Stochastik

Barth, Friedrich

München, [20]03

18. 6. Die Stichprobenvarianz

urn:nbn:de:hbz:466:1-83580

Auf Grund der bei den Beweisen verwendeten Sätze 212.1 und 212.2 gilt die Erwartungstreue von $\bar{X}_n$ auch für Stichproben, bei denen die X_i nicht unabhängig sind, und die Konsistenz von $\bar{X}_n$ für Stichproben, bei denen die X_i nur paarweise unabhängig sind.

18.6. Die Stichprobenvarianz

Wir wollen nun eine Schätzgröße für die Varianz σ^2 einer Zufallsgröße X ausfindig machen. Wenn möglich, soll sie erwartungstreu sein. Dazu ziehen wir aus X eine Stichprobe $(X_1 | X_2 | \ldots | X_n)$. Die X_i sind Kopien von X; also ist $\mathscr{E} X_i = \mathscr{E} X = \mu$ und $\sigma^2 = \operatorname{Var} X = \operatorname{Var} X_i = \mathscr{E}[(X_i - \mu)^2]$.

Da wir in **18.5.** unbekannte Erwartungswerte durch das Stichprobenmittel geschätzt haben, ist es naheliegend, als Schätzfunktion für σ^2 das Stichprobenmittel $V := \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2$ der Zufallsgrößen $(X_i - \mu)^2$ zu verwenden. Wie in Aufgabe 384/**4** gezeigt werden soll, ist V eine erwartungstreue Schätzgröße für σ^2. Darüber hinaus ist sie konsistent und effizient.

In den meisten Fällen ist aber auch μ nicht bekannt. Wir schätzen dann μ gemäß **18.5.** durch das Stichprobenmittel $\bar{X}_n$, das wir an Stelle von μ in den Ausdruck für V einführen. Wir erhalten somit als Schätzfunktion für σ^2 die Stichprobenfunktion

$$\tilde{S}^2 := \frac{1}{n}[(X_1 - \bar{X}_n)^2 + \ldots + (X_n - \bar{X}_n)^2] = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 .$$

Zur Vereinfachung der Schreibweise läßt man üblicherweise beim Stichprobenmittel $\bar{X}_n$ den Index n weg und schreibt

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

Um zu prüfen, ob $\tilde{S}^2$ eine erwartungstreue Schätzgröße ist, formen wir die Summe um. Wir gehen dabei genauso vor wie bei der Herleitung des Verschiebungssatzes 207.1:

$$\begin{aligned}
\tilde{S}^2 &= \frac{1}{n} \sum_{i=1}^{n} [(X_i - \mu) - (\bar{X} - \mu)]^2 = \\
&= \frac{1}{n} \sum_{i=1}^{n} [(X_i - \mu)^2 + (\bar{X} - \mu)^2 - 2(\bar{X} - \mu)(X_i - \mu)] = \\
&= \frac{1}{n} \left[\sum_{i=1}^{n} (X_i - \mu)^2 + \sum_{i=1}^{n} (\bar{X} - \mu)^2 - 2(\bar{X} - \mu) \sum_{i=1}^{n} (X_i - \mu) \right] = \\
&= \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2 + \frac{1}{n} \cdot n(\bar{X} - \mu)^2 - \frac{2}{n} \cdot (\bar{X} - \mu)(n\bar{X} - n\mu) = \\
&= \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2 + (\bar{X} - \mu)^2 - 2(\bar{X} - \mu)^2 = \\
&= \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2 - (\bar{X} - \mu)^2 .
\end{aligned}$$

Wir prüfen $\tilde{S}^2$ auf Erwartungstreue:

$$\mathscr{E}(\tilde{S}^2) = \mathscr{E}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 - (\bar{X}-\mu)^2\right] =$$

$$= \frac{1}{n}\cdot\mathscr{E}\left[\sum_{i=1}^{n}(X_i-\mu)^2\right] - \mathscr{E}[(\bar{X}-\mu)^2] =$$

$$= \frac{1}{n}\cdot\sum_{i=1}^{n}\mathscr{E}[(X_i-\mu)^2] - \mathscr{E}[(\bar{X}-\mu)^2] =$$

$$= \frac{1}{n}\sum_{i=1}^{n}\operatorname{Var}X_i - \mathscr{E}[(\bar{X}-\mu)^2] =$$

$$= \frac{1}{n}\sum_{i=1}^{n}\sigma^2 - \mathscr{E}[(\bar{X}-\mu)^2] =$$

$$= \sigma^2 - \mathscr{E}[(\bar{X}-\mu)^2].$$

Da $\mu = \mathscr{E}\bar{X}$, ist der Subtrahend nichts anderes als $\operatorname{Var}\bar{X}$, wofür nach Satz 212.2 gilt $\operatorname{Var}\bar{X} = \frac{1}{n}\operatorname{Var}X = \frac{1}{n}\sigma^2$. Damit wird

$$\mathscr{E}(\tilde{S}^2) = \sigma^2 - \frac{1}{n}\sigma^2 =$$

$$= \frac{n-1}{n}\sigma^2.$$

Wegen des Faktors $\frac{n-1}{n}$ ist $\tilde{S}^2$ keine erwartungstreue Schätzgröße für σ^2. Man würde σ^2 stets zu klein schätzen, und dies um so mehr, je kleiner die Länge n der Stichprobe ist. Der Mangel läßt sich aber für $n > 1$ leicht beheben: Wir multiplizieren $\tilde{S}^2$ mit $\frac{n}{n-1}$ und erhalten eine erwartungstreue Schätzgröße für σ^2, die sogenannte Stichprobenvarianz S^2.
Wir halten fest

Satz 382.1 Ist $(X_1 | X_2 | \ldots | X_n)$ eine Stichprobe aus der Zufallsgröße X mit $\bar{X} := \frac{1}{n}\cdot\sum_{i=1}^{n} X_i$, dann ist für $n \geqq 2$ die **Stichprobenvarianz**

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

eine erwartungstreue Schätzgröße für die Varianz σ^2 der Verteilung von X.

Für große n ist der Unterschied zwischen S^2 und $\tilde{S}^2$ natürlich unerheblich. Sowohl die Stichprobenvarianz S^2 wie auch $\tilde{S}^2$ sind konsistent (vgl. Aufgabe 385/**14**), aber nicht effizient, was wir ohne Beweis mitteilen.
Es liegt nun die Vermutung nahe, man könne die **Stichprobenstreuung** $S := \sqrt{S^2}$

als erwartungstreue Schätzgröße für die Standardabweichung σ benützen. *Dies ist leider falsch,* wie wir an einem Beispiel zeigen:
Es sei $X :=$ »Zahl der Adler beim Wurf einer evtl. unsymmetrischen Münze«. X ist nach $\mathrm{B}(1; p)$ verteilt. $(X_1 | X_2)$ sei eine Stichprobe der Länge 2 aus X. Mit $\bar{X} = \frac{1}{2}(X_1 + X_2)$ erhält man

$$S^2 = (X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 = \tfrac{1}{2}(X_1 - X_2)^2,$$
$$S \;= \tfrac{1}{2}\sqrt{2}\,|X_1 - X_2|.$$

S nimmt offensichtlich nur die Werte 0 und 1 an. Damit errechnet man

$$\mathscr{E}S = \tfrac{1}{2}\sqrt{2}[P(X_1 = 1 \wedge X_2 = 0) + P(X_1 = 0 \wedge X_2 = 1)] =$$
$$= p(1-p)\sqrt{2}.$$

Dagegen ist $\sigma(X) = \sqrt{p(1-p)}$.
Die Verschiedenheit von $\mathscr{E}S$ und σ läßt sich nicht so wie die von $\mathscr{E}\tilde{S}^2$ und σ^2 durch einen für alle p gültigen Faktor beseitigen.

Die Berechnung des Wertes von S^2. Bei mehrfacher Messung einer (z.B. physikalischen) Größe unter gleichen Bedingungen ist es üblich, aus der Meßreihe die Werte $\bar{x}$ und s von $\bar{X}$ bzw. S zu bestimmen und das Meßresultat in der Form

$\bar{x} \pm s$ zu schreiben.

So findet man z.B. für das Wirkungsquantum h die Angabe

$$\frac{h}{2\pi} = (1{,}05443 \pm 0{,}00004) \cdot 10^{-34}\,\mathrm{Js}.$$

s ist meist sehr mühsam zu berechnen, wenn man die in Satz 382.1 enthaltene Definition von S^2 direkt benützt, da $\bar{x}$ im allgemeinen keine glatte Zahl ist. Man hat viele unbequeme Subtraktionen auszuführen. Die folgende Umformung erleichtert die Arbeit. Sie entspricht genau der oben vorgeführten Umformung von $\tilde{S}^2$.

$$(n-1)\cdot s^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 =$$
$$= \sum_{i=1}^{n} (x_i - a)^2 + 2\sum_{i=1}^{n} (x_i - a)(a - \bar{x}) + n(a - \bar{x})^2.$$

Der zweite Term ist gleich

$$2(a - \bar{x})\sum_{i=1}^{n} (x_i - a) = 2(a - \bar{x})(n\bar{x} - na) = -2n(a - \bar{x})^2.$$

Also erhält man schließlich mit $\bar{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$

$$\boxed{s^2 = \frac{1}{n-1}\cdot\sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} (x_i - a)^2 - n(\bar{x} - a)^2\right)}$$

Diese Formel gilt für eine beliebige Zahl a. Man wählt a als glatte Zahl in der Nähe von $\bar{x}$ und kann damit die rechte Seite verhältnismäßig einfach ausrechnen. Ein Beispiel zeigt Tabelle 384.1.

Tab. 384.1 Punktbewertungen x_i von 10 Personen bei einem Gedächtnistest. Die Hilfszahl $a = 35$ wird bereits zur Berechnung von $\bar{x}$ mit Vorteil verwendet. (Vgl. Aufgabe 215/**20**)

$\bar{x} = 35 - \frac{5}{10} = 34{,}5;$
$s^2 = \frac{1}{9}(1473 - 10 \cdot 0{,}5^2);$
$s = 12{,}8.$

x_i	$x_i - a$	$(x_i - a)^2$
12	−23	529
21	−14	196
28	− 7	49
30	− 5	25
34	− 1	1
37	+ 2	4
39	+ 4	16
39	+ 4	16
49	+14	196
56	+21	441
Summe:	− 5	1473

Aufgaben

Zu 18.3. – 18.6.

1. Bestimme bei folgenden Stichproben aus einer Zufallsgröße Schätzwerte für den Erwartungswert und die Varianz:
a) 3; 5; 3; 6; 9. **b)** 0,3; 0,7; −0,4; 0,8; −0,2.
c) 300; 700; −400; 800; −200 **d)** 1; 1; 1; 1; 1.

2. Bestimme Schätzwerte für Erwartungswert und Varianz der Zufallsgröße »Bremsweg bei einer Geschwindigkeit von 50 km/h«, wenn sich folgende Meßwerte in m ergaben:
15,5 14,0 14,1 14,9 13,4 15,0 14,4 14,4 15,8 15,9.

3. Welchen Erwartungswert und welche Varianz erhält man bei n Messungen für das Stichprobenmittel $\bar{X}$, wenn für die Zufallsgröße X gilt:
a) $\mathscr{E}X = 2{,}71;\ \operatorname{Var}X = 1{,}5;\ n = 10.$ **b)** $\mathscr{E}X = 1;\ \operatorname{Var}X = 1;\ n = 100.$
c) $\mathscr{E}X = 1;\ \operatorname{Var}X = 1;\ n = 1000.$ **d)** $\mathscr{E}X = 0;\ \operatorname{Var}X = 1;\ n = 1000.$

4. Zeige, daß $V := \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$ eine erwartungstreue Schätzfunktion für σ_X^2 ist.

5. a) Eine Zufallsgröße X ist binomial nach $B(1; 0{,}25)$ verteilt. Bestimme Erwartungswert, Varianz und Standardabweichung von X. Berechne für eine Stichprobe der Länge 3 die Wahrscheinlichkeitsverteilungen und die Erwartungswerte des Stichprobenmittels $\bar{X}_3$, der Stichprobenvarianz S^2 und der Stichprobenstreuung S.
b) Löse **a)** für eine Stichprobe der Länge 2, wenn X nach $B(2; 0{,}25)$ verteilt ist.

6. Eine Zufallsgröße X besitze die Verteilung $P(X = x_j) =: p_j$ für $j = 1, \ldots, s$. $(X_1|X_2|\ldots|X_n)$ sei eine Stichprobe der Länge n, ferner sei $N_j :=$ Anzahl der X_i, welche den Wert x_j annehmen. A_j bedeute das Ereignis »$\left|\frac{1}{n}N_j - p_j\right| \geqq a$«.
a) Formuliere für $P(A_j)$ die *Tschebyschow*-Ungleichung.
b) Formuliere mit den A_j das Ereignis $A :=$ »Die Stichprobenverteilung weicht nirgends um a oder mehr von der Wahrscheinlichkeitsverteilung ab«.
•**c)** Finde für $P(\bar{A})$ eine Abschätzung, die zeigt, daß das Ereignis A für genügend große n eine beliebig nahe an 1 gelegene Wahrscheinlichkeit hat. *Die Stichprobenverteilung konvergiert »in Wahrscheinlichkeit« gegen die Wahrscheinlichkeitsverteilung.*

7. Eine Zufallsgröße sei nach $B(100; 0{,}4)$ verteilt. Wie lang muß eine Stichprobe mindestens sein, damit die Varianz des Stichprobenmittels kleiner als 0,01 ist?

8. Die Wahrscheinlichkeit für eine Knabengeburt ist 0,514. Die Zufallsgröße X nehme den Wert 1 an, wenn ein Knabe geboren wird, sonst habe sie den Wert 0.