

Statistik

Dr. Thomas Zehrt

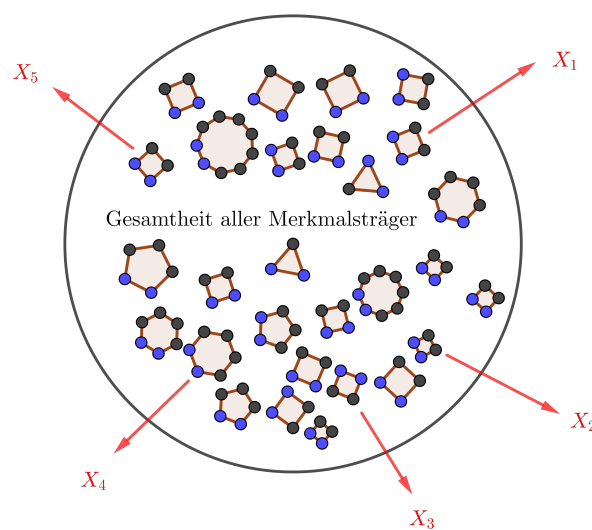
(Punkt)Schätzen

Motivation

Eine vollständige Information über die Verteilung eines Merkmals X in einer Grundgesamtheit kann nur durch eine Totalerhebung gewonnen werden. Es ist aber in der Praxis oft so, dass eine Totalerhebung nicht möglich bzw. zu teuer ist, oder dass uns ein Schätzwert völlig ausreicht. In diesem Fall versucht man mit Hilfe einer so genannten Teilerhebung (oder Stichprobe) Anhaltspunkte über die unbekannte Verteilung zu gewinnen.

Eine Stichprobe der Grösse $n(=5)$ wird dabei durch n Zufallsvariablen X_1, X_2, \dots, X_n mathematisch modelliert (**bevor** man tatsächlich zieht):

- X_1 : erste Ziehung eines Elementes aus der Gesamtheit,
- X_2 : zweite Ziehung eines Elementes aus der Gesamtheit,
- ...
- X_n : n -te Ziehung eines Elementes aus der Gesamtheit.



Die einzelnen Ziehungen sollen **unabhängig** voneinander vollzogen werden, d.h. das Ergebnis einer Ziehung soll die Ergebnisse anderer Ziehungen nicht stochastisch beeinflussen. Ausserdem wünschen wir uns, dass hinter jeder Ziehung die selbe Grundgesamtheit (an Möglichkeiten) steht. Erreichbar wäre das sicher, falls wir stets zufällig und mit Zurücklegen ziehen würden. Mathematisch drücken wir das wie folgt aus: Die (Ziehungs)Zufallsvariablen X_1, X_2, \dots, X_n sind **identisch verteilt und unabhängig**.

Benötigtes Schulwissen

- Summenzeichen, etwas Verständnis für Grenzwerte

Beispiele:

$$\begin{aligned}\lim_{n \rightarrow \infty} \frac{\sigma^2}{n} &= \sigma^2 \cdot \lim_{n \rightarrow \infty} \frac{1}{n} = \sigma^2 \cdot 0 = 0 \\ \lim_{n \rightarrow \infty} \frac{n}{n-1} \cdot \sigma^2 &= \sigma^2 \cdot \lim_{n \rightarrow \infty} \frac{n}{n-1} = \sigma^2 \cdot \lim_{n \rightarrow \infty} \frac{n}{n(1-1/n)} = \sigma^2 \cdot 1\end{aligned}$$

- Logarithmusfunktion, Rechenregeln für Logarithmen, Ableiten von Logarithmen (die Ableitung nach p wird hier stets als $\frac{\partial}{\partial p}$ geschrieben)

$$\begin{aligned}\ln((1-p)^k) &= k \cdot \ln(1-p) \\ \frac{\partial}{\partial p} \ln(1-p) &= \frac{1}{1-p} \cdot (-1)\end{aligned}$$

1 Grundproblem

Ausgangspunkt ist wieder eine endliche oder unendliche Grundgesamtheit, wobei wir im endlichen Fall aber davon ausgehen wollen, dass die Anzahl N ihrer Elemente so gross ist, dass eine vollständige Untersuchung praktisch unmöglich ist. An den Elementen dieser Grundgesamtheit interessiert uns wieder ein Merkmal (Zufallsvariable) X (z.B. Alter, Masse, Länge, ...) mit der Verteilung $P(X = x)$ (bzw. der Verteilungsfunktion F) und falls wir dieses Merkmal untersuchen wollen, haben wir keine andere Möglichkeit als eine (Zufalls)Stichprobe vom Umfang $n \ll N$ zu entnehmen, diese zu untersuchen und von diesen Untersuchungsergebnissen auf die Grundgesamtheit zu schliessen. Wir wollen dieses einleuchtende Verfahren mittels Zufallsvariablen präzisieren.

Definition 1.1 *Eine Zufallsstichprobe vom Umfang n ist eine Folge von unabhängigen Zufallsvariablen (X_1, X_2, \dots, X_n) , wobei X_i die Merkmalsausprägung des i -ten Elementes in der Stichprobe bezeichnet. Dabei sollen alle diese Zufallsvariablen die selbe Verteilung wie X haben. Jede Realisierung (x_1, x_2, \dots, x_n) dieser Zufallsvariablen heisst eine konkrete Stichprobe.*

Es gibt offensichtlich zwei (eigentlich vier) Möglichkeiten, eine Stichprobe von n Elementen zu nehmen:

1. mit Zurücklegen

In diesem Fall sind die Zufallsvariablen (X_1, X_2, \dots, X_n) tatsächlich unabhängig und identisch verteilt.

2. ohne Zurücklegen

In diesem Fall sind die Zufallsvariablen (X_1, X_2, \dots, X_n) nicht identisch verteilt und auch nicht unabhängig. Falls N sehr gross ist, können wir aber noch glauben, dass die obigen Zufallsvariablen **fast** unabhängig und identisch verteilt sind.

Um also die Unabhängigkeit und identische Verteilung der Zufallsvariablen zu garantieren, beachten wir:

N klein	\longrightarrow	Stichprobe mit Zurücklegen
N gross oder unendlich	\longrightarrow	Stichprobe mit oder ohne Zurücklegen

Die Information, die man auf Grund der Stichprobe über die Grundgesamtheit gewinnen möchte, lässt sich meistens in Form so genannter Parameter darstellen. Das ist eine feste aber unbekannt reelle Zahle, die wir meist mit dem griechischen Buchstaben θ abkürzen werden, und im weiteren wird stets angenommen, dass die Verteilung von X von diesem unbekannt Parameter abhängt.

Schätzprinzip der schliessenden Statistik

Ein unbekannter Parameter der Grundgesamtheit wird durch den entsprechenden Parameter der Stichprobe geschätzt.

Es ist klar, dass die Qualität der Schätzung von der konkreten Stichprobe abhängt.

2 Punktschätzung

Zur Schätzung eines Parameters θ wird aus den n Stichprobenvariablen ein Wert berechnet.

Definition 2.1 *Eine Funktion*

$$g = g(X_1, X_2, \dots, X_n)$$

der Stichprobenvariablen heisst Stichprobenfunktion (und ist selbst wieder eine Zufallsvariable). Wird eine Stichprobenfunktion zur Schätzung eines Parameters θ genutzt, so heisst sie Schätzfunktion oder Schätzer für θ und wird mit $\hat{\theta}$ bezeichnet.

Beachte: θ ist ein fester Wert also insbesondere **keine** Zufallsvariable, $\hat{\theta}$ ist eine Zufallsvariable.

2.1 Einige einfache Beispiele

Stichprobenmittel

Es sei $\mu := E(X)$ mit $X :=$ „Merkmalsausprägung eines zufällig ausgewählten Elementes der Grundgesamtheit,“

Ist die Grundgesamtheit endlich mit den N Elementen $1, 2, \dots, N$ und den jeweils zugehörigen Merkmalsausprägungen x_1, x_2, \dots, x_N , so gilt hier natürlich

$$\mu = \frac{1}{N}(x_1 + x_2 + \dots + x_N)$$

und μ ist das arithmetische Mittel aller Merkmalsausprägungen.

Dann ist das Stichprobenmittel für eine Stichprobe (X_1, X_2, \dots, X_n) vom Umfang n

$$\hat{\mu} := \frac{1}{n}(X_1 + X_2 + \dots + X_n)$$

eine Schätzfunktion für μ .

Es soll hier nochmals erwähnt werden, dass die Konstruktion von Schätzfunktionen und deren Untersuchung stets vollzogen werden sollte, **bevor** man eine konkrete Stichprobe genommen hat. Alle Schätzfunktionen sind Zufallsvariablen und hinter jeder Zufallsvariablen X_i stehen **alle** möglichen Realisierungen des Merkmals mit ihren zugehörigen Wahrscheinlichkeiten.

Anteilssatz

Sei p der prozentuale Anteil der Elemente der Grundgesamtheit, die eine gewisse fixierte Eigenschaft A haben. Dann ist der Anteilssatz

$$\hat{p} := \frac{\text{Anzahl der Elemente in der Stichprobe mit Eigenschaft } A}{n}$$

ein Schätzer für p .

Stichprobenvarianzen

Sei X wie oben definiert und $\sigma^2 := \text{Var}(X)$ die Varianz der Grundgesamtheit. Zunächst können wir zwei Fälle unterscheiden: Wir könnten den Erwartungswert $\mu = E(X)$ kennen oder eben nicht. Im zweiten Fall müssten wir also auch diesen Erwartungswert schätzen. Ich biete Ihnen nun vier verschiedene Schätzer für σ^2 an:

$$\begin{aligned}\hat{\sigma}_1^2 &:= \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 \\ \hat{\sigma}_2^2 &:= \frac{1}{n-1} \sum_{i=1}^n (X_i - \mu)^2 \\ \hat{\sigma}_3^2 &:= \frac{1}{n} \sum_{i=1}^n (X_i - \hat{\mu})^2 \\ \hat{\sigma}_4^2 &:= \frac{1}{n-1} \sum_{i=1}^n (X_i - \hat{\mu})^2\end{aligned}$$

Beispiel 2.1 Grundgesamtheit und Stichprobe *Eine endliche Grundgesamtheit bestehe aus drei Elementen, nämlich 1, 2, 3. Eine Zufallsstichprobe vom Umfang $n = 2$ ist eine Folge von unabhängigen Zufallsvariablen (X_1, X_2) (X_i ist die Merkmalsausprägung des i -ten Elementes in der Stichprobe).*

Für die Gesamtheit gilt (bitte nachrechnen) $\mu = 2$ und $\sigma^2 = 2/3$.

Listen Sie alle möglichen Realisierungen (x_1, x_2) von (X_1, X_2) auf und bestimmen Sie jeweils den Mittelwert sowie $\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$, $\hat{\sigma}_3^2$ und $\hat{\sigma}_4^2$ jeder Realisierung (konkrete Stichprobe).

i	i -te Real.	$\hat{\mu}$	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$	$\hat{\sigma}_3^2$	$\hat{\sigma}_4^2$
1	(1, 1)	1	1	2	0	0
2	(2, 2)	2	0	0	0	0
3	(3, 3)	3	1	2	0	0
4	(1, 2)	1.5	0.5	1	0.25	0.5
5	(2, 1)	1.5	0.5	1	0.25	0.5
6	(1, 3)	2	1	2	1	2
7	(3, 1)	2	1	2	1	2
8	(2, 3)	2.5	0.5	1	0.25	0.5
9	(3, 2)	2.5	0.5	1	0.25	0.5

Für die vierte konkrete Stichprobe $X_1 = 1$ und $X_2 = 2$ gilt z.B.:

$$\begin{aligned}\hat{\mu} &= \frac{1}{2}(1+2) = 1.5 \\ \hat{\sigma}_1^2 &= \frac{1}{2} [(1-2)^2 + (2-2)^2] = 0.5 \\ \hat{\sigma}_2^2 &= \frac{1}{1} [(1-2)^2 + (2-2)^2] = 1 \\ \hat{\sigma}_3^2 &= \frac{1}{2} [(1-1.5)^2 + (2-1.5)^2] = 0.25 \\ \hat{\sigma}_4^2 &= \frac{1}{1} [(1-1.5)^2 + (2-1.5)^2] = 0.5\end{aligned}$$

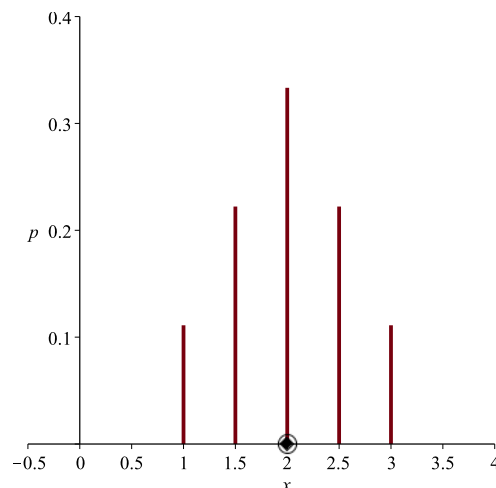
Wir wollen nun die Erwartungswerte der fünf Schätzfunktionen $\hat{\mu}$ sowie $\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$, $\hat{\sigma}_3^2$ und $\hat{\sigma}_4^2$ bestimmen und uns dann überlegen, welche Schätzfunktionen man benutzen sollte.

Für der Erwartungswert der Schätzfunktion $\hat{\mu}$ gilt

$$\begin{aligned} E(\hat{\mu}) &= \frac{1}{9}(1 + 2 + 3 + 1.5 + 1.5 + 2 + 2 + 2.5 + 2.5) \\ &= \frac{1}{9}(1 \cdot 1 + 2 \cdot 1.5 + 3 \cdot 2 + 2 \cdot 2.5 + 1 \cdot 3) \\ &= 1 \cdot \frac{1}{9} + 1.5 \cdot \frac{2}{9} + 2 \cdot \frac{3}{9} + 2.5 \cdot \frac{2}{9} + 3 \cdot \frac{1}{9} = 2 \end{aligned}$$

Wir nehmen diese Umformungen hier nur vor, um ein Stabdiagramm der Situation zeichnen zu können. Durch den Diamanten wird der echte zu schätzende Parameter gekennzeichnet, dagegen ist der Kreis der Erwartungswert der Schätzfunktion (Mittelwert der Häufigkeitsverteilung).

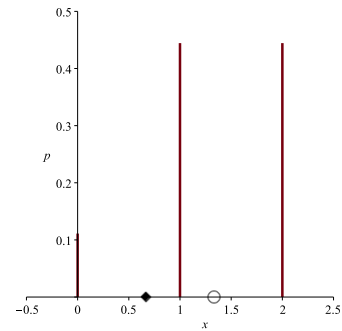
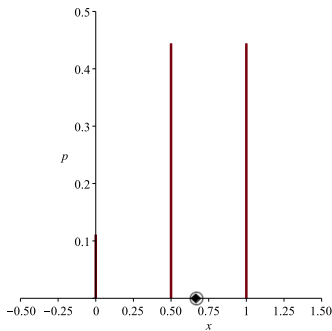
Wenn wir auch nicht erwarten können, dass der Mittelwert einer konkreten Stichprobe auch der Mittelwert der Gesamtheit ist, so können wir doch sehen, dass wir **im Durchschnitt** gut schätzen! Der Erwartungswert der Schätzfunktion ist der zu schätzende Parameter der Gesamtheit!



Nach dem gleichen Prinzip können wir nun auch für die vier Varianzschätzer den Erwartungswert bestimmen (und das entsprechende Stabdiagramm zeichnen). Wieder ist der Diamant der echte zu schätzende Parameter (immer $2/3$) und der Kreis der Erwartungswert der jeweiligen Schätzfunktion.

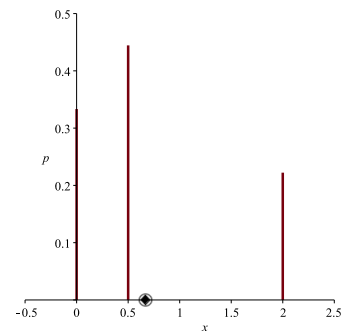
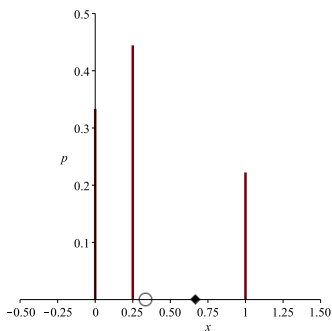
$$E(\hat{\sigma}_1^2) = \frac{1}{9}(4 \cdot 0.5 + 4 \cdot 1) = \frac{2}{3}$$

$$E(\hat{\sigma}_2^2) = \frac{12}{9} = \frac{4}{3}$$



$$E(\hat{\sigma}_3^2) = \frac{1}{3}$$

$$E(\hat{\sigma}_4^2) = \frac{2}{3}$$



Wir sehen, dass zwei der Schätzfunktionen im Durchschnitt (man sagt auch: systematisch) falsch schätzen! Solche Schätzer sollte man besser nicht benutzen.

2.2 Erwartungstreue und Konsistenz

2.2.1 Grundlagen

Eine Schätzfunktion ist eine Zufallsvariable, sie kann also bei jedem Versuch (Stichprobe) anders ausfallen. Schon gar nicht kann man erwarten, dass $\hat{\theta}$ den Parameter θ exakt trifft.

Dagegen sollte eine gute Schätzfunktion zumindest **im Mittel** richtig schätzen, d.h. ihr Erwartungswert sollte gleich dem zu schätzenden Parameter sein.

Definition 2.2 Eine Schätzfunktion $\hat{\theta}$ für den Parameter θ heisst erwartungstreu, wenn $E(\hat{\theta}) = \theta$ gilt.

Der Erwartungswert des Schätzfehlers $E(\theta - \hat{\theta})$ heisst Verzerrung oder Bias $bias(\hat{\theta})$ der Schätzfunktion.

Sie heisst konsistent, wenn ihre Varianz mit wachsendem Stichprobenumfang n gegen 0 konvergiert:

$$\lim_{n \rightarrow \infty} \text{Var}(\hat{\theta}) = 0.$$

2.2.2 Stichprobenmittel und Stichprobenvarianz

Satz 1 Das Stichprobenmittel $\hat{\mu}$ für μ ist erwartungstreu und konsistent.

Beweis:

Wir nutzen die allgemeinen Rechenregeln für Erwartungswerte und Varianzen.

$$\begin{aligned} E(\hat{\mu}) &= E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \\ &= E\left(\frac{1}{n} X_1 + \frac{1}{n} X_2 + \dots + \frac{1}{n} X_n\right) \\ &= \frac{1}{n} E(X_1) + \frac{1}{n} E(X_2) + \dots + \frac{1}{n} E(X_n) \\ &= \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} \cdot n \cdot \mu = \mu \end{aligned}$$

Damit ist die Erwartungstreue gezeigt. Weiterhin gilt:

$$\text{Var}(\hat{\mu}) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{1}{n} \cdot \sigma^2 \rightarrow 0$$

und die Konsistenz ist damit bewiesen. □

Bemerkung: Würden wir aus einer N -elementigen Grundgesamtheit eine Stichprobe ohne Zurücklegen ziehen, wären die Zufallsvariablen X_i nicht unabhängig. Wir dürften also die Varianz der Summe nicht einfach als Summe der Einzelvarianzen schreiben. Ausserdem können wir auch nicht beliebig grosse Stichproben mehr nehmen. In diesem Fall gilt:

$$\text{Var}(\hat{\mu}) = \frac{N-n}{N-1} \frac{\sigma^2}{n}$$

Mit wachsendem n wird $\text{Var}(\hat{\mu})$ kleiner und im Fall $n = N$ verschwindet die Varianz.

Satz 2 Die Stichprobenvarianzen $\hat{\sigma}_1^2$ und $\hat{\sigma}_4^2$ für σ^2 sind erwartungstreu und konsistent.

Beweis:

Da für jede Zufallsvariable X der Verschiebungssatz $Var(X) = E(X^2) - E(X)^2$ gilt, folgt weiterhin für jede der Zufallsvariablen X_i und für die Zufallsvariable $\hat{\mu}$:

$$\begin{aligned} E(X_i^2) &= Var(X_i) + E(X_i)^2 = \sigma^2 + \mu^2 \\ E(\hat{\mu}^2) &= Var(\hat{\mu}) + E(\hat{\mu})^2 = \frac{1}{n} \sigma^2 + \mu^2 \end{aligned}$$

- Wir wollen zeigen, dass $\hat{\sigma}_1^2$ erwartungstreu ist. Es gilt zunächst:

$$\begin{aligned} \hat{\sigma}_1^2 &= \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 \\ &= \frac{1}{n} \sum_{i=1}^n (X_i^2 - 2\mu X_i + \mu^2) \\ &= \frac{1}{n} \sum_{i=1}^n X_i^2 - \frac{1}{n} 2\mu \underbrace{\sum_{i=1}^n X_i}_{=n \cdot \hat{\mu}} + \frac{1}{n} \sum_{i=1}^n \mu^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - 2\mu \hat{\mu} + \mu^2 \end{aligned}$$

Daraus folgt dann für den Erwartungswert:

$$E(\hat{\sigma}_1^2) = \frac{1}{n} \sum_{i=1}^n E(X_i^2) - 2\mu E(\hat{\mu}) + \mu^2 = \frac{1}{n} \sum_{i=1}^n (\sigma^2 + \mu^2) - \mu^2 = \sigma^2$$

- Wir wollen zeigen, dass $\hat{\sigma}_4^2$ erwartungstreu ist, was insbesondere daran liegt, dass man bei der Definition der Varianz durch $(n-1)$ anstatt durch n teilt. Es gilt zunächst:

$$\hat{\sigma}_4^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \hat{\mu})^2 = \frac{1}{n-1} \sum_{i=1}^n X_i^2 - \frac{n}{n-1} \hat{\mu}^2$$

und somit insgesamt:

$$\begin{aligned} E(\hat{\sigma}_4^2) &= \frac{1}{n-1} \sum_{i=1}^n E(X_i^2) - \frac{n}{n-1} E(\hat{\mu}^2) \\ &= \frac{1}{n-1} \sum_{i=1}^n (\sigma^2 + \mu^2) - \frac{n}{n-1} \left(\frac{1}{n} \sigma^2 + \mu^2 \right) \\ &= \frac{1}{n-1} n (\sigma^2 + \mu^2) - \frac{n}{n-1} \left(\frac{1}{n} \sigma^2 + \mu^2 \right) \\ &= \sigma^2 \end{aligned}$$

Auf die schwierigeren Beweise der Konsistenz verzichten wir. □

2.2.3 *Linearkombinationen der X_i *

Seien X_1, X_2, \dots, X_n unabhängige und gleichverteilte Zufallsvariablen mit $E(X_i) = \mu$ und $Var(X_i) = \sigma^2$ (z.B. aus einer Zufallsstichprobe (X_1, X_2, \dots, X_n)). Wir hatten im vorherigen Kapitel schon das Stichprobenmittel

$$\hat{\mu} = \frac{1}{n}(X_1 + X_2 + \dots + X_n) = \frac{1}{n}X_1 + \frac{1}{n}X_2 + \dots + \frac{1}{n}X_n$$

auf Erwartungstreue und Konsistenz untersucht. In der Praxis entspricht die Nutzung der Schätzfunktion $\hat{\mu}$ der Situation, in der wir jeder der n Messungen das selbe Vertrauen schenken. Jede einzelne Messung X_i geht also mit dem gleichen Gewicht $1/n$ in das Resultat $\hat{\mu}$ ein.

In der Praxis sind aber auch Situationen denkbar, in denen man nicht jeder Messung das gleiche Vertrauen schenkt! Wir wollen uns deshalb Schätzfunktionen für μ von folgender Art ansehen:

$$T := \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_n X_n = \sum_{j=1}^n \alpha_j X_j$$

wobei $\alpha_i \in \mathbb{R}$ gut gewählte und nicht-negative reelle Zahlen sind. Beachte, dass $\alpha_i = \alpha_i(n)$. Vertraut man jeder Messung gleich stark, so gilt natürlich $\alpha_i = 1/n$ für alle i .

Für den Erwartungswert von T ergibt eine direkte Rechnung

$$\begin{aligned} E(T) &= E(\alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_n X_n) \\ &= \alpha_1 E(X_1) + \alpha_2 E(X_2) + \dots + \alpha_n E(X_n) \\ &= \alpha_1 \mu + \alpha_2 \mu + \dots + \alpha_n \mu \\ &= (\alpha_1 + \alpha_2 + \dots + \alpha_n) \mu = \left(\sum_{j=1}^n \alpha_j \right) \mu \end{aligned}$$

Wir sehen sofort: T ist genau dann erwartungstreu, wenn $\sum_{j=1}^n \alpha_j = 1$ gilt.

Für die Varianz von T gilt

$$\begin{aligned} Var(T) &= Var(\alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_n X_n) \\ &= \alpha_1^2 Var(X_1) + \alpha_2^2 Var(X_2) + \dots + \alpha_n^2 Var(X_n) \\ &= \alpha_1^2 \sigma^2 + \alpha_2^2 \sigma^2 + \dots + \alpha_n^2 \sigma^2 \\ &= (\alpha_1^2 + \alpha_2^2 + \dots + \alpha_n^2) \sigma^2 = \left(\sum_{j=1}^n \alpha_j^2 \right) \sigma^2 \end{aligned}$$

und

$$\begin{aligned}\lim_{n \rightarrow \infty} \text{Var}(T) &= \sigma^2 \lim_{n \rightarrow \infty} [\alpha_1^2 + \alpha_2^2 + \dots + \alpha_n^2] \\ &= \sigma^2 \left[\left(\lim_{n \rightarrow \infty} \alpha_1 \right)^2 + \left(\lim_{n \rightarrow \infty} \alpha_2 \right)^2 + \dots + \left(\lim_{n \rightarrow \infty} \alpha_n \right)^2 \right]\end{aligned}$$

T ist somit genau dann konsistent, wenn für alle $i = 1, 2, \dots, n$

$$\lim_{n \rightarrow \infty} \alpha_i(n) = 0$$

gilt.

2.2.4 *Mittlerer quadratischer Fehler*

Meist sollte man bei Schätzfunktionen die Erwartungstreue verlangen. Allerdings kann es Probleme geben, wo man die Erwartungstreue durch eine unverhältnismässig grosse Varianz erkaufen muss. Hier kann es sein, dass eine kleine Verzerrung gerne in Kauf genommen wird, falls man dafür die Varianz der Schätzfunktion deutlich reduzieren kann. Wir wollen hier noch ein Qualitätsmerkmal für Schätzfunktionen vorstellen, das sowohl Verzerrung als auch Varianz berücksichtigt.

Definition 2.3 *Der mittlere quadratische Fehler (Mean Squared Error) einer Schätzfunktion $\hat{\theta}$ für einen Parameter θ ist definiert als*

$$MSE(\hat{\theta}) = E((\hat{\theta} - \theta)^2).$$

Mit dem Verschiebungssatz gilt dann

$$MSE(\hat{\theta}) = E((\hat{\theta} - \theta)^2) = \text{Var}(\hat{\theta}) + (E(\hat{\theta}) - \theta)^2 = \text{Var}(\hat{\theta}) + \text{bias}(\hat{\theta})^2.$$

Es kann manchmal auch sinnvoll sein einen Schätzer zu wählen, der einen möglichst kleinen mittleren quadratischen Fehler realisiert.

3 Die Maximum-Likelihood-Methode

Sei X wieder eine Zufallsvariable (Merkmal) deren Verteilung $P(X = x)$ (bzw. deren Verteilungsfunktion F) bis auf einen Parameter θ bekannt sei. Es soll weiterhin bekannt sein, welche möglichen Werte θ annehmen kann, was wir durch $\theta \in \Theta$ abkürzen wollen. Meistens ist Θ eine endliche Menge $\Theta = \{\theta_1, \theta_2, \dots, \theta_k\}$ oder ein (offenes) Teilintervall der reellen Zahlen $\Theta = (a, b)$.

Der Parameter θ soll anhand einer konkreten Stichprobe (x_1, x_2, \dots, x_n) geschätzt werden. Dazu schreiben wir zunächst $P(x_i; \theta) := P(X = x_i)$ für $i = 1, 2, \dots, n$ um daran zu erinnern, dass die Verteilung von X vom Parameter θ abhängt. Die Funktion

$$L(x_1, x_2, \dots, x_n; \theta) := \prod_{i=1}^n P(x_i; \theta) := P(x_1; \theta) \cdot P(x_2; \theta) \cdots P(x_n; \theta)$$

heisst Likelihood-Funktion zur konkreten Stichprobe (x_1, x_2, \dots, x_n) .

Natürlich gilt: Die Wahrscheinlichkeit genau die konkrete Stichprobe (x_1, x_2, \dots, x_n) zu ziehen ist $L(x_1, x_2, \dots, x_n; \theta)$ (als Funktion des unbekanntes Parameters θ).

Ein Parameterwert $\theta_{ML} = \hat{\theta}$ mit

$$L(x_1, x_2, \dots, x_n; \theta_{ML}) \geq L(x_1, x_2, \dots, x_n; \theta)$$

für alle erlaubten $\theta \in \Theta$ heisst Maximum-Likelihood-Schätzer oder einfach ML-Schätzer für den Parameter θ .

Wir suchen also den Wert θ_{ML} , der unsere konkrete (tatsächlich realisierte) Stichprobe (x_1, x_2, \dots, x_n) mit der grössten Wahrscheinlichkeit hervorbringt.

- Ist nun $\Theta = \{\theta_1, \theta_2, \dots, \theta_k\}$ so können wir einen ML-Schätzer für θ leicht finden:
 1. Bestimme $\max\{L(x_1, x_2, \dots, x_n; \theta_1), \dots, L(x_1, x_2, \dots, x_n; \theta_k)\}$
 2. Bestimme ein (nicht notwendigerweise eindeutiges) θ_i , welches dieses Maximum realisiert. Es gilt dann $\theta_{ML} := \theta_i$.
- Ist $\Theta = (a, b)$, so können wir einen ML-Schätzer in vielen Fällen durch das Lösen der

1. Maximum-Likelihood-Gleichung

$$\frac{\partial}{\partial \theta} L(x_1, x_2, \dots, x_n; \theta) = 0$$

2. oder der Ln-Maximum-Likelihood-Gleichung

$$\underbrace{\frac{\partial}{\partial \theta}}_{\text{Ableiten nach } \theta} \ln L(x_1, x_2, \dots, x_n; \theta) = 0$$

Ableiten nach θ

bestimmen. Beide Gleichungen sind im Hinblick auf unser Problem gleichwertig. Die logarithmierte Likelihood-Funktion lässt sich aber einfacher ableiten, da das Produkt in eine Summe transformiert wird.

Beispiel 3.1 p Prozent der Elemente einer Grundgesamtheit haben die Eigenschaft E . Bestimmen Sie an Hand der konkreten Stichproben (E, E^c, E, E, E^c) einen ML-Schätzer für p .

Lösung:

Das Merkmal ist sicher zweipunktverteilt, d.h. unsere Zufallsvariable X kann nur die beiden Werte $1 = E$ und $0 = E^c$ mit den Wahrscheinlichkeiten $P(X = 1) = P(1; p) = p$ und $P(X = 0) = P(0, p) = 1 - p$ annehmen.

Wir nehmen wieder an, dass auch alle unsere fünf Stichprobenvariablen (X_1, \dots, X_5) unabhängig und identisch zweipunktverteilt wie X sind.

Mit welcher Wahrscheinlichkeit zieht man dann genau die obige konkrete Stichprobe? Die Antwort gibt natürlich die Likelihood-Funktion:

$$L(1, 0, 1, 1, 0; p) = P(1; p) \cdot P(0; p) \cdot P(1; p) \cdot P(1; p) \cdot P(0; p) = p^3 \cdot (1 - p)^2.$$

Ln-Maximum-Likelihood-Gleichung:

$$\begin{aligned} 0 &= \frac{\partial}{\partial p} \ln(L(1, 0, 1, 1, 0; p)) \\ &= \frac{\partial}{\partial p} \ln(p^3 \cdot (1 - p)^2) \\ &= \frac{\partial}{\partial p} (3 \cdot \ln(p) + 2 \cdot \ln(1 - p)) \\ &= \frac{3}{p} - \frac{2}{1 - p} \end{aligned}$$

und diese Gleichung hat die eindeutige Lösung $p = 3/5$.

Aufgabe 3.1 *Es gelte $X \sim Po(\lambda)$. Bestimmen Sie einen allgemeinen ML-Schätzer für λ bezüglich der konkreten Stichprobe $(3, 5, 2)$.*

Lösung:

4 Übungsaufgaben

1. Ein Verlag führt eine Leseranlyse unter den $N = 40'000$ Beziehern seiner Zeitschrift „Spannende Statistik“ durch. Eine einfache Zufallsstichprobe vom Umfang $n = 400$ ergab unter anderem Folgendes:
 - (a) 97.82% aller Befragten sind darüber zufrieden, dass eine Befragung durchgeführt wurde.
 - (b) 63.10% aller weiblichen Befragten (Anteil: 37% aller Befragten) schätzen das jährliche Durchschnittseinkommen aller männlichen Bezieher der Zeitschrift auf über 150'000, –. Dabei darf den weiblichen Befragten (natürlich) eine sehr hohe Kompetenz unterstellt werden!
 - (c) Ein Durchschnittseinkommen von 120'000.- jährlich mit einer Standardabweichung von 34'000.-.
 - (d) Weiterhin gab der sich unter den Befragten befindliche (weltweit anerkannte) Wahrsager *Mike Brahma-Shiba* an, dass alle (also 100%) Leser glücklich (also auch zufrieden) sind, auch wenn sie sich dessen teilweise (etwa 5% aller Leser) noch nicht bewusst sind.
 - (e) Einen Anteil der Leser, die mit der Zeitschrift zufrieden sind, von 80%, obwohl die Befragung an einem Montag (!) stattfand. Meiner Meinung nach sollten Befragungen über Zufriedenheit niemals an einem Montag stattfinden, bzw. alternativ muss der erfragte (Montags)Zufriedenheitsgrad hier auf 91.6% angehoben werden.
 - (f) Ein befragter Statistik-Dozent (Name bekannt) zitiert aus einer seiner Veröffentlichungen, in der er (statistisch) beweist, dass der Grad der Zufriedenheit normalverteilt ist. Somit **müssen** genau 50% der Leser mindestens durchschnittlich zufrieden sein.

Schätzen Sie anhand dieser Informationen

- (a) das Durchschnittseinkommen aller $N = 40'000$ Bezieher,
- (b) den Anteil der zufriedenen Leser unter allen Beziehern.

Welches Prinzip haben Sie angewendet? Was besagt dieses Prinzip allgemein?

2. Eine Stadt wird zur Wahl des Bürgermeisters in zwei Wahlbezirke *I* und *II* mit 1'200 bzw. 1'500 Wählern geteilt. Einige Tage vor der Wahl möchte ein Institut den relativen Anteil p für einen bestimmten Kandidaten schätzen. Dazu werden n_1 Wähler im Bezirk *I* und n_2 Wähler im Bezirk *II* befragt, von denen x_1 bzw. x_2 für diesen Kandidaten stimmen. Die Zufallsvariablen $\hat{p}_1 = \frac{X_1}{n_1}$ bzw. $\hat{p}_2 = \frac{X_2}{n_2}$ beschreiben die relativen Anteile für den Kandidaten in den Stichproben.
 - (a) Zeigen Sie, dass die Schätzfunktion $\hat{p} = \frac{1200}{2700}\hat{p}_1 + \frac{1500}{2700}\hat{p}_2$ eine erwartungstreue Schätzung für p ist.
 - (b) Schätzen Sie p speziell mit $\frac{x_1}{n_1} = 65\%$ und $\frac{x_2}{n_2} = 58\%$
 - (c) Wann ist die Schätzung $\hat{p} = \frac{1}{2}(\hat{p}_1 + \hat{p}_2)$ erwartungstreu für p ?

3. Sei X ein Merkmal (einer gewissen Grundgesamtheit) mit (unbekanntem) Erwartungswert μ und Varianz σ^2 . Der Parameter μ soll geschätzt werden. X_1, \dots, X_n seien unabhängige, identisch wie X verteilte Stichproben(zufalls)variablen. Es stehen folgende Schätzfunktionen für μ zur Verfügung:

$$\bullet T_1 = \frac{1}{2}(X_1 + X_2)$$

$$\bullet T_2 = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} X_i + \frac{1}{2} X_n$$

$$\bullet T_3 = \frac{1}{n+1} \sum_{i=1}^n X_i$$

$$\bullet T_4 = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

- (a) Bestimmen Sie die Erwartungswerte der vier Schätzfunktionen.
(b) Bestimmen Sie die Varianzen der vier Schätzfunktionen.
(c) Untersuchen Sie die Schätzfunktionen auf Erwartungstreue und Konsistenz.

4. Es gelte $X \sim B(100; p)$. Bestimmen Sie auf direktem Weg einen allgemeinen ML-Schätzer für p bezüglich der (konkreten) Stichprobe

- $(3, 6, 4)$
- $(2, 4, 5, 4, 4)$
- (x_1, x_2, \dots, x_k) .

Hinweis: Der Weg der Bestimmung des Schätzers ist entscheidend, das Resultat ist überhaupt nicht überraschend und kann für Binomialverteilungen sicher einfacher gefunden werden.

Resultate einiger Übungsaufgaben

1. Nutzen Sie das Schätzprinzip der schliessenden Statistik und ignorieren Sie irrelevante Informationen.

2. (a)

$$E(\hat{p}) = E\left(\frac{1200}{2700} \cdot \hat{p}_1 + \frac{1500}{2700} \cdot \hat{p}_2\right) = \frac{1200}{2700} \cdot \underbrace{E(\hat{p}_1)}_{p_1} + \frac{1500}{2700} \cdot \underbrace{E(\hat{p}_2)}_{p_2} = p$$

(b) $\hat{p} = 61.1\%$

(c) Falls $p_1 = p_2$ gilt, so folgt zunächst:

$$p = \frac{1200}{2700} \cdot p_1 + \frac{1500}{2700} \cdot p_1 = p_1 = p_2$$

und

$$E\left(\frac{1}{2}(\hat{p}_1 + \hat{p}_2)\right) = \frac{1}{2}(E(\hat{p}_1) + E(\hat{p}_2)) = \frac{1}{2} \cdot 2p_1 = p_1 = p$$

Falls $n_1 = n_2$ und $p_1 \neq p_2$ gilt:

$$E\left(\frac{1}{2}(\hat{p}_1 + \hat{p}_2)\right) = E\left(\frac{1}{2}\left(\frac{X_1}{n_1} + \frac{X_2}{n_1}\right)\right) = E\left(\frac{X_1 + X_2}{2n_1}\right) = p$$

3. Direkte Rechnung...

4.

$$\hat{p} = \frac{x_1 + x_2 + \dots + x_k}{n \cdot k}.$$