

Análise Bayesiana na Teoria de Resposta ao Item Gradual

Marcos Alves dos Santos

Marcia D'Elia Branco.

O modelo de teoria de resposta ao item binário é amplamente aplicado em várias áreas, em especial na área educacional e psicológica. Contudo, comumente estamos interessados em modelos que incluam categorias intermediárias à resposta. Estas categorias podem ser ordenadas de tal forma que a categoria mais baixa contribuiria menos para o escore do indivíduo e a categoria mais alta contribuiria mais. Neste sentido, a escala de Likert é um clássico exemplo de resposta gradual em que as categorias poderiam ser definidas como: concordo completamente, concordo parcialmente, indiferente e não concordo. Sobre isso, Baker (1992) comenta que o modelo de resposta gradual tem uma importante contribuição a TRI (teoria de resposta ao item) uma vez que ele é uma extensão do modelo binário e desta forma abre precedentes a análise de escalas do tipo Likert. Como resultado, a resposta gradual pode ser modelada usando uma análise muito parecida com o caso dicotômico.

Assim, supondo que P_{ijk} é a probabilidade de no j -ésimo item ($j = 1, \dots, p$) o indivíduo i ($i = 1, \dots, n$) escolher a categoria k ($k = 1, \dots, m_j$),

$$\sum_{k=1}^{m_j} P_{ijk} = 1$$

para todo i e j e sem perda de generalidade podemos escrever $m_j = m$. Assumindo que um dado item tem 4 categorias, construímos a Curva Característica do Item (CCI) apresentado na figura 1, que nada mais é do que as probabilidades P_1, P_2, P_3, P_4 em função de uma variável latente relacionada como a habilidade do indivíduo. Podemos observar que as curvas de probabilidade não têm uma forma comum. Na categoria mais baixa $k = 1$ as probabilidades mais altas são obtidas quando a habilidade é baixa representando uma função estritamente decrescente. Já na categoria mais alta $k = 4$ ocorre o contrário. Altas probabilidades estão associadas com altas habilidades. E desta maneira, a probabilidade para esta categoria é uma função crescente. As categorias intermediárias $k = 2$ e $k = 3$ apresentam maiores probabilidades para níveis de habilidades intermediárias, tendo como característica funções unimodais.

Por causa da falta de inconsistência das curvas características é muito complicado interpretar os parâmetros a partir deste gráfico. Então, uma maneira mais fácil de interpretá-los é usar as probabilidades limites. A idéia é agrupar as 4 categorias em apenas duas e tratá-las como se fossem no caso dicotômico. Ou seja, a categoria $k = 4$ poderia ser considerada como a categoria "correta" e as outras categorias como a "incorreta". Se utilizássemos como ponto de corte $k = 3$ a categoria "correta" seria as respostas $k = 3$ ou $k = 4$ e a categoria "incorreta" as respostas 1 ou 2 e assim por diante. O interessante é que quando acumulamos as categorias das mais altas em direção as mais baixas, as curvas se tornam estritamente crescente e igual a CCI no caso da resposta binária. Estas probabilidades acumuladas foram plotadas conjuntamente na figura 2, em que,

$$\begin{aligned} P_{ij4}^+ &= 0 \\ P_{ij3}^+ &= P_{ij4} \\ P_{ij2}^+ &= P_{ij3} + P_{ij4} \end{aligned}$$

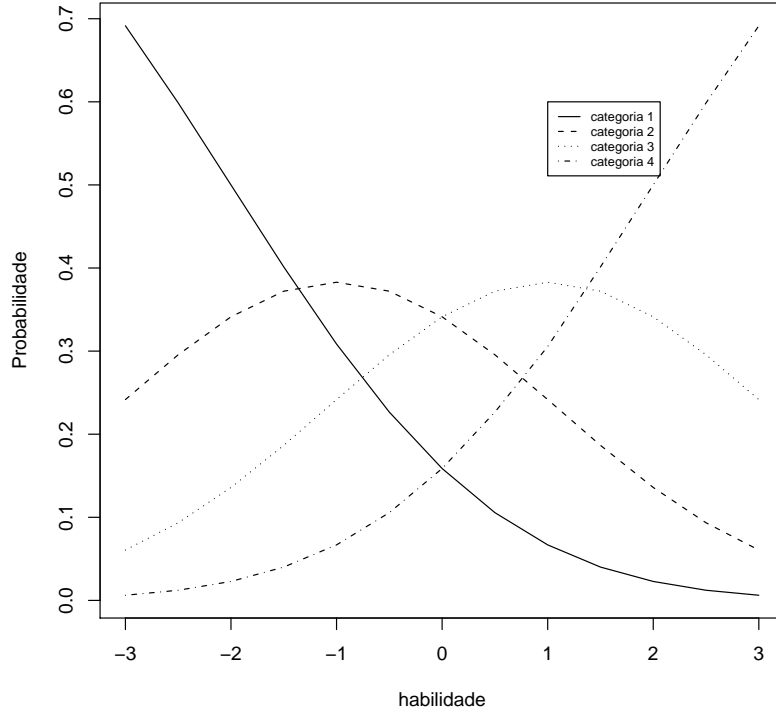


Figura 1: CCI do modelo gradual

$$P_{ij1}^+ = P_{ij2} + P_{ij3} + P_{ij4}$$

$$P_{ij0}^+ = P_{ij1} + P_{ij2} + P_{ij3} + P_{ij4} = 1$$

Em geral,

$$P_{ijk}^+ = \sum_{k'=k+1}^m P_{ijk'}$$

e

$$P_{ijk} = P_{ijk-1}^+ - P_{ijk}^+$$

Naturalmente, qualquer distribuição acumulada pode ser usada como probabilidade limite. As distribuições amplamente aplicada na TRI e em outras áreas são as distribuições ogiva normal e a logística caracterizadas respectivamente por,

$$P_{ijk}^+ = \phi(\eta_{ijk})$$

$$P_{ijk}^+ = \frac{1}{1 + e^{-\eta_{ijk}}}$$

cujo, $\eta_{ijk} = \alpha_j(\theta_i - \beta_{jk})$. Como as distribuições acumuladas anteriores representam a soma das probabilidades de resposta a k -ésima categoria, a diferença entre as curvas não assumem valores negativos. Portanto, pela figura 2, as curvas não podem se cruzar e têm que ter a mesma inclinação α_j . Ainda,

$$\beta_{j1} \leq \dots \leq \beta_{jk} \leq \dots \leq \beta_{jm-1}$$

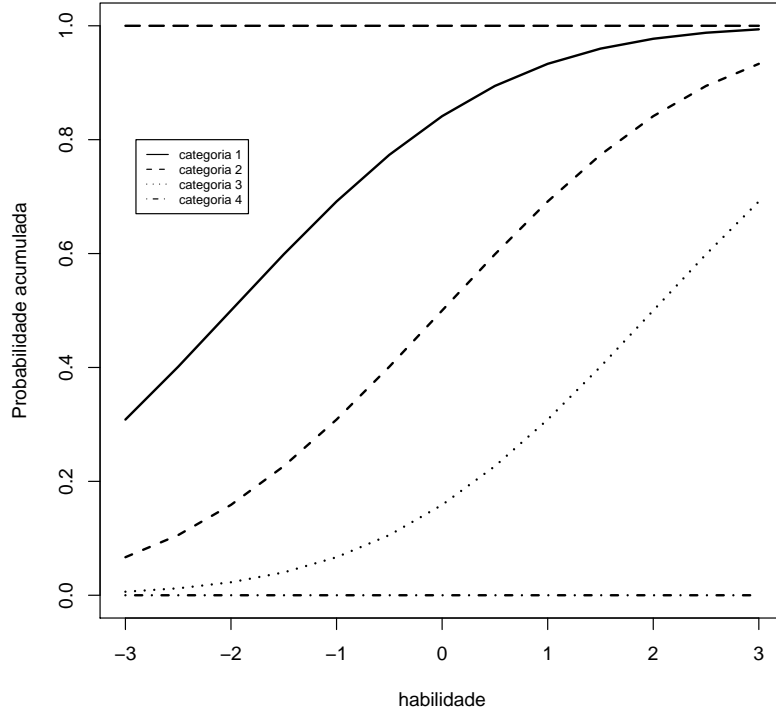


Figura 2: Limite da curva característica

com $-\infty \leq \beta_{ijk} \leq \infty$ e $\alpha_j \geq 0$. Por conta da restrição $\sum_{k=1}^{m_j} P_{ijk} = 1$, sempre haverá um parâmetro de locação a menos do que o número de categorias. Retornando à figura 2 e associando-a com o modelo dicotômico, as interpretações são praticamente as mesmas. A diferença deve-se aos parâmetros de dificuldade β_{jk} . Supondo quatro categorias de respostas: completamente certo, parcialmente certo, parcialmente errado e completamente errado os parâmetros podem ser interpretados como

- α_j é o parâmetro de discriminação. Quanto maior o seu valor maior é o poder de discriminação dos indivíduos.
- θ_i é a habilidade do indivíduo. Este parâmetro associa a capacidade do indivíduo acertar os ítems.
- β_{j1} é dificuldade em acertar completamente o item
- β_{j2} é a dificuldade em acertar no mínimo parcialmente
- β_{j3} é a dificuldade em não errar.

Uma das principais características da Teoria de Resposta ao Item é que os itens e as habilidades são variáveis latentes usualmente desconhecidas. Portanto, precisam ser estimadas. Basicamente, o foco deste trabalho é a estimação Bayesiana como uso de variáveis latentes. Esta abordagem permite que a posteriori exata seja obtida facilmente a partir do uso de prioris adequadas.

A idéia principal dos dados aumentados é introduzir variáveis latentes no modelo, em que Z_{ij} tem distribuição definida pela seguinte classe de distribuições, que de acordo

com Farias (2007) é comumente chamada de classe das distribuições obtidas através de misturas de normais em relação ao parâmetro de escala,

$$F_s = \left\{ F(\cdot) = \int_0^\infty \Phi\left(\frac{\cdot}{\delta}\right) dG(\delta) \right\} \quad (1)$$

cujo $G(\delta)$ é uma fda (função densidade acumulada) no intervalo $[0, \infty)$. Algumas das distribuições mais conhecidas dessa classe são probito, t-Student e a logística.

A introdução de variáveis latentes auxilia na obtenção de formas conhecidas das distribuições condicionais completas, o que facilita a implementação do amostrador de Gibbs. Albert (1992) baseou-se nessa idéia para estimar os parâmetros do modelo da TRI. Entretanto, o método foi desenvolvido somente para o modelo probito. Em Albert e Chib (1993) sob o contexto de regressão binária e politômica, além do modelo probito foi também utilizado a t-Student, que de acordo com os autores, se aproxima da distribuição logística quando o grau de liberdade é igual a 8. Chris e Held (2006) estenderam essa análise utilizando-se de uma mistura de normais para a variável latente. Como resultado, foi utilizado a distribuição exata da logística sem a necessidade de aproximação para uma distribuição t-Student.

A vantagem de trabalhar com a representação de dados aumentados é que com uma escolha apropriada de prioris podemos ter uma eficiente simulação usando o Gibbs. Para o vetor de habilidade θ Guzmán (2005) considera uma distribuição normal e independente com média 0 e variância 1, ou seja,

$$\theta_i \sim N(0, 1), i = 1, \dots, n$$

A escolha de $\mu = 0$ e $\sigma^2 = 1$ resolve o problema de indeterminação da variável latente e portanto o problema de identificação dos parâmetros dos itens (Guzmán (2005)). Seguindo a proposta das prioris os parâmetros dos itens $\xi = (\alpha, \beta)$ também são independentes

$$\pi(\xi) = \pi(\beta)\pi(\alpha)$$

Observe que a distribuição dos β s têm que estar na mesma escala do que a habilidade pois $\eta_{ij} = \alpha_j(\theta_i - \beta_j)$. Assim,

$$\beta_j \sim N(0, 1), j = 1, \dots, p$$

Porém, como $\alpha_j > 0$ a priori do item de discriminação tem densidade normal truncada no zero,

$$\alpha_j \sim N(\mu_\alpha, \sigma_\alpha^2) I_{(0, \infty)}, j = 1, \dots, p$$

Considerando o modelo probito, obtido quando a classe de distribuições definida em 1 δ_i tem distribuição degenerada e utilizando as prioris previamente discutidas, podemos representar o modelo de resposta gradual como

$$\begin{aligned} Y_{ij} &= k \text{ se } \gamma_{k-1} \leq Z_{ij} \leq \gamma_k \\ Z_{ij} | \delta_i = 1 &\sim N(\phi_{ij}; 1) \\ \theta_i &\sim N(0, 1), i = 1, \dots, n \\ \beta_j &\sim N(0, 1), j = 1, \dots, p \\ \alpha_j &\sim N(\mu_\alpha, \sigma_\alpha^2) I(\alpha_j > 0), j = 1, \dots, p \\ (\gamma_2, \dots, \gamma_{m-1}) &\propto \text{Uniforme} \end{aligned}$$

Nessa nova representação por dados aumentados a variável Y_{ij} fica totalmente condicionada a variável latente sendo limitada pelas variáveis $\gamma = (\gamma_0, \dots, \gamma_m)$ com $\gamma_0 = -\infty$, $\gamma_1 = 0$ e $\gamma_m = \infty$. Logo, a verossimilhança aumentada é dada por

$$L(\alpha; \beta; \theta | \mathbf{Z}; \mathbf{Y}) = \prod_{i=1}^n \prod_{j=1}^p \Phi(z_{ij} - \eta_{ij}) \left[\sum_{k=1}^m I(Y_{ij} = k) I(\gamma_{k-1} < Z_{ij} < \gamma_k) \right]$$

Apesar de a distribuição a posteriori ainda não ser tratável, com o auxílio das variáveis latentes é possível determinar as posteriores condicionais, que são proporcionais a

$$\pi(\alpha; \beta; \theta; \gamma | \mathbf{Z}; \mathbf{Y}) \propto L(\alpha; \beta; \theta | \mathbf{Z}; \mathbf{Y}) \pi(\alpha; \beta; \theta; \delta)$$

Desta forma, a distribuição condicional a posteriori são obtidas facilmente

$$\begin{aligned} Z_{ij} | Y, \theta, \alpha, \beta &\sim N(\eta_{ij}, 1) \left[\sum_{k=1}^m I(Y_{ij} = k) I(\gamma_{k-1} < Z_{ij} < \gamma_k) \right] \\ \theta_i | Z, \alpha, \beta &\sim N \left(\frac{\sum_{j=1}^p \alpha_j (z_{ij} + \alpha_j \beta_j)}{1 + \sum_{j=1}^p \alpha_j^2}; \frac{1}{1 + \sum_{j=1}^p \alpha_j^2} \right) \\ \beta_j | Z, \alpha, \theta &\sim N \left(\frac{\sum_{i=1}^n \alpha_j (\alpha_j \theta_i - z_{ij})}{1 + n \alpha_j^2}; \frac{1}{1 + n \alpha_j^2} \right) \\ \alpha_j | Z, \beta, \theta &\sim N \left(\frac{\sum_{i=1}^n \sigma_{\alpha}^2 (\theta_i - \beta_j) + \mu_{\alpha}}{1 + n \sigma_{\alpha}^2 \sum_{i=1}^n (\theta_i - \beta_j)^2}; \frac{\sigma_{\alpha}^2}{1 + \sigma_{\alpha}^2 \sum_{i=1}^n (\theta_i - \beta_j)^2} \right) \end{aligned}$$

A distribuição condicional completa de γ_t dado $(Y, Z, \gamma_0, \dots, \gamma_{t-1}, \gamma_{t+1}, \dots, \gamma_m)$ é proporcional a

$$\begin{aligned} &\prod_{i=1}^n \prod_{j=1}^p \left[\sum_{k=1}^m I(Y_{ij} = k) I(\gamma_{k-1} < Z_{ij} < \gamma_k) \right] \\ &= \prod_{i=1}^n \prod_{j=1}^p [I(Y_{ij} = t) I(\gamma_{t-1} < Z_{ij} < \gamma_t) + I(Y_{ij} = t+1) I(\gamma_t < Z_{ij} < \gamma_{t+1})] \end{aligned}$$

Obrigatoriamente γ_t tem que ser maior do que todos os Z_{ij} para os quais $Y_{ij} = t$ e não pode ser menor do que γ_{t-1} . Ou seja, γ_t sempre será maior que

$$\max \{ \max \{ Z_{ij} : Y_{ij} = t \}; \gamma_{t-1} \} \quad (2)$$

Além disso, γ_t tem que ser menor do que qualquer Z_{ij} quando $Y_{ij} = t+1$ e também menor do que γ_{t+1} . Isto quer dizer que γ_t é limitado superiormente por

$$\min \{ \min \{ Z_{ij} : Y_{ij} = t+1 \}; \gamma_{t+1} \} \quad (3)$$

Desta forma, podemos ver que a distribuição condicional é uma uniforme com limite inferior e superior dado respectivamente por 2 e 3.

Neste trabalho, estudamos o modelo de resposta gradual ao item que pode ser visto como uma extensão do modelo de resposta ao item binário. Apesar desse modelo não apresentar uma posteriori conhecida podemos ver que com a inclusão de variáveis auxiliares a estimação dos parâmetros dos itens e das habilidades se tornam simples uma vez que as condicionais completas são facilmente obtidas e assim o algoritmo de Gibbs pode ser implementado.

Referências

Albert, J. H. *Bayesian Estimation of Normal Ogive Item Response Curves Using Gibbs Sampling*. 1992. *Journal of Educational Statistics*, Vol. 17, No 3, 251-269 p.

Albert, J. H.; Chib, S. *Bayesian Analysis of Binary and Polychotomous Response Data*. 1993. *Journal of the American Statistical Association*, Vol 88, No 422, 669-679 p.

Baker, F. B. *Item Response Theory: Parameter estimation techniques*. New York: M Dekker, 1992. 440p.

Farias, R. B. A. *Regressão Binária Bayesiana com Uso de Variáveis Auxiliares*. 2007. 89 p.

Guzmán, J. L. B. *Uma Família de Modelos de Resposta Ao Ítem Normal Assimétrica*. 2005. 117 p.

Holmes, C. C.; Held, L. *Bayesian Auxiliary Variable Models for Binary and Multinomial Regression*. 2006. *International Society for Bayesian Analysis*, Vol 1, No 1, 145-168 p.