Die Streuung der Stichprobenmittelwerte

STREUUNG VON PARAMETERSCHÄTZERN

Neben Erwartungstreue und Konsistenz wünscht man sich von einer guten Schätzfunktion auch, dass ihre Streuung möglichst gering ist. Im folgenden Beispiel soll der Parameter α geschätzt werden (der wahre Wert von α sei hier 200). Es stehen zwei Schätzfunktionen zur Verfügung: $\hat{\lambda}$ und $\hat{\varphi}$. $\hat{\lambda}$ sei erwartungstreu und konsistent, $\hat{\varphi}$ sei ebenfalls konsistent, aber nicht erwartungstreu (leichter negativer Bias). Betrachtet man nun die Schätzwerte...

Beispiel

i	$\hat{\lambda}_i$	$\hat{\varphi}_i$
1	112	178
2	223	202
3	166	197
4	198	189
5	308	192
6	211	199
7	244	183
8	148	200
9	127	191
10	265	189
$\sum_i$	2002	1920

... so sieht man Folgendes: Obwohl $\hat{\lambda}$ durchschnittlich deutlich besser schätzt ($\bar{\hat{\lambda}} = 200.2$; $\bar{\hat{\varphi}} = 192.0$), sind die meisten $\hat{\lambda}_i$ sehr weit vom wahren Wert entfernt, wogegen $\hat{\varphi}_i$ großteils passable Schätzungen ergibt. Der Grund dafür ist, dass die $\hat{\lambda}$-Werte zu stark streuen.

Falls man α also aufgrund einer einzelnen Stichprobe schätzen müsste, so wäre man wohl mit dem Schätzer $\hat{\varphi}$ besser beraten, obwohl dieser nicht erwartungstreu ist. *(Alternative: Da $\hat{\lambda}$ konsistent ist, könnte man auch versuchen, die Genauigkeit zu erhöhen, indem man den Stichprobenumfang vergrößert, was in der Praxis aber oft schwierig oder gar nicht durchführbar ist.)*

VARIANZ DER STICHPROBENMITTELWERTE

Wie stark streuen die Stichprobenmittelwerte $\bar{x}$ um den wahren Mittelwert μ_x?

$$\begin{aligned}\sigma^2_{\bar{X}} &= E[(\bar{X}-\mu_X)^2] = \\ &= E(\bar{X}^2) - 2\cdot\mu_X\cdot E(\bar{X}) + E(\mu_X^2) =\end{aligned}$$

Wegen $E(\bar{X}) = \mu_X$ und $E(\mu_X^2) = \mu_X^2$ gilt:

$$\begin{aligned}&= E(\bar{X}^2) - 2\cdot\mu_X^2 + \mu_X^2 \\ \sigma^2_{\bar{X}} &= E(\bar{X}^2) - \mu_X^2\end{aligned}$$

Wir wissen *(siehe Punkt 1.3. Bias und Biaskorrektur: Bias der Stichprobenvarianz, Einschub 2)*:

$$E(\bar{X}^2) = \frac{1}{n}\cdot E(X^2) + \frac{n-1}{n}\cdot\mu_X^2$$

Eingesetzt in obige Gleichung erhalten wir:

$$\begin{aligned}\sigma^2_{\bar{X}} &= \frac{1}{n}\cdot E(X^2) + \frac{n-1}{n}\cdot\mu_X^2 - \mu_X^2 = \\ &= \frac{1}{n}\cdot E(X^2) + \frac{n-1}{n}\cdot\mu_X^2 - \frac{n}{n}\cdot\mu_X^2 = \\ &= \frac{1}{n}\cdot E(X^2) - \frac{1}{n}\cdot\mu_X^2 = \\ &= \frac{1}{n}\cdot[E(X^2) - \mu_X^2] = \\ &= \frac{1}{n}\cdot\sigma_X^2 \\ \sigma^2_{\bar{X}} &= \frac{\sigma_X^2}{n}\end{aligned}$$

In Worten: *Die Varianz der Stichprobenmittelwerte erhält man, indem man die Populationsvarianz durch die Stichprobengröße dividiert.*

Man erkennt hier auch, dass $\bar{x}$ ein konsistenter Schätzer ist: Je größer die Stichproben sind, umso genauer ist der Schätzer und umgekehrt.