

Teorie odpovědi na položku

PSY028 | JARO 2018 | BLOK 3

IRT VS. RASCHŮV MODEL; VÍCEPARAMETROVÉ MODEL; MODEL PRO POLYTOMICKÉ ODPOVĚDI.

Raschův model: opakování

Předpoklady základního Raschova modelu:

- Existuje spojitý, intervalový latentní rys, který „způsobuje“ pozorované binární odpovědi.
- Tyto odpovědi záleží dále na parametrech položky.
- Odpověď lze predikovat prostřednictvím tzv. **charakteristické funkce položky**.
- $P_i(\theta) = \frac{e^{(\theta-b_i)}}{1+e^{(\theta-b_i)}} = \frac{1}{1+e^{-(\theta-b_i)}}$; analogicky pak $\ln \frac{P_i(\theta)}{1-P_i(\theta)} = \theta - b_i$
- Pozorované odpovědi jsou navzájem lokálně nezávislé vzhledem ke specifikovanému modelu.

Při fundamentálním měření musí být atributy měřené proměnné nezávislé na měřícím nástroji.

- Raschovské estimátory (např. JMLE) bez předpokladu o distribuci lat. rysu.
- „Data musí vyhovovat modelu.“

RM: Analýza IDS v R

```
install.packages("mirt")
library(psych)
library(mirt)

ids <-
read.csv2("ids.csv")

ids <- ids[,2:21]
describe(ids)
alpha(ids)
omega(ids); omega(ids,
                    nfactors=1)

irt1 <- mirt(ids, 1,
             itemtype = "Rasch")

irt1
summary(irt1)
M2(irt1)
itemfit(irt1, fit_stats =
"infitem")
coef(irt1, simplify=T)
itemplot(irt1, 10)
plot(irt1, type="trace")
```

Polytomické Raschovy modely

Prof. David Andrich právě
obdržel Benjamin Drake
Wright Senior Scholar
Award

<https://goo.gl/DL1zbw>



RM: Polytomické odpovědi

Základní Raschův model pracuje s binárními odpověďmi.

- Ty jsou dobré u výkonových testů, ale u posuzovacích škál nesou příliš málo informace; delší odpověďové škály s sebou nesou vyšší reliabilitu měření.

Historicky dvě hlavní rozšíření:

- Partial Credit Model (PCM; Masters, 1982).
- Rating Scale Model (RSM; Andrich, 1978).

Ačkoliv historický vývoj obou modelů byl odlišný, lze je přepsat do velmi podobného tvaru, přičemž RSM má o něco silnější předpoklady.

- Jsou tedy uhnížděné („nested“).

3 hlavní kategorie polytomních modelů:

- **kategorie mohou být ordinální (PCM, RSM)**
- kategorie jsou ordinální (GRM, modified GRM)
- kategorie jsou nominální (NRM) – 2PL only

Rating Scale Model

Původní účel byl pro modelování Likertových škál (Andrich, 1978; Masters, 1982).

Jednotlivé odpověďové kategorie jsou od sebe odděleny prahy („thresholds“), umístěnými na kontinuu latentního rysu.

Pravděpodobnost pozorované odpovědi pak záleží na tom, mezi kterými prahy se nachází hodnota respondentova latentního rysu.

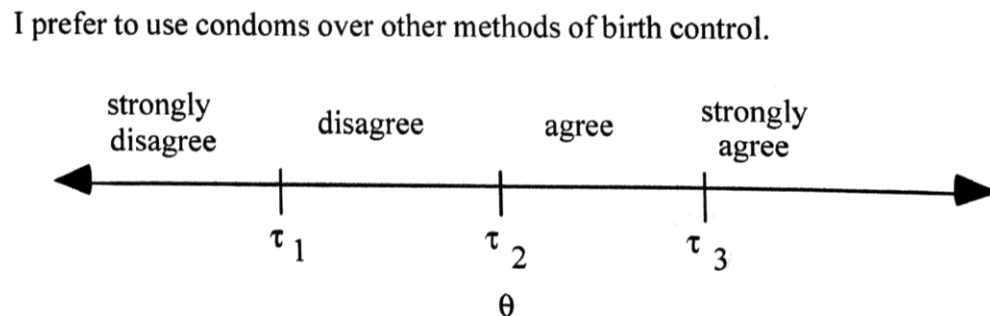


FIGURE 7.8. Representation of item parameter characterization for the RS model ($m = 3$).

Rating Scale Model (RSM)

Jednotlivé položky pak lze seřadit na té stejné škále latentního rysu.

Obtížnost položek je pak definována jako průměr jednotlivých prahů.

Všechny položky mají stejnou strukturu prahů.

Odvození modelu viz PCM dále.

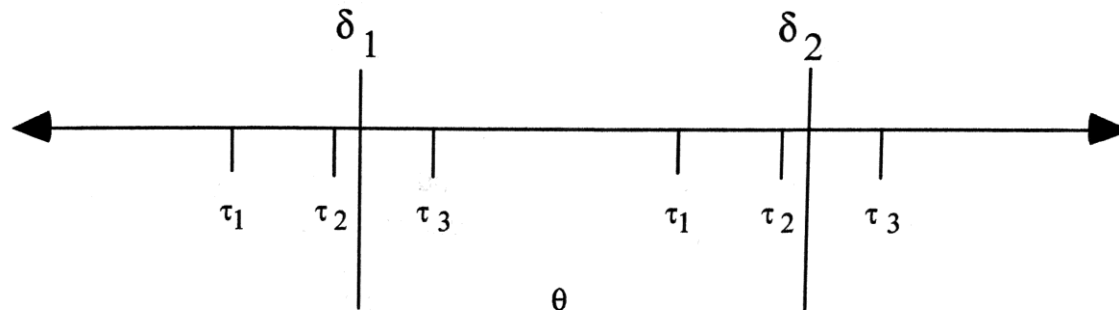


FIGURE 7.9. Graphical representation of item locations and thresholds for two items.

Rating Scale Model (RSM)

Charakteristická křivka Rating Scale Modelu je definována pomocí série rovnic:

$$P\{X_{ni} = x; x > 0\} = \frac{e^{\sum_{k=1}^x [\theta_n - (b_i - \tau_k)]}}{\sum_{j=1}^m e^{\sum_{k=1}^j [\theta_n - (b_i - \tau_k)]}}$$

- Pro všechny částečně správné odpovědi

$$P\{X_{ni} = 0\} = \frac{1}{1 + \sum_{j=1}^m e^{\sum_{k=1}^j [\theta_n - (b_i - \tau_k)]}}$$

- Pro špatnou odpověď.

$$\sum_{k=1}^m (\tau_k) = 0$$

- Pro účely identifikace modelu, aby bylo možné pracovat s obtížností položky

X_{ni} – predikovaná odpověď respondentu n na položku i .

x – odpověďová kategorie v rozmezí $0-m$.

θ_n – míra latentního rysu respondentu n .

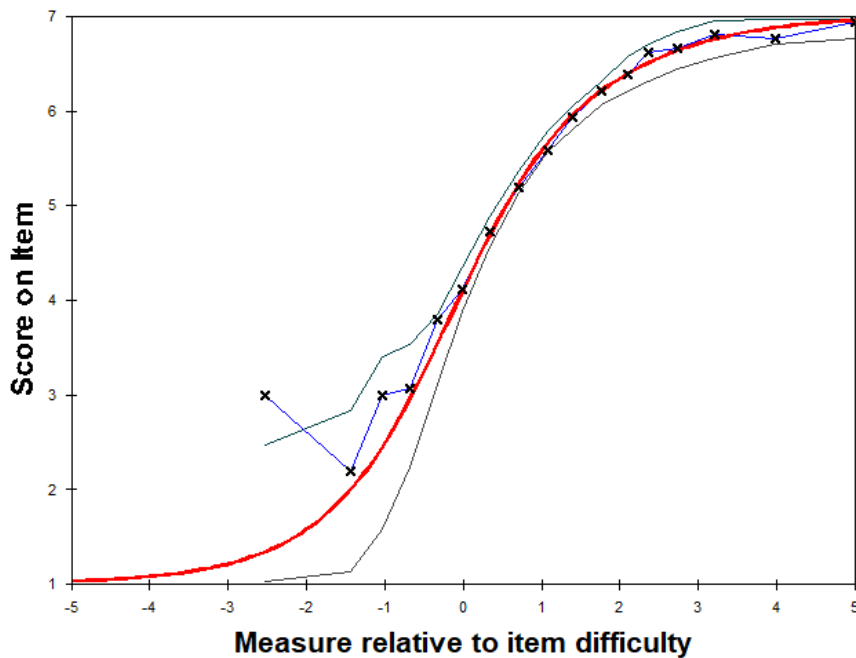
b_i – obtížnost položky i .

τ_k – hodnota „ k -tého“ prahu.

- 1. práh odděluje odpověďové kategorie 0 a 1.

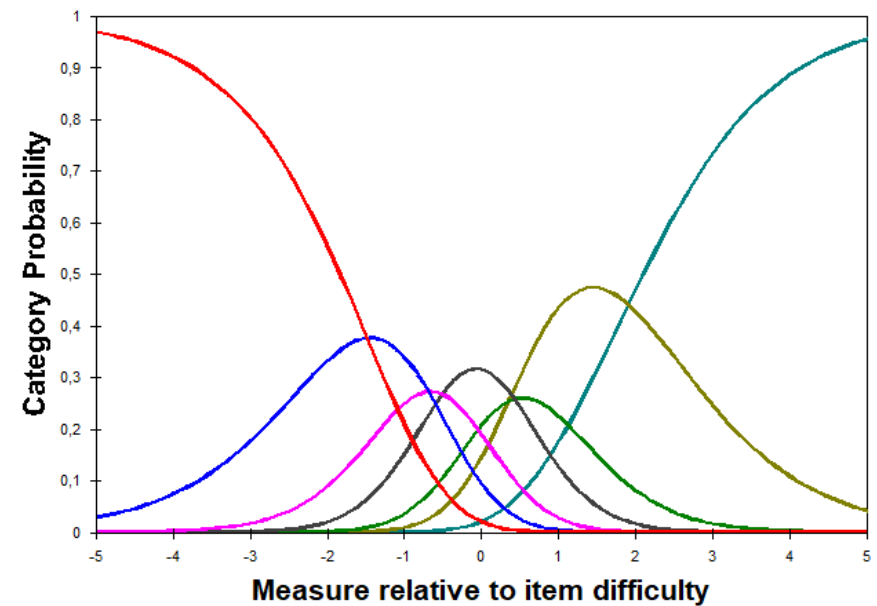
Rating Scale Model (RSM)

1. Můžu se sám/sama rozhodovat, jak chci žít svůj život.



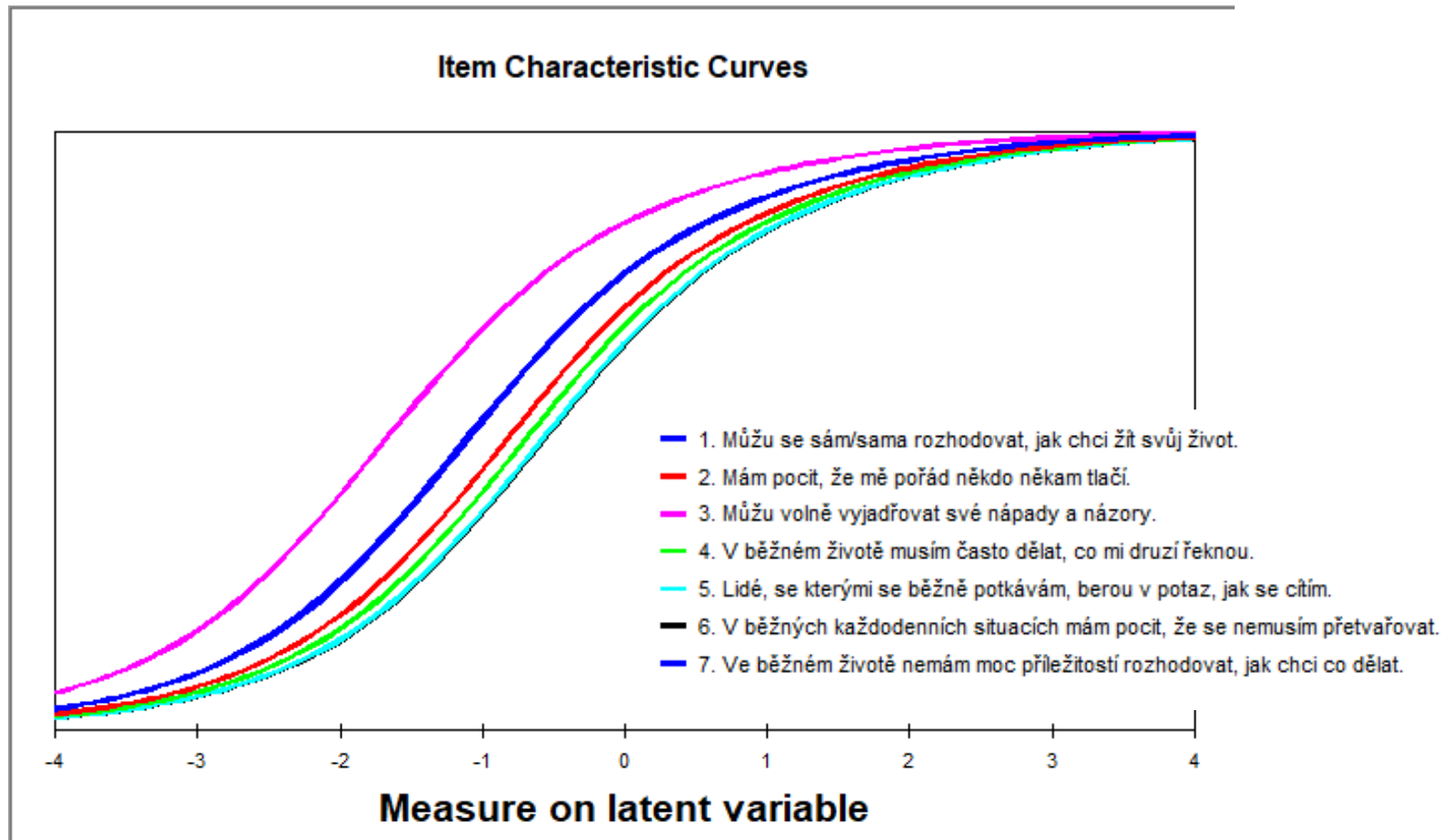
— Expected score ogive: Model ICC
 — Upper 95% 2-sided confidence interval
 * Empirical ICC
 — Lower 95% 2-sided confidence interval

1. Můžu se sám/sama rozhodovat, jak chci žít svůj život.



— Category probability: 1 Vůbec mě to nevystihuje
 — Category probability: 2
 — Category probability: 3
 — Category probability: 4 do určité míry mě to vystihuje
 — Category probability: 5
 — Category probability: 6
 — Category probability: 7 úplně mě to vystihuje

Rating Scale Model (RSM)



Partial Credit model (PCM)

Původní účel PCM modelu byl v případě výkonových testů, kde pro správnou odpověď je nutné provést řadu dílčích úkonů.

Typický příklad 3+2×6:

- 0 bodů – zcela chybně
- 1 bod – správná identifikace přednosti násobení před sčítáním
- 2 body – správný výpočet.

Jde tedy o sérii položek či „kroků“, oddělených tzv. „step difficulty“.

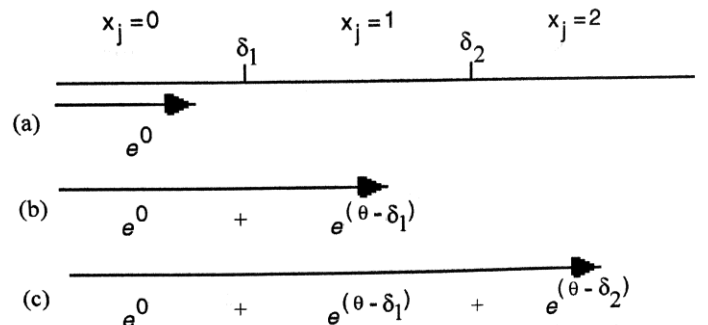


FIGURE 7.1. Schematic representation of the processes for obtaining the category scores of 0, 1, and 2.

Partial Credit model (PCM)

PCM nám říká: „Pokud by respondent odpověděl kategorií k , jaká je pravděpodobnost, že odpoví i $k+1$?

- „Jaká je pravděpodobnost „tranzice“ do vyšší kategorie“?
- Jaká je pravděpodobnost, že odpoví $k+1$, a ne k ?

V případě odpovědi za **0 bodů** jde vlastně o pravděpodobnost „chybné“ odpovědi ze základního RM: $P(x_i = 0|\theta) = 1 - P(x_i = 1|\theta) = 1 - \frac{e^{\theta - \delta_1}}{1 + e^{\theta - \delta_1}} = \frac{1}{1 + e^{\theta - \delta_1}} = \frac{e^0}{e^0 + e^{\theta - \delta_1}}$, což lze zobecnit jako $P(x_i = 1|\theta) = \frac{e^0}{\varphi}$, kde φ je označuje všechny možné výstupy.

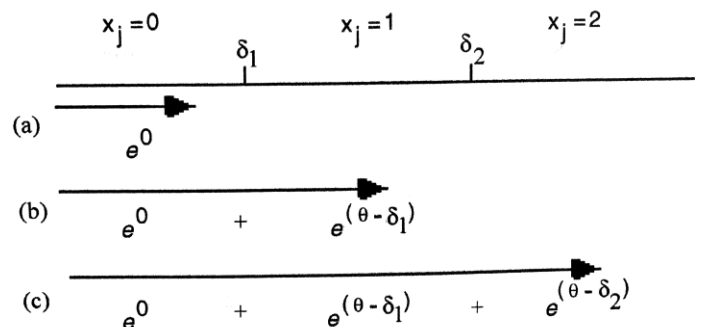


FIGURE 7.1. Schematic representation of the processes for obtaining the category scores of 0, 1, and 2.

Partial Credit model (PCM)

Předchozí příklad lze zobecnit pro **c bodů**: $P[(x_i = c | x_i = c - 1) | \theta] = \frac{e^{\theta - \delta_c}}{\varphi}$, tedy jaká je pravděpodobnost, že respondent odpoví c vs. $c-1$.

To δ_c je potom tzv. „step difficulty“, „step parameter“, „transition parameter“.

Většinou nás ale nezajímá, jaká je podmíněná pravděpodobnost „2 nebo 1 podle théta“, ale jaká je pravděpodobnost konkrétní odpovědi.

Toto zobecnění provedl právě Masters (1982).

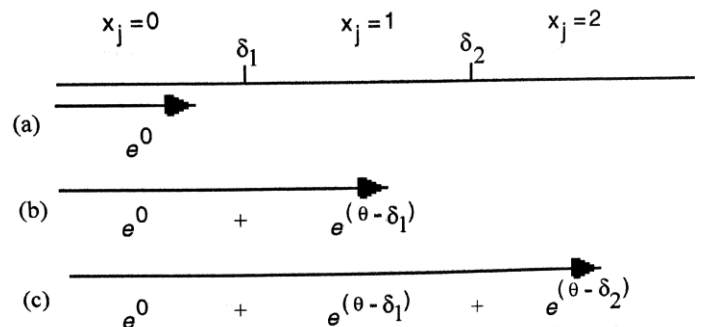


FIGURE 7.1. Schematic representation of the processes for obtaining the category scores of 0, 1, and 2.

Partial Credit model (PCM)

Pro libovolnou odpověď x_j na položku j :

$$P(x_j | \theta, \delta_j) = \frac{e^{\sum_{h=0}^{x_j} (\theta - \delta_{jh})}}{e^0 + \sum_{k=1}^{m_j} e^{\sum_{h=0}^k (\theta - \delta_{jh})}} = \frac{e^{\sum_{h=0}^{x_j} (\theta - \delta_{jh})}}{\sum_{k=0}^{m_j} e^{\sum_{h=0}^k (\theta - \delta_{jh})}}$$

- kde m_j je maximální možný skór na položce j

Na rozdíl od RSM tedy chybí celková obtížnost položky. Identicky ale lze model definovat tak, že obtížnost položek je průměrem kroků, a definovat prahy jako rozdíl původních kroků a této obtížnosti.

- Tedy $\delta_{jh} = b_j - \tau_{jk}$
- Aby byla obtížnost identifikovaná, je nutné fixovat jeden z prahů nebo (zpravidla) jejich průměr (na nulu).

Partial Credit model (PCM)

Potom tedy:

$$P\{X_{ni} = x; x > 0\} = \frac{e^{\sum_{k=1}^x [\theta_n - (b_i - \tau_{ik})]}}{\sum_{j=1}^{m_i} e^{\sum_{k=1}^j [\theta_n - (b_i - \tau_{ik})]}}$$

$$P\{X_{ni} = 0\} = \frac{1}{1 + \sum_{j=1}^{m_i} e^{\sum_{k=1}^j [\theta_n - (b_i - \tau_{ik})]}}$$

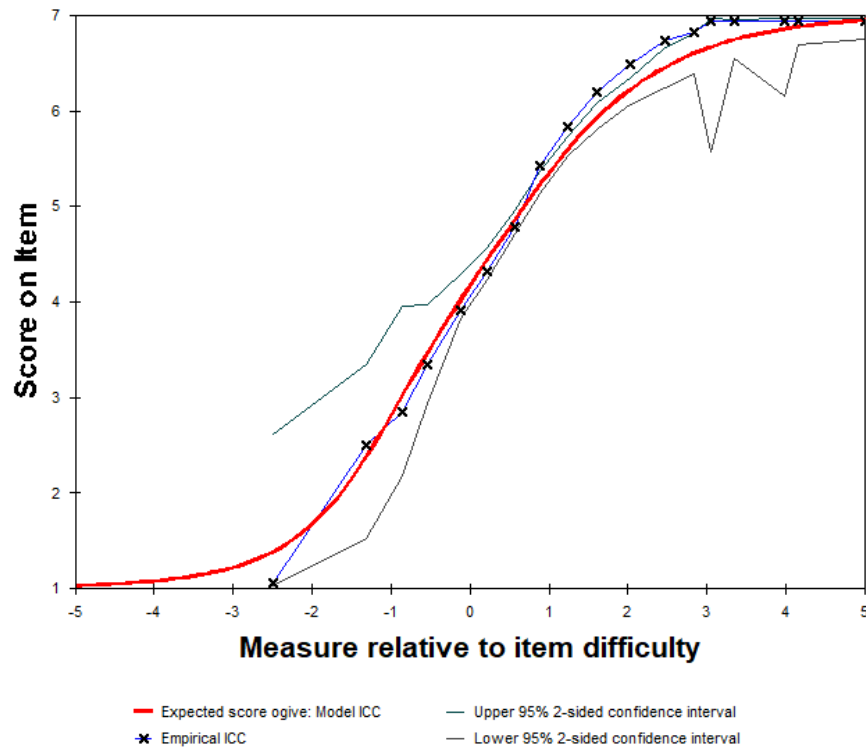
$$\sum_{k=1}^{m_i} (\tau_{ik}) = 0$$

Rozdíl RSM a PCM je ten, že PCM má strukturu prahů rozdílnou pro každou položku, kdežto RSM je má stále shodné.

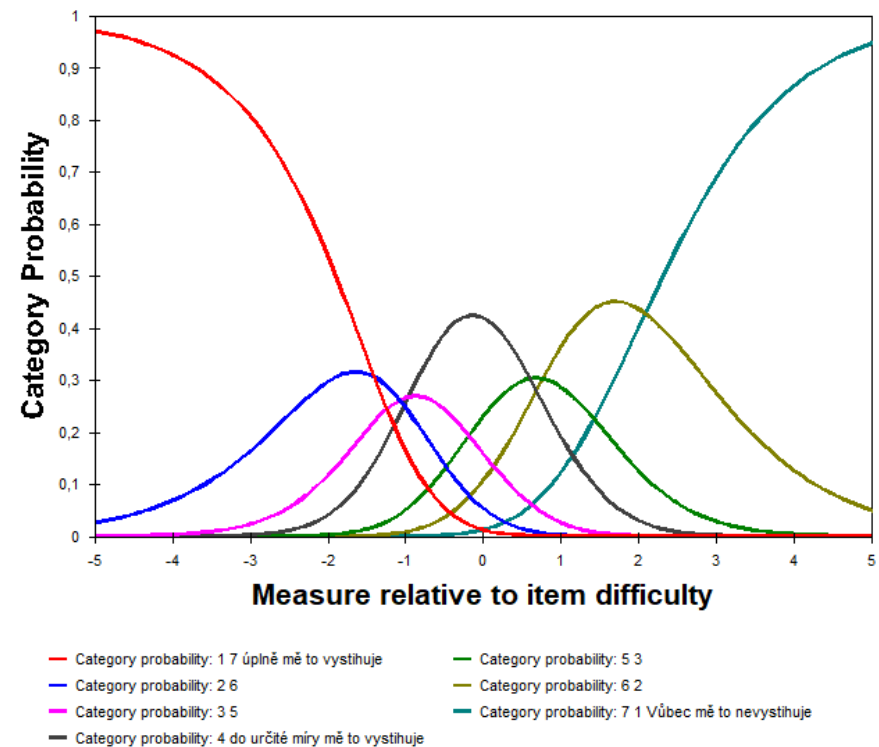
- Pro RSM platí vždy $\tau_{ik} = \tau_k$

Partial Credit model (PCM)

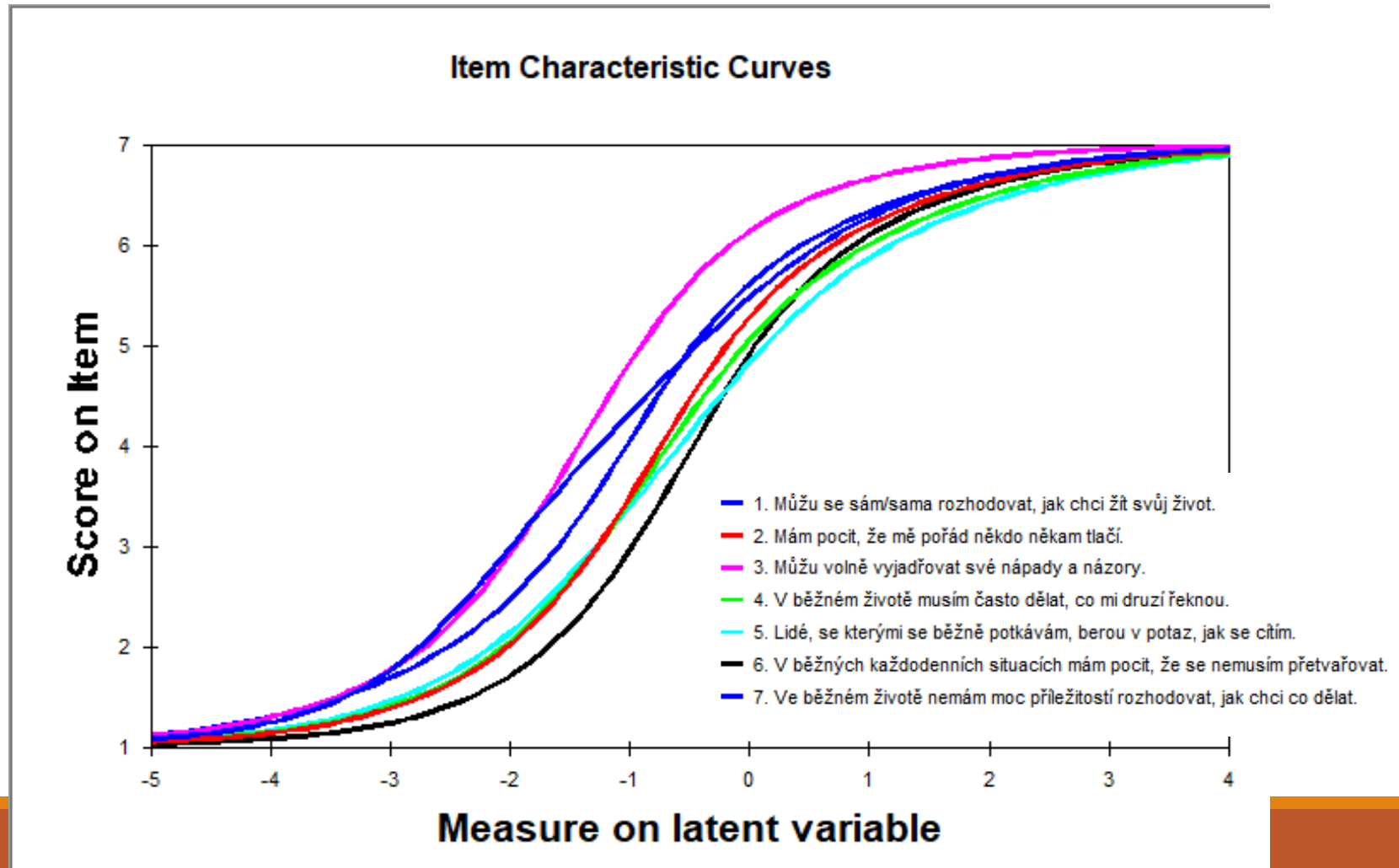
7. Ve běžném životě nemám moc příležitostí rozhodovat, jak chci co dělat.



7. Ve běžném životě nemám moc příležitostí rozhodovat, jak chci co dělat.



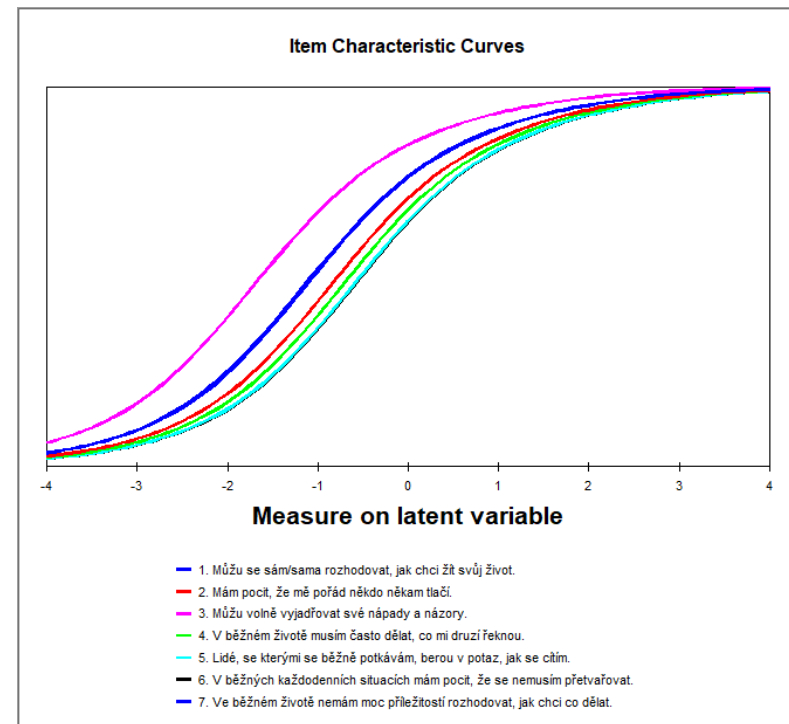
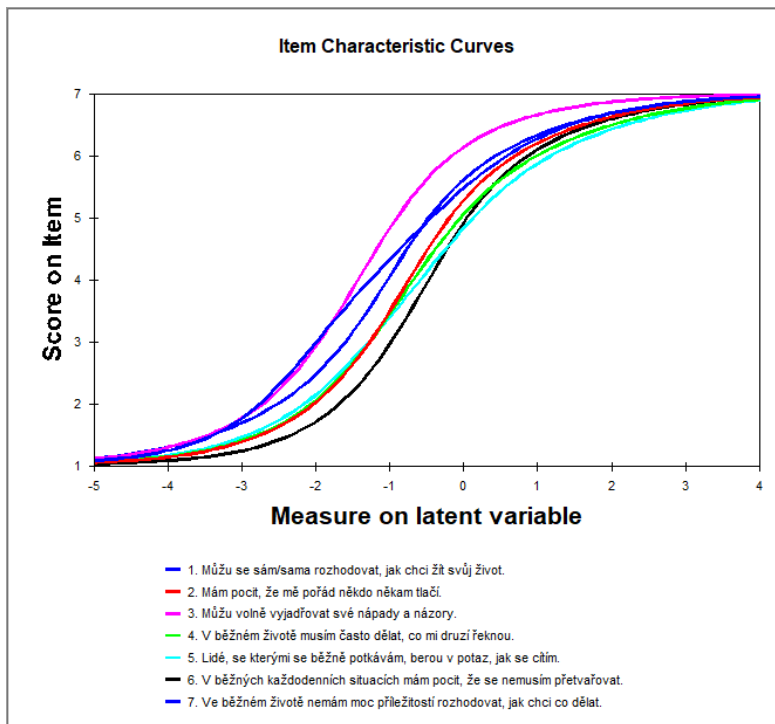
Partial Credit model (PCM)



PCM

×

RSM



Různé definice prahů

Zatím jsme pracovali s tzv. strukturními prahy (Andrichovy nebo Rasch-Andrichovy). Lze definovat ale další, které zjednodušují interpretaci.

- Tyto jsou používány pro konstrukci ICC.
- Ovšem prahy mohou ležet „mimo“ charakteristické křivky kategorií.

Raschův „půlbodový“ práh („Rasch-half-point thresholds“).

- Místo na škále latentního rysu, kde je pravý skór položky c-0,5 (např. práh mezi kategoriemi 2 a 3 je právě tam, kde očekáváme odpověď 2,5 hrubého skóre)
- Tedy $E(X_{ni}|\theta_n) = k - 0,5$, pokud $\theta_n = b_i - \tau_{ik}$.

Rasch-Thurstonův práh.

- Místo, kde očekáváme shodnou pravděpodobnost nižší, jako pravděpodobnost vyšší odpovědi.
- Tedy $P\{X_{ni} \leq k - 1\} = P\{X_{ni} \geq k\}$, pokud $\theta_n = b_i - \tau_{ik}$.

Důležité vlastnosti

(Platí pro všechny polytomní modely)

Skórovací funkce (scoring function): analogie ICC u binární položky.

- Odhad pravého skóru na položce, tedy očekávaný pozorovaný hrubý skór.
- Součet charakteristických funkcí kategorií, tedy pravděpodobností všech možných kategorií, tedy $E(x_i|\theta) = \bar{x}_i(\theta) = \sum_{j=0}^{m_i} P_j(x = j|\theta)$.

Informační funkce: součet informačních funkcí kategorií.

- $I_i(\theta) = \sum_{j=1}^{m_i} I_{ij}(\theta)$. Ty se na celkové informační funkci podílejí takovou měrou, s jakou pravděpodobností nastanou, tedy lze rozparcelovat:
- $I_{ij}(\theta) = P_{ij}(\theta)I_i(\theta)$

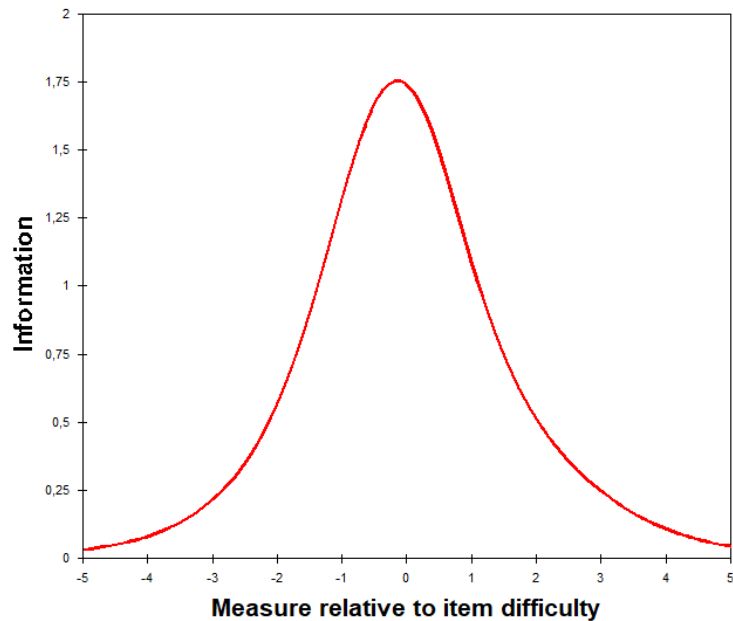
Informační funkce celé položky je potom:

- $I_i(\theta) = \sum_{j=0}^{m_i} \{ [x_{ij} - \bar{x}_i(\theta)]^2 P_{ij}(\theta) \}$
- $\bar{x}_i(\theta) = \sum_{j=0}^{m_i} [x_{ij} P_{ij}(\theta)]$, kde $\bar{x}_i(\theta)$ je skórovací funkce.

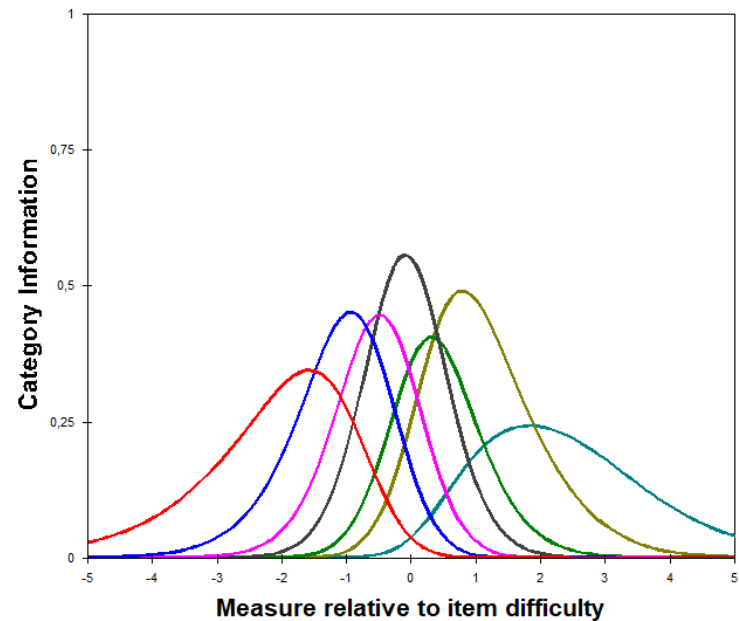
Fit položek lze spočítat pro každou kategorii, a nebo pro celou položku pomocí součtů obdobně jako u informační funkce.

Informační funkce (RSM)

1. Můžu se sám/sama rozhodovat, jak chci žít svůj život.

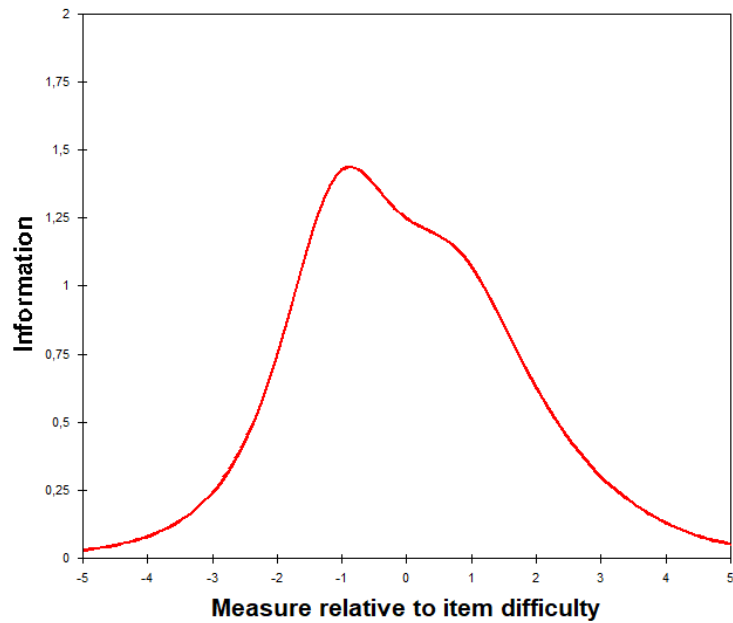


1. Můžu se sám/sama rozhodovat, jak chci žít svůj život.

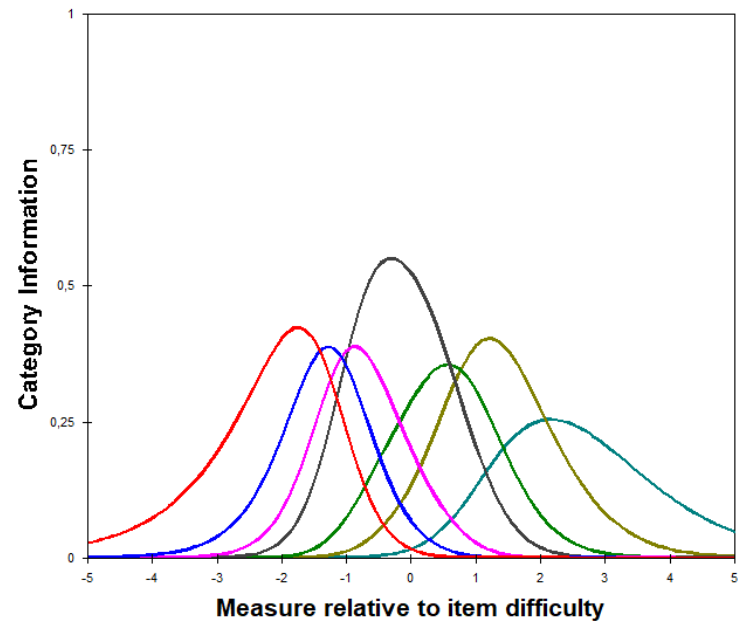


Informační funkce (PCM)

1. Můžu se sám/sama rozhodovat, jak chci žít svůj život.



1. Můžu se sám/sama rozhodovat, jak chci žít svůj život.



Prakticky: Autonomie

POLOŽKY

Můžu se sám/sama rozhodovat, jak chci žít svůj život.

*Mám pocit, že mě pořád někdo někam tlačí.**

Můžu volně vyjadřovat své nápady a názory.

*V běžném životě musím často dělat, co mi druzí řeknou.**

Lidé, se kterými se běžně setkávám, berou v potaz, jak se cítím.

V běžných každodenních situacích mám pocit, že se nemusím přetvařovat.

*Ve běžném životě nemám moc příležitostí rozhodovat, jak chci co dělat.**

LIKERTOVA ŠKÁLA

1 Vůbec mě to nevystihuje

2

3

4 do určité míry mě to vystihuje

5

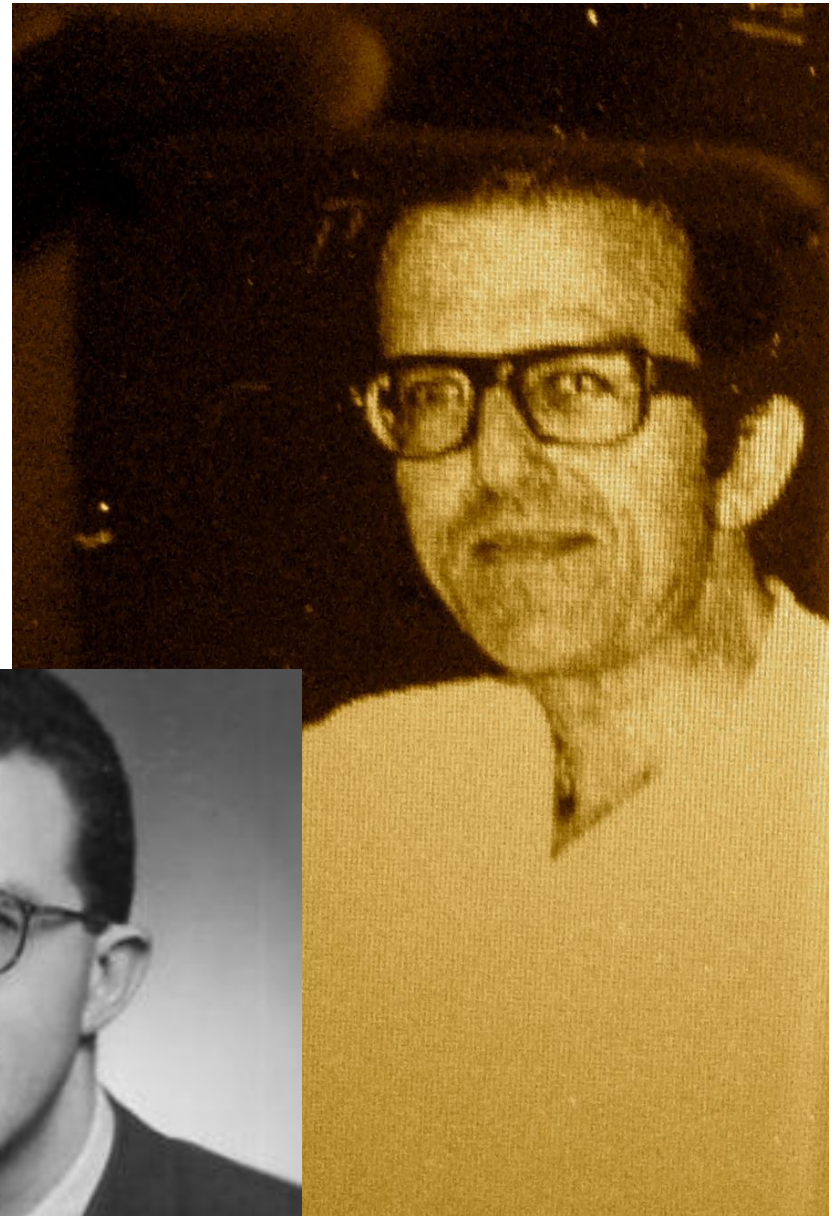
6

7 úplně mě to vystihuje

Vícepara- metrové IRT modely

Allan Birnbaum (3PL model)

Náhodička: Birnbaum, A. (1968).
Some latent trait models and their
use in inferring an examinee's
ability. In Lord, F. M. and Novick, M.
R. (eds.) *Statistical Theories of
Mental Test Scores*, pp 397–479.
Reading, MA: Addison-Wesley.



Raschův model: nedostatky

RM nikdy nepopíše data dobře, protože má dost striktní předpoklady.

- Všechny položky diskriminují stejně (tau-ekvivalence položek).
- Respondent s extrémní mírou rysu má 0%, resp. 100% pravděpodobnost správné odpovědi (a co hádání?).
- Respondenti jsou lokálně nezávislí (a co když jde o žáky ze tříd?).
- Lze identifikovat jednodimenzionální měřené rysy, resp. každá položka je sycena jen jedním latentním rysem (nelze měřit „dvě věci“ jednou položkou).
 - Raschista by řekl, že RM jednodimenzionalitu „vyrobí“, a že nechci měřit „výškováhu“, ale „výšku“ a pak „váhu“.

Raschův model je veden snahou o vytvoření fundamentální škály. Ostatní IRT přístupy jsou vedeny o co nejlepší popis dat.

- Menší epistemologické nároky, vyšší míra pragmatismu.

Dvouparametrový model (2PL)

Diskriminační parametr a_i je rozlišovací schopnost položky: ukazuje, jak dobře rozdělí položka lepší/horší respondenty

$$P_i(\theta) = \frac{e^{a_i(\theta - b_i)}}{1 + e^{a_i(\theta - b_i)}}$$

- Analogií je faktorový náboj v FA.
- Model je identifikován buď rozptylem latentního rysu ($\text{var}(\theta) = 1$) a nebo (výjimečně) loadingem první položky ($a_1 = 1$) – tedy stejně, jako v FA.
- V případě, že všechny položky mají stejný diskriminační parametr ($a_i = a$), jde o 1PL model; rozdíl s RM je ten, že 1PL volně odhaduje „škálu“ rysu (fixací rozptylu), kdežto RM škálu odhaduje prostřednictvím fixace loadingů (zpravidla na 1).

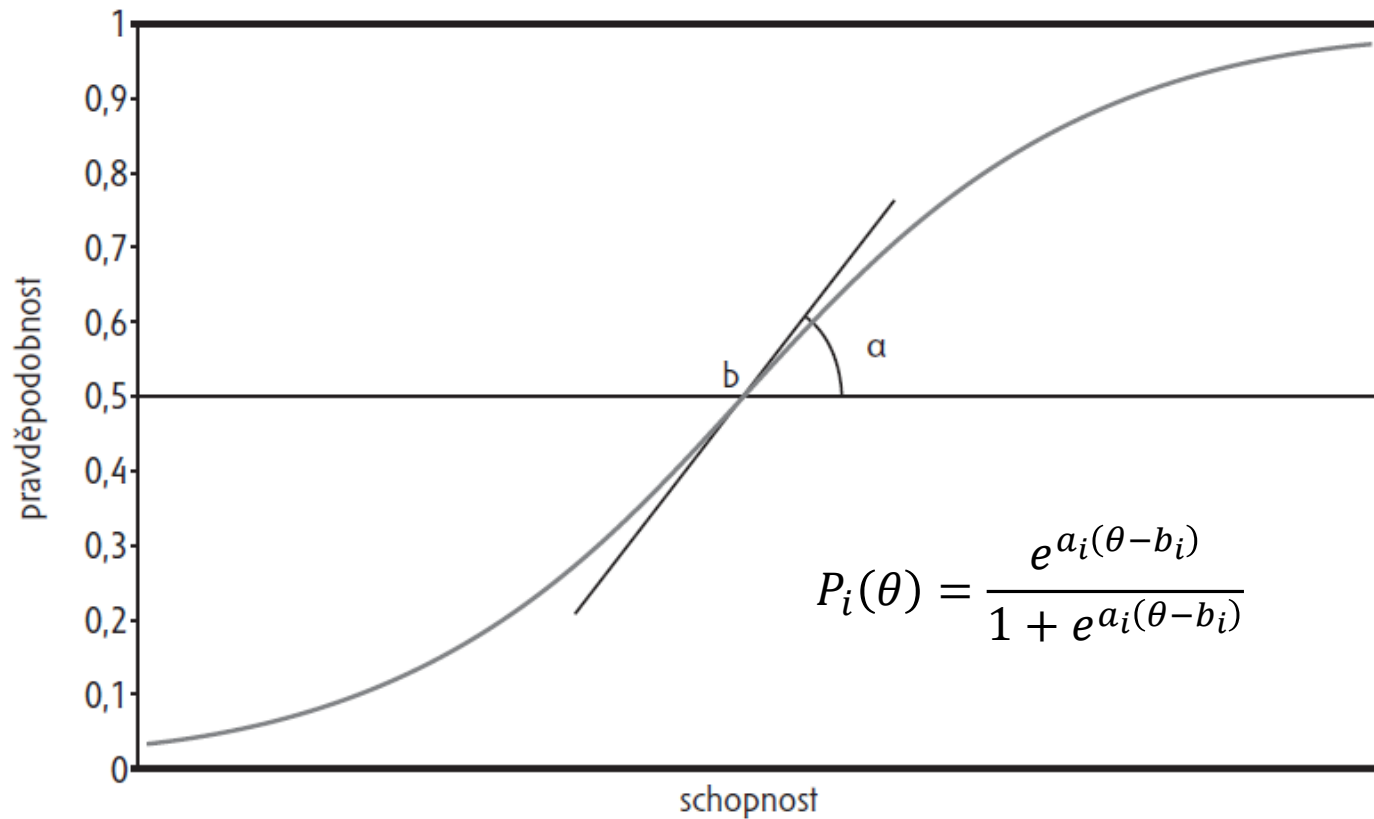
Informační funkce:

$$I_i(\theta) = a_i P_i (1 - P_i)$$

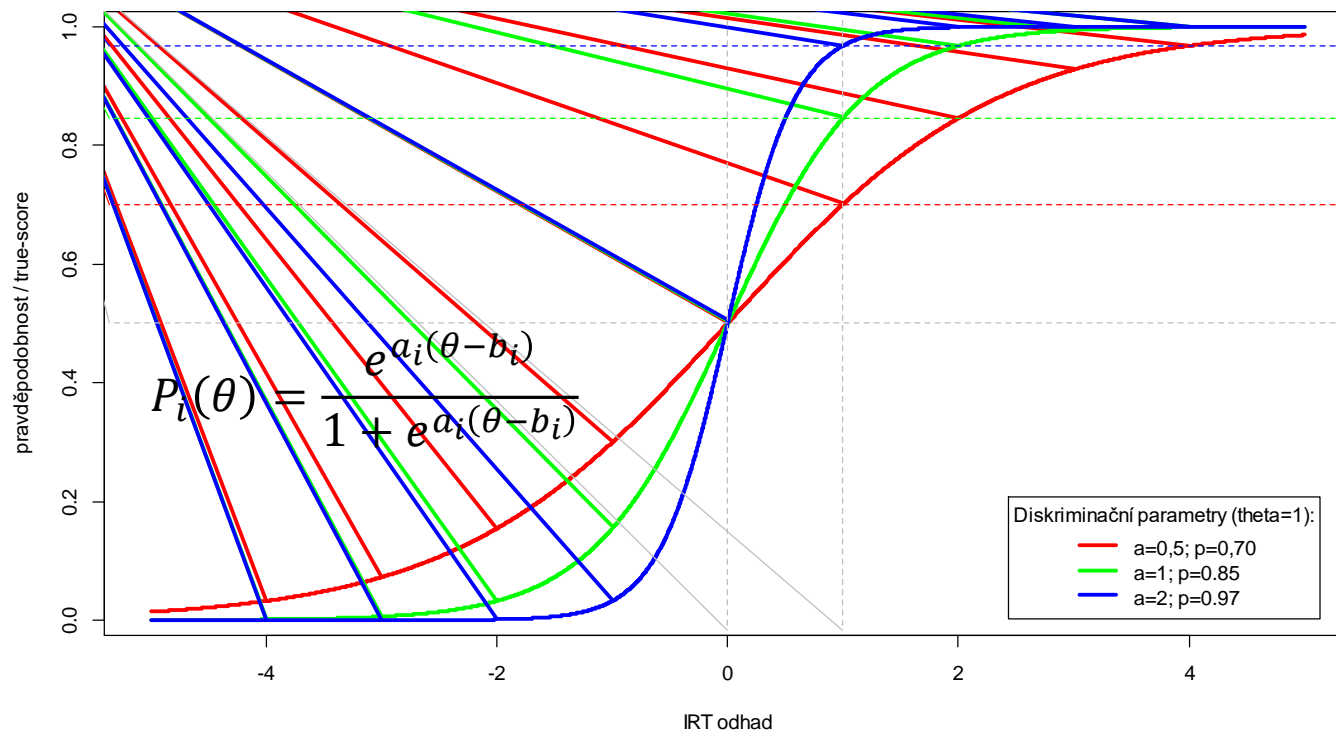
V případě 2PL modelu přestávají dávat smysl infit a outfit statistiky.

- Položka prakticky nemůže „overfitovat“.
- Používá se velké množství jiných statistik.

ICC 2PL modelu



ICC 2PL modelu



Tříparametrový model (3PL)

Přidává parametr pseudouhádnutelnosti c_i :

$$P_i(\theta) = c_i + (1 - c_i) \frac{e^{a_i(\theta - b_i)}}{1 + e^{a_i(\theta - b_i)}}$$

- c_i je spodní asymptota: pravděpodobnost správné odpovědi respondenta s extrémně malou mírou schopnosti.
- Fixace $c_i = 0$ vede k 2PL modelu.
- Ačkoliv by v zásadě mělo platit že $c_i = 1/k$ kde k je počet kategorií, je pseudouhádnutelnost modelovaná (ale u malého vzorku lze fixovat).
- Přestává platit, že obtížnost položky je bod na škále rysu, se kterým má respondent 50% pravděpodobnost správné odpovědi.

Derivace je náročnější a tedy informační funkce:

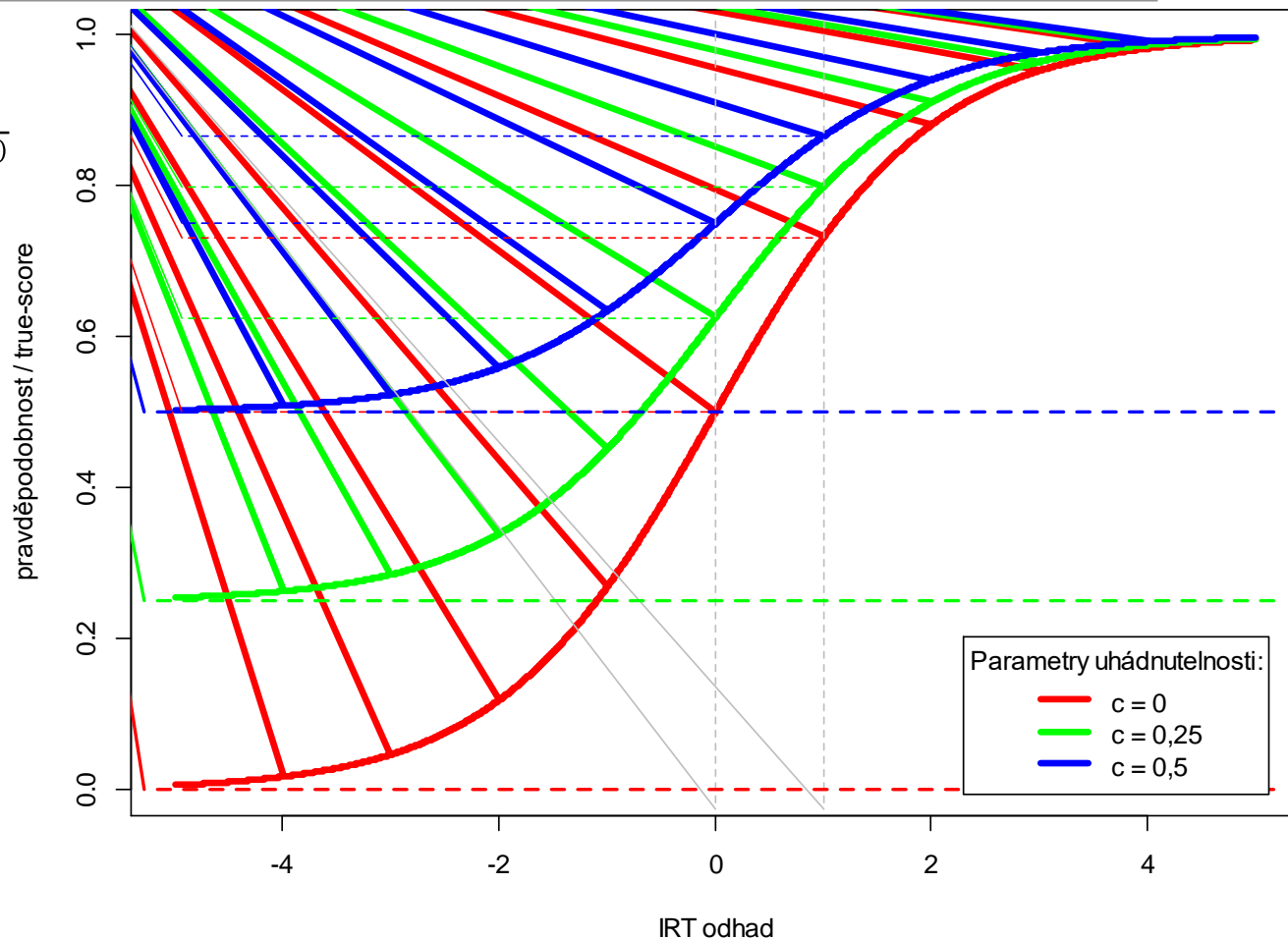
$$I_i(\theta) = a_i^2 \frac{(P_i - c_i)^2}{(1 - c_i)^2} \frac{1 - P_i}{P_i}$$

- Zároveň položka nenese maximum informace v bodě obtížnosti, ale o něco „výše“.

ICC 3PL modelu

$$P_i(\theta) = c_i + (1 - c_i) \frac{e^{a_i(\theta - b_i)}}{1 + e^{a_i(\theta - b_i)}}$$

c	P(θ=0)	P(θ=1)
0	0,5	0,73
0,25	0,63	0,80
0,5	0,75	0,87



4parametrový model (4PL)

Zavádí horní asymptotu, „ledabylost“ d_i :

$$P_i(\theta) = c_i + (d_i - c_i) \frac{e^{a_i(\theta - b_i)}}{1 + e^{a_i(\theta - b_i)}}$$

- Fixace $d_i = 1$ vede ke 3PL modelu.
- Ledabylost je pravděpodobnost správné odpovědi respondent s extrémně vysokou mírou rysu.
- Informační funkce je fakt složitá 😊

Občas se používá tzv. 3PLu model: tedy 3PL model bez dolní, zato s horní asymptotou.

- Užití ve specifických případech, např. TAT (nespuštění projektivního procesu).

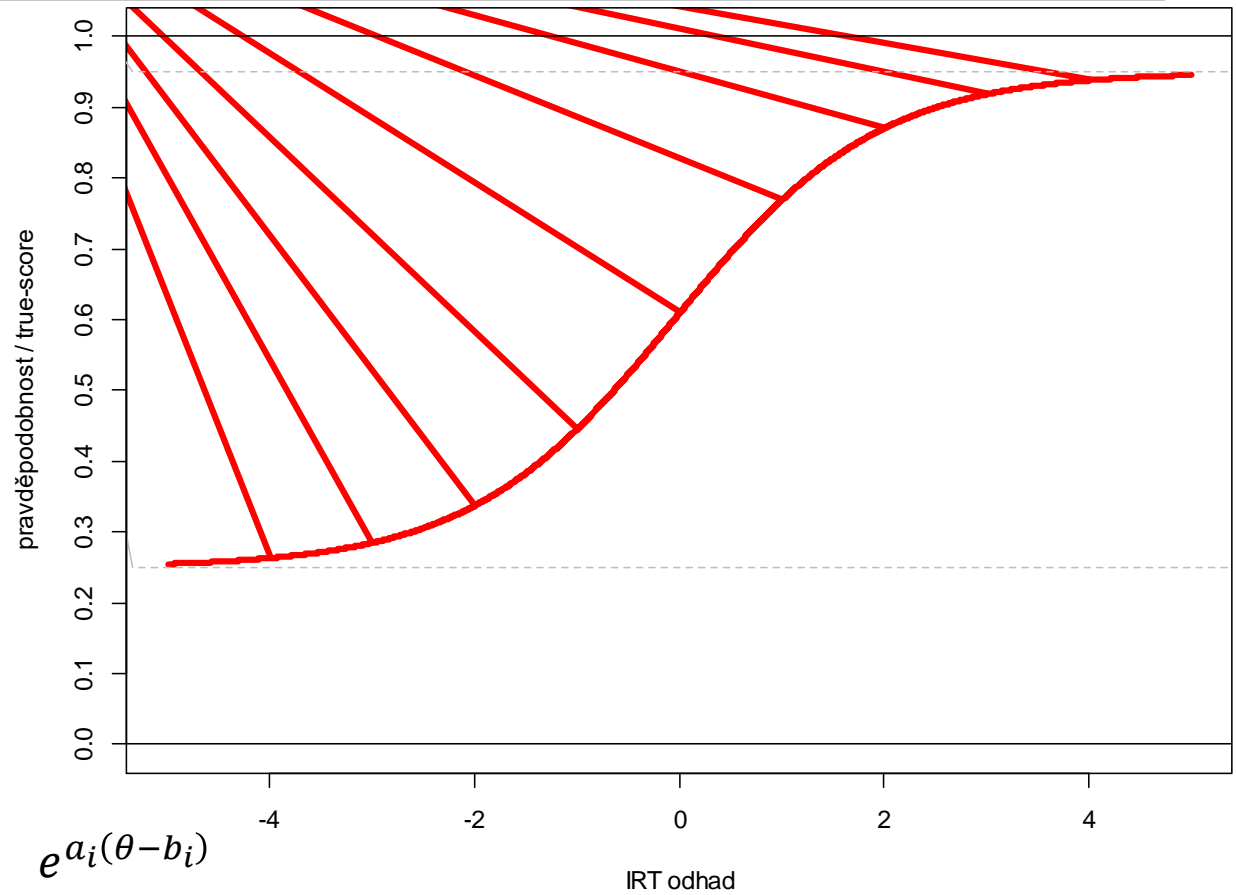
ICC 4PL modelu

▶ Parametry:

- ▶ $a = 1$
- ▶ $b = 0$
- ▶ $c = 0,25$
- ▶ $d = 0,95$

▶ True-score:

- ▶ $P_i(\theta=0)=0,61$
- ▶ $P_i(\theta=1)=0,77$



$$P_i(\theta) = c_i + (d_i - c_i) \frac{e^{a_i(\theta - b_i)}}{1 + e^{a_i(\theta - b_i)}}$$

5parametrový model (5PL)

Použití v přírodních vědách; v sociálních většinou není nutné tak přesné modelování ICC.

Zavádí parametr asymetrie e_i :

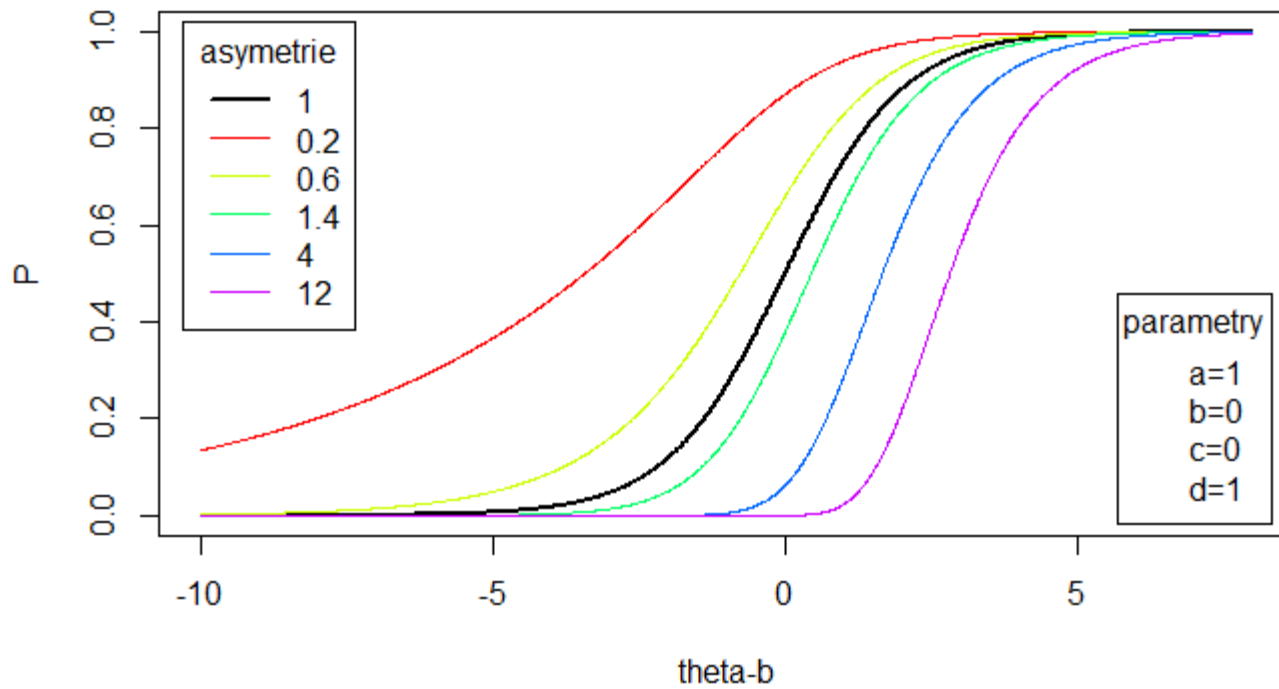
$$P_i(\theta) = c_i + (d_i - c_i) \frac{e^{a_i(\theta - b_i)^{e_i}}}{1 + e^{a_i(\theta - b_i)^{e_i}}}$$

- Fixace $e_i = 1$ vede zpět ke 4PL modelu.

Výhodnější, než zavádět tolik parametrů, je ale většinou použít nějaké neparametrické IRT, které při obdobném množství parametrů neklade takové předpoklady na průběh funkce.

- Mokken scale (neumím, neučím).
- Spline (ale může vést k paradoxní, non-monotónní ICC).
- Monotónní polynomická regrese (zpravidla s frakčními polynomy).

ICC 5PL modelu



Non-rasch vícepara- metrové IRT modely

Fumiko Samejima



<https://www.psychometricsociety.org/content/past-present-and-incoming-presidents>

Non-rasch vícepara- metrové IRT modely

Fumiko Samejima



<https://www.psychometricsociety.org/content/past-present-and-incoming-presidents>

Generalizovaný partial credit model (GPCM)

Murakiho (1992) zobecnění raschovského PCM modelu. Každý z kroků má ten stejný, avšak volně odhadovaný diskriminační parametr:

$$P(x_{ji}|\theta, a_i, \delta_{ji}) = \frac{e^{\sum_{h=0}^{x_j} a_i(\theta - \delta_{jh})}}{\sum_{k=0}^{m_j} e^{\sum_{h=0}^k a_i(\theta - \delta_{jh})}}$$

- Pokud pro všechny položky platí $a_i = 1$, pak jde o klasický raschovský PCM.

Logika je stejná, jako u PCM: tedy „postupné sbírání“ kreditu.

- Odhad místa na škále, kde je shodná pravděpodobnost odpovědi na nižší a vyšší kategorii.

Graded Response Model (GRM)

Odlišné pojetí polytomických položek, vyvinuté Samejimou (1969) jako ordinální rozšíření Thurstonovy metody stejně se jevících intervalů.

Jde o „navrstvení“ 2PL modelů za sebe, přičemž na rozdíl od GPCM modelu nejsou srovnávány 1 vyšší a 1 nižší kategorie, ale pravděpodobnost odpovědi na *jakoukoliv vyšší a jakoukoliv nižší* kategorii P_{ix}^* :

$$P_{ix}^*(\theta) = \frac{e^{a_i(\theta - \delta_{xi})}}{1 + e^{a_i(\theta - \delta_{xi})}}$$

- Kde $P_{ix}^*(\theta)$ je pravděpodobnost získat v položce i skór x nebo vyšší.

Logicky potom pravděpodobnost získat právě skór x je:

$$P_{ix}(\theta) = P_{ix}^*(\theta) - P_{i(x+1)}^*(\theta)$$

- GRM je proto tzv. dvoukrokový IRT model; v prvním kroku jsou odhadnuty $P_{ix}^*(\theta)$ a až na jejich základě lze sekundárně odvodit přímo $P_{ix}(\theta)$.

Modifikovaný GRM (MGRM)

Muraki (1990) při svém zobecňování IRT modelů zaměnil parametry kroků za prahy a obtížnost položky podobně, jako v RSM/PCM modelu.

Zavedl prahy c_i , jejichž průměr je nula, a původní krok je funkcí obtížnosti položky a prahu: $\delta_{xi} = c_i + \tau_i$.

MGRM potom:

$$P_{ix}^*(\theta) = \frac{e^{a_i[\theta - (b_i - c_j)]}}{1 + e^{a_i[\theta - (b_i - c_j)]}}$$

GRM vs. PCM

Modely jsou velmi podobné.

- V GRM jsou ale prahy „násilně“ seřazené.

Přestože predikované pravděpodobnosti a výsledky jsou velmi podobné, logika je diametrálně odlišná.

- Partial credit: Série navazujících kroků/znalostí nutných pro získání vyššího „skóre“.
 - Musím získat 1 bod, abych mohl získat 2 body; musím získat 2 body, abych mohl získat 3 body...
- Graded response: latentní kontinuum je rozčleněné na dílčí binární 2PL modely.
 - Určí se pravděpodobnost překročení každého ze „stupňů“ separátně a ty se pak „složí“ dohromady.

Nominal Response Model (NRM)

Bock (1972): Obecný model pro položky s více odpověďmi, které nejsou (nemusí být) ordinálně seřazené:

$$P_{ix}(\theta) = \frac{e^{a_{ix}\theta + c_{ix}}}{\sum_{x=0}^m e^{a_{ix}\theta + c_{ix}}}$$

- kde z důvodů identifikace pro každou položku položku platí $\sum a_{ix} = \sum c_{ix} = 0$.
- Každý práh x položky i má tedy vlastní diskriminační koeficient a_{ix} a vlastní obtížnost c_{ix} .

NRM je zobecněným GPCM modelem:

$$a_{ix}\theta + c_{ix} = a_{ix} \left(\theta + \frac{c_{ix}}{a_{ix}} \right)$$

- Zlomek v závorce lze vyjádřit jako $\frac{c_{ix}}{a_{ix}} = \delta_{ix}$, kde δ_{ix} je tranzitivní parametr GPCM modelu. Pak se NRM a GPCM liší tím, že NRM má zvláštní diskriminační parametr pro každý práh separátně.

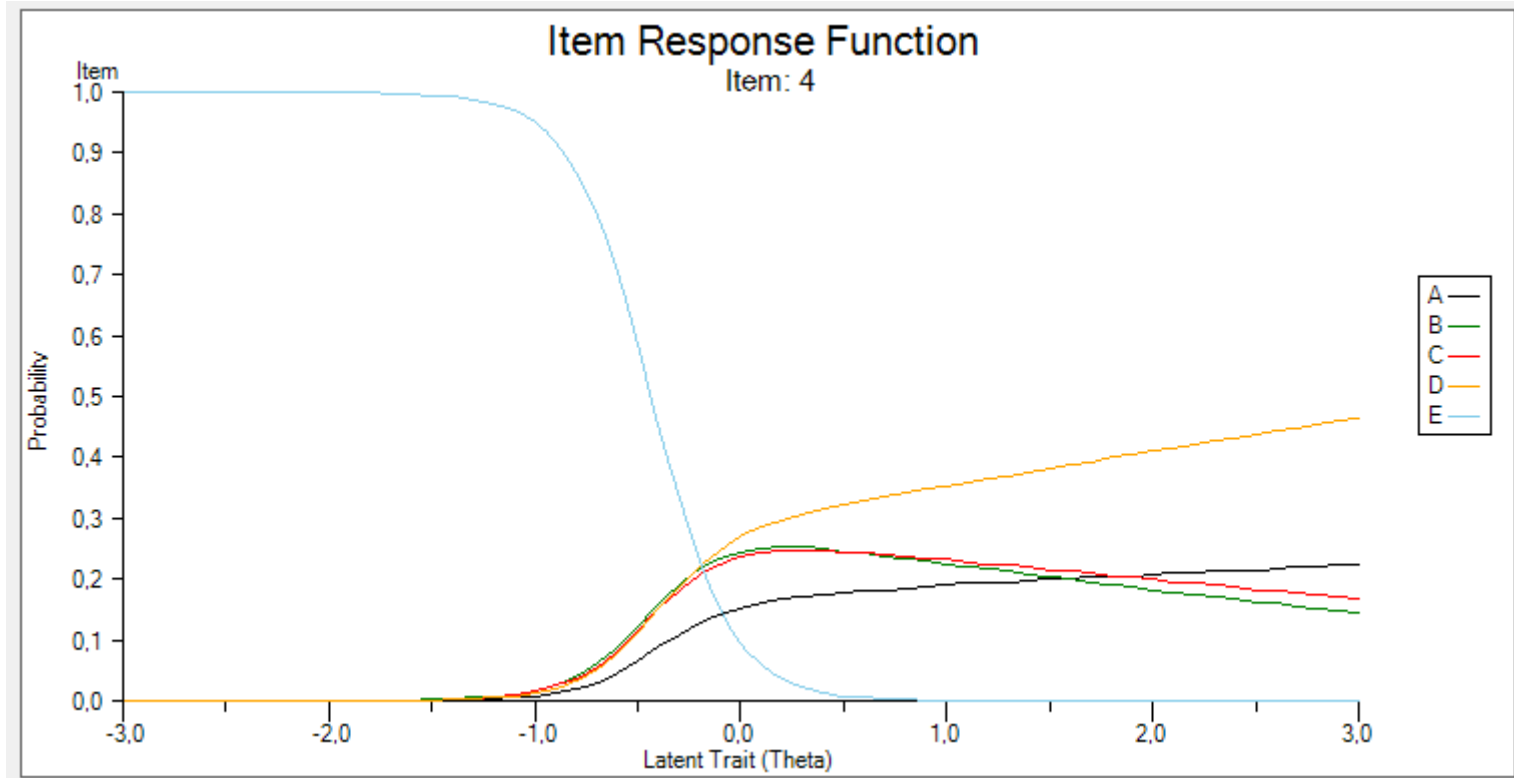
Nominal Response Model (NRM)

Vhodný pro multiple-choice testy (s jednou správnou) či výběr z odpovědí, kdy každá má jiný vztah s latentním rysem (rysy).

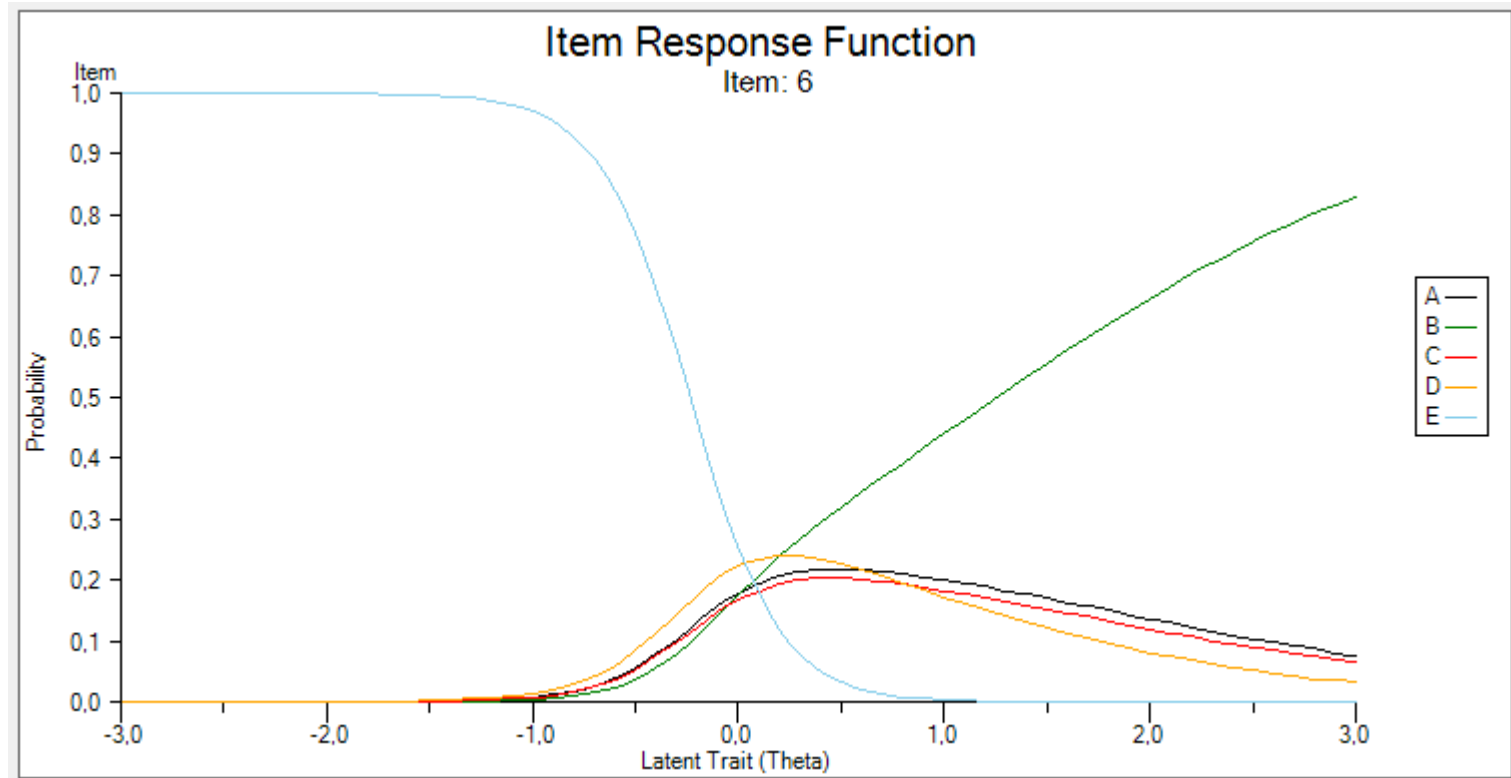
- Výhodou je, že jsou pro odhad lokálního rysu využity i konkrétní typy chybných odpovědí (ale má zase více parametrů...).
- Nevýhodou je, že zpravidla jedna kategorie se stává dominantní u velmi nízkých hodnot latentního rysu (volba tedy není náhodná). Existují různé úpravy, které to řeší.

Lze ale použít i pro dotazníková data (neobsahuje předpoklad o konkrétním seřazení položek), zde ale nemá mnoho výhod.

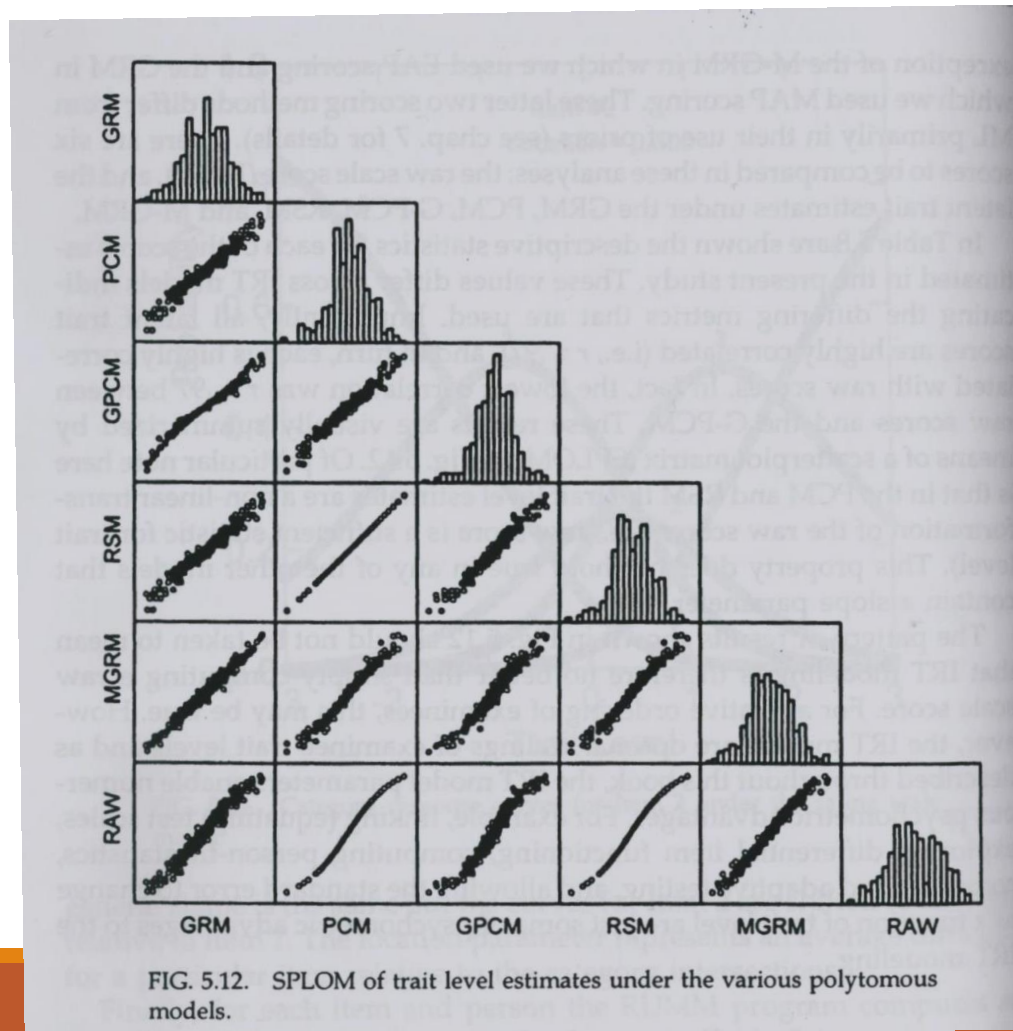
Nominal Response Model (NRM)



Nominal Response Model (NRM)



Srovnání polytomních IRT modelů



Embretson a Reise, 2009

Estimátory

Klíčovým parametrem IRT modelů je použitý estimátor. Každý má jiné předpoklady o datech a své nevýhody.

Joint Maximum Likelihood (JMLE), resp. Weighted Mean Likelihood (WMLE)

- Použitelný pouze pro Raschův model (1PL IRT model).
- Parametry položek i respondentů jsou odhadovány jako fixed, žádné předpoklady o jejich rozložení (fundamentální měření!).
- Iterativní „ping-pong“ odhadů, kdy jsou odhadovány společně na střídačku parametry osob a položek.

Marginal Maximum Likelihood (MMLE), Conditional Maximum Likelihood (CMLE)

- CMLE neumí pracovat s missing daty (ale různé úpravy).
- Podstatné je, že parametry osob jsou zvažovány jako random; je zde tedy předpoklad normálního rozdělení latentního rysu.
- V případě 1PL modelu jde tedy o generalizovaný mixed model (GLMM); v případě 2PL modelu o mixture model s random parametry.

A další 😊

Odhad latentního rysu

JMLE: Parametry osob i položek jsou přímo součástí modelu a není je nutné odhadovat poté, co je odhadnut samotný model.

- V tomto případě jsou ale odhady osob zatíženy chybou; jde o analogii pozorovaného skóru v CTT. Nezahrnují „regresi k průměru“ v důsledku chyby měření.
- Konstrukce intervalu spolehlivosti kolem tohoto odhadu proto není úplně vhodná.

MMLE, CMLE: Pouze parametry položek jsou součástí modelu; parametry osob je nutné odhadnout dodatečně.

- expected a posteriori (EAP), bayesovské maximum a posteriori (MAP):
- maximum likelihood (MLE) a weighted likelihood (WLE) – moc se nepoužívá
- expected a posteriori pro součtový skór (průměr EAP pro jednotlivé „observed patterns“ odpovědí).
- plausible values

EAP, MAP

Odhad posteriorní distribuce měřeného rysu je funkcí apriorní distribuce (tedy rozložení všech latentních rysů v populaci) a likelihood funkce.

Z posteriorní distribuce lze spočítat interval spolehlivosti a dva různé bodové odhady:

- EAP: **Průměr** posteriorní distribuce.
- MAP: Bayesovský přístup, jde o **modus** posteriorní distribuce.
- Oba vedou k podobnému závěru.

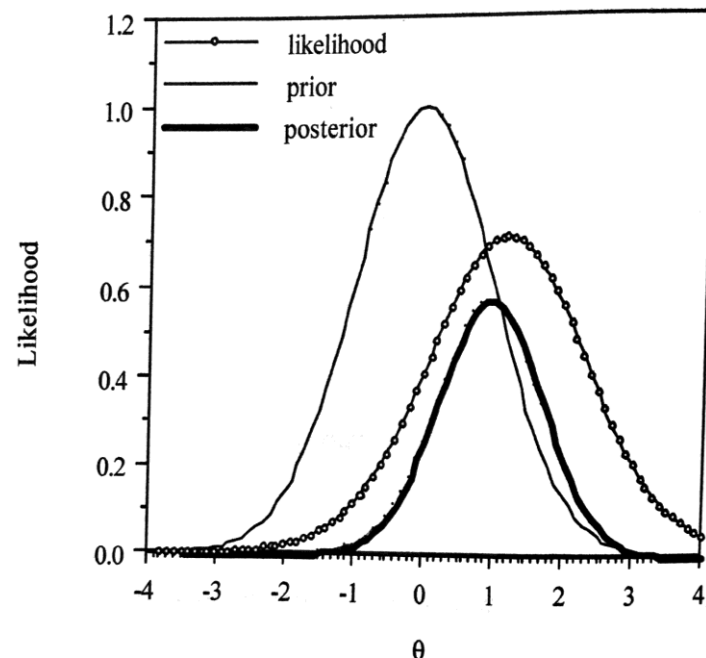


FIGURE 4.2. Conceptual presentation of the incorporation of a prior distribution into a likelihood function to produce the posterior distribution.

MLE, WLE

Odhad posteriorní pomocí funkce maximální věrohodnosti (maximum likelihood).

- `irtDemo("mle")`

MLE odhad je nepatrně zkreslený, používá se proto vážený odhad (WLE).

Plausible values (PV)

Bodové odhady latentního rysu neobsahují informaci o chybě tohoto odhadu. Z toho důvodu mohou být některé sekundární analýzy, které používají bodové odhady, zkreslené.

Z toho důvodu se někdy používají tzv. **plauzibilní hodnoty** (plausible values): náhodné vzorky z posteriorní distribuce.

- Namísto jednoho bodového odhadu pro každého respondenta je vylosováno určité množství (zpravidla pět) náhodných hodnot (PV) z posteriorní distribuce latentního rysu.
- Hodnoty nejbližší EAP/MAP mají pochopitelně nejvyšší pravděpodobnost, že budou vylosovány.
- Využití je výhradně pro sekundární analýzy nad daty exportovanými z modelu (srovnání skupin, korelace s jinými rysy atp.). PV umožňují očistit vztahy o chybu měření bez nutnosti fitovat celý model s latentními rysy (SEM, mixed model aj. přístupy).
- Zároveň vedou k vhodnějšímu ex-post odhadu distribuce latentního rysu v populaci než prosté bodové hodnoty.