Beispiel Shrinkage

Betrachte das Random-Intercept-Modell ohne Kovariablen:

$$y_{ij} = b_i + \varepsilon_{ij}, \quad b_i \overset{iid}{\sim} N(0, \tau^2),\ \varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2),$$
$$j = 1, \dots, n_i, i = 1, \dots, N.$$

1 KQ-Schätzer für die b_i

Behandeln wir die b_i wie feste Effekte β_i, so ist das Modell

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{\varepsilon}$$

mit festen Effekten $\boldsymbol{b}$,

$$\boldsymbol{X} = \begin{pmatrix} 1 & & \\ \vdots & & \\ 1 & & \\ & \ddots & \\ & & 1 \\ & & \vdots \\ & & 1 \end{pmatrix}, \qquad \mathrm{Cov}(\boldsymbol{\varepsilon}) = \boldsymbol{R} = \sigma^2 \boldsymbol{I}_n.$$

Es ergibt sich der KQ-Schätzer

$$\hat{\boldsymbol{b}}^{KQ} = (\boldsymbol{X'}\boldsymbol{R}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X'}\boldsymbol{R}^{-1}\boldsymbol{y} = (\frac{1}{\sigma^2}\,\mathrm{diag}(n_i))^{-1}\frac{1}{\sigma^2}\begin{pmatrix} n_1\bar{y}_1 \\ \vdots \\ n_N\bar{y}_N \end{pmatrix} = \begin{pmatrix} \bar{y}_1 \\ \vdots \\ \bar{y}_N \end{pmatrix}$$

mit $\bar{y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}$, also $\hat{b}_i^{KQ} = \bar{y}_i$. Damit ist

- $\mathrm{E}(\hat{b}_i^{KQ}) = \frac{1}{n_i}\sum_{j=1}^{n_i} b_i = b_i$ (kein Bias)
- $\mathrm{Var}(\hat{b}_i^{KQ}) = \frac{1}{n_i^2}\sum_{j=1}^{n_i}\sigma^2 = \frac{1}{n_i}\sigma^2$
- $\mathrm{MSE}(\hat{b}_i^{KQ}) = E[(\hat{b}_i^{KQ} - b_i)^2] = \mathrm{Bias}(\hat{b}_i^{KQ})^2 + \mathrm{Var}(\hat{b}_i^{KQ}) = 0 + \frac{1}{n_i}\sigma^2 = \frac{1}{n_i}\sigma^2.$

2 BLUPs für die b_i

Behandeln wir die b_i als zufällige Effekte mit Verteilungsannahme $b_i \overset{iid}{\sim} N(0, \tau^2)$, so ist das Modell

$$\boldsymbol{y} = \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{\varepsilon}$$

mit zufälligen Effekten $\boldsymbol{b}$, $\boldsymbol{Z}$ wie $\boldsymbol{X}$ oben und $\boldsymbol{G} = \tau^2 \boldsymbol{I}_N$. Es ergibt sich der BLUP für $\boldsymbol{b}$ als

$$\hat{\boldsymbol{b}}^{BLUP} = (\boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{G}^{-1})^{-1}\boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{y} = (\frac{1}{\sigma^2}\operatorname{diag}(n_i) + \frac{1}{\tau^2}\boldsymbol{I}_N)^{-1}\frac{1}{\sigma^2}\begin{pmatrix} n_1\bar{y}_1 \\ \vdots \\ n_N\bar{y}_N \end{pmatrix} = \operatorname{diag}(\frac{n_i}{n_i + \sigma^2/\tau^2})\begin{pmatrix} \bar{y}_1 \\ \vdots \\ \bar{y}_N \end{pmatrix},$$

also $\hat{b}_i^{BLUP} = \frac{n_i}{n_i+\sigma^2/\tau^2}\bar{y}_i = c_i\bar{y}_i$ mit $0 < c_i := \frac{n_i}{n_i+\sigma^2/\tau^2} < 1$ (Shrinkage). Damit ist

- $\mathrm{E}(\hat{b}_i^{BLUP}|b_i) = c_i \frac{1}{n_i}\sum_{j=1}^{n_i} b_i = c_i b_i \neq b_i$ mit $|c_i b_i| < |b_i|$ (Bias durch Shrinkage)
  Die BLUPs streuen also weniger als die KQ-Schätzer.
- $\mathrm{Var}(\hat{b}_i^{BLUP}|b_i) = c_i^2 \frac{1}{n_i^2}\sum_{j=1}^{n_i}\sigma^2 = c_i^2\frac{1}{n_i}\sigma^2 < \frac{1}{n_i}\sigma^2$. Die Varianz ist also kleiner als beim KQ-Schätzer.
- Der MSE betrachtet einen Kompromiss aus Varianz und Bias:

$$\begin{aligned}
\mathrm{MSE}(\hat{b}_i^{BLUP}) &= \mathrm{E}[(\hat{b}_i^{BLUP} - b_i)^2] \\
&= \mathrm{E}_b[\mathrm{Var}(\hat{b}_i^{BLUP}|b_i) + (E(\hat{b}_i^{BLUP}|b_i) - b_i)^2] \\
&= c_i^2\frac{1}{n_i}\sigma^2 + \mathrm{E}_b[(1-c_i)^2 b_i^2] = c_i^2\frac{1}{n_i}\sigma^2 + (1-c_i)^2\tau^2 \\
&= \frac{n_i\sigma^2}{(n_i+\sigma^2/\tau^2)^2} + \frac{(\sigma^2/\tau^2)^2\tau^2}{(n_i+\sigma^2/\tau^2)^2} \\
&= \frac{\sigma^2(n_i+\sigma^2/\tau^2)}{(n_i+\sigma^2/\tau^2)^2} = \frac{\sigma^2}{n_i+\sigma^2/\tau^2} < \frac{1}{n_i}\sigma^2,
\end{aligned}$$

  da $\sigma^2/\tau^2 > 0$.

  Im MSE wird also der Bias des BLUP durch die kleinere Varianz mehr als aufgehoben.

3 Zusammenfassung

Für den BLUP $\hat{\boldsymbol{b}}^{BLUP}$ und den KQ-Schätzer $\hat{\boldsymbol{b}}^{KQ}$, der $\boldsymbol{b}$ als feste Effekte schätzt, gilt

	$\hat{\boldsymbol{b}}^{BLUP}$	$\hat{\boldsymbol{b}}^{KQ}$	Kommentar zu $\hat{\boldsymbol{b}}^{BLUP}$
$\hat{b_i}$	$\frac{n_i}{n_i+\sigma^2/\tau^2}\bar{y}_i$	$\bar{y}_i$	Shrinkage
$\mathrm{E}(\hat{b_i}\|b_i)$	$\frac{n_i}{n_i+\sigma^2/\tau^2}b_i$	b_i	Bias
$\mathrm{Var}(\hat{b_i}\|b_i)$	$(\frac{n_i}{n_i+\sigma^2/\tau^2})^2\frac{1}{n_i}\sigma^2$	$\frac{1}{n_i}\sigma^2$	Varianz kleiner
$\mathrm{MSE}(\hat{b_i})$	$\frac{\sigma^2}{n_i+\sigma^2/\tau^2}$	$\frac{\sigma^2}{n_i}$	MSE kleiner trotz Bias

Dabei ist der Unterschied zwischen den beiden Schätzern größer, je kleiner n_i ist und je größer σ^2/τ^2. Für diese Fälle ist die Information in den y_{ij} pro i am geringsten und man gewinnt am meisten durch den Shrinkage-Effekt. Der KQ-Schätzer ($c_i \to 1$) ergibt sich als Grenzfall des BLUPs für $\tau^2 \to \infty$.