

## Schätzung des Erwartungswerts

Um den unbekanntem Erwartungswert einer Zufallsvariablen zu schätzen, führen wir das zugehörige Zufallsexperiment  $n$ -mal durch und bilden den Mittelwert  $\bar{X}$  der Stichprobenwerte. Es soll gezeigt werden, dass gilt:

$$E(\bar{X}) = E(X) \quad \text{und}$$

$$V(\bar{X}) = \frac{1}{n} V(X), \quad \text{d.h.} \quad \sigma_{\bar{X}} = \frac{1}{\sqrt{n}} \sigma_X$$

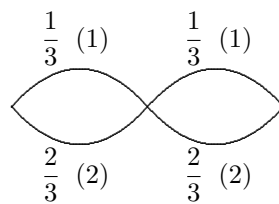
Die mittlere Abweichung von  $\bar{X}$  zum Erwartungswert ist also nur  $\frac{1}{\sqrt{n}}$ -mal so groß wie die Standardabweichung von  $X$ . Dies kennzeichnet die Güte der Schätzung.

Die Zufallsvariable  $X$  mit der Verteilung:

$k$	1	2
$P(X = k)$	$\frac{1}{3}$	$\frac{2}{3}$

hat den Erwartungswert  $E(X) = \frac{5}{3}$  und die Varianz  $V(X) = \frac{2}{9}$ .

Für die zweimalige unabhängige Durchführung des Zufallsexperiments ermitteln wir  $E(\bar{X})$  und  $V(\bar{X})$



<i>Pfade</i>	(1   1)	(1   2)	(2   1)	(2   2)
<i>Wahrscheinlichkeit</i>	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{2}{9}$	$\frac{4}{9}$
$\bar{X} = \frac{X_1 + X_2}{2}$	1	$\frac{3}{2}$	$\frac{3}{2}$	2

und erhalten das Erwartete:  $E(\bar{X}) = \frac{5}{3}$ ,  $V(\bar{X}) = \frac{1}{9}$ .

$X_1, X_2$  sind hierbei unabhängige (d. h. nur von der Stelle abhängige) Zufallsvariable, definiert auf der Menge aller Wertepaare.  $X_1, X_2$  haben dieselbe Verteilung wie  $X$ .

Insbesondere ist  $E(X_1) = E(X_2) = E(X)$ .

*Begründungen für  $n = 2$  (für beliebige  $n$  entsprechend):*

$$E(\bar{X}) = E\left(\frac{X_1 + X_2}{2}\right) = \frac{E(X_1) + E(X_2)}{2} = \frac{E(X) + E(X)}{2} = E(X)$$

Leicht zu zeigen ist:  $V(aY) = a^2 V(Y)$

$$V(\bar{X}) = V\left(\frac{X_1 + X_2}{2}\right) = \frac{V(X_1) + V(X_2)}{4} = \frac{V(X) + V(X)}{4} = \frac{V(X)}{2}$$

Es fehlt noch der Nachweis von:

$$V(X_1 + X_2) = V(X_1) + V(X_2)$$

## Schätzung des Erwartungswerts, Fortsetzung

Wir verwenden (siehe Varianz der Binomialverteilung):  $V(Y) = E(Y^2) - (E(Y))^2$ .

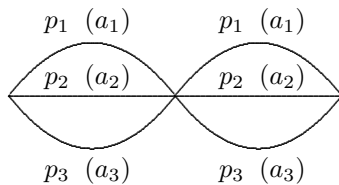
$$\begin{aligned}
 V(X_1 + X_2) &= E((X_1 + X_2)^2) - (E(X_1 + X_2))^2 \\
 &= E(X_1^2) + 2E(X_1 \cdot X_2) + E(X_2^2) - (2E(X))^2 \\
 &= \underbrace{E(X^2) - (E(X))^2}_{V(X_1)} + \underbrace{E(X^2) - (E(X))^2}_{V(X_2)}, \quad \text{weil } 2E(X_1 \cdot X_2) - 2(E(X))^2 = 0 \\
 &= V(X_1) + V(X_2)
 \end{aligned}$$

Zum Schluss fehlt noch die Begründung der verwendeten (allgemein für unabhängige Zufallsvariablen gültigen) Beziehung:

$$E(X_1 \cdot X_2) = E(X_1) \cdot E(X_2)$$

Sei hierzu die Verteilung von  $X$  gegeben durch

$k$	$a_1$	$a_2$	$a_3$
$P(X = k)$	$p_1$	$p_2$	$p_3$



$X_1, X_2$  haben wieder die gewohnte Bedeutung.

Mit  $E(X) = a_1p_1 + a_2p_2 + a_3p_3$  ergibt sich:

$$\begin{aligned}
 E(X_1) \cdot E(X_2) &= (a_1p_1 + a_2p_2 + a_3p_3) \cdot (a_1p_1 + a_2p_2 + a_3p_3) \\
 &= \text{Klammern auflösen} \\
 E(X_1 \cdot X_2) &= \text{Hier stehen dieselben Summanden.}
 \end{aligned}$$

$X_1 \cdot X_2$  nimmt z.B. das Produkt  $a_1a_3$  mit der Wahrscheinlichkeit  $p_1p_3$  an.

Nun ist zu erkennen, woher die Übereinstimmung rührt.

## Schätzung der Varianz

Die Varianz ist der Erwartungswert der Zufallsvariablen  $(X - \mu)^2$ , wobei  $E(X) = \mu$  ist.

Es liegt daher nahe, die Varianz mit dem Mittelwert

$$S^* = \frac{1}{n} [(X_1 - \mu)^2 + \dots + (X_n - \mu)^2]$$

zu schätzen.

Jedoch ist im Allgemeinen der Erwartungswert  $\mu$  nicht bekannt. Daher wird  $\mu$  durch  $\bar{X}$  ersetzt.

Wir vermuten für

$$S^{**} = \frac{1}{n} [(X_1 - \bar{X})^2 + \dots + (X_n - \bar{X})^2]$$

$E(S^{**}) = V(X)$ .

Für das Beispiel am Anfang ist  $E(S^{**}) = \frac{1}{9}$ , jedoch  $V(X) = \frac{2}{9}$ .

Formulierung: Die Schätzgröße  $S^{**}$  ist nicht erwartungstreu.

Nun ist es erfreulicherweise möglich  $S^{**}$  geringfügig abzuändern, damit ihr Erwartungswert  $V(X)$  ist.

Um die Änderung zu erkennen, muss  $E(S^{**})$  allgemein ausgerechnet werden. Das ist etwas aufwändig.

$$\begin{aligned} S^{**} &= \frac{1}{n} [(X_1 - \bar{X})^2 + \dots + (X_n - \bar{X})^2] \\ &= \frac{1}{n} [(X_1^2 - 2X_1\bar{X} + \bar{X}^2) + \dots + (X_n^2 - 2X_n\bar{X} + \bar{X}^2)] \\ &= \frac{1}{n} [X_1^2 + \dots + X_n^2 - \underbrace{2(X_1 + \dots + X_n) \cdot \bar{X}}_{n \cdot \bar{X}} + n \cdot \bar{X}^2] \\ &= \frac{1}{n} [X_1^2 + \dots + X_n^2 - n \cdot \bar{X}^2] \end{aligned}$$

$$\begin{aligned} E(S^{**}) &= \sigma^2 + \mu^2 - E(\bar{X}^2) \quad \text{beachte: } \sigma^2 = E(X^2) - \mu^2 \implies E(X^2) = \sigma^2 + \mu^2 \\ &= \sigma^2 + \mu^2 - \frac{1}{n^2} E(X_1^2 + \dots + X_n^2 + \underbrace{2X_1X_2 + \dots + 2X_{n-1}X_n}_{\binom{n}{2} = \frac{n(n-1)}{1 \cdot 2}}) \quad \text{Summanden} \\ &= \sigma^2 + \mu^2 - \frac{1}{n^2} [n \cdot (\sigma^2 + \mu^2) + \frac{n(n-1)}{2} \cdot 2\mu^2] \quad \text{beachte: } E(X_iX_k) = \mu^2 \quad (i \neq k) \\ &= \frac{n-1}{n} (\sigma^2 + \mu^2) - \frac{n-1}{n} \mu^2 \\ &= \frac{n-1}{n} \sigma^2 \end{aligned}$$

Der aufgedeckte Mangel lässt sich für  $n > 1$  leicht beheben. Für große  $n$  kann er vernachlässigt werden.

Wir multiplizieren  $S^{**}$  mit  $\frac{n}{n-1}$  und erhalten für die Varianz die erwartungstreue Schätzgröße:

$$S^2 = \frac{1}{n-1} [(X_1 - \bar{X})^2 + \dots + (X_n - \bar{X})^2] \quad \text{mit}$$

$$\bar{X} = \frac{1}{n} (X_1 + \dots + X_n)$$