Part 3: Parameter Estimation in the Rasch Model

Parameter Estimation - general

given a data vector $x = (x_1, x_2, ..., x_n)$
is a random sample from an unknown population
goal of data analysis is to acquire knowledge about population

each population is identified by a probability distribution, specified as a function of (usually unknown) parameters

2 situations:

statistical tests:
if parameters were known, probability for specific data can be calculated
– assumptions on parameters are made under H_0
– statistical test: evaluates sample data given these assumptions

estimation:
- try to get knowledge about unknown parameters

Parameter Estimation

example: ball and urn experiment
– red and black balls, let $P(\text{red}) = \pi$
– draw $n = 10$ balls (with replacement)
– data: $\{X_1 = \text{red}, X_2 = \text{red}, \ldots, X_{10} = \text{black}\}$
$k = 3$ had been red, $n - k = 7$ had been black

probability distribution function (PDF):

$$P(\text{observing exactly this sample}) = \pi \cdot \pi \cdot \cdots \cdot (1 - \pi) \cdot (1 - \pi) = \pi^3 \, (1 - \pi)^7$$

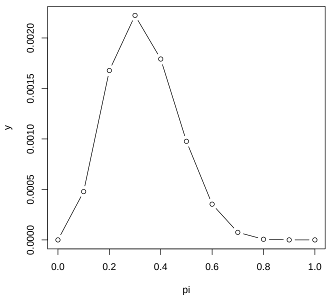
general: $\pi^k \, (1 - \pi)^{n-k}$

we do not know π - how can we calculate it?

Maximum Likelihood (ML) Method: we choose that π that is most likely to have generated the sample

let's try different values of π:

pi	P(pi|k=3,n-k=7)
0.0	0.00000
0.1	0.00048
0.2	0.00168
0.3	0.00222
0.4	0.00179
0.5	0.00098
0.6	0.00035
0.7	0.00008
0.8	0.00001
0.9	0.00000
1.0	0.00000



`P(pi|k=3,n-k=7)` is called likelihood, generally $L(\theta | x_1, \ldots, x_n)$

if we want to estimate π, we look for the maximal value of the likelihood function – maximum likelihood (ML) estimation

general: to obtain the maximum of a function:
set first derivative to zero - solve equation(s)
derivative often easier found, when using the log likelihood

$$L(\pi|x_1,\dots,x_n) = \prod_{i=1}^{n} \pi^{x_i}\,(1-\pi)^{(1-x_i)} = \pi^k\,(1-\pi)^{n-k}$$

$$\log L = k \log \pi + (n-k)\log(1-\pi)$$

$$\frac{d\log L}{d\pi} = k\frac{1}{\pi} + (n-k)\frac{1}{1-\pi}\cdot(-1) = \frac{k}{\pi} - \frac{n-k}{1-\pi} = 0$$

ML estimator (function to estimate the parameter - rule):

$$\frac{k}{\pi} = \frac{n-k}{1-\pi} \quad \rightarrow \quad k - k\pi = n\pi - k\pi \quad \rightarrow \quad \pi = \frac{k}{n}$$

ML estimate (result of applying the rule):

$$\hat{\pi} = \frac{k}{n} = 0.3$$

Parameter Estimation in the RM

Item Parameter Estimation

- likelihood based methods:
  differ in their treatment of person parameters
  - joint ML estimation (JML)
  - conditional ML estimation (CML)
  - marginal ML estimation(MML)
- other methods available:
  less often used
  not covered here

Person Parameter Estimation

- ML and weighted ML estimation
- Bayes approaches

Joint Maximum Likelihood (JML)

or 'unconditional' ML

$$L_u = \frac{\exp(\sum_v \theta_v r_v)\exp(-\sum_i \beta_i s_i)}{\prod_v \prod_i (1+\exp(\theta_v - \beta_i))}$$

joint estimation of item and person parameters
sufficient statistics are: $r_v = \sum_i x_{vi}$ for θ_v and $s_i = \sum_v x_{vi}$ for β_i

problem:
as $n \to \infty$ estimates for item parameters are inconsistent
and biased in finite samples with $k(k-1)$

Conditional Maximum Likelihood (CML)

condition on r_v

$$L_c = \exp(-\sum_i \beta_i s_i) / \prod_r \sum_{x|r} \exp(-\sum_i x_i \beta_i)^{n_r}$$

– person parameters do not occur in the conditional likelihood
– items can be compared independent of persons (separation)
– leads to specific objectivity
– person free item calibration
– 'sample-independence':
  actual sample not of relevance for inference on item parameters

CML estimates are unbiased and consistent as $n \to \infty$
for estimability set $\beta_1 = 0$ or $\sum \beta_i = 0$
items with score $s_i = 0$ or n and person with $r_v = 0$ or k are removed prior to estimation

Derivation of the conditional likelihood

basic idea:
use the conditional probability of observing a certain response pattern $\boldsymbol{x}_v$ given the marginal sum r_v

$$P(\boldsymbol{x}_v|r_v;\theta_v,\boldsymbol{\beta}) = \frac{P(\boldsymbol{x}_v|\theta_v,\boldsymbol{\beta})}{P(r_v|\theta_v,\boldsymbol{\beta})}$$

example: 4 items, $r = 3$, observed pattern $\boldsymbol{x}_v$ is

1101

all possible response patterns $\boldsymbol{x}$ with $r = 3$:

1110 1101 1011 0111

Derivation of the conditional likelihood (ctd.)

rewrite RM in multiplicative form

$$P(X_{vi} = 1) = \frac{\xi_v \epsilon_i}{1 + \xi_v \epsilon_i}, \qquad \xi_v = \exp(\theta_v), \epsilon_i = \exp(-\beta_i)$$

probability for the response pattern $\boldsymbol{x}_v$ for a certain subject v

$$P(\boldsymbol{x}_v|\xi_v,\boldsymbol{\epsilon}) = \prod_{i=1}^{k} \frac{(\xi_v\epsilon_i)^{x_{vi}}}{1+\xi_v\epsilon_i} = \frac{{\theta_v}^{r_v}\prod_{i=1}^{k}{\epsilon_i}^{x_{vi}}}{\prod_{i=1}^{k}(1+\xi_v\epsilon_i)}$$

probability for a fixed raw score r_v is

$$P(r_v|\xi_v,\boldsymbol{\epsilon}) = \sum_{\boldsymbol{y}|r_v}\prod_{i=1}^{k} \frac{(\xi_v\epsilon_i)^{x_{vi}}}{1+\xi_v\epsilon_i} = \frac{{\theta_v}^{r_v}\sum_{\boldsymbol{y}|r_v}\prod_{i=1}^{k}{\epsilon_i}^{x_{vi}}}{\prod_{i=1}^{k}(1+\xi_v\epsilon_i)}$$

Derivation of the conditional likelihood (cont'd)

collecting all terms

$$P(\boldsymbol{x}_v|r_v;\theta_v,\boldsymbol{\beta}) = \frac{P(\boldsymbol{x}_v|\theta_v,\boldsymbol{\beta})}{P(r_v|\theta_v,\boldsymbol{\beta})} = \frac{\frac{{\theta_v}^{r_v}\prod_{i=1}^{k}{\epsilon_i}^{x_{vi}}}{\prod_{i=1}^{k}(1+\xi_v\epsilon_i)}}{\frac{{\theta_v}^{r_v}\sum_{\boldsymbol{y}|r_v}\prod_{i=1}^{k}{\epsilon_i}^{x_{vi}}}{\prod_{i=1}^{k}(1+\xi_v\epsilon_i)}} = \frac{\prod_{i=1}^{k}{\epsilon_i}^{x_{vi}}}{\sum_{\boldsymbol{y}|r_v}\prod_{i=1}^{k}{\epsilon_i}^{x_{vi}}}$$

crucial term is: $\sum_{\boldsymbol{y}|r_v}\prod_{i=1}^{k}{\epsilon_i}^{x_{vi}} \equiv \gamma_r(\epsilon_i)$

the γ's are called *elementary symmetric functions* (of order r)

$$\begin{aligned}
\gamma_0 &= 1\\
\gamma_1 &= \epsilon_1 + \ldots + \epsilon_k\\
\gamma_2 &= \epsilon_1\epsilon_2 + \epsilon_1\epsilon_3 + \ldots + \epsilon_{k-1}\epsilon_k\\
&\vdots\\
\gamma_k &= \epsilon_1\epsilon_2\epsilon_3\cdots\epsilon_{k-1}\epsilon_k
\end{aligned}$$

Marginal Maximum Likelihood (MML)

instead of conditioning integrate out the person parameter

$$L_m = \prod_r \left[\exp(-\sum_i \beta_i s_i)) \int \frac{\exp(\theta r)}{\prod_{i=1}^{k}(1+\exp(\theta-\beta_i))} dG(\theta)\right]^{n_r}$$

distribution for θ , i.e., $G(\theta)$ must be specified
usually it is assumed that $\theta \sim N(0,1)$

Marginal Maximum Likelihood (MML) (cont'd)

Advantages:

- gives also estimates for persons with $r_v = 0$ or $r_v = k$
- advantageous if research aims at person distribution
- allows estimation of additional parameters (2PL, 3PL models)

Disadvantages:

- parameters can be grossly biased if $G(\theta)$ incorrectly specified
- CML closer to concept of person-free assessment
- no argument for specific objectivity
- several goodness-of-fit tests not available

distributional properties of CML and MML estimated are asymtotically the same

can be estimated in R using the **ltm** package (Rizopoulos,2009)

Derivation of the marginal likelihood

probability of observing a certain response pattern $\boldsymbol{x}_v$

$$\begin{aligned} P(\boldsymbol{x}_v|\theta_v,\boldsymbol{\beta}) &= P(\boldsymbol{x}_v|r_v;\theta_v,\boldsymbol{\beta})P(r_v|\theta_v,\boldsymbol{\beta}) \\ &= \int P(\mathbf{x}_v|\theta_v,\boldsymbol{\beta})dG(\theta) \end{aligned}$$

inserting the RM parameters gives

$$P(\mathbf{x}_v) = \exp(-\sum_i \beta_i s_i) \int \frac{\exp(\theta_v r_v)}{\prod_{i=1}^k (1+\exp(\theta_v - \beta_i))} dG(\theta)$$

product over all subjects gives L_m

Person Parameter Estimation

using the unconditional likelihood

$$L_u = \frac{\exp(\sum_v \theta_v r_v)\exp(-\sum_i \beta_i s_i)}{\prod_v \prod_i (1+\exp(\theta_v - \beta_i))}$$

and assuming the βs to be known (from prior estimation)

slightly biased (bias smaller than s.e.'s of estimates)
no estimates for $r_v = 0$ and $r_v = k$
can be approximated using, e.g., spline interpolation

weighted ML estimation:
likelihood function is skewed, additional source of estimation bias
Warm suggests unbiasing correction, computationally unfeasible