Generalizações e Extensões de algumas Distribuições Simétricas

Cícero Carlos Ramos de Brito - IFPE/DEINFO-UFRPE [1]
David Augusto Silva - DEINFO-UFRPE [2]
Gabriel Rivas de Melo - DEINFO-UFRPE [3]

1 Introdução

A família de distribuições simétricas com suporte na reta real tem grande aplicabilidade na modelagem estatística. Esta família forma uma classe geral de distribuições com a mesma simetria que a distribuição normal padrão. As distribuições Cauchy, Logística tipos I e II, generalizações e extensões de algumas distribuições simétricas, entre outras, pertencem a esta classe.

A classe de distribuições simétricas tem recebido crescente atenção na literatura estatística , nos últimos anos, pois ela assume para os erros, distribuições com caudas mais pesadas do que a normal, afim de reduzir a influência de pontos aberrantes. Existem ainda várias generalizações e extensões da distribuição normal pertencente à classe simétrica. Uma revisão de diferentes áreas em que distribuições simétricas são aplicadas pode ser encontrada em [1],[3],[4],[5]. A importância desses modelos é que dependendo da situação, precisamos de modelos mais sensíveis a massa de dados ou de modelos mais robustos para não detectar pontos aberrantes ou out-line.

O objetivo deste trabalho é portanto generalizar algumas distribuições conhecidas pertencentes a essa classe e desta forma poder avaliar seu comportamento de modo a identificar situações de modelagem mais especificas para o conjunto de dados.

2 Metodologia e Resultados

2.1 Generalizações e Extensões da Distribuição Normal

A normal é a distribuição pertencente à classe simétrica mais utilizada, devido a todo desenvolvimento teórico e aplicado estabelecido no decorrer dos anos. A distribuição normal possui várias propriedades que permitem caracterizá-la dentro da classe das distribuições simétricas nas chamadas distribuições normais compostas. Alguns resultados nesse sentido podem ser vistos em [2],[7].Existem várias generalizações e extensões da distribuição normal pertencente à classe simétrica. A importância desses modelos é que dependendo da situação, precisamos de modelos mais sensíveis a massa de dados ou de modelos mais robustos para não detectar pontos aberrantes ou out-line. Vejamos o seguinte modelo: Seja $Y \sim S(\mu, \phi, m)$ de tal forma, que a função

[1] *cicerocarlosbrito@yahoo.com.br*
[2] *expressosub@yahoo.com.br*
[3] *rivas@ufrpe.br*

característica e densidade,respectivamente, é dada por

$$\psi_y(t) = \frac{1}{C(0,m)} e^{it\mu} \int_{-\infty}^{+\infty} \exp\left\{ it\phi x - \frac{x^{2m}}{2m} \right\} dx.$$

$$\pi(y;\mu,\phi,m) = \frac{1}{c(0,m)\sqrt{\phi}} \exp\left\{ -\frac{1}{2m} (\frac{y-\mu}{\sqrt{\phi}})^{2m} \right\}.$$

Se Y tem distribuição acima definido, então sua esperança e variância são respectivamente $E(y) = \mu$ e $Var(y) = \phi\frac{C(2,m)}{C(0,m)}$, e os momentos centrais de ordem r são

$$E\{(Y-\mu)^r\} = \begin{cases} \phi^{r/2}\dfrac{C(r,m)}{C(0,m)} = 0, & r \text{ ímpar} \\ \phi^{r/2}\dfrac{C(r,m)}{C(0,m)}, & r \text{ par} \end{cases}$$

$$em \quad que \quad C(r,m) = \int_{-\infty}^{+\infty} x^r \exp\left\{ -\frac{x^{2m}}{2m} \right\} dx.$$

Considere a seguinte expressão $Cg(r) = \frac{E\{(Y-\mu)^r\}}{E\{(Y-\mu)^2\}^{R/2}}$, ou ainda, podemos escrever $Cg(r) = \frac{C(r,m)C(0,m)^{r/2}}{C(0,m)C(2,m)^{r/2}}$, assim para r=4, temos o coeficiente de curtose dada pela expressão $Cc = Cg(4) = \frac{E\{(Y-\mu)^4\}}{E\{(Y-\mu)^2\}^2}$ e portanto, o coeficiente de curtose é $Cc = Cg(4) = \frac{C(4,m)C(0,m)}{C(2,m)C(2,m)}$. Já para r=3, temos o coeficiente de assimetria dada pela expressão $Ca = Cg(3) = \frac{E\{(Y-\mu)^3\}}{E\{(Y-\mu)^2\}^{3/2}}$, e portanto $Ca = Cg(3) = 0$.

2.2 Generalizações e Extensões das Distribuições Logísticas

De maneira generalizada, expressamos que a variável aleatória $Y \sim S(\mu,\phi,m)$ tem distribuição logística m, se sua função densidade é dada por

$$\pi(y;\mu,\phi,\alpha,m) = \frac{1}{C(0,\alpha,m)\sqrt{\phi}} \frac{\exp(-(\frac{y-\mu}{\sqrt{\phi}})^{2/m})}{[1+\exp(-(\frac{y-\mu}{\sqrt{\phi}})^{2/m})]^\alpha}, \quad \alpha > 1 \quad , \quad m > 0 \quad e$$

$$C(0,\alpha,m) = \int_0^\infty \frac{\exp(-x^{2/m})}{[1+\exp(-x^{2/m})]^\alpha} dx,$$

de tal forma que para m=1 e $\alpha = 2$, temos a distribuição logística I, e para m=2 e $\alpha = 2$ temos a distribuição logística II.

Temos que $E(Y) = \mu$, $Var(Y) = \phi\frac{C(2,\alpha,m)}{C(0,\alpha,m)}$ e os momentos centrais de ordem r são

$$E\{(Y-\mu)^r\} = \begin{cases} \phi^{r/2}\dfrac{C(r,\alpha,m)}{C(0,\alpha,m)} = 0, & r \text{ ímpar} \\ \phi^{r/2}\dfrac{C(r,\alpha,m)}{C(0,\alpha,m)}, & r \text{ par} \end{cases}$$

$$em \quad que \quad C(r,\alpha,m) = \int_{-\infty}^{+\infty} x^r \frac{\exp(-x^{2/m})}{[1+\exp(-x^{2/m})]^\alpha} dx.$$

Considere a seguinte expressão $Cg(r) = \frac{E\{(Y-\mu)^r\}}{E\{(Y-\mu)^2\}^{r/2}}$, ou seja, podemos escrever que $Cg(r) = \frac{C(r,m)C(0,m)^{r/2}}{C(0,m)C(2,m)^{r/2}}$, assim para r=4, temos o coeficiente de curtose dada pela expressão $Cc = Cg(4) = \frac{E\{(Y-\mu)^4\}}{E\{(Y-\mu)^2\}^2}$ e portanto, o coeficiente de curtose é $Cc = Cg(4) = \frac{C(4,\alpha,m)C(0,\alpha,m)}{C(2,\alpha,m)C(2,\alpha,m)}$. Já para r=3, temos o coeficiente de assimetria dada pela expressão $Ca = Cg(3) = \frac{E\{(Y-\mu)^3\}}{E\{(Y-\mu)^2\}^{3/2}}$, e portanto $Ca = Cg(3) = 0$.

2.3 Generalizações e Extensões da Distribuição de Cauchy

Há várias generalizações e extensões da distribuição de Cauchy pertencente à classe simétrica nas mais diversas literaturas. A importância desses modelos depende da sensibilidade em explicar ou equacionar os fenômenos através da massa de dados. De maneira generalizada sendo $Y \sim S(\mu,\phi,m)$

sua função densidade é dada por

$$\pi(y;\mu,\phi,m) = \frac{1}{C(0,1,m)\sqrt{\phi}}[1 + (\frac{y-\mu}{\sqrt{\phi}})^{2m}]^{-1}, \quad e \quad m \geq 1$$

para m=1, temos a distribuição Cauchy

Se Y tem distribuição definido acima, então sua esperança e variância são respectivamente $E(y) = \mu$ e $Var(y) = \phi\frac{C(2,1,m)}{C(0,1,m)}$, e os momentos centrais de ordem r são

$$E\{(Y-\mu)^r\} = \begin{cases} \phi^{r/2}\dfrac{C(r,1,m)}{C(0,1,m)} = 0, & r \text{ ímpar} \\ \phi^{r/2}\dfrac{C(r,1,m)}{C(0,1,m)}, & r \text{ par} \end{cases}$$

$$com \quad C(r,1,m) = \int_{-\infty}^{+\infty} \frac{x^r}{[1+x^{2m}]} dx.$$

Por outro lado analisando $Y \sim S(\mu,\phi,\alpha,m)$ com função densidade dada por

$$\pi(y;\mu,\phi,\alpha,m) = \frac{1}{C(0,\alpha,m)\sqrt{\phi}}[1 + (\frac{y-\mu}{\sqrt{\phi}})^{2m}]^{-\alpha}, \quad \alpha > 0, \quad m > 0$$

$$e \quad m\alpha > \frac{1}{2}.$$

temos assim uma distribuição mais geral do que a anterior, uma vez que a expressão anterior é um caso particular desta para $\alpha = 1$. Sendo $\alpha = 1$ e m=1, temos a distribuição Cauchy.

Se Y tem distribuição dada pela expressão acima, então sua esperança e variância se existirem são respectivamente $E(y) = \mu$ e $Var(y) = \phi\frac{C(2,\alpha,m)}{C(0,\alpha,m)}$, e os momentos centrais de ordem r se existirem são

$$E\{(Y-\mu)^r\} = \begin{cases} \phi^{r/2}\dfrac{C(r,\alpha,m)}{C(0,\alpha,m)} = 0, & r \text{ ímpar} \\ \phi^{r/2}\dfrac{C(r,\alpha,m)}{C(0,\alpha,m)}, & r \text{ par} \end{cases}$$

$$em \quad que \quad C(r,\alpha,m) = \int_{-\infty}^{+\infty} \frac{x^r}{[1+x^{2m}]^\alpha} dx$$

e as condições para que existam os momentos centrais de ordem r são $2m\alpha - r > 1$ e $m\alpha > \frac{1}{2}$. Considere a seguinte expressão $Cg(r) = \frac{E\{(Y-\mu)^r\}}{E\{(Y-\mu)^2\}^{R/2}}$, ou seja, $Cg(r) = \frac{C(r,\alpha,m)C(0,\alpha,m)^{r/2}}{C(0,\alpha,m)C(2,\alpha,m)^{r/2}}$. Assim para r=4, temos o coeficiente de curtose dada pela expressão $Cc = Cg(4) = \frac{E\{(Y-\mu)^4\}}{E\{(Y-\mu)^2\}^2}$ e portanto, o coeficiente de curtose é $Cc = Cg(4) = \frac{C(4,\alpha,m)C(0,\alpha,m)}{C(2,\alpha,m)C(2,\alpha,m)}$. Já para r=3, temos o coeficiente de assimetria dada pela expressão $Ca = Cg(3) = \frac{E\{(Y-\mu)^3\}}{E\{(Y-\mu)^2\}^{3/2}}$, e portanto $Ca = Cg(3) = 0$.

2.4 Generalizações e Extensões de algumas Distribuições simétricas

Vejamos o seguinte modelo: Seja $Y \sim S(\mu,\phi,a,b,c,\alpha,\beta,m_1,m_2,m_3)$ de tal forma, que a função densidade é dada por

$$\pi(y;\mu,\phi,a,b,c,\alpha,\beta,m_1,m_2,m_3) = \frac{1}{C(0,a,b,c,\alpha,\beta,m_1,m_2,m_3)\sqrt{\phi}}$$

$$\frac{exp[-a(\frac{y-\mu}{\sqrt{\phi}})^{2m_1}]}{b[1+(\frac{y-\mu}{\sqrt{\phi}})^{2m_2}]^\alpha + c[1+exp(-(\frac{y-\mu}{\sqrt{\phi}})^{2m_3})]^\beta},$$

$$em \quad que \quad C(0,a,b,c,\alpha,\beta,m_1,m_2,m_3) = \int_{-\infty}^{+\infty} \frac{exp(-ax^{2m_1})}{b[1+x^{2m_2}]^\alpha + c[1+exp(-x^{2m_3})]^\beta} dx,$$

$$a \geq 0, \quad b \geq 0, \quad c \geq 0, \quad \alpha \geq 0, \quad \beta \geq 0, \quad m_1 > 0,$$

$$m_2 > 0, \quad m_3 > 0, \quad b + c > 0, \quad etc,$$

as condições adequadas e o mais geral possível sobre os parâmetros, para $C(\cdot)$ seja convergente. Se Y tem distribuição conforme expressão acima, então sua esperança e variância são respectivamente

$$E(y) = \mu \quad e \quad Var(y) = \phi\frac{C(2,a,b,c,\alpha,\beta,m_1,m_2,m_3)}{C(0,a,b,c,\alpha,\beta,m_1,m_2,m_3)},$$

e os momentos centrais de ordem r são

$$E\{(Y-\mu)^r\} = \begin{cases} \phi^{r/2}\dfrac{C(r,a,b,c,\alpha,\beta,m_1,m_2,m_3)}{C(0,a,b,c,\alpha,\beta,m_1,m_2,m_3)} = 0, & r \text{ ímpar} \\ \phi^{r/2}\dfrac{C(r,a,b,c,\alpha,\beta,m_1,m_2,m_3)}{C(0,a,b,c,\alpha,\beta,m_1,m_2,m_3)}, & r \text{ par} \end{cases}$$

$$em \quad que \quad C(r,a,b,c,\alpha,\beta,m_1,m_2,m_3) = \int_{-\infty}^{+\infty} \frac{x^r exp(-ax^{2m_1})}{b[1+x^{2m_2}]^{\alpha} + c[1+exp(-x^{2m_3})]^{\beta}} dx,$$

Por outro lado, analisando $Y \sim S(\mu,\phi,a,b,c,\alpha,\beta,m_1,m_2,m_3)$ e impondo certas condições sobre os parâmetros, obtemos as distribuições normais, logísticas, etc. Vejamos algumas dessas distribuições: Para $a = 0, \quad b = 1, \quad c = 0$ e $m_2\alpha > \frac{1}{2}$, temos uma generalização da distribuição de Cauchy. Para $a = \frac{1}{2m_1}, \quad b = 0$, $c = 1$ e $\beta = 0$, temos uma generalização da distribuição normal. Para $a = 1, \quad b = 1, \quad c = 1$, $\alpha = 0, \quad \beta = 0 \quad e \quad m_1 = \frac{1}{2}$, temos a distribuição de Laplace. Para $a = 1, \quad b = 0, \quad c = 1$ e $m_1 = m_3 = \frac{1}{m}$, temos uma generalização das distribuições Logísticas e assim por diante. É fácil perceber que essa expressão pode ser ampliada de modo abranger mais distribuições, e sua função característica é dada por:

$$\psi_y(t) = \frac{e^{it\mu}}{C(0,a,b,c,\alpha,\beta,m_1,m_2,m_3)}$$

$$\int_{-\infty}^{+\infty} \frac{exp[itx\sqrt{\phi} - ax^{2m_1}]}{b[1+x^{2m_2}]^{\alpha} + c[1+exp(-x^{2m_3})]^{\beta}} dx.$$

Considere a seguinte expressão $Cg(r) = \frac{E\{(Y-\mu)^r\}}{E\{(Y-\mu)^2\}^{R/2}}$, ou seja, podemos escrever que $Cg(r) = \frac{C(r,\cdot)C(0,\cdot)^{r/2}}{C(0,\cdot)C(2,\cdot)^{r/2}}$. Assim para r=4, temos o coeficiente de curtose dada pela expressão $Cc = Cg(4) = \frac{E\{(Y-\mu)^4\}}{E\{(Y-\mu)^2\}^2}$ e portanto, o coeficiente de curtose é $Cc = Cg(4) = \frac{C(4,\cdot)C(0,\cdot)}{C(2,\cdot)C(2,\cdot)}$. Já para r=3, temos o coeficiente de assimetria dada pela expressão $Ca = Cg(3) = \frac{E\{(Y-\mu)^3\}}{E\{(Y-\mu)^2\}^{3/2}}$, e portanto $Ca = Cg(3) = 0$.

3 Conclusão

A principal contribuição teórica deste trabalho consistiu em gerar novas distribuições simétricas, generalizações e extensões da mesma com o intuito de encontrar modelos mais robustos ou mais sensíveis que possam modelar as informações que possivelmente seriam mais complexos do que os modelos já existentes.

Referências

[1] Chmielewski, M.A. (1981). Elliptically symmetric distributions: a review and bibliography. *International Statistical Review.* **49**, 67-74.

[2] Devlin, S.J., Gnanadesikan, R., Kettenring, J.R. (1976). Some multivariate applications of elliptical distributions. *Essays in Probability and Statistics*, **24**, 365 - 393. Tokio: Shinko Tsusho Co.

[3] Fang, K.T., Anderson, T.W. (1990). *Statistical Inference in elliptical Contoured and Related Distribuitions.* New York: Allerton Press.

[4] Fang, K.T., Kotz, S., Ng, K.W. (1990). *Symmetric Multivariate and Related Distributions.* London: Chapman and Hall.

[5] Fang, K.T., Zhang, Y. T. (1990). *Generealized Multivariate Analysis.* London: Springer-Verlag.

[6] Gupta, A.K., Varga, T. (1993). *Elliptically Contoured Models in Statistics.* Kluwer Academic Publishers.

[7] Muirhead,M. (1980). The effects of elliptical ditributions on some standart procedures involving correlation coefficients. *Multivariate Statistical Analysis*, North-Holland, 143 - 159.