Stat 315c: Transposable Data
Rasch model and friends

Art B. Owen

Stanford Statistics

Categorical data analysis

- Anova has a problem with too much of the df being in the interaction.
- Generalizations of Anova for categorical data analysis face the same issue.
- They have some good solutions. We'll look at them.
- The Rasch model is a famous waypoint on our journey.

Contingency tables

Setting

- Tables of counts
- 2×2 or $I \times J$ or $\ldots$ $I \times J \times K \times \cdots \times Z$

Contingency tables

Setting

- Tables of counts
- 2×2 or $I \times J$ or $\ldots$ $I \times J \times K \times \cdots \times Z$

Notation

- $n_{ijk\ell}$ = number of obs in cell (i,j,k,ℓ)
- $n_{ij\bullet\bullet} = \sum_{k=1}^{K} \sum_{\ell=1}^{L} n_{ijk\ell}$ etc.
- $\mu_{ijk\ell} = E(n_{ijk\ell})$
- $\pi_{ijk\ell} = \mu_{ijk\ell}/\mu_{\bullet\bullet\bullet\bullet}$
- $\pi_{ij|k\ell} = \pi_{ijk\ell}/\pi_{\bullet\bullet k\ell}$

Default sampling models

- The simplest probability model has independent Poisson random variables

$$n_{ijk\ell} \sim \mathrm{Poi}(\mu_{ijk\ell}) \quad \text{Indep}$$

Default sampling models

- The simplest probability model has independent Poisson random variables

$$n_{ijk\ell} \sim \mathrm{Poi}(\mu_{ijk\ell}) \quad \text{Indep}$$

- Conditioning $N = n_{\bullet\bullet\bullet\bullet}$ we get a multinomial

$$n_{ijk\ell} \sim \mathrm{Mult}(N; \mu_{ijk\ell})$$

Default sampling models

- The simplest probability model has independent Poisson random variables

$$n_{ijk\ell} \sim \mathrm{Poi}(\mu_{ijk\ell}) \quad \text{Indep}$$

- Conditioning $N = n_{\bullet\bullet\bullet\bullet}$ we get a multinomial

$$n_{ijk\ell} \sim \mathrm{Mult}(N; \mu_{ijk\ell})$$

- Conditioning (holding fixed) the $K \times \cdots \times L$ margin we get a product of multinomials on $I \times \cdots \times J$:

$$n_{ij|k\ell} \sim \mathrm{Mult}(n_{\bullet\bullet k\ell}; \pi_{ij|k\ell}) \quad \text{Indep over } k \text{ and } \ell$$

These usually lead to the same likelihoods and same inferences.
There are also hypergeometric models and generalizations that hold fixed eg the I and J margins without fixing the $I \times J$ margin.

Multinomial model

Sum N independent observations from Mult$(1; \pi_{ijk\ell})$.
Clearly $\pi_{ijk\ell} \geq 0$.
Suppose $\pi_{ijk\ell} > 0$.
For $I \times J$, write

$$\log \pi_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$$

Suppose that $(\alpha\beta)_{ij} = 0$ for all i and j
Then

$$\pi_{ij} = \exp(\mu + \alpha_i + \beta_j) = \exp(\mu) \times \exp(\alpha_i) \times \exp(\beta_j)$$

So rows and columns are independent.
And conversely.

Log linear models

Anova expansion

$$\log \pi_{ijk} = \mu + \alpha_k + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk}$$

Hierarchical models

- Dropping interactions amounts to simplifying dependence
- EG I×J indep of K, or,
- EG I and J indep given K, etc.
- Tests come from Poisson likelihood

Downside

- Huge numbers of parameters
- that are hard to interpret
- response is a count, not one of the marginal variables
- rejecting independence is often uninteresting

Modeling the interaction

Residence in 1966 and 1971

	CC	ULY	WM	GL
CC	118	12	7	23
ULY	14	2127	86	130
WM	8	69	2548	107
GL	12	110	88	7712

4 regions of UK
Reject independence
But so what?

Mover-Stayer models (Quasi-Independence)

Some people won't move, some might. Model via

$$\log \pi_{ij} = \mu + \alpha_i + \beta_j + \delta_i 1_{i=j}$$

using $\sim 3K$ params
Might even use

$$\log \pi_{ij} = \mu + \alpha_i + \beta_j + \delta 1_{i=j}$$

More models (from Agresti)

Quasi-Symmetry

$$\log \pi_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} \quad \text{with} \quad (\alpha\beta)_{ij} = (\alpha\beta)_{ji}$$

Symmetry

$$\log \pi_{ij} = \mu + \alpha_i + \alpha_j + (\alpha\beta)_{ij} \quad \text{with} \quad (\alpha\beta)_{ij} = (\alpha\beta)_{ji}$$

Equilibrium

$$\pi_{i\bullet} = \pi_{\bullet i}$$

$$\sum_j \exp(\mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}) = \sum_i \exp(\mu + \alpha_i + \beta_j + (\alpha\beta)_{ij})$$

Bradley-Terry models

Preferences

- Customer prefers wine i to wine j
- Team i beats team j

Models

- Basic model

$$\Pr(\text{i beats j}) = \frac{e^{\alpha_i - \alpha_j}}{1 + e^{\alpha_i - \alpha_j}}$$

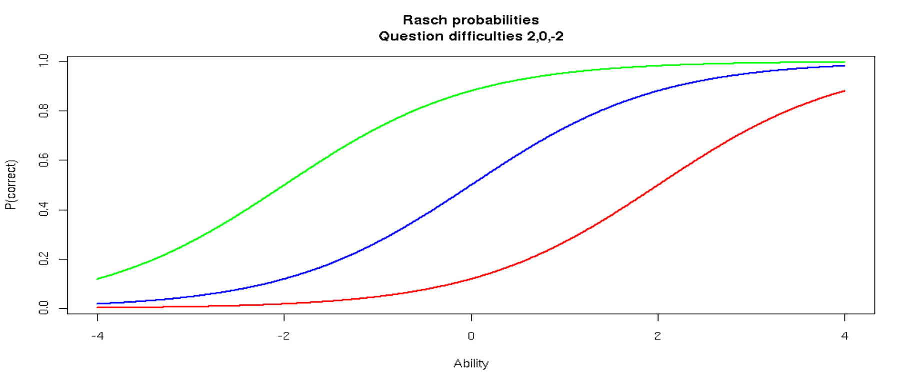
- Fit by logistic regression (no intercept)
- Overidentified ... so take $\alpha_1 = 0$ or other constraint
- With covariates x eg home field advantage

$$\Pr(\text{i beats j} \mid X = x) = \frac{e^{x\beta + \alpha_i - \alpha_j}}{1 + e^{x\beta + \alpha_i - \alpha_j}}$$

Rasch model

Educational testing

- Student i with ability θ_i
- Question j with difficulty β_j
- Probability correct $\frac{e^{\theta_i - \beta_j}}{1+e^{\theta_i - \beta_j}}$



Rasch model ctd

Latent variables

- Threshold model

$$Y_{ij} = 1_{\theta_i - \beta_j + \varepsilon_{ij} > 0}$$

with latent variables $\varepsilon_{ij} \sim$ logistic

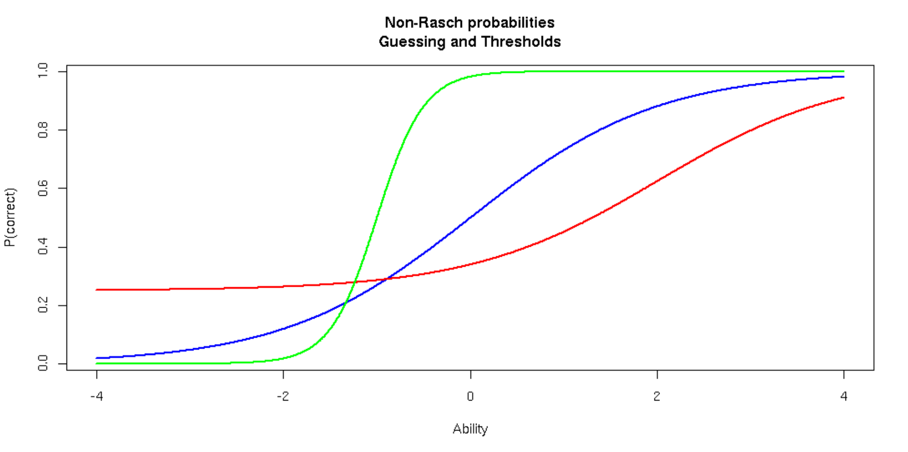
Effect types

- Students random and questions fixed
- (But sometimes questions are drawn from a pool)

Estimation

- Conditional maximum likelihood
- Huge number of student params
- Small number of item params
- Condition on estimates of ability so e^{θ_i}'s cancel out
- See Agresti and web site for his book

Rasch problems



- Green question has steeper threshold
- Red question is hard, so they guess

Alternative models

Two parameters for items

$$\Pr(\text{Correct}_{ij}) = \frac{e^{(\theta_i - \beta_j)\delta_j}}{1 + e^{(\theta_i - \beta_j)\delta_j}}$$

...with guessing

$$\Pr(\text{Correct}_{ij}) = \gamma_i + (1 - \gamma_i)\frac{e^{(\theta_i - \beta_j)\delta_j}}{1 + e^{(\theta_i - \beta_j)\delta_j}}$$

...and probit link

$$\Pr(\text{Correct}_{ij}) = \gamma_i + (1 - \gamma_i)\Phi((\theta_i - \beta_j)\delta_j)$$

More item response theory

Followups

- Local School of Ed: David Rogosa, Ingram Olkin, Ed Haertel
- Lord and Novick (1968) “Statistical Theories of Mental Tests”
- Thissen and Wainer (Eds) (2001) “Test Scoring”

Check an item

by plotting correctness vs $\hat{\theta}_i$ (and smoothing)

Interactions

- Can look at subtests
- Can't fit many params per student
- Can pool students e.g. CA vs NY vs ...