LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Raschmodelle und generalisierte Regression

Sven Hilbert

Bestandteile generalisierter linearer Modelle

- Zufällige Komponente Y mit zugehöriger Wahrscheinlichkeitsverteilung
  - Mit unabhängigen Beobachtungen $(y_1, \ldots, y_N)$
- Sytematische Komponente (erklärende Größe) θ
  - Benutzt in linearer Vorhersagefunktion
- Linkfunktion spezifiziert die Funktion von $E(Y)$, die das Modell mit der systematischen Komponente verbindet

Generalisiertes lineares Modell (2)

Allgemeine Form Wahrscheinlichkeitsfunktion von Y (Exponentialfamilie):

$$f(y_i;\theta_i) = a(\theta_i)b(y_i)\exp[y_i Q(\theta_i)]$$

- $Q(\theta_i)$ wird natürlicher Parameter genannt
- Systematische Komponente verbindet einen Vektor $(\eta_1, \ldots, \eta_N)$ mit den erklärenden Variablen durch ein lineares Modell mit Prädiktoren j ($j = 1, 2, \ldots, p$):

$$\eta_i = \sum_{j=1}^{p} \beta_j x_{ij}$$

- Die Linearkombination der erklärenden Variablen wird linearer Prädiktor genannt

Generalisiertes lineares Modell (3)

Linkfunktion: $g(\mu_i) = \eta_i$

- $\mu_i = E(Y_i)$ mit $i = 1, \ldots, N$
- Verbindet zufälligen mit systematischen Komponenten

Identitätslink : $g(\mu) = \mu$

Kanonische (natürliche) Linkfunktion:

$$Q(\theta_i) = \sum_{j=1}^{p} \beta_j x_{ij} = \eta_i = g(\mu_i)$$

- Transformiert μ in den natürlichen Parameter η
  - Abhängig von Verteilung der y_i

Wahrscheinlichkeitsfunktion der Bernoulliverteilung:

$$f(y;\pi) = \pi^{y}(1-\pi)^{1-y}$$

- $P(Y = 1) = \pi$ und $P(Y = 0) = 1 - \pi$ mit $E(Y) = \pi$
- Kanonischer Link: $Q(\pi) = \ln\left(\frac{\pi}{1-\pi}\right)$
- Für Logistische Regression folgt: $\pi(x) = \frac{\exp(\eta)}{1+\exp(\eta)}$
  - Hierbei $P(Y = 1) = \pi(x)$
  - Gilt für Bernoulli- und Binomialverteilung

Zweidimensionale Kontingenztafel

- Testung Unabhängigkeitshypothese
  - Systematische kategoriale Quanitität
  - Multiple binäre Quantitäten

Eisaffinität abhängig von Studiengang?

- Drei Eissorten
  - Binärer Response (ja / nein)
- Zwei Studiengänge

2-dimens.	Eissorten
Studien-gänge	

2x3 Kontingenztafel

- Studiengang systematische Variable
  - 2 Level
  - Kreuzklassifiziert mit Eissorte
- Jeder Student kann 0 – 3 Sorten kaufen
- Randsummen Reihen Eissorten in Tabelle ≠ Anzahl Studenten

	Eissorte			
Studiengang	Schokolade (A)	Kirsche (B)	Gummibär (C)	Anzahl Studenten
Statistik	32	65	86	160
Psychologie	87	148	157	220
Total	119	213	243	380

Abhängigkeit prüfen

Spalten schließen sich nicht gegenseitig aus

- Pearson χ^2-Test kann nicht angewandt werden

Durch Kreuzklassifikation kann 2 x 2 x 2 Tabelle für beide Studiengänge erstellt werden

- Jeweils $2^3 = 8$ Zellen
  - Möglichkeiten:
    000; 100; 010; … ; 111
    - 0 = „nein“, 1 = „ja“

- Jeweils 6 Randsummen (3 Sorten Eis **x** „ja“ / „nein“)

Randsummentabelle

Tabelle enthält also eigentlich Randsummen

- „ja“-Responses für beide Schultypen

Unabhängigkeitshypothese bezieht sich auf Randsummen

- Marginal Independence

	Eissorte			
Studiengang	Schokolade (A)	Kirsche (B)	Gummibär (C)	Anzahl Studenten
Statistik	32	65	86	160
Psychologie	87	148	157	220
Total	119	213	243	380

Marginales Logit Modell

Zufällige Variable: Eissorte

- Level: $i = 1, 2, 3$

Systematische Variable: Studiengang

- Level: $j = 1, 2$

Annahme: Alle Antwortmuster unabhängig multinomial verteilt

- ➢ Gemeinsame Verteilung der Häufigkeiten der Antwortmuster ist produkt-multinomial verteilt

Parameter und Modelle

π_{ij} bezeichnet Randwahrscheinlichkeit

- Person aus Studiengang *j* kauft Sorte *i*
- Parameter der Binomialverteilung, welcher Häufigkeit modelliert

Marginales Logit Modell:

Unabhängig (Modell 1):

$$\beta_i = \ln\left(\frac{\pi_i}{1-\pi_i}\right)$$

Abhängig (Modell 2):

$$\beta_{ij} = \ln\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right)$$

Zwei grundlegende Methoden für marginales Modell

1. Likelihood basiert
   - Restriktion: Randwahrscheinlichkeiten genügen marginalem Modell
   - Produkt-multinomiale Likelihood Funktion wird maximiert
     - Erlaubt wenn alle möglichen Responsemuster unabhängig multinomialverteilt sind
   - ➢ Allerdings sind In den vorgestellten Modellen sind Responsemuster abhängig

2. Quasi-Likelihood basiert
   - Keine Annahmen zur Verteilung nötig
   - Nur erste beide Momente werden genutzt (Mean-Variance Relationship)
   - Bei Annahme einer aus der Exponentialfamilie bekannten Mean-Variance-Relatioship äquivalent zur vollständigen Likelihood

Vergleich Likelihood und Quasi-Likelihood

Parameterschätzung bei ML:

$$\frac{\partial \ell_i}{\partial \beta_j} = \frac{(y_i - \mu_i) x_{ij}}{var(\mathrm{Y}_\mathrm{i})} \frac{\partial \mu_i}{\partial \eta_i}, j = 1, \dots, p$$

- $var(Y_i)$ ist aufgrund der angenommenen Verteilung der Exponentialfamilie bekannt

Parameterschätzung bei QML

$$\frac{\partial \ell_i}{\partial \beta_j} = \frac{(y_i - \mu_i) x_{ij}}{\upsilon(\mu_i)} \frac{\partial \mu_i}{\partial \eta_i}, j = 1, \dots, p$$

- Varianz wird nur über die angenommenen Mean-Variance-Relationship $var(Y_i) = \upsilon(\mu_i)$ spezifiziert

Parameterschätzung bei GEE:

$$\frac{\partial \ell_i}{\partial \boldsymbol{\beta}} = \mathbf{D}_i' \mathbf{V}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}))$$

mit $\boldsymbol{\Delta}_i = \frac{\partial \theta_{it}}{\partial \eta_{it}}$ als Diagonalemente, $\mathbf{B}_i = b''(\theta_{it})$ als Diagonalelemente,

$\mathbf{D}_i = \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}_i}$, $\mathbf{R}(\alpha)$ als Arbeitskovarianzmatrix und $\mathbf{V}_i = \mathbf{B}_i^{1/2} \mathbf{R}(\boldsymbol{\alpha}) \mathbf{B}_i^{1/2} \phi$.

Eigenschaften:

- Multivariate Erweiterung der Quasi-Likelihood
- Computational leichter aber weniger effizient als Likelihood
- Likelihood basierte Tests können nicht verwendet werden

Schätzung am Beispiel von Modell (1)

Generalisierte Schätzgleichung (GEE) wird verwendet

$$\hat{\beta}_i = \ln\left(\frac{n_i}{N - n_i}\right) \quad \text{und} \quad se_i = \sqrt{\left[N\hat{\pi}_i(1-\hat{\pi}_i)\right]^{-1}} \quad \text{mit} \quad \hat{\pi}_i = \frac{\exp(\hat{\beta}_i)}{1+\exp(\hat{\beta}_i)}$$

Werte aus Randsummentabelle:

	Eissorte			
Studiengang	Schokolade (A)	Kirsche (B)	Gummibär (C)	Anzahl Studenten
Statistik	32	65	86	160
Psychologie	87	148	157	220
Total	119	213	243	380

Schätzung am Beispiel von Modell (1)

Resultierende Parameter für Modell (1):

$$\hat{\beta}_1 = \ln\left(\frac{119}{380-119}\right) \qquad \hat{\beta}_2 = \ln\left(\frac{213}{380-213}\right) \qquad \hat{\beta}_3 = \ln\left(\frac{243}{380-243}\right)$$

Studiengang	Eissorte: Schokolade (A)	Eissorte: Kirsche (B)	Eissorte: Gummibär (C)	Anzahl Studenten
Statistik	32	65	86	160
Psychologie	87	148	157	220
Total	119	213	243	380

Parameter und Standardfehler für Modelle (1) und (2):

	Modell (1)	Modell (2)	
Eis		Statistik	Psychologie
Schokolade (A)	-.773 (.111)	-1.386 (.198)	-.424 (.138)
Kirsche (B)	.254 (.103)	-.379 (.161)	.721 (.144)
Gummibär (C)	.573 (.107)	.150 (.158)	1.913 (.149)

Wald-Test: $W = 46.84$

- $df = k\,(l - 1) = 3$
- $p = 1.85 * 10^{-10}$

→ Eiskauf ist abhängig vom Studiengang

Erweiterte Fragestellung

- Gibt es eine andere Erklärung für den gefundenen Unterschied?
  - ➢ Könnte Eitelkeit die entscheidende Rolle spielen?
- Eitelkeit als zweite systematische Variable aufnehmen
  - – Verschwindet der Effekt „Studiengang“ bedingt auf Eitelkeit?
- Eitelkeit ist nicht-beobachtbare (latente) Größe
  - ➢ Fragebogen (1 – 4 Punkte)

Formalisierung

Modell (3) mit zwei Einflussgrößen:

$$\theta_v + \beta_i = \ln\left(\frac{\pi_{vi}}{1-\pi_{vi}}\right)$$

π_{vi} ist Wahrscheinlichkeit dass Studierender v Sorte i wählt

Bedingtes Modell

- Kauf der Eissorten darf von Eitelkeit abhängen

Bedingtes Modell

Vorteil: Modell determiniert die gemeinsame Verteilung der Responses aller Items (Eissorten)

- Gegeben $\beta_1, \ldots, \beta_k$ und $\theta_1, \ldots, \theta_k$ sind alle Bernoulli Variablen (Antworten) lokal unabhängig

➢ cML Schätzung kann verwendet werden

Gemeinsame Verteilung ist gegeben durch das Produkt der Studierenden-spezifischen Wahrscheinlichkeiten aus Modell (3)

- Über alle Eissorten und Studierende

Eigenschaften bedingtes Modell

Bedeutung des Fehlens von Studiengangsindex j :

Gegeben β_i hängt die Wahrscheinlichkeit ein Eis zu kaufen von der Eitelkeit ab

- Unabhängig vom Studiengang

Hypothese ist simultane (auf Eitelkeit) bedingte Unabhängigkeit zwischen Schultyp und Response (Eiskauf)

Problem

Eitelkeit ist unbeobachtete latente Größe und muss für jeden Studierenden geschätzt werden

- ➢ Neyman-Scott Problem
- ➢ Lösung von Andersen (1970) wird angewandt

→ Anzahl der gekauften Eissorten ist suffizient für jeden Studierenden

$r \in \{0,\ldots,k\}$ ist Anzahl der gekauften Eissorten

- Wird als „Eismenge“ definiert mit k Level
  - Ersetzt die angenommene stetige latente Variable „Eisaffinität“

Bedingte Wahrscheinlichkeit

Modell (4) (bedingtes Rasch Modell):

$$\pi_{ir} = \frac{\exp(\beta_i)\gamma_{r-1}^{(i)}}{\gamma_r}$$

$i = 1, \ldots, k$ und $r = 1, \ldots, k\text{-}1$ mit $\beta_k = 0$ zur Modellidentifizierung

γ_r bezeichnet symmetrische Grundfunktion der Ordnung r der Parameter $\exp(\beta_1), \ldots, \exp(\beta_k)$

$\gamma^{(i)}_{r-1}$ bezeichnet die erste partielle Ableitung von γ_r nach β_i

Parameterschätzung mit cML

Durch Bedingen auf Eismenge r kann die bedingte Wahrscheinlichkeitsverteilung hergeleitet werden

- Anteil, den die Wahrscheinlichkeit für ein bestimmtest Antwortmuster an der Gesamtwahrscheinlichkeit aller Muster mit Summenscore r hat

Bed. Likelihood (cML) ist Funktion der Parameter $\beta_1, \ldots, \beta_k$

- Also nicht mehr abhängig von $\theta_1, \ldots, \theta_k$
- Theoretisch könnte auch umgekehrt bedingt werden

Analog zu den bisher bekannten Personen- und Itemparametern im Rasch Modell

Alternatives Modell

In Modell (4) wird implizit festgelegt:

$$\pi_{ir1} = \ldots = \pi_{irl} = \pi_{ir}$$

→Studiengang *j* spielt keine Rolle

Alternativ kann folgendes Modell (5) aufgestellt werden:

$$\pi_{irj} = \frac{\exp(\beta_{ij})\gamma_{r-1}^{(ij)}}{\gamma_r}$$

Abgeleitet wird hier nach den β_{ij} → cML Schätzung

→Studiengang *j* wird ebenfalls modelliert

Aufgrund der cML Schätzung können der Andersen Likelihood Ratio Test und der Wald Test angewendet werden

Vergleich der Parameter der Modell (4) und (5):

	Modell (4)	Modell (5)	
Eis		Statistik	Psychologie
Schokolade (A)	0	0	0
Kirsche (B)	-1.089 (.160)	-.961 (.253)	-1.187 (.210)
Gummibär (C)	-1.428 (.164)	-1.472 (.256)	-1.378 (.215)

$G^2 = 1.167$ $p = .558$ $df = 2$

$W = .220$; $p = .390$ mit

→ Bedingt auf Eitelkeit ist Eisaffinität unabh. vom Studiengang

Zusammenfassung

Erweiterung der Anwendbarkeit des Rasch Modells

Rasch Modell wird nicht als Messmodell sondern als Logit Modell für die Analyse von Häufigkeiten von zweidimensionalen Kontingenztafeln verwendet

- Bedingte Unabhängigkeit wird getestet

Hypothese der simultanen bedingten Unabhängigkeit wird getestet

- Zwischen systematischer (Studiengang) und dem binären Response auf multiple Items, gegeben eine zweite Variable (Eitelkeit)

Literatur

Agresti, A. (2002). *Categorical data analysis*. Hoboken: John Wiley & Sons.

Agresti, A., & Liu, I. (1999). Modeling a categorical variable allowing arbitrarily many category choices. *Biometrics, 55*, 936–943.

Andersen, E.B. (1970). Asymptotic properties of conditional maximum likelihood estimators. *Journal of the Royal Statistical Society, Series B, 32*, 283–301.

Draxler, C. (2011). Logit models for the analysis of two-way categorial data. *Educational Research and Evaluation, 17*(5), 351–360.

Appendix

- Zusatzmaterial zur Erklärung und Weiterführung

Beweis der Suffizienten Statistik

p(0,1,0) | r = 1

$$\frac{\dfrac{1}{\cancel{1+\xi_1\cdot\varepsilon_1}}\cdot\dfrac{\xi_1\cdot\varepsilon_2}{\cancel{1+\xi_1\cdot\varepsilon_2}}\cdot\dfrac{1}{\cancel{1+\xi_1\cdot\varepsilon_3}}}{\dfrac{\xi_1\cdot\varepsilon_2\cdot 1\cdot 1}{\cancel{(1+\xi_1\cdot\varepsilon_1)+(1+\xi_1\cdot\varepsilon_2)+(1+\xi_1\cdot\varepsilon_2)}}+\dfrac{1\cdot\xi_1\cdot\varepsilon_2\cdot 1}{\cancel{(1+\xi_1\cdot\varepsilon_1)+(1+\xi_1\cdot\varepsilon_2)+(1+\xi_1\cdot\varepsilon_2)}}+\dfrac{1\cdot 1\cdot\xi_1\cdot\varepsilon_3}{\cancel{(1+\xi_1\cdot\varepsilon_1)+(1+\xi_1\cdot\varepsilon_2)+(1+\xi_1\cdot\varepsilon_2)}}}$$

$$=\frac{\cancel{\xi_1}\cdot\varepsilon_2}{\cancel{\xi_1}\cdot\varepsilon_1+\cancel{\xi_1}\cdot\varepsilon_2+\cancel{\xi_1}\cdot\varepsilon_3}\qquad=\frac{\varepsilon_2}{\varepsilon_1+\varepsilon_2+\varepsilon_3}$$

← Bedingte Wahrscheinlichkeit

➢ Nur noch der Itemparameter (ε) ist in der Formel enthalten

Angewandt auf das Rasch-Modell

Die Antworten einer Person *v* auf alle *k* Items wird auch als Antwortmuster $\boldsymbol{x}_v = (x_{v1}, \ldots, x_{vk})$ bezeichnet.

Formal erhalten wir dann für das oben behauptete

$$P = (\boldsymbol{x}_v | r_v) = \frac{\exp\left(\sum_{i=1}^{k} -x_{vi}\beta_i\right)}{\gamma_{r_v}(\boldsymbol{\beta})},$$

für $v = 1, \ldots, n$, wobei $\gamma_r(\boldsymbol{\beta})$ die so genannte symmetrische Grundfunktion *r*-ter Ordnung ist.

Die Symmetrische Grundfunktion:

$$\gamma_1(\boldsymbol{\beta}) = \exp(-\beta_1) + \exp(-\beta_2) + \ldots + \exp(-\beta_k)$$
$$\gamma_2(\boldsymbol{\beta}) = \exp(-\beta_1 - \beta_2) + \exp(-\beta_1 - \beta_3) + \ldots + \exp(-\beta_{k-1} - \beta_k)$$
$$\gamma_3(\boldsymbol{\beta}) = \exp(-\beta_1 - \beta_2 - \beta_3) + \exp(-\beta_1 - \beta_2 - \beta_4) + \ldots + \exp(-\beta_{k-2} - \beta_{k-1} - \beta_k)$$

.

.

.

$$\gamma_k(\boldsymbol{\beta}) = \exp(-\beta_1 - \beta_2 - \ldots - \beta_k),$$

wobei $\gamma_0(\boldsymbol{\beta}) = 1$ und $\gamma_r(\boldsymbol{\beta}) = 0$ für $r < 0$ und $r > k$.

(zur Erinnerung: $\varepsilon_i = \exp(-\beta_k)$)

Symmetrische Grundfunktion

Formel: $\sum_{x|r} \prod_{i=1}^{k} \exp(-x_i \sigma_i)$

Kurzform: $\gamma_r(\exp(-\sigma))$

Eingesetzt im Nenner der cML: $cL = \prod_{v=1}^{N} \frac{\exp(-\sum_{i=1}^{k} x_{vi}\sigma_i}{\gamma_r(\exp(-\sigma))}$

Generalized Linear Mixed Model

Linearer Prädiktor: $g(\mu_{it}) = \mathbf{x}'_{it}\boldsymbol{\beta} + \mathbf{z}'_{it}\mathbf{u}_i$

- y_{it} ist Beobachtung t in cluster i
- $\mathbf{u}_\mathrm{i} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ angenommen
- $\mathrm{var}(Y_{\mathrm{it}}|u_\mathrm{i}) = \Phi_{it}\upsilon(\mu_{it})$
- Bedingt auf $\mathbf{u}_i$, behandelt das Modell y_{it} unabhänging über i und t

Als logit Modell: $\mathrm{logit}[P\left(Y_{it} = 1 | u_i\right) = \beta_t + u_i$

- u_i wird im Raschmodell zwar als fester Effekt behandelt, allerdings durch cML eliminiert
- Andere Autoren: Random Effects Modell mit Probit-Link