Karlsruher Institut für Technologie

Institut für Volkswirtschaftslehre (ECON)

Lehrstuhl für Ökonometrie und Statistik

Kapitel V - Erwartungstreue Schätzfunktionen

Induktive Statistik

Prof. Dr. W.-D. Heller

Hartwig Senska

Carlo Siebenschuh

Aufgabe: Angabe eines Schätzwertes für den Parameter einer Verteilung auf der Grundlage einer einfachen Stichprobe mit Zurücklegen.

Gesucht: "Optimale" Entscheidungsfunktion $\delta : \mathcal{X} \rightarrow \Gamma$

$\delta(x_1, x_2, \ldots, x_n)$ Schätzwert beim Stichprobenergebnis $(x_1, \ldots, x_n)$ bzw. bei Verwendung einer (suffizienten und vollständigen) Stichprobenfunktion $T : \mathcal{X} \rightarrow \mathcal{X}^{\mathcal{T}}$

$\delta^T : \mathcal{X}^{\mathcal{T}} \rightarrow \Gamma$

$\delta^T(T(x))$ Schätzwert beim Auswertungsergebnis $T(x)$

Zur Frage der Optimalität wird ein Beurteilungskriterium benötigt.

Entscheidungsfunktionen bei Parameterpunktschätzungen werden als Schätzfunktionen bezeichnet.

Beispiele für sinnvolle Schätzfunktionen:

1. Bernoulliverteilung mit Parameter $p \in [0,1]$
   Stichprobenraum ist $\mathcal{X} = \{0,1\}^n$
   Wegen $p = P(Y = 1)$ ist die relative Häufigkeit des Wertes 1 ein sinnvoller Ansatz für eine Schätzfunktion
   $$\delta(x_1, \ldots, x_n) = \frac{\sum\limits_{i=1}^{n} x_i}{n}$$
   $\sum x_i$ Anzahl der Stichprobeneinheiten mit Wert 1

2. Binomialverteilung $B(m, p)$ mit Parameter $p \in [0, 1]$
   Stichprobenraum ist $\mathcal{X} = \{0,1, \dots, m\}^n$
   Gesetz der großen Zahlen:
   arithmetisches Mittel $\approx$ Erwartungswert mp

Ansatz:

$$mp \approx \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x}$$

Schätzfunktion:

$$\delta(x_1, \dots, x_n) = \frac{1}{m} \cdot \frac{1}{n} \sum_{i=1}^{n} x_i$$

3. Exponentialverteilung mit Parameter $\lambda > 0$:
   Stichprobenraum ist $\mathcal{X} = \mathbb{R}^n_{++} = \{(x_1, \dots, x_n) | x_i > 0\}$
   Gesetz der großen Zahlen:
   arithmetisches Mittel $\approx$ Erwartungswert

Ansatz:

$$\bar{x} \approx \frac{1}{\lambda}$$

Schätzfunktion:

$$\delta(x_1, \dots, x_n) = \frac{1}{\frac{1}{n} \sum_{i=1}^{n} x_i} = \frac{n}{\sum_{i=1}^{n} x_i}$$

4. Normalverteilung:
   a) Parameter μ
      Gesetz der großen Zahlen:
      arithmetisches Mittel $\approx$ Erwartungswert
      Schätzfunktion:

$$\delta(x_1, \ldots, x_n) = \frac{1}{n} \sum_{i=1}^{n} x_i$$

4. Normalverteilung:

b) Parameter σ^2, μ bekannt

Varianz $= E[(Y - E(Y))^2]$

$\approx$ arithmetisches Mittel

der quadratischen Abweichungen vom Erwartungswert

Schätzfunktion:

$$\delta(x_1, \ldots, x_n) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2$$

4. Normalverteilung:

c)Parameter σ^2, μ unbekannt

Ersetzen von μ durch den Schätzwert $\bar{x}$ ergibt die Schätzfunktion:

$$\delta(x_1, \ldots, x_n) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

Bemerkung:
Bei den Beispielen 4. a) und c) wurde nicht der Parameter selbst geschätzt, sondern eine aus dem Parameter ermittelte Größe.

Erläuterung:

Bei 4. ist bei unbekanntem μ und σ^2 der Parameterraum

$$\Gamma = \{(\mu, \sigma^2) \mid \mu \in \mathbb{R},\ \sigma^2 > 0\}$$

Geschätzt wird jeweils nur eine der Komponenten, also ein Funktionswert:

bei a) $\theta(\mu, \sigma^2) = \mu$ bei c) $\theta(\mu, \sigma^2) = \sigma^2$

Weiteres Beispiel dazu:

Exponentialverteilte Lebensdauer

Zu schätzen sei nicht der Parameter λ, sondern die "mittlere Lebensdauer"

$$E(Y) = \frac{1}{\lambda} = \theta(\lambda)$$

Dies kann meist auf das Schätzen eines Parameters zurückgeführt werden, indem man die Verteilung auf den Parameter λ^{-1} umparametrisiert.

$\alpha = \frac{1}{\lambda}$ Dichtefunktion

$$f(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{\alpha} e^{-\frac{1}{\alpha}x} & x \geq 0 \end{cases}$$

Sind die Schätzfunktionen in den Beispielen wirklich gut?
Benötigt: Beurteilungskriterium für Schätzfunktionen
Das Ergebnis der Schätzung (Entscheidung) hängt ab von

- dem Schätzwert $\hat{\gamma}$
- dem wahren Parameter γ

$S(\hat{\gamma}, \gamma)$ *"Schaden" beim Schätzwert* $\hat{\gamma}$
und wahrem Parameter γ
Funktion mit den Argumenten $\hat{\gamma}$ *und* γ

S wird als **Schadensfunktion** bezeichnet.

Wie kann eine Schadensfunktion aussehen?

Ein Parameter γ aus $\Gamma \subset \mathbb{R}$
Wesentlich ist die Abweichung

$$(\hat{\gamma} - \gamma)$$

des Schätzwerts vom wahren Wert.

Die genaue Gestaltung der Schadensfunktion hängt von der Anwendungssituation ab.

Beispiele für Schadensfunktionen:
1. Toleranzkriterium: d "Toleranz"

$$S(\hat{\gamma},\gamma)=\begin{cases} 0, & |\hat{\gamma}-\gamma|\leq d \quad \text{"}\hat{\gamma}\text{ ist ein "guter" Schätzwert"} \\ 1, & |\hat{\gamma}-\gamma|> d \quad \text{"}\hat{\gamma}\text{ ist ein "schlechter" Schätzwert"} \end{cases}$$

2. Quadratische Schadensfunktion (Taguchi):

$$S(\hat{\gamma},\gamma)=(\gamma-\hat{\gamma})^2$$

Verallgemeinerung für mehrere Parameter:

Parameter $\gamma = (\gamma_1, \ldots, \gamma_k)$; Schätzwert $\hat{\gamma} = (\hat{\gamma}_1, \ldots, \hat{\gamma}_k)$

1. Toleranzkriterium: d_i Toleranz für Komponente i

$$S(\hat{\gamma}, \gamma) = \begin{cases} 0, & |\hat{\gamma}_i - \gamma_i| \leq d_i \text{ für } i = 1, 2, \ldots, k \\ 1, & |\hat{\gamma}_i - \gamma_i| > d_i \text{ für ein } i \in \{1, \ldots, k\} \end{cases}$$

2. Quadratische Schadensfunktion:

$$S(\hat{\gamma}, \gamma) = \sum_{i=1}^{k} (\gamma_i - \hat{\gamma}_i)^2$$

Das Schätzen mehrerer Parameter kann nacheinander, separat voneinander erfolgen, indem die jeweils übrigen als fest betrachtet werden.

Im Folgenden: ein Parameter, also $\Gamma \subset \mathbb{R}$

Beurteilung einer Schätzfunktion mit einer Schadensfunktion:

$\delta : \mathcal{X} \rightarrow \Gamma$ Schätzfunktion

$\delta(x_1, \dots, x_n)$ Schätzwert beim Stichprobenergebnis $(x_1, \dots, x_n)$

$(X_1, \dots, X_n)$ zufällig $\Rightarrow \delta(X_1, \dots, X_n)$ zufällig

Konsequenz:

$$S(\delta(X_1, \dots, X_n), \gamma) \text{ ist zufällig.}$$

Bei der Beurteilung einer Schätzfunktion ist diese Zufälligkeit zu berücksichtigen.

Damit: Beurteilung der Schätzfunktion δ durch Analyse der Zufallsvariable $S(\delta(X_1, \ldots, X_n), \gamma)$

Beispiel

Lineare Schätzfunktion für den Mittelwert einer Normalverteilung

$(x_1, x_2, \ldots, x_n)$ Stichprobenergebnis

$\delta(x_1, x_2, \ldots, x_n) = a_0 + a_1 x_1 + \ldots + a_n x_n$ ist eine lineare Funktion

Ist diese Funktion als Schätzfunktion für den Mittelwert geeignet

und wie sind gegebenenfalls die Koeffizienten $a_0, a_1, \ldots, a_n$ zu wählen?

$$X_i \quad N(\mu, \sigma^2)\text{-verteilt} \Rightarrow a_i X_i \qquad N(a_i\mu, a_i^2\sigma^2)\text{-verteilt}$$

$$\Rightarrow a_0 + \sum_{i=1}^{n} a_i X_i \qquad N(\tilde{\mu}, \tilde{\sigma^2})\text{-verteilt}$$

$$\text{mit } \tilde{\mu} = a_0 + \sum_{i=1}^{n} a_i\mu, \quad \tilde{\sigma^2} = \sum_{i=1}^{n} a_i^2\sigma^2$$

Verteilung von $S(\delta(X_1, \ldots, X_n), \gamma)$ ergibt sich aus der Schadensfunktion S.
Toleranzkriterium: Werte der Schadensfunktion sind 0 und 1

$$P(S(\delta(X), \mu) = 0) = ?$$

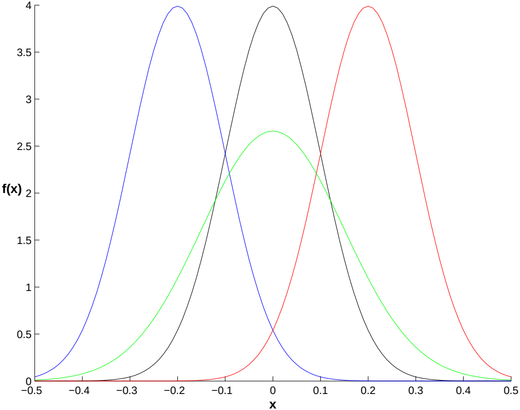
$$P(S(\delta(X), \mu) = 1) = 1 - ?$$

Es gilt:

$$
\begin{aligned}
P(S(\delta(X), \mu) = 0) &= P(|\delta(X) - \mu| \leq d) \\
&= P(\mu - d \leq \delta(X) \leq \mu + d) \\
&= F_{\tilde{\mu},\tilde{\sigma^2}}(\mu + d) - F_{\tilde{\mu},\tilde{\sigma^2}}(\mu - d)
\end{aligned}
$$

Zielsetzung: Maximierung dieser Wahrscheinlichkeit
Dies wird erreicht durch:

$\tilde{\mu} = \mu$ und $\tilde{\sigma^2}$ minimal (siehe Graphik für Begründung)



"Dichte für $\widetilde{\mu} = \mu$ und $\widetilde{\sigma^2}$ min"

"**Zusätzlich:** Dichte für $\widetilde{\mu} < \mu$ und σ^2 min."

"**Zusätzlich:** Dichte für $\widetilde{\mu} > \mu$ und $\widetilde{\sigma^2}$ min."

"**Zusätzlich:** Dichte für $\widetilde{\mu} = \mu$ und $\widetilde{\sigma^2} >$ min."

$$\tilde{\mu} = \mu : \qquad a_0 + \sum_{i=1}^{n} a_i \mu = \mu$$

Zwei Lösungen (Dies sind nicht die einzigen Lösungen. Jede Linearkombination dieser Lösungen ist ebenfalls eine Lösung):

1. $a_0 = \mu, a_1 = \ldots = a_n = 0$,
2. $a_0 = 0, a_1 + \ldots + a_n = 1$.

Lösung 1 kommt nicht in Frage, da μ unbekannt ist.
Lösung 2:

$$\sum_{i=1}^{n} a_i X_i \text{ ist } N(\mu, \tilde{\sigma^2})\text{-verteilt mit } \tilde{\sigma^2} = \sum_{i=1}^{n} a_i^2 \sigma^2$$

Aufgabe:

$$\text{Minimiere } \sum_{i=1}^{n} a_i^2 \text{ unter } \sum_{i=1}^{n} a_i = 1$$

Lösung:

$$a_1 = a_2 = \ldots = a_n = \frac{1}{n}$$

Das Stichprobenmittel ist die "beste lineare Schätzfunktion bezüglich des Toleranzkriterium".

Übungsaufgabe: Analyse der linearen Schätzfunktionen bzgl. der quadratischen Schadensfunktion.

Analyse der Zufallsvariable $S(\delta(X_1, X_2, \ldots, X_n), \gamma)$

$\hookrightarrow$ an Hand der Wahrscheinlichkeitsverteilung (s.Beispiel)
$\hookrightarrow$ oder mit Kenngrößen

- Erwartungswert: $E_\gamma[S(\delta(X_1, X_2, \ldots, X_n), \gamma)]$
- Varianz: $Var_\gamma[S(\delta(X_1, X_2, \ldots, X_n), \gamma)]$
- ......

Der Erwartungswert $E_\gamma[S(\delta(X), \gamma)]$ zu gegebener Schadensfunktion hängt ab von:

- dem Parameter γ
- der Wahrscheinlichkeitsverteilung von X, die aber auch mit γ variiert
- der Schätzfunktion δ

Definition:

$R(\delta, \gamma) = E_\gamma[S(\delta(X), \gamma)]$ heißt **Risikofunktion** zur Schadensfunktion S,
oder -ausführlicher- **erwarteter Schaden bei Verwendung der Schätzfunktion δ und dem wahren Parameter γ**.

Beispiele:

1. Schadensfunktion Toleranzkriterium
   $E_\gamma[S(\delta(X),\gamma)] = 1P_\gamma(|\delta(X)-\gamma| > d) + 0P_\gamma(|\delta(X)-\gamma| \leq d)$
   Wahrscheinlichkeit für einen Schätzwert außerhalb der Toleranz.
2. Quadratische Schadensfunktion
   $E_\gamma[S(\delta(X),\gamma)] = E_\gamma[(\delta(X)-\gamma)^2]$
   Mittlerer quadratischer Fehler(MQF),
   Mean Square Error(MSE)

Für den MQF gilt:

$$
\begin{aligned}
& E_\gamma[(\delta(X)-\gamma)^2] \\
= \quad & E_\gamma[(\delta(X)-E_\gamma(\delta(X))+E_\gamma(\delta(X))-\gamma)^2] \\
= \quad & E_\gamma[(\delta(X)-E_\gamma(\delta(X)))^2+2(\delta(X)-E_\gamma(\delta(X)))(E_\gamma(\delta(X))-\gamma) \\
& +(E_\gamma(\delta(X))-\gamma)^2] \\
= \quad & E_\gamma[(\delta(X)-E_\gamma(\delta(X)))^2]+2E_\gamma[(\delta(X)-E_\gamma(\delta(X)))(E_\gamma(\delta(X))-\gamma)] \\
& +E_\gamma[(E_\gamma(\delta(X))-\gamma)^2] \\
= \quad & Var_\gamma(\delta(X))+2(E_\gamma(\delta(X))-\gamma)\underbrace{E_\gamma[\delta(X)-E_\gamma(\delta(X))]}_{=0}+(E_\gamma(\delta(X))-\gamma)^2 \\
= \quad & Var_\gamma(\delta(X))+(E_\gamma(\delta(X))-\gamma)^2
\end{aligned}
$$

Der Mittlere Quadratische Fehler

$$E_\gamma[(\delta(X) - \gamma)^2] = Var_\gamma(\delta(X)) + (E_\gamma(\delta(X)) - \gamma)^2$$

setzt sich also zusammen aus:

- der Varianz der Schätzfunktion $Var_\gamma \ (\delta(X))$ und
- dem Quadrat der Abweichung des Erwartungswertes der Schätzfunktion vom wahren Parameter: $(E_\gamma(\delta(X)) - \gamma)^2$

Die Abweichung $E_\gamma(\delta(X)) - \gamma$ wird auch **Verzerrung(Bias) der Schätzfunktion** genannt.

Zielsetzung:
Verwendung einer Schätzfunktion mit minimalem Risiko für jedes γ.

z.B.:
"Ratefunktionen" $\delta_{a_0}(X) = a_0$, d.h. der Schätzwert hängt nicht vom Stichprobenergebnis ab, wird also "geraten".

MQF einer Ratefunktion:

$$R(\delta_{a_0}, \gamma) = E_\gamma[(\delta_{a_0}(X) - \gamma)^2] = E_\gamma(a_0 - \gamma)^2 = (a_0 - \gamma)^2$$

Der MQF ist 0 für $\gamma = a_0$, d.h. eine Schätzfunktion, die der Zielsetzung genügt, müsste auch an der Stelle a_0 mindestens so gut sein wie die Ratefunktion mit dem Wert a_0, also an der Stelle a_0 den MQF 0 haben. Da a_0 beliebig gewählt werden kann, müsste diese Schätzfunktion überall den MQF 0 haben. Eine solche Schätzfunktion gibt es aber nicht.

Folgerung:
Die oben angegebene Zielsetzung hat keine Lösung: Es gibt keine gleichmäßig (d.h. für alle γ) beste Schätzfunktion unter allen Schätzfunktionen.

Ausweg:
Die Verzerrung wird als systematischer Fehler betrachtet.
Man fordert daher $E_\gamma(\delta(X)) - \gamma = 0$ oder $E_\gamma(\delta(X)) = \gamma$.

Definition:
Eine Schätzfunktion $\delta : \mathcal{X} \to \Gamma$ heißt **erwartungstreu(unverzerrt, unbiased)**, wenn

$$E_\gamma(\delta(X)) = \gamma \text{ für alle } \gamma \in \Gamma$$

gilt.

Also: Einschränkung auf erwartungstreue Schätzfunktionen.

Damit ergibt sich die -neue- Zielsetzung:
Unter allen erwartungstreuen Schätzfunktionen wird die gesucht, deren MQF

$$Var_\gamma(\delta(X)) + (E_\gamma(\delta(X)) - \gamma)^2 = Var_\gamma(\delta(X))$$

minimal ist, also mit minimaler Varianz für alle $\gamma \in \Gamma$. Eine solche Schätzfunktion ist dann die **gleichmäßig**(d.h. für alle γ) **beste Schätzfunktion unter den erwartungstreuen Schätzfunktionen,** kurz die **gleichmäßig beste erwartungstreue Schätzfunktion.**

Definition:
Eine Schätzfunktion $\delta : \mathcal{X} \to \Gamma$ für einen Parameter $\gamma \in \Gamma$ heißt **gleichmäßig beste erwartungstreue Schätzfunktion für den Parameter** γ, wenn sie erwartungstreu ist und für jede erwartungstreue Schätzfunktion δ'

$$Var_\gamma(\delta(X)) \leq Var_\gamma(\delta'(X)) \text{ für alle } \gamma \in \Gamma$$

gilt.

Beispiele für erwartungstreue Schätzfunktionen:

1. Bernoulliverteilung
   Vorgeschlagene Schätzfunktion
   $$\delta(x_1, \ldots, x_n) = \frac{\sum\limits_{i=1}^{n} x_i}{n}$$
   $\sum\limits_{i=1}^{n} X_i$ ist binomialverteilt mit Erwartungswert np.
   Erwartungswert der Schätzfunktion ist damit p, sie ist also erwartungstreu.

2. Binomialverteilung $B(m, p)$ mit Parameter $p \in [0, 1]$
   Vorgeschlagene Schätzfunktion

$$\delta(x_1, \ldots, x_n) = \tfrac{1}{m} \cdot \tfrac{1}{n} \sum_{i=1}^{n} x_i$$

   $\sum\limits_{i=1}^{n} X_i$ ist $B(nm, p)-$ verteilt mit Erwartungswert nmp.
   Erwartungswert der Schätzfunktion ist damit p, sie ist also erwartungstreu.

3. Exponentialverteilung mit Parameter $\lambda > 0$:
   Vorgeschlagene Schätzfunktion:
$$\delta(x_1, \ldots, x_n) = \frac{1}{\frac{1}{n}\sum\limits_{i=1}^{n} x_i}$$
   $Z_n = \sum\limits_{i=1}^{n} X_i$ ist Erlang-verteilt mit der Dichte:

$$f_{Z_n,\lambda}(z) = \begin{cases} \frac{\lambda^n}{(n-1)!} \cdot z^{n-1} \cdot e^{-\lambda z} & z > 0 \\ 0 & z \leq 0 \end{cases}$$

Damit ist

$$
\begin{aligned}
E_\lambda\left(\frac{n}{Z_n}\right) &= \int\limits_0^\infty \frac{n}{z} \cdot \frac{\lambda^n}{(n-1)!} \cdot z^{n-1} \cdot e^{-\lambda z} dz \\
&= \frac{n\lambda}{(n-1)!} \cdot \int\limits_0^\infty \lambda \cdot (\lambda z)^{n-2} \cdot e^{-\lambda z} dz
\end{aligned}
$$

Mit der Substitution $w = \lambda z$:

$$
= \frac{n\lambda}{(n-1)!} \int_0^\infty w^{n-2} \cdot e^{-w} dw = \frac{n}{n-1}\lambda
$$

Die Schätzfunktion ist folglich **nicht** erwartungstreu für λ.

4. Normalverteilung

   Parameter μ

   Vorgeschlagene Schätzfunktion:

$$\delta(x_1, \ldots, x_n) = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$\delta(X_1, \ldots, X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i \qquad \text{ist } N(\mu, \tfrac{\sigma^2}{n})\text{-verteilt,}$$

hat also den Erwartungswert μ und ist damit erwartungstreu.

5. Parameter ist der Erwartungswert der Zufallsvariable: $E_\gamma(Y) = \gamma$:
Der Stichprobenmittelwert
$$\delta(X_1, \ldots, X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i$$
hat den Erwartungswert $E_\gamma(Y) = \gamma$, ist also erwartungstreu.
Aber auch jede lineare Schätzfunktion
$$\delta(X_1, \ldots, X_n) = \sum_{i=1}^{n} a_i X_i \text{ mit } \sum_{i=1}^{n} a_i = 1$$
ist erwartungstreu.

6. Parameter ist die Varianz der Zufallsvariable: $Var_\gamma(Y) = \gamma$:
   Vorgeschlagene Schätzfunktion:
   $$\delta(x_1, \ldots, x_n) = \tfrac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$$
   Ist diese Schätzfunktion erwartungstreu?

$$\begin{aligned}
E_\gamma(\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2) &= E_\gamma(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X^2}) \\
&= E_\gamma(\frac{1}{n}\sum_{i=1}^{n} X_i^2) - E_\gamma(\bar{X^2}) \\
&= E_\gamma(\frac{1}{n}\sum_{i=1}^{n} X_i^2) - E_\gamma(Y)^2 - E_\gamma(\bar{X^2}) + E_\gamma(Y)^2
\end{aligned}$$

$$
\begin{aligned}
&= E_\gamma(\frac{1}{n}\sum_{i=1}^{n} X_i^2) - E_\gamma(Y)^2 - (E_\gamma(\bar{X^2}) + E_\gamma(Y)^2) \\
&= \frac{1}{n}\sum_{i=1}^{n} E_\gamma(X_i^2) - E_\gamma(Y)^2 - Var_\gamma(\bar{X}) \\
&= \frac{1}{n}\sum_{i=1}^{n}(E_\gamma(Y^2) - E_\gamma(Y)^2) - \frac{1}{n}Var_\gamma(Y) \\
&= \frac{1}{n}(\sum_{i=1}^{n} Var_\gamma(Y) - Var_\gamma(Y)) \\
&= \frac{1}{n}(nVar_\gamma(Y) - Var_\gamma(Y)) = \frac{n-1}{n}Var_\gamma(Y)
\end{aligned}
$$

Die Stichprobenvarianz ist **nicht** erwartungstreu für die Varianz einer Verteilung.

Bemerkung:

In den Beispielen 3. und 6. erhalten wir keine erwartungstreuen Schätzfunktionen. Da aber die Verzerrung bekannt ist, kann die Schätzfunktion entsprechend korrigiert werden, indem wir mit dem Kehrwert des jeweiligen "Fehlerfaktors" multiplizieren.

Beispiel: Exponentialverteilung

$$\frac{n-1}{n} \cdot \frac{1}{\frac{1}{n}\sum_{i=1}^{n} x_i} = \frac{n-1}{\sum_{i=1}^{n} x_i}$$

ist erwartungstreu für λ.

Beispiel: Varianz

$$\frac{n}{n-1} \cdot \frac{1}{n} \sum_{i=1}^{n} (x_1 - \bar{x})^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_1 - \bar{x})^2$$

Die korrigierte Stichprobenvarianz ist damit erwartungstreu für die Varianz.

Wie finden wir eine gleichmäßig beste erwartungstreue Schätzfunktion?

Wenn wir eine vollständige und suffiziente Statistik bezüglich dem gesuchten Parameter kennen, so sollten wir versuchen, diese zu benutzen.

Sei $T : \mathcal{X} \to \mathcal{X}^{\mathcal{T}}$ eine suffiziente und vollständige Statistik bzgl. γ.

$\delta^T : \mathcal{X}^{\mathcal{T}} \to \Gamma$ eine Schätzfunktion unter Verwendung von T:
$\delta^T(T(x))$ ist der Schätzwert zum Auswertungsergebnis $T(x)$ des Stichprobenergebnisses x.
δ^T ist erwartungstreu, wenn $E_\gamma(\delta^T(T(x))) = \gamma$ für alle $\gamma \in \Gamma$ gilt.

Sind δ_1^T und δ_2^T zwei erwartungstreue Schätzfunktionen, so gilt

$$E_\gamma[\delta_1^T(T(X))] = \gamma = E_\gamma[\delta_2^T(T(X))], \quad \text{für alle } \gamma$$

und wegen der Vollständigkeit von T folgt

$$P_\gamma\left(\delta_1^T(T(X)) = \delta_2^T(T(X))\right) = 1.$$

Die beiden Schätzfunktionen stimmen also fast sicher überein. Es gibt also im wesentlichen nur eine erwartungstreue Schätzfunktion unter Verwendung von T, falls es überhaupt eine erwartungstreue Schätzfunktion unter Verwendung von T gibt.

Damit ergeben sich folgende Fragen:

1. Gibt es eine erwartungstreue Schätzfunktion unter Verwendung von T?
2. Wenn ja, ist diese dann -wie zu erwarten- die gleichmäßig beste erwartungstreue Schätzfunktion?

Zur ersten Frage:

Behauptung: Wenn es überhaupt eine erwartungstreue Schätzfunktion gibt, dann auch eine unter Verwendung von T.

Begründung:
Sei δ eine erwartungstreue Schätzfunktion. Betrachten wir die Teilbereiche von $\mathcal{X}$ mit jeweils übereinstimmendem Auswertungsergebnis $T(x) = t$ bei T. Variiert δ in diesen Teilbereichen, so kann die Schätzfunktion möglicherweise verbessert werden, indem wir sie in diesen Teilbereichen glätten, d.h. die Variation und damit die Varianz reduzieren.

Sei Y eine diskrete Zufallsvariable und X der zugehörige Stichprobenvektor. Sei t ein Auswertungsergebnis bei T und $\{x \in \mathcal{X} | T(x) = t\}$ der zugehörige Teilbereich von $\mathcal{X}$ mit

$$P_\gamma(T(X) = t) \neq 0$$

Da T suffizient ist, ist $P(X = x \mid T(X) = t)$ unabhängig von γ.
Wir können δ in diesem Bereich glätten, indem wir die Werte auf das mittlere Niveau, also den Erwartungswert in diesem Bereich nivellieren:

$$E(\delta(X) \mid T(X) = t) = \sum_{x:\ T(x)=t} \delta(x) P(X = x \mid T(X) = t)$$

$$E(\delta(X) \mid T(X) = t) = \sum_{x:\ T(x)=t} \delta(x) P(X = x \mid T(X) = t)$$

Die rechte Seite ist unabhängig von γ, also auch die linke Seite und wir setzen

$$u(t) = \begin{cases} E[\delta(X) \mid T(X) = t] & P_\gamma(T(X) = t) \neq 0 \\ 0 & \text{sonst} \end{cases}$$

Dann ist u eine Schätzfunktion unter Verwendung von T:

$\tilde{\delta}(X) = u(T(X))$ ist erwartungstreu und

$Var_\gamma(\tilde{\delta}(X)) \leq Var_\gamma(\delta)$ (Antwort auf Frage 2)

Haben wir also eine erwartungstreue Schätzfunktion gefunden, so können wir durch Glättung in den Teilbereichen $T(x) = t$ **die** erwartungstreue Schätzfunktion finden, die die suffiziente und vollständige Stichprobenfunktion benutzt. Diese ist dann gleichmäßig beste erwartungstreue Schätzfunktion.

Beispiel: Poissonverteilung

$$T(x_1, \ldots, x_n) = \sum_{i=1}^{n} x_i$$

ist eine suffiziente und vollständige Statistik für den Parameter λ der Poissonverteilung.

X_1 ist identisch verteilt wie Y, d.h. $\delta(X_1, \ldots, X_n) = X_1$ ist eine erwartungstreue Schätzfunktion für λ, sicherlich keine sehr gute, da nur der erste Stichprobenwert berücksichtigt wird.

Wie lautet die geglättete Schätzfunktion?

$$
\begin{aligned}
u(t) &= E(\delta(X) \mid T(X) = t) \\
&= \sum_{x:\ T(x)=t} \delta(x) P(X = x \mid T(X) = t) \\
&= \sum_{x:\ T(x)=t} x_1 P(X = x \mid T(X) = t) \\
&= \sum_{x_1 \leq t} x_1 \sum_{x_2,\ldots,x_n: \sum\limits_{i=2}^{n} x_i = t - x_1} P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n \mid T(X) = t) \\
&= \sum_{x_1 \leq t} x_1 P(X_1 = x_1 \mid T(X) = t)
\end{aligned}
$$

Es soll zweimal $\sum_{x_1 \leq t}$ durch $\sum_{0 \leq x_1 \leq t}$ ersetzt werden.

$$
\begin{aligned}
P(X_1 = x_1 \mid T(X) = t) &= \frac{P(X_1 = x_1, T(X) = t)}{P(T(X) = t)} \\
&= \frac{P(X_1 = x_1, \sum\limits_{i=1}^{n} X_i = t)}{P(\sum\limits_{i=1}^{n} X_i = t)} \\
&= \frac{P(X_1 = x_1, \sum\limits_{i=2}^{n} X_i = t - x_1)}{P(\sum\limits_{i=1}^{n} X_i = t)}
\end{aligned}
$$

$$
\begin{aligned}
P(X_1 = x_1 \mid T(X) = t) &= \frac{P(X_1 = x_1) P(\sum\limits_{i=2}^{n} X_i = t - x_1)}{P(\sum\limits_{i=1}^{n} X_i = t)} \\
&= \frac{\frac{\lambda^{x_1}}{x_1!} \cdot e^{-\lambda} \cdot \frac{((n-1)\lambda)^{t-x_1}}{(t-x_1)!} \cdot e^{-(n-1)\lambda}}{\frac{(n\lambda)^t}{t!} \cdot e^{-n\lambda}} \\
&= \binom{t}{x_1} \cdot \frac{(n-1)^{t-x_1}}{n^t} \\
&= \binom{t}{x_1} \cdot \Big(\frac{n-1}{n}\Big)^{t-x_1} \cdot \Big(\frac{1}{n}\Big)^{x_1}
\end{aligned}
$$

ist die Wahrscheinlichkeit der $B(t, \frac{1}{n})$-Verteilung mit Parameter $p = \frac{1}{n}$ an der Stelle x_1.

Damit ist $u(t)$ der Erwartungswert einer $B(t, n^{-1})$-Verteilung, also

$$u(t) = t \cdot \tfrac{1}{n}$$

und die geglättete Schätzfunktion lautet

$$\tilde{\delta}(x) = u(\sum_{i=1}^{n} x_i) = \tfrac{1}{n} \sum_{i=1}^{n} x_i = \overline{x}$$

Für die Varianzen der beiden Schätzfunktionen gilt:

$$Var_\lambda[\tilde{\delta}(X)] = \tfrac{1}{n} Var(Y) = \tfrac{\lambda}{n} < \lambda = Var_\lambda(X_1) = Var_\lambda(\delta(X))$$

Gleichmäßig beste erwartungstreue Schätzfunktionen:
Nach Kapitel 4 gilt

$$T(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} x_i$$

ist suffizient und vollständig für den Parameter

a) p der Bernoulli-Verteilung
b) p der Binomialverteilung $B(m, p)$
c) λ der Poisson-Verteilung
d) λ der Exponentialverteilung
e) μ der Normalverteilung bei fester Varianz

Weiter gilt:
$\bar{X} = \frac{1}{n} \sum\limits_{i=1}^{n} X_i$ ist erwartungstreu für den Erwartungswert der Verteilung von Y.

Damit ist

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

gleichmäßig beste erwartungstreue Schätzfunktion in den Fällen a), c) und e).
Bei der Binomialverteilung (Fall b)) ist

$$\frac{1}{m}\bar{x} = \frac{1}{mn} \sum_{i=1}^{n} x_i$$

erwartungstreu und damit GBE-Schätzfunktion.

Bei der Exponentialverteilung (Fall d)) ist

$$\frac{n-1}{\sum\limits_{i=1}^{n} x_i}$$

erwartungstreu und damit GBE-Schätzfunktion.

Bemerkung:

Sei Y die exponentialverteilte Lebensdauer eines Geräts, es gilt also

$$P(Y \leq t) = 1 - e^{-\lambda t} \text{ für } t > 0$$

Dann ist λ die Ausfallrate, d.h. der Anteil der pro Zeiteinheit ausfallenden Geräte, und λ^{-1} die erwartete Lebensdauer. Die Aufgabe kann jetzt sowohl

- Schätze die Ausfallrate mittels einer Stichprobe
  als auch
- Schätze die erwartete Lebensdauer $\tau = \lambda^{-1}$ mittels einer Stichprobe
  lauten.

Für die Schätzwerte

$$\hat{\lambda} = \frac{n-1}{\sum\limits_{i=1}^{n} x_i} \text{ und } \hat{\tau} = \bar{x}$$

gilt also nicht wie bei den zu schätzenden Werten, dass die erwartete Lebensdauer der Kehrwert der Ausfallrate ist:

$$\tau = \frac{1}{\lambda} \qquad \text{aber} \qquad \hat{\tau} \neq \frac{1}{\hat{\lambda}}$$

GBE-Schätzfunktionen für die Parameter der Normalverteilung:

1. μ ist zu schätzen, σ^2 fest: s.o. $\bar{X}$ ist GBE-Schätzfunktion.

2. σ^2 ist zu schätzen, μ ist bekannt:

$$\sum_{i=1}^{n}(x_i - \mu)^2$$

ist suffizient und vollständig bezüglich σ^2 (siehe Kap.4)

$\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$ ist erwartungstreu für σ^2:

$$\begin{aligned}
E_{\sigma^2} \quad & (\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2) = \frac{1}{n}\sum_{i=1}^{n} E_{\sigma^2}(X_i - E_{\sigma^2}(X_i) + E_{\sigma^2}(X_i) - \mu)^2 \\
= \quad & \frac{1}{n}\sum_{i=1}^{n}[E_{\sigma^2}((X_i - E_{\sigma^2}(X_i))^2) + 2E_{\sigma^2}((X_i - E_{\sigma^2}(X_i))(\underbrace{E_{\sigma^2}(X_i)-\mu}_{0})) \\
& +(\underbrace{E_{\sigma^2}(X_i)-\mu}_{0})^2] \\
= \quad & \frac{1}{n}\sum_{i=1}^{n} Var_{\sigma^2}(X_i) = \sigma^2
\end{aligned}$$

Damit ist

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$

GBE-Schätzfunktion für σ^2 bei bekanntem μ.

3. μ und σ^2 sind unbekannt und zu schätzen:

$T(x) = (\sum_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i^2)$ ist suffizient und vollständig für (μ, σ^2).

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

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

sind erwartungstreu für μ und σ^2, verwenden die Statistik und sind damit GBE-Schätzfunktionen für μ und σ^2.