

Aplicação da distribuição Burr XII discreta na análise de dados de produção animal

Application of discrete Burr XII distribution in the analysis of animal production data

Josmar Mazucheli^{*1}, Ricardo Puziol Oliveira², Danielle Peralta¹ e Isabele Picada Emanuelli^{3,4}

¹Departamento de Estatística, Universidade Estadual de Maringá, Maringá, P.R., Brasil

²Departamento de Medicina Social, Universidade de São Paulo, Ribeirão Preto, S.P., Brasil

³Centro Universitário de Maringá/UniCesumar, Maringá, P.R. Brasil

⁴Instituto Cesumar de Ciência, Tecnologia e Inovação/ICETI Maringá, P.R., Brasil

Resumo

Na produção animal os modelos estatísticos que mimetizam a realidade biológica são de grande importância para otimização e sustentabilidade de sistemas produtivos. A distribuição Burr XII contínua é bastante usada na análise de dados de sobrevivência, entretanto, o mesmo não ocorre com a sua versão discreta a qual foi recentemente proposta na literatura. A proposta deste trabalho é utilizar a distribuição Burr XII discreta, obtida pelo procedimento de discretização proposto por Nakagawa e Osaki (1975), na análise de dados relacionados a produção animal. Os dados analisados descrevem o tempo, em dias, contados do nascimento até a primeira postura de codornas amarelas (*Coturnix coturnix japonica*) submetidas a duas dietas alimentares. Para tanto foram utilizadas as versões discretizadas de cinco distribuições: Burr XII discreta, Weibull discreta, gama discreta, inversa-gaussiana discreta e log-normal discreta. Em todas elas, as estimativas dos parâmetros foram obtidas pelo método da máxima verossimilhança. Apesar da similaridade entre as estimativas é natural optar pela versão discreta dada a natureza dos dados uma vez que, pela adoção da versão discreta, pode-se calcular exatamente, por exemplo, a probabilidade do tempo até a primeira postura, o que não é possível caso seja adotado uma distribuição contínua. Entre as distribuições discretas, o teste qui-quadrado de aderência mostrou que a distribuição Burr XII foi a única indicada para descrever o comportamento dos dados utilizados.

Palavras-chave: Análise de sobrevivência, discretização, função de probabilidade, função de verossimilhança, produção animal, seleção de modelos.

Abstract

In animal production, the models that mimic the biological reality are of great importance for optimization and sustainability of the productive system. The continuous Burr XII distribution is widely used in survival data analysis, however, the same does not occur with its discrete version, recently proposed in the literature. The purpose of this work is to use the discrete Burr XII distribution, obtained by the discretization method proposed by Nakagawa and Osaki (1975), in the analysis of data related to animal production. The data analyzed describe the time, in days, from birth to first laying of yellow quail (*Coturnix coturnix japonica*) submitted to two diets. For this purpose the discretized versions of five distributions were used: the discrete Burr XII, the discrete Weibull, the discrete gamma, the discrete inverse-Gaussian and the discrete log-normal. For all distributions, the parameter estimates were obtained by the maximum likelihood method. Despite the similarity between the estimates it is natural to choose the discrete given the nature of the data and assuming the discrete distribution, it could be calculated exactly, for example, the probability of the time to the first posture, which is not possible if a continuous distribution is assumed. Thus, among the discrete distributions, the chi-square goodness-of-fit test showed that the Burr XII distribution was the only one indicated to describe the behavior of the data considered.

Keywords: Animal production, discretization, likelihood function, model selection, probability function, survival analysis.

*Corresponding Author: jmazucheli@gmail.com

1 Introdução

De maneira geral, em análise de sobrevivência a variável resposta indica o tempo transcorrido entre o início de um estudo e a ocorrência de algum evento. Este tempo é chamado de tempo de vida ou de sobrevivência e é medido em horas, dias, metros, ciclos entre várias outras unidades. Como exemplo, na área médica pode-se ter o tempo até a morte do paciente, da cura ou da recidiva de alguma doença (Klein e Moeschberger, 1997). Na engenharia pode-se ter o tempo até a falha ou até o mal funcionamento de um equipamento (Meeker e Escobar, 1998). Na produção animal, distribuições de probabilidade que mimetizem de maneira plausível o tempo até a ocorrência de algum evento são de extrema importância para otimização e sustentabilidade de sistemas de produção. Entretanto, este tipo de análise ainda é pouco utilizada. Pode-se citar o caso em que a resposta indica o tempo de lactação de vacas (Mrode, 2014) e o tempo até a primeira postura (Peralta et al., 2017).

Independente da área de aplicação, em geral, assume-se uma distribuição de probabilidade contínua para a análise do tempo de vida, sendo este tempo contínuo, discreto ou discretizado (Lai, 2014). É comum, por exemplo, utilizar a distribuição Weibull contínua na análise do número de ciclos que um equipamento suporta antes da sua quebra. Os tempos de vida desta natureza, ou seja, discretos ou discretizados, e analisados segundo alguma distribuição de probabilidade são comuns nos principais livros de Análise de Sobrevivência, por exemplo, Rupert (2011); Hamada et al. (2008); Collett (2003); Lawless (2003); Lee e Wang (2003); Kalbfleisch e Prentice (2002); Meeker e Escobar (1998); Klein e Moeschberger (1997). Vale ressaltar que o tempo de vida pode não ser exatamente observado, o que caracteriza a presença de um tempo censurado (à esquerda, intervalarmente, à direita, etc).

Nos últimos anos, para se evitar o uso de distribuições contínuas na análise de dados discretos ou discretizados, tem-se usado muito funções de probabilidade discretizadas. O processo de discretização pode gerar funções de probabilidade concorrentes das distribuições Poisson e binomial negativa, úteis na análise de dados de contagem na presença de equidispersão e sobredispersão, respectivamente. Existem na literatura diversos métodos de discretização, dentre eles destacam-se: o baseado na função de sobrevivência (Nakagawa e Osaki, 1975), o baseado em uma série infinita (Good, 1953; Kulasekera e Tonkyn, 1992; Kemp, 1997; Sato et al., 1999), o baseado na função de risco (Stein e Dattero, 1984), o das duas fases compostas (Chakraborty, 2015a), 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). Uma relação de alguns métodos de discretização e das distribuições discretizadas pode ser encontrada, por exemplo, em Chakraborty (2015a), Lai (2013) e Bracquemond e Gaudoin (2003).

Neste artigo, utilizando o método de discretização baseado na função de sobrevivência, apresenta-se a distribuição Burr XII discreta usada na descrição do número de dias transcorridos entre o nascimento e a primeira postura de codornas da linhagem amarela (*Coturnix coturnix japonica*) divididas em dois tratamentos (diets alimentares). Em tempo, é importante destacar que estes dados foram analisados por Peralta et al. (2017) considerando a versão contínua das distribuições Burr XII, Weibull, gama, inversa-gaussiana, log-logística e log-normal. No mesmo trabalho, além do tempo médio até a ocorrência do evento (postura) também foram estimados o tempo mais provável da ocorrência do evento. O que este artigo introduz são distribuições discretas alternativas àquelas usadas por Peralta et al. (2017). Apesar da similaridade entre as estimativas é natural optar pela versão discreta dada a natureza dos dados. Ainda, pela adoção da versão discreta pode-se calcular exatamente, por exemplo, a probabilidade do tempo até a primeira postura, o que não é possível caso seja adotado uma distribuição contínua.

Este trabalho está organizado da seguinte forma: a Seção 2 e 3 trazem uma breve revisão de literatura sobre o método de discretização utilizado e a versão contínua e discreta da distribuição Burr XII com três parâmetros. A descrição do dados, as distribuições alternativas e os procedimentos de discriminação, usados para determinar a distribuição mais apropriada, são apresentados na Seção 4. Os resultados numéricos são expostos e discutidos na Seção 5. Algumas considerações são expostas na Seção 6 que finaliza este artigo.

2 Discretização via função de sobrevivência

As distribuições discretas mais conhecidas como a geométrica, Poisson e binomial negativa são pouco utilizadas na análise de dados de sobrevivência. É notório que a distribuição geométrica é a versão discreta da distribuição exponencial; que na presença de observações censuradas as funções de verossimilhanças envolvem uma soma; que a distribuição Poisson somente é recomendada em situações de equidispersão e que as estimativas de máxima verossimilhança dos parâmetros da distribuição binomial negativa são únicos somente em situações de sobredispersão. Estas e outras peculiaridades, discutidas em Bakouch et al. (2014), intrínsecas às distribuições discretas por natureza, têm motivado a geração de distribuições discretas a partir de distribuições contínuas. Por exemplo, a versão discreta

da distribuição de uma variável aleatória contínua X , segundo a proposta de Nakagawa e Osaki (1975), utiliza a função de sobrevivência, $S_X(x | \Theta) = P(X \geq x | \Theta)$, e tem função de probabilidade escrita na forma:

$$P(T = k | \Theta) = S_X(k | \Theta) - S_X(k + 1 | \Theta) \quad (1)$$

em que $k = 0, 1, 2, \dots$, Θ é o vetor de parâmetros da distribuição de X e $T = \lfloor X \rfloor$ indica o maior inteiro menor ou igual a X .

Segundo este método, por exemplo, considerando X exponencialmente distribuída com $S_X(x | \mu) = e^{-\frac{x}{\mu}}$ tem-se como função de probabilidade a distribuição geométrica dada por:

$$P(T = k | \phi) = \phi^k - \phi^{k+1} = \phi^k (1 - \phi)$$

em que $\mu > 0$ e $\phi = e^{-\frac{1}{\mu}}$. Já a distribuição Weibull com $S_X(x | \mu, \beta) = e^{-\left(\frac{x}{\mu}\right)^\beta}$ tem na sua versão discreta a função de probabilidade:

$$P(T = k | \phi, \beta) = \phi^{k^\beta} - \phi^{(k+1)^\beta} \quad (2)$$

em que $\mu > 0$, $\beta > 0$ e $\phi = e^{-\frac{1}{\mu}}$. Como casos particulares, da mesma forma que no caso contínuo, tem-se a distribuição exponencial discreta quando $\beta = 1$ e a distribuição Rayleigh discreta quando $\beta = 2$ (Roy, 2004).

É importante destacar que o método de discretização baseado na função de sobrevivência é seguramente o mais simples e o mais discutido na literatura. Uma propriedade interessante é que o mesmo preserva a expressão da função de sobrevivência, ou seja $S_{\text{discreta}}(k) = S_{\text{contínua}}(k + 1)$, $k = 0, 1, \dots$. Esta propriedade faz, na presença de observações censuradas, que função de verossimilhança não dependa de uma soma.

Recentemente, Chakraborty (2015b), utilizando (1), apresenta a versão discreta da distribuição gama generalizada (Stacy, 1962) a qual tem como casos particulares 10 outras distribuições, entre elas a gama, Weibull, exponencial, Rayleigh, log-normal e seminormal discretas.

3 As Distribuições Burr XII Contínua e Discreta

A distribuição Burr XII ou Burr Tipo XII (Burr, 1942) é membro de um sistema com 12 funções de distribuições de probabilidade. Os tipos III e XII são os mais usadas na análise de dados de sobrevivência. Uma variável aleatória não negativa X tem distribuição Burr XII se sua função de sobrevivência é escrita na forma:

$$S_X(x | \theta, \alpha, \gamma) = \left[1 + \left(\frac{x}{\theta}\right)^\alpha\right]^{-\gamma} \quad (3)$$

em que $\theta > 0$ é o parâmetro de escala e $\alpha > 0$ é o parâmetro de forma. O parâmetro $\gamma > 0$ não influencia o comportamento das funções de densidade e de risco, apesar de em muitos trabalhos ser definido como parâmetro de forma. Alguns casos particulares da Burr XII incluem: a distribuição Lomax (Pareto do segundo tipo) quando $\alpha = 1$, a distribuição log-logística quando $\gamma = 1$ e a distribuição Weibull quando $\gamma \rightarrow \infty$. Os trabalhos de Rodriguez (1977), Tadikamalla (1980) e Zimmer et al. (1998) exploram em detalhes a conexão entre a Burr XII e outras distribuições.

A partir de Para e Jan (2016), substituindo a expressão (3) em (1), tem-se a função de probabilidade Burr XII discreta e escrita na forma:

$$P(T = k | \theta, \alpha, \phi) = \phi^{\log\left[1 + \left(\frac{k}{\theta}\right)^\alpha\right]} - \phi^{\log\left[1 + \left(\frac{k+1}{\theta}\right)^\alpha\right]} \quad (4)$$

em que $\phi = e^{-\gamma}$. Uma versão simplificada de (4) com $\theta = 1$ é apresentada em Krishna e Pundir (2009), Akdoğan et al. (2014) e Kamari et al. (2015). Ainda em (4), para $\alpha = 1$ tem-se a distribuição Lomax discreta (Para e Jan, 2016) enquanto que para $\gamma = 1$, ou seja $\phi = e^{-1}$, tem-se a distribuição log-logística discreta (Khorashadizadeh et al., 2013).

Seja T uma variável aleatória com distribuição Burr XII discreta. O momento de ordem r da variável aleatória T é dado pela seguinte expressão:

$$E(T^r) = \sum_{k=0}^{\infty} k^r \left\{ \phi^{\log\left[1 + \left(\frac{k}{\theta}\right)^\alpha\right]} - \phi^{\log\left[1 + \left(\frac{k+1}{\theta}\right)^\alpha\right]} \right\} \quad (5)$$

e, em particular, para $r = 1$ tem-se $E(T) = \sum_{k=1}^{\infty} \phi^{\log\left[1 + \left(\frac{k}{\theta}\right)^\alpha\right]}$ e para $r = 2$, $E(T^2) = \sum_{k=1}^{\infty} (2k - 1) \phi^{\log\left[1 + \left(\frac{k}{\theta}\right)^\alpha\right]}$.

Table 1: Distribuição do número de dias até a primeira postura (nº de dias, frequência).

| Dieta I | |
|---|--|
| (31, 4), (32, 2), (33, 5), (34, 9), (35, 4), (36, 13), (37, 14), (38, 16), (39, 15), (40, 13), (41, 12), (42, 9), (43, 8), (44, 5), (45, 4), (46, 8), (47, 5), (48, 2), (49, 4), (50, 1), (53, 1), (54, 2), (56, 1), (59, 2), (63, 1), (75, 1), (76, 1) | |
| Dieta II | |
| (31, 2), (32, 2), (33, 4), (34, 7), (35, 7), (36, 16), (37, 5), (38, 13), (39, 14), (40, 18), (41, 10), (42, 8), (43, 6), (44, 5), (45, 3), (46, 3), (47, 4), (48, 1), (49, 1), (51, 2), (52, 1), (53, 1), (54, 3), (55, 1), (56, 1), (58, 5), (73, 1), (80, 1) | |

Embora não seja possível obter uma expressão para $E(T^r)$ que não envolva uma soma infinita, a mesma é facilmente calculada atribuindo-se valores para θ , α e ϕ , somando-se um número grande de termos.

Considere T_1, \dots, T_n uma amostra aleatória de uma variável aleatória T com distribuição Burr XII discreta e função de probabilidade definida em (4). Observado $\mathbf{k} = (k_1, \dots, k_n)$, os valores de $\hat{\theta}$, $\hat{\alpha}$ e $\hat{\phi}$, respectivamente, de θ , α e ϕ maximizam a função de verossimilhança ou a função log-verossimilhança escritas, respectivamente, como:

$$L(\theta, \alpha, \phi | \mathbf{k}) = \prod_{i=1}^n \left\{ \phi \log \left[1 + \left(\frac{k_i}{\theta} \right)^\alpha \right] - \phi \log \left[1 + \left(\frac{k_i+1}{\theta} \right)^\alpha \right] \right\} \quad (6)$$

e

$$\ell(\theta, \alpha, \phi | \mathbf{k}) = \sum_{i=1}^n \log \left\{ \phi \log \left[1 + \left(\frac{k_i}{\theta} \right)^\alpha \right] - \phi \log \left[1 + \left(\frac{k_i+1}{\theta} \right)^\alpha \right] \right\}. \quad (7)$$

Como maximizar (7) é mais simples e numericamente mais estável que maximizar (6), as estimativas de θ , α e ϕ são as raízes do sistema formado pelas derivadas parciais de $\ell(\theta, \alpha, \phi | \mathbf{k})$ com respeito a θ , α e ϕ , ou seja, deve-se resolver, neste caso numericamente, um sistema homogêneo formado pelas equações $\frac{\partial}{\partial \theta} \ell(\theta, \alpha, \phi | \mathbf{k}) = 0$, $\frac{\partial}{\partial \alpha} \ell(\theta, \alpha, \phi | \mathbf{k}) = 0$ e $\frac{\partial}{\partial \phi} \ell(\theta, \alpha, \phi | \mathbf{k}) = 0$.

Uma vez encontradas as estimativas de $\hat{\theta}$, $\hat{\alpha}$ e $\hat{\phi}$, os intervalos de confiança para θ , α e ϕ podem ser construídos a partir da normalidade assintótica dos estimadores de máxima verossimilhança. Pela propriedade de invariância destes estimadores facilmente estima-se $E(T^r)$, $r = 1, 2, \dots$, definida em (5). Os intervalos de confiança para $E(T^r)$ podem ser construídos via *Bootstrap* (Efron, 1979; Davison e Hinkley, 1997), ou também, pela normalidade assintótica com $\text{Var}[E(T^r)]$ estimada pelo método delta (Davison, 2003).

4 Materiais e Métodos

Os dados analisados neste artigo descrevem o tempo, em dias, contados do nascimento até a primeira postura de codornas amarelas (*Coturnix coturnix japonica*) submetidas a duas dietas alimentares. Aos 28 dias de idade as aves passaram por um processo de sexagem e as fêmeas foram divididas em dois grupos. Um grupo passou a receber uma dieta de 2.900 kcal/kg (dieta I, com 163 aves) e o outro de 2.500 kcal/kg (dieta II, com 145 aves). Estes dados foram obtidos em um experimento conduzido entre agosto de 2006 e janeiro de 2007 no Setor de Coturnicultura da Fazenda Experimental da Universidade Estadual de Maringá. A distribuição de frequência do número de dias até a primeira postura em cada dieta é apresentada na Tabela 1.

Outros detalhes a respeito destes dados e suas análises via distribuições contínuas estão disponíveis em Peralta et al. (2017).

Além das distribuições Burr XII discreta e Weibull discreta (definidas pelas equações 4 e 2, respectivamente) considera-se na aplicação as versões discretas das seguintes distribuições:

- Gama discreta (Chakraborty e Chakravarty, 2012)

$$P(T = k | \theta, \alpha) = \frac{1}{\Gamma(\alpha)} \Gamma\left(\alpha, \frac{k}{\theta}\right) - \frac{1}{\Gamma(\alpha)} \Gamma\left(\alpha, \frac{k+1}{\theta}\right)$$

em que $\theta > 0$, $\alpha > 0$, $\Gamma\left(\alpha, \frac{k}{\theta}\right) = \int_{\frac{k}{\theta}}^{\infty} u^{\alpha-1} e^{-u} du$ é a função gama incompleta e $\Gamma(\alpha) = \int_0^{\infty} u^{\alpha-1} e^{-u} du$ é a função gama completa.

Table 2: Estimativas de máxima verossimilhança (erros-padrão), valor da estatística do teste qui-quadrado de aderência, valores de AIC e BIC — dieta I.

| Distribuição | θ | α | γ | μ | χ^2 | valor-p | AIC | BIC |
|--------------|-------------------|-----------------------|------------------|-------------------|----------|---------|---------|---------|
| 1 | 37.125 (0.703) | 20.464 (2.956) | 0.407 (0.094) | | 1.53 | 0.464 | 1014.36 | 1023.62 |
| 2 | 43.855 (4.863) | 1.062 (0.118) | | | 44.57 | < 0.01 | 1054.44 | 1060.61 |
| 3 | 44.248 (0.734) | 5.046 (0.246) | | | 74.21 | < 0.01 | 1137.90 | 1144.08 |
| 4 | | 1895.890 (197.831) | | 41.277 (0.479) | 30.62 | < 0.01 | 1043.81 | 1049.99 |
| 5 | 0.146 (0.008) | | | 3.708 (0.011) | 30.31 | < 0.01 | 1042.98 | 1049.15 |

1: Burr XII, 2: gama, 3: Weibull, 4: inversa-gaussiana e 5: log-normal.

- Inversa-Gaussiana discreta

$$P(T = k | \mu, \alpha) = \Phi\left(\sqrt{\frac{\alpha}{k+1}}\left(\frac{k+1}{\mu} - 1\right)\right) + e^{\frac{2\alpha}{\mu}}\Phi\left(-\sqrt{\frac{\alpha}{k+1}}\left(\frac{k+1}{\mu} + 1\right)\right) - \Phi\left(\sqrt{\frac{\alpha}{k}}\left(\frac{k}{\mu} - 1\right)\right) + e^{\frac{2\alpha}{\mu}}\Phi\left(-\sqrt{\frac{\alpha}{k}}\left(\frac{k}{\mu} + 1\right)\right)$$

em que $\Phi(\cdot)$ é a função de distribuição acumulada de uma variável aleatória com distribuição normal padrão.

- Log-Normal discreta (Chakraborty, 2015b)

$$P(T = k | \mu, \theta) = \Phi\left(\frac{\log(k+1) - \mu}{\theta}\right) - \Phi\left(\frac{\log(k) - \mu}{\theta}\right).$$

É importante mencionar que os parâmetros $\theta > 0$, $\alpha > 0$ e $\mu > 0$ descrevem, respectivamente, a escala, a forma e a localização. Somente na distribuição log-normal o parâmetro μ tem suporte nos números reais.

Na aplicação, apresentada na Seção 5, a qualidade do ajuste da distribuição Burr XII discreta, Weibull discreta, gama discreta, inversa-gaussiana discreta e log-normal discreta foi avaliada pelo valor-p associado ao valor da estatística do teste qui-quadrado de aderência (Voinov et al., 2013). São reportados também os valores do Critério de Informação de Akaike, $AIC = -2\ell(\hat{\Theta} | \mathbf{k}) + 2p$ e do Critério de Informação Bayesiano, $BIC = -2\ell(\hat{\Theta} | \mathbf{k}) + p \log(n)$, em que $\ell(\hat{\Theta} | \mathbf{k})$ é a função log-verossimilhança maximizada, n o tamanho da amostra e p é o número de parâmetros (Held e Sabanés Bové, 2014).

5 Resultados e Discussões

Nas Tabelas 2 e 3 são apresentados, para as dietas I e II respectivamente, alguns valores associados ao ajuste das distribuições Burr XII, Weibull, gama, inversa-gaussiana e log-normal, discretizadas pelo método de Nakagawa e Osaki (1975). Todos os valores foram obtidos no software *R*, (R Core Team, 2016), utilizando a função *fitdist* da biblioteca *fitdistrplus* (Delignette-Muller e Dutang, 2015). O script utilizado para a obtenção dos resultados associados a distribuição Burr XII é apresentado no Apêndice.

É importante observar que pelo teste qui-quadrado de aderência, considerando níveis de significância usuais — 5% e 1% — somente a distribuição Burr XII discreta é apropriada para descrever o tempo até a primeira postura das codornas. Os valores observados e esperados, usados no cálculo do valor da estatística do teste qui-quadrado, são apresentados na Tabela 4. Nas Tabelas 2 e 3 omitimos os resultados oriundos dos ajustes das distribuições Poisson e binomial negativa. Isto se deve ao fato do teste qui-quadrado também ter rejeitado a aderência de ambas (valor-p < 0.0001). O mesmo fato é verdade para o ajuste das distribuições Lomax discreta e log-logística discreta.

A Tabela 5 apresenta as estimativas pontuais e os intervalos de confiança *Bootstrap*, baseados nos percentis da distribuição *Bootstrap*, do tempo médio até a primeira postura (Davison e Hinkley, 1997; Efron e Tibshirani, 1993). Foram consideradas 10.000 amostras *Bootstrap* não paramétrico e paramétrico. O número de termos usados em (5) fixado em 100. Apesar da estimação de (5) e dos intervalos de confiança ser computacionalmente intensivo, estes são facilmente obtidos pela função *bootdist* da biblioteca *fitdistrplus* do software *R* (ver, Apêndice).

Table 3: Estimativas de máxima verossimilhança (erros-padrão), valor da estatística do teste qui-quadrado de aderência, valores de AIC e BIC — dieta II.

| Distribuição | θ | α | γ | μ | χ^2 | valor-p | AIC | BIC |
|--------------|-------------------|-----------------------|------------------|-------------------|----------|---------|---------|---------|
| 1 | 36.545 (0.657) | 23.388 (3.841) | 0.314 (0.076) | | 2.45 | 0.292 | 917.74 | 926.68 |
| 2 | 38.825 (4.547) | 0.932 (0.109) | | | 32.17 | < 0.01 | 964.29 | 970.24 |
| 3 | 44.815 (0.822) | 4.821 (0.252) | | | 65.25 | < 0.01 | 1036.13 | 1042.09 |
| 4 | | 1701.508 (139.635) | | 41.650 (0.541) | 24.36 | < 0.01 | 953.65 | 959.61 |
| 5 | 0.155 (0.009) | | | 3.716 (0.011) | 24.30 | < 0.01 | 952.96 | 958.92 |

1: Burr XII, 2: gama, 3: Weibull, 4: inversa-gaussiana e 5: log-normal.

Table 4: Frequências observadas (Freq. Obs) e frequências esperadas (Freq. Esp.) segundo o ajuste da distribuição Burr XII.

| Intervalos | Dieta I | | Dieta II | |
|----------------|------------|------------|------------|------------|
| | Freq. Obs. | Freq. Esp. | Freq. Obs. | Freq. Esp. |
| < 31 dias | 4 | 3.05 | 2 | 1.99 |
| 31 ≤ dias < 39 | 78 | 78.61 | 68 | 71.07 |
| 39 ≤ dias < 47 | 64 | 61.32 | 57 | 52.47 |
| 47 ≤ dias < 55 | 10 | 13.74 | 10 | 13.21 |
| 55 ≤ dias < 63 | 4 | 3.54 | 6 | 3.92 |
| ≥ 63 dias | 2 | 1.73 | 2 | 2.34 |

A fim de se ter uma outra medida de comparação entre as duas dietas estimamos o parâmetro $R = P(X > Y)$. Neste caso, sejam X e Y variáveis aleatórias independentes indicando o tempo até a primeira postura sob as dietas I e II, respectivamente, com distribuições Burr XII($\theta_X, \alpha_X, \gamma_X$) e Burr XII($\theta_Y, \alpha_Y, \gamma_Y$). Define-se o parâmetro estresse-resistência (Kotz et al., 2003), como:

$$R = P(X > Y) = \sum_{k=0}^{\infty} P_X(X = k | \theta_X, \alpha_X, \gamma_X) [1 - S_Y(k + 1 | \theta_Y, \alpha_Y, \gamma_Y)]$$

em que $P_X(\cdot)$ e $S_Y(\cdot)$ indicam a função de probabilidade e de sobrevivência de X e Y , respectivamente.

Na Tabela 6 são apresentadas as estimativas pontuais e intervalares de R . As estimativas intervalares foram obtidas utilizando o mesmo procedimento usado na estimação do tempo médio até a primeira postura.

A partir das estimativas intervalares do tempo médio (Tabela 5) e de R (Tabela 6) é fácil verificar que não existem evidências suficientes nos dados para garantir o efeito da dieta. Estas conclusões se pautam no fato do intervalo de confiança para a diferença entre as médias conter o valor zero e também pelo intervalo de confiança para R conter o valor 0.5 ou 50%.

O fato da expressão dos momentos da distribuição Burr XII discreta depender de uma série infinita (ver equação 5) pode trazer dúvidas com respeito ao valor de k a ser usado na estimação de $E(T)$ ou $Var(T)$. A convergência da série para a estimação de $E(T)$ e $Var(T)$ sob as dietas I e II, considerando as estimativas do parâmetros da distribuição Burr XII e vários valores para k é ilustrada na Figura 1. Em nosso processo de estimação consideramos $k = 100$.

Table 5: Estimativas pontuais e intervalares para o tempo médio até a primeira postura.

| Dieta | Bootstrap Não Paramétrico | | | Bootstrap Paramétrico | | | |
|-------|---------------------------|----------|-------|-----------------------|----------|-------|-------|
| | Média | I.C. 95% | | Média | I.C. 95% | | |
| | | L.I. | L.S. | | L.I. | L.S. | |
| I | 40.78 | 39.76 | 41.85 | I | 40.77 | 39.78 | 41.81 |
| II | 41.15 | 39.98 | 42.40 | II | 41.16 | 40.01 | 42.39 |

I.C.: intervalo de confiança, L.I.: limite inferior e L.S.: limite superior.

Table 6: Estimativas pontuais e intervalares para $R = P(X > Y)$.

| <i>Bootstrap</i> Não Paramétrico | | | <i>Bootstrap</i> Paramétrico | | |
|----------------------------------|----------|--------|------------------------------|----------|--------|
| R | I.C. 95% | | R | I.C. 95% | |
| | L.I. | L.S. | | L.I. | L.S. |
| 0.5228 | 0.4576 | 0.5883 | 0.5244 | 0.4603 | 0.5888 |

I.C.: intervalo de confiança, L.I.: limite inferior e L.S.: limite superior.

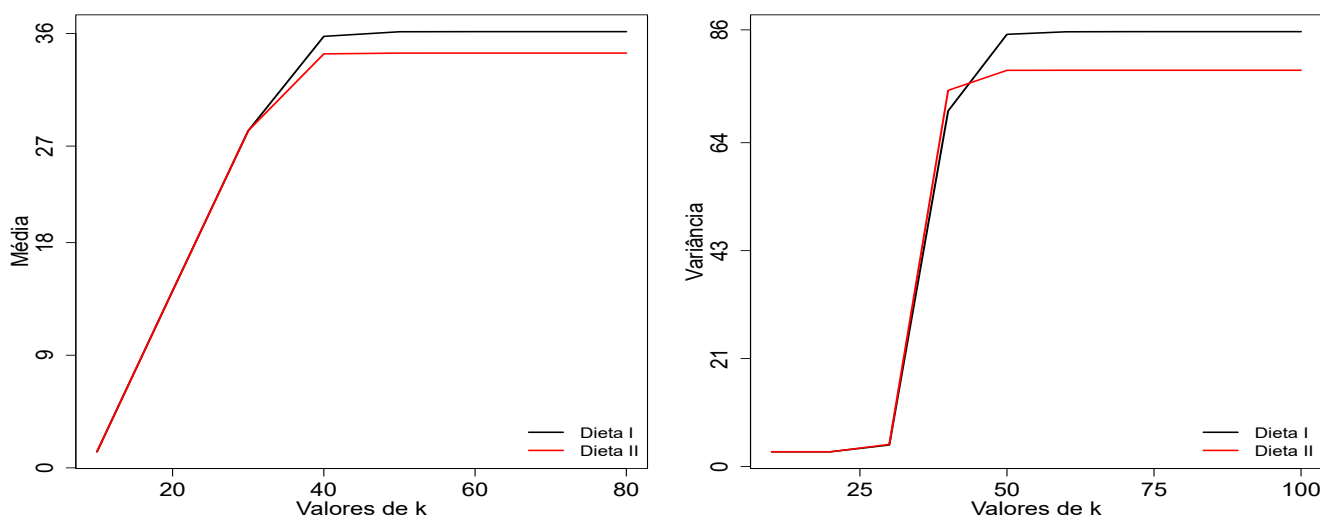


Figure 1: Média e variância segundo a distribuição Burr XII discreta para ambas as dietas.

6 Considerações Finais

Neste artigo comparamos a distribuição Burr XII discretizada com as distribuições Weibull, gama, inversa-gaussiana e log-normal também nas versões discretizadas. Nesta investigação, o teste qui-quadrado de aderência mostrou que a distribuição Burr XII foi a única indicada para descrever o comportamento dos dados relacionados a produção animal. As versões discretas foram obtidas pelo método baseado na função de sobrevivência o qual preserva a expressão da função de sobrevivência contínua em sua parte inteira. Verificamos na literatura que a maioria das distribuições consideradas neste estudo foram recentemente propostas e com pouca ou nenhuma aplicação em dados de produção animal. Com o intuito de não usar distribuições contínuas na análise de dados estritamente discretos, reanalisamos parte dos dados de produção animal utilizados por Peralta et al. (2017). Comparando as estimativas dos parâmetros obtidas via distribuições contínuas com as obtidas aqui, observamos estimativas bastante próximas, entretanto, das distribuições discretas, somente a distribuição Burr XII mostrou aderência segundo o critério usado para avaliar a qualidade do ajuste. Em adição, as estimativas intervalares do tempo médio até a primeira postura e do parâmetro R apontaram que não existem evidências suficientes nos dados para garantir o efeito da dieta alimentar sobre a primeira postura. Apesar da similaridade entre as estimativas é natural optar pela versão discreta dada a natureza dos dados. Ainda, pela adoção da versão discreta pode-se calcular exatamente, por exemplo, a probabilidade do tempo até a primeira postura, o que não é possível caso seja adotado uma distribuição contínua.

Agradecimentos

Os autores agradecem os revisores pelas valiosas sugestões. A terceira autora agradece ao Instituto Cesumar de Ciência, Tecnologia e Inovação (ICETI) pela Bolsa de Produtividade em Pesquisa concedida.

Apêndice

Neste apêndice apresentamos o código R (R Core Team, 2016) usados para obter os valores associados ao ajuste da distribuição Burr XII discreta. A título de ilustração consideramos apenas os dados referentes a dieta I (ver, Tabela 1).

Para a obtenção dos valores associados às demais distribuições basta definir as respectivas funções de sobrevivência contínua, funções de probabilidade e as funções de distribuição acumulada.

```
library(fitdistrplus)

# Funcao de sobrevivencia - Burr XII continua
sburr <- function(x, alpha, theta, gamma)
{
  (1 + (x / theta) ^ alpha) ^ (-gamma)
}

# Funcao de probabilidade - Burr XII
ddburr <- function(x, alpha, theta, gamma)
{
  sburr(x, alpha, theta, gamma) - sburr(x + 1, alpha, theta, gamma)
}

# Funcao de distribuicao acumulada - Burr XII
pdburr <- function(q, alpha, theta, gamma)
{
  1 - sburr(q + 1, alpha, theta, gamma)
}

# Dados (ver Tabela 1)
tempos <- c(31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42,
43, 44, 45, 46, 47, 48, 49, 50, 53, 54, 56, 59, 63, 75, 76)

freqs <- c(4, 2, 5, 9, 4, 13, 14, 16, 15, 13, 12, 9,
8, 5, 4, 8, 5, 2, 4, 1, 1, 2, 1, 2, 1, 1, 1)

x <- rep(tempos, freqs)

# Ajuste
fit <- fitdistr(x, 'dburr', start = list(gamma = 0.04, alpha = 30, theta = 37), discrete = T)

# Teste qui-quadrado de aderencia
gofstat(fit, discrete = T, meancount = 5, chisqbreaks = seq(31, 65, 8))

# Bootstrap nao parametrico
np <- bootdist(fit, niter = 10000, bootmethod = 'nonparam')
bp <- bootdist(fit, niter = 10000, bootmethod = 'param')
```

Referências

- Akdoğan, Y., Kuş, Klnacl, I. (2014). Point estimation of parameters in discrete Burr distribution based on type I censored sample. *Journal of the Turkish Statistical Association*, 10(69), 80–86.
- 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.
- Burr, I. W. (1942). Cumulative frequency functions. *Annals of Mathematical Statistics*, 13(2), 215–232.
- Chakraborty, S. (2015a). Generating discrete analogues of continuous probability distributions - A survey of methods and constructions. *Journal of Statistical Distributions and Applications*, 2(1), 1–30.
- Chakraborty, S. (2015b). A new discrete distribution related to generalized Gamma distribution and its properties. *Communications in Statistics - Theory and Methods*, 44(8), 1691–1705.
- Chakraborty, S., Chakravarty, D. (2012). Discrete Gamma distributions: properties and parameter estimations. *Communications in Statistics - Theory and Methods*, 41(18), 3301–3324.
- Collett, D. (2003). *Modelling Survival Data in Medical Research*, 2^o edn. Chapman and Hall, New York.
- Davison, A. C. (2003). *Statistical Models*. Cambridge: Cambridge University Press, New York.
- Davison, A. C., Hinkley, D. V. (1997). *Bootstrap Methods and Their Applications*. Cambridge: Cambridge University Press, New York.
- Delignette-Muller, M. L., Dutang, C. (2015). fitdistrplus: An R package for fitting distributions. *Journal of Statistical Software*, 64(4), 1–34.
- Efron, B. (1979). Bootstrap methods: Another look at the Jackknife. *Annals of Statistics*, 7(1), 1–26.
- Efron, B., Tibshirani, R. J. (1993). *An introduction to the Bootstrap*, Monographs on Statistics and Applied Probability, vol 57. Chapman and Hall, New York.

- 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.
- Good, I. J. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika*, 40(3-4), 237–264.
- Hamada, M. S., Wilson, A. G., Reese, C. S., Martz, H. F. (2008). *Bayesian Reliability*. Springer Series in Statistics, Springer, New York.
- Held, L., Sabanés Bové, D. (2014). *Applied Statistical Inference - Likelihood and Bayes*. Springer, Heidelberg.
- Kalbfleisch, J. D., Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*, 2^o edn. Wiley, New York, NY.
- Kamari, H., Bevrani, H., Kamary, K. (2015). Bayesian estimate of discrete Burr distribution with two parameters. *Research & Reviews: Journal of Statistics and Mathematical Sciences*, 1(2), 62–68.
- Kemp, A. W. (1997). Characterizations of a discrete Normal distribution. *Journal of Statistical Planning and Inference*, 63(2), 223 – 229.
- Khorashadzadeh, M., Roknabadi, R. A. H., G. R. Borzadaran, M. G. R. (2013). Characterization of life distributions using Log-odds rate in discrete aging. *Communications in Statistics - Theory and Methods*, 42(1), 76–87.
- Klein, J. P., Moeschberger, M. L. (1997). *Survival Analysis: Techniques for Censored and Truncated Data*. Springer-Verlag, New York.
- Kotz, S., Y. L., Pensky, M. (2003). *The Stress-Strength Model and its Generalizations: Theory and Applications*. Academic Press.
- Krishna, H., Pundir, P. S. (2009). Discrete Burr and discrete Pareto distributions. *Statistical Methodology*, 6(2), 177 – 188.
- Kulasekera, K. B., Tonkyn, D. W. (1992). A new discrete distribution, with applications to survival, dispersal and dispersion. *Communications in Statistics - Simulation and Computation*, 21(2), 499–518.
- Lai, C. (2013). Issues concerning constructions of discrete lifetime models. *Quality Technology & Quantitative Management*, 10(2), 251–262.
- Lai, C. (2014). *Discrete Weibull Distributions and Their Generalizations*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 97–113.
- 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.
- Meeker, W. Q., Escobar, L. A. (1998). *Statistical Methods for Reliability Data*. John Wiley & Sons, New York.
- Mrode, R. A. a. (2014). *Linear Models for the Prediction of Animal Breeding Values*, 3^o edn. CABI, Nosworthy Way, UK.
- Nakagawa, T., Osaki, S. (1975). The discrete Weibull distribution. *IEEE Transactions on Reliability*, 24(5), 300–301.
- Para, B. A., Jan, T. R. (2016). On discrete three parameter Burr type XII and discrete Lomax distributions and their applications to model count data from medical science. *Biometrics & Biostatistics International Journal*, 4(2), 1–15.
- Pearson, K. (1895). Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London A*, pp. 343–414.
- Peralta, D., Mazucheli, J., Emanuelli, I., Rossi, R. (2017). Aplicação da distribuição Burr XII na análise do tempo até a primeira postura de codornas. *Revista Brasileira de Biometria*, 35(1), 1–16.
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, URL <https://www.R-project.org/>.
- Rodriguez, R. N. (1977). A guide to the Burr type XII distributions. *Biometrika*, 64(1), 129–134.

- Roy, D. (2004). Discrete Rayleigh distribution. *IEEE Transactions on Reliability*, 53(2), 255–260.
- 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.
- Rupert, M. (2011). *Survival Analysis*, vol 66. John Wiley & Sons.
- Sato, H., Ikota, M., Sugimoto, A., Masuda, H. (1999). A new defect distribution metrology with a consistent discrete exponential formula and its applications. *Semiconductor Manufacturing, IEEE Transactions on*, 12(4), 409–418.
- Stacy, E. W. (1962). A generalization of the Gamma distribution. *The Annals of Mathematical Statistics*, 33(3), 1187–1192.
- Stein, W. E., Dattero, R. (1984). A new discrete Weibull distribution. *IEEE Transactions on Reliability*, 33(2), 196–197.
- Tadikamalla, P. R. (1980). A look at the Burr and related distributions. *International Statistical Review Revue Internationale de Statistique*, 48(3), 337–344.
- Voinov, V., Nikulin, M., Balakrishnan, N. (2013). *Chi-Squared Goodness of Fit Tests with Applications*. Academic Press.
- Zimmer, W. J., Keats, J. B., Wang, F. K. (1998). The Burr XII distribution in reliability. *Journal of Quality Technology*, 30(4), 389–394.

Danielle Peralta

Departamento de Estatística . Universidade Estadual de Maringá, PR, Brasil
Email: danielleperalta@outlook.com

Contribuição do autor

Josmar Mazucheli,

Departamento de Estatística . Universidade Estadual de Maringá, PR, Brasil

Email:jmazucheli@gmail.com

Contribuição do autor

Isabele P. Emanuelli

Program de Pós-Graduação em Tecnologias Limpas, Unicesumar, Brasil

Email: isabelevet@hotmail.com

Contribuição do autor

Ricardo Puziol Oliveira

Departamento de Medicina Social, Universidade de São Paulo, Brasil

Email: rpuziol.oliveira@gmail.com

Contribuição do autor