

# Modelagem Bayesiana e Aplicações

Márcia D'Elia Branco

Universidade de São Paulo  
Instituto de Matemática e Estatística  
<http://www.ime.usp.br/~mbranco>

## Modelos de Regressão 3

## Exemplo 3.1. Consumo permanente e renda.

- Foram analisados dados de  $n = 26$  países, apresentados originalmente em Zellner e Moulton (1985).
- A variável resposta  $y$  é uma transformação da variável consumo permanente ( $c$ ).
- A variável explicativa é o logaritmo da renda ( $x$ ) .
- O modelo assumido é linear simples

$$y_i = \beta_0 + \beta_1 \log(x_i) + \epsilon_i \quad , \quad \epsilon_i \sim N(0, \sigma^2).$$

- Foi considerada a priori não informativa  $h(\beta, \log(\sigma)) \propto 1$ .

# Regressão linear normal: exemplo

- O principal objetivo do exemplo é verificar o comportamento de algumas medidas de diagnóstico de adequação do modelo.
- O programa escrito em BUGS é apresentado em notas no final do capítulo.
- Análise de presença de valores discrepantes. Malta e Japão foram claramente identificados como discrepantes, usando a medida

$$P(Y_{rep, Malta} > y_{obs, Malta} | y) = 0.99.$$

- Verificando a assimetria. Foi obtida a proporção de vezes que o coeficiente de assimetria é maior que zero, resultando no valor 0.14. Isso indica assimetria negativa. A medida utilizada é

$$\frac{1}{T} \sum_{t=1}^T I \left( \frac{1}{n} \sum_{i=1}^n \epsilon_i^{3(t)} / \sigma^{3(t)} > 0 \right).$$

# Regressão linear normal: exemplo

- Também foi considerada uma medida global de discrepância que compara a função empírica dos erros com a função acumulada da normal. Assim

$$D(y_{obs}, \theta) = \sum_{i=1}^n \frac{\hat{F}_i - F_i^{exp}}{F_i^{exp}(1 - F_i^{exp})}.$$

- Em que  $F_i^{exp} = \Phi(\epsilon_i^p)$  e

$$\hat{F}_i = \frac{1}{n} \sum_{k \neq i} I(\epsilon_k^p \leq \epsilon_i^p).$$

- A mesma medida é calculada para cada uma das  $T$  réplicas .

# Regressão linear normal: exemplo

- O valor-P é obtido fazendo

$$\frac{1}{T} \sum_{t=1}^T I \left( D(y_{rep}^{(t)}, \theta^{(t)}) > D(y_{obs}, \theta^{(t)}) \right).$$

- Resultando num valor 0.29 , o qual não indica falha do modelo normal.

## Exemplo 3.2. Aprovação dos estudantes.

- O conjunto de dados originalmente analisados em Johnson and Albert (1999), consiste do resultado  $y_i = 1$  aprovado e  $y_i = 0$  reprovado,  $i = 1, \dots, 30$ .  $p_i = P(y_i = 1)$ .
- Como variável explicativa foi considerado o escore SAT-M centrada, isto é, o valor do escore menos a média.
- Foi ajustado o seguinte modelo

$$\text{logito}(p_i) = \beta_0 + \beta_1(x_i - \bar{x}).$$

- O programa escrito em BUGS está nas notas do final do capítulo.

- Um dos objetivos da análise foi comparar o cálculo da  $CPO_i$  usando a definição e sua aproximação dada por

$$G_i = \left[ \frac{1}{T} \sum_{t=1}^T \frac{1}{L_i(\theta^{(t)})} \right]^{-1}$$

- em que  $\theta^{(t)}$  são simuladas da posteriori completa.
- Lembre que para obtenção da verdadeira  $CPO_i$  é necessário obter amostras das posteriores condicionais a  $y_{[-i]}$  (dados sem a  $i$ -ésima observação)

$$CPO_i = f(y_i \mid y_{[-i]}).$$

# Regressão binária: exemplo

- Uma outra análise feita no exemplo foi a avaliação da precisão de classificação.
- Inicialmente, vamos definir sensibilidade e especificidade de um classificador.
- Considere a seguinte regra de classificação: se  $\hat{p}_i > \pi^T$  o indivíduo é classificado no grupo 1 (positivo); caso contrário, é classificado no grupo 0 (negativo). Em que  $\pi^T$  é um limiar previamente fixado.
- Note que verificar se  $\hat{p}_i = F(\eta_i) > \pi^T = F(\eta^T)$  é equivalente a verificar se  $\eta_i > \eta^T$  devido a  $F$  ser estritamente crescente.
- Se tivermos uma amostra de treinamento e uma amostra de teste, podemos comparar a classificação feita com os verdadeiros valores observados.



# Regressão binária: exemplo

Abaixo ilustramos uma tabela de classificação

Real valor	Classifica 1	como 0	Total
1	$n_{11}$	$n_{01}$	$n_1$
0	$n_{01}$	$n_{00}$	$n_0$

A *sensibilidade* de um classificador é dada pela proporção de classificação correta positiva, isto é,  $n_{11}/n_1$

A *especificidade* de um classificador é dada pela proporção de classificação correta negativa, isto é,  $n_{00}/n_0$ .

# Regressão binária: exemplo

- No exemplo, foram considerados os seguintes limiares para  $\eta^T$  :  $-1.0, -0.8, -0.6, \dots, 0.6, 0.8, 1.0$
- As medidas de sensibilidade e especificidade foram calculadas com base em validação cruzada e nos valores médios de  $n_{11}$  e  $n_{00}$  para as várias réplicas de MC.
- Nota-se que a sensibilidade decai e a especificidade cresce, conforme aumenta o limiar.
- Como medida resumo de precisão do classificador foi usada *Sensibilidade + Especificidade*. O valor mais alto dessa medida foi observado para  $\eta^T = -0.4$  .
- O limiar ótimo em termos de probabilidade é

$$\pi^T = F(-0.4) = \frac{e^{-0.4}}{1 + e^{-0.4}} = 0.4$$

# Regressão binária: dados aumentados

- Modelos de regressão binários, ordinais e multinomiais podem ser simplificados usando um conjunto de variáveis latentes contínuas que permitam a implementação do algoritmo de dados aumentados.
- Conforme já discutido em exemplo da aula8 desta disciplina, para a ligação probito definimos

$$z_i = x_i^t \beta + u_i \quad u_i \sim N(0, 1) \text{ ind..}$$

- Considere  $y_i = I(z_i > 0)$  , então  $y_i$  segue uma Bernoulli com probabilidade  $p_i = \Phi(x_i^t \beta)$  .
- Este esquema pode ser usado para outras ligações substituindo a distribuição dos erros  $u_i$  de forma adequada. Na regressão logito, usar a distribuição logística.

# Regressão binária: dados aumentados

- Para o modelo probito com priori  $\beta \sim N(b_0, V_0)$ , as condicionais completas são

$$\beta \mid y, z \sim N(b_n, V_n)$$

$$z_i \mid \beta, y_i = 1 \sim N(x_i^t \beta, 1) I(0, \infty)$$

$$z_i \mid \beta, y_i = 0 \sim N(x_i^t \beta, 1) I(-\infty, 0)$$

- Portanto, um algoritmo de simulação do tipo GS pode ser implementado.
- No BUGS pode ser usada a função  $y[i] \sim dbern.aux(z[i])$  e  $z[i] \sim dnorm(eta[i], 1)$ .

- Com o uso da abordagem de dados aumentados podemos definir resíduos latentes, que tem a vantagem de ter distribuição conhecida.
- Assim  $u_i = z_i - x_i^t \beta$  tem distribuição  $N(0, 1)$  no modelo probito e Logística no modelo logito.
- Com o esquema de amostragem de MC é fácil controlar esses resíduos . Para o modelo probito, por exemplo,

$$P(|u_i| > 1.96 | y)$$

pode ser comparada com a probabilidade a priori que é 0.05.

# Reversible Jump MCMC

- RJMCMC é um algoritmo computacional usado para seleção de modelos.
- A grande diferença desta metodologia em relação aos outros algoritmos MCMC já estudados, é a possibilidade de transitar por espaços paramétricos de diferentes dimensões.
- Considere  $\{M_k, k = 1, \dots, K\}$  o conjunto de possíveis modelos e  $\Theta_k$  o espaço paramétrico associado ao modelo  $M_k$ , cuja dimensão é  $d_k$ .
- O espaço conjunto de parâmetros que o algoritmo deve alcançar é dado por  $\cup_k (M_k \times \Theta_k)$ .
- A distribuição alvo é dada por

$$h(M_k, \theta_k \mid y) \propto f(y \mid M_k, \theta_k) h(\theta_k \mid M_k) h(M_k) = h^*(M_k, \theta_k)$$

- O processo de transição de  $(M_k, \theta_k)$  para  $(M_k^*, \theta_k^*)$  envolve um algoritmo de MH com dois tipos de propostas.
- A primeira associada a escolha do modelo, denotada por  $J(k, k^*)$ .
- A segunda associada aos parâmetros do modelo, denotada por  $q(\theta_{k^*} \mid \theta_k, M_k, M_{k^*})$ .
- O problema reside na escolha desta segunda proposta, considerando o fato de a dimensão  $d_k \neq d_{k^*}$ .

Por exemplo, considere

$M_1$  o MRL simples sem o intercepto com  $\theta_1 = \beta_1$  e  
 $M_2$  o MRL usual com  $\theta_2 = (\beta_0, \beta_1)$  .

Se estamos em  $M_1$  só temos amostras de um dos parâmetros, como adicionar  $\beta_0$  ?

- Uma importante estratégia do algoritmo é o dito "preenchimento" da dimensão. Isto é feito com o uso de uma variável auxiliar  $u \sim q(u)$  .
- O segundo ponto chave do algoritmo é o uso de uma transformação bijetora  $T$  tal que  $T(\beta_1, u) = (\beta_0, \beta_1)$  e vice-versa.



*Move up* :  $(M_1, \theta_1) \rightarrow (M_2, \theta_2)$

- 1 Com probabilidade  $q_{12}$  escolha mover para  $M_2$ . Se o movimento não for aceito, continue em  $M_1$  e simule uma nova amostra de  $\theta_1$ . Caso contrário,
- 2 Simular  $u \sim q(u)$  e fazer  $\theta_2^* = (\beta_0^*, \beta_1^*) = T(\beta_1, u)$ .
- 3 Calcular o jacobiano da transformação

$$J = \frac{dT(\beta_1, u)}{d\beta du}$$

- 4 Determine a probabilidade de aceitação para esse valor proposto, dada por  $\min\{1, \rho_{up}\}$  com

$$\rho_{up} = \frac{h^*(M_2, \theta_2^*)}{h^*(M_1, \theta_1)} \frac{q_{21}}{q_{12}q(u)} |J|$$

*Move down* :  $(M_2, \theta_2) \rightarrow (M_1, \theta_1)$

- 1 Com probabilidade  $q_{21}$  escolha mover para  $M_1$ . Se o movimento não for aceito, continue em  $M_2$  e simule uma nova amostra de  $\theta_2$ . Caso contrário,
- 2 Simular  $u \sim q(u)$  e fazer  $\theta_1^* = \beta_1^*$  com  $(\beta_1^*, u) = T^{-1}(\theta_0, \theta_1)$ .
- 3 Determine a probabilidade de aceitação para esse valor proposto, dada por  $\min\{1, \rho_{up}\}$  com

$$\rho_{down} = \frac{h^*(M_1, \theta_1^*)}{h^*(M_2, \theta_2)} \frac{q_{12} q(u)}{q_{21}} |J|^{-1}$$

- 4 Se o movimento for aceito fazer  $\theta_1 = \theta_1^*$ .

- Green, P.J. (1995). RJMCMC computation and Bayesian model determination. *Biometrika*, 82, 711-732.
- Lunn, D.J., Best, N. and Whittaker, J. (2009). Generic RJMCMC using graphical models. *Statistics and Computing*, 19, 395-408.
- Funções no BUGS para implementar RJMCMC em seleção de variáveis em modelos lineares: *jump.lin.pred()* e *jump.model.id()*.
- Ver Exemplo 3.3 do livro do Congdon (2014).

# Modelo de Regressão multinomial

- Considere agora  $K$  possíveis categorias de respostas.
- $X_i$  é o vetor de covariáveis associado a  $i$ -ésima unidade amostral.
- $y_{ij} = 1$  se o  $i$ -ésima unidade esta na categoria  $j$  e  $y_{ik} = 0$  para  $k \neq j$ .
- $p_{ij} = P(y_{ij} = 1)$  . Vários tipos de logitos podem ser definidos.
- Um dos mais populares é o logito categoria de referência

$$\text{logito}R_{ij} = \log \left( \frac{p_{ij}}{p_{i1}} \right)$$

# Modelo de Regressão multinomial

- O modelo de regressão é dado por

$$\text{logito} R_{ij} = \alpha_j + X_i^t \beta_j \quad j = 2, \dots, K.$$

- Uma reta de regressão para cada uma das  $K - 1$  categorias.
- As probabilidades são obtidas por

$$p_{ij} = \frac{e^{\eta_{ij}}}{1 + e^{\eta_{ij}}} \quad j = 2, \dots, K$$

$$p_{i1} = \frac{1}{1 + e^{\eta_{ij}}}$$

- em que  $\eta_{ij}$  é o preditor linear.