Part 8: Polytomous Models

Partial Credit Model (PCM) and Rating Scale Model (RSM)

Polytomous Models

extension to more than two response categories $h = 0, 1, \ldots, m$

nominal responses (multidimensional):

The Polytomous Multidimensional RM

$$P(X_{vi} = h|\theta_{vh}, \beta_{ih}) = \frac{\exp(\theta_{vh} - \beta_{ih})}{1 + \exp(\theta_{vh} - \beta_{ih})}$$

there are h latent dimensions

X_{vi} . . . person v scores in category h of item i
θ_{vh} . . . location of person v on latent trait h
β_{ih} . . . location of item i on h-th latent trait

ordinal responses (unidimensional):

Partial Credit Model (PCM; Masters, 1982)

$$P(X_{vih} = 1) = \frac{\exp[h\theta_v - \beta_{ih}]}{\sum_{l=0}^{m_i} \exp[l\theta_v - \beta_{il}]}$$

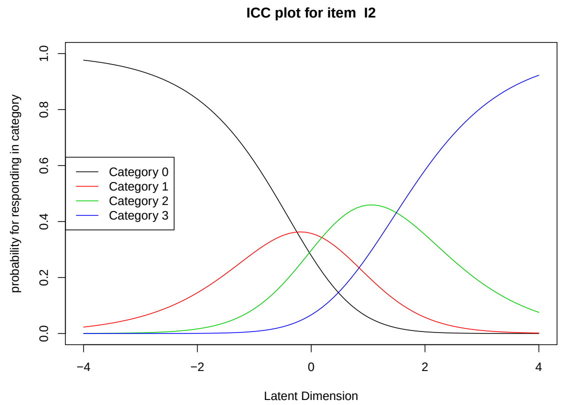
introducing the restrictions $\theta_{vh} = h\theta_v$
β_{ih}'s describe item-category combinations
number of categories may vary across items (m_i)

alternative formulation:

$$P(X_{vi} = h) = \frac{\exp[h(\theta_v - \beta_i) + \omega_{hi}]}{\sum_{l=0}^{m_i} \exp[l(\theta_v - \beta_i) + \omega_{li}]}$$

ω_{hi} are category parameter, have also interpretation as cumulative thresholds

ICCs for the PCM



Derivation of the PCM

basic idea: ordered performance levels

example: three-steps mathematic item: $\sqrt{\frac{7.5}{0.3} - 16} = ?$

	Performance Levels 0	1	2	3
$7.5/0.3 = ?$	0 $\xrightarrow{\text{first step}}$	1		
$25 - 16 = ?$		1 $\xrightarrow{\text{second step}}$	2	
$\sqrt{9} = ?$			2 $\xrightarrow{\text{third step}}$	3

each step can be modelled by a RM:
$\phi_{vi1} = \pi_1/(\pi_0 + \pi_1)$, $\phi_{vi2} = \pi_2/(\pi_1 + \pi_2)$, ...

Derivation of the PCM (cont'd)

first step: response is 0 or 1

$$\phi_{vi1} = \frac{\exp[\theta_v + \tau_{i1}]}{1 + \exp[\theta_v + \tau_{i1}]} = \frac{\pi_1}{(\pi_0 + \pi_1)}$$

second step: response is 1 or 2 (cannot be 0 or 3)

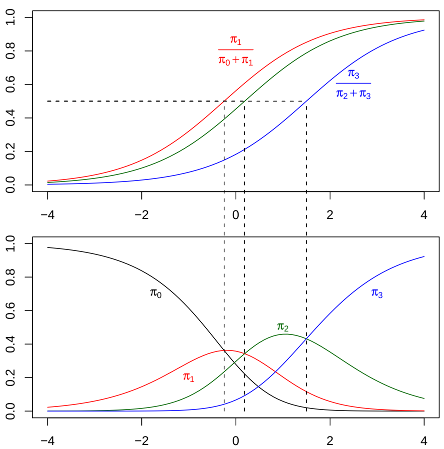
$$\phi_{vi2} = \frac{\exp[\theta_v + \tau_{i2}]}{1 + \exp[\theta_v + \tau_{i2}]} = \frac{\pi_2}{(\pi_1 + \pi_2)}$$

...

τ_{ih} are difficulties of reaching level h in item i
must always be ordered since sufficient statistics s_{ih} are ordered (there cannot be more persons reaching level h than $h - 1$)

only difference to the RM is that $\pi_h + \pi_{h-1} < 1$ since $\sum_h \pi_h = 1$

Category Probability Curves



Derivation of the PCM (cont'd)

general expression:

$$\pi_{vih} = \frac{\exp[\sum_{j=0}^{h}(\theta_v - \tau_{ij})]}{\sum_{l=0}^{m_i} \exp[\sum_{j=0}^{l} \theta_v - \tau_{ij}]} \qquad h = 0, 1, \ldots, m_i \qquad (\sum_{l=0}^{0} \cdots \equiv 0)$$

this is the PCM or alternatively

$$\pi_{vih} = \frac{\exp[h\theta_v - \sum_{j=0}^{h} \tau_{ij}]}{\sum_{l=0}^{m_i} \exp[l\theta_v - \sum_{j=0}^{l} \tau_{ij}]} \qquad \beta_{ih} = \sum_{j=1}^{h} \tau_{ij}, \qquad \tau_{i0} = 0$$

π_{vih} is the probability that person v reaches level h
step difficulties β_{ih} can be estimated independently of θ_v (CML)
the sufficient statistic for θ is h, the count of successfully completed steps

Derivation of the PCM (cont'd)

how about hierarchical dependence?
RM requires one parameter for each item and probabilities being independent

alternative view of β:
instead of ordered level difficulty it can be seen as difficulty of each successive step

- third step, e.g., is from level 2 to level 3
- difficulty of this step governs probability to complete this step (to level 3)
- i.e., the probability of making 3 rather than 2 (once having reached 2)

it says nothing about other steps, they depend on θ and the other β's

Threshold Formulation

β_{ij} can be rewritten as $h\beta_i + \omega_{ih}$ giving

$$\pi_{vih} = \frac{\exp[h\theta_v - \beta_{ih}]}{\sum_{l=0}^{m_i} \exp[l\theta_v - \beta_{il}]} = \frac{\exp[h(\theta_v - \beta_i) + \omega_{ih}]}{\sum_{l=0}^{m_i} \exp[l(\theta_v - \beta_i) + \omega_{il}]}$$

ω's can be interpreted as category 'difficulty' parameters

when using

$$\phi_{vij} = \frac{\exp[\theta_v + \tau_{ij}]}{j + \exp[\theta_v + \tau_{ij}]} = \frac{\exp[\theta_v - (\beta_i + \tau_j)]}{j + \exp[\theta_v + (\beta_i + \tau_j)]}$$

in the derivation of the model and normalise τ as $\sum \tau_j = 0$ then
– β_i is mean of the threshold locations
– τ_j are the distances to the thresholds

ω's are cumulative τ's, i.e., $\omega_{ih} = \sum_{j=1}^{h} \tau_{ij}$

The Rating Scale Model (RSM)

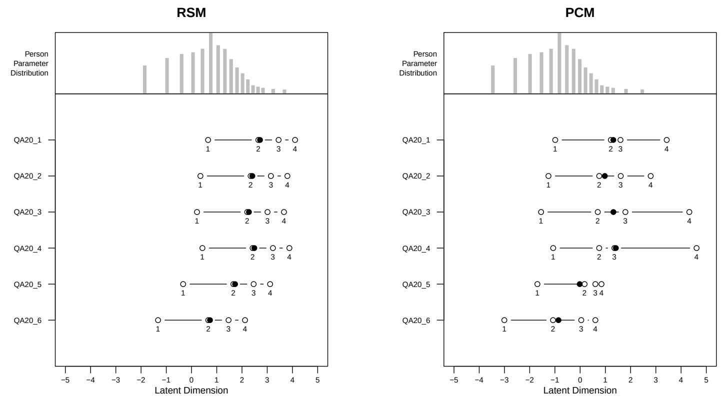
derived in a different context (Andrich, 1978)
can be seen as special case of the PCM

if we simplify the PCM by $\omega_{hi} = \omega_h$ for all i and m_i is m

$$\pi_{vih} = \frac{\exp[h(\theta_v - \beta_i) + \omega_h]}{\sum_{l=0}^{m_i} \exp[l(\theta_v - \beta_i) + \omega_l]}$$

this is sometimes called 'equidistant' scoring
we assume, that the distances between the categories are equal across all items
used for 'Likert Scales'
often too restrictive
cannot detect possible violation of 'ordinality'

Comparison RSM vs PCM



R commands

main functions concerning fit of polytomous models:

- `PCM(data)` fits the PCM and generates object of class `Rm`
- `RSM(data)` fits the RSM and generates object of class `Rm`
- `thresholds(rmobj)` displays the itemparameter estimates as thresholds
- all other functions are the same as previously presented (except for `plotjointICC()`)

PCM Example

Data: Eurobarometer 71.1 (Jan/Feb 2009)

Question Q20:
6 Items on satisfaction with aspects of everyday life

```
qa20_1:  HOUSING
qa20_2:  AREA
qa20_3:  LIVING STANDARD
qa20_4:  STATE OF HEALTH
qa20_5:  MEDICAL SERVICES
qa20_6:  JOB OPPORTUNITIES
```

responses recoded (for this example):
(0) not at all satisfactory . . . (3) very satisfactory

Italian subsample, $n = 1009$ (NAs removed)

PCM Example

ASK ALL

QA20 I am now going to read out different aspects of everyday life. For each, could you tell me if this aspect of your life is very satisfactory, fairly satisfactory, not very satisfactory or not at all satisfactory?

(SHOW CARD WITH SCALE - ONE ANSWER PER LINE)

		(READ OUT)	Very satisfactory	Fairly satisfactory	Not very satisfactory	Not at all satisfactory	DK
(334)	1	Your house or flat	1	2	3	4	5
(335)	2	The quality of life in the area where you live	1	2	3	4	5
(336)	3	Your standard of living	1	2	3	4	5
(337)	4	Your state of health	1	2	3	4	5
(338)	5	The medical services in your local area	1	2	3	4	5
(339)	6	The job opportunities in your local area	1	2	3	4	5

EB66.3 QA3 TREND MODIFIED

Analysis using eRm

```
# data
load(file="zacatI.Rdata")
#
pM<-PCM(zacatI[,1:6])
thresholds(pM)
plotPImap(pM)
#
# check the model
LRtest(pM)
# items 1, 3, 4 inappropriate response patterns
# let's have a look at the distribution of the response patterns
apply(zacatI[,1:6],2,table) # response distribution
#
# rawscores
r<-rowSums(zacatI[,1:6])
median(r)
mean(r)
attach(zacatI)
table(r,QA20_1)  # suggests to split: <=8,>8
```

Analysis using eRm (cont'd)

```
# look at possible split values for other items
table(r,QA20_3)  # 'too good' to split
table(r,QA20_4)  # either at 6, 7 or 8
#
# let's try sex
lrs<-LRtest(pM,splitcr=SEX)
lrs
#
# let's try age
lra<-LRtest(pM,splitcr=AGE) # significant
lra
# again inappropriate response patterns
#
# now Item 1
table(QA20_1,AGE) # no cat 1 response for youngest category
# collapse age categories
age3<-ifelse(AGE>1,AGE-1,AGE)
lra3<-LRtest(pM,splitcr=age3) # still significant
lra3
```


Analysis using eRm (cont'd)

```
# plot estimates
beta3<-as.matrix(as.data.frame(lra$betalist))
beta3<- -beta3 # difficulty parameters
pairs(beta3, lower.panel=panel.smooth, upper.panel=panel.smooth)
#
# one with value very large value
table(age3,QA20_6) # few responses in category 3
#
# check if RSM is possible
rM<-RSM(zacatI[,1:6])
devdiff<-2*(pM$loglik-rM$loglik)
dfdiff<-pM$npar-rM$npar
1-pchisq(devdiff,dfdiff)
# no
#
# further steps can be taken by collapsing categories, covariate levels,...
#
detach(zacatI) # don't forget
```