Modelos Lineares Generalizados

Modelo de regressão binomial negativa

Terezinha K. A. Ribeiro

terezinha.ribeiro@unb.br

Instituto de Ciências Exatas
Departamento de Estatística
Universidade de Brasília

Organização

- Distribuição binomial negativa
- Construção da distribuição
- Distribuição de Pascal
- Estrutura de regressão
- Estimação dos parâmetros

Distribuição binomial negativa

Seja Y uma variável aleatória com distribuição binomial negativa de parâmetros $\mu > 0$ e $\nu > 0$. A função de probabilidades de Y é dada por

$$f(y;\mu,\nu) = \frac{\Gamma(\nu+y)}{\Gamma(y+1)\Gamma(\nu)}\left(\frac{\mu}{\mu+\nu}\right)^{y}\left(\frac{\nu}{\mu+\nu}\right)^{\nu}, \quad y \in \{0,1,2,\ldots\}.$$

Denotaremos por $Y \sim \mathrm{BN}(\mu,\nu)$.

Veja que

$$\begin{aligned} f(y;\mu,\nu) &= \exp\left\{\log\left[\frac{\Gamma(\nu+y)}{\Gamma(y+1)\Gamma(\nu)}\right] + y\log\left(\frac{\mu}{\mu+\nu}\right) + \nu\log\left(\frac{\nu}{\mu+\nu}\right)\right\} \\ &= \exp\left\{\nu\left[\frac{y}{\nu}\log\left(\frac{\mu}{\mu+\nu}\right) + \log\left(\frac{\nu}{\mu+\nu}\right)\right] + \log\left[\frac{\Gamma(\nu+y)}{\Gamma(y+1)\Gamma(\nu)}\right]\right\}. \end{aligned}$$

Construção da distribuição

Suponha que $Y|z \sim \mathrm{Poisson}(z)$ e $Z \sim \mathrm{gama}(\mu, \nu)$, $\mu > 0$ e $\nu > 0$. Qual será a distribuição da variável aleatória Y?

Sabemos que

$$f_{Y|Z=z}(y) = \frac{\mathrm{e}^{-z} z^y}{y!}, \quad y \in \{0, 1, 2, \ldots\},$$

$$g_Z(z; \mu, \nu) = \frac{1}{z} \frac{1}{\Gamma(\nu)} \left(\frac{z\nu}{\mu} \right)^{\nu} \mathrm{e}^{-\nu z/\mu}, \quad z > 0.$$

Daí, a função de probabilidades de Y é

$$f(y; \mu, \nu) = \int_0^{\infty} f(y, z; \mu, \nu) dz,$$

em que $f(y, z; \mu, \nu)$ é a função densidade conjunta de Y e Z.

Por definição,

$$
\begin{aligned}
f(y, z; \mu, \nu) &= f_{Y|Z=z}(y) g_Z(z; \mu, \nu) \\
&= \frac{e^{-z} z^y}{y!} \frac{1}{z} \frac{1}{\Gamma(\nu)} \left(\frac{z\nu}{\mu} \right)^{\nu} e^{-\nu z/\mu} \\
&= \frac{1}{y!} \left(\frac{\nu}{\mu} \right)^{\nu} \frac{1}{\Gamma(\nu)} e^{-z(1+\nu/\mu)} z^{\nu+y-1} \\
&= \frac{1}{\Gamma(y+1)\Gamma(\nu)} \left(\frac{\nu}{\mu} \right)^{\nu} e^{-z(1+\nu/\mu)} z^{\nu+y-1}
\end{aligned}
$$

com $y \in \{0, 1, 2, \ldots\}$, $z > 0$, $\nu > 0$.

Construção da distribuição

Assim, a função de probabilidades de Y é

$$
\begin{aligned}
f(y;\mu,\nu) &= \int_0^\infty \frac{1}{\Gamma(y+1)\Gamma(\nu)}\left(\frac{\nu}{\mu}\right)^\nu \mathrm{e}^{-z(1+\nu/\mu)} z^{\nu+y-1} dz \\
&= \frac{1}{\Gamma(y+1)\Gamma(\nu)}\left(\frac{\nu}{\mu}\right)^\nu \int_0^\infty \mathrm{e}^{-z(1+\nu/\mu)} z^{\nu+y-1} dz.
\end{aligned}
$$

Faça

$$
t = z(1+\nu/\mu) \Rightarrow z = \frac{t}{1+\nu/\mu} \Rightarrow \frac{dz}{dt} = \frac{1}{1+\nu/\mu} \Rightarrow dz = \frac{1}{1+\nu/\mu} dt.
$$

Logo, obtemos que

$$
\begin{aligned}
f(y;\mu,\nu) &= \frac{1}{\Gamma(y+1)\Gamma(\nu)}\left(\frac{\nu}{\mu}\right)^{\nu}\int_0^{\infty} \mathrm{e}^{-t}\left(\frac{t}{1+\nu/\mu}\right)^{\nu+y-1}\frac{1}{1+\nu/\mu}dt \\
&= \frac{1}{\Gamma(y+1)\Gamma(\nu)}\left(\frac{\nu}{\mu}\right)^{\nu}\frac{1}{(1+\nu/\mu)^{\nu+y}}\int_0^{\infty} \mathrm{e}^{-t}t^{\nu+y-1}dt \\
&= \frac{\Gamma(\nu+y)}{\Gamma(y+1)\Gamma(\nu)}\left(\frac{\nu}{\mu}\right)^{\nu}\frac{1}{(1+\nu/\mu)^{\nu+y}} \\
&= \frac{\Gamma(\nu+y)}{\Gamma(y+1)\Gamma(\nu)}\left(\frac{\nu}{\mu}\right)^{\nu}\left(\frac{\mu}{\mu+\nu}\right)^{(y+\nu)} \\
&= \frac{\Gamma(\nu+y)}{\Gamma(y+1)\Gamma(\nu)}\left(\frac{\mu}{\mu+\nu}\right)^{y}\left(\frac{\nu}{\mu+\nu}\right)^{\nu}.
\end{aligned}
$$

Distribuição binomial negativa

Inicialmente, considere o parâmetro ν conhecido. Assim, tomando como variável $Y^* = Y/\nu$, defina $\phi = \nu$,

$$\theta = \log\left(\frac{\mu}{\mu+\nu}\right) \Rightarrow \mu = \frac{\nu e^{\theta}}{1-e^{\theta}}; \quad \theta < 0;$$

$$b(\theta) = -\log\left(\frac{\nu}{\mu+\nu}\right) = -\log(1-e^{\theta});$$

$$c(y;\phi) = \log\left[\frac{\Gamma(\nu+y)}{\Gamma(y+1)\Gamma(\nu)}\right] = \log\left[\frac{\Gamma(\phi+y)}{\Gamma(y+1)\Gamma(\phi)}\right].$$

O suporte $y^* \in \{0, 1/\nu, 2/\nu, \ldots\}$ não depende de parâmetros desconhecidos.

Neste caso, para ν um valor conhecido, $Y^* = Y/\nu \in \mathrm{FE}(\mu, \phi)$.

Distribuição binomial negativa

Segue que

$$\mu^* = \mathrm{E}(Y^*) = \mathrm{E}\left(\frac{Y}{\nu}\right) = b'(\theta) = \frac{\mathrm{e}^\theta}{1-\mathrm{e}^\theta} = \frac{\mu/(\mu+\nu)}{1-\mu/(\mu+\nu)} = \frac{\mu}{\nu}.$$

Assim, $\mathrm{E}(Y) = \mu$. Portanto, o parâmetro μ da distribuição binomial negativa é diretamente a média do modelo.

A função de variância é

$$\begin{aligned} V(\mu^*) &= b''(\theta) \\ &= \frac{\mathrm{e}^\theta(1-\mathrm{e}^\theta) - \mathrm{e}^\theta \cdot -\mathrm{e}^\theta}{(1-\mathrm{e}^\theta)^2} \\ &= \frac{\mathrm{e}^\theta}{1-\mathrm{e}^\theta}\frac{1}{1-\mathrm{e}^\theta} \\ &= \frac{\mu}{\nu}\frac{\mu+\nu}{\nu} \\ &= \frac{\mu(\mu+\nu)}{\nu^2}. \end{aligned}$$

Distribuição binomial negativa

$$
\begin{aligned}
\text{Var}(Y^*) &= \text{Var}\left(\frac{Y}{\nu}\right) \\
&= \phi^{-1} V(\mu^*) \\
&= \frac{1}{\nu}\frac{\mu(\mu+\nu)}{\nu^2} \\
&= \frac{\mu(\mu+\nu)}{\nu^3}.
\end{aligned}
$$

Isto implica que

$$
\text{Var}(Y) = \nu^2 \, \text{Var}(Y^*) = \nu^2 \frac{\mu(\mu+\nu)}{\nu^3} = \frac{\mu(\mu+\nu)}{\nu}.
$$

Observe que

$$
\text{Var}(Y) = \frac{\mu^2}{\nu} + \mu > \mu.
$$

Distribuição binomial negativa

Portanto, o modelo binomial negativa acomoda dados com sobredispersão.

Note que quanto maior for o parâmetro $\nu > 0$, menor será $\text{Var}(Y)$, e consequentemente, menor será a sobredispersão.

Apesar da distribuição de Poisson não ser um caso especial deste modelo, quando $\nu \to \infty$, a distribuição binomial negativa converge para a distribuição de Poisson de média μ.

Na prática, o parâmetro ν é desconhecido e deve estimado. Por esta razão, o modelo de regressão binomial negativa não é um MLG.

Distribuição de Pascal

A distribuição de Pascal é obtida como um caso especial da distribuição BN. Para tanto, considere

- $\nu = r > 0$ é um valor inteiro;
- tome $y' = y + r$, isto é, $y' \in \{r, r+1, \ldots\}$;
- $\pi = r/(\mu + r) \Rightarrow 1 - \pi = \mu/(\mu + r)$ com $0 < \pi < 1$.

Observe que

$$\frac{\Gamma(r+y)}{\Gamma(y+1)\Gamma(r)} = \frac{(r+y-1)!}{y!(r-1)!} = \binom{y+r-1}{y} = \binom{y'-1}{y'-r} = \binom{y'-1}{r-1}.$$

A função de probabilidade da binomial negativa fica expressa por

$$f(y; r, \pi) = \binom{y'-1}{r-1} \pi^r (1-\pi)^{y'-r}, \quad y' \in \{r, r+1, \ldots\}.$$

Denota-se $Y' \sim \text{Pascal}(r, \pi)$.

Estrutura de regressão

Considere $Y_1, Y_2, \ldots, Y_n$ variáveis aleatórias independentes tais que $Y_i \sim \mathrm{BN}(\mu_i, \phi)$. A função de probabilidade de Y_i é

$$f(y_i; \mu_i, \phi) = \frac{\Gamma(\phi + y_i)}{\Gamma(y_i + 1)\Gamma(\phi)} \left(\frac{\mu_i}{\mu_i + \phi}\right)^{y_i} \left(\frac{\phi}{\mu_i + \phi}\right)^{\phi},$$

em que $y_i \in \{0, 1, 2, \ldots\}$, $\mu_i > 0$, $\phi > 0$ tais que

$$\mathrm{E}(Y_i) = \mu_i, \quad \mathrm{Var}(Y_i) = \mu_i + \frac{\mu_i^2}{\phi}.$$

Neste caso, temos

- média da resposta igual a μ_i;
- variância dependente da média (heteroscedasticidade);
- variância maior do que a média;
- adequação para dados de contagem (alternativa ao MLG Poisson).

Estrutura de regressão

O modelo de regressão binomial negativa é definido por

i) $Y_i \mid \boldsymbol{x}_i \overset{\text{ind}}{\sim} \mathrm{BN}\left(\mu_i, \phi\right)$

ii) $g\left(\mu_i\right) = \eta_i \Leftrightarrow \mu_i = g^{-1}\left(\eta_i\right)$ com $\eta_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}$.

Como $\mu_i > 0$, analogamente ao MLG Poisson, as ligações mais utilizadas são

- $g(x) = \log(x)$;
- $g(x) = \sqrt{x}$;
- $g(x) = x$.

Estimação dos parâmetros

O logaritmo da função de verossimilhança para $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top \ \phi)^\top$ é

$$
\begin{aligned}
\ell(\boldsymbol{\theta}) &= \sum_{i=1}^{n} \left\{ \log \left[\frac{\Gamma(\phi + y_i)}{\Gamma(y_i + 1)\Gamma(\phi)} \right] + y_i \log \left(\frac{\mu_i}{\mu_i + \phi} \right) + \phi \log \left(\frac{\phi}{\mu_i + \phi} \right) \right\} \\
&= \sum_{i=1}^{n} \Big\{ \log \left(\Gamma(\phi + y_i) \right) - \log(\Gamma(y_i + 1)) - \log(\Gamma(\phi)) \\
&+ \ y_i \log(\mu_i) + \phi \log(\phi) - (y_i + \phi) \log(\mu_i + \phi) \Big\},
\end{aligned}
$$

com $\mu_i = g^{-1}(\eta_i)$.

O vetor escore para $\boldsymbol{\beta}$ é

$$
U_{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}}.
$$

Estimação dos parâmetros

As entradas do vetor escore para $\boldsymbol{\beta}$ são obtidas de

$$
\begin{aligned}
\frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_j} &= \sum_{i=1}^{n} \left\{ \frac{y_i}{\mu_i} \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_j} - \frac{(y_i+\phi)}{\mu_i+\phi} \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_j} \right\} \\
&= \sum_{i=1}^{n} \left\{ \frac{y_i}{\mu_i} - \frac{(y_i+\phi)}{\mu_i+\phi} \right\} \frac{d\mu_i}{d\eta_i} x_{ij} \\
&= \sum_{i=1}^{n} \left\{ \frac{y_i(\mu_i+\phi) - \mu_i(y_i+\phi)}{\mu_i(\mu_i+\phi)} \right\} \frac{d\mu_i}{d\eta_i} x_{ij} \\
&= \sum_{i=1}^{n} \left\{ \frac{y_i\mu_i + y_i\phi - \mu_i y_i - \mu_i\phi)}{\mu_i(\mu_i+\phi)} \right\} \frac{d\mu_i}{d\eta_i} x_{ij} \\
&= \sum_{i=1}^{n} \frac{\phi(y_i-\mu_i)}{\mu_i(\mu_i+\phi)} \frac{d\mu_i}{d\eta_i} x_{ij} \\
&= \sum_{i=1}^{n} \frac{(y_i-\mu_i)}{(\mu_i^2/\phi+\mu_i)} \frac{d\mu_i}{d\eta_i} x_{ij}.
\end{aligned}
$$

Estimação dos parâmetros

Matricialmente,

$$U_{\boldsymbol{\beta}}(\boldsymbol{\theta}) = X^\top W F^{-1}(\boldsymbol{y} - \boldsymbol{\mu}),$$

em que X é a matriz do modelo $n \times p$, $\boldsymbol{y}$ é o vetor de respostas, $\boldsymbol{\mu}$ é o vetor de médias, $W = \mathrm{diag}\{\omega_1, \omega_2, \ldots, \omega_n\}$, $F = \mathrm{diag}\{f_1, f_2, \ldots, f_n\}$,

$$\omega_i = \frac{(d\mu_i/d\eta_i)^2}{\mu_i + \mu_i^2/\phi}, \quad f_i = \frac{d\mu_i}{d\eta_i}.$$

A função escore para ϕ é dada por

$$\begin{aligned} U_\phi = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \phi} &= \sum_{i=1}^{n} \left\{ \psi(\phi + y_i) - \psi(\phi) + \log(\phi) + \frac{\phi}{\phi} - \log(\mu_i + \phi) - \frac{(y_i + \phi)}{(\mu_i + \phi)} \right\} \\ &= \sum_{i=1}^{n} \left\{ \psi(\phi + y_i) - \psi(\phi) + \log\left(\frac{\phi}{\mu_i + \phi}\right) - \frac{(y_i + \phi)}{(\mu_i + \phi)} + 1 \right\}, \end{aligned}$$

em que $\psi(y) = d\log(\Gamma(y))/dy$ é a função digama.

Estimação dos parâmetros

Para encontrar $\widehat{\theta}$, fazemos

$$U_{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \mathbf{0}, \quad U_{\phi}(\boldsymbol{\theta}) = 0.$$

Observe que $\widehat{\boldsymbol{\beta}}$ e $\widehat{\phi}$ não possuem forma fechada. Por esta razão, utiliza-se processo iterativo escore de Fisher para encontrar o valor de $\widehat{\boldsymbol{\beta}}$.

As demais quantidades inferenciais, testes de hipóteses, medidas de qualidade de ajuste e técnicas de diagnóstico podem ser desenvolvidas de forma análoga ao feito para os MLGs.