Programa Interinstitucional de Pós-Graduação em Estatística UFSCar-USP (PIPGEs)

Éder Silva de Brito | 03 de Dezembro de 2020

Estimação e Predição para Distribuição Flexible Weibull baseada em Censura Tipo-II Progressiva

Sumário

Proposta do Artigo

Referência: Bdair, O. M., Awwad, R. A., Abufoudeh, G. K., Naser, M. F. M. (2020). *Estimation and prediction for flexible Weibull distribution based on progressive type II censored data*. Communications in Mathematics and Statistics, 8(3), 255-277.

Objetivos:

- Computar o estimador de Bayes para os parâmetros da distribuição Flexible Weibull baseada em uma amostra sob censura do tipo-II progressiva, considerando prioris Gama e função de perda quadrática, bem como comparar suas performances com os estimadores de máxima verossimilhança (EMV) para diferentes esquemas de censura.
- Realizar predições para dados não observados (censurados) baseadas nos dados observados, utilizando algoritmos de importância e Metropolis-Hastings (M-H).

Censura do Tipo-II Progressiva

É uma generalização da censura do Tipo II. Nesse caso, determina-se um número máximo de falhas (eventos de interesse) a serem observadas (m) e determina-se um esquema de censura $(R_1, \ldots, R_m)$ onde, após a ocorrência da i-ésima falha, R_i indivíduos serão retirados do estudo (censurados).

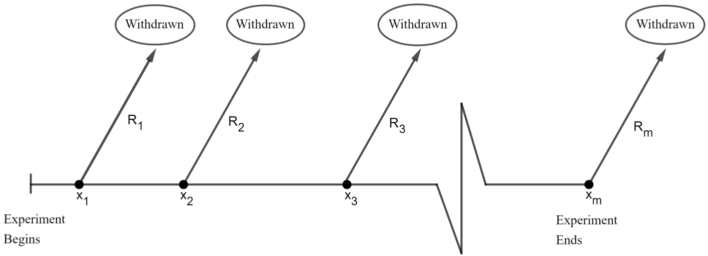


Figura: Esquema de Censura Tipo-II Progressiva

Censura do Tipo-II Progressiva

Suponhamos que n indivíduos estejam em teste e define-se um esquema de censura do tipo-II progressiva $(R_1, \ldots, R_m)$ de modo que $m + \sum_{i=i}^{m} R_i = n$. Suponhamos ainda que os tempos de vida são observações i.i.d. de uma variável aleatória X com função de densidade $f(x;\theta)$ e função de distribuição acumulada $F(x;\theta)$, onde θ é o parâmetro (ou vetor de parâmetros) da distribuição.

Se $x_1, \ldots, x_m$ são os m tempos ordenados de falha observados, a função de verossimilhança é dada por

$$L(\theta) = C \prod_{i=1}^{m} f(x_i;\theta)[1 - F(x_i;\theta)]^{R_i}, \qquad (1)$$

onde $C = n(n - R_1 - 1)(n - R_1 - R_2 - 2)\cdots(n - R_1 - \cdots - R_{m-1} - m + 1)$ é a constante normalizadora.

- Apresentada por Bebbington et.al. (2007), é uma generalização da distribuição Weibull tradicional.
- Trata-se de uma generalização da distribuição Weibull usual, contando ainda com dois parâmetros.
- A flexibilização em relação à Weibull usual, se dá pelo comportamento da função de risco, já que no modelo de Flexible Weibull proposto, esta função poderá ser não monótona.
- Possui várias possibilidades de aplicação em modelagem de tempo de vida, sobretudo para componentes eletrônicos e outros ligados às diversas áreas de Engenharia.

Distribuição Flexible Weibull

Para todas as funções abaixo, temos $x \geq 0$ e os parâmetros $\alpha > 0$ e $\lambda > 0$.

- A função de densidade da distribuição Flexible Weibull é dada por:

$$f(x|\alpha,\lambda) = \left(\alpha + \frac{\lambda}{x^2}\right)\exp\left(\alpha x - \frac{\lambda}{x}\right)\exp\left(-e^{\alpha x - \frac{\lambda}{x}}\right).$$

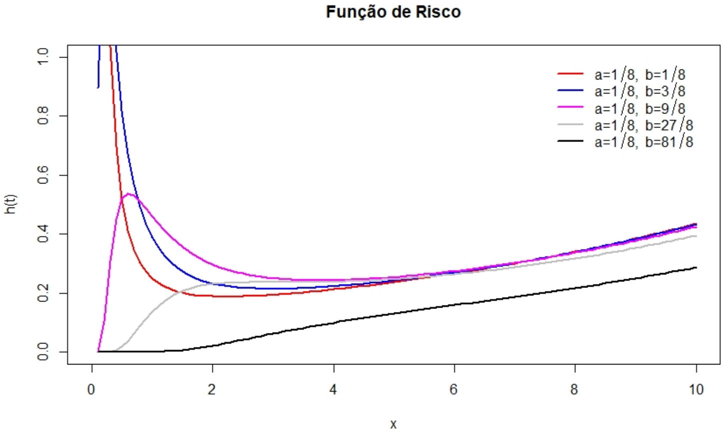
- A função de distribuição acumulada é dada por:

$$F(x|\alpha,\lambda) = 1 - \exp\left(-e^{\alpha x - \frac{\lambda}{x}}\right).$$

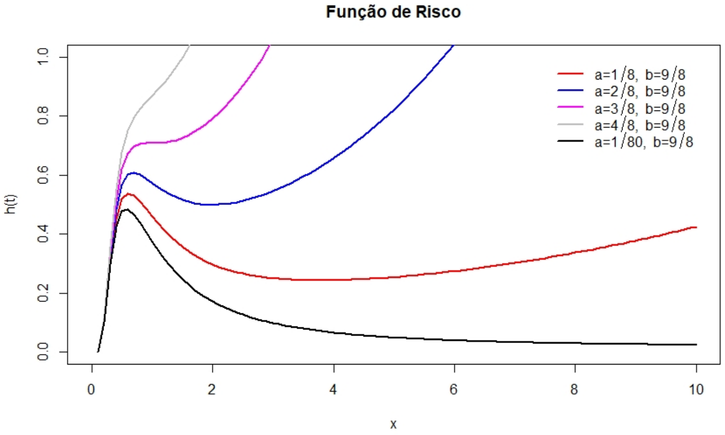
- E por fim, a função de risco é dada por:

$$h(x|\alpha,\lambda) = \frac{f(x)}{1-F(x)} = \left(\alpha + \frac{\lambda}{x^2}\right)\exp\left(\alpha x - \frac{\lambda}{x}\right).$$

1º caso: Mantendo α fixo igual a $1/8$ e alterando os valores de λ entre $1/8$, $3/8$, $9/8$, $27/8$ e $81/8$:



2º caso: Mantendo λ fixo igual a $9/8$ e alterando os valores de α entre $1/8$, $2/8$, $3/8$, $4/8$ e $1/80$:



A função de verossimilhança para a distribuição Flexible Weibull sob censura do tipo-II progressiva e esquema $(R_1, \ldots, R_m)$ é dada por

$$\begin{aligned} L(\alpha, \lambda) &= C \prod_{i=1}^{m} \left(\alpha + \frac{\lambda}{x_i^2} \right) \exp\left(\alpha x_i - \frac{\lambda}{x_i} \right) \exp\left(-(1 + R_i) e^{\alpha x_i - \frac{\lambda}{x_i}} \right) \\ &= C \left[\prod_{i=1}^{m} \left(\alpha + \frac{\lambda}{x_i^2} \right) \right] \exp\left(\sum_{i=1}^{m} \alpha x_i - \frac{\lambda}{x_i} \right) \exp\left(- \sum_{i=1}^{m} (1 + R_i) e^{\alpha x_i - \frac{\lambda}{x_i}} \right). \end{aligned}$$

Prioris

- As prioris de α e λ serão Gama com hiperparâmetros (a, b) e (c, d), respectivamente:

$$\pi_1(\alpha|a, b) = \frac{b^a}{\Gamma(a)}\alpha^{a-1}e^{-b\alpha} \quad \text{e} \quad \pi_2(\lambda|c, d) = \frac{d^c}{\Gamma(c)}\lambda^{c-1}e^{-d\lambda}.$$

- A priori conjunta de α e λ é dada por

$$\pi(\alpha, \lambda) = \pi_1(\alpha|a, b)\pi_2(\lambda|c, d).$$

- A densidade a posteiori conjunta de α e λ é dada por

$$\pi(\alpha, \lambda|\mathbf{x}) = \frac{L(\alpha, \lambda|\mathbf{x})\pi(\alpha, \lambda)}{\int_0^\infty \int_0^\infty L(\alpha, \lambda|\mathbf{x})\pi(\alpha, \lambda)d\alpha d\lambda}.$$

Considerando a função de perda quadrática, o estimador de Bayes para uma função qualquer $\theta = g(\alpha, \lambda)$ é dado por

$$\theta_{BE} = E_{\pi}[\theta|\mathbf{x}] = \frac{\int_0^\infty \int_0^\infty g(\alpha, \lambda) L(\alpha, \lambda|\mathbf{x}) \pi(\alpha, \lambda) d\alpha d\lambda}{\int_0^\infty \int_0^\infty L(\alpha, \lambda|\mathbf{x}) \pi(\alpha, \lambda) d\alpha d\lambda}$$

Ideia: desenvolver um algoritmo usando método de amostragem de Gibbs e o algoritmo de M-H para computar o estimador de Bayes.

- A distribuição condicional completa a posteriori de α dados λ e $\mathbf{x}$ é dada por

$$\pi_1(\alpha|\lambda,\mathbf{x}) \propto \left[\prod_{i=1}^{m}\left(\alpha+\frac{\lambda}{x_i^2}\right)\right] e^{-\alpha\left(b-\sum_{i=1}^{m} x_i\right)} e^{-\sum_{i=1}^{m}(1+R_i)e^{\alpha x_i-\frac{\lambda}{x_i}}} \alpha^{a-1}.$$

- A distribuição condicional completa a posteriori de λ dados α e $\mathbf{x}$ é dada por

$$\pi_2(\lambda|\alpha,\mathbf{x}) \propto \left[\prod_{i=1}^{m}\left(\alpha+\frac{\lambda}{x_i^2}\right)\right] e^{-\lambda\left(d+\sum_{i=1}^{m}\frac{1}{x_i}\right)} e^{-\sum_{i=1}^{m}(1+R_i)e^{\alpha x_i-\frac{\lambda}{x_i}}} \lambda^{c-1}.$$

- Como π_1 e π_2 não podem ser reduzidas à uma distribuição conhecida e não é possível gerar α e λ diretamente dessas distribuições por métodos padrão, foi utilizado o algoritmo M-H com distribuição proposta normal.

Estimador de Bayes

Após obter uma amostra MCMC $\{(\alpha_i, \lambda_i); i = 1, \ldots, N\}$, o estimador de Bayes de $\theta = g(\alpha, \lambda)$ baseado na função de perda quadrática e a variância a posteriori são dados por

$$\hat{\theta}_{BE} = \frac{1}{N} \sum_{i=1}^{N} g(\alpha_i, \lambda_i) \quad \text{e} \quad V\hat{a}r[\theta|\mathbf{x}] = \frac{i=1}{N} \sum_{1}^{N} (\theta_i - \hat{\theta}_{BE})^2$$

Então

- se $g(\alpha, \lambda) = \alpha$, temos:

$$\hat{\alpha}_{BE} = \frac{1}{N} \sum_{i=1}^{N} \alpha_i \quad \text{e} \quad V\hat{a}r[\alpha|\mathbf{x}] = \frac{1}{N} \sum_{i=1}^{N} (\alpha_i - \hat{\alpha}_{BE})^2$$

- se $g(\alpha, \lambda) = \lambda$, temos:

$$\hat{\lambda}_{BE} = \frac{1}{N} \sum_{i=1}^{N} \lambda_i \quad \text{e} \quad V\hat{a}r[\lambda|\mathbf{x}] = \frac{1}{N} \sum_{i=1}^{N} (\lambda_i - \hat{\lambda}_{BE})^2$$

- Ideia: Predizer um tempo de falha não observado ou censurado baseado na amostra de tempos de falha observados (dita amostra *informativa*).

- Consideraremos a estimação da densidade preditiva a posteriori da k-ésima ordem $Y_{k:R_j}$, $k = 1, \ldots, R_j$, $j = 1, \ldots, m$, baseada em uma amostra sob censura tipo-II progressiva $\mathbf{x} = (x_1, \ldots, x_m)$.

- Usando a função de perda quadrática, temos que o preditor de Bayes (BP) de $Y = Y_{k:R_j}$ pode ser obtido por

$$
\begin{aligned}
Y_{k:R_j}^{BP} &= E_p[Y|\mathbf{x}] \\
&= \int_{x_j}^{\infty} y \int_0^{\infty} \int_0^{\infty} f_{k:R_j|\mathbf{x}}(y_{k:R_j}|\alpha, \lambda)\pi(\alpha, \lambda|\mathbf{x}) d\alpha d\lambda dy,
\end{aligned}
$$

onde $p(y_{k:R_j}|\mathbf{x})$ é a densidade preditiva a posteriori de $Y_{k:R_j}$

Utilizando novamente o amostrador de Gibbs e o algoritmo M-H proposto anteriormente para gerar uma amostra MCMC $\{(\alpha_l, \lambda_l), l = 1, \ldots, M\}$, podemos também obter uma estimativa $\hat{Y}_{k:R_j}^{BP}$ de $Y_{k:R_j}$ baseado na amostra:

$$\begin{aligned}\hat{Y}_{k:R_j}^{BP} &= \frac{C}{M}\sum_{l=1}^{M}\sum_{i=0}^{k-1}\binom{k-1}{i}(-1)^{k-i-1}e^{(R_j-i)e^{\alpha_l x_j-\frac{\lambda_l}{x_j}}} \\ &\quad\times\int_{x_j}^{\infty} y\left(\alpha_l+\frac{\lambda_l}{y^2}\right)e^{\alpha_l y-\frac{\lambda_l}{y}}e^{(i-R_j)e^{\alpha_l y-\frac{\lambda_l}{y}}}\,dy\end{aligned}$$

Obs: A integral pode ser reparametrizada de modo a ficar definida num intervalo $(0, 1)$ e, assim, ser calculada.

Estudo de Simulação

- Um estudo de simulação considerando diferentes tamanhos de amostra observada e diferentes esquemas de censura foi realizado para verificar o comportamento dos estimadores de Bayes e compará-los com os EMV.
- Foram geradas amostras sob censura do tipo-II progressiva utilizando o algoritmo apresentado por Balakrishnan(1995). Aqui consideraremos apenas o caso de $\alpha = 2$ e $\lambda = 1$
- Para as prioris, foram consideradas duas possibilidades. Na primeira, assumindo $a = b = c = d = 0$ (uma priori não informativa), denotada por *Priori 0*. Na segunda, assumindo $a = b = d = 1$ e $c = 2$, denotada por *Priori 1*.
- Para o algoritmo de M-H, a distribuição proposta é uma normal que depende estado anterior. Especificamente, o valor dos parâmetros propostos será gerado de uma normal centrada no parâmetro anterior e a variância considerada é a mesma obtida através da matriz de informação de Fisher.

Estudo de Simulação

n	m	R	$\hat{\theta}$	EMV	Priori 0	Priori 1
25	10	(15, 0*9)	α	2.27 (0.47)	2.09 (0.37)	2.02 (0.24)
			λ	1.09 (0.91)	1.07 (0.95)	1.07 (0.92)
25	10	(0*4,8,7,0*4)	α	2.48 (1.13)	2.11 (0.84)	2.00 (0.34)
			λ	1.11 (0.88)	1.05 (0.98)	1.04 (0.96)
25	10	(0*9,15)	α	2.84 (2.94)	1.98 (2.30)	1.82 (0.59)
			λ	1.18 (0.81)	1.04 (1.05)	1.01 (1.02)
25	15	(10,0*14)	α	2.21 (0.27)	2.09 (0.22)	2.05 (0.17)
			λ	1.08 (0.91)	1.05 (0.96)	1.06 (0.93)
25	15	(0*7,5,5,0*6)	α	2.26 (0.40)	2.09 (0.32)	2.03 (0.22)
			λ	1.08 (0.90)	1.04 (0.96)	1.05 (0.95)
25	15	(0*14,10)	α	2.32 (0.68)	2.06 (0.57)	1.97 (0.31)
			λ	1.11 (0.87)	1.05 (0.97)	1.04 (0.97)
40	20	(20,0*19)	α	2.14 (0.18)	2.05 (0.15)	2.03 (0.13)
			λ	1.06 (0.94)	1.04 (0.97)	1.05 (0.94)
40	20	(1*20)	α	2.25 (0.36)	2.10 (0.28)	2.03 (0.21)
			λ	1.06 (0.91)	1.03 (0.97)	1.03 (0.97)
40	20	(0*9,20)	α	2.30 (0.59)	2.07 (0.51)	1.98 (0.30)
			λ	1.08 (0.89)	1.04 (0.97)	1.02 (0.98)
40	30	(10,0*29)	α	2.10 (0.11)	2.04 (0.10)	2.03 (0.09)
			λ	1.04 (0.96)	1.02 (0.99)	1.03 (0.97)
40	30	(5,0*28,5)	α	2.11 (0.13)	2.04 (0.12)	2.02 (0.10)
			λ	1.04 (0.95)	1.02 (0.99)	1.03 (0.97)
40	30	(0*29,10)	α	2.12 (0.17)	2.04 (0.15)	2.01 (0.13)
			λ	1.04 (0.95)	1.02 (0.99)	1.02 (0.98)

Foi utilizado um conjunto de dados que contém os tempos de falha de 50 dispositivos eletrônicos colocados em teste. Uma censura do tipo-II progressiva foi aplicada nesse conjunto de $n = 50$ dados, tomando-se uma amostra observada de tamanho $m = 11$. O esquema de censura e os tempos de falha observados estão elencados na tabela abaixo.

i	1	2	3	4	5	6	7	8	9	10	11
R_i	0	0	5	0	6	5	5	5	0	5	8
x_i	0.1	0.2	1	3	6	18	45	63	72	75	84

As estimativas obtidas pelos EMV e EB utilizando-se a priori 1 estão elencados abaixo:

	α	λ
EMV	0.4971 (0.1610)	0.0047 (0.0261)
Priori 1	0.6031 (0.0602)	0.0054 (0.0160)

Aplicação em Dados Reais - Predição

$n = 50, m = 11$	Valores Preditivos	$95\% IC$
$Y_{1,3}$	1.153	(0.069, 3.917)
$Y_{3,3}$	1.329	(0.070, 4.124)
$Y_{1:5}$	11.987	(6.711, 18.884)
$Y_{5,5}$	18.803	(13.982, 28.029)
$Y_{1:6}$	18.113	(14.823, 27.376)
$Y_{3:6}$	31.925	(25.728, 35.757)
$Y_{3:7}$	55.247	(49.923, 61.728)
$Y_{5,7}$	57.949	(53.086, 68.117)
$Y_{1,8}$	66.098	(59.097, 73.392)
$Y_{5,8}$	73.022	(62.020, 84.991)
$Y_{3,10}$	80.726	(72.628, 93.167)
$Y_{5:10}$	86.196	(81.923, 101.427)
$Y_{1:11}$	85.752	(80.236, 100.562)
$Y_{7:11}$	90.936	(84.876, 111.139)

Considerações Finais

- Nesse estudo foi apresentada uma proposta Bayeasiana de inferência e predição para dados não observados de uma amostra da distribuição Flexible Weibull sob censura do tipo-II progressiva;
- Considerou-se prioris Gama para os parâmetros α e λ e obteve-se um método de estimação desses parâmetros utilizando-se método de amostragem de Gibbs e o algoritmo de M-H;
- As simulações indicaram vantagens da proposta do artigo em relação aos EMV e também vantagens da priori informativa em relação à priori não informativa. Além disso, foi possível realizar predições de dados não observados em uma amostra sob censura tipo-II progressiva.

Referências

- Balakrishnan, N., Sandhu, R. A. (1995). *A simple simulational algorithm for generating progressive Type-II censored samples*. The American Statistician 49(2), 229-230.
- Bdair, O. M., Awwad, R. A., Abufoudeh, G. K., Naser, M. F. M. (2020). *Estimation and prediction for flexible Weibull distribution based on progressive type II censored data*. Communications in Mathematics and Statistics, 8(3), 255-277.
- Bebbington, M., Lai, C. D., Zitikis, R. (2007). *A flexible Weibull extension*. Reliability Engineering and System Safety, 92(6), 719-726.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., Rubin, D. B. (2013). *Bayesian data analysis*. CRC press.

ufscar

ICMC USP
SÃO CARLOS
Serviço de Pós-Graduação