

A Distribuição Half-Normal Generalizada Discreta: uma distribuição alternativa para a análise de dados de contagem

The Generalized Discrete Half-Normal Distribution: an alternative distribution for analysing count data

Josmar Mazucheli^I
 Ricardo Puziol de Oliveira^{II}
 Jean Carlos Cardoso^{III}

Resumo

Geralmente, dados que são obtidos por processos de contagem, estritamente discretos ou discretizados (provenientes de trunca- mentos e/ou arredondamentos), são analisados, sem exaustão, pelas distribuições Geométrica, Logarítmica, Poisson e Binomial negativa. Nos *últimos* anos um grande número de distribuições discretas vêm sendo propostas na literatura a partir da discretização de variáveis aleatórias contínuas. Muitos dos métodos de discretização preservam uma ou mais características da versão contínua, sendo que a proposta de Nakagawa e Osaki (1975) é a mais utilizada. Neste artigo, a partir dessa metodologia, a qual faz uso da função de sobrevivência, é proposta a versão discreta da distribuição Half-Normal generalizada contínua, introduzida na literatura por Cooray e Ananda (2008). Discute-se algumas de suas propriedades e via simulações Monte Carlo avalia-se o viés e a acurácia das estimativas obtidas pelo método da máxima verossimilhança e pelo método dos momentos. Alguns conjuntos de dados discretos da literatura são considerados para ilustrar a aplicabilidade da distribuição proposta.

Palavras-chave: Discretização; Distribuição Half-Normal generalizada; Método dos momentos; Simulação Monte Carlo; Verossimilhança

Abstract

In general, data that are obtained by counting processes, strictly discrete or discretized (from truncations and/or rounding), are analyzed, without exhaustion, by the Geometric, Logarithmic, Poisson and Negative Binomial distributions. In recent years a large number of discrete distributions have been proposed in the literature from the discretization of continuous random variables. Many of the discretization methods preserve one or more characteristics of the continuous version, with the proposal of Nakagawa e Osaki (1975) being the most used. In this paper, from this methodology, which makes use of the survival function, we propose the discrete version of the continuous generalized Half-Normal distribution, introduced in the literature by Cooray e Ananda (2008). Some of its properties are discussed and Monte Carlo simulations evaluate the bias and accuracy of the estimates obtained by the maximum likelihood method and method of moments. Some discrete data sets found in the literature are considered to illustrate the applicability of the proposed distribution.

Keywords: Discretization; Generalized Half-Normal distribution; Method of moments; Monte Carlo simulation; Likelihood

^I Departamento de Estatística, Universidade Estadual de Maringá, Maringá, PR, Brasil - jmazucheli@gmail.com

^{II} Departamento de Medicina Social, Universidade de São Paulo, Ribeirão Preto, SP, Brasil - rpuziol.oliveira@gmail.com

^{III} Departamento de Estatística, Universidade Estadual de Maringá, Maringá, PR, Brasil- jeancarlos.card@gmail.com



1 Introdução

O número de vezes que um dispositivo eletrônico é ligado antes de apresentar algum defeito, o número de dias que um paciente permanece internado na UTI antes de receber alta e o número de semanas, meses ou anos que um transplantado sobrevive são exemplos de dados de sobrevivência discretos na maioria das vezes analisados por meio de distribuições contínuas. Klein e Moeschberger (1997) traz em seu primeiro capítulo exemplos de dados de sobrevivência discretos, entretanto, analisados no decorrer do texto via distribuições contínuas. O mesmo ocorre com os vários exemplos apresentados, entre outros, em Collett (2003), Lawless (2003) e Lee e Wang (2003).

Nos últimos anos várias distribuições discretas têm sido propostas na literatura, e, apesar de suas atraentes potencialidades, não são muito usadas na análise de dados de sobrevivência discretos por natureza, discretizados ou em dados de contagem. Um dos primeiros, senão o primeiro artigo científico que propõe uma distribuição discreta (derivado da distribuição Weibull) data da década de 1970. Desde então uma infinidade de novas distribuições discretas vem sendo propostas. Uma revisão dos vários métodos de discretização e das distribuições discretizadas por eles pode ser encontrada em Chakraborty (2015), Lai (2013) e Bracquemond e Gaudoin (2003). Nestes trabalhos são discutidos os métodos baseado na função de sobrevivência (Nakagawa e Osaki, 1975), o baseado em uma série infinita (Good, 1953), o baseado na função de risco (Stein e Dattero, 1984), o composto por dois estágios (Chakraborty, 2015), o baseado na função de risco reversa (Ghosh et al., 2013), o via função de distribuição acumulada (Roy e Dasgupta, 2001) e o método análogo ao de Pearson (Pearson, 1895).

O objetivo deste artigo é, a partir do método de discretização baseado na função de sobrevivência (Nakagawa e Osaki, 1975), introduzir a distribuição discreta análoga a distribuição Half-Normal generalizada contínua, introduzida na literatura por Cooray e Ananda (2008). Na Seção 2 o método de discretização é brevemente descrito. A Seção 3 apresenta a versão contínua da distribuição Half-Normal generalizada (HNG) e logo em seguida é introduzida a versão discreta. Na Seção 4 o método da máxima verossimilhança e o método dos momentos são apresentados. Um estudo de simulação Monte Carlo na Seção 5 é usado para avaliar o viés e a acurácia das estimativas obtidos por esses dois métodos. A Seção 6 apresenta três aplicações considerando dados reais da literatura e a Seção 7 encerra este artigo com algumas conclusões.

2 Discretização via Função de Sobrevivência

Um dos principais métodos disponível na literatura para se obter a distribuição discreta análoga a distribuição de uma variável aleatória contínua é o que utiliza a função de sobrevivência. Este método foi proposto por Nakagawa e Osaki (1975) e tem como principal característica a

preservação da expressão da função de sobrevivência original em sua parte inteira (Kemp, 2004; Chakraborty, 2015). Alguns exemplos de distribuições discretas geradas por este método são: a distribuição Weibull discreta (Nakagawa e Osaki, 1975), a distribuição Weibull geométrica discreta (Bracquemond e Gaudoin, 2003), a distribuição Gumbel discreta (Chakraborty e Chakravarty, 2014), a distribuição gama discreta (Chakraborty e Chakravarty, 2012), a distribuição Lindley discreta (Gómez-Déniz e Calderín-Ojeda, 2011; Bakouch et al., 2014; Oliveira et al., 2017), a distribuição Rayleigh inversa discreta (Hussain e Ahmad, 2014), a distribuição Weibull inversa discreta (Aghababaei Jazi et al., 2010), a distribuição Maxwell discreta (Krishna e Pundir, 2007), e as distribuições Burr e Pareto discretas (Krishna e Pundir, 2009; Mazucheli et al., 2018).

Definição 2.1. *Seja X uma variável aleatória contínua. Se X tem função de sobrevivência $S_X(x) = P(X \geq x)$, então a variável aleatória $Y = \lfloor X \rfloor$, em que $\lfloor X \rfloor$ indica a parte inteira menor ou igual a X , tem função de probabilidade definida por:*

$$\begin{aligned} P(Y = k) &= P(k < X < k + 1) = \\ &= P(X \geq k) - P(X \geq k + 1) = \\ &= S_X(k) - S_X(k + 1). \end{aligned} \quad (1)$$

As principais vantagens desse método de discretização é que o mesmo mantém, em termos de estrutura, boa parte das propriedades da distribuição contínua após a discretização. Por exemplo, na distribuição obtida por (1) tem-se $S_Y(k) = S_X(k)$ e, por consequência, a função de distribuição acumulada é dada por $F_Y(k) = F_X(k)$. Além disso, o quantil da distribuição discreta, neste caso, é dado por $Q_Y(k) = \lfloor Q_X(k) \rfloor$, ou seja, a amplitude interquartilica e a mediana são as mesmas nas versões discretas e contínuas. Por fim, a função de risco por este método é definida por:

Definição 2.2. *Seja Y uma variável aleatória discreta, com função de probabilidade e de sobrevivência $p(k)_e$, $S_Y(k)$ respectivamente. A função de risco de Y é definida como a probabilidade condicional de que a falha é observada em k , dado que esta não ocorreu antes de k , e é escrita na forma:*

$$\begin{aligned} h_Y(k) &= \frac{p(k)}{S_Y(k)} = \frac{S_Y(k) - S_Y(k + 1)}{S_Y(k)} = \\ &= 1 - \frac{S_Y(k + 1)}{S_Y(k)} \quad k \in \mathbb{Z}_+ \end{aligned} \quad (2)$$

sendo limitada ao intervalo (0,1) para todo $k \in \mathbb{Z}_+$.

A proposta neste artigo é usar em (1) a função de sobrevivência da distribuição Half-Normal generalizada (HNG) e assim obter a função de probabilidade Half-Normal generalizada discreta (HNGD). A distribuição HNG tem várias propriedades interessantes e no melhor de nosso conhecimento sua versão discreta não foi explorada na literatura.

3 A Distribuição Half-Normal Generalizada

3.1 Versão Contínua

Cooray e Ananda (2008) introduziram a distribuição HNG com dois parâmetros que tem como caso particular a distribuição Half-Normal (HN). Para uma variável aleatória X não negativa que segue a distribuição HNG com parâmetro de escala $\theta > 0$ e de forma $\alpha > 0$ tem-se que a função densidade de probabilidade e a função distribuição acumulada são escritas, respectivamente, como:

$$f(x | \alpha, \theta) = \frac{2\alpha}{\theta} \left(\frac{x}{\theta}\right)^{\alpha-1} \phi\left[\left(\frac{x}{\theta}\right)^\alpha\right] \text{ e } F(x | \alpha, \theta) = 2\Phi\left[\left(\frac{x}{\theta}\right)^\alpha\right] - 1 \quad (3)$$

em que $x \geq 0$; $\phi(\cdot)$ e $\Phi(\cdot)$ são, respectivamente, a função densidade de probabilidade e de distribuição acumulada de uma variável aleatória com distribuição normal padrão. Para $\alpha = 1$ se tem distribuição HN como caso particular.

Outras características da distribuição HNG são discutidas em Cooray e Ananda (2008). Estes autores também apresentaram o método da máxima verossimilhança e estudaram, por meio de simulações, as probabilidades de cobertura dos intervalos de confiança assintóticos. Recentemente, os autores Mazucheli e Dey (2018) deduziram, pela metodologia de Cox e Snell (1968), as expressões analíticas para a correção do viés das estimativas de α e θ obtidas pelo método da máxima verossimilhança. Cooray e Ananda (2008) também mostraram que a função no lado esquerdo de (3) é monotonicamente decrescente para $\alpha < 1$ e unimodal para $\alpha \geq 1$ com ponto de inflexão em $\theta\left(\frac{\alpha-1}{\alpha}\right)^{\frac{1}{2\alpha}}$. A função de risco, escrita como:

$$h(x | \alpha, \theta) = \frac{\alpha}{\theta} \left(\frac{x}{\theta}\right)^{\alpha-1} \phi\left[\left(\frac{x}{\theta}\right)^\alpha\right] \Phi\left[-\left(\frac{x}{\theta}\right)^\alpha\right]^{-1} \quad x \geq 0 \quad (4)$$

pode ser monotonicamente crescente, monotonicamente decrescente ou assumir a forma de banheira. Essas formas da função de risco fazem com que a distribuição HNG seja uma boa alternativa a distribuição Weibull para a análise de dados de sobrevivência.

3.2 Versão Discreta

Definição 3.1. Uma variável aleatória X que toma valores no conjunto \mathbb{Z}_+ segue a distribuição HNGD com parâmetro de escala α e forma θ se sua função de probabilidade é escrita na forma:

$$P(X = x | \alpha, \theta) = 2\phi\left[\left(\frac{x+1}{\theta}\right)^\alpha\right] - 2\Phi\left[\left(\frac{x}{\theta}\right)^\alpha\right]. \quad (5)$$

O comportamento de (5) para alguns valores de α e θ é ilustrado na Figura 1 em que pode-se notar que (5) é monótona decrescente para $\alpha < 1$ e unimodal caso contrário.

As funções de distribuição acumulada e de sobrevivência são dadas, respectivamente, por:

$$F(x | \alpha, \theta) = 2\Phi\left[\left(\frac{\lfloor x \rfloor}{\theta}\right)^\alpha\right] - 1 \text{ e } S(x | \alpha, \theta) = 2\left(1 - \Phi\left[\left(\frac{\lfloor x \rfloor}{\theta}\right)^\alpha\right]\right). \quad (6)$$

É importante mencionar que para uma variável aleatória discreta a função de risco é limitada, $h(x) < 1$, logo ela não é equivalente a $-\log S(x)$ como no caso contínuo (Hussain e Ahmad, 2014). Neste caso, foi considerado neste artigo uma segunda função de risco à distribuição HNGD dada por:

$$h^*(x | \alpha, \theta) = \log\left[\frac{S(x | \alpha, \theta)}{S(x+1 | \alpha, \theta)}\right] = \log\left\{\Phi\left[-\left(\frac{x}{\theta}\right)^\alpha\right] - \log\left\{\Phi\left[-\left(\frac{x+1}{\theta}\right)^\alpha\right]\right\}, \quad x = 0, 1, 2, \dots \quad (7)$$

Figura 1 - Comportamento da função de probabilidade da distribuição HNGD

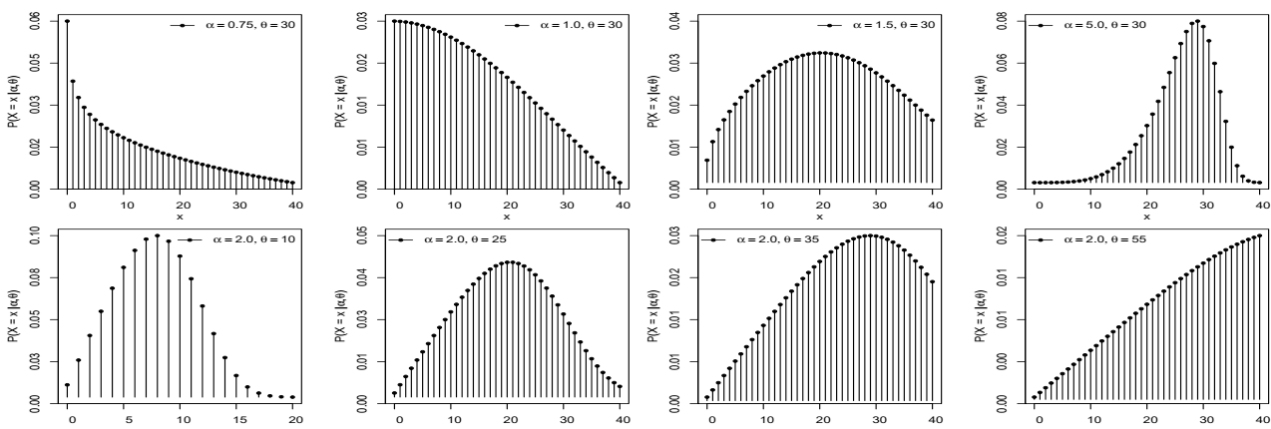
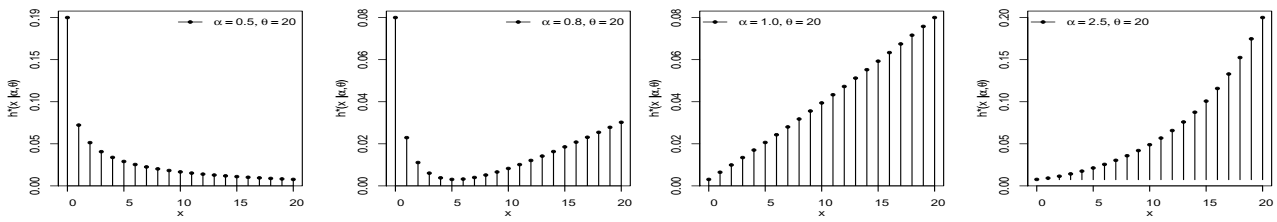


Figura 2 – Comportamento da função de risco da distribuição HNGD



O comportamento de (7) para alguns valores de α e θ é ilustrado na Figura 2 em que pode-se notar que (7) é monótona decrescente, monótona crescente e em forma de banheira. Isto é, a função de risco da distribuição proposta possui uma flexibilidade maior comparada aos modelos tradicionais apresentados na literatura como Geométrico, Poisson e Binominal Negativa.

Para a distribuição HNGD a função quantil é obtida pela transformação inversa como no caso contínuo com restrição aos números inteiros, isto é, $Q(p) = \lfloor F^{-1}(p) \rfloor$. Logo, a função quantil é escrita como:

$$Q(p | \alpha, \theta) = \left\lfloor \theta \left[\Phi^{-1} \left(\frac{p+1}{2} \right)^{\frac{1}{\alpha}} \right] \right\rfloor$$

em que $0 < p < 1$. Neste caso, se $p = 0.5$, a mediana é dada por:

$$Q(0.5 | \alpha, \theta) = \left\lfloor \theta \left[\Phi^{-1}(0.75)^{\frac{1}{\alpha}} \right] \right\rfloor.$$

3.3 Função Geradora de Momentos

Seja X uma variável aleatória discreta com distribuição HNGD. Por definição, a função geradora de momentos é escrita como:

$$M_X(t) = \mathbb{E}(e^{tx}) = \sum_{x=0}^{\infty} 2e^{tx} \left\{ \Phi \left[\left(\frac{x+1}{\theta} \right)^{\alpha} \right] - \Phi \left[\left(\frac{x}{\theta} \right)^{\alpha} \right] \right\}. \quad (8)$$

Embora a série em (8) não tenha forma analítica fechada sua convergência pode ser verificada observando que $\Phi(u) = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{u}{\sqrt{2}} \right) \right]$,

$$\text{em que } \operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n z^{2n+1}}{n!(2n+1)}$$

é uma série de Maclaurin. A esperança e variância, neste caso, são dadas, respectivamente, por:

$$\mathbb{E}(X) = \sum_{x=0}^{\infty} 2x \left\{ \Phi \left[\left(\frac{x+1}{\theta} \right)^{\alpha} \right] - \Phi \left[\left(\frac{x}{\theta} \right)^{\alpha} \right] \right\}$$

Tabela 1 – Índice de dispersão da distribuição HNGD para alguns valores de α e θ

$\theta \backslash \alpha$	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	5.5	6.0
0.5	1.66	0.96	1.00	1.00	0.61	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.0	2.68	0.91	0.71	0.68	0.68	0.68	0.68	0.68	0.68	0.68	0.68	0.68
1.5	3.70	1.07	0.66	0.47	0.35	0.26	0.20	0.16	0.13	0.10	0.09	0.07
2.0	4.72	1.27	0.72	0.50	0.37	0.30	0.26	0.23	0.21	0.20	0.19	0.18
2.5	5.74	1.48	0.81	0.54	0.40	0.30	0.24	0.19	0.16	0.13	0.11	0.10
3.0	6.75	1.69	0.91	0.60	0.43	0.33	0.27	0.23	0.20	0.18	0.16	0.15
3.5	7.77	1.92	1.02	0.66	0.47	0.36	0.29	0.24	0.20	0.16	0.14	0.12
4.0	8.78	2.14	1.13	0.73	0.52	0.39	0.31	0.26	0.22	0.19	0.17	0.15
4.5	9.78	2.36	1.24	0.80	0.57	0.43	0.34	0.28	0.23	0.20	0.17	0.15
5.0	10.79	2.59	1.35	0.87	0.62	0.46	0.37	0.30	0.25	0.21	0.18	0.16
5.5	11.80	2.81	1.46	0.94	0.66	0.50	0.39	0.32	0.27	0.23	0.19	0.17
6.0	12.80	3.04	1.58	1.01	0.71	0.54	0.42	0.34	0.28	0.24	0.21	0.18

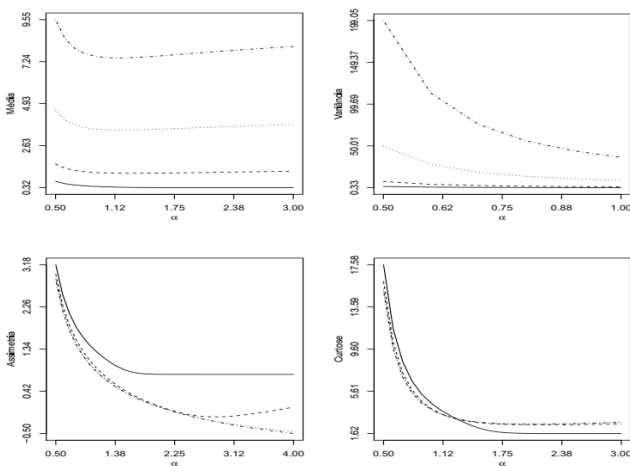
e

$$\text{Var}(X) = \sum_{x=0}^{\infty} 2x^2 \left\{ \Phi \left[\left(\frac{x+1}{\theta} \right)^\alpha \right] - \Phi \left[\left(\frac{x}{\theta} \right)^\alpha \right] \right\} - \left\{ \sum_{i=0}^{\infty} 2x \left[\Phi \left[\left(\frac{x+1}{\theta} \right)^\alpha \right] - \Phi \left[\left(\frac{x}{\theta} \right)^\alpha \right] \right] \right\}^2$$

que também não possuem forma analítica fechada, porém podem ser calculadas numericamente. Estas quantidades existem e são ilustradas na Tabela 1 por meio do cálculo do índice de dispersão que é uma medida usada para verificar a subdispersão, equidispersão e superdispersão. Dados com superdispersão são comuns em aplicações. Um número reduzido de modelos comportam a subdispersão, no caso, da distribuição HNGD verifica-se que a mesma pode ser uma alternativa para modelar dados com subdispersão, equidispersão ou superdispersão uma vez que seu índice de dispersão pode ser menor que 1 (subdispersão), igual a 1 (equidispersão) e maior que 1 (superdispersão).

O comportamento da média, variância, assimetria e curtose, em função de α é ilustrado na Figura 3. Pode-se observar que o coeficiente de assimetria assume valores positivos ou negativos que é uma propriedade importante na análise de dados com assimetria negativa ou positiva.

Figura 3 – Comportamento da média, variância, assimetria e curtose da distribuição HNGD. $\theta = 1.0$ (—), $\theta = 2.0$ (---), $\theta = 5$ (···) e $\theta = 10.0$ (----).



4 Métodos de Estimação

4.1 Estimação Via O Método da Máxima Verossimilhança

Seja X_1, X_2, \dots, X_n uma amostra aleatória de tamanho n da distribuição HNGD com parâmetros α e θ e função de probabilidade dada por (5). A função de verossimilhança pode ser escrita na forma:

$$L(\alpha, \theta | \mathbf{x}) = \prod_{i=1}^n 2\Phi \left[\left(\frac{x_i + 1}{\theta} \right)^\alpha \right] - 2\Phi \left[\left(\frac{x_i}{\theta} \right)^\alpha \right] \quad (9)$$

cujos o logaritmo da função de verossimilhança é dado por:

$$\ell(\alpha, \theta | \mathbf{x}) = \prod_{i=1}^n \log \left[2\Phi \left[\left(\frac{x_i + 1}{\theta} \right)^\alpha \right] - 2\Phi \left[\left(\frac{x_i}{\theta} \right)^\alpha \right] \right] \quad (10)$$

As estimativas de máxima verossimilhança de α e θ são obtidos como soluções das equações não-lineares

$$\frac{\partial}{\partial \alpha} \ell(\alpha, \theta | \mathbf{x}) = 0 \text{ e } \frac{\partial}{\partial \theta} \ell(\alpha, \theta | \mathbf{x}) = 0 \text{ em que:}$$

$$\frac{\partial \ell(\alpha, \theta | \mathbf{x})}{\partial \alpha} = \sum_{i=1}^n \frac{\Phi \left[\left(\frac{x_i + 1}{\theta} \right)^\alpha \right] \left(\frac{x_i + 1}{\theta} \right)^\alpha \log \left(\frac{x_i + 1}{\theta} \right) - \Phi \left[\left(\frac{x_i}{\theta} \right)^\alpha \right] \left(\frac{x_i}{\theta} \right)^\alpha \log \left(\frac{x_i}{\theta} \right)}{\Phi \left[\left(\frac{x_i + 1}{\theta} \right)^\alpha \right] - \Phi \left[\left(\frac{x_i}{\theta} \right)^\alpha \right]}$$

$$\frac{\partial \ell(\alpha, \theta | \mathbf{x})}{\partial \theta} = \sum_{i=1}^n \frac{\Phi \left[\left(\frac{x_i}{\theta} \right)^\alpha \right] \left(\frac{x_i}{\theta} \right)^\alpha \frac{-\alpha}{\theta} - \Phi \left[\left(\frac{x_i + 1}{\theta} \right)^\alpha \right] \left(\frac{x_i + 1}{\theta} \right)^\alpha \frac{-\alpha}{\theta}}{\Phi \left[\left(\frac{x_i + 1}{\theta} \right)^\alpha \right] - \Phi \left[\left(\frac{x_i}{\theta} \right)^\alpha \right]}$$

As estimativas de máxima verossimilhança $\hat{\theta}$ e $\hat{\alpha}$ de α e θ não podem ser obtidas analiticamente. Ainda, a matrix informação de Fisher é aproximada pelas segundas derivadas do logaritmo da função de verossimilhança com respeito a α e θ localmente nas estimativas de máxima verossimilhança $\hat{\theta}$ e $\hat{\alpha}$.

4.2 Estimação Via o Método dos Momentos

Para aplicar o método de momentos para estimar α e θ temos que equacionar os momentos da população com os momentos amostrais e resolver duas equações simultaneamente. Denotemos então por μ_1 e μ_2 o primeiro e segundo momento em torno da origem, respectivamente. Logo:

$$\mu_1 = 2 \sum_{i=0}^{\infty} x_i \left\{ \Phi \left[\left(\frac{x_i + 1}{\theta} \right)^\alpha \right] - \Phi \left[\left(\frac{x_i}{\theta} \right)^\alpha \right] \right\} \text{ e } \mu_2 = 2 \sum_{i=0}^{\infty} x_i^2 \left\{ \Phi \left[\left(\frac{x_i + 1}{\theta} \right)^\alpha \right] - \Phi \left[\left(\frac{x_i}{\theta} \right)^\alpha \right] \right\} \quad (11)$$

Os momentos amostrais correspondentes aos momentos populacionais, são dados por:

$$M_1 = \frac{1}{n} \sum_{i=1}^n x_i \text{ e } M_2 = \frac{1}{n} \sum_{i=1}^n x_i^2. \quad (12)$$

Para obtermos as estimativas de α e θ , podemos utilizar a proposta em Khan et al. (1989), que consiste em encontrar α e θ , que minimizem a função:

$$\ell(\alpha, \theta | M_1, M_2) = (M_1 - \mu_1)^2 + (M_2 - \mu_2)^2. \quad (12)$$

5 Estudo de Simulação

Nesta seção são apresentados os resultados de um estudo de simulação Monte Carlo usado para avaliar o viés e o erro-quadrático médio das estimativas obtidas pelo método da máxima verossimilhança e pelos métodos dos momentos. Foram gerados amostras de tamanho $n = 20, 50, \dots, 170, 200$, $\alpha = 0.5, 1.0, 3.0, 5.0, 10.0$ e $\theta = 1.0$ e 10.0 . Para cada uma das combinações de n , α e θ foram simuladas, pelo método da transformação inversa, $N = 10000$ amostras pseudo-

aleatória da distribuição HNGD. Foram estimados o viés e a raiz quadrada do erro quadrático médio. Estas estimativas são apresentadas nas Tabelas 2 e 3.

Ao observarmos o vício de $\hat{\alpha}$, percebemos que θ exerce influência sobre o mesmo. Para $\theta = 1.0$ e $\alpha > 1$, temos oscilações na estimativa do vício de α para ambos os métodos de estimação, fazendo com que o vício se torne instável, isto é, decresça e cresça a cada acréscimo no tamanho da amostra. De forma semelhante, o vício $\hat{\theta}$ de também oscila, entretanto, a amplitude dessa oscilação é pequena e centrada em zero. Mediante isto, podemos

Tabela 2 – Estimativas do viés e da raiz quadrada do erro-quadrático médio (REQM) dos estimadores de máxima verossimilhança (MV) e dos estimadores obtidos pelo método dos momentos (MM) para $\theta = 1.0$.

α	v	MV				MM			
		Viés		REQM		Viés		REQM	
		α	θ	α	θ	α	θ	α	θ
0.5	20	0.0747	0.0282	0.0407	0.1451	0.1139	0.0960	0.0417	0.1319
	50	0.0317	0.0150	0.0129	0.0689	0.0526	0.0511	0.0141	0.0689
	80	0.0189	0.0093	0.0070	0.0436	0.0333	0.0336	0.0080	0.0451
	110	0.0135	0.0067	0.0048	0.0320	0.0247	0.0252	0.0057	0.0337
	140	0.0106	0.0057	0.0037	0.0252	0.0197	0.0205	0.0044	0.0270
	170	0.0087	0.0045	0.0030	0.0207	0.0164	0.0169	0.0036	0.0224
	200	0.0075	0.0041	0.0025	0.0176	0.0141	0.0147	0.0030	0.0193
1.0	20	-0.0134	0.0300	0.0716	0.0532	0.0145	0.0278	0.0794	0.0494
	50	0.0321	0.0024	0.0389	0.0186	0.0326	0.0030	0.0380	0.0181
	80	0.0355	-0.0010	0.0289	0.0114	0.0349	-0.0009	0.0283	0.0113
	110	0.0308	-0.0012	0.0222	0.0083	0.0302	-0.0012	0.0218	0.0082
	140	0.0257	-0.0012	0.0176	0.0065	0.0251	-0.0013	0.0173	0.0064
	170	0.0215	-0.0009	0.0142	0.0054	0.0209	-0.0011	0.0140	0.0053
	200	0.0180	-0.0008	0.0118	0.0046	0.0175	-0.0009	0.0116	0.0045
3.0	20	0.0619	-0.0029	0.0095	0.0046	0.0716	0.0011	0.0150	0.0052
	50	0.0175	0.0007	0.0013	0.0021	0.0175	0.0007	0.0013	0.0021
	80	0.0618	0.0004	0.0092	0.0012	0.0618	0.0004	0.0092	0.0012
	110	-0.0096	0.0004	0.0003	0.0010	-0.0112	0.0004	0.0004	0.0010
	140	0.0126	-0.0017	0.0102	0.0006	0.0114	0.0011	0.0097	0.0007
	170	0.0491	0.0002	0.0057	0.0006	0.0492	0.0002	0.0057	0.0006
	200	0.0139	0.0002	0.0006	0.0005	0.0137	0.0002	0.0006	0.0005
5.0	20	-0.0006	0.0003	0.0001	0.0016	-0.0005	0.0008	0.0005	0.0022
	50	-0.0160	0.0003	0.0052	0.0010	-0.0160	0.0006	0.0072	0.0008
	80	-0.0079	0.0001	0.0005	0.0006	-0.0079	0.0002	0.0002	0.0005
	110	0.0494	0.0001	0.0046	0.0002	0.0494	0.0002	0.0066	0.0003
	140	0.0148	0.0001	0.0004	0.0002	0.0148	0.0002	0.0007	0.0003
	170	-0.0211	0.0001	0.0018	0.0001	-0.0211	0.0002	0.0011	0.0002
	200	-0.0260	0.0002	0.0047	0.0001	-0.0260	0.0004	0.0082	0.0002
10.0	20	-0.0002	0.0001	0.0001	0.0005	-0.0002	0.0005	0.0003	0.0015
	50	-0.0301	0.0001	0.0024	0.0002	-0.0301	0.0005	0.0035	0.0012
	80	-0.0042	0.0001	0.0000	0.0001	-0.0042	0.0004	0.0003	0.0011
	110	0.0270	0.0000	0.0022	0.0001	0.0270	0.0004	0.0032	0.0008
	140	0.0078	0.0002	0.0002	0.0001	0.0078	0.0003	0.0005	0.0008
	170	-0.0114	0.0000	0.0004	0.0001	-0.0114	0.0001	0.0001	0.0004
	200	-0.0309	0.0000	0.0028	0.0000	-0.0311	0.0001	0.0018	0.0001

perceber que a influência de α em θ é mínima.

Já para $\theta = 1.0$ e $\alpha > 1$, o REQM de $\hat{\alpha}$ possui oscilações em suas estimativas. Para $\theta = 10$ e $\alpha > 1$, podemos observar que a convergência do REQM para zero é mais demorada. No entanto, se $\theta = 10$ e $\alpha = 1$ o REQM de $\hat{\theta}$ não convergiu diretamente para zero, mesmo com tamanho amostral 200.

Em geral, o comportamento do vício de $\hat{\alpha}$ foi o mesmo em ambos os métodos, tendo como única diferença as amplitudes. Apesar do vício de $\hat{\theta}$ ser menor no mé-

todo da máxima verossimilhança, temos mais cenários convergindo para zero nos métodos dos momentos. O REQM de $\hat{\alpha}$ tem comportamento idêntico em ambos os métodos, sendo a convergência para zero mais rápida no método da máxima verossimilhança. Para o REQM de $\hat{\theta}$, novamente, temos um comportamento muito semelhante entre os dois métodos, sendo a convergência para zero pelo método da máxima verossimilhança pouca coisa mais rápida em quase todos os cenários.

Tabela 3 – Estimativas do vício e da raiz quadrada do erro-quadrático médio (REQM) dos estimadores de máxima verossimilhança (MV) e dos estimadores obtidos pelo método dos momentos (MM) para $\theta = 10.0$.

α	n	MV				MM			
		Viés		REQM		Viés		REQM	
		α	θ	α	θ	α	θ	α	θ
0.5	20	0.0477	-0.0243	0.0148	0.3315	0.0818	-0.0123	0.0210	0.3284
	50	0.0173	-0.0252	0.0044	0.3225	0.0361	-0.0101	0.0065	0.3232
	80	0.0108	-0.0252	0.0025	0.3206	0.0238	-0.0074	0.0038	0.3196
	110	0.0077	-0.0282	0.0018	0.3147	0.0179	-0.0059	0.0027	0.3141
	140	0.0058	-0.0224	0.0014	0.3087	0.0142	-0.0040	0.0021	0.3120
	170	0.0046	-0.0243	0.0011	0.3045	0.0116	-0.0054	0.0017	0.3039
	200	0.0038	-0.0244	0.0009	0.2992	0.0098	-0.0063	0.0015	0.3021
	20	0.0817	-0.0201	0.0503	0.3182	0.0825	-0.0178	0.0484	0.3168
1.0	50	0.0309	-0.0210	0.0157	0.2987	0.0319	-0.0202	0.0158	0.2992
	80	0.0188	-0.0157	0.0093	0.2781	0.0195	-0.0161	0.0095	0.2788
	110	0.0133	-0.0146	0.0066	0.2587	0.0139	-0.0134	0.0068	0.2602
	140	0.0103	-0.0139	0.0052	0.2417	0.0107	-0.0142	0.0054	0.2428
	170	0.0082	-0.0131	0.0042	0.2256	0.0086	-0.0128	0.0044	0.2263
	200	0.0069	-0.0104	0.0036	0.2106	0.0072	-0.0101	0.0038	0.2117
	20	0.0830	-0.0431	0.2107	0.2194	0.0458	-0.0433	0.2228	0.2220
	50	0.0726	-0.0214	0.1278	0.1261	0.0585	-0.0248	0.1410	0.1282
3.0	80	0.0543	-0.0122	0.0863	0.0831	0.0482	-0.0144	0.0986	0.0846
	110	0.0403	-0.0085	0.0624	0.0608	0.0372	-0.0100	0.0726	0.0619
	140	0.0321	-0.0070	0.0487	0.0479	0.0301	-0.0082	0.0570	0.0487
	170	0.0267	-0.0051	0.0397	0.0392	0.0252	-0.0060	0.0468	0.0399
	200	0.0223	-0.0042	0.0334	0.0333	0.0210	-0.0050	0.0394	0.0339
	20	0.0187	-0.0497	0.2859	0.1139	-0.0028	-0.0524	0.2991	0.1219
	50	0.0368	-0.0228	0.2284	0.0487	0.0147	-0.0242	0.2507	0.0517
	80	0.0464	-0.0133	0.1854	0.0307	0.0309	-0.0148	0.2131	0.0323
5.0	110	0.0444	-0.0092	0.1524	0.0225	0.0341	-0.0106	0.1812	0.0235
	140	0.0422	-0.0064	0.1275	0.0177	0.0365	-0.0079	0.1564	0.0184
	170	0.0380	-0.0047	0.1083	0.0146	0.0348	-0.0060	0.1360	0.0152
	200	0.0342	-0.0037	0.0933	0.0124	0.0331	-0.0046	0.1189	0.0129
	20	-0.0084	-0.0370	0.3083	0.0363	-0.0073	-0.0333	0.3097	0.0384
	50	0.0000	-0.0127	0.3091	0.0140	-0.0134	-0.0143	0.3108	
	80	-0.0029	-0.0082	0.2933	0.0088	0.0034	-0.0088	0.3118	
	110	0.0072	-0.0058	0.2804	0.0064	-0.0010	-0.0060	0.2966	
10.0	140	0.0145	-0.0048	0.2677	0.0050	0.0059	-0.0050	0.2876	0.0056
	170	0.0171	-0.0040	0.2549	0.0042	0.0080	-0.0040	0.2774	0.0047
	200	0.0200	-0.0036	0.2417	0.0036	0.0107	-0.0035	0.2699	0.0040

Tabela 3 – Estimativas do viés e da raiz quadrada do erro-quadrático médio (REQM) dos estimadores de máxima verossimilhança (MV) e dos estimadores obtidos pelo método dos momentos (MM) para $\theta = 10.0$.

α	MV					MM			
	n	Viés		REQM		Viés		REQM	
		α	θ	α	θ	α	θ	α	θ
0.5	20	0.0477	-0.0243	0.0148	0.3315	0.0818	-0.0123	0.0210	0.3284
	50	0.0173	-0.0252	0.0044	0.3225	0.0361	-0.0101	0.0065	0.3232
	80	0.0108	-0.0252	0.0025	0.3206	0.0238	-0.0074	0.0038	0.3196
	110	0.0077	-0.0282	0.0018	0.3147	0.0179	-0.0059	0.0027	0.3141
	140	0.0058	-0.0224	0.0014	0.3087	0.0142	-0.0040	0.0021	0.3120
	170	0.0046	-0.0243	0.0011	0.3045	0.0116	-0.0054	0.0017	0.3039
	200	0.0038	-0.0244	0.0009	0.2992	0.0098	-0.0063	0.0015	0.3021
1.0	20	0.0817	-0.0201	0.0503	0.3182	0.0825	-0.0178	0.0484	0.3168
	50	0.0309	-0.0210	0.0157	0.2987	0.0319	-0.0202	0.0158	0.2992
	80	0.0188	-0.0157	0.0093	0.2781	0.0195	-0.0161	0.0095	0.2788
	110	0.0133	-0.0146	0.0066	0.2587	0.0139	-0.0134	0.0068	0.2602
	140	0.0103	-0.0139	0.0052	0.2417	0.0107	-0.0142	0.0054	0.2428
	170	0.0082	-0.0131	0.0042	0.2256	0.0086	-0.0128	0.0044	0.2263
	200	0.0069	-0.0104	0.0036	0.2106	0.0072	-0.0101	0.0038	0.2117
3.0	20	0.0830	-0.0431	0.2107	0.2194	0.0458	-0.0433	0.2228	0.2220
	50	0.0726	-0.0214	0.1278	0.1261	0.0585	-0.0248	0.1410	0.1282
	80	0.0543	-0.0122	0.0863	0.0831	0.0482	-0.0144	0.0986	0.0846
	110	0.0403	-0.0085	0.0624	0.0608	0.0372	-0.0100	0.0726	0.0619
	140	0.0321	-0.0070	0.0487	0.0479	0.0301	-0.0082	0.0570	0.0487
	170	0.0267	-0.0051	0.0397	0.0392	0.0252	-0.0060	0.0468	0.0399
	200	0.0223	-0.0042	0.0334	0.0333	0.0210	-0.0050	0.0394	0.0339
5.0	20	0.0187	-0.0497	0.2859	0.1139	-0.0028	-0.0524	0.2991	0.1219
	50	0.0368	-0.0228	0.2284	0.0487	0.0147	-0.0242	0.2507	0.0517
	80	0.0464	-0.0133	0.1854	0.0307	0.0309	-0.0148	0.2131	0.0323
	110	0.0444	-0.0092	0.1524	0.0225	0.0341	-0.0106	0.1812	0.0235
	140	0.0422	-0.0064	0.1275	0.0177	0.0365	-0.0079	0.1564	0.0184
	170	0.0380	-0.0047	0.1083	0.0146	0.0348	-0.0060	0.1360	0.0152
	200	0.0342	-0.0037	0.0933	0.0124	0.0331	-0.0046	0.1189	0.0129
10.0	20	-0.0084	-0.0370	0.3083	0.0363	-0.0073	-0.0333	0.3097	0.0384
	50	0.0000	-0.0127	0.3091	0.0140	-0.0134	-0.0143	0.3108	0.0159
	80	-0.0029	-0.0082	0.2933	0.0088	0.0034	-0.0088	0.3118	0.0098
	110	0.0072	-0.0058	0.2804	0.0064	-0.0010	-0.0060	0.2966	0.0072
	140	0.0145	-0.0048	0.2677	0.0050	0.0059	-0.0050	0.2876	0.0056
	170	0.0171	-0.0040	0.2549	0.0042	0.0080	-0.0040	0.2774	0.0047
	200	0.0200	-0.0036	0.2417	0.0036	0.0107	-0.0035	0.2699	0.0040

6 Aplicações

Nesta seção, utilizando dados da literatura, verificamos a primazia da distribuição HNGD em relação às tradicionais distribuições Poisson, Binomial negativa (Bin. Neg.) e Weibull discreta (DWeibull) como distribuição candidata

para à análise de dados de contagem. São considerados três conjuntos de dados e os parâmetros das distribuições candidatas estimados pelo método de máxima verossimilhança. Para comparar os desempenhos das várias distribuições, além do teste χ^2 de aderência, também é utilizado o AIC (Akaike Information Criterion) e o BIC

(Bayesian Information Criterion), dados, respectivamente, por: $AIC = -2 \log(L) + 2m$ e $AIC = -2 \log(L) + m \log(n)$, em que m é o número de parâmetros, L é o valor da função log-verossimilhança maximizada e n o tamanho da amostra.

O primeiro conjunto de dados se refere ao número de gols por partida considerando os jogos “em casa” e os jogos “fora de casa” na temporada 2013/2014 da *Premier League*. As Tabelas 4 e 5 mostram a distribuição do número de gols observado em comparação com o número de gols esperados segundo as distribuições Half-Normal generalizada discreta, Weibull Discreta, Poisson e Binomial Negativa. A média e a variância do número de gol são, respectivamente, 1.56 gols e 1.77 gols² para os jogos “em casa” e 1.19 gols e 1.42 gols² para os jogos “fora de casa”. Dos resultados apresentados, de acordo com o teste qui-quadrado, os modelos HNGD, DWeibull e Binomial Negativa podem ser usados na análise desse conjunto de dados em relação aos gols “em casa” e apenas o modelo HNGD em relação aos gols “fora de casa”. Além disso, com base nos critérios AIC e BIC, nota-se que o modelo HNGD possui o menor valor em ambos os critérios e pode-se notar também que o modelo HNGD também possui os valores das frequências esperadas bem mais próximo aos valores das frequências observadas quando comparado aos outros modelos. Sendo assim, concluímos que a distribuição HNGD fornece um ajuste semelhante aos outros modelos considerados na análise do conjunto de dados e pode ser usada como alternativa na modelagem de dados discretos.

O segundo exemplo utiliza os dados retirado de Al-malki e Nadarajah (2014) que foram coletado pelo Ministério da Saúde Hospitalar na Arábia Saudita. Os mesmos indicam o tempo de vida em dias de 43 pacientes que foram diagnosticados com Leucemia (ver, Tabela 6). A média e variância são dadas, respectivamente, por 1191.63 dias e 256679 dias². Os resultados são apresentados na Tabela 7 em que, novamente, pode-se notar que a distribuição HNGD apresenta ajuste semelhante de acordo com os critérios adotados na análise.

Tabela 4 – Distribuição do número de gols “em casa” observados e esperados por partida na *Premier League*.

Número de Gols	Frequência Observada	Frequência Esperada			
		HNGD	DWeibull	Bin. Neg.	Poisson
0	95	95.1	90.2	90.0	80.4
1	113	110.6	122.4	121.7	125.3
2	85	89.1	88.8	90.3	97.6
3	49	53.8	48.6	48.7	50.7
4	28	23.8	21.8	21.3	19.9
5	10	7.6	8.2	8.0	6.1
Total	380	380	380	380	380
χ^2		1.4	3.3	3.6	9.4
valor-p		0.705	0.352	0.305	0.051
AIC		1221.6	1226.1	1226.8	1228.2
BIC		1229.5	1234.0	1234.7	1232.2

Tabela 5 – Distribuição do número de gols “fora de casa” observados e esperados por partida na *Premier League*.

Número de Gols	Frequência Observada	Frequência Esperada			
		HNGD	DWeibull	Bin. Neg.	Poisson
0	137	133.7	131.0	130.7	115.5
1	114	117.3	126.4	126.6	137.7
2	66	75.6	72.3	73.2	82.1
3	49	36.6	33.3	32.7	32.6
4	10	13.3	12.6	12.5	9.7
5	4	3.5	4.4	4.3	2.4
Total	380	380	380	380	380
χ^2		6.3	11.3	11.6	19.6
valor-p		0.052	0.003	0.003	<0.01
AIC		1106.8	1112.4	1112.8	1117.7
BIC		1114.7	1120.2	1120.6	1121.7

Tabela 6 – Dados de Leucemia.

115	181	255	418	441	461	516	739	743	789	807
865	924	983	1025	1062	1063	1165	1191	1222	1222	1251
1277	1290	1357	1369	1408	1455	1478	1519	1578	1578	1599
1603	1605	1696	1735	1799	1815	1852	1899	1925	1965	

Tabela 7 – Estimativas de máxima verossimilhança (erros-padrão), valor da estatística do teste qui-quadrado de aderência e os valores de AIC e BIC.

Distribuições	α	θ	χ^2	valor-p	AIC	BIC
HNGD	2.15 (0.29)	1431.70 (75.88)	3.25	0.661	657.99	661.51
DWeibull	2.56 (0.33)	1335.05 (82.53)	5.26	0.385	662.83	666.36
Bin. Neg.	3.61 (0.75)	1191.47 (95.82)	8.06	0.153	671.66	675.19
Poisson	1191.63 (5.26)		Inf.	<0.001	11069.55	11071.31

7 Conclusões

Neste artigo foi introduzida a distribuição HNGD formulada a partir do método baseado na função de sobrevivência proposto por Nakagawa e Osaki (1975). Algumas características e propriedades matemáticas da distribuição proposta foram estudadas. Além disso, a distribuição HNGD pode ser usada na análise de dados com subdispersão e superdispersão. Um estudo de simulação Monte Carlo evidenciou que os parâmetros são assintoticamente não viesados considerando o método da máxima verossimilhança e também o método dos momentos. Nas aplicações consideradas, a discriminação entre as distribuições Weibull discreta, Binomial Negativa

e Poisson foi realizada utilizando os critérios AIC e BIC. A qualidade do ajuste foi avaliada pela estatística dos teste qui-quadrado de aderência. Os resultados obtidos mostraram que a distribuição de Poisson não se adequou aos dados o que era esperado uma vez que os dados não possuem equidispersão. Além disso, verificou-se que a distribuição HNGD possui um ajuste semelhante aos das distribuições supracitadas o que evidencia sua utilização na análise de dados de contagem e de sobrevivência.

Agradecimentos

Os autores são gratos aos pareceristas por suas contribuições. Josmar Mazucheli agradece o apoio financeiro parcial da Fundação Araucária (Protocolo nº 49353, Convênio nº 064/2019 – UEM/Fundação Araucária).

Referências

- JAZI, M. A., LAI, C. D., ALAMATSAZ, M. H., (2010). A discrete inverse Weibull distribution and estimation of its parameters. **Statistical Methodology**, 7, 121-132.
- ALMALKI, S. J., NADARAJAH, S. (2014). A new discrete modified Weibull distribution. **IEEE Transactions on Reliability**, 63(1), 68-80.
- BAKOUCH, H. S., JAZI, M. A., NADARAJAH, S. (2014). A new discrete distribution. **Statistics**, 48(1), 200-240.
- BRACQUEMOND, C., GAUDOIN, O. (2003). A survey on discrete lifetime distributions. **International Journal of Reliability, Quality and Safety Engineering**, 10(1), 69-98.
- CHAKRABORTY, S. (2015). Generating discrete analogues of continuous probability distributions - **A survey of methods and constructions**. **Journal of Statistical Distributions and Applications**, 1(1), 1-30.
- CHAKRABORTY, S., CHAKRAVARTY, D. (2012). Discrete gamma distributions: Properties and parameter estimations. **Communications in Statistics-Theory and Methods**, 41(18), 3301-3324.
- CHAKRABORTY, S., CHAKRAVARTY, D. (2014). A discrete Gumbel distribution. **arXiv preprint arXiv:14107568**.
- COLLETT, D. (2003). **Modelling Survival Data in Medical Research**, 2^o edn. Chapman and Hall, New York.
- COORAY, K., ANANDA, M. M. A. (2008). A generalization of the Half-Normal distribution with applications to lifetime data. **Communications in Statistics - Theory and Methods**, 37(9), 1323-1337.
- COX, D. R., SNELL, E. J. (1968). A general definition of residuals. **Journal of the Royal Statistical Society Series B (Methodological)**, 30(2), 248-275.
- GHOSH, T., ROY, D., CHANDRA, N. K. (2013). Reliability approximation through the discretization of random variables using reversed hazard rate function. **International Journal of Mathematical, Computational, Statistical, Natural and Physical Engineering**, 7(4), 96 - 100.
- GÓMEZ-DÉNIZ, E., CALDERÍN-OJEDA, E. (2011). The discrete Lindley distribution: properties and applications. **Journal of Statistical Computation and Simulation**, 81(11), 1405-1416.
- GOOD, I. J. (1953). The population frequencies of species and the estimation of population parameters. **Biometrika**, 40(3-4), 237-264.
- HUSSAIN, T., AHMAD, M. (2014). Discrete inverse Rayleigh distribution. **Pakistan Journal of Statistics**, 30(2), 203-222.
- KEMP, A. W. (2004). Classes of discrete lifetime distributions. **Communications in Statistics - Theory and Methods**, 33(12), 3069-3093.
- KHAN, M., KHALIQUE, A., ABOUAMMOH, A. (1989). On estimating parameters in a discrete Weibull distribution. **IEEE Transactions on Reliability**, 38, 348-350.
- KLEIN, J. P., MOESCHBERGER, M. L. (1997). **Survival Analysis: Techniques for Censored and Truncated Data**. Springer-Verlag, New York.
- KRISHNA, H., PUNDIR, P. S. (2007). Discrete Maxwell distribution. **Interstat**.
- KRISHNA, H., PUNDIR, P. S. (2009). Discrete Burr and discrete Pareto distributions. **Statistical Methodology**, 6(2), 177-188.
- LAI, C. D. (2013). Issues concerning constructions of discrete lifetime models. **Quality Technology & Quantitative Management**, 10(2), 251-262.
- LAWLESS, J. F. (2003). **Statistical Models and Methods for Lifetime Data**, 2^o edn. Wiley Series in Probability and Statistics, John Wiley & Sons, Hoboken, NJ.
- LEE, E. T., WANG, J. W. (2003). **Statistical Methods for Survival Data Analysis**, 3^o edn. Wiley Series in Probability and Statistics, Wiley-Interscience [John Wiley & Sons], Hoboken, NJ.
- MAZUCHELI, J., DEY, S. (2018). Bias-corrected maximum likelihood estimation of the parameters of the generalized Half-Normal distribution. **Journal of Statistical Computation and Simulation**, 88(6), 1027-1038.

MAZUCHELI, J., OLIVEIRA, R. P., PERALTA, D., EMANUELLI, I. P. (2018). Application of discrete Burr XII distribution in the analysis of animal production data. **Ciência & Natura**, **40**(1), 1-10.

NAKAGAWA, T., OSAKI, S. (1975). The discrete Weibull distribution. **IEEE Transactions on Reliability**, **R-24**(5), 300-301.

OLIVEIRA, R. P., MAZUCHELI, J., ACHCAR, J. A. (2017). A comparative study between two discrete Lindley distributions. **Ciência e Natura**, **39**(3), 539-552.

PEARSON, K. (1895). Contributions to the mathematical theory of evolution. II. Skew variation in homogeneous material. **Philosophical Transactions of the Royal Society of London A**, **186**, 343-414.

ROY, D., DASGUPTA, T. (2001). A discretizing approach for evaluating reliability of complex systems under stress-strength model. **IEEE transactions on reliability**, **50**(2), 145-150.

STEIN, W. E., DATTERO, R. (1984). A new discrete Weibull distribution. **IEEE Transactions on Reliability**, **33**(2), 196-197.