2. Punktschätzung von Parametern einer Grundgesamtheit

Nur durch eine **Totalerhebung** kann man die Verteilung einer Zufalls-variablen X in einer Grundgesamtheit vollständig beschreiben.

Häufig ist jedoch eine Totalerhebung entweder unmöglich oder einfach zu teuer. Deswegen versucht man mit Hilfe von **Teilerhebungen ((Zu-falls-)Stichproben)** Anhaltspunkte über die Verteilung einer Grund-gesamtheit und über deren Parameter zu gewinnen.

Unterscheidungen der Schätzungen:

- mit oder ohne Zurücklegen

- homograder Fall vs. heterograder Fall

- **homograde Variablen** unterscheiden, ob ein Objekt eine bestimmte Eigenschaft besitzt oder nicht, (0,1)-Variable: Schätzung von Anteilssätzen

- **heterograde Variablen** (metrisch skalierte ZV): Schätzung von Mittelwerten und Varianzen

- Punkt- oder Intervallschätzungen

Bei der **Punktschätzung** handelt es sich um die Bestimmung eines **einzelnen** Wertes zur Schätzung eines unbekannten Parameters.

Da Punktschätzungen mit sehr hoher Wahrscheinlichkeit mit dem wahren Wert des Parameters nicht übereinstimmen, ergänzt man eine Punktschätzung oft durch eine **Intervallschätzung.** Dabei berechnet man ein Intervall, das den wahren unbekannten Wert des Parameters der Grundgesamtheit mit einer vorgegebenen Wahrscheinlichkeit über-deckt.

2.1. Punktschätzung für Mittelwert und Anteilswerte

2.1.1. Punktschätzung für den Mittelwert (heterograder Fall)

Der unbekannte Mittelwert μ der metrischen Zufallsvariablen X einer Grundgesamtheit soll anhand einer Zufallsstichprobe von **Umfang *n*** geschätzt werden.

Das arithmetische Mittel der Stichprobenelemente stellt den Schätzwert $\hat{\mu}$ für den unbekannten Mittelwert μ dar:

$$\{x_1, x_2, ..., x_n\} \Rightarrow \bar{x} = \frac{1}{n}\sum x_i \qquad \hat{\mu} = \bar{x}$$

Die Punktschätzung $\hat{\mu} = \bar{x}$ ist die Realisation der Zufallsvariablen $\bar{X}_n$.

Der geschätzte Wert $\hat{\mu}$ ist von dem wahren Mittelwert μ der Grundgesamtheit sorgfältig zu unterscheiden. Die Abweichung des geschätzten Wertes $\hat{\mu}$ vom wahren Mittelwert μ wird **Schätzfehler** genannt: $e = \mu - \hat{\mu}$

Bei den meisten Schätzungen treten Schätzfehler auf. Entscheidend ist aber, ob die Punktschätzung den wahren Wert **im Durchschnitt** trifft.

$$E(\mu) = E(\overline{X}_n) = \frac{1}{n} E(\sum X_i) = \frac{1}{n} \sum E(X_i) = \frac{1}{n} n\mu \quad \Rightarrow \quad E(\hat{\mu}) = \mu$$

$\Rightarrow$ $\hat{\mu}$ ist eine **erwartungstreue Schätzung** für μ .

Die für die Schätzung des Mittelwertes der Grundgesamtheit verwendete Schätzmethode nennt man **Momentenmethode**. Sie schätzt das unbekannte 1. unzentrierte Moment der Grundgesamtheit (Mittelwert) mit dem entsprechenden 1. unzentrierten Moment der Stichprobe.

2.1.2 Punktschätzung für den Anteilswert (homograder Fall)

Analog zur Schätzung des unbekannten Mittelwertes einer Grundgesamtheit durch das arithmetische Mittel der Stichprobe ist zur Schätzung eines unbekannten **Anteilwertes *p*** einer Grundgesamtheit die Punktschätzung durch den **Stichprobenanteilswert** sehr gut geeignet:

$$\hat{p} = h \qquad mit \quad \hat{p}: Schätzwert\ für\ den\ Anteilsatz\ in\ der\ Grundgesamtheit$$

$$h: \ Anteilsatz\ in\ der\ Stichprobe$$

Diese Schätzung ist eine **erwartungstreue** und **konsistente** Schätzung.

$$E(\hat{p}) = E(H_n) = \frac{1}{n}E(\sum B_i) = \frac{1}{n}\sum E(B_i) = \frac{1}{n}np \quad \Rightarrow \quad E(\hat{p}) = p \quad (Erwartungstreue)$$

$$V(\hat{p}) = V(H_n) = V(\frac{1}{n}\sum B_i) = \frac{1}{n^2}\sum V(B_i) = \frac{1}{n^2}npq = pq/n \quad \Rightarrow \quad (Konsistenz)$$

Mit zunehmendem Stichprobenumfang wird die Varianz immer kleiner und die Schätzung genauer.

2.2 Punktschätzung für Varianzen und Standardabweichungen

2.2.1 Punktschätzung für die Varianz (heterograder Fall)

Wenn man bei der Punktschätzung der Varianz die Momentenmethode direkt anwendet (d.h. als Schätzwert für die unbekannte Varianz der Grundgesamtheit $\hat{\sigma}^2$ die empirische Varianz der Stichprobe s^2 verwendet), bekommt man nicht erwartungstreue, sondern **verzerrte** Schätzer:

Das 2. zentrale Moment (= empirische Varianz der Stichprobe):

$$s^2 = \frac{1}{n}\sum(x_i - \bar{x})^2$$

$$E(S^2) = E\left[\frac{1}{n}\sum(X_i - \mu)^2\right] - E(\bar{X}_n - \mu)^2$$

$$= \frac{1}{n}\sum E(X_i - \mu)^2 - E(\bar{X}_n - \mu)^2 = \frac{1}{n}\sum V(X_i) - V(\bar{X}_n) = \sigma^2 - \sigma^2/n = \frac{n-1}{n}\sigma^2$$

Diese Punktschätzung ist somit um den Faktor *(n-1)/n* **verzerrt.**

(S^2: Zufallsvariable, s^2: Realisation der Zufallsvariablen).

Da die Verzerrung durch den Faktor *(n-1)/n* ausgedrückt wird, korrigiert man mit dem Kehrwert dieses Faktors:

$$\hat{\sigma}^2 = \frac{n}{n-1} s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2$$

(*n-1* ist die Zahl der Freiheitsgrade)

Mit dieser Formel erhält man **erwartungstreue** Schätzer für die Varianz.

Die obige Formel gilt für den Fall **mit Zurücklegen**.

Im Fall **ohne Zurücklegen** muss noch der **Korrekturfaktor** berücksichtigt werden:

$$Kf = \frac{N-n}{N-1} < 1$$

Schätzformel für die Varianz im Fall **ohne Zurücklegen**:

$$\hat{\sigma}^2 = \frac{N-1}{N} \frac{n}{n-1} s^2$$

2.3 Eigenschaften von Punktschätzungen

Schätzwert einer Punktschätzung eines Parameters ist eine **Zufallsvariable** und kann viele mögliche Werte annehmen.

Der unbekannte Parameter selbst ist eine **konstante Größe**.

Allgemeine Schreibweise:

$$\theta : \ der \ \ wahre \ Wert \ \ der \ \ Grundgesamtheit$$

$$\hat{\theta} : Schätzwert \ \ des \ Parameters$$

$$\hat{\theta} = \hat{\theta}(X_1, ..., X_n) : Schätzfunktion$$

Schätzfunktionen sind durch bestimmte **stochastische Eigenschaften** charakterisiert, die Auskunft darüber geben, wie gut Schätzfunktionen für bestimmte Zwecke geeignet sind.

Erwartungstreue

Ein Schätzer $\hat{\theta}$ ist **erwartungstreu** oder **unverzerrt** („unbiased“), wenn sein Erwartungswert dem wahren Wert des zu schätzenden Parameters entspricht:

$$E(\hat{\theta}) = \theta$$

Asymptotische Erwartungstreue

Die asymptotische Erwartungstreue ist eine schwächere Eigenschaft als die Erwartungstreue. Bei nicht erwartungstreuen Schätzern impliziert sie, dass die Verzerrung mit zunehmendem Stichprobenumfang geringer wird und für $n \to \infty$ verschwindet:

$$\lim_{n \to \infty} E(\hat{\theta}) = \theta$$

Konsistenz

Ein Schätzer ist **konsistent**, wenn er erwartungstreu (oder mindestens asymptotisch erwartungstreu) ist und seine Varianz bei zunehmendem Stichprobenumfang gegen Null geht:

$$P(|\hat{\theta} - \theta| > \varepsilon) \rightarrow 0 \quad \textit{für} \quad n \rightarrow \infty$$

$$\textit{oder} \quad \underset{n \rightarrow \infty}{p\lim}\ \hat{\theta} = \theta$$

Effizienz

Ein Schätzer heißt **effizient**, wenn er in der Klasse aller erwartungstreuen Schätzer die kleinste Varianz besitzt:

$$V(\theta^*) < V(\hat{\theta})$$

Mittlerer Quadratischer Fehler

Schätzprobleme: Erwartungstreue wird mit zu hoher Varianz erkauft.

Der mittlere quadratische Fehler (mean squared error MSE) eines Schätzers ist definiert als der Erwartungswert der quadrierten Differenz zwischen Schätzwert und wahren Wert:

$$MSE(\hat{\theta}) := E[(\hat{\theta} - \theta)^2]$$

Bei erwartungstreuen Schätzern ist dies die Varianz, bei den verzerrten erhält man nach dem Verschiebungssatz mit $d = \theta$ die Formel

$$MSE(\hat{\theta}) := V(\hat{\theta}) + bias^2$$

d.h. mittlerer quadratischer Fehler berücksichtigt beides, die Varianz und den Bias.

2.4 Schätzprinzipien

= wichtigsten allgemeine Ansätze zur Gewinnung von Schätzmethoden und Schätzformeln

a) Momentenmethode

Man schätzt die Momente der Verteilung der Grundgesamtheit mit entsprechenden Momenten der Stichprobe

Grundgesamtheit	Stichprobe
$E(X) = \mu$	$\frac{1}{n}\sum X_j$
$E(X^2) = \mu^2 + \sigma^2$	$\frac{1}{n}\sum X_j^2$
$\vdots$	$\vdots$
$E(X^m)$	$\frac{1}{n}\sum X_j^m$

Die Schätzformeln, die sich hieraus ergeben lauten:

$$\hat{\mu} = \overline{x}$$

$$\hat{\mu}^2 + \hat{\sigma}^2 = \frac{1}{n}\sum x_j^2 \text{, d.h. } \hat{\sigma}^2 = \frac{1}{n}\sum x_j^2 - \overline{x}^2 = s^2$$

Schätzung für Varianz nur asymptotisch erwartungstreu, jedoch wenn μ bekannt wäre, würde $\hat{\sigma}^2 = \frac{1}{n}\sum x_j^2 - \mu^2$ eine erwartungstreue Schätzung für die Varianz darstellen.

b) Schätzprinzip der kleinsten Quadrate (KQ-Methode)

= Wähle diejenige Größe $\hat{\mu}_{KQ}$ als Schätzwert für den unbekannten Mittelwert μ, von der die Summe der quadrierten Abstände zu den Stichprobenmittelwerten x_i minimal ist:

$$\sum_{j=1}^{n}(x_j - \hat{\mu}_{KQ})^2 \xrightarrow[\hat{\mu}_{KQ}]{} Minimum$$

Minimierung durch Ableiten und Nullsetzen:

$$\sum_{j=1}^{n} 2(x_j - \hat{\mu}_{KQ})(-1) = (-2)\sum (x_j - \hat{\mu}_{KQ}) = 0$$

$$\sum x_j - \sum \hat{\mu}_{KQ} = \sum_{j=1}^{n} x_j - n\hat{\mu}_{KQ} = 0$$

$$\hat{\mu}_{KQ} = \frac{1}{n}\sum x_j = \overline{x}$$

= Minimierung der Fehlervarianz, wobei die Bewertung der Fehler nicht proportional, sondern quadratisch erfolgt, d.h. doppelt so große Fehler bringen vierfachen Verlust in die Bewertung (quadratische Verlustfunktion)

c) Maximum-Likelihood-Prinzip

= Prinzip der größten Mutmaßlichkeit

Wähle $\hat{\theta}_{ML}$ als Schätzwert für einen unbekannten Parameter θ, welcher angesichts des Stichprobenergebnisses die größte Likelihood hat bzw. verglichen mit allen anderen Werten das Stichprobenergebnis mit der größten Wahrscheinlichkeit hervorgebracht hätte.

Beispiel: Urne mit Kugeln mit Anteilswert p = 0,1 und p = 0,5, wobei letzterer eine größere Likelihood hat.

Zur Bestimmung eines Maximum-Likelihood-Schätzers benötigt man eine sogenannte *Likelihood-Funktion L*.

= Massenfunktion oder Dichtefunktion der gemeinsamen Verteilung der X_i, wobei der Parameter θ als Variable angesehen und für x_i die beobachteten Stichprobenwerte verwendet werden. L ist adbei eine Funktion von θ:

$$L(\theta; x_1, x_2, \ldots, x_n) := f(x_1, x_2, \ldots, x_n; \theta)$$

Zur Bestimmung der Schätzfunktion löst man folgendes Maximierungs-problem:

$$L(\hat{\theta}_{ML}, x_1, x_2, \ldots, x_n) \xrightarrow[\hat{\theta}_{ML}]{} Minimum$$

wobei ein Verteilungstyp der X_i *a priori* vorgegeben wird

Beispiele:

Gesucht: Maximum-Likelihood-Schätzer für den Erwartungswert μ und die Standardabweichung σ einer Normalverteilung mit der gemeinsamen Dichtefunktion der *i.i.d.* ZV:

$$\begin{aligned} f(x_1, x_2, \ldots, x_n; \mu, \sigma) &= f_N(x_1; \mu, \sigma) \ldots f_N(x_n; \mu, \sigma) \\ &= \frac{1}{\sigma^n \sqrt{2\pi}^n} \exp\left[\frac{-1}{2\sigma^2} \sum_{j=1}^{n} (x_j - \mu)^2 \right] \\ &= L(\mu, \sigma; x_1, \ldots, x_n) \end{aligned}$$

Logarithmus der Likelihood-Funktion wird maximiert:

$$\ln L = -n \ln \hat{\sigma} - n \ln \sqrt{2\pi} - \frac{1}{2\hat{\sigma}^2} \sum_{j=1}^{n} (x_j - \hat{\mu})^2 \xrightarrow[\hat{\mu},\hat{\sigma}]{} Maximum$$

(monoton steigende Funktion)

I. ML-Schätzer für μ:

Partielles Ableiten von $\ln L$ nach $\hat{\mu}$ und Nullsetzen:

$$\frac{\partial}{\partial \hat{\mu}} \ln L = -\frac{1}{2\hat{\sigma}^2} \sum_{j=1}^{n} 2(x_j - \hat{\mu})(-1) = 0$$

$$\sum_{j=1}^{n}(x_j - \hat{\mu}) = 0$$

$$\hat{\mu}_{ML} = \overline{x} \text{ (Stichprobenmittelwert)}$$

II. ML-Schätzer für σ:

ML-Schätzung für μ, nach σ ableiten und Nullsetzen:

$$\frac{\partial}{\partial \hat{\sigma}} \ln L = -n\frac{1}{\hat{\sigma}} - \frac{1}{2\hat{\sigma}^3}(-2)\sum_{j=1}^{n}(x_j - \overline{x})^2 = 0$$

$$\hat{\sigma}_{ML} = \frac{1}{n}\sum_{j=1}^{n}(x_j - \overline{x})^2 \text{ (asymptotisch erwartungstreu)}$$

Beispiele:

Gesucht: ML-Schätzer für unbekannten Anteilswert p einer Grundgesamtheit.

Zufallsstichprobe vom Umfang n mit x Elementen für das dichotome Merkmal. Likelihood-Funktion lautet:

$$f(x;p) = p^x(1-p)^{n-x} = L(p;X=x)$$

Nach p ableiten und Ableitung Nullsetzen:

$$xp^{x-1}(1-p)^{n-x} + p^x(n-x)(1-p)^{n-x-1}(-1)$$

$$= xp^{x-1}(1-p)^{n-x-1}\left[(1-p)-\frac{n-x}{x}p\right]=0$$

Da der erste Faktor für $p=0$ und $p=1$ Null ist, minimieren beide Lösungen die Likelihood-Funktion. Setzt man jedoch den zweiten Faktor gleich Null, erhält man die Bestimmungsgleichung für den ML-Schätzer

$$1-\hat{p}_{ML}-\frac{n-x}{x}\hat{p}_{ML}=1-\frac{n}{x}\hat{p}_{ML}=0$$

$$\hat{p}_{ML}=\frac{x}{n}=h$$