CE043 - GAMLSS

Família GAMLSS - Tópicos Complementares

Silva, J.P; Taconeli, C.A.

10 de setembro, 2020

Conteúdo

Introdução

Introdução

- Variáveis aleatórias contínuas com suporte no intervalo (0,1) são resultantes de índices, porcentagens e taxas, dentre outros. Alguns exemplos:

Introdução

- Variáveis aleatórias contínuas com suporte no intervalo (0,1) são resultantes de índices, porcentagens e taxas, dentre outros. Alguns exemplos:

  - Proporção das áreas de folhas de uma planta danificadas por certa doença;

Introdução

- Variáveis aleatórias contínuas com suporte no intervalo (0,1) são resultantes de índices, porcentagens e taxas, dentre outros. Alguns exemplos:
  - Proporção das áreas de folhas de uma planta danificadas por certa doença;
  - Índices, usados por exemplo para medir o desenvolvimento de municípios, a desigualdade de renda de países...

Introdução

- Variáveis aleatórias contínuas com suporte no intervalo (0,1) são resultantes de índices, porcentagens e taxas, dentre outros. Alguns exemplos:

    - Proporção das áreas de folhas de uma planta danificadas por certa doença;

    - Índices, usados por exemplo para medir o desenvolvimento de municípios, a desigualdade de renda de países...

    - Desempenho (notas) de candidatos em exames escolares ou provas de processos seletivos...

Introdução

- Caso a variável original (X) seja definida num intervalo (a, b), podemos mapeá-la para o intervalo $(0, 1)$ da seguinte maneira:

$$Y = \frac{X - a}{b - a}.$$

Introdução

- Como alternativas de modelagem para variáveis contínuas com suporte no intervalo (0,1), destacam-se:

Introdução

- Como alternativas de modelagem para variáveis contínuas com suporte no intervalo (0,1), destacam-se:
    - Distribuições de probabilidades que originalmente tem suporte no intervalo (0,1);

Introdução

- Como alternativas de modelagem para variáveis contínuas com suporte no intervalo (0,1), destacam-se:

  - Distribuições de probabilidades que originalmente tem suporte no intervalo (0,1);

  - Criar uma distribuição no intervalo (0,1) aplicando a transformação inversa da logito a qualquer distribuição contínua com suporte em $(-\infty, \infty)$, ou inversa exponencial para variáveis em $(0, \infty)$;

Introdução

- Como alternativas de modelagem para variáveis contínuas com suporte no intervalo (0,1), destacam-se:

  - Distribuições de probabilidades que originalmente tem suporte no intervalo (0,1);

  - Criar uma distribuição no intervalo (0,1) aplicando a transformação inversa da logito a qualquer distribuição contínua com suporte em $(-\infty, \infty)$, ou inversa exponencial para variáveis em $(0, \infty)$;

  - Criar uma distribuição no intervalo (0,1) truncando alguma distribuição contínua com suporte em $(-\infty, \infty)$ ou $(0, \infty)$.

Modelos probabilísticos para variáveis contínuas definidos em (0,1)

Modelos probabilísticos para variáveis contínuas definidos em (0,1)

Tabela 1: Distribuições definidas em (0,1)

Distribuição	Nome gamlss	Parâmetro (função de ligação)			
		μ	σ	ν	τ
beta	`BE`	logit	logit	-	-
beta (orig)	`BEo`	log	log	-	-
gen beta type I	`GB1`	logit	logit	log	log
logit-normal	`LOGITNO`	logit	log	-	-
simplex	`SIMPLEX`	logit	log	-	-

Distribuição beta (BE)

- A distribuição beta, em sua formulação original, é definida pela seguinte função densidade de probabilidade:

$$f(y|\alpha,\beta) = \frac{1}{B(\alpha,\beta)} y^{\alpha-1}(1-y)^{\beta-1} \text{para } 0 < y < 1, \alpha > 0, \beta > 0,$$

e $f(y|\alpha,\beta) = 0$, caso contrário. Além disso, $B(a,b)$ representa a função matemática beta, definida por:

$$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx.$$

Distribuição beta (BE)

- O pacote `gamlss` dispõe de uma versão reparametrizada da distribuição beta, com parâmetros μ e σ, tais que:

$$\mu = E(y) = \frac{\alpha}{\alpha + \beta}; \quad \sigma = \frac{1}{\alpha + \beta + 1},$$

com $0 < \mu < 1$, sendo $0 < \sigma < 1$ um parâmetro de dispersão e

$$Var(y) = \sigma^2 \mu(1 - \mu).$$

Distribuição simplex (SIMPLEX)

- A distribuição simplex é definida pela seguinte função densidade de probabilidade:

$$f(y|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2(y(1-y))^3}} \exp\left\{-\frac{1}{2\sigma^2}\frac{(y-\mu)^2}{y(1-y)\mu^2(1-\mu)^2}\right\},$$

com $0 < y < 1$, $0 < \mu < 1$ e $\sigma > 0$.

Distribuição simplex (SIMPLEX)

- A distribuição simplex é definida pela seguinte função densidade de probabilidade:

$$f(y|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2(y(1-y))^3}} \exp\left\{-\frac{1}{2\sigma^2}\frac{(y-\mu)^2}{y(1-y)\mu^2(1-\mu)^2}\right\},$$

com $0 < y < 1$, $0 < \mu < 1$ e $\sigma > 0$.

- Neste caso, $E(y) = \mu$ é a média da distribuição e $\sigma > 0$ um parâmetro de dispersão.

Distribuição beta generalizada (GB1)

- A distribuição beta generalizada é definida pela seguinte função densidade de probabilidade:

$$f(y|\mu,\sigma,\nu,\tau) = \frac{\tau\nu^{\beta}y^{\tau\alpha-1}(1-y^{\tau})^{\beta-1}}{B(\alpha,\beta)\left[\nu+(1-\nu)y^{\tau}\right]^{\alpha+\beta}},$$

com $0 < y < 1$, $0 < \mu < 1$, $\alpha > 0$, $\beta > 0$, $\nu > 0$, $\tau > 0$.

Distribuição beta generalizada (GB1)

- A distribuição beta generalizada é definida pela seguinte função densidade de probabilidade:

$$f(y|\mu,\sigma,\nu,\tau) = \frac{\tau\nu^{\beta}y^{\tau\alpha-1}(1-y^{\tau})^{\beta-1}}{B(\alpha,\beta)\left[\nu+(1-\nu)y^{\tau}\right]^{\alpha+\beta}},$$

com $0 < y < 1$, $0 < \mu < 1$, $\alpha > 0$, $\beta > 0$, $\nu > 0$, $\tau > 0$.

- No pacote `gamlss`, a distribuição beta generalizada está parametrizada em função de μ e σ:

$$\mu = E(y) = \frac{\alpha}{\alpha+\beta}; \sigma = \frac{1}{\sqrt{\alpha+\beta+1}},$$

além de ν e τ.

Distribuição logit-normal (LOGITNO)

- A distribuição logit-normal é produzida pela inversa da transformação logito aplicada a uma variável com distribuição normal.

Distribuição logit-normal (LOGITNO)

- A distribuição logit-normal é produzida pela inversa da transformação logito aplicada a uma variável com distribuição normal.

- Seja $Z \sim \texttt{NO}(\text{logit}(\mu), \sigma)$. Então, a v.a. Y definida por

$$Y = \frac{1}{1 + \exp(-Z)}$$

tem distribuição logit-normal com mediana μ e parâmetro de escala σ.

Distribuição logit-normal (LOGITNO)

- A distribuição logit-normal é produzida pela inversa da transformação logito aplicada a uma variável com distribuição normal.

- Seja $Z \sim \texttt{NO}(\text{logit}(\mu), \sigma)$. Então, a v.a. Y definida por

$$Y = \frac{1}{1 + \exp(-Z)}$$

tem distribuição logit-normal com mediana μ e parâmetro de escala σ.

- Na próxima seção vamos ver uma regra geral para criar distribuições com suporte em (0,1) usando a inversa da transformação logito.

Produzindo novas distribuições em (0,1)

Distribuições geradas pela transformação logito

- Distribuições definidas em (0,1) podem ser criadas a partir de variáveis originalmente definidas em $(-\infty, \infty)$ usando a inversa da transformação logito.

Distribuições geradas pela transformação logito

- Distribuições definidas em (0,1) podem ser criadas a partir de variáveis originalmente definidas em $(-\infty, \infty)$ usando a inversa da transformação logito.

- Seja Z uma v.a. definida em $(-\infty, \infty)$, e Y uma função de Z, definida pelo inverso da função logito:

$$Y = \frac{1}{1 + \exp(-Z)}$$

.

Distribuições geradas pela transformação logito

- Distribuições definidas em (0,1) podem ser criadas a partir de variáveis originalmente definidas em $(-\infty, \infty)$ usando a inversa da transformação logito.

- Seja Z uma v.a. definida em $(-\infty, \infty)$, e Y uma função de Z, definida pelo inverso da função logito:

$$Y = \frac{1}{1 + \exp(-Z)}$$

.

- A variável resultante tem suporte no intervalo (0,1) e pode ser usada como alternativa à distribuição beta e demais distribuições definidas no intervalo unitário.

Distribuições geradas pela transformação logito

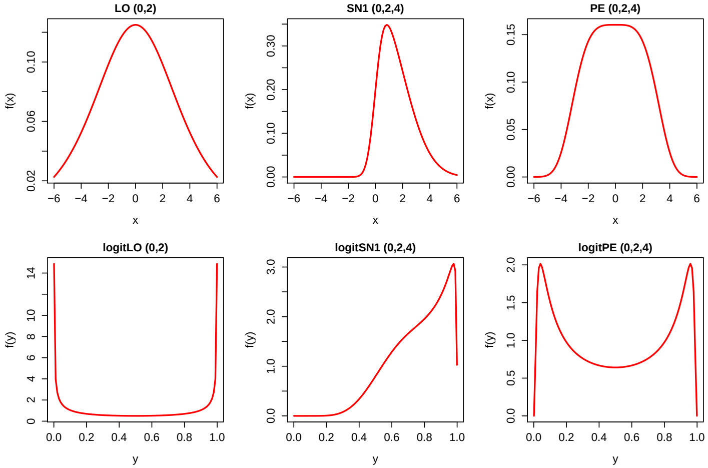

Figura 1: Distribuições geradas pela transformação logito

Distribuições geradas pela transformação logito

- Na biblioteca `gamlss`, podemos gerar uma distribuição com suporte em (0,1) usando a transformação logito com a função `gen.Family`.

Distribuições geradas pela transformação logito

- Na biblioteca `gamlss`, podemos gerar uma distribuição com suporte em (0,1) usando a transformação logito com a função `gen.Family`.

- Como resultado, teremos a distribuição implementada com os recursos usuais `d`, `p`, `q` e `r`.

Distribuições geradas pela transformação logito

- Na biblioteca `gamlss`, podemos gerar uma distribuição com suporte em (0,1) usando a transformação logito com a função `gen.Family`.

- Como resultado, teremos a distribuição implementada com os recursos usuais `d`, `p`, `q` e `r`.

- Outra alternativa para a criação de distribuições em (0,1) é mediante truncamento de uma variável originalmente definida nos reais ou reais positivos.

Distribuições produzidas por truncamento

- Seja X uma v.a. contínua, com suporte em $(-\infty, \infty)$, ou $(0, \infty)$.

Distribuições produzidas por truncamento

- Seja X uma v.a. contínua, com suporte em $(-\infty, \infty)$, ou $(0, \infty)$.
- Considere $f_X(\cdot)$ e $F_X(\cdot)$ a f.d.p. e a f.d.a de X, respectivamente.

Distribuições produzidas por truncamento

- Seja X uma v.a. contínua, com suporte em $(-\infty, \infty)$, ou $(0, \infty)$.
- Considere $f_X(\cdot)$ e $F_X(\cdot)$ a f.d.p. e a f.d.a de X, respectivamente.
- Então, a v.a. Y, produzida pelo truncamento de X no intervalo (0,1), tem f.d.p.:

$$f_Y(y) = \frac{f_X(y)}{F_X(1) - F_X(0)}, \quad \text{para } 0 < y < 1.$$

Distribuições produzidas por truncamento

- Seja X uma v.a. contínua, com suporte em $(-\infty, \infty)$, ou $(0, \infty)$.
- Considere $f_X(\cdot)$ e $F_X(\cdot)$ a f.d.p. e a f.d.a de X, respectivamente.
- Então, a v.a. Y, produzida pelo truncamento de X no intervalo (0,1), tem f.d.p.:

$$f_Y(y) = \frac{f_X(y)}{F_X(1) - F_X(0)}, \quad \text{para } 0 < y < 1.$$

- Na biblioteca `gamlss.tr`, a função `gen.trun()` permite criar distribuições truncadas em (0,1) ou em qualquer outro intervalo de interesse.

Distribuições produzidas por truncamento

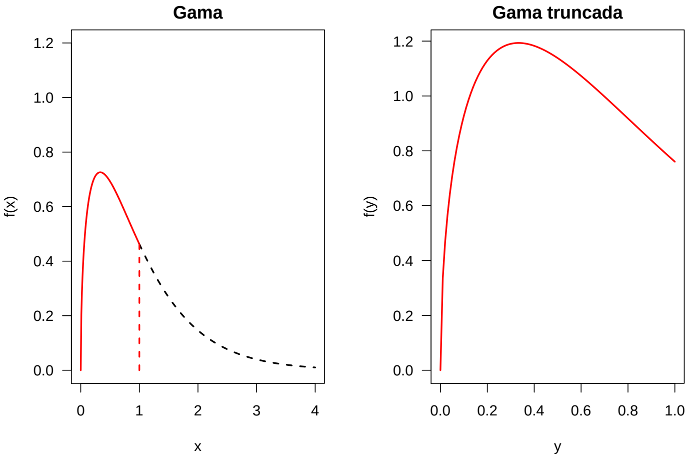

Figura 2: Distribuição gama truncada em (0,1)

Distribuições geradas por truncamento

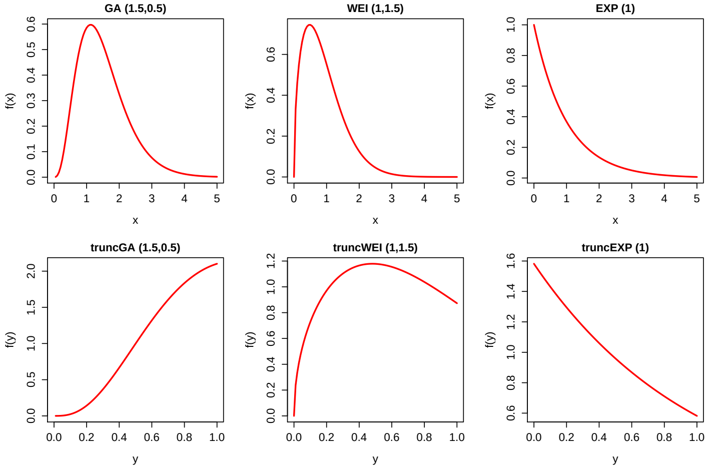

Figura 3: Distribuições geradas por truncamento

Distribuições contínuas zero-ajustadas

Introdução

- Em algumas análises podemos nos deparar com uma variável resposta contínua que assume valor zero com probabilidade não nula (massa de probabilidades em zero);

Introdução

- Em algumas análises podemos nos deparar com uma variável resposta contínua que assume valor zero com probabilidade não nula (massa de probabilidades em zero);

- Um exemplo é a modelagem da precipitação (quantidade de chuva) em diferentes regiões;

Introdução

- Em algumas análises podemos nos deparar com uma variável resposta contínua que assume valor zero com probabilidade não nula (massa de probabilidades em zero);

- Um exemplo é a modelagem da precipitação (quantidade de chuva) em diferentes regiões;

- Nas regiões em que for registrado chuva, a precipitação configura uma variável aleatória contínua;

Introdução

- Em algumas análises podemos nos deparar com uma variável resposta contínua que assume valor zero com probabilidade não nula (massa de probabilidades em zero);
- Um exemplo é a modelagem da precipitação (quantidade de chuva) em diferentes regiões;
- Nas regiões em que for registrado chuva, a precipitação configura uma variável aleatória contínua;
- Nas regiões sem registro de chuva, no entanto, a precipitação será igual a zero.

Distribuições contínuas zero-ajustadas

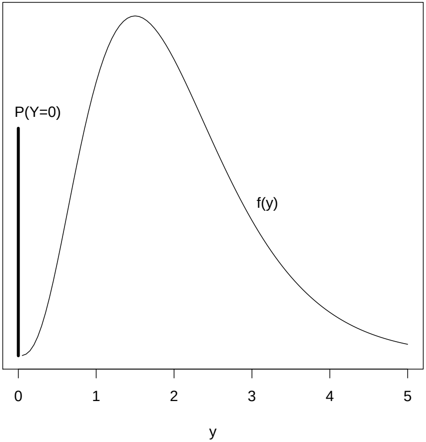

Figura 4: Distribuição contínua zero-ajustada

Distribuições contínuas zero-ajustadas

- Seja Y uma variável aleatória contínua não negativa, com probabilidade não nula $P(Y = 0) = \epsilon_0$;

Distribuições contínuas zero-ajustadas

- Seja Y uma variável aleatória contínua não negativa, com probabilidade não nula $P(Y = 0) = \epsilon_0$;

- Vamos assumir que a parte da distribuição de Y correspondente a valores diferentes de zero seja modelada por uma v.a. X com f.d.p. $f_X(x|\boldsymbol{\theta})$, em que $\boldsymbol{\theta}' = (\mu, \sigma, \nu, \tau)$, no contexto de GAMLSS.

Distribuições contínuas zero-ajustadas

- Seja Y uma variável aleatória contínua não negativa, com probabilidade não nula $P(Y = 0) = \epsilon_0$;

- Vamos assumir que a parte da distribuição de Y correspondente a valores diferentes de zero seja modelada por uma v.a. X com f.d.p. $f_X(x|\boldsymbol{\theta})$, em que $\boldsymbol{\theta}' = (\mu, \sigma, \nu, \tau)$, no contexto de GAMLSS.

- Como resultado, a função de probabilidade mista de Y fica definida por:

$$f_Y(y|\boldsymbol{\theta}, \epsilon_0) = \begin{cases} \epsilon_0, & \text{se } y = 0 \\ (1 - \epsilon_0) f_X(y|\boldsymbol{\theta}), & \text{se } 0 < y < \infty \end{cases}$$

Distribuição gama zero-ajustada (ZAGA)

- A distribuição gama zero-ajustada tem a parte contínua definida pela densidade da distribuição gama, sendo definida por:

$$f(y|\mu,\sigma,\nu) = \nu, \qquad \text{para } y = 0;$$

$$f(y|\mu,\sigma,\nu) = (1-\nu)\left[\frac{1}{(\sigma^2\mu)^{1/\sigma^2}}\frac{y^{\frac{1}{\sigma^2}-1}e^{\frac{-y}{\sigma^2\mu}}}{\Gamma(1/\sigma^2)}\right], \qquad \text{para } y > 0,$$

com $\mu > 0$, $\sigma > 0$, $0 < \nu < 1$.

Distribuição gama zero-ajustada (ZAGA)

- A distribuição gama zero-ajustada tem a parte contínua definida pela densidade da distribuição gama, sendo definida por:

$$f(y|\mu, \sigma, \nu) = \nu, \quad \text{para } y = 0;$$

$$f(y|\mu, \sigma, \nu) = (1-\nu)\left[\frac{1}{(\sigma^2\mu)^{1/\sigma^2}} \frac{y^{\frac{1}{\sigma^2}-1} e^{\frac{-y}{\sigma^2\mu}}}{\Gamma(1/\sigma^2)}\right], \quad \text{para } y > 0,$$

com $\mu > 0$, $\sigma > 0$, $0 < \nu < 1$.

- Assim, $E(y) = (1-\nu)\mu$ e $Var(y) = (1-\nu)\mu^2(\nu + \sigma^2)$.

Distribuição normal inversa zero-ajustada (ZAIG)

- A distribuição normal inversa zero-ajustada tem a parte contínua definida pela densidade da distribuição normal inversa, sendo dada por:

$$f(y|\mu, \sigma, \nu) = \nu, \qquad \text{para } y = 0;$$

$$f(y|\mu, \sigma, \nu) = (1-\nu)\frac{1}{\sqrt{2\pi\sigma^2 y^3}} \exp\left\{-\frac{(y-\mu)^2}{2\mu^2\sigma^2 y}\right\}, \qquad \text{para } y > 0,$$

com $\mu > 0$, $\sigma > 0$, $0 < \nu < 1$.

Distribuição normal inversa zero-ajustada (ZAIG)

- A distribuição normal inversa zero-ajustada tem a parte contínua definida pela densidade da distribuição normal inversa, sendo dada por:

$$f(y|\mu, \sigma, \nu) = \nu, \qquad \text{para } y = 0;$$

$$f(y|\mu, \sigma, \nu) = (1-\nu)\frac{1}{\sqrt{2\pi\sigma^2 y^3}} \exp\left\{-\frac{(y-\mu)^2}{2\mu^2\sigma^2 y}\right\}, \qquad \text{para } y > 0,$$

com $\mu > 0$, $\sigma > 0$, $0 < \nu < 1$.

- Assim, $E(y) = (1-\nu)\mu$ e $Var(y) = (1-\nu)\mu^2(\nu + \mu\sigma^2)$.

Gerando distribuições zero-ajustadas em $(0, \infty)$

- Qualquer distribuição contínua com suporte em $(0, \infty)$ implementada na biblioteca `gamlss` pode ser estendida para uma correspondente versão zero-ajustada.

Gerando distribuições zero-ajustadas em $(0, \infty)$

- Qualquer distribuição contínua com suporte em $(0, \infty)$ implementada na biblioteca `gamlss` pode ser estendida para uma correspondente versão zero-ajustada.
- Neste caso, dependendo do número de parâmetros da distribuição original, a versão zero-ajustada pode ter até cinco parâmetros, caso em que o modelo mais geral fica definido por:

$$Y \sim D(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau}, \boldsymbol{\epsilon_0})$$

Gerando distribuições zero-ajustadas em $(0, \infty)$

- Qualquer distribuição contínua com suporte em $(0, \infty)$ implementada na biblioteca `gamlss` pode ser estendida para uma correspondente versão zero-ajustada.

- Neste caso, dependendo do número de parâmetros da distribuição original, a versão zero-ajustada pode ter até cinco parâmetros, caso em que o modelo mais geral fica definido por:

$$Y \sim D(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau}, \boldsymbol{\epsilon_0})$$

$$\begin{aligned}
g_1(\boldsymbol{\mu}) &= \eta_1 = \boldsymbol{X_1}\boldsymbol{\beta_1} + s_{11}(\boldsymbol{x}_{11}) + ... + s_{1J1}(\boldsymbol{x}_{1_J1}) \\
g_2(\boldsymbol{\sigma}) &= \eta_2 = \boldsymbol{X_2}\boldsymbol{\beta_2} + s_{21}(\boldsymbol{x}_{21}) + ... + s_{2J2}(\boldsymbol{x}_{2_J2}) \\
g_3(\boldsymbol{\nu}) &= \eta_3 = \boldsymbol{X_3}\boldsymbol{\beta_3} + s_{31}(\boldsymbol{x}_{31}) + ... + s_{3J3}(\boldsymbol{x}_{3_J3}) \\
g_4(\boldsymbol{\tau}) &= \eta_4 = \boldsymbol{X_4}\boldsymbol{\beta_4} + s_{41}(\boldsymbol{x}_{41}) + ... + s_{4J4}(\boldsymbol{x}_{4_J4}) \\
g_5(\boldsymbol{\epsilon_0}) &= \eta_5 = \boldsymbol{X_5}\boldsymbol{\beta_5} + s_{51}(\boldsymbol{x}_{51}) + ... + s_{5J5}(\boldsymbol{x}_{5_J5})
\end{aligned}$$

Gerando distribuições zero-ajustadas em $(0, \infty)$

- A função `gen.Zadj()` da biblioteca `gamlss.inf` deve ser usada para gerar distribuições zero-ajustadas em $(0, \infty)$.

Gerando distribuições zero-ajustadas em $(0, \infty)$

- A função `gen.Zadj()` da biblioteca `gamlss.inf` deve ser usada para gerar distribuições zero-ajustadas em $(0, \infty)$.

- Devido à ortogonalidade de ϵ_0 com os parâmetros do modelo original, o modelo resultante é ajustado em duas etapas:

Gerando distribuições zero-ajustadas em $(0, \infty)$

- A função `gen.Zadj()` da biblioteca `gamlss.inf` deve ser usada para gerar distribuições zero-ajustadas em $(0, \infty)$.

- Devido à ortogonalidade de ϵ_0 com os parâmetros do modelo original, o modelo resultante é ajustado em duas etapas:

  - O modelo original (parte contínua) é ajustado usando todas as observações tais que $y > 0$;

Gerando distribuições zero-ajustadas em $(0, \infty)$

- A função `gen.Zadj()` da biblioteca `gamlss.inf` deve ser usada para gerar distribuições zero-ajustadas em $(0, \infty)$.

- Devido à ortogonalidade de ϵ_0 com os parâmetros do modelo original, o modelo resultante é ajustado em duas etapas:

  - O modelo original (parte contínua) é ajustado usando todas as observações tais que $y > 0$;

  - Um modelo para resposta binária (ligação logito, probito, cloglog...) é ajustado para a resposta binária $y = 0$ *vs* $y > 0$.

Gerando distribuições zero-ajustadas em $(0, \infty)$

- A função `gen.Zadj()` da biblioteca `gamlss.inf` deve ser usada para gerar distribuições zero-ajustadas em $(0, \infty)$.

- Devido à ortogonalidade de ϵ_0 com os parâmetros do modelo original, o modelo resultante é ajustado em duas etapas:
    - O modelo original (parte contínua) é ajustado usando todas as observações tais que $y > 0$;
    - Um modelo para resposta binária (ligação logito, probito, cloglog...) é ajustado para a resposta binária $y = 0$ *vs* $y > 0$.

- As verossimilhanças produzidas pelos dois modelos são então combinadas, e os resíduos recalculados para o modelo zero-ajustado resultante.

Distribuições inflacionadas em [0,1]

Introdução

- Dados contínuos com suporte no intervalo (0,1) também podem apresentar probabilidade não nula em um ou ambos os extremos ($y = 0$ e/ou $y = 1$).

Introdução

- Dados contínuos com suporte no intervalo (0,1) também podem apresentar probabilidade não nula em um ou ambos os extremos ($y = 0$ e/ou $y = 1$).

- Por exemplo, uma planta pode ter 100% de sua folha comprometida por uma praga; um atleta pode atingir nota máxima numa escala de 0 a 10 em sua apresentação; ou a proporção do tempo gasto diariamente por indivíduos com certa atividade pode ser igual a zero.

Introdução

- Dados contínuos com suporte no intervalo (0,1) também podem apresentar probabilidade não nula em um ou ambos os extremos ($y = 0$ e/ou $y = 1$).

- Por exemplo, uma planta pode ter 100% de sua folha comprometida por uma praga; um atleta pode atingir nota máxima numa escala de 0 a 10 em sua apresentação; ou a proporção do tempo gasto diariamente por indivíduos com certa atividade pode ser igual a zero.

- A análise de dados desse tipo baseia-se em alguma distribuição definida em (0,1) com inflação (probabilidade não nula) em 0 e/ou 1.

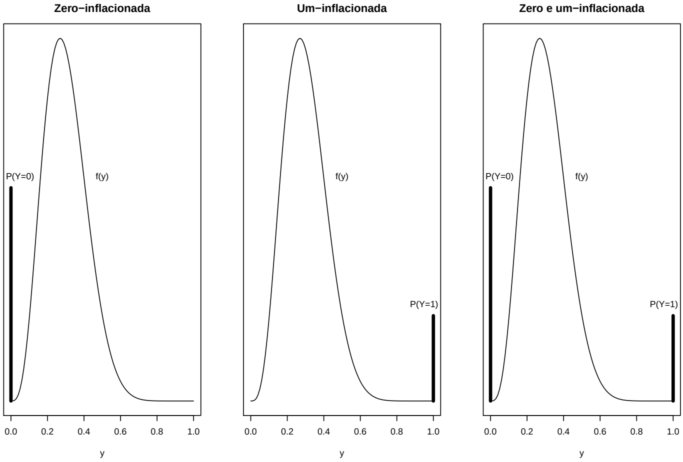

Figura 5: Distribuições inflacionadas em [0,1]

Distribuições inflacionadas em [0,1]

- Seja X uma v.a. contínua com suporte no intervalo (0,1). As seguintes extensões inflacionadas podem ser consideradas:

- Seja X uma v.a. contínua com suporte no intervalo (0,1). As seguintes extensões inflacionadas podem ser consideradas:

- **Distribuições zero-inflacionadas:** A v.a. Y, com suporte em $[0, 1)$, é definida por uma função de probabilidade mista da seguinte forma:

$$f_Y(y|\boldsymbol{\theta}, \epsilon_0) = \begin{cases} \epsilon_0, & \text{se } y = 0 \\ (1 - \epsilon_0) f_X(y|\boldsymbol{\theta}), & \text{se } 0 < y < 1 \end{cases}$$

onde $\epsilon_0 = P(Y = 0)$, $0 < \epsilon_0 < 1$.

Distribuições inflacionadas em [0,1]

- **Distribuições um-inflacionadas:** A v.a. Y, com suporte em $(0, 1]$, é definida por uma função de probabilidade mista da seguinte forma:

$$f_Y(y|\boldsymbol{\theta}, \epsilon_1) = \begin{cases} (1 - \epsilon_1) f_X(y|\boldsymbol{\theta}), & \text{se } 0 < y < 1 \\ \epsilon_1, & \text{se } y = 1 \end{cases}$$

onde $\epsilon_1 = P(Y = 1)$, $0 < \epsilon_1 < 1$.

Distribuições inflacionadas em [0,1]

- **Distribuições um-inflacionadas:** A v.a. Y, com suporte em $(0,1]$, é definida por uma função de probabilidade mista da seguinte forma:

$$f_Y(y|\boldsymbol{\theta},\epsilon_1)=\begin{cases}(1-\epsilon_1)f_X(y|\boldsymbol{\theta}), & \text{se } 0<y<1\\ \epsilon_1, & \text{se } y=1\end{cases}$$

onde $\epsilon_1=P(Y=1)$, $0<\epsilon_1<1$.

- **Distribuições zero e um-inflacionadas:** A v.a. Y, com suporte em $[0,1]$, é definida por uma função de probabilidade mista da seguinte forma:

$$f_Y(y|\boldsymbol{\theta},p_0,p_1)=\begin{cases}p_0, & \text{se } y=0\\ (1-p_0-p_1)f_X(y|\boldsymbol{\theta}), & \text{se } 0<y<1\\ p_1, & \text{se } y=1\end{cases}$$

Distribuições inflacionadas em [0,1]

- A parametrização apresentada para a distribuição inflacionada de zeros e uns não é conveniente para modelagem, pois os parâmetros têm a restrição $p_0 + p_1 < 1$, e os correspondentes MLE's são negativamente correlacionados.

Distribuições inflacionadas em [0,1]

- A parametrização apresentada para a distribuição inflacionada de zeros e uns não é conveniente para modelagem, pois os parâmetros têm a restrição $p_0 + p_1 < 1$, e os correspondentes MLE's são negativamente correlacionados.

- Uma formulação mais apropriada para modelos desse tipo é a seguinte:

$$f_Y(y|\boldsymbol{\theta}, \epsilon_0, \epsilon_1) = \begin{cases} \frac{\epsilon_0}{1+\epsilon_0+\epsilon_1}, & \text{se } y = 0 \\ \frac{1}{1+\epsilon_0+\epsilon_1} f_X(y|\boldsymbol{\theta}), & \text{se } 0 < y < 1 \\ \frac{\epsilon_1}{1+\epsilon_0+\epsilon_1}, & \text{se } y = 1 \end{cases}$$

onde $\epsilon_0 > 0$, $\epsilon_1 > 0$.

- Nesta parametrização:

$$p_0 = P(Y = 0) = \frac{\epsilon_0}{1 + \epsilon_0 + \epsilon_1} \;\; ; \;\; p_1 = P(Y = 1) = \frac{\epsilon_1}{1 + \epsilon_0 + \epsilon_1}$$

- Nesta parametrização:

$$p_0 = P(Y=0) = \frac{\epsilon_0}{1+\epsilon_0+\epsilon_1} \;\; ; \;\; p_1 = P(Y=1) = \frac{\epsilon_1}{1+\epsilon_0+\epsilon_1}$$

- Desta forma,

$$\epsilon_0 = \frac{p_0}{p_2} \quad \text{e} \quad \epsilon_1 = \frac{p_1}{p_2}$$

Distribuições inflacionadas em [0,1]

- A forma geral de um GAMLSS para dados inflacionados em [0,1], assumindo $\boldsymbol{\theta} = (\mu, \sigma, \nu, \tau)$, é dada por:

Distribuições inflacionadas em [0,1]

- A forma geral de um GAMLSS para dados inflacionados em [0,1], assumindo $\boldsymbol{\theta} = (\mu, \sigma, \nu, \tau)$, é dada por:

$$Y \sim D(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau}, \boldsymbol{\epsilon_0})$$

Distribuições inflacionadas em [0,1]

- A forma geral de um GAMLSS para dados inflacionados em [0,1], assumindo $\boldsymbol{\theta} = (\mu, \sigma, \nu, \tau)$, é dada por:

$$Y \sim D(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau}, \boldsymbol{\epsilon_0})$$

$$g_1(\boldsymbol{\mu}) = \eta_1 = \boldsymbol{X_1}\boldsymbol{\beta_1} + s_{11}(\boldsymbol{x}_{11}) + ... + s_{1J1}(\boldsymbol{x}_{1_J1})$$

$$g_2(\boldsymbol{\sigma}) = \eta_2 = \boldsymbol{X_2}\boldsymbol{\beta_2} + s_{21}(\boldsymbol{x}_{21}) + ... + s_{2J2}(\boldsymbol{x}_{2_J2})$$

$$g_3(\boldsymbol{\nu}) = \eta_3 = \boldsymbol{X_3}\boldsymbol{\beta_3} + s_{31}(\boldsymbol{x}_{31}) + ... + s_{3J3}(\boldsymbol{x}_{3_J3})$$

$$g_4(\boldsymbol{\tau}) = \eta_4 = \boldsymbol{X_4}\boldsymbol{\beta_4} + s_{41}(\boldsymbol{x}_{41}) + ... + s_{4J4}(\boldsymbol{x}_{4_J4})$$

$$g_5(\boldsymbol{\epsilon_0}) = \eta_5 = \boldsymbol{X_5}\boldsymbol{\beta_5} + s_{51}(\boldsymbol{x}_{51}) + ... + s_{5J5}(\boldsymbol{x}_{5_J5})$$

$$g_6(\boldsymbol{\epsilon_1}) = \eta_6 = \boldsymbol{X_6}\boldsymbol{\beta_6} + s_{61}(\boldsymbol{x}_{61}) + ... + s_{6J6}(\boldsymbol{x}_{6_J6})$$

Distribuição beta inflacionada (BEZI)

- A distribuição beta zero-inflacionada tem suporte no intervalo [0,1) e, no intervalo aberto (0,1) é definida pela densidade da distribuição beta:

$$f(y|\mu,\sigma,\nu) = \nu, \qquad \text{para } y = 0;$$

$$f(y|\mu,\sigma,\nu) = (1-\nu)\frac{\Gamma(\sigma)}{\Gamma(\mu\sigma)\Gamma((1-\mu)\sigma)} y^{\mu\sigma}(1-y)^{((1-\mu)\sigma)-1},$$

para $0 < y < 1$, com $0 < \mu < 1$, $\sigma > 0$, $0 < \nu < 1$.

Distribuição beta inflacionada (BEZI)

- A distribuição beta zero-inflacionada tem suporte no intervalo [0,1) e, no intervalo aberto (0,1) é definida pela densidade da distribuição beta:

$$f(y|\mu,\sigma,\nu) = \nu, \quad \text{para } y = 0;$$

$$f(y|\mu,\sigma,\nu) = (1-\nu)\frac{\Gamma(\sigma)}{\Gamma(\mu\sigma)\Gamma((1-\mu)\sigma)} y^{\mu\sigma}(1-y)^{((1-\mu)\sigma)-1},$$

para $0 < y < 1$, com $0 < \mu < 1$, $\sigma > 0$, $0 < \nu < 1$.

- Assim, $E(y) = (1-\nu)\mu$ e $Var(y) = (1-\nu)\frac{\mu(1-\mu)}{\sigma+1} + \nu(1-\nu)\mu^2$.

Distribuição beta inflacionada (BEZI)

- A distribuição beta zero-inflacionada tem suporte no intervalo [0,1) e, no intervalo aberto (0,1) é definida pela densidade da distribuição beta:

$$f(y|\mu,\sigma,\nu) = \nu, \quad \text{para } y = 0;$$

$$f(y|\mu,\sigma,\nu) = (1-\nu)\frac{\Gamma(\sigma)}{\Gamma(\mu\sigma)\Gamma((1-\mu)\sigma)} y^{\mu\sigma}(1-y)^{((1-\mu)\sigma)-1},$$

para $0 < y < 1$, com $0 < \mu < 1$, $\sigma > 0$, $0 < \nu < 1$.

- Assim, $E(y) = (1-\nu)\mu$ e $Var(y) = (1-\nu)\frac{\mu(1-\mu)}{\sigma+1} + \nu(1-\nu)\mu^2$.
- De maneira similar, a biblioteca `gamlss` dispõe ainda das versões inflacionadas de uns e de zeros e uns para a distribuição beta (ver `BEOI` e `BEINF`).

Geração de novas distribuições inflacionadas em [0,1]

- A função `gen.Inf0to1`, disponível na biblioteca `gamlss.inf`, permite produzir distribuições inflacionados em [0,1] a partir de distribuições contínuas implementadas na biblioteca `gamlss`.

Geração de novas distribuições inflacionadas em [0,1]

- A função `gen.Inf0to1`, disponível na biblioteca `gamlss.inf`, permite produzir distribuições inflacionados em [0,1] a partir de distribuições contínuas implementadas na biblioteca `gamlss`.

- O argumento `type.of.Inflation` deve ser configurado adequadamente em uma das três opções:

Geração de novas distribuições inflacionadas em [0,1]

- A função `gen.Inf0to1`, disponível na biblioteca `gamlss.inf`, permite produzir distribuições inflacionados em [0,1] a partir de distribuições contínuas implementadas na biblioteca `gamlss`.

- O argumento `type.of.Inflation` deve ser configurado adequadamente em uma das três opções:

  - `Zero` - para inflação em zero;

Geração de novas distribuições inflacionadas em [0,1]

- A função `gen.Inf0to1`, disponível na biblioteca `gamlss.inf`, permite produzir distribuições inflacionados em [0,1] a partir de distribuições contínuas implementadas na biblioteca `gamlss`.

- O argumento `type.of.Inflation` deve ser configurado adequadamente em uma das três opções:

  - `Zero` - para inflação em zero;

  - `One` - para inflação em um;

Geração de novas distribuições inflacionadas em [0,1]

- A função `gen.Inf0to1`, disponível na biblioteca `gamlss.inf`, permite produzir distribuições inflacionados em [0,1] a partir de distribuições contínuas implementadas na biblioteca `gamlss`.

- O argumento `type.of.Inflation` deve ser configurado adequadamente em uma das três opções:

  - `Zero` - para inflação em zero;
  - `One` - para inflação em um;
  - `Zero&One` - para inflação em zero e um.

Resumindo

Família GAMLSS - Tópicos complementares

- Neste módulo estudamos a modelagem de alguns problemas específicos usando gamlss:

Família GAMLSS - Tópicos complementares

- Neste módulo estudamos a modelagem de alguns problemas específicos usando gamlss:
  - Distribuições com suporte em (0,1);

Família GAMLSS - Tópicos complementares

- Neste módulo estudamos a modelagem de alguns problemas específicos usando gamlss:
  - Distribuições com suporte em (0,1);
  - Distribuições zero ajustadas com suporte em $[0,\infty)$;

Família GAMLSS - Tópicos complementares

- Neste módulo estudamos a modelagem de alguns problemas específicos usando gamlss:
  - Distribuições com suporte em (0,1);
  - Distribuições zero ajustadas com suporte em $[0,\infty)$;
  - Distribuições zero e/ou um inflacionadas com suporte em [0,1].

Família GAMLSS - Tópicos complementares

- Neste módulo estudamos a modelagem de alguns problemas específicos usando gamlss:
  - Distribuições com suporte em (0,1);
  - Distribuições zero ajustadas com suporte em $[0,\infty)$;
  - Distribuições zero e/ou um inflacionadas com suporte em [0,1].
- Além das distribuições já implementadas, novas distribuições podem ser geradas através de transformações, truncamentos e misturas, dentre outros;

Família GAMLSS - Tópicos complementares

- Neste módulo estudamos a modelagem de alguns problemas específicos usando gamlss:
  - Distribuições com suporte em (0,1);
  - Distribuições zero ajustadas com suporte em $[0,\infty)$;
  - Distribuições zero e/ou um inflacionadas com suporte em [0,1].
- Além das distribuições já implementadas, novas distribuições podem ser geradas através de transformações, truncamentos e misturas, dentre outros;
- Os parâmetros de inflação, assim como os demais parâmetros das distribuições, podem ser modelados em função de covariáveis.

Próximos passos

Próximos passos

- Efeitos aleatórios em gamlss

Próximos passos

- Efeitos aleatórios em gamlss
    - Diferentes especificações de efeitos aleatórios;

Próximos passos

- Efeitos aleatórios em gamlss
  - Diferentes especificações de efeitos aleatórios;
  - Implementações na biblioteca `gamlss`;

Próximos passos

- Efeitos aleatórios em gamlss
  - Diferentes especificações de efeitos aleatórios;
  - Implementações na biblioteca `gamlss`;
  - Exemplos de motivação.