Statistik IV für Nebenfachstudierende

4.4. Random Utilities

(alternative Motivation mehrkategorieller Regression)

Prof. Dr. Andreas Mayr

Institut für Statistik, LMU München

Sommersemester 2017

Wie kommt man zum Multinomialen Modell?

Zu jeder Alternative $1, ..., k$ existiere eine *latente* Variable

Random Utility (Zufallsnutzen Funktion) von der Form:

Link zwischen latenten U (Nutzen, Attraktivität) und beobachteten Klassen $Y \in \{1, ...k\}$ erfolgt nach Prinzip des maximalen Nutzens:

Daraus folgt:

$$\mathbb{P}(Y = r) = \mathbb{P}(u_r \geq u_1, u_r \geq u_2, ...)$$

Wenn ε einer Gumbel-Verteilung folgt mit

$$F(x) = \exp(-\exp(-x))$$

erhält man die explizite Form

$$\mathbb{P}(Y = r) = \frac{\exp(u_r)}{\sum_{j=1}^{k} \exp(u_j)}$$

Nun als Modell

$$u_r = \gamma_{r0} + \underline{x}^\top \gamma_r$$

$$\Rightarrow \mathbb{P}(Y = r) = \frac{\exp(\gamma_{r0} + \underline{x}^\top \gamma_r)}{\sum_{j=1}^{k} \exp(\gamma_{j0} + \underline{x}^\top \gamma_j)}$$

Eingesetzt ins Logit-Modell

$$
\begin{aligned}
\log \frac{\mathbb{P}(Y=r)}{\mathbb{P}(Y=k)} &= \log\left(\frac{\exp(\gamma_{r0}+\underline{x}^\top\gamma_r)}{\sum_{j=1}^k \exp(\gamma_{j0}+\underline{x}^\top\gamma_j)}\cdot\frac{\sum_{j=1}^k \exp(\gamma_{j0}+\underline{x}^\top\gamma_j)}{\exp(\gamma_{k0}+\underline{x}^\top\gamma_k)}\right)\\
&= \log\left(\frac{\exp(\gamma_{r0}+\underline{x}^\top\gamma_r)}{\exp(\gamma_{k0}+\underline{x}^\top\gamma_k)}\right)\\
&= \gamma_{r0}+\underline{x}^\top\gamma_r-\gamma_{k0}+\underline{x}^\top\gamma_k\\
&= (\gamma_{r0}-\gamma_{k0})+\underline{x}^\top(\gamma_r-\gamma_k)
\end{aligned}
$$