

# Einführung in die Statistik

Dr. C.J. Luchsinger

## 7 Crash-Course in Statistics II: Schätztheorie und Konfidenzintervalle

### Literatur Kapitel 7

- \* Statistik in Cartoons: Kapitel 6 und 7
- \* Krengel: § 4 und 13
- \* Storrer: 43

### 7.1 Schätzer

**Ziele von 7.1:** Die StudentInnen wissen, welche Anforderungen an Schätzer sinnvollerweise gestellt werden. Sie kennen die ML-Schätzmethode.

Was wir in 7.1 erarbeiten werden, ist vielen StudentInnen vielleicht schon einmal durch den Kopf gegangen - es geht jetzt vor allem darum, diese Gedanken zu ordnen.

#### 7.1.1 Motivierendes Beispiel

Gegeben sei eine Stichprobe vom Umfang 11; wir wissen, sie stammt aus einer Normalverteilung mit Varianz 1. Der Mittelwert ist leider unbekannt. Wir wollen diesen Mittelwert schätzen. Die Stichprobe sieht folgendermassen aus (bereits der Reihe nach geordnet und auf 2 Stellen nach dem Komma gerundet):

-0.21, 0.11, 0.39, 0.64, 1.24, 1.46, 1.48, 1.51, 1.95, 2.89, 3.53

Wie könnten wir  $\mu$  schätzen?

Wann wollen wir einen Schätzer für  $\mu$  als "gut" bezeichnen?

Bezeichnen wir einen Schätzer für  $\mu$  mit  $\hat{\mu}_n$ , wenn wir eine Stichprobe vom Umfang  $n$  haben. Des weiteren bezeichnen wir mit  $x_{(i)}$  die  $i$ 't kleinste Zahl aus der Stichprobe, wenn wir die Stichprobe der Grösse nach ordnen. Intuitiv denkt man vielleicht an folgende Vorschläge, wie man  $\mu$  schätzen könnte.

\*  $\hat{\mu}_{11}^{(1)} = \bar{x} := (1/11) \sum_{j=1}^{11} x_j$  [hier 1.36] (R Befehl: mean(x))

\*  $\hat{\mu}_{11}^{(2)} := \text{median}(x_1, \dots, x_{11})$  [hier 1.46] (R Befehl: median(x))

\*  $\hat{\mu}_{11}^{(3)} := (1/9) \sum_{j=2}^{10} x_{(j)}$  [hier 1.30] (R Befehl: mean(x, trim=1/11))

\*  $\hat{\mu}_{11}^{(4)} := (x_{(1)} + x_{(11)})/2$  [hier 1.66]

\*  $\hat{\mu}_{11}^{(5)} := 1.47$  [hier 1.47]

$\hat{\mu}_n^{(1)}$  ist derjenige Schätzer, welcher irgendwie "kanonisch" erscheint und einem meist zuerst in den Sinn kommt. In der Tat ist dieser Schätzer in einem noch zu spezifizierenden Sinn optimal (mehr dazu in der Vorlesung "Statistische Methoden").  $\hat{\mu}_n^{(3)}$  (und Abwandlungen davon) wird häufig verwendet mit dem (in dieser Kürze zweifelhaften) Argument, man wolle "Ausreisser" ausschliessen.  $\hat{\mu}_n^{(5)}$  erscheint wie ein Druckfehler in dieser Auflistung. Wenn man sukzessive den Umfang der Stichprobe (das  $n$ ) erhöht, so wird  $\hat{\mu}_n^{(1)}$  immer wieder schwanken (wenn auch immer weniger stark), ebenso alle anderen Schätzer ausser  $\hat{\mu}_n^{(5)}$ .  $\hat{\mu}_n^{(5)}$  hat also den gewaltigen Vorteil, dass er exakt *eine* Zahl ist mit Varianz 0. Trotzdem ist uns vielleicht unwohl,  $\hat{\mu}_n^{(5)}$  als Schätzer für  $\mu$  zu bezeichnen. Aber in der Tat:  $\hat{\mu}_n^{(5)}$  ist ein Schätzer für  $\mu$ . Damit kommen wir an die Stelle, wo wir definieren müssen, was wir unter einem Schätzer verstehen wollen. ( $\mu$  war übrigens oben  $\sqrt{2}$ ).

### 7.1.2 Definition Schätzer

**Definition 7.1 [Schätzer, engl. Estimator]** Ein Schätzer  $\hat{\mu}_n := \hat{\mu}_n(x_1, \dots, x_n)$  für einen unbekannt Parameter  $\mu$  ist eine beliebige Funktion der Daten. Insbesondere kann die Schätzfunktion die Daten vollständig ignorieren ( $\hat{\mu}_n^{(5)}$ ).

**Bemerkung zu Definition 7.1:**  $\hat{\mu}_n := \hat{\mu}_n(x_1, \dots, x_n)$  ist eine Zahl und keine Zufallsgrösse, da die  $(x_1, \dots, x_n)$  ja auch feste Zahlen sind (Bsp:  $(1/n) \sum_{j=1}^n x_j$ ). Aber:  $\hat{\mu}_n(X_1, \dots, X_n)$  ist eine Zufallsgrösse (Bsp:  $(1/n) \sum_{j=1}^n X_j$ ). Wir werden hier beide mit  $\hat{\mu}_n$  bezeichnen - es gibt keine Verwechslungen.

### 7.1.3 Anforderungen an Schätzer

Obige Vorschläge zeigen, dass es unendlich viele Möglichkeiten gibt,  $\mu$  zu schätzen. Wir wollen uns jetzt darum bemühen, einen “guten” Schätzer zu finden. Wir müssen zuerst definieren, was wir unter “gut” verstehen wollen. Wir müssen sinnvolle Anforderungen oder Kriterien finden. Dies ist aber nicht so klar, wie man a priori denken könnte. Zum Teil ist es nicht möglich, dass ein Schätzer alle nachfolgenden Kriterien einhält; andererseits folgen aus einigen Kriterien andere. Man beachte nochmals, dass ein Schätzer als Funktion der Zufallsgrössen auch eine Zufallsgrösse ist und damit Erwartungswerte, Konvergenzverhalten und die Varianz eines Schätzers betrachtet werden können.  $P_\mu$  bezeichne dabei die Wahrscheinlichkeit, wenn der Parameter  $\mu$  der richtige ist, also wenn  $X_1, \dots, X_n$  Zufallsgrössen mit einer Verteilung sind, welche vom Parameter  $\mu$  abhängt (z.B.  $\mathcal{N}(\mu, 1)$ ); analog  $E_\mu$ . Die folgenden Anforderungen an einen Schätzer sind denkbar; wir werden in der Vorlesung mit dem Erwartungswert einer Normalverteilung die Problematik illustrieren (in den Übungen sind dann auch andere Verteilungen zu untersuchen):

**Definition 7.2** [erwartungstreu (unverfälscht), engl. unbiased; Bias] Ein Schätzer  $\hat{\mu}_n$  für  $\mu$  heisst erwartungstreu oder unverfälscht, wenn

$$E_\mu[\hat{\mu}_n] = \mu$$

für alle  $\mu$  im Parameterraum. Ansonsten spricht man davon, dass der Schätzer einen Bias  $b$  hat; wir definieren

$$b := E_\mu[\hat{\mu}_n - \mu].$$

Untersuchen Sie den Schätzer  $\hat{\mu}_{11}^{(1)}$  aus 7.1.1 auf Erwartungstreue, wenn die Stichprobe aus einer  $\mathcal{N}(\mu, 1)$ -Verteilung stammt mit unbekanntem  $\mu$ .

Welche der 5 Schätzer aus 7.1.1 sind erwartungstreu? Welche nicht? Warum?

**Definition 7.3 [Konsistenz]** Ein Schätzer  $\hat{\mu}_n$  für  $\mu$  ist konsistent, wenn für alle  $\mu$  im Parameterraum,  $\epsilon > 0$  gilt:

$$\lim_{n \rightarrow \infty} P_\mu [|\hat{\mu}_n - \mu| > \epsilon] = 0.$$

Untersuchen Sie den Schätzer  $\hat{\mu}_n^{(1)}$  aus 7.1.1 auf Konsistenz, wenn die Stichprobe aus einer  $\mathcal{N}(\mu, 1)$ -Verteilung stammt mit unbekanntem  $\mu$ . Geben Sie die notwendigen Argumente exakt an.

**Definition 7.4 [kleinste Varianz]** *Ein Schätzer  $\hat{\mu}_n$  für  $\mu$  hat kleinste Varianz für  $n$  fix, wenn die Varianz von  $\hat{\mu}_n$  unter  $P_\mu$  minimal ist für alle  $\mu$  im Parameterraum.*

**Bemerkungen zu Definition 7.4:** Diese Anforderung bedeutet, dass die Realisationen des Schätzers alle möglichst beieinander liegen; sie müssen aber nicht nahe beim zu schätzenden Wert liegen (siehe  $\hat{\mu}_n^{(5)}$  in 7.1.1).

Man kann zeigen (Vlsg SM): Wenn wir eine Stichprobe aus einer Normalverteilung haben, dann hat unter den erwartungstreuen Schätzern  $\hat{\mu}_n^{(1)}$  die kleinstmögliche Varianz.

Um den Schätzer  $\hat{\mu}_n^{(5)}$  hier auszuschalten, könnte man fordern, dass der Schätzer *gleichzeitig* erwartungstreu *und* von kleinster Varianz ist. Der nachfolgende MSE zielt in diese Richtung (siehe auch Lemma 7.6); der MSE ist unter mathematischen StatistikerInnen am etabliertesten:

**Definition 7.5 [minimaler Mean-Square-Error (MSE)]** *Ein Schätzer  $\hat{\mu}_n$  für  $\mu$  minimiert den Mean-Square-Error (MSE) für  $n$  fix, wenn*

$$MSE(\hat{\mu}_n, \mu) := E_\mu[(\hat{\mu}_n - \mu)^2]$$

*minimal ist für alle  $\mu$  im Parameterraum.*

Auf die Schnelle könnte man meinen, MSE sei die Varianz des Schätzers. Dies stimmt jedoch nur, wenn der Schätzer erwartungstreu ist (Bias  $b = 0$ ), wie wir auch aus folgendem Lemma schliessen können:

**Lemma 7.6** [ $MSE = V + b^2$ ] *Mit obigen Bezeichnungen gilt:*

$$MSE(\hat{\mu}_n, \mu) = V_\mu[\hat{\mu}_n] + b^2.$$

**Beweis von Lemma 7.6** Aufgabenblatt 12.

□

Man kann zeigen: Der MSE-minimale Schätzer in Situation 7.1.1 ist  $\hat{\mu}_n^{(1)}$ .

Daneben gibt es noch die schwieriger zu definierende Eigenschaft der **Robustheit**. Es geht dabei um die Anforderung, dass falsche Messungen die Schätzungen für  $\mu$  nicht stark beeinflussen dürfen. Wir sind dann aber eigentlich gar nicht mehr z.B. im Modell  $\mathcal{N}(\mu, \sigma^2)$ .  $\hat{\mu}_n^{(2)}$  ist ein sehr robuster Schätzer. Betrachten wir dazu einen Vergleich von  $\hat{\mu}_n^{(1)}$  und  $\hat{\mu}_n^{(2)}$ . Wenn man in  $(x_1, \dots, x_n)$  selbst bei grossem  $n$  auch nur einen Wert total verfälscht ("nach  $10^9$  und mehr schickt"), dann wird  $\hat{\mu}_n^{(1)}$  beliebig falsch. Hingegen ändert sich  $\hat{\mu}_n^{(2)}$  kaum. In der Tat kann man bei  $\hat{\mu}_n^{(2)}$  bis zu  $< 50\%$  aller Daten total verfälschen; die Schätzung hat immer noch "etwas mit dem richtigen Wert zu tun". Dieses "etwas mit dem richtigen Wert zu tun" haben lässt sich mathematisch sauber ausformulieren (Bruchpunkt), wir verzichten hier aus Zeitgründen darauf. Das Forschungsgebiet heisst *robuste Statistik*, ist sehr wichtig, komplex, kompliziert, aktuell und benötigt sehr gute Vorkenntnisse in *reiner* Mathematik.

Abschliessend kann man sagen, dass bei  $\mathcal{N}(\mu, 1)$   $\hat{\mu}_n^{(1)}$  bei 100%-Datenqualität (keine fehlerhafte Daten) der beste Schätzer ist (unverfälscht, konsistent, kleinste Varianz unter den unverfälschten Schätzern, minimaler MSE). Man kann zeigen:  $\sigma$  muss dabei nicht 1 sein -  $\sigma$  muss nicht mal bekannt sein.

#### 7.1.4 MLE: Maximum Likelihood Estimator

Wir behandeln hier lediglich den MLE, in der Vorlesung "Statistische Methoden" werden wir weitere allgemeine Schätzverfahren kennenlernen.

In Kapitel 6 haben wir bei der Behandlung des Lemmas von Neyman-Pearson (NP) bereits kurz den Begriff der "Likelihood" erwähnt (den Test nach NP nennt man auch Likelihood-Ratio-Test). Dort haben wir den Quotienten der Dichten (oder der Wahrscheinlichkeitsfunktionen) angeschaut. Die Daten waren dabei fest, der Parameter war variabel, einmal z.B.  $\mu_0$ , dann  $\mu_1$ . Dies ist genau die Likelihood:

**Definition 7.7 [Likelihood]** *Wenn wir eine Dichtefunktion  $f_\mu(x_1, \dots, x_n)$  oder Wahrscheinlichkeitsfunktion  $p_\theta(x_1, \dots, x_n)$  als Funktion des Parameters  $\mu$  resp.  $\theta$  auffassen bei konstanten Daten, dann spricht man von der Likelihood.*

Man kann auch sagen, dass der Begriff der Likelihood aus der Not entstanden ist: wir haben ja nur die Daten (die können wir nicht mehr variieren) und gehen mal davon aus, dass die Daten zum Beispiel aus einer Normalverteilung stammen. Dann können wir nur noch die Parameter variieren und schauen, was passiert.

Wir haben noch nicht die *Methode*:

$$\hat{\mu}_n^{MLE} := \operatorname{argmax}_{\mu \in \mathbb{R}} f_\mu(x_1, \dots, x_n),$$

analog im Fall von Wahrscheinlichkeitsfunktionen.

Wie ist im Fall  $\mathcal{N}(\mu, 1)$  der MLE? Von 6.2.2 haben wir mit  $\sigma = 1$

$$f_\mu(x_1, \dots, x_n) = \left( \frac{1}{\sqrt{2\pi}} \right)^n \exp \left[ -\frac{1}{2} \left( \sum_{i=1}^n x_i^2 - 2\mu n\bar{x} + n\mu^2 \right) \right].$$

Da wir nur das argmax suchen, können wir ebenso gut den Ausdruck

$$\exp \left[ -\frac{1}{2} \left( \sum_{i=1}^n x_i^2 - 2\mu n\bar{x} + n\mu^2 \right) \right]$$

maximieren. Da der Logarithmus die Ordnung nicht verändert (und wir nur argmax suchen), können wir ebenso gut den Ausdruck

$$-\frac{1}{2} \left( \sum_{i=1}^n x_i^2 - 2\mu n\bar{x} + n\mu^2 \right)$$

maximieren. Dieser Schritt (Logarithmieren) ist zentral wichtig in vielen Rechnungen der Statistik, um das Leben einfacher zu machen. Wir müssen letztendlich den Ausdruck

$$-2\mu\bar{x} + \mu^2$$

*minimieren*, Daten fest,  $\mu$  variabel.

Dieses Resultat gilt übrigens für alle  $\sigma$ .

Auf Blatt 12 müssen Sie den MLE bei der Exponentialverteilung finden. Die ML-Methode kann man analog auch auf diskrete Verteilungen anwenden. Man nimmt dann anstelle der Dichte die Wahrscheinlichkeitsfunktion. Auch dazu ist auf Blatt 12 eine Aufgabe zu lösen.

Wie gut die ML-Methode für Schätzprobleme ist (Vor- und Nachteile), wird in der Vorlesung "Statistische Methoden" genauer untersucht. Kurz: die Methode hat sich bewährt.



### 7.1.5 Die Schätzung von $\sigma^2$ und $\sigma$ im Modell $\mathcal{N}(\mu, \sigma^2)$

Im Modell  $\mathcal{N}(\mu, \sigma^2)$  haben wir also mit  $\hat{\mu}_n^{(1)}$  einen ausgezeichneten Schätzer für den unbekannt Parameter  $\mu$ . Falls wir uns aber für den unbekannt Parameter  $\sigma^2$  interessieren (also die Varianz), so sieht die Sache anders aus. Bekannt sind für dieses Problem etwa folgende Schätzer:

$\mu$  unbekannt

$$* \hat{\sigma}_n^{2(1)} := (1/n) \sum_{j=1}^n (x_j - \bar{x})^2$$

$$* \hat{\sigma}_n^{2(2)} := [1/(n-1)] \sum_{j=1}^n (x_j - \bar{x})^2 \text{ (R Befehl: var(x))}$$

$$* \hat{\sigma}_n^{2(3)} := [1/(n+1)] \sum_{j=1}^n (x_j - \bar{x})^2$$

$$* \hat{\sigma}_n^{2(4)} = (1.4826 \times \text{MAD})^2 := (1.4826 \times \text{median}(|x_i - \text{median}(x_1, \dots, x_n)| \ i = 1, \dots, n))^2$$

“MAD” steht für “Median of Absolut Deviations” (R Befehl: mad(x)).

Hier gestaltet sich die Sache von den Kriterien her beinahe *chaotisch* (meiste Aussagen ohne Beweis):

$\hat{\sigma}_n^{2(2)}$  ist ein erwartungstreuer Schätzer (Aufgabenblatt 12). Der “naiv” sich aufdrängende Schätzer  $\hat{\sigma}_n^{2(1)}$  ist also insbesondere *nicht* erwartungstreu.  $\hat{\sigma}_n^{2(4)}$  ist (fast) erwartungstreu. Der Faktor “1.4826” wird gerade deshalb eingesetzt, damit der Schätzer  $\hat{\sigma}_n^{2(4)}$  auch (fast) erwartungstreu ist (gilt nur im Fall der Normalverteilung). Das Wort “fast” kommt daher, dass dieser vorgeschaltete Faktor (1.4826) unendlich viele Stellen nach dem Komma hat. Alle obigen Schätzer sind zumindest asymptotisch erwartungstreu und konsistent.  $\hat{\sigma}_n^{2(3)}$  ist derjenige Schätzer, welcher den MSE minimiert.  $\hat{\sigma}_n^{2(4)}$  ist ein sehr robuster Schätzer.  $\hat{\sigma}_n^{2(1)}$  ist der Maximum-Likelihood-Schätzer. Es gibt in dieser Situation also insbesondere keinen herausragenden Schätzer, welcher fast alle Kriterien als einziger erfüllt, wie wir das mit  $\bar{x}$  im Fall der Schätzung von  $\mu$  hatten.

$\mu$  bekannt

Will man hingegen  $\sigma^2$  schätzen und  $\mu$  ist bekannt, so ist es sinnvoll, in der Summe der Quadrate  $\bar{x}$  durch  $\mu$  zu ersetzen, also die Summe

$$\sum_{j=1}^n (x_j - \mu)^2$$

zu betrachten. Wenn man diese Summe durch  $n$  teilt, so ist dieser Schätzer erwartungstreu und hat minimale Varianz unter den erwartungstreuen Schätzern - es ist übrigens der MLE. Den MSE-minimale Schätzer erhält man, indem man diese Summe durch  $n + 2$  teilt. Selbst wenn  $\mu$  bekannt ist, haben wir keinen herausragenden Schätzer; immerhin gibt es hier (bezüglich obiger Kriterien) nur 2 Kandidaten: entweder

$$\frac{1}{n} \sum_{j=1}^n (x_j - \mu)^2 \quad \text{oder} \quad \frac{1}{n+2} \sum_{j=1}^n (x_j - \mu)^2.$$

Man kann zeigen, dass

$$\sqrt{\frac{1}{n-1} \sum_{j=1}^n (x_j - \bar{x})^2} \quad \text{beziehungsweise} \quad \sqrt{\frac{1}{n} \sum_{j=1}^n (x_j - \mu)^2}$$

*keine* erwartungstreuen Schätzer für  $\sigma$  selber sind. Sie werden aber eingesetzt (vgl. Kapitel 6, z.B. t-Test).

### 7.1.6 Abschliessende Bemerkungen zum Schätzproblem

Es stellt sich immer die Frage, welchen Schätzer man einsetzen soll. Wichtig ist sicher die Forderung nach Konsistenz. Früher wurde immer die Unverfälschtheit gefordert, dieses Kriterium ist nicht mehr so en vogue. Sinnvoll ist sicher die Minimierung des MSE. Heutzutage wird auch mehr auf Robustheit geachtet.

Bei der Frage, welchen Schätzer man einsetzen soll, muss man sich folgende Fragen stellen:

“Wozu wird der Schätzer gebraucht?”

“Was soll mit dem Schätzer gemessen werden?”

Welches Kriterium auch immer gewählt wird: wichtig ist, dass man angibt, nach welcher Methode ein Parameter geschätzt wird (und welche Annahmen bei der Modellbildung gemacht werden). Dies ist sowieso bei allen statistischen Methoden zentral.

Hat man keine Normalverteilung oder Abhängigkeitsstrukturen in den Daten (Zeitreihen), so müssen obige Schätzvorschläge für Lage- und Varianzparameter oft durch andere Schätzer ersetzt werden.

Andere Vorschläge für Schätzer müssen immer zuerst auf obige Kriterien hin untersucht werden.

## 7.2 Konfidenzintervalle (Intervall-Schätzer)

... geben uns eine Vorstellung von der Präzision eines Schätzers.

**Ziele von 7.2:** Die StudentInnen wissen, was ein Konfidenzintervall (KI) ist - und was es *nicht* ist. Sie kennen das Konstruktionsrezept für KI und erkennen Formeln für KI in 2 wichtigen Fällen wieder und können diese anwenden.

### 7.2.1 Was ist ein Konfidenzintervall (KI) - was ist es *nicht*

Was ist es *nicht*: In den Nachrichten und wissenschaftlichen Publikationen liest man oft Sätze der Art (Zahlen frei erfunden): "Aufgrund einer Befragung mit einer Stichprobe von 10'000 Personen kam man zum Schluss, dass der Anteil der Anhänger von Frau Merkel mit 95 % Wahrscheinlichkeit in einem Konfidenzintervall von [46%, 48%] liegt." Was ist hier falsch?

**Definition 7.8 [Konfidenzintervalle]** Ein Konfidenzintervall  $KI$  für  $\theta$  mit Konfidenzkoeffizient  $(1-\alpha)$  ist eine zufällige Teilmenge des Parameterraums mit der Eigenschaft, dass

$$P_{\theta}[\theta \in KI] = 1 - \alpha$$

für alle  $\theta$  des Parameterraums (z.B.  $\forall \theta \in \mathbb{R}$ ). [siehe auch Cartoon Guide p. 120]

**Bemerkungen zu Definition 7.8:** Im Teil "was ist es *nicht*" haben wir bereits betont, dass  $\theta$  nicht zufällig ist. Das Konfidenzintervall  $KI$  ist zufällig (*vor* der Realisation). Wenn danach die Realisation mit konkreten Zahlen vorliegt, sollte man Formulierungen brauchen wie: "[46%, 48%] ist eine Realisation eines 95 % Konfidenzintervalles". Aber auch vor der Realisation sollte man nicht sagen: " $\theta$  liegt mit Wahrscheinlichkeit  $(1 - \alpha)$  in  $KI$ ", sondern "KI deckt mit Wahrscheinlichkeit  $(1 - \alpha)$  den Parameter  $\theta$  ab" - um zu betonen, dass der Zufall (die Bewegung) in  $KI$  liegt und nicht in  $\theta$ .

### 7.2.2 Konstruktionsrezept für KI

Wie in 6.4 "Heuristische Methoden" (Testtheorie) gibt es auch hier optimale Konstruktionen von KI und heuristische Ansätze. Es würde zu weit führen, jetzt zu definieren, was bei KI optimal bedeutet. Wir werden in der Vorlesung "Statistische Methoden" die optimalen Methoden vertieft studieren. Das nachfolgende Kochrezept liefert jedoch in beiden Fällen (7.2.3 und 7.2.4) trotz heuristischem Ansatz optimale KI.

#### Kochrezept für ein KI für $\theta$

1. Mit welcher Statistik würde man  $\theta$  schätzen?
2. Umformen (zum Beispiel zentrieren / normieren), bis man (meist stabile) bekannte Verteilung hat
3. kritische Werte (untere/obere 2.5 %) von *bekannter* Verteilung
4. Umformen, bis man es als KI verkaufen kann

### 7.2.3 $\mathcal{N}(\mu, \sigma^2)$ : KI für $\mu$ wenn $\sigma^2$ bekannt (optimal)

Daten  $x_1, \dots, x_n$ , berechnen arithmetisches Mittel  $\bar{x}$ . Wechseln zu Zufallsgrößen: betrachten  $\bar{X}$ . Verteilung von  $\bar{X}$  ist:

Erwartungswert=

Varianz=

Verteilungsart:

Zusammen:

Also ist

$$\frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$$

wegen der “Z-Transform” eine  $\mathcal{N}(0, 1)$ -Zufallsgröße. Aber das ist ja fantastisch: wir kennen die kritischen Werte bei einer  $\mathcal{N}(0, 1)$ -Verteilung:

$$P\left[-1.96 \leq \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \leq 1.96\right] = 95\%. \quad (7.1)$$

Ziel ist offensichtlich ein  $(1 - \alpha) = 95\%$ -KI! Aber: wir können offensichtlich (7.1) (noch) nicht als KI verkaufen. Machen wir ein paar algebraische Umformungen (Resultat ist dann (7.6)):

$$P\left[-1.96\sqrt{\frac{\sigma^2}{n}} \leq \bar{X} - \mu \leq 1.96\sqrt{\frac{\sigma^2}{n}}\right] = 95\%. \quad (7.2)$$

Wechseln zu den absoluten Werten:

$$P\left[\left|\bar{X} - \mu\right| \leq 1.96\sqrt{\frac{\sigma^2}{n}}\right] = 95\%, \quad (7.3)$$

schreiben es ein bisschen eleganter als

$$P\left[\left|\bar{X} - \mu\right| \leq \frac{1.96\sigma}{\sqrt{n}}\right] = 95\% \quad (7.4)$$

und bekommen schlussendlich

$$P\left[\mu \in \left[\bar{X} - \frac{1.96\sigma}{\sqrt{n}}, \bar{X} + \frac{1.96\sigma}{\sqrt{n}}\right]\right] = 95\%. \quad (7.5)$$

Wir vergleichen voller Stolz und Befriedigung (7.5) mit Definition 7.8.

Also gilt: *ein* 95 % KI für  $\mu$  bei bekanntem  $\sigma^2$  und Daten  $x_1, \dots, x_n$  in Modell  $\mathcal{N}(\mu, \sigma^2)$  ist:

$$\left[ \bar{x} - \frac{1.96\sigma}{\sqrt{n}}, \bar{x} + \frac{1.96\sigma}{\sqrt{n}} \right]. \quad (7.6)$$

Kleine Übung:

99 % KI in dieser Situation ist:

Was geschieht wenn  $n$  wächst?

Was geschieht wenn  $\sigma^2$  wächst?

Offenbar sind beide der obigen Eigenschaften einleuchtend!

Kleine Aufgabe zu KI I: Es wird angenommen, dass die Durchmesser der auf einer bestimmten Anlage hergestellten Stahlkugeln durch die Realisationen einer normalverteilten Zufallsgrösse mit  $\sigma = 1.04$  mm beschrieben werden können. Aus einer Stichprobe vom Umfang  $n = 30$  ergab sich  $\bar{x} = 12.14$  mm. Bestimmen Sie für die Vertrauenswahrscheinlichkeit von 0.95 und 0.99 die Grenzen des KI für den mittleren Durchmesser dieser Kugeln.

### 7.2.4 $\mathcal{N}(\mu, \sigma^2)$ : KI für $\mu$ wenn $\sigma^2$ unbekannt (optimal)

Gegeben Daten  $x_1, \dots, x_n$ , berechnen das arithmetische Mittel  $\bar{x}$ . Wechseln auf die Ebene der Zufallsgrößen: betrachten  $\bar{X}$ . Verteilung von  $\bar{X}$  ist:

Erwartungswert=

Varianz=

Art der Verteilung:

Insgesamt:

Also hat wegen der “Z-Transform” der Ausdruck

$$\frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$$

eine  $\mathcal{N}(0, 1)$ -Verteilung. Ooooops?  $\sigma^2$  ist unbekannt! Keine Panik: Wir werden es einfach schätzen:

$$\hat{\sigma}^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Das Tier

$$\frac{\bar{X} - \mu}{\sqrt{\frac{\hat{\sigma}^2}{n}}} = \frac{\bar{X} - \mu}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{n}}} \quad (7.7)$$

hat eine  $t_{n-1}$ -Verteilung (vgl. mit 4.3.8 - braucht bisschen algebraische Umformung, bis man das erkennt)!



Mit den gleichen Schritten wie in 7.2.3 folgt dann analog zu (7.6):

$$\left[\bar{x} - \frac{CV\hat{\sigma}}{\sqrt{n}}, \bar{x} + \frac{CV\hat{\sigma}}{\sqrt{n}}\right], \quad (7.8)$$

wo

$$\hat{\sigma} := \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}. \quad (7.9)$$

Der kritische Wert (Critical Value CV) ist jetzt nicht mehr nur von  $(1 - \alpha)$  abhängig (wie in 7.2.3), sondern auch von  $n$  (vgl. z.B. Tabelle I in Krengel). Zum Beispiel mit  $n = 20$  sind die kritischen Werte für ein 95%-KI:

zu vergleichen mit den 1.96 von 7.2.3.

Kleine Aufgabe zu KI II: gleiche Ausgangslage wie bei Aufgabe I, aber wir kennen  $\sigma$  nicht und haben es mit (7.9) genau auf 1.12 mm geschätzt. Berechnen Sie das 0.95-KI nochmals in dieser Situation.

Was passiert wenn  $n$  grösser wird (2 Sachen)?

Was wenn  $\sigma^2$  grösser wird?