Ludwig-Maximilians-Universität München
Institut für Statistik

Seminar: Psychometrische Modelle: Theorie und Anwendungen

Paarvergleichsmodelle

Mila Petkova

07. Juni 2014

Gliederung

Motivation

Idee der Paarvergleichsmodellen:

ein Paar von Reizen, Objekten oder Subjekten wird verglichen, um zu einer Bewertung der einzelnen Elementen zu gelangen.

$\hookrightarrow$ **Ziel:** eine Präferenzordnung zu machen!

Anwendung

- sensorische Wahrnehmung
- Marktforschung
- Sport
- Politik

Beispiel:

4 verschiedene Weinsorten $Wein1$, $Wein2$, $Wein3$, $Wein4$
$\rightarrow$ 6 Paaren zum Vergleich:
$(W1, W2), (W1, W3), (W1, W4), (W2, W3), (W2, W4), (W3, W4)$
Zu jedem Paar werden 15 Personen gefragt welcher von den beiden Weinen sie präferieren.

	Wein1	Wein2	Wein3	Wein4	Total
Wein1	-	3	2	2	7
Wein2	12	-	11	3	26
Wein3	13	4	-	5	22
Wein4	13	12	10	-	35

- Erster Ansatz in Zermelo (1929): Spielstärke von Schachspielern aus ihren Turnierergebnissen zu folgern
- Nach der Idee von Thurstone (1927): das psychologische Kontinuum $\rightarrow$ Messung wie Stimuli (Reize) wahrgenommen werden (und nicht deren physikalische Eigenschaften)
- Wiederentdeckung des Modells durch Bradley and Terry (1952)
- Auswahlaxiom von Luce (1959)

$\Rightarrow$ Das Model auch als Bradley-Terry-Luce-Modell (BTL-Modell) bekannt.

Grundgleichung des BT-Modells:

$$\pi_{ij} := P(\ i \text{ wird gegenüber } j \text{ präferiert}) = \frac{\pi_i}{\pi_i + \pi_j}$$

- Paarweise Vergleich mit $t \geq 2$ Treatments, $T_1, \cdots, T_t$
- Treatments haben ein wahres Rating $\pi_1, \cdots, \pi_t \in \mathscr{P}(t)$ mit $\mathscr{P}(t) = \{\pi \in \mathbb{R}^t | \sum_{i=1}^{t} \pi_i = 1, \forall i : \pi_i \geq 0\}$
- π_{ij} und $\pi_{ji} = 1 - \pi_{ij}$, also $\dfrac{t(t-1)}{2}$ Wahrscheinlichkeiten zu schätzen
- n_{ij} Anzahl (unabhängige) Vergleiche zwischen T_i und T_j

Die Wahrscheinlichkeit aus n_{ij} Vergleiche a_{ij} zu beobachten:

$$\binom{n_{ij}}{a_{ij}} \left(\frac{\pi_i}{\pi_i + \pi_j}\right)^{a_{ij}} \left(\frac{\pi_j}{\pi_i + \pi_j}\right)^{a_{ji}} = \binom{n_{ij}}{a_{ij}} \frac{\pi_i^{a_{ij}} \pi_j^{a_{ji}}}{(\pi_i + \pi_j)^{n_{ij}}}$$

wobei $a_{ij} + a_{ji} = n_{ij}$, a_{ij} Anzahl Präferenzen für T_i gegenüber T_j.

Die Likelihood Funktion hat die Form:

$$L = \prod_{i<j} \binom{n_{ij}}{a_{ij}} \frac{\pi_i^{a_{ij}} \pi_j^{a_{ji}}}{(\pi_i + \pi_j)^{n_{ij}}}$$

$$= \underbrace{\prod_{i<j} \binom{n_{ij}}{a_{ij}}}_{C} \prod_{i<j} \frac{1}{(\pi_i + \pi_j)^{n_{ij}}} \pi_1^{a_{12}} \pi_2^{a_{21}} \pi_1^{a_{13}} \pi_3^{a_{31}} \cdots \pi_2^{a_{23}} \pi_3^{a_{32}} \cdots \pi_{t-1}^{a_{t-1,t}} \pi_t^{a_{t,t-1}}$$

$$= C \frac{\prod_{i=1}^{t} \pi_i^{a_i}}{\prod_{i<j} (\pi_i + \pi_j)^{n_{ij}}}$$

mit $a_i = \sum_{\substack{j \\ j \neq i}} a_{ij}$ die gesamte Anzahl Präferenzen für T_i[1]

[1] a_i ist suffiziente Statistik für die Schätzung von π_i (Buhlmann and Huber, 1963)

Die log-Likelihood Funktion hat die Form

$$\log L \propto \sum_i a_i \log \pi_i - \sum_{i<j} n_{ij} \log(\pi_i + \pi_j)$$

ML-Schätzer, p für π durch

$$\operatorname*{argmax}_{\pi \in \mathscr{P}(t)} \log L$$

mit $\mathscr{P}(t) = \{\pi \in \mathbb{R}^t | \sum_{i=1}^{t} \pi_i = 1, \forall i : \pi_i \geq 0\}$

Nach Ableiten der $\log L$ unter Nebenbedingung $\sum_{i=1}^{t} \pi_i = 1$ ergeben sich:

$$\frac{a_i}{p_i} - \sum_{\substack{j \\ j \neq i}} \frac{n_{ij}}{p_i + p_j} = 0, \; i = 1, \cdots, t$$

und

$$\sum_{i=1}^{t} p_i = 1$$

Für p_i lösen:

$$p_i = \frac{a_i}{\sum_{\substack{j \\ j \neq i}} \frac{n_{ij}}{p_i + p_j}}$$

Iterative Lösung:

Wähle Startwerte für $p_i^{(0)} = 1/t,\ i = 1, \cdots, t$

Berechne:

$$p_i^{(k)} = \frac{a_i}{\sum\limits_{\substack{j \\ j \neq i}} \frac{n_{ij}}{p_i^{(k-1)} + p_j^{(k-1)}}}$$

$$p_i^{(1)} = a_i \left(\frac{n_{i1}}{p_i^{(0)} + p_1^{(0)}} + \cdots + \frac{n_{i,i-1}}{p_i^{(0)} + p_{i-1}^{(0)}} + \frac{n_{i,i+1}}{p_i^{(0)} + p_{i+1}^{(0)}} + \cdots + \frac{n_{it}}{p_i^{(0)} + p_t^{(0)}} \right)^{-1}$$

$p_i^{(2)} = \cdots$
$p_i^{(3)} = \cdots$

Der Prozess wird beendet wenn Konvergenz erreicht: $p_i^{(k+1)} = p_i^{(k)}$

- Bradley and Terry (1952) haben Tabellen mit den Werten für p_i berechnet
- Ford (1957): Prozess konvergiert
- Dykstra (1956, 1960): bessere Startwerte

1. Test über Präferenzengleichheit

$$H_0 : \pi_1 = \pi_2 = \cdots = \pi_t = 1/t$$
$$H_1 : \pi_i \neq \pi_j \text{ für mind. ein Paar (i,j)}$$

Teststatistik:

$$-2\log\lambda_1 = 2N\log 2 - 2B_1,\ N = \sum_{i<j} n_{ij},$$

$$B_1 = \sum_{i<j} n_{ij}\log(p_i + p_j) - \sum_i a_i \log p_i$$

Für n_{ij} groß, $-2\log\lambda_1 \overset{H_0}{\sim} \chi^2_{t-1}$
Testentscheidung: H_0 ablehnen falls $-2\log\lambda_2 > \chi^2_{t-1}$

2. Model-Fit Test

$$H_0 : \pi_{ij} = \frac{\pi_i}{\pi_i + \pi_j} \text{ für } i \neq j,\ i,j = 1, \cdots, t$$
$$H_1 : \pi_{ij} \neq \frac{\pi_i}{\pi_i + \pi_j} \text{ für eine } i,j,\ i \neq j$$

Teststatistik:

$$-2 \log \lambda_2 = 2 \left(\sum_{i \neq j} a_{ij} \log a_{ij} - \sum_{i<j} n_{ij} \log_{ij} + B_1 \right)$$

Für groß n_{ij}, $-2 \log \lambda_2 \overset{H_0}{\sim} \chi^2_{\frac{1}{2}(t-1)(t-2)}$

Testentscheidung: H_0 ablehnen falls $-2 \log \lambda_2 > \chi^2_{\frac{1}{2}(t-1)(t-2)}$

Weitere Tests

3. Test über Präferenzenunterschiede von Subgruppen in Bradley and Terry (1952), Bradley (1976)
4. Model-fit Test in Mosteller (1951)

Sei A endliche Menge von Möglichkeiten, T bel. Teilmenge, $T \subseteq A$, $i \in A$ eine Alternative

Auswahlaxiom von Luce

$$P_A(i) = P_A(T)P_T(i) \tag{1}$$

Beispiel:
Es soll aus einer Speisekarte ($\widehat{=}A$) ein Gericht ($\widehat{=}i$) ausgewählt werden. Im ersten Schritt ($\widehat{=}P_A(T)$) entscheidet sich die Person, ob sie eher ein Fisch- oder ein Fleischgericht essen möchten. Im zweiten Schritt ($\widehat{=}P_T(i)$) wird nur noch aus dem Fischgerichten ein Gericht ausgewählt.

Aus dem Auswahlaxion ergibt sich, dass[2]:

$$P_T(i) = \frac{P_A(i)}{P_A(T)} = P_A(i|T)$$

Beispiel:
Die Wahrscheinlichkeit ein Gericht ($\widehat{=} i$) zu wählen, wenn das Restaurant nur z.B. Fisch- und Fleischgerichte ($\widehat{=} T$) anbietet ist gleich die bedingte Wahrscheinlichkeit auf T, wenn die gesamte Speisekarte ($\widehat{=} A$) angeboten ist.

[2]Beweis in Luce (1959)

Konsequenz:

$$P_T(i) = \frac{P_A(i)}{P_A(T)} = \frac{P_A(i)}{\sum_k P_A(k)} = \frac{\pi_i}{\sum_k \pi_k}$$

mit $\pi_i = P_A(i)$ und $\pi_k = P_A(k)$

Für den Fall von Paarvergleichen, also $T = \{i, j\}$

$$P_T(i) = P(i \text{ gegenüber } j \text{ präferiert}) = \frac{\pi_i}{\pi_i + \pi_j}$$

GLM Notation

$$g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\theta}$$

$\mathbf{Y}$: Vektor der Beobachtungen
$\mathbf{X}$: Design-Matrix
$\boldsymbol{\mu}$: Vektor der Erwartungswerte von Y
$\boldsymbol{\theta}$: Vektor der unbekannten Parameter
$g(.)$: Link-Funktion

Für $i < j$, Y_{ij} Anzahl Präferenzen von i gegenüber j,
n_{ij} Anzahl Vergleiche zwischen i und j

$$\mathbf{Y_{ij}} \sim B(n_{ij}, \pi_{ij})$$

$$\mathbf{Y} = (Y_{12}, Y_{13}, \cdots, Y_{1t}, Y_{23}, \cdots, Y_{t-1,t})'$$

Indices nach den Paarvergleiche (i, j)

$\mathbf{Y}$ hat die $(N \times 1)$-Dimension, $N = \dfrac{t(t-1)}{2}$

$$\boldsymbol{\mu} = (\pi_{12}, \pi_{13}, \cdots, \pi_{1t}, \pi_{23}, \cdots, \pi_{t-1,t})'$$

mit Dimension $(N \times 1)$

Mit Logit als Link-Funktion:

$$g(\boldsymbol{\mu}) = \log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right)$$

Zur Errinering: die Grudgleichung von dem BTL-Modell:

$$\pi_{ij} = \frac{\pi_i}{\pi_i + \pi_j}$$

Mit Logit-Link:

$$\begin{aligned}
\text{logit}(\pi_{ij}) &= \log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) \\
&= \log\left(\frac{\frac{\pi_i}{(\pi_i+\pi_j)}}{1-\frac{\pi_i}{(\pi_i+\pi_j)}}\right) \\
&= \log\left(\frac{\pi_i}{\pi_j}\right) \\
&= \log \pi_i - \log \pi_j \\
&\overset{\theta_i = \log \pi_i}{=} \theta_i - \theta_j
\end{aligned}$$

$$\mathbf{X} = \begin{pmatrix} 1 & -1 & 0 & \dots & \dots & 0 \\ 1 & 0 & -1 & \dots & \dots & 0 \\ \vdots & & & & & \vdots \\ 1 & 0 & 0 & & \dots & -1 \\ 0 & 1 & -1 & & \dots & 0 \\ 0 & 1 & 0 & -1 & \dots & 0 \\ \vdots & & & & & \vdots \\ 0 & 1 & 0 & 0 & \dots & -1 \\ \vdots & & & & & \vdots \\ 0 & \dots & \dots & \dots & 1 & -1 \end{pmatrix},$$

mit Dimension $(N \times t)$ und $rg(\mathbf{X}) = t - 1$

$$\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_{t-1}, 0)'$$

mit Dimension $(t \times 1)$

	Wein1	Wein2	Wein3	Wein4	Total
Wein1	-	3	2	2	7
Wein2	12	-	11	3	26
Wein3	13	4	-	5	22
Wein4	13	12	10	-	35

```
> y.1 <-c(3, 2, 2, 11, 3, 5)
> y.2 <- c(12, 13, 13, 4, 12, 10)
```

```
> X <- matrix(c(1, -1, 0, 0,
+               1, 0, -1, 0,
+               1, 0, 0, -1,
+               0, 1, -1, 0,
+               0, 1, 0, -1,
+               0, 0, 1, -1), nrow=6, byrow=TRUE)
> X
     [,1] [,2] [,3] [,4]
[1,]    1   -1    0    0
[2,]    1    0   -1    0
[3,]    1    0    0   -1
[4,]    0    1   -1    0
[5,]    0    1    0   -1
[6,]    0    0    1   -1
```

```
> btl <- glm(cbind(y.1,y.2) ~ X[,(1:3)] - 1, family = binomial)
> summary(btl)
Call:
glm(formula = cbind(y.1, y.2) ~ X[, (1:3)] - 1, family = binomial)
Deviance Residuals:
      1         2         3         4         5         6
 0.3433  -0.8047   0.6020   1.2555  -1.0562   0.6481
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
X[, (1:3)]1  -2.3571     0.5123  -4.601 4.21e-06 ***
X[, (1:3)]2  -0.7441     0.4208  -1.768   0.0771 .
X[, (1:3)]3  -1.0561     0.4290  -2.462   0.0138 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for binomial family taken to be 1)
    Null deviance: 34.6890  on 6  degrees of freedom
Residual deviance:  4.2399  on 3  degrees of freedom
AIC: 26.768
Number of Fisher Scoring iterations: 4
```

```
> coef <- c(btl$coefficients, 0)
> coef

X[, (1:3)]1 X[, (1:3)]2 X[, (1:3)]3
 -2.3571159  -0.7440733  -1.0561245   0.0000000
```

$$\hat{\boldsymbol{\theta}} = \begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \hat{\theta}_3 \\ \hat{\theta}_4 \end{pmatrix} = \begin{pmatrix} -2.3571159 \\ -0.7440733 \\ -1.0561245 \\ 0 \end{pmatrix}$$

Wir haben $\sum_{i=1}^{4} p_i = 1$

$$\text{mit } \log(p_i) = \hat{\theta}_i \Leftrightarrow p_i = \exp(\hat{\theta}_i)$$

$$\Leftrightarrow \sum_{i=1}^{4} \exp(\hat{\theta}_i) \overset{!}{=} 1$$

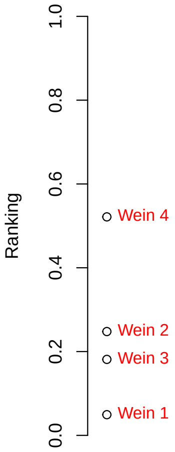
$$\Rightarrow \frac{\exp(\hat{\theta}_1)}{\sum_{i=1}^{4} \exp(\theta_i)} + \frac{\exp(\hat{\theta}_2)}{\sum_{i=1}^{4} \exp(\theta_i)} + \frac{\exp(\hat{\theta}_3)}{\sum_{i=1}^{4} \exp(\theta_i)} + \frac{\exp(\hat{\theta}_4)}{\sum_{i=1}^{4} \exp(\theta_i)} = 1$$

```
> p_i <- exp(coef)/sum(exp(coef))
> p_i
```

```
X[, (1:3)]1 X[, (1:3)]2 X[, (1:3)]3
  0.0493792   0.2477876   0.1813666   0.5214666
```

$$\mathbf{p} = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{pmatrix} = \begin{pmatrix} 0.0493792 \\ 0.2477876 \\ 0.1813666 \\ 0.5214666 \end{pmatrix}$$

Numerisches Beispiel



```
> summary(btl)

Call:
glm(formula = cbind(y.1, y.2) ~ X[, (1:3)] - 1, family = binomial)

Deviance Residuals:
      1        2        3        4        5        6
 0.3433  -0.8047   0.6020   1.2555  -1.0562   0.6481

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
X[, (1:3)]1  -2.3571     0.5123  -4.601 4.21e-06 ***
X[, (1:3)]2  -0.7441     0.4208  -1.768   0.0771 .
X[, (1:3)]3  -1.0561     0.4290  -2.462   0.0138 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 34.6890  on 6  degrees of freedom
Residual deviance:  4.2399  on 3  degrees of freedom
```

Das dichotome Rasch-Modell lässt sich als ein unvollständiges BTL-Modell für den Vergleich von Personen und Aufgaben verstehen.

- die Person dominiert die Aufgabe (die Aufgabe wird gelöst)
- die Aufgabe dominiert die Person (die Aufgabe wird nicht gelöst)

Grundgleichung des BTL-Modells:

$$\text{logit}(\pi_{ij}) = \log(\pi_{ij}/(1-\pi_{ij})) = \theta_i - \theta_j$$

$$\Leftrightarrow \pi_{ij} = \frac{\exp(\theta_i - \theta_j)}{1 + \exp(\theta_i - \theta_j)}$$

Vgl. Grundgleichung des Rasch-Modells:

$$P(u_{ij} = 1 | \theta_i, \beta_j) = \frac{\exp(\theta_i - \beta_j)}{1 + \exp(\theta_i - \beta_j)}$$

- Davidson (1970): Ties = Unentschieden
- Davidson and Beaver (1977): Ordnungsabhängigkeit
- Subjektspezifische Kovariablen
- Objektspezifische Kovariablen
- Triple Comparisons

Bradley, R. A. (1976). A biometrics invited paper. science, statistics, and paired comparisons, *Biometrics* **32**(2): 213–239.

Bradley, R. A. and Terry, M. E. (1952). Rank analysis of incomplete block design: I. the method of paired comparisons, *Biometrika* **39**: 324–345.

Buhlmann, H. and Huber, P. J. (1963). Pairwise comparison and ranking in tournaments, *The Annals of Mathematical Statistics* **34**(2): 501–510.

Davidson, R. R. (1970). On extending the bradley-terry model to accommodate ties in paired comparison experiments, *Journal of the American Statistical Association* **65**: 317–328.

Davidson, R. R. and Beaver, R. (1977). On extending the bradley-terry model to incorporate within-pair order effect, *Biometrics* **33**: 693–702.

Dykstra, Otto, J. (1956). A note on the rank analysis of incomplete block designs – applications beyond the scope of existing tables, *Biometrics* **12**(3): 301–306.

Dykstra, Otto, J. (1960). Rank analysis of incomplete block designs: A method of paired comparisons employing unequal repetitions on pairs, *Biometrics* **16**(2): 176–188.

Ford, L. R., J. (1957). Solution of a ranking problem from binary comparisons, *The American Mathematical Monthly* **64**(8): 28–33.

Luce, R. D. (1959). *Individual Choice Behavior*, Wiley, New York.

Mosteller, F. (1951). Remarks on the method of paired comparisons: Iii. a test of significance for paired comparisons when equal standard deviations and equal correlations are assumed, *Psychometrika* **16**: 207–218.

Thurstone, L. L. (1927). A law of comparative judgment, *Psychological Review* **34**: 273–286.

Zermelo, E. (1929). Die berechnung der turnier-ergebnisse als ein maximum-problem der wahrscheinlichkeitsrechnung, *Math. Zeitschrift* **29**: 436–460.