Statistik IV für Nebenfachstudierende

4.5. Ordinale Regression

Prof. Dr. Andreas Mayr

Institut für Statistik, LMU München

Sommersemester 2017

Ordinaler Response

$$\underline{x} \longrightarrow Y \in \{1, ..., k\}$$

bilden jetzt geordnete Kategorien.

Beispiel:

Motivation

... über latente random utilities:

$$U = \underline{x}^\top \underline{\gamma} + \varepsilon$$

Beobachtet wird eine kategorisierte ("vergröberte")

Version:

Motivation

Graphik:

Kumulatives Modell

Daraus folgt: $\mathbb{P}(Y \leq r) =$

Kumulatives Modell

$$\mathbb{P}(Y \leq r \mid \underline{x}) = F(\gamma_{r0} + \underline{x}^\top \gamma)$$

- Vorteil gegenüber multinomialen Modell:
  - Globale Koeffizienten über alle Kategorien für einzelne Einflussgrößen.
  - Nur der Intercept hängt von der jeweiligen Kategorie ab und ist geordnet.
- Begründung: Im Hintergrund metrisches Regressionsmodell für $U = \underline{x}^\top \underline{\beta}$.

Kumulatives Logit Modell

$$\mathbb{P}(Y \leq r \mid \underline{x}) = \frac{\exp(\gamma_{r0} + \underline{x}^\top \gamma)}{1 + \exp(\gamma_{r0} + \underline{x}^\top \gamma)}$$

$\Rightarrow$ $F(\cdot)$ ist logistische Verteilungsfunktion

$$\log \frac{\mathbb{P}(Y \leq r \mid \underline{x})}{\mathbb{P}(Y > r \mid \underline{x})} = \gamma_{r0} + \underline{x}^\top \gamma$$

Interpretation:

- γ wirken auf Dichotomisierung $\{Y \leq r\}, \{Y > r\}$ wie im binären Logit Modell.
- Betrachte zwei Populationen bestimmt durch $\underline{x}_1$ und $\underline{x}_2$:

$$\frac{\mathbb{P}(Y \leq r \mid \underline{x}_1)/\mathbb{P}(Y > r \mid \underline{x}_1)}{\mathbb{P}(Y \leq r \mid \underline{x}_2)/\mathbb{P}(Y > r \mid \underline{x}_2)} = \exp\left((\underline{x}_1^\top - \underline{x}_2^\top)\gamma\right)$$

Sequentielles Modell: Modellierung als Prozess

(1) Entscheidung zwischen 1 und Kategorien $2, ..., k$ durch:

$\mathbb{P}(Y = 1|\underline{x}) = F(\gamma_{10} + \underline{x}^\top \gamma_1)$

Falls $Y = 1$: Prozess beendet...

(2) Entscheidung zwischen 2 und Kategorien $3, ..., k$ durch:

$\mathbb{P}(Y = 2|Y \geq 2) = F(\gamma_{20} + \underline{x}^\top \gamma_2)$

Falls $Y = 2$: Prozess beendet...

$\vdots$

(r) Entscheidung zwischen r und Kategorien $r + 1, ..., k$ durch:

$\mathbb{P}(Y = r|Y \geq r) = F(\gamma_{r0} + \underline{x}^\top \gamma_r)$

Falls $Y = r$: Prozess beendet...

Sequentielles Modell

$$\mathbb{P}(Y = r \mid Y \geq r,\ \underline{x}) = F(\gamma_{r0} + \underline{x}^\top \gamma_r)$$

- Basiert auf multiplen binären Entscheidungen.
  - Prinzipiell können γ_r Kategorie-spezifisch sein.
  - Oft nimmt man jedoch $\gamma_r = \gamma$, $r = 1, ..., k-1$ an.
  - Häufigste Wahl für $F(\cdot)$?
  - $\Rightarrow$

Logistisches Sequentielle Modell

$$\mathbb{P}(Y = r \mid Y \geq r,\ \underline{x}) = \frac{\exp(\gamma_{r0} + \underline{x}^\top \gamma)}{1 + \exp(\gamma_{r0} + \underline{x}^\top \gamma)}$$

$\Rightarrow$ $F(\cdot)$ ist logistische Verteilungsfunktion

$$\log \frac{\mathbb{P}(Y = r \mid Y \geq r,\ \underline{x})}{1 - \mathbb{P}(Y = r \mid Y \geq r,\ \underline{x})} = \gamma_{r0} + \underline{x}^\top \gamma$$

$$\log \frac{\mathbb{P}(Y = r \mid Y \geq r,\ \underline{x})}{1 - \mathbb{P}(Y = r \mid Y \geq r,\ \underline{x})} = \gamma_{r0} + \underline{x}^\top \gamma$$

$\Rightarrow$ **continuation ratio logit**