigma^2}(\tilde{S}^2) = \frac{n-1}{n}\sigma^2 - \sigma^2 = -\frac{1}{n}\sigma^2$$

Aber

$$\begin{aligned}
\mathbb{E}_{\sigma^2}(S^2) &= \mathbb{E}_{\sigma^2}\Big(\frac{1}{n-1}\sum_{i=1}^n \big(X_i - \bar{X}\big)^2\Big) \\
= \quad & \mathbb{E}_{\sigma^2}\Big(\frac{1}{n-1}\cdot\frac{n}{n}\sum_{i=1}^n \big(X_i - \bar{X}\big)^2\Big) = \\
= \quad & \mathbb{E}_{\sigma^2}\Big(\frac{n}{n-1}S^2\Big) = \frac{n}{n-1}\cdot\frac{n-1}{n}\sigma^2 = \sigma^2
\end{aligned}$$

Also S^2 erwartungstreu für σ^2.

iii) Vorsicht: Im Allgemeinen ist für beliebige, nichtlineare Funktionen g

$$\mathbb{E}g(X) \neq g(\mathbb{E}(X))\ ,$$

man kann also nicht einfach z.B. $\sqrt{\cdot}$ und $\mathbb{E}$ vertauschen. In der Tat gilt: S^2 ist zwar erwartungstreu für σ^2, aber $\sqrt{S^2}$ nicht erwartungstreu für $\sqrt{\sigma^2} = \sigma$.

Bsp 2.8 (Wahlumfrage)
Gegeben sei eine Stichprobe der wahlberechtigten Bundesbürger. Geben Sie einen erwartungstreuen Schätzer des Anteils der rot-grün Wähler an.

Aber: Erwartungstreue ist ein schwaches Kriterium!

2.2.3 Effizienz

Bsp 2.9

Zurück zu Bsp 2.8 (Wahlumfrage)

Drei erwartungstreue Schätzer, n sei gerade

$$\begin{aligned} T_1 &= \frac{1}{n} \sum_{i=1}^{n} X_i \\ T_2 &= X_1 \\ T_3 &= \frac{1}{n/2} \sum_{i=1}^{n/2} X_i \end{aligned}$$

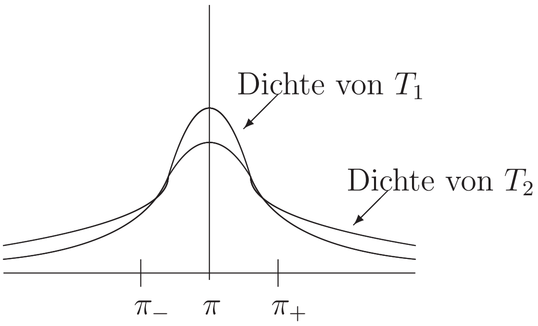
Was unterscheidet T_1 von den unsinnigen Schätzern, die die in der Stichprobe enthaltene Information nur unvollständig ausnutzen?

$$\frac{1}{\sqrt{n}}\left(\frac{\sum_{i=1}^{n}(X_i-\pi)}{\sqrt{\pi(1-\pi)}}\right) \quad\sim\quad N(0;1)$$

$$\text{bzw.}\qquad \frac{\sum_{i=1}^{n} X_i - n\cdot\pi}{\sqrt{\pi(1-\pi)}} \quad=\quad \frac{\frac{1}{n}\sum_{i=1}^{n} X_i - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} \quad\sim\quad N(0;1)$$

$$T_1 = \frac{1}{n}\sum_{i=1}^{n} X_i \quad\sim\quad N\left(\pi;\frac{\pi(1-\pi)}{n}\right)$$

analog $T_2 = \frac{1}{n/2}\sum_{i=1}^{n/2} X_i \sim N\left(\pi, \frac{\pi(1-\pi)}{n/2}\right)$.



T_1 und T_2 sind approximativ normalverteilt, wobei T_1 eine deutlich kleinere Varianz als T_2 (und erst recht als T_3) hat.
T_1 und T_2 treffen beide im Durchschnitt den richtigen Wert π;
T_1 schwankt aber weniger um das wahre π, ist also „im Durchschnitt genauer".

Für jeden Punkt $\pi_+ > \pi$ ist damit $P(T_1 > \pi_+) < P(T_2 > \pi_+)$
und für jeden Punkt $\pi_- < \pi$ ist $P(T_1 < \pi_-) < P(T_2 < \pi_-)$.
Es ist also die Wahrscheinlichkeit, mindestens um $\pi_+ - \pi$ bzw. $\pi - \pi_-$ dane-

ben zu liegen, bei T_2 (und T_3) stets größer als bei T_1. (Ein konkreter Wert ist damit verlässlicher, wenn er von T_1, als wenn er von T_2 (T_3) stammt.) Diese Überlegung gilt ganz allgemein: Ein erwartungstreuer Schätzer ist umso besser, je kleiner seine Varianz ist.

$$\text{Varianz}(T) \;=\; \text{Erwartete quadratische Abweichung von } T \text{ von } \underbrace{\mathbb{E}(T)}_{=\vartheta\,!}$$

Je kleiner die Varianz, umso mehr konzentriert sich die Verteilung eines erwartungstreuen Schätzers um den wahren Wert. Dies ist umso wichtiger, da der Schätzer den wahren Wert i.A. nur selten exakt trifft.

Def 2.10 *Effizienz*

i) Gegeben seien zwei erwartungstreue Schätzfunktionen T_1 und T_2 für einen Parameter ϑ.

$$\begin{array}{llcll} \text{Ist} & Var_\vartheta(T_1) & \leq & Var_\vartheta(T_2) & \text{für alle} \quad \vartheta \\ \text{und} & Var_{\vartheta^*}(T_1) & < & Var_{\vartheta^*}(T_2) & \text{für ein} \quad \vartheta^*, \end{array}$$

so heißt T_1 *effizienter als* T_2.

ii) Eine für ϑ erwartungstreue Schätzfunktion T^* heißt *UMVU-Schätzfunktion* für ϑ (*u*niformly *m*inimum *v*ariance *u*nbiased), falls

$$Var_\vartheta(T^*) \leq Var_\vartheta(T)$$

für alle ϑ und für alle erwartungstreuen Schätzfunktionen T.

Bemerkung:

i) *inhaltliche Bemerkung:*
Der (tiefere) Sinn von Optimalitätskriterien wird klassischerweise insbesondere auch in der *Gewährleistung von Intersubjektivität* („Objektivität“) gesehen. Ohne wissenschaftlichen Konsens darüber, welcher Schätzer in welcher Situation zu wählen ist, wäre die Auswertung einer Stichprobe willkürlich und der Manipulation Tür und Tor geöffnet. Dieser Aspekt tritt aber bei komplexeren Modellen im Rahmen des „Statistical Modelling“ etwas zurück.

ii) Ist $X_1, \ldots, X_n$ eine i.i.d. Stichprobe mit $X_i \sim \mathcal{N}(\mu, \sigma^2)$, dann ist

- $\overline{X}$ UMVU-Schätzfunktion für μ und
- S^2 UMVU-Schätzfunktion für σ^2.

iii) Ist $X_1, \ldots, X_n$ mit $X_i \in \{0, 1\}$ eine i.i.d. Stichprobe mit $\pi = P(X_i = 1)$, dann ist die relative Häufigkeit $\overline{X}$ UMVU-Schätzfunktion für π.

iv) Bei nicht erwartungstreuen Schätzern macht es keinen Sinn, sich ausschließlich auf die Varianz zu konzentrieren:
(unsinniger Schätzer $T = g(X_1, \ldots, X_n) = 1783$, der die Stichprobe nicht beachtet, hat Varianz 0 !)
Man zieht den sogenannten *Mean Square Error*

$$MSE_\vartheta(T) \; := \; \mathbb{E}_\vartheta(T - \vartheta)^2 \tag{2.2.9}$$

zur Beurteilung heran. Es gilt

$$MSE_\vartheta(T) \; = \; Var_\vartheta(T) + (Bias_\vartheta(T))^2 \, . \tag{2.2.10}$$

Der MSE kann als Kompromiss zwischen zwei Auffassungen von Präzision gesehen werden, möglichst geringer systematischer Verzerrung (Bias) und möglichst geringer Schwankung (Varianz).

2.2.4 Konstruktionsprinzipien guter Schätzer

Typische Fragestellung der „wissenschaftlichen Statistik“; hier nur Einblick in die prinzipielle Argumentation. Ganz wichtig für die Behandlung von Nichtstandardsituationen.
Es gibt eine Reihe von Prinzipien, die wichtigsten sind:

a) Die Momentenmethode

Schätze den Mittelwert der Grundgesamtheit durch den Mittelwert der Stichprobe, die Varianz der Grundgesamtheit durch die Varianz der Stichprobe (analog für höhere Momente) und löse nach den Parametern auf.

Problem: Gerade bei etwas komplizierteren Modellen lassen sich die interessierenden Größen häufig nicht als Mittelwert / Varianz ausdrücken. Auflösen ist oft schwierig und liefert dann oft einen nicht erwartungstreuen Schätzer. Zudem: keine direkte Verallgemeinerung auf Regressionsmodelle.

b) Die Methode der kleinsten Quadrate

$\longrightarrow$ Regressionsanalyse

c) Das Maximum-Likelihood-Prinzip

Sehr allgemein, liegt eigentlich fast allen in Software implementierten Prozeduren zugrunde:

Aufgabe: Schätze Parameter ϑ eines parametrischen Modells anhand einer i.i.d. Schätzprobe $X_1, \ldots, X_n$ mit der konkreten Realisation $x_1, \ldots, x_n$.

Idee:

i) Man kann für jedes ϑ die Dichte / Wahrscheinlichkeit ausrechnen, genau diese Stichprobe $x_1, \ldots, x_n$ zu erhalten:

$$P_\vartheta(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \prod_{i=1}^{n} P_\vartheta(X_i = x_i)$$

bzw. analog

$$f_\vartheta(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_\vartheta(x_i)$$

ii) Je größer für ein festes ϑ_0 diese Wahrscheinlichkeit / Dichte ist, umso plausibler ist es, dass tatsächlich ϑ_0 der wahre Wert ist.

Man nennt daher $P_\vartheta(X_1 = x_1, \ldots, X_n = x_n)$, nun als Funktion von ϑ gesehen, die *Likelihood* (deutsch: Plausibilität, Mutmaßlichkeit) von ϑ gegeben die Realisation $x_1, \ldots, x_n$.

Beispiel Wahlumfrage, gesucht π
Stichprobe mit $\overline{X} = 0.47$ beobachtet.

$P(\overline{X} = 0.47)$ bei $\pi = 0.5$ wesentlich größer als $P(\overline{X} = 0.47)$ bei $\pi = 0.25$; also $\pi = 0.5$ plausibler als $\pi = 0.25$.

iii) Def 2.11

- Gegeben sei die Realisation $x_1, \ldots, x_n$ einer i.i.d. Stichprobe. Die Funktion in ϑ

$$Lik(\vartheta||x_1, \ldots, x_n) = \begin{cases} \prod_{i=1}^{n} P_\vartheta(X_i = x_i) & X_i \text{ diskret} \\ & \text{falls} \\ \prod_{i=1}^{n} f_\vartheta(x_i) & X_i \text{ stetig} \end{cases} \tag{2.2.11}$$

heißt *Likelihood* des Parameters ϑ bei der Beobachtung $x_1, \ldots, x_n$.
- Derjenige Wert $\widehat{\vartheta} = \widehat{\vartheta}(x_1, \ldots, x_n)$, der $Lik(\vartheta||x_1, \ldots, x_n)$ maximiert, heißt *Maximum-Likelihood-Schätzwert*; die zugehörige Schätzfunktion $T(X_1, \ldots, X_n)$ *Maximum-Likelihood-Schätzer*.

Einige Bemerkungen

i) nochmals: zwei Sichtweisen auf $P_\vartheta(X_1 = x_1, \ldots, X_n = x_n)$

deduktiv (Wahrscheinlichkeitsrechnung)	ϑ bekannt,	$x_1, \ldots, x_n$ „unbekannt“
induktiv (Statistik)	ϑ unbekannt,	$x_1, \ldots, x_n$ bekannt

ii) In dem Wahlbeispiel ist $\overline{X}$ in der Tat Maximum-Likelihood-Schätzer für den Anteil π.

iii) Für die praktische Berechnung maximiert man statt (5.3.1) meistens

$$\ln \prod_{i=1}^{n} P_\vartheta(X_i = x_i) = \sum_{i=1}^{n} \ln P_\vartheta(X_i = x_i)$$

bzw.

$$\ln \prod_{i=1}^{n} f_\vartheta(x_i) = \sum_{i=1}^{n} \ln f_\vartheta(x_i)\ .$$

Dies liefert denselben Schätzwert $\hat{\vartheta}$ und erspart beim Differenzieren die Anwendung der Produktregel.

$$\left(\begin{array}{l} x_0 \text{ Stelle des Maximums von } g(x) > 0 \iff \\ x_0 \text{ Stelle des Maximums von } \ln(g(x)) \end{array} \right)$$

2.3 Intervallschätzung/Konfidenzintervalle

Fahrmeir et al. (2003, Kap. 9.4)
Jann (2002, Kap. 5.5.2)

2.3.1 Motivation und Hinführung

Bsp. 2.12 (Wahlumfrage)

Der wahre Anteil der rot-grün Wähler 2002 war genau 47.1%. Wie groß ist die Wahrscheinlichkeit, in einer Zufallsstichprobe von 1000 Personen genau einen relativen Anteil von 47.1% von rot-grün Anhängern erhalten zu haben?

D.h.: Mit Wahrscheinlichkeit von etwa 100%-2.567%, also etwa 97.5%, verfehlt der (UMVU!)-Schätzer den wahren Wert.
Hat man ein stetiges Merkmal, so ist sogar $P(\bar{X} = a) = 0$ für jedes a, d.h. der wahre Wert wird mit Wahrscheinlichkeit 1 verfehlt.

Was tun?

- Insbesondere Vorsicht bei der Interpretation „knapper Ergebnisse“ (z.B Anteil 50.2%)
- vgl. Ausführungen in 2.2: Suche Schätzer mit möglichst kleiner Varianz, um „im Durchschnitt möglichst nahe dran zu sein“
- Ferner: es ist meist auch gar nicht nötig, sich genau auf einen Wert festzulegen; oft reicht die Angabe eines Intervalls, von dem man hofft, dass es den wahren Wert überdeckt: *Intervallschätzung*

Symmetrische Intervallschätzung basierend auf einer Schätzfunktion $T = g(X_1, \ldots X_n)$

$$I(T) = [T - a, T + a]$$

z.B. $I(\overline{X}) = [\bar{X} - a, \bar{X} + a]$ (etwa in obigem Beispiel relativer Anteil $\pm 2\%$)

„Trade off“ bei der Wahl von a:

Wie soll man a wählen?

Typisches Vorgehen:

i) Man gebe sich einen Sicherheitsgrad (*Konfidenzniveau*) γ vor

ii) und konstruiere dann das Intervall so, dass es mindestens mit der Wahrscheinlichkeit γ den wahren Parameter überdeckt.

2.3.2 Definition von Konfidenzintervallen

Def 2.13
Gegeben sei eine i.i.d. Stichprobe $X_1, \ldots, X_n$ zur Schätzung eines Parameters ϑ und eine Zahl $\gamma \in (0; 1)$. Ein zufälliges Intervall $\mathcal{C}(X_1, \ldots, X_n)$ heißt *Konfidenzintervall* zum *Sicherheitsgrad* γ (Konfidenzniveau γ), falls für jedes ϑ

$$P_\vartheta(\vartheta \in \underbrace{\mathcal{C}(X_1, \ldots, X_n)}_{\text{zufälliges Intervall}}) \geq \gamma. \qquad (2.3.1).$$

Bem. 2.14:

i) In (2.3.1) steht die Wahrscheinlichkeit, dass das zufällige Intervall den festen, wahren Parameter überdeckt. (Streng genommen darf man, von dem üblichen objektivistischen Verständnis von Wahrscheinlichkeit ausgehend, nicht von der *Wahrscheinlichkeit* sprechen, „dass ϑ in dem Intervall liegt“, da ϑ nicht zufällig ist und somit keine Wahrscheinlichkeitsverteilung besitzt.)

ii) Typischerweise konstruiert man Konfidenzintervalle wie in Abschnitt 2.3.1, also symmetrisch um einen Schätzer T.
Es sind aber auch manchmal z.B. einseitige Konfidenzintervalle, etwa der Form
$$[\bar{X}, \bar{X} + b],$$
sinnvoll. (Beispielsweise: Bei sozial unerwünschten Ereignissen (Antwortverzerrung!))

2.3.3 Typische Beispiele für die Konstruktion von Konfidenzintervallen

Für die Konstruktion praktische Vorgehensweise:
Suche Zufallsvariable Z, die

a) den gesuchten Parameter ϑ enthält und

b) deren Verteilung aber nicht mehr von dem Parameter abhängt, (*„Pivotgröße“*, dt. Angelpunkt)

dann:

c) Wähle Bereich C_Z so, dass $P_\vartheta(Z \in C_Z) = \gamma$

d) „nach ϑ auflösen“.

Bsp 2.15 Konfidenzintervall für den Mittelwert eines normalverteilten Merkmals bei *bekannter Varianz*
$X_1, \dots X_n$ i.i.d., $X_i \sim N(\mu, \sigma^2)$, $\quad \sigma$ bekannt.

ceteris paribus Betrachtung (eine Größe verändern, Rest festhalten;)

- Je größer σ, desto

- Je größer γ, desto

- Je größer n und damit $\sqrt{n}$, desto

Bsp 2.16: Konfidenzintervall für den Mittelwert eines normalverteilten Merkmals bei unbekannter Varianz

Was tun, wenn auch σ^2 unbekannt?
Man kann σ^2 schätzen, nämlich durch den UMVU-Schätzer

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

aber (mit $S = \sqrt{S^2}$)

$$Z = \frac{\bar{X} - \mu}{S} \cdot \sqrt{n}$$

ist nicht mehr normalverteilt, denn

neues Verteilungsmodell:

Def 2.17 Gegeben sei eine i.i.d. Stichprobe $X_1, \ldots, X_n$ mit $X_i \sim \mathcal{N}(\mu, \sigma^2)$. Dann heißt die Verteilung von

$$Z = \frac{\overline{X} - \mu}{S} \cdot \sqrt{n}$$

t-Verteilung (oder Student-Verteilung) mit $\nu = n - 1$ *Freiheitsgraden.*
In Zeichen: $Z \sim t(\nu)$

Wichtige Werte der t-Verteilung sind tabelliert; angegeben ist, für verschiedene δ, die Lösung t_δ der Gleichung

$$P(Z \leq t_\delta^{(\nu)}) = \delta,$$

wobei $t_\delta^{(\nu)}$ von der Anzahl ν der Freiheitsgrade abhängt.

Die Dichte einer t-Verteilung ist der Dichte der Standardnormalverteilung sehr ähnlich: sie ist auch symmetrisch um 0, besitzt aber etwas breitere „Flanken".

Je größer ν ist, umso ähnlicher sind sich die $t(\nu)$-Verteilung und die Standardnormalverteilung; für $\nu \to \infty$ sind sie gleich, ab $\nu = 30$ gilt der Unterschied als vernachlässigbar.

<u>Ansatz:</u> Konfidenzintervall zum Konfidenzniveau γ:

- Analoge ceteris paribus Betrachtung zu oben!

- Ferner: Vergleich mit (2.3.2)

- Je größer ν, umso kleiner ist der Unterschied; wie gesagt, für $n \geq 30$ rechnet man einfach auch bei der t-Verteilung mit $z_{\frac{1+\gamma}{2}}$.

Bsp. 2.18a

Eine Maschine füllt Gummibärchen in Tüten ab, die laut Aufdruck 250g Füllgewicht versprechen. Wir nehmen im folgenden an, dass das Füllgewicht normalverteilt ist. Bei 16 zufällig aus der Produktion herausgegriffenen Tüten wird ein mittleres Füllgewicht von 245g und eine Stichprobenstreuung (Standardabweichung) von 10g festgestellt.

a) Berechnen Sie ein Konfidenzintervall für das mittlere Füllgewicht zum Sicherheitsniveau von 95%.

b) Wenn Ihnen zusätzlich bekannt würde, daß die Stichprobenstreuung gleich der tatsächlichen Streuung ist, wäre dann das unter a) zu berechnende Konfidenzintervall für das mittlere Füllgewicht breiter oder schmäler?
Begründen Sie ihre Antwort ohne Rechnung.

b) Jetzt ist auch σ bekannt.
Man kann mit dem Konfidenzintervall aus (2.3.2) rechnen:

$$[\bar{X} \pm z_{\frac{1+\gamma}{2}} \cdot \frac{\sigma}{\sqrt{n}}]$$

Da jetzt σ bekannt, ist die Unsicherheit geringer und damit das Konfidenzintervall schmäler.
In der Tat ist $z_{\frac{1+\gamma}{2}} < t_{\frac{1+\gamma}{2}}$.
(Rechnerisch ergäbe sich mit $z_{\frac{1+\gamma}{2}} = 1.96$:
$[240.100; 249.900]$.)

Approximative Konfidenzintervalle

Ist der Stichprobenumfang groß genug, so kann wegen des zentralen Grenzwertsatzes die Formel (2.3.2) auf beliebige Merkmale (mit existierender Varianz) angewendet werden und erhält approximative Konfidenzintervalle.

Insbesondere:

Bsp 2.18b Konfidenzintervall für den Anteil π
Situation $X_1, \ldots X_n$ i.i.d. mit

$$X_i = \begin{cases} 1 \\ 0 \end{cases} , \quad P(X_i = 1) = \pi.$$

z.B. Wahlumfrage: Intervall angeben, das mit 95% Sicherheit den wahren Anteil π der rot-grün Wähler überdeckt.

Beispielsweise: Wahlumfrage

$n = 500$, $\bar{X} = 46.5\%$ (Anteil rot-grün Wähler), $\gamma = 95\%$