Modelagem Bayesiana e Aplicações

Márcia D'Elia Branco

Universidade de São Paulo
Instituto de Matemática e Estatística
http:www.ime.usp.br/m̃branco

Modelos de Regressão 2

Regressão linear normal: diagnóstico

O modelo linear normal homocedástico

$$y = X\beta + \epsilon$$

$y = (y_1, \dots, y_n)^t$ é um vetor aleatório $n \times 1$ de variáveis respostas, X uma matrix de valores fixados $n \times p$ (variáveis explicativas) e $\beta = (\beta_1, \dots, \beta_p)^t$ é o vetor de parâmetros de dimensão $p \times 1$.

Além disso,

$$\epsilon \mid \sigma^2 \sim N_n(0, \frac{1}{\gamma} I_n)$$

condicionamente independentes.

$$y_i = x_i^t \beta + \sigma \epsilon_i^p \quad , \quad \epsilon_i^p \sim N(0,1).$$

Análise de resíduos:

$$\epsilon_i^p = \frac{y_i - x_i^t \beta}{\sigma} \quad i = 1, \ldots, n$$

- A partir da amostra simulada da distribuição a posteriori $h(\beta, \gamma \mid y)$, obtemos uma amostra dos resíduos padronizados.
- Podemos construir um *qq-plot* (gráfico de quantis-quantis) para analisar o comportamento dos resíduos.
- Também podemos monitorar $I(|\epsilon_i^p| > 1.96)$.
- Outras medidas de diagnóstico inclue o monitoramento de medidas de assimetria e curtose dos resíduos. Sob normalidade, é esperado uma assimétria próxima de zero e curtose próximo de 3 .

Regressão linear normal: diagnóstico

- Medida de assimetria:
$$\frac{(y_i - x_i^t\beta)^3}{\sigma^3}.$$
- Medida de curtose:
$$\frac{(y_i - x_i^t\beta)^4}{\sigma^4}.$$

- Essas medidas e algumas outras são implementadas no Exemplo 3.1, pag. 101 de Congdon (2014).
- Os códigos de programa estão nas Notas do capítulo 3 do livro.

A distribuição *a priori* proposta por Zellner para os coeficientes regressores, usa o fato de conhecermos as covariáveis X.

$$\beta \mid g, \gamma \sim N_p\left(c_0, \frac{g}{\gamma}[X^t X]^{-1}\right).$$

em que γ é a precisão e $g > 0$ pode ser interpretado como o inverso do "tamanho de amostra a priori". Valores altos de g indicam pouca confiança na priori, conduzindo à uma priori mais vaga. Note que esta priori está dentro da classe conjugada. Portanto, a distribuição a posteriori condicional é

$$\beta \mid g, \gamma, y \sim N_p\left(c_1, \frac{C_1}{\gamma}\right)$$

Regressão linear normal: g-priori

$$C_1^{-1} = \frac{1}{g}[X^tX] + X^tX = \frac{1}{q}X^tX,$$

$$c_1 = C_1(C_0^{-1}c_0 + X^ty) = qc_0 + (1-q)\hat{\beta},$$

$$q = \frac{g}{g+1}.$$

- A g-priori tem sido usada para seleção de variáveis, considerando-se $c_0 = 0$ e monitorando g.
- Para evitar problemas com a escolha de g, há uma sugestão de condiderar
$$q \sim Beta(1, \frac{a}{2} - 1) \quad 2 < a < 4.$$

- Considere agora que nossa variável resposta é dicotomica, $P(y_i = 1) = p_i$ e $P(y_i = 0) = 1 - p_i$. Isto é, $y_i \sim Ber(p_i)$.
- Claramente o MRN proposto anteriormente não é adequado.
- No entanto, podemos manter a suposição de linearidade em alguma transformação da média, denotada por $g(p_i)$ e denominada função de ligação. Assim

$$g(p_i) = x_i^t \beta \quad \Leftrightarrow \quad p_i = g^{-1}(x_i^t \beta)$$

- As funções de ligação mais populares são a logito, probito e cloglog . As duas primeiras são simétricas e a última assimétrica.

Regressão binária e binomial

- De modo geral, podemos considerar

$$p_i = F(x_i^t \beta) \quad \text{com} \quad F \text{ uma fda}$$

- Quando observamos n_i unidades amostrais da mesma $Ber(p_i)$, usamos a estatística suficiente dada pela soma dessas observações. Mantemos a mesma notação, agora considerando $y_i \sim Binomial(n_i, p_i)$.
- A função de verossimilhança é proporcional a

$$exp\{\sum_{i=1}^{k} y_i ln(p_i) + (n_i - y_i) ln(1 - p_i)\}$$

- No caso do modelo binário, basta considerar $n_i = 1$.

Se $\beta \sim N(c_0, C_0)$ a distribuição a posteriori é proporcional a

$$exp\left\{-\frac{1}{2}(\beta - c_0)^t C_0^{-1}(\beta - c_0) + \right.$$

$$\left. + \sum_{i=1}^{k} y_i ln(F(x_i^t \beta)) + (n_i - y_i) ln(1 - F(x_i^t \beta))\right\}$$

e não tem solução padrão.

- Métodos MCMC são considerados. Em particular, para a ligação logito é usado o algoritmo de rejeição adaptativa combinado com o GS.
- Podemos usar a linguagem BUGS. Exemplos de codigos são dados nas Notas do capítulo 3 do livro.

Regressão binária e binomial

- Uma proposta interessante de construção de priori para β em modelos binários/binomiais é partir da elicitação de prioris para as probabilidades.
- As probabilidades são mais facilmente interpretáveis no contexto do problema e portanto, é mais facil construir prioris informativas para essas quantidades.
- Considere π_r um valor de probabilidade associado a um preditor x_r ou um grupo de preditores.
- Para poder induzir uma distribuição para o vetor de β a partir das probabilidades, é necessário ter uma transformação bijetora. Portanto, é preciso considerar $\pi_r \;\; r = 1, \ldots, p$. Sendo p a dimensão do vetor β.

Supondo

$$\pi_r \sim Beta(a_r, b_r) \quad r = 1, \ldots, p \quad \text{independentes}$$

como

$$\pi_r = F(x_r^t \beta)$$

usando o método jacobiano de transformação de variáveis obtemos

$$h(\beta) \propto \prod_{r=1}^{p} [F(x_r^t \beta)]^{a_r - 1} [1 - F(x_r^t \beta)]^{b_r - 1} \mid \frac{dF}{d\beta} \mid$$

Caso especial $p = 2$: $x_i^t \beta = \beta_1 + \beta_2 x_i$ e ligação logito.

$$F(u) = \frac{e^u}{1 + e^u}$$

Temos que

$$\frac{dF}{du} = \frac{e^u}{(1 + e^u)^2} = F(u)[1 - F(u)]$$

É preciso especificar duas probabilidades $\pi_{r1} = F(\beta_1 + \beta_2 x_{r1})$ e $\pi_{r2} = F(\beta_1 + \beta_2 x_{r2})$. Neste caso

$$\frac{dF}{d\beta} = \begin{bmatrix} \frac{dF(\beta_1+\beta_2 x_{r1})}{d\beta_1} & \frac{dF(\beta_1+\beta_2 x_{r1})}{d\beta_2} \\ \\ \frac{dF(\beta_1+\beta_2 x_{r2})}{d\beta_1} & \frac{dF(\beta_1+\beta_2 x_{r2})}{d\beta_2} \end{bmatrix} =$$

Regressão binária e binomial

$$\begin{bmatrix} F(\beta_1+\beta_2 x_{r1})[1-F(\beta_1+\beta_2 x_{r1})] & x_{r1}F(\beta_1+\beta_2 x_{r1})[1-F(\beta_1+\beta_2 x_{r1})] \\ F(\beta_1+\beta_2 x_{r2})[1-F(\beta_1+\beta_2 x_{r2})] & x_{r2}F(\beta_1+\beta_2 x_{r2})[1-F(\beta_1+\beta_2 x_{r2})] \end{bmatrix}$$

O determinante é dado por

$$(x_{r2}-x_{r1})F(\beta_1+\beta_2 x_{r1})[1-F(\beta_1+\beta_2 x_{r1})]F(\beta_1+\beta_2 x_{r2})[1-F(\beta_1+\beta_2 x_{r2})]$$

Resulta na seguinte distribuição a priori

$$h(\beta) \propto \prod_{j=1}^{2}[F(x_{rj}^t\beta)]^{a_{rj}}[1-F(x_{rj}^t\beta)]^{b_{rj}}$$

Por exemplo, fixando $x_{r1}^t = (1, x_{q1})$ e $x_{r2}^t = (1, x_{q3})$. Em que $q1$ e $q3$ são os primeiro e terceiro quartis da variável explicativa.

Sob o modelo logito, temos que

$$F^{-1}(\pi_{q1}) = logito(\pi_{q1}) = \beta_1 + \beta_2 x_{q1}$$
$$F^{-1}(\pi_{q3}) = logito(\pi_{q3}) = \beta_1 + \beta_2 x_{q3}$$

Invertendo, resulta

$$\beta_2 = \frac{logito(\pi_{q3}) - logito(\pi_{q1})}{x_{q3} - x_{q1}}$$

$$\beta_1 = logito(\pi_{q1}) - \beta_2 x_{q1}$$

Regressão Poisson

- Associada a dados de contagem,

$$y_i \mid \lambda_i \sim Poisson(\lambda_i).$$

- A função de ligação frequentemente utilizada é a logaritmica, isto é,

$$ln(\lambda_i) = x_i^t \beta.$$

- Novamente poderiamos supor uma distribuição normal multivariada para β.
- Alternativamente, podemos usar uma técnica análoga a feita anteriormente para elicitar indiretamente uma priori para β.

Regressão Poisson

Considere $\rho_r = g^{-1}(x_r^t \beta)$ em que g é a função de ligação. Além disso,

$$\rho_r \sim Gama(a_r, b_r) \quad , r = 1, \ldots, p \;\; \text{ind.}$$

Então,

$$h(\beta) \propto \prod_{r=1}^{p} \left(g^{-1}(x_r^t \beta)\right)^{a_r - 1} \exp\{b_r g^{-1}(x_r^t \beta)\} | \frac{dg^{-1}}{d\beta} |$$

Considerando agora $g^{-1}(u) = \exp\{u\}$ temos

$$h(\beta) \propto \prod_{r=1}^{p} \left(\exp\{x_r^t \beta\}\right)^{a_r} \exp\left\{b_r \exp\{x_r^t \beta\}\right\}.$$

Caso especial $p = 2$: $x_i^t\beta = \beta_1 + \beta_2 x_i$ e ligação logaritmo.
Vamo ilustrar com outra escolha para os x.

Considere que a variável x esta padronizada (subtrair a média e dividir pelo desvio). Seja $x_1^t = (1, x_L) = (1, -1)$ e $x_1^t = (1, x_U) = (1, 1)$, então

$$ln(\rho_1) = \beta_1 + \beta_2 x_L = \beta_1 - \beta_2$$

$$ln(\rho_2) = \beta_1 + \beta_2 x_U = \beta_1 + \beta_2$$

Portanto,

$$\beta_2 = [ln(\rho_2) - ln(\rho_1)]/2$$

$$\beta_1 = [ln(\rho_2) + ln(\rho_1)]/2$$

Regressão com respostas discretas

- Quando a variável resposta é Binomial/Bernoulli ou Poisson, a medida de ajuste $I(y_{rep,i} > y_i)$ deve ser substituida por $I(y_{rep,i} = y_i)$.
- Para obtenção das réplicas $y_{rep,i}$ devemos simular de $f(y_i \mid \beta^{(j)}$ em que $\beta^{(j)}$, $j = 1, \ldots, M$ amostra da posteriori.
- O uso de simulação da posteriori completa é mais simples mas menos realista. Pois, o próprio valor a ser predito, y_i, está sendo usado na estimação dos parâmetros.
- A alternativa mais eficiente é considerar validação cruzada, isto é, simular de $h(\beta \mid y_{[i]})$. Isso pode ser feito com o conjunto de dados não é muito grande.