

4. Punkt-Schätzer

- Punktschätzung
- Konstruktion von Schätzfunktionen
- Maximum-Likelihood-Methode
- Momentenmethode
- Gütekriterien für Punktschätzer
 - Erwartungstreue
 - Effizienz
 - Konsistenz
 - Asymptotische Effizienz
- Zusammenfassung

4 Punktschätzer

Statistik wird häufig eingesetzt, um Informationen über bestimmte Charakteristika einer Grundgesamtheit zu beschaffen.

z.B.: Durchschnittseinkommen aller Bachelorstudenten, durchschnittliche erwartete Verspätung einer Zuglinie, Defektrate im Fertigungsprozess, Basisreproduktionszahl bei einer Epidemie etc.

- ↪ Vollerhebung der Grundgesamtheit i.A. nicht möglich.
- ↪ Idee: Ziehe aus einer „repräsentativen“ Stichprobe Rückschlüsse auf die Grundgesamtheit.
- ↪ Dabei wichtig: homogene Grundgesamtheit muss vorliegen (Partitionierung von Grundgesamtheit bzw. Stichprobe durch Clusterverfahren)

1) Punktschätzverfahren

- ↪ liefern einen einzelnen Schätzwert für einen unbekannten Parameter einer Grundgesamtheit/Verteilung.
- ↪ Stichprobenwerte werden mit Hilfe einer Schätzfunktion aggregiert. Die dabei verwendeten Datenkennzahlen (gemäß DuW) korrespondieren oft zum unbekannten Parameter (z.B. aufgrund der Gesetze großer Zahlen).

2) Intervallschätzverfahren

- ↪ Schätzergebnis ist ein Intervall, in das der unbekannte Parameter der Grundgesamtheit mit einer bestimmten Wahrscheinlichkeit fällt.
- ↪ Oft ist ein „plausibler“ Punktschätzer die Intervallmitte, und die Intervallbreite reflektiert die Unsicherheit über die Genauigkeit des Schätzers.

3) Statistische Tests

- ↪ Überprüfung einer bzgl. der Grundgesamtheit erstellten Hypothese
- ↪ Konstruktion derart, dass eine korrekte Hypothese nur mit einer sehr geringen Wahrscheinlichkeit verworfen wird.

4.1 Punktschätzung

Möglichst genaue Annäherung eines unbekannten Grundgesamtheitsparameters.

Kennwerte einer beliebigen, unbekannten Verteilung

- ↪ Erwartungswert bzw. Varianz einer Zufallsvariablen
- ↪ Median oder allg. Quantil einer Zufallsvariablen
- ↪ Korrelation zweier Zufallsvariablen ...

Spezifische Parameter eines zugrundegelegten Verteilungsmodells

- ↪ z.B. X : Anzahl Schadensmeldungen innerhalb eines Monats, $X \sim \text{Poisson}(\lambda)$, interessierender Parameter λ
- ↪ z.B. X : Durchmesser von produzierten Schrauben, $X \sim \mathcal{N}(\mu, \sigma^2)$, interessierende Parameter μ und σ^2
- ↪ z.B. X : Lebensdauer von Glühbirnen, $X \sim \text{Exp}(\lambda)$, interessierender Parameter λ

Ausgangspunkt:

- ↪ n Stichprobenziehungen/Zufallsexperimente, repräsentiert durch ZV X_1, \dots, X_n
- ↪ X_1, \dots, X_n heißen auch Stichprobenvariablen
- ↪ häufig: Stichprobenvariablen sind unabhängige Wiederholungen von X :
 - Experimente, die den ZVen X_1, \dots, X_n zugrundeliegen, sind unabhängig,
 - jedes Mal wird dasselbe Zufallsexperiment durchgeführt

Für Konsistenzuntersuchung angenommen: u.i.v.-**Folge** X_1, \dots, X_n, \dots

- ↪ auf Basis der Realisierungen x_1, \dots, x_n soll auf θ geschlossen werden.

Eine **Punktschätzung** für θ ist eine Funktion $t = g(x_1, \dots, x_n)$.

- ↪ z.B. $g(x_1, \dots, x_n) = \sum_{i=1}^n \frac{x_i}{n}$ ist (sinnvolle) Punktschätzung für $\theta = E(X)$.

- ↪ die Stichproben sind Realisationen von Zufallsvariablen X_1, \dots, X_n
- ↪ die Punktschätzung ist als Funktion von ZVen auch eine ZVe
- ↪ Variabilität wird durch die Variabilität der ZVen X_1, \dots, X_n bestimmt

Eine **Schätzfunktion** oder **Schätzstatistik** für den Grundgesamtheitsparameter θ ist eine Funktion der Stichprobenvariablen X_1, \dots, X_n :

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

Der **Schätzwert** ergibt sich aus den Realisationen x_1, \dots, x_n :

$$\hat{\theta} = g(x_1, \dots, x_n)$$

Beispiele

$$\bar{X} = g(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n X_i \quad (\text{Schätzfkt. für } E(X))$$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad X_i \in \{0, 1\} \quad (\text{Schätzfkt. für Anteilswert } \pi = P(X = 1))$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (\text{Schätzfkt. für } \sigma^2 = \text{Var}(X))$$

$$\tilde{S}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (\text{Schätzfkt. für } \sigma^2 = \text{Var}(X))$$

Beispiele

X : Verspätung der Rückgabe eines Mathebuches in der Bibliothek in Tagen.

X_i : X bei i -ter Messung.

Seien $(x_1, \dots, x_{10}) = (2, 14, 10, 0, 9, 20, 8, 2, 3, 2)$. $\mu = E(X) = ???$

Annahme: $X \sim F$ mit unbekannter Verteilung $F \Rightarrow X_1, \dots, X_n \stackrel{\text{uiv}}{\sim} F$

Mögliche Schätzungen:

$$\begin{aligned}\hat{\mu}_1 &= \bar{x} = 7 \\ \hat{\mu}_2 &= x_1 = 2 \\ \hat{\mu}_3 &= 3 \cdot x_8 = 6 \\ \hat{\mu}_4 &= \frac{1}{9} \sum_{i=1}^9 x_i = 7.56\end{aligned}$$

Welche ist die "Beste"?

Antwort allgemein abhängig vom verwendeten WS-Modell und Gütekriterium.

Wie findet man geeignete Schätzungen für unbekannte Parameter?

Idee: Suche für Realisationen x_1, \dots, x_n denjenigen Parameter $\hat{\theta} = \hat{\theta}(x_1, \dots, x_n)$, der die plausibelste Erklärung für die Beobachtung dieser Realisationen liefert.

- **ML-Schätzer** $\hat{\theta}_{ML}$: Durch Maximierung der gemeinsamen WS-Dichte der Stichprobenvariablen als Funktion des Parameters.
- **MM-Schätzer** $\hat{\theta}_{MM}$: Durch Gleichsetzung von theoretischen und Stichproben-Kennzahlen und Auflösen dieser Gleichung(en) nach θ .
- **KQ-Schätzer (Regression)**: „bestmögliche Erklärung des Zusammenhangs zwischen abhängiger und unabhängiger Variable durch Minimierung der „Residuenquadratsumme“ der abhängigen Variable
- **MD-Schätzer**: Gegenüberstellung von Stichprobenverteilung und theoretischer Verteilung (z.B. deren Verteilungsfunktion, Quantilfunktion) und Minimierung der „Distanzen“ dazwischen. Hier nicht behandelt.
- **Bayes-Schätzer**: Gewichtetes Mittel aus datenabhängiger Schätzung und „Vorinformation“ des Schätzers (Idee: von der a-priori-Verteilung zur a-posteriori-Verteilung). Hier nicht behandelt.

4.3 Maximum-Likelihood-Methode (ML)

- ↪ Betrachtung der „Plausibilität“ innerhalb einer parametrischen Verteilungsklasse anhand einer konkreten Stichprobe x_1, \dots, x_n .
- ↪ geht v.a. auf Sir Ronald Aylmer Fisher zurück (Anfang des 20. Jahrhunderts)



- ↪ Genetiker, Evolutionstheoretiker, Eugeniker, Statistiker,
* 17.02.1890 in London; + 29.07.1962 in Adelaide
- ↪ wichtige Beiträge zur Statistik: ML-Methode, Varianzanalyse, F-Test, Versuchsplanung, Suffizienz, Verteilungsfreiheit, Fisher-Information, Diskriminanzanalyse

Maximum-Likelihood-Prinzip

Seien X_1, \dots, X_n unabhängige und identische Wiederholungen eines Experimentes.

↪ **Likelihood-Funktion:** Fasse die gemeinsame Dichte als Funktion des unbekannten Parameters θ bei festen Realisationen x_1, \dots, x_n auf:

$$L(\theta) = L(\theta|x_1, \dots, x_n) := f(x_1, \dots, x_n|\theta) = f(x_1|\theta) \cdots f(x_n|\theta)$$

↪ **Maximum-Likelihood-Schätzung**

$\hat{\theta}_{ML} = \hat{\theta}_{ML}(x_1, \dots, x_n)$ ist erklärt als Maximalstelle der Funktion $\theta \mapsto L(\theta)$

↪ **Log-Likelihood:** Statt L wird meist $\ln(L(\theta)) = \sum_{i=1}^n \ln(f(x_i|\theta))$ maximiert.

- **Maximum-Likelihood-Schätzer:** $g(X_1, \dots, X_n) = \hat{\theta}(X_1, \dots, X_n)$
(ist eine Zufallsvariable, die ML-Schätzung ist deren Realisation)
- Durch die Logarithmierung wird das Maximierungsproblem i.A. einfacher.
Das maximierende θ selbst ist identisch, da der Logarithmus eine streng monotone Transformation ist.

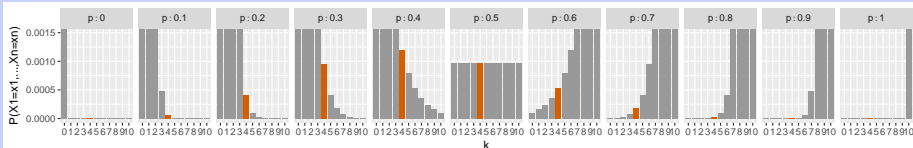
Bsp: Elfmeter (aus *DuW*)

- ↪ Ergebnis werde als Bernoulli-Experiment mit Treffer-WS p angesehen (problematisch: u.i.v-Annahme)
- ↪ Bei $n = 10$ Schüssen folgende Ergebnisse: 1,1,0,0,1,0,0,0,1,0 ($k = 4$ Treffer)
- Ziel: Schätzung von p mit ML-Methode.

Likelihood=Gemeinsame Dichte, hier allgemein (mit $k = x_1 + \dots + x_n$)

$$\begin{aligned} f_p(x_1, \dots, x_n) &= p^{x_1} (1-p)^{1-x_1} \dots p^{x_n} (1-p)^{1-x_n} \\ &= p^{x_1 + \dots + x_n} (1-p)^{n - (x_1 + \dots + x_n)} = p^k (1-p)^{n-k} \end{aligned}$$

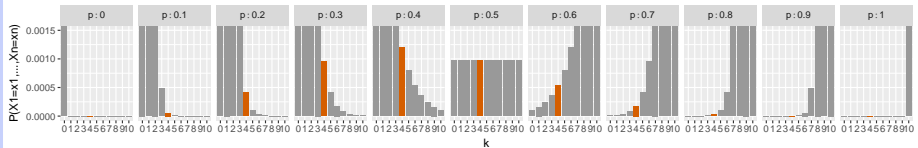
Plot der WS ($n = 10$, diverse p) für $k = 0, \dots, 10$



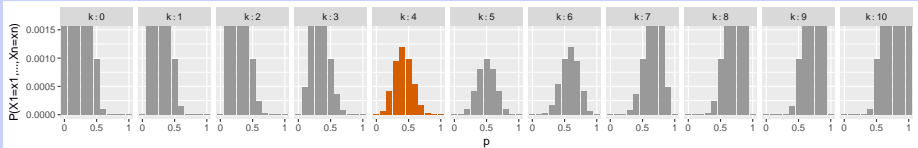
Hervorgehoben: jeweils $k = 4$

Likelihoodfunktion der konkreten Stichprobe ($k = x_1 + \dots + x_{10} = 4$):

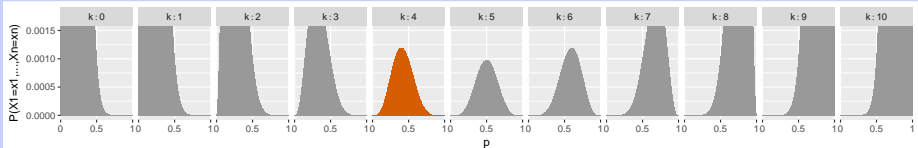
Dichte (s.o.)



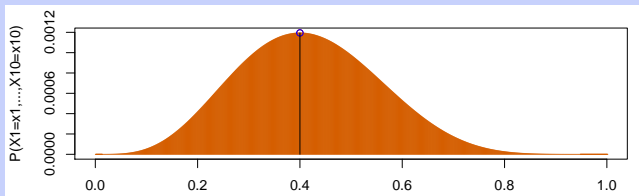
Likelihood: Stelle Werte der Dichte gebündelt nach k dar:



vollständige Graphen der Likelihood (Polynome 10. Grades):



Bsp (Elfmeter): Likelihoodfunktion der konkreten Stichprobe ($k = 4$)



↪ Likelihood mit $k = \sum x_i$ rechnerisch: $L(p) = L(x_1, \dots, x_n, p) = p^k(1-p)^{n-k}$

↪ Log-Likelihood: $\ln(L(p)) = \ln(p^k(1-p)^{n-k}) = k \ln(p) + (n-k) \ln(1-p)$

↪ Maximierung der Log-Likelihood: $\frac{\partial \ln(L(p))}{\partial p} = \frac{k}{p} - \frac{n-k}{1-p}$

□ Notwendig: $0 \stackrel{!}{=} \frac{k}{p} - \frac{n-k}{1-p} \Leftrightarrow \frac{k}{p} = \frac{n-k}{1-p} \Leftrightarrow k - kp = np - kp \Leftrightarrow p = \frac{k}{n}$

□ Hinreichend: Zielfunktion ist konkav: $\frac{\partial^2 \ln(L(p))}{\partial^2 p} = -\frac{k}{p^2} - \frac{n-k}{(1-p)^2} \leq 0$

Maximal ist die (Log-)Likelihood für $p = \frac{k}{n} = \frac{x_1 + \dots + x_n}{n} = \bar{x}$

Der ML-Schätzer ist $g(X_1, \dots, X_n) = \hat{p}_{ML}(X_1, \dots, X_n) = \hat{p}_{ML} = \bar{X}$

Bei Ergebnisfolge 1,1,0,0,1,0,0,0,1,0 ($k = 4$ Treffer) ML-Schätzung $\hat{p}_{ML} = 0.4$

Übung: In einem Fertigungsprozess werden 5x wöchentlich je 4 Items auf Fehler überprüft. Die Defektzahlen X_1, \dots, X_5 seien st.u., $\mathcal{L}(X_i) = \text{Bin}(4, p)$.

In einer bestimmten Woche werden die Defektzahlen 2,0,0,1,0 beobachtet.

Berechnen Sie eine ML-Schätzung für die Defektwahrscheinlichkeit p eines Items.

↪ Log-Likelihood (mit $k = x_1 + \dots + x_5$):

$$\begin{aligned}
 \ln(L(p)) &= \ln(L(x_1, \dots, x_5, p)) \\
 &= \ln \left(\binom{4}{x_1} p^{x_1} (1-p)^{4-x_1} \dots \binom{4}{x_5} p^{x_5} (1-p)^{4-x_5} \right) \\
 &= \ln \left(\binom{4}{x_1} p^{x_1} (1-p)^{4-x_1} \right) + \dots + \ln \left(\binom{4}{x_5} p^{x_5} (1-p)^{4-x_5} \right) \\
 &= \ln \left(\binom{4}{x_1} \right) + \ln(p^{x_1} (1-p)^{4-x_1}) + \dots + \ln \left(\binom{4}{x_5} \right) + \ln(p^{x_5} (1-p)^{4-x_5}) \\
 &= \ln \left(\binom{4}{x_1} \right) + \dots + \ln \left(\binom{4}{x_5} \right) + x_1 \ln(p) + (4-x_1) \ln(1-p) + \dots + x_5 \ln(p) + (4-x_5) \ln(1-p) \\
 &= \ln \left(\binom{4}{x_1} \right) + \dots + \ln \left(\binom{4}{x_5} \right) + (x_1 + \dots + x_5) \ln(p) + (20 - (x_1 + \dots + x_5)) \ln(1-p) \\
 &= \ln \left(\binom{4}{x_1} \right) + \dots + \ln \left(\binom{4}{x_5} \right) + k \ln(p) + (20 - k) \ln(1-p)
 \end{aligned}$$

↪ Maximierung: $\frac{\partial \ln(L(p))}{\partial p} = \frac{k}{p} - \frac{20-k}{1-p}$

□ Notwendig: $0 = \frac{k}{p} - \frac{20-k}{1-p} \Leftrightarrow \frac{k}{p} = \frac{20-k}{1-p} \Leftrightarrow k - kp = 20p - kp \Leftrightarrow p = \frac{k}{20} = \frac{x_1 + \dots + x_5}{20}$

□ Hinreichend: $\ln(L(p))$ ist konkav wegen $\frac{\partial^2 \ln(L(p))}{\partial^2 p} = -\frac{k}{p^2} - \frac{20-k}{(1-p)^2} \leq 0$.

ML-Schätzer ist $\hat{p}_{ML}(X_1, \dots, X_5) = \frac{X_1 + \dots + X_5}{20}$

ML-Schätzung ist $\hat{p}_{ML}(x_1, \dots, x_5) = \frac{2+0+0+1+0}{20} = 0.15$

Beachte: Die Rechnung ist i.w. identisch zum Elfmeter-Beispiel. Die Terme mit den Binomialkoeffizienten werden zu Null differenziert, weil sie nicht von p abhängen.

Memo (DuW) Normalverteilung $\mathcal{N}(\mu, \sigma^2)$ mit $\mu \in \mathbb{R}, \sigma > 0$

↪ Träger: $\mathcal{X} = \mathbb{R}$

↪ Dichte: $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

↪ VF $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$

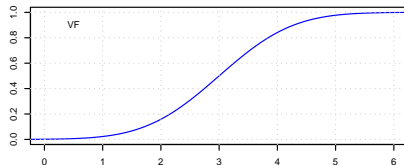
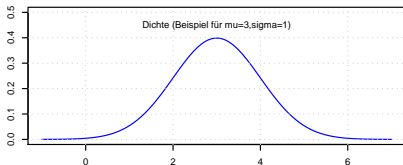
↪ $E(X) = \text{med}(X) = \mu$

↪ $\text{var}(X) = \sigma^2$

↪ R: **dnorm, pnorm, qnorm, rnorm**

↪ Speziell: VF, Dichte zu $\mathcal{N}(0, 1)$:
 $\Phi(x) = \int_{-\infty}^x \phi(t) dt = 1 - \Phi(-x),$
 $\phi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}$

Schreibweise: $u_\alpha := \Phi^{-1}(\alpha)$



- ↪ Als Modellverteilung stetiger Merkmale verwendet (z.B. Regressionsanalyse), aber oft ungerechtfertigt – fehlende Symmetrie, begrenzter Träger der Daten
- ↪ Approximative Stichprobenverteilung standardisierter Summen (Zentraler Grenzwertsatz), für derart kumulierte Daten dann auch Modellverteilung.

Beispiel: Statistico (DuW)

- ↪ Der Spielehersteller „R-Games“ benötigt zum Testen seines Spiels „Statistico“ eine Gruppe von Testspielern, deren IQ demjenigen der Grundgesamtheit entspricht (Annahme: normalverteilt, $\mu = 100$, $\sigma = 15$).
- ↪ Von 20 möglichen Testpersonen liegen die folgenden IQ-Ergebnisse vor:
119, 95, 120, 119, 106, 77, 86, 96, 100, 136, 111, 88, 83, 96, 96, 94, 104, 87, 107, 81
- ↪ Kennzahlen der Stichprobe: $\bar{x} = 100.05$, $\tilde{s}^2 = \frac{1}{n} \sum (x_i - \bar{x})^2 = 221.8475$
- ↪ Kann man bei dieser Stichprobe davon ausgehen, dass sie die Grundgesamtheit ausreichend genau repräsentiert?
- ↪ Unter der Annahme, dass die IQ-Werte über die Grundgesamtheit normalverteilt sind, sind zunächst (ML-)Schätzungen für μ und σ gesucht.

Ziel: ML-Schätzung von $\theta = (\mu, \sigma)$ in u.i.v. Stichprobe $X_1, \dots, X_n \stackrel{\text{u.i.v.}}{\sim} \mathcal{N}(\mu, \sigma^2)$

1.) Likelihood: Schreibe $L(\mu, \sigma)$ für $L(\mu, \sigma, x_1, \dots, x_n)$

$$L(\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_1-\mu)^2}{2\sigma^2}\right) \cdots \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_n-\mu)^2}{2\sigma^2}\right)$$

2.) Log-Likelihood:

$$\ln L(\mu, \sigma) = \sum_{i=1}^n \left[\ln\left(\frac{1}{\sqrt{2\pi}\sigma}\right) - \frac{(x_i-\mu)^2}{2\sigma^2} \right] = \sum_{i=1}^n \left[-\ln(\sqrt{2\pi}) - \ln(\sigma) - \frac{(x_i-\mu)^2}{2\sigma^2} \right]$$

3.) Maximierung (hier nur notwendige Bedingungen)

$$\square \quad \frac{\partial \ln L(\mu, \sigma)}{\partial \mu} = \sum_{i=1}^n \frac{x_i - \mu}{\sigma^2} \stackrel{!}{=} 0 \quad \text{d.h.} \quad \sum_{i=1}^n x_i - n \cdot \mu = 0 \Leftrightarrow \boxed{\hat{\mu} = \bar{x}}$$

$$\square \quad \frac{\partial \ln L(\mu, \sigma)}{\partial \sigma} = \sum_{i=1}^n \left(-\frac{1}{\sigma} + \frac{2(x_i - \mu)^2}{2\sigma^3} \right) \stackrel{!}{=} 0,$$

$$\text{d.h.} \quad -\frac{n}{\sigma} + \sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma^3} = 0 \Leftrightarrow \sum_{i=1}^n (x_i - \mu)^2 = n \cdot \sigma^2,$$

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

ML-Schätzer für (μ, σ) sind die Schätzstatistiken (\bar{X}, \tilde{S}) .

Statistico-Stichprobe: $\hat{\mu} = \bar{x} = 100.05$, $\hat{\sigma} = \tilde{s} \approx 14.89$

↪ ML-Schätzer sind oft nicht (nur) mittels Differentialrechnung zu bestimmen:

Doppelexponentialverteilung

↪ X_1, \dots, X_n u.i.v. mit Dichte (!) $f(x) = \frac{1}{2\sigma} e^{-|x-\mu|/\sigma}$, $\mu \in \mathbb{R}$, $\sigma > 0$.

↪ Verwendung anstelle Normalverteilung (höhere WS extremer Ereignisse, Verteilungen mit „schweren Flanken“, z.B. bei Finanzdaten)

↪ Log-Likelihood: $\ln(L(\mu, \sigma)) = -n \ln(2\sigma) - \frac{1}{\sigma} \sum_{i=1}^n |x_i - \mu|$

□ Maximierung in μ bei festem σ :

L maximal, wenn $\sum_{i=1}^n |x_i - \mu|$ minimal, d.h. (DuW) für $\hat{\mu} = \text{med}(x)$

(dieser Schritt zwangsläufig ohne Differentialrechnung)

□ Maximierung in σ bei $\mu = \hat{\mu}$:

$\ln(L(\hat{\mu}, \sigma)) = -n \ln(2\sigma) - \frac{n \cdot MA(x)}{\sigma}$ mit $MA(x) = \frac{1}{n} \sum_{i=1}^n |x_i - \text{med}(x)|$

Notwendig: $\frac{\partial L(\hat{\mu}, \sigma)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{n \cdot MA(x)}{\sigma^2} \stackrel{!}{=} 0 \Rightarrow \hat{\sigma} = MA(x)$

MA-Schätzer sind also $\hat{\mu}_{ML} = \text{med}(X)$ und $\hat{\sigma}_{ML} = MA(X)$

Übung: Berechnen Sie im vorigen Beispiel mit $f(x) = \frac{1}{2\sigma} e^{-|x-\mu|/\sigma}$ die angegebene Log-Likelihood $\ln(L(\mu, \sigma)) = -n \ln(2\sigma) - \frac{1}{\sigma} \sum_{i=1}^n |x_i - \mu|$

$$\begin{aligned}
 \ln(L(\mu, \sigma)) &= \ln(f(x_1) \cdots f(x_n)) \\
 &= \ln(f(x_1)) + \cdots + \ln(f(x_n)) \\
 &= \ln\left(\frac{1}{2\sigma} e^{-|x_1-\mu|/\sigma}\right) + \cdots + \ln\left(\frac{1}{2\sigma} e^{-|x_n-\mu|/\sigma}\right) \\
 &= \ln\left(\frac{1}{2\sigma}\right) + \ln(e^{-|x_1-\mu|/\sigma}) + \cdots + \ln\left(\frac{1}{2\sigma}\right) + \ln(e^{-|x_n-\mu|/\sigma}) \\
 &= n \ln\left(\frac{1}{2\sigma}\right) + \ln(e^{-|x_1-\mu|/\sigma}) + \cdots + \ln(e^{-|x_n-\mu|/\sigma}) \\
 &= -n \ln(2\sigma) - \frac{|x_1-\mu|}{\sigma} - \cdots - \frac{|x_n-\mu|}{\sigma} \\
 &= -n \ln(2\sigma) - \frac{1}{\sigma} \sum_{i=1}^n |x_i - \mu|
 \end{aligned}$$

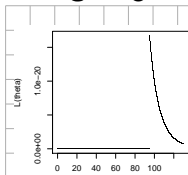
Fritzbrötchen

Bäcker Kalkoves neueste Kreation ist das Fritzbrötchen, welches er in der Bäckerei neben dem Hörsaal verkaufen möchte. Unter 10 Studierenden hat er folgende Preisbereitschaften (PB) erfragt (in EuroCent): 20, 70, 85, 50, 75, 0, 40, 90, 95, 60.

Übung: Die Likelihood-Funktion für (mit $\theta > 0$) u.i.v. $\text{Re}(0, \theta)$ -verteilte PB lautet $L(\theta) = \frac{1}{\theta^n} \mathbb{1}_{[v; \infty[}(\theta)$ mit $v = \max(x_1, \dots, x_n)$

$L(\theta)$	$= \prod_{i=1}^n f(x_i \theta) = \prod_{i=1}^n \frac{\mathbb{1}_{[0; \theta]}(x_i)}{\theta}$	
$= \frac{1}{\theta^n} \prod_{i=1}^n \mathbb{1}_{[0; \theta]}(x_i)$		$\mathbb{1}_{[0; \theta]}(x_i) = 1 \Leftrightarrow \theta \geq x_i \Leftrightarrow \mathbb{1}_{[x_i; \infty[}(\theta)$
$= \frac{1}{\theta^n} \prod_{i=1}^n \mathbb{1}_{[x_i; \infty[}(\theta)$		$\mathbb{1}_A \mathbb{1}_B = \mathbb{1}_{A \cap B}$
$= \frac{1}{\theta^n} \mathbb{1}_{[x_1; \infty[} \cap \dots \cap [x_n; \infty[}(\theta)$		$\theta \in [x_1; \infty[\cap \dots \cap [x_n; \infty[\Leftrightarrow \theta \geq v$
$= \frac{1}{\theta^n} \mathbb{1}_{[v; \infty[}(\theta)$		

Übung: Begründen Sie $\hat{\theta}_{ML} = \max(X_1, \dots, X_n)$ mit einer Skizze von L .



Die Likelihood ist bis v konstant Null, dann hat sie die Form $1/\theta^n$, d.h. ist streng monoton fallend. Der Maximalwert wird daher für $\theta = v$ angenommen, Hier $\hat{\theta} = 95$.

4.4 Momentenmethode (MM)

- ↪ Vergleich der Momente der Stichprobenvariablen X_1, \dots, X_n mit den empirischen Momenten der konkreten Stichprobe x_1, \dots, x_n .
- ↪ geht auf Karl Pearson zurück (Ende des 19. Jahrhunderts)



- ↪ britischer Mathematiker, * 27.03.1857 in London; + 27.04.1936 in Coldharbour
- ↪ wissenschaftliche Schwerpunkte: Mathematik, Physik, Deutsche Literatur, Recht, Theologie und Geschichte, Statistik, Psychologie
- ↪ Gründung der weltweit ersten Statistik-Fakultät 1911, University College of London
- ↪ wichtige Beiträge: Korrelationskoeffizient, Momentenmethode, statistische Testtheorie, ...

Modellannahmen: Wir wollen $\theta = (\theta_1, \dots, \theta_m)$ schätzen. Betrachte hierzu eine Stichprobe x_1, \dots, x_n zu den Stichprobenvariablen $X_1, \dots, X_n \stackrel{u.i.v.}{\sim} F$.

Vorgehen (Kochrezept)

- 1.) Bestimme k -te empirische Momente $\hat{m}_k(x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n x_i^k \quad \forall k$
- 2.) Gleichsetzen empirischer und theoretischer Momente:

$$\hat{m}_k(x_1, \dots, x_n) = g_k(\theta) = E(X_1^k) \quad \forall k \quad (\star)$$

Beachte: $E(X_1^k)$ hängt von $\theta_1, \dots, \theta_m$ ab.

- 3.) Bestimmung der Lösung $\hat{\theta}(x_1, \dots, x_n)$ des Gleichungssystems.
(vorausgesetzt werden die eindeutige Lösbarkeit sowie die „Messbarkeit“ von $\hat{\theta} : \mathbb{R}^n \rightarrow \Theta \subset \mathbb{R}^m$ als Stichprobenfunktion.

Momentenschätzer

Der durch (\star) gegebene Zufallsvektor $\hat{\theta}_{MM}(X_1, \dots, X_n)$ heißt **Momentenschätzer** des Parametervektors θ .

Beispiel: Normalverteilte Stichprobenvariablen

$X_1, \dots, X_n \stackrel{i.i.d.}{\sim} \mathcal{N}(\mu, \sigma^2)$, μ und σ^2 unbekannt

$\theta = (\theta_1, \theta_2) = (\mu, \sigma^2)$, $\Theta = \mathbb{R} \times (0, \infty)$

1.) Bestimmung der 1. und 2. empirischen Momente der Stichprobe:

$$\hat{m}_1 = \frac{1}{n} \sum_{i=1}^n x_i, \quad \hat{m}_2 = \frac{1}{n} \sum_{i=1}^n x_i^2$$

2.) Gleichsetzen der empirischen und theoretischen Momente:

$$g_1(\mu, \sigma^2) = E(X_1) = \mu = \frac{1}{n} \sum_{i=1}^n x_i,$$

$$g_2(\mu, \sigma^2) = E(X_1^2) = \sigma^2 + \mu^2 = \frac{1}{n} \sum_{i=1}^n x_i^2$$

Beispiel: Normalverteilte Stichprobenvariablen

3.) Auflösen des Gleichungssystems:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \hat{\mu}^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \left(\frac{1}{n} \sum_{i=1}^n x_i \right)^2$$

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

Beispiel: Normalverteilte Stichprobenvariablen

4.) Konstruktion der Momentenschätzer

$$\hat{\mu}_{MM}(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n X_i$$

$$(\hat{\sigma}^2)_{MM}(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

$$\hookrightarrow (\hat{\sigma}^2)_{MM} \stackrel{\wedge}{=} \tilde{S}^2$$

Beispiel: Binomialverteilte Stichprobenvariablen

$X_1, \dots, X_n \stackrel{u.i.v.}{\sim} \text{Bin}(k, p), \theta = (\theta_1, \theta_2) = (k, p), \Theta = \mathbb{N} \times (0, 1)$

1.) Bestimmung des 1./2. emp. Moments: $\hat{m}_1 = \frac{1}{n} \sum_{i=1}^n x_i, \quad \hat{m}_2 = \frac{1}{n} \sum_{i=1}^n x_i^2$

2.) Gleichsetzen der empirischen und theoretischen Momente:

$$\square g_1(k, p) = E(X_1) = k \cdot p \stackrel{!}{=} \frac{1}{n} \sum_{i=1}^n x_i$$

$$\square g_2(k, p) = E(X_1^2) = \text{var}(X_1) + E(X_1)^2 = k \cdot p \cdot (1-p) + (k \cdot p)^2 \stackrel{!}{=} \frac{1}{n} \sum_{i=1}^n x_i^2$$

3.) Auflösen des Gleichungssystems:

$$\square \hat{k} = \frac{1}{n \cdot \hat{p}} \sum_{i=1}^n x_i = \frac{1}{\hat{p}} \bar{x}$$

$$\square \frac{1}{n} \sum_{i=1}^n x_i^2 = kp(1-p) + (kp)^2$$

$$\Leftrightarrow \frac{1}{n} \sum_{i=1}^n x_i^2 = \bar{x}(1-p) + \bar{x}^2$$

$$\Leftrightarrow \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 = \bar{x}(1-p)$$

$$\Leftrightarrow \hat{p} = 1 - \frac{\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2}{\bar{x}} = \frac{\bar{x} - \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}{\bar{x}}$$

Beispiel: Binomialverteilte Stichprobenvariablen

4.) Konstruktion der Momentenschätzer

$$\hat{k}(X_1, \dots, X_n) = \frac{\bar{X}^2}{\bar{X} - \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2}$$

$$\hat{p}(X_1, \dots, X_n) = \frac{\bar{X}}{\hat{k}(X_1, \dots, X_n)}$$

für $(X_1, \dots, X_n) \neq (0, \dots, 0)$.

Fritzbrötchen

Bäcker Kalkoves neueste Kreation ist das Fritzbrötchen, welches er in der Bäckerei neben dem Hörsaal verkaufen möchte. Unter 10 Studierenden hat er folgende Preisbereitschaften (PB) erfragt (in EuroCent): 20, 70, 85, 50, 75, 0, 40, 90, 95, 60.

Übung: Berechnen Sie $E(X_i)$, $\text{var}(X_i)$ für (mit $\theta > 0$) u.i.v. $\text{Re}(0, \theta)$ -verteilte PB

$$E(X_i) = \int_0^\theta x \cdot \frac{1}{\theta} dx = \left[\frac{1}{2\theta} x^2 \right]_0^\theta = \frac{1}{2\theta} \theta^2 = \frac{\theta}{2}$$

$$\text{var}(X_i) = \int_0^\theta x^2 / \theta dx - \frac{\theta^2}{4} = \theta^2 / 12$$

Übung: Berechnen Sie einen MM-Schätzer für θ und vergleichen Sie mit $\hat{\theta}_{ML}$.

Ansatz $E(X_i) \stackrel{!}{=} \bar{x} \Leftrightarrow \frac{\theta}{2} = \bar{x} \Leftrightarrow \theta = 2\bar{x}$

Der MM-Schätzer ist $\hat{\theta}_{MM} = 2\bar{X}$. Die MM-Schätzung ist $\hat{\theta} = 2 \cdot 56.5 = 113$.

MM- und ML-Schätzer sind verschieden, letzterer wählt bekanntlich die maximale Beobachtung.

4.5 Gütekriterien für Punktschätzer

Nachfolgend betrachten wir folgende Problemstellungen:

Fragestellung 1

Gegeben: Punktschätzer T für μ_X

Frage: Wie „gut“ ist T ?

Beispiel: Gegeben Körpergrößen X_1, \dots, X_n in cm. Wir schätzen die mittlere Körpergröße konstant durch $T = 150$. Gut oder schlecht?

Fragestellung 2

Gegeben: Mehrere Punktschätzer T_1, \dots, T_k für μ_X

Frage: Welcher Schätzer ist „der beste“? Bzw. wie kann man überhaupt Schätzer vergleichen?

Beispiel: Betrachte erneut Körpergrößen X_1, \dots, X_n und die zwei Schätzer $T_1 = 150$ und $T_2 = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$. Welcher ist besser?

Zu Fragestellung 1

Betrachten wir erneut Körpergrößen X_1, \dots, X_n . Die Schätzung des Erwartungswertes μ_X mit Hilfe des Schätzers

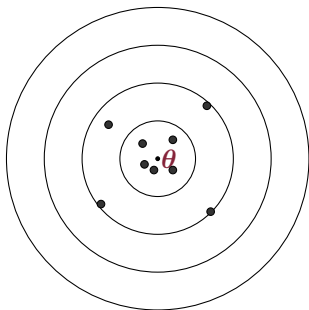
$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

„macht intuitiv Sinn“!

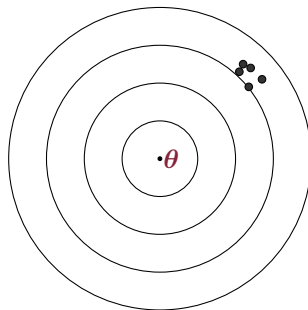
Wir werden sehen, dass \bar{X} tatsächlich ein „guter“ Schätzer (für μ_X) ist. Aber was heißt denn nun überhaupt gut?

Eine Analogie: Darts

Wir stellen uns eine Stichprobe eines Schätzers T für den Parameter θ als Pfeilwurf auf eine Dartscheibe mit **Bullseye** θ vor:



Im Mittel richtig



Systematisch daneben

Zur Erinnerung

Eine **Schätzfunktion** oder **Schätzstatistik** für einen Parameter θ ist eine **Funktion der Stichprobenvariablen** X_1, \dots, X_n :

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

D. h. der Schätzer ist selbst eine Zufallsvariable!

D. h. wiederum, dass wir etwa Erwartungswert und Varianz von T betrachten können!

Erwartungstreue

Eine Statistik $T = g(X_1, \dots, X_n)$ heißt **erwartungstreu** oder unverzerrt, wenn gilt

$$E_{\theta}(T) = \theta$$

Erwartungstreue des Stichprobenmittels einer u.i.v.-Stichprobe

\bar{X} ist erwartungstreue Schätzstatistik für den Erwartungswert $\mu = E(X)$:

$$E_{\mu}(\bar{X}) = \frac{1}{n} \sum_{i=1}^n E_{\mu}(X_i) = \frac{1}{n} \cdot n \cdot \mu = \mu$$

Beispiel Buchrückgabe: $\mu_1 = \bar{X}$

$$E(\hat{\mu}_1) = \frac{1}{10} \sum_{i=1}^n E(X_i) = \frac{1}{10} \cdot 10 \cdot \mu = \mu \quad \text{erwartungstreu}$$

Übung: Sind $\hat{\mu}_2 = X_1$, $\hat{\mu}_3 = 3 \cdot X_8$ bzw. $\hat{\mu}_4 = \frac{1}{9} \sum_{i=1}^9 X_i$ erwartungstreu für μ ?

$$E(\hat{\mu}_2) = E(X_1) = \mu \quad \text{erwartungstreu}$$

$$E(\hat{\mu}_3) = E(3 \cdot X_8) = 3 \cdot \mu \neq \mu \quad \text{nicht erwartungstreu}$$

$$E(\hat{\mu}_4) = \frac{1}{9} \cdot \sum_{i=1}^9 E(X_i) = \frac{1}{9} \cdot 9 \cdot \mu = \mu \quad \text{erwartungstreu}$$

→ "Erwartungstreue" als einziges Kriterium ist nicht ausreichend - und auch nicht immer sinnvoll.

Weitere Beispiele (u.i.v.-Wiederholungen)

$$E_p(\bar{X}) = p \quad \text{mit } X_i \in \{0, 1\}, p = P(X_i = 1) \quad \text{erwartungstreu}$$

$$E_{\sigma^2}(S^2) = E_{\sigma^2} \left(\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \right) = \sigma^2 \quad \text{erwartungstreu}$$

$$\begin{aligned} E_{\sigma^2}(\tilde{S}^2) &= E_{\sigma^2} \left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right) \\ &= \frac{n-1}{n} \cdot E_{\sigma^2}(S^2) = \frac{n-1}{n} \cdot \sigma^2 \quad \text{nicht erwartungstreu} \end{aligned}$$

(dabei $\sigma^2 = \text{var}(X_i)$)

Beachten Sie: Wie in diesen Beispielen entstehen Schätzfunktionen oft aus deskriptiven Lage- und Streuungskennzahlen, denen eine Zufallstichprobe mit geeigneten Zufallsvariablen (d.h. WS-Modell) zugrundeliegt.

Übung: Zeigen Sie in verschiedenen Schritten $E_{\sigma^2}(S^2) = \sigma^2$

$$E\left(\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2\right) = \frac{1}{n-1} \sum_{i=1}^n E(X_i - \bar{X})^2 = \frac{1}{n-1} \sum_{i=1}^n (E(X_i^2) - 2E(X_i \bar{X}) + E(\bar{X}^2))$$

1. Zeigen Sie zunächst: $E(X_i^2) = \sigma^2 + \mu^2$ (mit $\mu = E(X_i)$)

$$\text{var}(X_i) = E(X_i^2) - (E(X_i))^2 \Rightarrow E(X_i^2) = \text{var}(X_i) + (E(X_i))^2 = \sigma^2 + \mu^2$$

2. Stellen Sie entsprechend den zweiten Erwartungswert dar

$$\begin{aligned} \square E(X_i \bar{X}) &= E\left(X_i \frac{1}{n} \sum_{k=1}^n X_k\right) = E\left(\frac{1}{n} \sum_{k=1}^n X_i X_k\right) = \frac{1}{n} \sum_{k=1}^n E(X_i X_k) \\ \square \text{Für } i \neq k \text{ sind } X_i, X_k \text{ st.u. und damit } E(X_i X_k) &= E(X_i)E(X_k) = \mu^2 \\ \square \text{Für } i = k \text{ ist } E(X_i X_k) &= \sigma^2 + \mu^2 \text{ (s.o.)} \\ \square E(X_i \bar{X}) &= \dots = \frac{1}{n} (E(X_i^2) + \sum_{k \neq i} E(X_i X_k)) = \frac{1}{n} (\sigma^2 + \mu^2 + (n-1)\mu^2) = \frac{\sigma^2 + n\mu^2}{n} \end{aligned}$$

3. Verfahren Sie ebenso mit dem dritten Erwartungswert.

$$E(\bar{X}^2) = \frac{1}{n} \sum_{i=1}^n E(X_i \bar{X}) \stackrel{\text{s.o.}}{=} \frac{\sigma^2 + n\mu^2}{n}$$

4. Zeigen Sie $E_{\sigma^2}(S^2) = \sigma^2$ mit den Ergebnissen aus 1. bis 3.

$$E(S^2) = \frac{1}{n-1} \sum_{i=1}^n (\sigma^2 + \mu^2 - 2 \frac{\sigma^2 + n\mu^2}{n} + \frac{\sigma^2 + n\mu^2}{n}) = \frac{1}{n-1} \sum_{i=1}^n \frac{n-1}{n} \sigma^2 = \sigma^2$$

Beispiel Buchrückgabe $((x_1, \dots, x_{10}) = (2, 14, 10, 0, 9, 20, 8, 2, 3, 2))$

X_i : Verspätung der Buchabgabe bei i -ter Messung ($i = 1, \dots, 10$),

$X_i \sim F$ (unbekannt)

Erwartungstreue Schätzung der Varianz σ^2 durch $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$.

$$\begin{aligned} S^2 &= \frac{1}{9} [(2-7)^2 + (14-7)^2 + (10-7)^2 + (0-7)^2 + (9-7)^2 \\ &\quad + (20-7)^2 + (8-7)^2 + (2-7)^2 + (3-7)^2 + (2-7)^2] \\ &= \frac{1}{9} (25 + 49 + 9 + 49 + 4 + 169 + 1 + 25 + 16 + 25) \\ &= \frac{1}{9} \cdot 372 = 41.\bar{3} \end{aligned}$$

```
> x = c(2, 14, 10, 0, 9, 20, 8, 2, 3, 2)
```

```
> var(x)
```

```
[1] 41.33333
```

Bias / Verzerrung

Eine Statistik $T = g(X_1, \dots, X_n)$ heißt **erwartungstreu**, wenn gilt $E_\theta(T) = \theta$.

Der *systematische* Fehler einer Schätzstatistik heißt **Verzerrung** oder **Bias**:

$$\text{Bias}_\theta(T) = E_\theta(T) - \theta$$

Verzerrung eines nicht erwartungstreuen Schätzers

$$E_{\sigma^2}(\tilde{S}^2) = E_{\sigma^2} \left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right) = \frac{n-1}{n} \cdot \sigma^2 \quad \text{nicht erwartungstreu}$$

$$\leadsto \text{Bias}_{\sigma^2}(\tilde{S}^2) = E_{\sigma^2}(\tilde{S}^2) - \sigma^2 = \left(\frac{n}{n} \cdot \sigma^2 - \frac{1}{n} \cdot \sigma^2 \right) - \sigma^2 = -\frac{1}{n} \cdot \sigma^2$$

Beispiel Buchrückgabe: Annahme: $E(X_i) = \mu$

$$E(\hat{\mu}_3) = E(3 \cdot X_8) = 3 \cdot \mu \neq \mu \quad \text{nicht erwartungstreu}$$

Übung: Bestimmen Sie den Bias des Schätzers $\hat{\mu}_3$.

$$\text{Bias}_\mu(\hat{\mu}_3) = E(3 \cdot X_8) - \mu = 3 \cdot \mu - \mu = 2 \cdot \mu$$

Asymptotische Erwartungstreue

Eine Schätzstatistik heißt **asymptotisch erwartungstreu** für θ , wenn gilt:

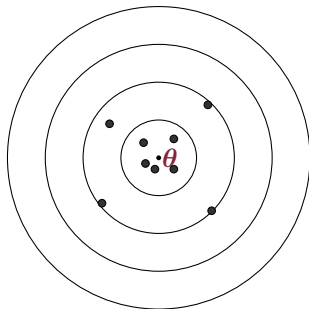
$$\lim_{n \rightarrow \infty} E_{\theta}(T) = \theta$$

$\hookrightarrow \tilde{S}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ ist asymptotisch erwartungstreu, da

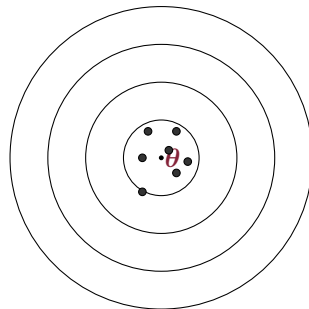
$$\lim_{n \rightarrow \infty} E_{\sigma^2}(\tilde{S}^2) = \lim_{n \rightarrow \infty} \frac{n-1}{n} \cdot \sigma^2 = \sigma^2$$

\hookrightarrow aber: für kleines n kann der Bias erheblich sein.

Eine Analogie: Darts



Im Mittel richtig, aber ungenau



im Mittel richtig und ziemlich genau

Standardfehler und Varianz einer Schätzstatistik

Die **Schätzervarianz** einer Schätzstatistik $T = g(X_1, \dots, X_n)$ ist

$$\text{var}(T) = E([T - E(T)]^2)$$

Der **Standardfehler** einer (meist erwartungstreuen) Schätzstatistik ist

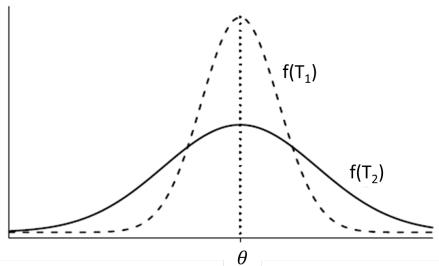
$$\sigma_T = \sqrt{\text{var}(T)}.$$

Beispiel: Standardfehler des Gruppenmittels

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \sqrt{\frac{\text{Var}(X)}{n}}.$$

- ↪ Gütekriterium für Schätzer: erwartungstreu mit minimaler Schätzervarianz (nur in Ausnahmefällen realisierbar).
- ↪ Der (theoretische) Standardfehler ist unbekannt. Für ihn werden in den Anwendungen Schätzungen benötigt, welche selber als (empirische) Standardfehler bezeichnet werden (z.B. für Intervallschätzer benötigt).

Effiziente Schätzstatistiken



Die Verteilung eines „guten“ Schätzers sollte:

- ☐ keine oder nur geringe systematische Abweichung nach oben und unten zum unbekannten Parameter aufweisen
- ☐ eine geringe Streuung besitzen.

Beide Kriterien lassen sich im MSE vereinen.

Fehler eines Schätzers (MSE)

Die **erwartete mittlere quadratische Abweichung** (MSE: *mean squared error*) bestimmt sich durch

$$\begin{aligned}
 \text{MSE}(T) &= E\left([T - \theta]^2\right) = E\left([T - E(T) + E(T) - \theta]^2\right) \\
 &= E\left([T - E(T)]^2 + 2 \cdot [T - E(T)] \cdot [E(T) - \theta] + [E(T) - \theta]^2\right) \\
 &= E\left([T - E(T)]^2\right) + 2 \cdot E\left([T - E(T)] \cdot [E(T) - \theta]\right) \\
 &\quad + E\left([E(T) - \theta]^2\right) \quad \text{beachte: } [E(T) - \theta] \text{ ist konstant} \\
 &= E\left([T - E(T)]^2\right) + \underbrace{[E(T) - \theta] \cdot 2 \cdot E\left([T - E(T)]\right)}_{=E(T)-E(T)=0} + [E(T) - \theta]^2 \\
 &= E\left([T - E(T)]^2\right) + [E(T) - \theta]^2 = \text{Var}(T) + [\text{Bias}(T)]^2
 \end{aligned}$$

- ↪ führt Schätzervarianz und Bias in einem Gütekriterium zusammen.
- ↪ meist nur in „linearen“ Ausnahmefällen (Existenz erwartungstreuer Schätzer) verwendbar, wird asymptotisch (Stichprobenumfang $\rightarrow \infty$) nützlich.

Eigenschaften erwartungstreuer Schätzer

Für einen zum Parameter θ erwartungstreuen Schätzer T gilt:

a) $E_{\theta}(T) = \theta$

b) $Bias_{\theta}(T) = E_{\theta}(T) - \theta \stackrel{a)}{=} 0$

c) $MSE_{\theta}(T) = Var_{\theta}(T) + [Bias_{\theta}(T)]^2 \stackrel{b)}{=} Var_{\theta}(T).$

Schätzervergleich durch MSE

Im Vergleich zweier Schätzstatistiken T_1 und T_2 heißt T_1 **MSE-effizienter**, wenn für alle zugelassenen Verteilungen gilt:

$$MSE(T_1) \leq MSE(T_2)$$

Eine Schätzstatistik heißt **MSE-effizient**, wenn ihr MSE den kleinsten möglichen Wert für alle zugelassenen Schätzer annimmt.

- ↪ Für erwartungstreue Statistiken reduziert sich der Vergleich auf die entsprechenden Varianzen der Schätzstatistiken.
- ↪ ansonsten alleiniger Varianzvergleich sinnlos (z.B. konstante Schätzstatistiken)

Beispiel: 200-facher Wurf einer verbogenen Münze

Seien X_1, \dots, X_{200} u.i.v. mit $X_i \sim \mathcal{B}(1, p)$ ($1 \equiv Z$, $0 \equiv W$)

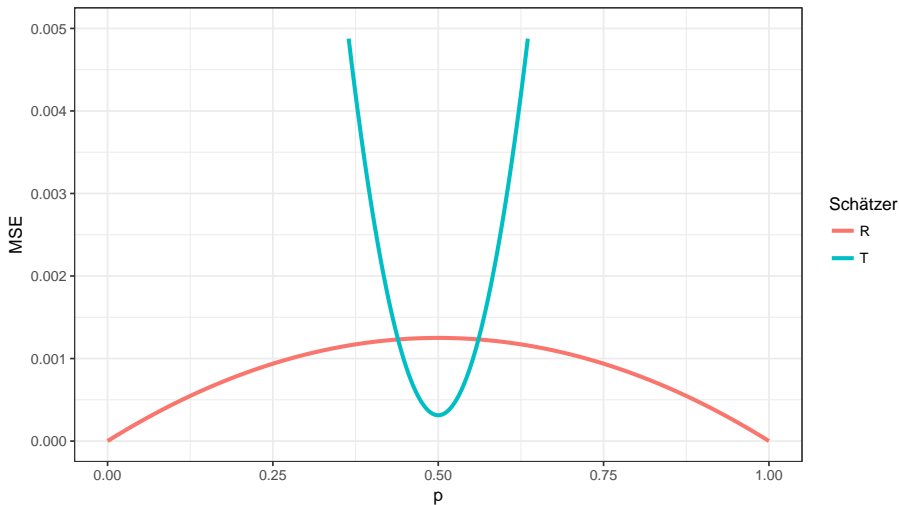
Mögliche Schätzer für p sind bspw.:

$$\square T = \bar{X}_{200} \text{ (hier: } t = 0.45\text{)} \quad (\text{e-treu})$$

$$MSE(T) = \text{var}(T) = \frac{p \cdot (1-p)}{200} \quad (\text{minimal unter e-treuen})$$

$$\square R = \frac{1}{2}T + \frac{1}{4} \text{ (hier: } r = 0.475\text{)} \quad (\text{nicht e-treu})$$

$$\begin{aligned} MSE(R) &= \text{var}(R) + [E(R) - p]^2 \\ &= \text{var}\left(\frac{1}{2} \cdot T + \frac{1}{4}\right) + \left[E\left(\frac{1}{2} \cdot T + \frac{1}{4}\right) - p\right]^2 \\ &= \frac{1}{4} \cdot \text{var}(T) + \left(\frac{1}{4} - \frac{p}{2}\right)^2 \\ &= \frac{1}{4} \cdot \frac{p \cdot (1-p)}{200} + \frac{1}{4} \cdot \left(p - \frac{1}{2}\right)^2 \end{aligned}$$



→ Beispiel für die Unvergleichbarkeit von Schätzern

Übung: Vergleichen Sie die drei erwartungstreuen Schätzer $\hat{\mu}_1 = \frac{1}{10} \sum_{i=1}^{10} X_i$, $\hat{\mu}_2 = X_1$ und $\hat{\mu}_4 = \frac{1}{9} \sum_{i=1}^9 X_i$ im Buch-Beispiel hinsichtlich der MSE-Effizienz.

$$MSE(\hat{\mu}_1) = Var(\hat{\mu}_1) = Var\left(\frac{1}{10} \sum_{i=1}^{10} X_i\right)$$

$$\stackrel{u.i.v.}{=} \frac{1}{100} \sum_{i=1}^{10} Var(X_i) = \frac{1}{100} \cdot 10 \cdot \sigma^2 = \frac{1}{10} \cdot \sigma^2$$

$$MSE(\hat{\mu}_2) = Var(\hat{\mu}_2) = Var(X_1) = \sigma^2$$

$$MSE(\hat{\mu}_4) = Var(\hat{\mu}_4) = Var\left(\frac{1}{9} \sum_{i=1}^9 X_i\right) \stackrel{u.i.v.}{=} \frac{1}{81} \cdot 9 \cdot \sigma^2 = \frac{1}{9} \cdot \sigma^2$$

\Rightarrow Für $\sigma^2 \neq 0$: $\hat{\mu}_1$ ist effizienter als $\hat{\mu}_4$ und $\hat{\mu}_4$ wiederum effizienter als $\hat{\mu}_2$.

Insbesondere ist \bar{X} unter allen erwartungstreuen Schätzern $\hat{\mu}$ für μ (d.h. $E(\hat{\mu}) = \mu$) MSE-effizienter Schätzer, denn es gilt $Var(\bar{X}) \leq Var(\hat{\mu})$.

Verallgemeinerung des MSE: Verlustfunktion, Risikofunktion

↪ Es bezeichne:

- Θ die Menge aller möglichen Parameter,
- \mathcal{T} den Wertebereich der (zur Verfügung stehenden) Schätzer.

↪ Eine **Verlustfunktion** ist eine Funktion $L : \Theta \times \mathcal{T} \rightarrow [0; \infty]$.

↪ Das **Risiko** einer Schätzfunktion $T = T(X)$ bei vorliegendem Parameter θ ist

$$R(\theta, T) = E_{\theta}(L(\theta, T(X))).$$

Die **Risikofunktion** von $T = T(X)$ ist die Funktion $R : \theta \mapsto R(\theta, T(X))$.

Beispiele von Verlustfunktionen

↪ $L(\theta, t) = (t - \theta)^2$ (= MSE)

↪ $L(\theta, t) = |\theta - t|^r$ mit $r > 0$

↪ Auch Fehlerwahrscheinlichkeiten 1./2. Art bei statistischen Tests lassen sich als Risiken darstellen (0/1-wertige Verlustfunktion)

- ↪ Konsistenz eines Schätzers bedeutet, dass dieser mit wachsendem Stichprobenumfang dem unbekannten Parameter immer näher kommt.
- ↪ Entsprechend den verschiedenen Konvergenzarten (s.o. Gesetze großer Zahlen) gibt es verschiedene Konzepte für Konsistenz:
 - ↪ Konsistenz im quadratischen Mittel
 - ↪ schwache Konsistenz
 - ↪ starke Konsistenz

Konsistenz im quadratischen Mittel (für einen Schätzer)

Eine Schätzstatistik T_n heißt **konsistent im quadratischen Mittel**, wenn gilt

$$MSE(T_n) = \text{Var}(T_n) + [\text{Bias}(T_n)]^2 \xrightarrow{n \rightarrow \infty} 0.$$

- ↪ sowohl Verzerrung als auch Varianz verschwinden für wachsende Stichprobenumfänge
- ↪ asymptotische Eigenschaft, kann für endliches n erhebliche Bias- und Varianzanteile aufweisen

Schwache Konsistenz eines Schätzers

Die Schätzstatistik $T_n = g(X_1, \dots, X_n)$ heißt **schwach konsistent**, wenn zu beliebigem $\varepsilon > 0$ gilt:

$$\lim_{n \rightarrow \infty} P(|T_n - \theta| < \varepsilon) = 1 \quad \text{bzw.}$$

$$\lim_{n \rightarrow \infty} P(|T_n - \theta| \geq \varepsilon) = 0$$

D.h. für wachsendes n konvergiert die Wahrscheinlichkeit, mit der die Schätzstatistik T_n höchstens um ε vom wahren Wert θ abweicht, gegen Null (*Konvergenz in Wahrscheinlichkeit*).

Schwache Konsistenz ist oft eine Folge des schwachen Gesetzes großer Zahlen bzw. der Tschebyscheff-Ungleichung (s.o.)

T_n ist konsistent im quadratischen Mittel $\Rightarrow T_n$ ist schwach konsistent.

Beweis: mit der Markoff-Ungleichung (mit $g(x) = x^2$) gilt für $\varepsilon > 0$

$$P(|T_n - \theta| \geq \varepsilon) \leq \frac{E([T_n - \theta]^2)}{\varepsilon^2} = \frac{1}{\varepsilon^2} \cdot \text{MSE}(T_n) \xrightarrow{n \rightarrow \infty} 0$$

Folglich ist auch eine Schätzstatistik schwach konsistent, wenn sie erwartungstreu ist und $\text{Var}(T_n) \rightarrow 0$ für $n \rightarrow \infty$ gilt.

Beweis: Wegen obiger Abschätzung (inkl. Markoff-Ungleichung) und $E(T_n) = \theta$ gilt

$$P(|T_n - \theta| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \cdot \underbrace{\text{MSE}(T_n)}_{= \text{Var}(T_n) + [E(T_n) - \theta]^2} = \frac{1}{\varepsilon^2} \cdot \text{Var}(T_n) \xrightarrow{n \rightarrow \infty} 0.$$

Beispiel

Sei $X \sim \mathcal{N}(\mu, \sigma^2)$. Für unabhängige Wiederholungen X_1, \dots, X_n erhält man

$$E(\bar{X}) = \mu, \text{ Var}(\bar{X}) = \frac{\sigma^2}{n}, \text{ d.h. } \bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

↪ \bar{X} ist erwartungstreu, abnehmende Varianz für wachsendes n
⇒ konsistent im quadratischen Mittel

↪ Überprüfung der schwachen Konsistenz:

$$\begin{aligned} P(|\bar{X} - \mu| \leq \varepsilon) &= P\left(\left|\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right| \leq \frac{\varepsilon}{\sigma/\sqrt{n}}\right) \\ &= P\left(-\frac{\varepsilon}{\sigma} \cdot \sqrt{n} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{\varepsilon}{\sigma} \cdot \sqrt{n}\right) \\ &= \Phi\left(\frac{\varepsilon}{\sigma} \cdot \sqrt{n}\right) - \Phi\left(-\frac{\varepsilon}{\sigma} \cdot \sqrt{n}\right) = 2 \cdot \Phi\left(\frac{\varepsilon}{\sigma} \cdot \sqrt{n}\right) - 1 \xrightarrow{n \rightarrow \infty} 1 \end{aligned}$$

⇒ \bar{X} ist schwach konsistent.

Übung: Ist folgende Aussage richtig? Eine asymptotisch erwartungstreue Schätzstatistik T_n ist schwach konsistent, wenn $\text{var}(T_n) \rightarrow 0$ für $n \rightarrow \infty$ gilt.

Mit der Tschebyscheff-Ungleichung gilt

$$\begin{aligned} P(|T_n - \theta| \geq \varepsilon) &\leq \frac{1}{\varepsilon^2} \cdot E((T_n - \theta)^2) \\ &= \frac{1}{\varepsilon^2} \cdot \text{MSE}(T_n) \\ &= \frac{1}{\varepsilon^2} \cdot (\text{Var}(T_n) + [E(T_n) - \theta]^2) \\ &\xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

Denn $\text{var}(T_n) \rightarrow 0$ nach Voraussetzung und $E(T_n) \rightarrow \theta$, deshalb auch $(E(T_n) - \theta)^2 \rightarrow 0$.
 T_n ist also schwach konsistent, die Aussage ist wahr.

Starke Konsistenz eines Schätzers

Die Schätzstatistik $T_n = g(X_1, \dots, X_n)$ heißt **stark konsistent**, falls:

$$P_\theta \left(\lim_{n \rightarrow \infty} |T_n - \theta| = 0 \right) = 1 \quad \forall \theta$$

\hookrightarrow *fast sichere Konvergenz* ($T_n \xrightarrow{f.s.} \theta$)

$\hookrightarrow T$ ist stark konsistent $\Rightarrow T$ ist schwach konsistent.

\hookrightarrow Unter bestimmten Zusatzannahmen gilt auch:
 T ist stark konsistent $\Rightarrow T$ ist konsistent im quadratischen Mittel.

\hookrightarrow Die starke Konsistenz eines Schätzers ergibt sich fast immer aus dem starken Gesetz großer Zahlen.

- Viele Schätzer T_n lassen sich bei festem n nicht vergleichen bzw. es existieren keine optimalen Schätzer bei festem n .
- Um die Schätzer zu „selektieren“, wird oft die Konsistenz von $T_n(X)$ für θ in einem geeigneten Sinne gefordert. Das reicht aber oft auch noch nicht aus.
- Zusätzlich zur Konsistenz wird dann die asymptotische Normalität gefordert:

$$\lim_{n \rightarrow \infty} P\left(\sqrt{n} \frac{T_n - \theta}{\sigma(\theta)} \leq t\right) = \Phi(t)$$

Diese ist oft aufgrund des zentralen Grenzwertsatzes erfüllt.

- $\sigma^2(\theta)$ wird als **asymptotische Schätzervarianz** bezeichnet. Je kleiner $\sigma^2(\theta)$, desto (asymptotisch) besser ist der Schätzer. Ein Schätzer mit minimalem $\sigma^2(\theta)$ heißt **asymptotisch effizient**.
- Untergrenze für $\sigma^2(\theta)$ ist oft die Cramer-Rao-Schranke $1/I(\theta)$ mit der Fisher-Information $I(\theta) = \text{var}_\theta(S_\theta)$ und Score-Funktion $S_\theta = \frac{\partial}{\partial \theta} \ln(f(x, \theta))$.
- Viele ML-Schätzer sind asymptotisch effizient, d.h. (stark) konsistent und ihre asymptotische Schätzervarianz ist $1/I(\theta)$

Elfmeter (Bernoullikette, $X_1, X_2, \dots \sim \text{Bin}(1, p)$, u.i.v.-Folge)

- ↪ Score-Funktion: $S_p(x) = \frac{\partial}{\partial p} \ln(p^x(1-p)^{1-x}) = \frac{x}{p} - \frac{1-x}{1-p} = \frac{x-p}{p(1-p)}$
- ↪ Fisher-Information: $I(p) = \text{var}(S_p(X_1)) = \text{var}\left(\frac{X_1-p}{p(1-p)}\right) = \frac{\text{var}(X_1)}{p^2(1-p)^2} = \frac{1}{p(1-p)}$
- ↪ Cramer-Rao-Schranke: $1/I(p) = p(1-p)$
- ↪ Asymptotische Schätzervarianz des ML-Schätzers $T_n = \bar{X}_n$:
 - T_n ist stark konsistent wegen des SGGZ: $\bar{X} \rightarrow p$ f.s.
 - T_n ist asymptotisch normal wegen des ZGS: $\sqrt{n} \frac{\bar{X}_n - p}{\sqrt{p(1-p)}} \rightarrow \mathcal{N}(0, 1)$
 - Die asymptotische Schätzervarianz ist dann $\sigma^2(p) = p(1-p)$
- ↪ Der ML-Schätzer hat die Cramer-Rao-Schranke als asymptotische Schätzervarianz.

Übung: X_1, X_2, \dots sei eine u.i.v.-Folge von $\text{Poi}(\lambda)$ -verteilten ZV (z.B. die Anzahl der Patienten in einer Notaufnahme an aufeinander folgenden Tagen). Bestimmen Sie zunächst einen ML-Schätzer für λ auf Basis von X_1, \dots, X_n und prüfen Sie, ob der Schätzer asymptotisch effizient ist. Hinweis: Dichte: $f_\lambda(x) = \frac{\lambda^x}{x!} e^{-\lambda}, x \in \mathbb{N}_0$

↪ ML-Schätzer: Es ergibt sich $T_n = \bar{X}_n$ wie folgt:

$$\square \ln L(x, \lambda) = \ln\left(\prod \frac{\lambda^{x_i}}{x_i!} e^{-\lambda}\right) = n(\bar{x} \ln(\lambda) - \lambda) - \sum \ln(x_i!)$$

$$\square \text{Notwendig } \frac{\partial}{\partial \lambda} L(x, \lambda) = n\left(\frac{\bar{x}}{\lambda} - 1\right) \stackrel{!}{=} 0 \Leftrightarrow \lambda = \bar{x}$$

$$\square \text{Hinreichend: } \frac{\partial^2}{\partial \lambda^2} L(x, \lambda) = -\bar{x}/\lambda^2 < 0, \text{ also ist die LL konkav, daher Max.}$$

↪ nach SGGZ gilt $E(T_n) \rightarrow E(X_1) = \lambda$ und $\text{var}(X_1) = \lambda$

↪ Nach ZGS gilt $\sqrt{n} \frac{\bar{X}_n - \lambda}{\sqrt{\lambda}} \rightarrow \mathcal{N}(0, 1)$, also $\sigma^2(\lambda) = \lambda$ (as. Schätzervarianz).

$$\hookrightarrow \text{Score-Funktion } S_\lambda(x) = \frac{\partial}{\partial \lambda} \ln(\lambda^x e^{-x}/x!) = \frac{\partial}{\partial \lambda} (x \ln(\lambda) - \lambda - \ln(x!)) = \frac{x}{\lambda} - 1$$

$$\hookrightarrow \text{Fisher-Information: } I(\lambda) = \text{var}(S_\lambda(X_1)) = \text{var}\left(\frac{X_1}{\lambda} - 1\right) = \frac{\text{var}(X_1)}{\lambda^2} = \frac{1}{\lambda}$$

↪ Cramer-Rao-Schranke: $1/I(\lambda) = \lambda$ ist die as. Schätzervarianz von T_n , also ist T_n as. effizient.

4.6 Zusammenfassung

Schätzer sind Zufallsvariablen $T(X_1, \dots, X_n)$ mit denen (z.B.) Parameter der Modellverteilung ermittelt werden sollen.

- ↪ Konstruktionsprinzipien: ML, MM, ...
- ↪ Nicht jeder Schätzer ist gut. Es gilt Vor- und Nachteile abzuwägen.
- ↪ Erwartungstreue als eine grundlegende Eigenschaft (zumindest asymptotisch)
- ↪ Unter den erwartungstreuen Schätzern sind die mit der kleinsten Varianz zu bevorzugen.
- ↪ Allgemein werden Schätzer anhand ihres Risikos verglichen.
- ↪ Schätzervergleich oft nur für wachsendes n durchführbar. Hier sind konsistente Schätzer vorzuziehen. Unter den konsistenten Schätzern sind diejenigen zu bevorzugen, welche asymptotisch effizient sind.

Ausblick: Weil Schätzer zufällig sind, können sie den „richtigen“ Wert eines Parameters nicht angeben. Statt dessen Bestimmung von Bereichen, welche den Parameter mit einer bestimmten (Mindest-)-WS „überdecken“.