2.2.3 Effizienz

Beispiel Wahlumfrage: Gegeben sind zwei erwartungstreue Schätzer (n sei zum einfacheren Rechnen gerade):

$$T_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$T_2 = \frac{1}{n/2}\sum_{i=1}^{n/2} X_i$$

Was unterscheidet formal T_1 von dem unsinnigen Schätzer T_2, der die in der Stichprobe enthaltene Information nicht vollständig ausnutzt?

Vergleiche die Schätzer über ihre Varianz, nicht nur über den Erwartungswert!

Wenn n so groß ist, dass der zentrale Grenzwertsatz angewendet werden kann, dann gilt approximativ

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

und damit

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

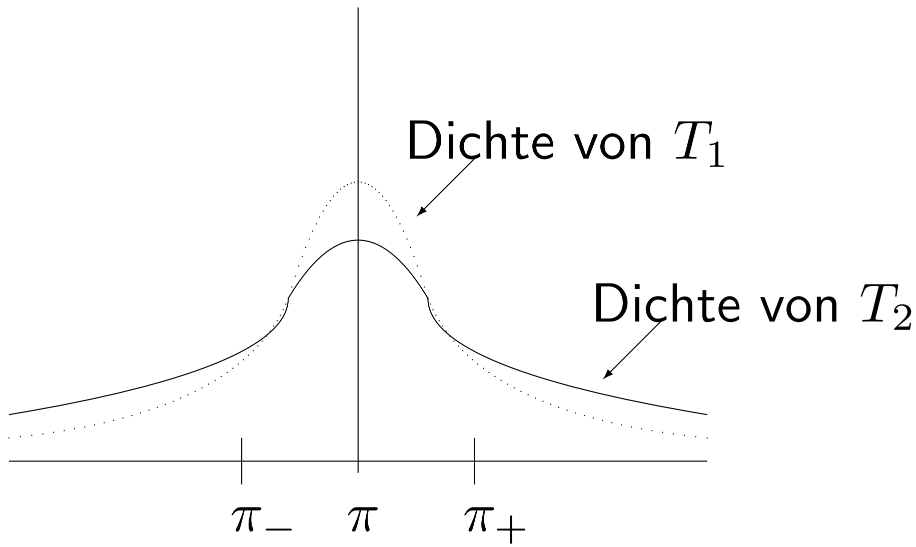
Analog kann man zeigen:

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

T_1 und T_2 sind approximativ normalverteilt, wobei T_1 eine deutlich kleinere Varianz als T_2 hat.

T_1 und T_2 treffen beide im Durchschnitt den richtigen Wert π. Der Schätzer T_1 schwankt aber weniger um das wahre π, ist also „im Durchschnitt genauer“.

Andere Interpretation:



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. $\pi - \pi_-$ daneben zu liegen, bei T_2 stets größer als bei T_1. Umgekehrt gesagt: Ein konkreter Wert ist damit verlässlicher, wenn er von T_1, als wenn er von T_2 stammt.

Diese Überlegung gilt ganz allgemein: Ein erwartungstreuer Schätzer ist umso besser, je kleiner seine Varianz ist.

$$\mathrm{Var}(T) = \text{Erwartete quadratische Abweichung von } T \text{ von } \underbrace{\mathrm{E}(T)}_{=\vartheta}$$

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.

Definition 2.7. *Effizienz*

- Gegeben seien zwei erwartungstreue Schätzfunktionen T_1 und T_2 für einen Parameter ϑ. Gilt

$$\mathrm{Var}_\vartheta(T_1) \leq \mathrm{Var}_\vartheta(T_2) \text{ für alle } \vartheta$$

und

$$\mathrm{Var}_{\vartheta^*}(T_1) < \mathrm{Var}_{\vartheta^*}(T_2) \text{ für mindestens ein } \vartheta^*$$

so heißt T_1 *effizienter als* T_2.

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

$$\mathrm{Var}_\vartheta(T^{\text{opt}}) \leq \mathrm{Var}_\vartheta(T)$$

für alle ϑ und für alle erwartungstreuen Schätzfunktionen T.

Bem. 2.8.

- *Inhaltliche Bemerkung:* Der „Sinn von Optimalitätskriterien“ bei der Auswertung der Stichprobe wird klassischerweise insbesondere auch in der „*Gewährleistung von Objektivität*“, besser Intersubjektivität, 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. Allerdings gibt es wirkliche Eindeutigkeit nur bei „idealen“, sauberen Daten. Z.B. sind ausreißerunempfindliche Verfahren bei „idealen“ Daten weniger effizient, haben aber den Vorteil, stabiler bei kleinen Abweichungen von den Verteilungsannahmen zu sein.

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

  * $\bar{X}$ UMVU-Schätzfunktion für μ bei bekanntem σ^2 und

  * S^2 UMVU-Schätzfunktion für σ^2 bei bekanntem μ.

- 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 $\bar{X}$ UMVU-Schätzfunktion für π.

- Bei nicht erwartungstreuen Schätzern macht es keinen Sinn, sich ausschließlich auf die Varianz zu konzentrieren.

  Man zieht dann den sogenannten *Mean Squared Error*

$$\mathrm{MSE}_\vartheta(T) \quad := \quad \mathrm{E}_\vartheta(T-\vartheta)^2$$

  zur Beurteilung heran. Es gilt

$$\mathrm{MSE}_\vartheta(T) \quad = \quad \mathrm{Var}_\vartheta(T) + (\mathrm{Bias}_\vartheta(T))^2.$$

  Der MSE kann als Kompromiss zwischen zwei Auffassungen von Präzision gesehen werden: möglichst geringe systematische Verzerrung (Bias) und möglichst geringe Schwankung (Varianz).

2.2.4 Asymptotische Gütekriterien

- **Asymptotische Erwartungstreue**

  * Eine Schätzfunktion heißt *asymptotisch erwartungstreu*, falls

$$\lim_{n \to \infty} E_\vartheta(\hat{\vartheta}) = \vartheta$$

  d.h.,

$$\lim_{n \to \infty} \mathrm{Bias}_\vartheta(\hat{\vartheta}) = 0$$

  gilt.

  * Abschwächung des Begriffs der Erwartungstreue: Gefordert wird die Erwartungstreue nur bei einer unendlich großen Stichprobe.

  * Erwartungstreue Schätzer sind auch asymptotisch erwartungstreu.

  * Sowohl S^2 als auch $\tilde{S}^2$ sind asymptotisch erwartungstreu.
  (vgl. vorne $\mathrm{Bias}_{\sigma^2}(\tilde{S}^2) = -\frac{1}{n}\sigma^2$ geht für $n \to \infty$ gegen 0.)

- Für komplexere Modelle ist oft die Erwartungstreue der Verfahren ein zu restriktives Kriterium. Man fordert deshalb oft nur, dass sich der Schätzer wenigstens für große Stichproben gut verhält. Hierzu gibt es verwandte aber „etwas" unterschiedliche Kriterien, z.B. das folgende:

- Ein Schätzer heißt *(MSE-)konsistent* oder *konsistent im quadratischen Mittel*, wenn gilt
$$\lim_{n\to\infty}(\mathrm{MSE}(T)) = 0.$$
Beispiel: Der MSE von $\bar{X}$ ist gegeben durch
$$\begin{aligned}\mathrm{MSE}(\bar{X}) &= \mathrm{Var}(\bar{X}) + \mathrm{Bias}^2(\bar{X})\\ &= \frac{\sigma^2}{n} + 0\\ &= \frac{\sigma^2}{n} \to 0.\end{aligned}$$
$\bar{X}$ ist also ein MSE-konsistenter Schätzer für den Erwartungswert.

- Anschaulich bedeutet die Konsistenz,

2.2.5 Konstruktionsprinzipien guter Schätzer

- **Die Methode der kleinsten Quadrate**
  $\Longrightarrow$ Regressionsanalyse

- **Das Maximum-Likelihood-Prinzip**

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

* Idee der Maximium-Likelihood (ML) Schätzung für diskrete Verteilungen:

- Man kann für jedes ϑ die Wahrscheinlichkeit ausrechnen, genau die Stichprobe

$x_1, \ldots, x_n$ zu erhalten:

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

- Je größer für ein gegebenes ϑ_0 die Wahrscheinlichkeit ist, die konkrete Stichprobe erhalten zu haben, umso plausibler ist es, dass tatsächlich ϑ_0 der wahre Wert ist (gute Übereinstimmung zwischen Modell und Daten).
- Man nennt daher $L(\vartheta) = P_\vartheta(X_1 = x_1, \ldots, X_n = x_n)$, nun als Funktion von ϑ gesehen, die *Likelihood* (deutsch: Plausibilität, Mutmaßlichkeit) von ϑ gegeben die Realisation $x_1, \ldots, x_n$.
- Derjenige Wert $\hat{\vartheta} = \hat{\vartheta}(x_1, \ldots, x_n)$, der $L(\vartheta)$ maximiert, heißt *Maximum-Likelihood-Schätzwert*; die zugehörige Schätzfunktion $T = g(X_1, \ldots, X_n)$ *Maximum-Likelihood-Schätzer* (siehe genauer Definition 2.11).

Bsp. 2.9.

i.i.d. Stichprobe vom Umfang $n = 5$ aus einer $B(10, \pi)$-Verteilung:

6 5 3 4 4

Wahrscheinlichkeit der Stichprobe für gegebenes π, hier notiert durch „$||\ \pi$“:
„$P(\ldots||\pi)$ Wahrscheinlichkeit, wenn π der wahre Parameter ist“

$$\begin{aligned} P(X_1 = 6, \ldots, X_5 = 4 \mid\mid \pi) &= P(X_1 = 6 \mid\mid \pi) \cdot \ldots \cdot P(X_5 = 4 \mid\mid \pi) \\ &= \binom{10}{6} \pi^6 (1-\pi)^4 \cdot \ldots \cdot \binom{10}{4} \pi^4 (1-\pi)^6 . \end{aligned}$$

Wahrscheinlichkeit für einige Werte von π:

π	$P(X_1 = 6, \ldots, X_5 = 4 \,\|\|\, \pi)$
0.1	0.0000000000001
0.2	0.0000000227200
0.3	0.0000040425220
0.4	0.0003025481000
0.5	0.0002487367000
0.6	0.0000026561150
0.7	0.0000000250490
0.8	0.0000000000055
0.9	0.0000000000000

Bem. 2.10.

- Zwei Sichtweisen auf

$$P_\vartheta(X_1 = x_1, \dots, X_n = x_n):$$

  * Deduktiv (Wahrscheinlichkeitsrechnung): ϑ bekannt, $x_1, \dots, x_n$ zufällig („unbekannt“).
  * Induktiv (Statistik): ϑ unbekannt, $x_1, \dots, x_n$ bekannt.

- Für stetige Verteilungen gilt

$$P_\vartheta(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = 0$$

für beliebige Werte ϑ. In diesem Fall verwendet man die Dichte

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

als Maß für die Plausibilität von ϑ.

- Für die praktische Berechnung maximiert man statt der Likelihood typischerweise die sog. *Log-Likelihood* $l(\vartheta)$, also den natürlichen Logarithmus der Likelihood.

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

bzw.

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

Dies liefert denselben Schätzwert $\hat{\vartheta}$ und erspart beim Differenzieren die Anwendung der Produktregel. Manchmal ist es noch geschickter, zunächst $\prod_{i=1}^{n} P_\vartheta(X_i = x_i)$ zu vereinfachen und dann zu logarithmieren.

- Bei den in Statistik II betrachteten „regulären“ Verteilungsmodellen reicht es, die erste Ableitung zu betrachten. Man kann zeigen, dass sie immer auf ein Maximum führt.

Definition 2.11. *(Maximum-Likelihood-Schätzung)*

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

$$L(\vartheta) = \begin{cases} \prod_{i=1}^{n} P_\vartheta(X_i = x_i) & \text{falls } X_i \text{ diskret} \\ \prod_{i=1}^{n} f_\vartheta(x_i) & \text{falls } X_i \text{ stetig.} \end{cases}$$

heißt $Likelihood$ des Parameters ϑ bei der Beobachtung $x_1, \ldots, x_n$.

Derjenige Wert $\hat{\vartheta} = \hat{\vartheta}(x_1, \ldots, x_n)$, der $L(\vartheta)$ maximiert, heißt $Maximum$-$Likelihood$-$Schätzwert$; die zugehörige Schätzfunktion $T = g(X_1, \ldots, X_n)$ $Maximum$-$Likelihood$-$Schätzer$.

Bsp. 2.12. *(Maximum-Likelihood-Schätzer bei der Bernoulli-Verteilung)*

Bsp. 2.13. [ML-Schätzung bei Normalverteilung]

- Der ML-Schätzer $\hat{\mu} = \bar{X}$ für μ stimmt mit dem üblichen Schätzer für den Erwartungswert überein.

- Der ML-Schätzer $\hat{\sigma}^2 = \tilde{S}^2$ für σ^2 ist die Stichprobenvarianz; diese ist nicht erwartungstreu.

Bem. 2.14. [Einige allgemeine Eigenschaften von ML-Schätzern]

Unter allgemeinen Regularitätsbedingungen gilt:

- ML-Schätzer $\hat{\theta}$ sind im Allgemeinen nicht erwartungstreu.
- ML-Schätzer $\hat{\theta}$ sind asymptotisch erwartungstreu.
- ML-Schätzer $\hat{\theta}$ sind konsistent (und meist in einem asymptotischen Sinne effizient).