A distribuição floor e suas propriedades[1]

Caroline Tenório Mendes de Aquino [2]
José Ailton Alencar Andrade [3]

1 Introdução

Este trabalho apresenta a distribuição floor, que está sendo proposta em [2] com o objetivo de ser uma alternativa à distribuição exponencial. Sua vantagem em relação à distribuição exponencial está na cauda, que é pesada, o que é uma característica de robustez para informações conflitantes na modelagem bayesiana [1]. Também pode-se dizer que a distribuição floor varia o-regularmente. O objetivo deste trabalho é exibir as principais propriedades da distribuição floor, como função de distribuição acumulada, momentos, coeficiente de assimetria e curtose; além da utilização do método de aceitação-rejeição para gerar valores aleatórios da distribuição.

2 Material e métodos

A partir da densidade da distribuição floor, suas propriedades são calculadas. Considere X uma variável aleatória com função densidade de probabilidade (f.d.p) dada por

$$f(x|a) = C(a)\, x^{-a}\, e^{\lfloor \ln x \rfloor}\, I_{(1,\infty)}(x), \tag{1}$$

em que $C(a)$ é a constante normalizadora (calculada na seção 3), $a > 2$ e $\lfloor . \rfloor$ é a função piso. Então diz-se que X segue a distribuição floor. A Figura 1(a) mostra a densidade quando $a = 3$.

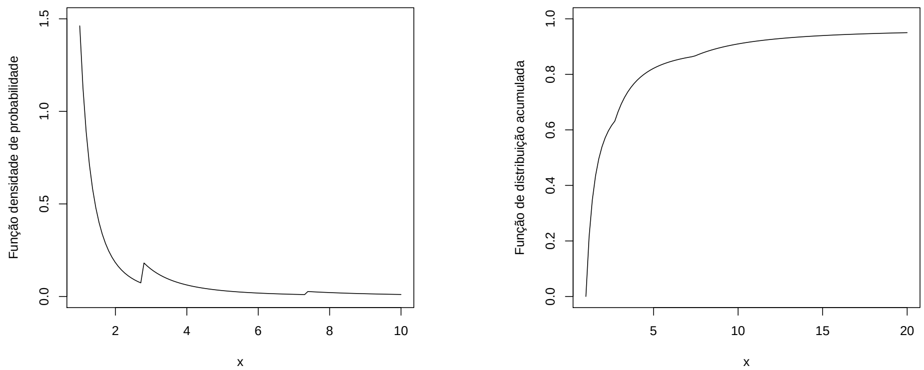
Para a obtenção de valores aleatórios da distribuição floor, o método de aceitação-rejeição foi utilizado [5], implementado no software R [4]. O algoritmo do método é como segue:

1. gere um valor de uma variável aleatória Y com distribuição g;
2. gere um valor de uma variável U com distribuição uniforme (independente de Y);
3. se $U \leq \frac{f(Y)}{cg(Y)}$, então faça $X = Y$ (aceite); caso contrário volte ao passo 1 (rejeite).

No algoritmo, $f(Y)$ é a distribuição alvo, da qual queremos gerar valores (neste caso, a floor). Além da distribuição alvo, é necessária outra distribuição com comportamento semelhante, a distribuição candidata. Para isso, a distribuição exponencial a dois parâmetros [3],

[1]Trabalho apresentado no evento: 58ª RBras e 15º SEAGRO. 2013. (Congresso).

[2]DEMA/UFC. e-mail: *mendes.aquino@gmail.com*
[3]DEMA/UFC. e-mail: *jaa.andrade@gmail.com*



(a) Densidade (b) Distribuição acumulada

Figura 1: Densidade e distribuição acumulada quando $a = 3$

com parâmetro de localização $\phi = 1$ foi utilizada. No algoritmo acima, ela ocupa o lugar de $g(Y)$. Sua f.d.p é definida por

$$g(y|\lambda, \phi = 1) = \frac{1}{\lambda}\, e^{-\frac{(y-1)}{\lambda}}\, I_{(1,\infty)}(y),$$

em que $\lambda > 0$ e $y > \phi = 1$.

3 Resultados e discussões

3.1 Geração de valores aleatórios

Como visto na seção anterior, o método de aceitação-rejeição foi utilizado para gerar valores aleatórios da distribuição floor. Para verificar se o método estava satisfatório, foi feito um histograma com a distribuição dos dez mil valores aleatórios gerados através do método para $a = 9$, com a curva da densidade sobre o histograma (ver Figura 2).

Além do histograma, foram calculadas medidas como média e variância dos valores gerados, e estes foram comparados com os valores da esperança e variância da distribuição. No caso para $a = 9$, a média dos valores foi igual a $1,147$ (com três casas decimais), enquanto que a esperança da distribuição floor é aproximadamente igual a $1,144$. A variância amostral foi igual a $0,020817$ (com seis casas decimais), enquanto a variância da distribuição floor quando $a = 9$ é aproximadamente igual a $0,029560$.

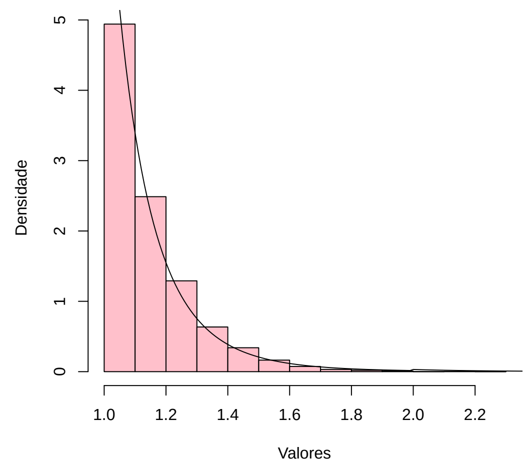


Figura 2: Histograma dos valores aleatórios da distribuição floor com $a = 9$

3.2 Constante normalizadora

Primeiramente, a constante normalizadora é calculada. Para que a equação 1 seja uma legítima função densidade de probabilidade, é necessário que

$$\int_1^\infty C(a) x^{-a} e^{\lfloor \ln x \rfloor} \, dx \quad = \quad 1. \tag{2}$$

Com a transformação $t = \ln(x)$,

$$\begin{aligned}
\frac{1}{C(a)} &= \int_0^\infty e^{-at} e^{\lfloor t \rfloor} e^t \, dt \\
&= \int_0^\infty e^{-t(a-1)} e^{\lfloor t \rfloor} \, dt \\
&= \int_0^1 e^{-t(a-1)} e^0 \, dt + \int_1^2 e^{-t(a-1)} e^1 \, dt + \cdots \\
&= \sum_{n=0}^\infty e^n \int_n^{n+1} e^{-t(a-1)} \, dt \\
&= \frac{1}{n-1} \left[\sum_{n=0}^\infty e^{-n(a-2)} - \sum_{n=0}^\infty e^{-(n+1)(a-1)+n} \right] \qquad (3)
\end{aligned}$$

As duas somas da equação 3 são séries geométricas com razão $e^{-(a-2)}$, de forma que

$$C(a) = \frac{(a-1)(e^2 - e^a)}{e - e^a}.$$

3.3 Distribuição acumulada

A função de distribuição acumulada da distribuição floor é dada por

$$F(t) = -\frac{C(a)}{a-1}\left\{e^{-\ln(t)(a-1)+\lfloor \ln(t)\rfloor} + (e^{-1}-1)\left[\frac{1-e^{-\lfloor \ln(t)\rfloor}(a-2)}{e^{a-2}-1}\right] - 1\right\},$$

para $t > 1$. A Figura 1(b) mostra a função de distribuição acumulada para $a = 3$.

3.4 Momento de ordem r

O momento de ordem r é calculado substituindo a por $(a-r)$ na integral da equação 2. Assim,

$$E(X^r) = C(a)\int_1^{\infty} x^{-(a-r)}\, e^{\lfloor \ln x\rfloor}\, dx = \frac{C(a)}{C(a-r)}, \tag{4}$$

que existe apenas quando $a > 2$ e $r < a-2$.

3.5 Variância, coeficiente de assimetria e curtose

Através da equação 4, a variância é obtida.

$$Var(X) = \frac{C(a)[C^2(a-1) - C(a)C(a-2)]}{C(a-2)C^2(a-1)} \tag{5}$$

Para encontrar o terceiro momento central (em torno da média), a equação 4 foi utilizada e o resultado obtido foi

$$m_3 = \frac{C(a)[C^3(a-1)C(a-2) - 3C(a)C^2(a-1)C(a-3) + 2C^2(a)C(a-2)C(a-3)]}{C^3(a-1)C(a-2)C(a-3)}$$

Assim, usando a equação 5, $Var^{3/2}(X) = \sigma^3 = \frac{C^{3/2}(a)[C^2(a-1)-C(a)C(a-2)]^{3/2}}{C^{3/2}(a-2)C^3(a-1)}$. Agora o coeficiente pode ser calculado: $V(X) = \frac{m_3}{\sigma^3}$:

$$\begin{aligned} V(X) &= \frac{m_3}{\sigma^3} \\ &= \frac{C^{1/2}(a-2)[C^3(a-1)C(a-2) - 3C(a)C^2(a-1)C(a-3) + 2C^2(a)C(a-2)C(a-3)]}{C^{1/2}(a)C(a-3)[C^2(a-1) - C(a)C(a-2)]^{3/2}} \end{aligned}$$

O valor de $V(X)$ sempre será maior do que *zero*, o que indica que a distribuição floor é assimétrica à direita, para qualquer valor de $a > 2$.

A curtose é obtida de forma semelhante à usada para calcular o coeficiente de assimetria, obtendo-se primeiro o quarto momento central e a variância ao quadrado, σ^4. Assim, a curtose é dada por

$$K(X) = \frac{m_4}{\sigma^4} - 3$$

$$= \frac{C(a-1)C(a-2)[.]}{C(a)C(a-3)C(a-4)[C^2(a-1)-C(a)C(a-2)]^2} - 3,$$

em que $[.] = [C^3(a-1)\prod_{i=2}^{3} C(a-i) - 4C(a-4)\prod_{i=0}^{2} C(a-i) + 6C^2(a)C(a-1)\prod_{i=3}^{4} C(a-i) - 4C(a)\prod_{i=0}^{4} C(a-i) + C^2(a)\prod_{i=2}^{4} C(a-i)]$.

4 Conclusões

A partir dos resultados vistos neste trabalho, algumas conclusões podem ser tomadas. A Figura 1(a) mostra que a densidade da distribuição floor decresce a *zero* oscilando, com alguns picos na cauda devido à função piso. Na Figura 1(b), também pode-se perceber que a função de distribuição acumulada tende a *um* oscilando.

O método de aceitação-rejeição simulou bem os valores da distribuição floor, como mostra a Figura 2. Como a distribuição floor é nova, são necessários mais estudos para verificar sua utilização em modelagem robusta bayesiana. Porém, este trabalho servirá como base para o conhecimento e difusão dessa distribuição.

Referências

[1] ANDRADE, J. A. A. **Bayesian Robustness Modelling Using Regularly Varying Distributions**. Sheffield: University of Sheffield, 2005. 117 p. Tese (Doutorado) - Department of Probability and Statistics. School of Mathematics and Statistics, University of Sheffield, Sheffield, 2005.

[2] ANDRADE, J. A. A.; OMEY, E. On the Characterization of some o-regularly varying distributions with applications on robustness modelling. (Em submissão). 2013

[3] DETZEL, D. H. M.; MINE, M. R. M. Modelagem de Quantidades Precipitadas em Escala Diária: Uma Análise Comparativa. **RBRH Revista Brasileira de Recursos Hídricos**. v. 16, n. 2, p. 101-110, 2011

[4] R CORE TEAM. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. 2012. ISBN 3-900051-07-0. Disponível em: http://www.R-project.org/. Vienna, 2012.

[5] SIGMAN, K. **Acceptance-Rejection Method**. Disponível em: http://www.columbia.edu/~ks20/4703-Sigman/4703-07-Notes-ARM.pdf. Acesso em: 25 de fevereiro de 2013.