Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”

O modelo Burr XII Geométrico: propriedades e aplicações

Beatriz Rezende Lanjoni

Dissertação apresentada para obtenção do título de Mestra
em Ciências. Área de concentração: Estatística e Experi-
mentação Agronômica

Piracicaba
2013

Beatriz Rezende Lanjoni
Bacharel em Estatística

O modelo Burr XII Geométrico: propriedades e aplicação

Versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientador:
Prof. Dr. **EDWIN MOISES ORTEGA**

Dissertação apresentada para obtenção do título de Mestra
em Ciências. Área de concentração: Estatística e Experi-
mentação Agronômica

Piracicaba
2013

**Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - ESALQ/USP**

Lanjoni, Beatriz Rezende

O modelo Burr XII Geométrico: propriedades e aplicações / Beatriz Rezende Lanjoni. -
versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2013.
89 p. : il.

Dissertação (Mestrado) - - Escola Superior de Agricultura "Luiz de Queiroz", 2013.
Bibliografia.

1. Dados censurados 2. Distribuição Burr XII 3. Distribuição Geométrica 4. Função de
sobrevivência 5. Máxima verossimilhança I. Título

CDD 519.54
L288M

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor"

DEDICATÓRIA

Primeiramente a DEUS por ter me guiado pelo melhor caminho nos momentos de dúvidas, me protegido durante minha caminhada pela vida e por ter me abençoado com familiares e amigos tão especiais.

Aos meus pais, Sidney José Lanjoni e Vera Rezende Lanjoni pelo amor a mim dedicado, pelo apoio incondicional e por confiarem no meu potencial muitas vezes mais que eu mesma.

Aos meus avós, José Lanjoni, Nilce de Oliveira Lanjoni e Iracema Alves de Rezende que com amor, ternura e paciência foram essenciais para eu ser quem eu sou hoje e que sempre estão de braços abertos para me receber no aconchego de suas casas.

AGRADECIMENTOS

Agradeço aos meus PAIS pelo lar de harmonia e confiança mútua que temos uns nos outros. O sorriso no rosto de vocês por cada conquista minha é meu maior estímulo para sempre continuar e cada dia chegar mais longe.

Aos meus FAMILIARES, que mesmo com a distância estão sempre presentes na minha vida.

Ao Prof. Dr. Edwin Moisés Ortega pela dedicação e apoio sem os quais a realização deste trabalho não seria possível.

Aos PROFESSORES do curso de Pós-graduação em Estatística e Experimentação Agronômica, Dra. Clarice Garcia Borges Demétrio, Dra. Sônia Maria de Stefano Piedade, Dra. Roseli Aparecida Leandro, Dr. César Gonçalves de Lima e Dr. Taciana Vilela Savian pela formação.

Aos FUNCIONÁRIOS do Departamento de Ciências Exatas da ESALQ/USP, em especial a Luciane Brajão pela disposição em sempre ajudar no que fosse necessário.

À CNPQ pelo apoio financeiro.

Agradeço também aos colegas e professores do COLÉGIO PEQUENO PRINCÍPE em São José do Rio Preto que fizeram parte de uma infância doce e uma adolescência tranquila, em especial à Irmã Mariluce e as grandes amigas Andrea Raphaela das Neves e Geisa Alice Brasilino. Aos professores e colegas da UNESP PRESIDENTE PRUDENTE que despertaram em mim o amor por essa ciência chamada Estatística, em especial ao professor Mário Tarumoto e as grande amigas Natalia Heimann e Livia Matos Garcia.

E finalmente à Felipe Alves Rodrigues Bergamo pelo apoio e compreensão nos momentos de dificuldade, por cada palavra de incentivo e pelo amor que a cada dia construímos juntos.

SUMÁRIO

RESUMO	9
ABSTRACT	11
LISTA DE FIGURAS	13
LISTA DE TABELAS	15
1 INTRODUÇÃO	17
2 DESENVOLVIMENTO	21
2.1 Conceitos básicos de análise de sobrevivência	21
2.2 A Distribuição BXII	22
2.3 A transformação geométrica	23
3 A DISTRIBUIÇÃO BXII GEOMÉTRICA	29
3.1 A distribuição BXIIGI	29
3.2 A distribuição BXIIGII	31
3.3 Expansão da distribuição BXII Geométrica como combinação linear da distribuição BXII	33
3.4 Momentos	34
3.5 Assimetria e Curtose	35
3.6 Função Geradora de Momentos	37
3.7 Desvio Médio	40
3.8 Função quantílica	44
3.9 Estimação dos parâmetros	49
4 MODELOS DE REGRESSÃO PARAMÉTRICOS	55
4.1 Modelos de Regressão locação-escala	55
5 MODELOS UNIFICADOS DE FRAÇÃO DE CURA	65
6 APLICAÇÕES	73
6.1 Churn Voluntário de clientes	73
6.2 Dados de entomologia	74
6.3 Dados de melanoma cutâneo	79
7 CONCLUSÕES	83
REFERÊNCIAS	85

RESUMO

O modelo Burr XII Geométrico: propriedades e aplicações

No presente trabalho são propostos dois modelos para dados censurados baseados na mistura da distribuição geométrica e na distribuição Burr XII considerando duas ativações latentes, máximo e mínimo. A distribuição Burr XII tem três parâmetros e é uma generalização da distribuição log-logística. Por sua vez a distribuição Burr XII Geométrica tipo I e tipo II tem quatro parâmetros e são generalizações da distribuição Burr XII relacionados as ativações latentes do mínimo e máximo respectivamente. Foram apresentadas algumas propriedades das duas novas distribuições tais como momentos, assimetria, curtose, função geradora de momentos e desvio médio. Além disso, foi introduzido os modelos de regressão correspondentes, log Burr XII Geométrica tipo I e log Burr XII Geométrica tipo II. Adicionalmente foi desenvolvido um modelo de sobrevivência com fração de cura assumindo que o número de causas competitivas do evento de interesse segue a distribuição geométrica e o tempo do evento segue a distribuição Burr XII. Para todos os modelos desenvolvidos foi utilizado o método da máxima verossimilhança para estimar os parâmetros, que possibilita a construção de intervalos de confiança e testes de hipóteses. Por fim, são apresentadas três aplicações para ilustrar os modelos propostos.

Palavras-chave: Dados censurados; Distribuição Burr XII; Distribuição geométrica; Função de sobrevivência; Máxima verossimilhança

ABSTRACT

The model Burr XII Geometric: properties and applications

In this paper are proposed two models for censored data based on the mixture of geometric distribution and Burr XII distribution considering two latent activations, maximum and minimum. The Burr XII distribution has three parameters and is a generalization of the log-logistic distribution. On the other hand Burr XII Geometric type I distribution and type II has four parameters and are a generalization of the Burr XII distribution related to minimum and maximum activations respectively. It were presented some properties of the news distributions such as moments, skewness, kurtosis, moment generating function and mean deviation. Furthermore, it was introduced two regression models, the log Burr XII Geometric type I and the log Burr XII Geometric type II. Additionally a new cure rate survival was formulated by assuming that the number of competing causes of the event of interest has the geometric distribution and the time to this event follows Burr XII distribution. For all models was developed the maximum likelihood method to estimate the parameters, which allows the construction of confidence intervals and hypothesis testing. Finally, three applications are presented to illustrate the proposed models.

Keywords: Data censored; Burr XII distribution; Geometric distribution; Survival function; Maximum likelihood

LISTA DE FIGURAS

Figura 1 - Exemplo de funções de taxa de falha	22
Figura 2 - Gráficos da fdp da distribuição BXXIGI	29
Figura 3 - Gráficos da função de sobrevivência da distribuição BXXIGI	30
Figura 4 - Gráficos da função taxa de falha da distribuição BXIIGI	30
Figura 5 - Gráficos da fdp da distribuição BXIIGII	31
Figura 6 - Gráficos da função de sobrevivência da distribuição BXIIGII	32
Figura 7 - Gráficos da função taxa de falha da distribuição BXIIGII	32
Figura 8 - Gráficos de assimetria e curtose da distribuição BXIIGI (p variando de 0 a 1)	36
Figura 9 - Gráficos de assimetria e curtose da distribuição BXIIGII (p variando de 0 a 1)	37
Figura 10 - Gráficos da fdp da distribuição LBXIIGI	57
Figura 11 - Gráficos da fdp da distribuição LBXIIGII	58
Figura 12 - Função de sobrevivência própria e imprópria	67
Figura 13 - Gráficos da função de sobrevivência da distribuição BXII considerando fração de cura	69
Figura 14 - EKM para o tempo do vida de clientes e sobrevivência estimada da distribuição BXIIGI	74
Figura 15 - (a) Densidade da distribuição BXII e BXIIGII ajustada para os dados de entomologia. (b) Função de sobrevivência da distribuição BXII e BXIIGII estimada e sobrevivência empírica para os dados de entomologia	77
Figura 16 - Função de sobrevivência do modelo LBXIIGII estimada e curva de sobrevivência empírica para os dados de entomologia	79
Figura 17 - Função de sobrevivência do modelo BXII com fração de cura estimada e curva de sobrevivência empírica para os dados de melanoma cutâneo (b) Função de sobrevivência do modelo BXII com fração de cura estimada e curva de sobrevivência empírica para os dados de melanoma cutâneo estratificado pelo status de ulceração	81

LISTA DE TABELAS

Tabela 1 - Estimativas de m.v. do modelo para os dados do tempo de vida de clientes	73
Tabela 2 - Teste da razão de verossimilhança	74
Tabela 3 - Estimativas dos parâmetros do modelo BXIIGII ajustado para dados de entomologia, correspondentes desvios padrões em parênteses e estatísticas AIC, BIC e CAIC	76
Tabela 4 - Teste da razão de verossimilhança	76
Tabela 5 - Estimativas dos parâmetros do modelo de regressão LBXIIGII ajustado para dados de entomologia, correspondentes desvios padrões em parênteses e estatísticas AIC, BIC e CAIC	78
Tabela 6 - Teste da razão de verossimilhança	78
Tabela 7 - Estimativas dos parâmetros do modelo de regressão LBXIIGII ajustado para dados de entomologia, correspondentes desvios padrões em parênteses e estatísticas AIC, BIC e CAIC	78
Tabela 8 - Estimativas de máxima verossimilhança dos parâmetros do modelo de regressão BXIIG considerando fração de cura ajustado para os dados de melanoma cutâneo	80

1 INTRODUÇÃO

Nas últimas décadas, a análise de sobrevivência teve um grande desenvolvimento e vem sendo aplicada nas mais diversas áreas, biologia, agronomia, engenharia, medicina entre outras. Na análise de sobrevivência a variável resposta é o tempo até a ocorrência de um evento de interesse. Esse tempo, em geral, é uma variável aleatória que assume valores reais positivos e com distribuição absolutamente contínua. Geralmente, este tipo de dados apresenta uma distribuição assimétrica não-negativa, então não se pode assumir a normalidade da distribuição subjacente. Assim sendo, não podem ser utilizados os modelos estatísticos usuais (Bastos J.; Rocha, C., 2006).

Outra característica especial dos dados de sobrevivência é a presença de censura, que pode ser interpretada como uma observação parcial da resposta. Existem diversos tipos de censura. Uma censura é dita à direita quando o evento de interesse ocorreu após o tempo registrado, por outro lado uma censura é dita à esquerda quando o evento de interesse já ocorreu quando a unidade experimental foi observada. Em relação aos mecanismos tem-se censura do tipo I, que é quando não é possível esperar que todas as unidades experimentais em teste falhem e é necessário que o experimento se encerre em um determinado tempo fixo; censura do tipo II é quando o experimento se encerra depois que um certo número de falhas tenha ocorrido; censura aleatória, quando ocorre a perda de informação de determinadas unidades experimentais pelos mais diversos motivos.

As estimativas da função de sobrevivência de uma variável aleatória T podem ser feitas através de métodos paramétricos e não paramétricos. Dentre os métodos não paramétricos para estimar a função de sobrevivência de T , destacam-se o estimador de Kaplan-Meier, estimador atuarial (tabela de vida) e o estimador de Nelson-Aalen.

Quando distribuições de probabilidade conhecidas são utilizadas na análise de sobrevivência, ou seja, quando recorre-se aos modelos paramétricos, é utilizado o método de máxima verossimilhança para a estimação dos parâmetros o que possibilita a construção de intervalos de confiança e teste de hipóteses. Vale ressaltar que a escolha do modelo probabilístico adequado para descrever o tempo de falha deve ser realizada com cuidado, pois a escolha errônea da distribuição de probabilidade para a variável aleatória T pode induzir a conclusões não corretas quanto a realização de inferências sobre a população amostrada.

Dentre as distribuições mais usadas para modelar tempo de vida destacam-se as distribuições exponencial, Weibull, log-normal, log-logística e gama, além de suas gene-

ralizações, distribuições exponencializadas, distribuições Betas, distribuições Kumaraswamy e distribuições McDonald. Muitos trabalhos mostram a aplicação de diversas outras distribuições não tão conhecidas ou transformações de distribuições já conhecidas.

Em muitas situações práticas as distribuições de probabilidade clássicas não fornecem estimativas adequadas para dados reais. Recentemente muitos métodos para introduzir um ou mais parâmetros em distribuições generalizadas tem sido estudados na literatura estatística e, entre estes métodos a composição de alguma distribuição discreta com alguma distribuição usual tem sido muito utilizada. Adamidis e Loukas (1998) foram pioneiros ao desenvolverem a distribuição exponencial geométrica (EG), distribuição esta que compõe a distribuição exponencial e geométrica. De maneira similar as distribuições exponencial Poisson (EP) e exponencial logarítmica (EL) foram introduzidas por Kus (2007) e Tahmasbi e Rezaei (2008) respectivamente. Barreto-Souza, Morais e Cordeiro (2010) e Lu e Shi (2011) propuseram as distribuições Weibull geométrica (WG) e Weibull Poisson (WP) que são extensões das distribuições EG e EP respectivamente. Rodrigues et al. (2011) definiu a distribuição Weibull binomial negativa (WNB) que inclui como sub modelos as distribuições WG e WP. Silva e Cordeiro (2013) ¹ definiram as distribuição Burr XII (BXII) séries de potência.

Com finalidade de flexibilizar as distribuições de tal forma que a função de taxa de falha possa acomodar formas do tipo monótona e unimodal, este trabalho apresenta duas novas distribuições baseadas na mistura da distribuição geométrica e na distribuição BXII considerando duas ativações latentes. Além de apresentar algumas de suas propriedades, tais como, momentos, assimetria, curtose, função geradora de momentos e desvio médio.

A influência de uma ou mais covariáveis no tempo de sobrevivência pode ser estudada através de modelos de regressão, estes se apresentam em duas classes de modelos de regressão, os modelos paramétricos, também chamados de modelos de tempo de vida acelerado ou modelos de locação e escala, e os semi-paramétricos, conhecidos como modelos de riscos proporcionais ou modelo de regressão de Cox, que é sem dúvida um dos modelos mais populares na análise de sobrevivência (Santana, T. V. F. de, 2010).

Como na maioria dos estudos de análise de sobrevivência, o interesse principal é prever e analisar os tempos de falha em função das covariáveis e verificar como essas variáveis dependentes influenciam no tempo observado. É de interesse deste trabalho

¹The Burr XII power series distributions: A new compounding family. Trabalho submetido ao Brazilian Journal of Probability and Statistics.

estudar as novas distribuições sob o enfoque de modelos de regressão paramétricos e conhecer em quais situações práticas da análise de sobrevivência eles seriam apropriadas.

Em muitos estudos de acompanhamento médico, pacientes que respondem bem ao tratamento podem tornar-se imunes aos sinais e sintomas da doença e assim serem considerados curados, ou seja, não suscetíveis ao evento de interesse. Sendo assim, é importante estudar a proporção de pacientes curados e analisar os efeitos das covariáveis sobre o tempo de sobrevivência, assim como também sobre a fração de cura (Paes, A. T., 2007).

Surge então a necessidade de considerar modelos com fração de cura, ou seja, modelos que tem função de sobrevivência imprópria. Esses modelos permitem a formulação de conclusões mais precisas e informativas e tem sido usado para modelar dados de tempo de vida relativos a vários tipo de câncer. Se um componente curado não está presente a análise é reduzida a abordagem padrão de análise de sobrevivência. Em muitos casos, um indivíduo é considerado curado após cinco anos imune ao evento de interesse. Para modelos com fração de cura, Berkson e Gage (1952) e Maller e Zhou (1996) construíram uma função de sobrevivência populacional imprópria na forma de mistura. Existe também um modelo biológico para modelar fração de cura, o modelo de tempo de promoção. Por último, o modelo unificado de fração de cura considera a função geradora de probabilidades de determinada distribuição discreta, uma vez que a função geradora de probabilidades só está definida para esta classe de distribuições.

Devido ao exposto acima, outro objetivo deste trabalho foi propor o modelo unificado de fração de cura baseado também na distribuição geométrica e BXII.

Finalmente algumas aplicações para dados reais foram apresentadas com o intuito de mostrar a superioridade das distribuições propostas em relação aos seus casos particulares.

Este presente estudo está organizado da seguinte maneira. Na seção 2, uma breve introdução de análise de sobrevivência é apresentada. Na seção 3, duas novas distribuições de probabilidade são obtidas a partir da distribuição BXII e da distribuição geométrica. Algumas caracterizações destas novas distribuições também são desenvolvidas. Na seção 4, são apresentados os modelos de regressão locação e escala. Na seção 5 é apresentado o modelo unificado de fração de cura. E finalmente na seção 6, estudou-se alguns conjuntos de dados reais afim de comparar os resultados com distribuições já conhecidas.

2 DESENVOLVIMENTO

2.1 Conceitos básicos de análise de sobrevivência

A origem da análise de sobrevivência ocorreu há vários séculos. Em 1662 o livro *Natural and Political Observations upon the Bills of Mortality* de John Graunt, pela primeira vez realiza um estudo sobre mortalidade de acordo com a idade, data de falecimento, sexo e causa de morte. Alguns anos depois, em 1693, Edmund Halley desenvolveu a primeira tabela de mortalidade, muito semelhante às utilizadas atualmente em demografia e em estudos atuariais (Bastos J.; Rocha, C., 2006).

No fim do século XX a análise de sobrevivência teve um grande desenvolvimento, com muitas publicações, como Miller (1981), Lawless (1982), Cox e Oakes (1984), Kalbfleisch e Prentice (2002), Collett (2003), Klein e Moeschberger (2003). Já para temas mais específicos como, extensões do modelo semiparamétrico de Cox, Therneau e Grambsch (2000) e análise de sobrevivência com enfoque bayesiano Ibrahim, Chen e Sinha (2000) (Ferreira, J. M., 2007).

Os dados de sobrevivência, em geral, são representados pelo par (t_i, δ_i) , para cada indivíduo $i = 1, \dots, n$ em que t_i é o tempo de falha ou censura e δ_i é uma variável indicadora definida por:

$$\delta_i = \begin{cases} 1, & \text{se corresponde ao tempo de falha} \\ 0, & \text{se corresponde ao tempo de censura.} \end{cases}$$

A variável aleatória T contínua e não-negativa representa o tempo de falha. Três funções são particularmente úteis em análise de sobrevivência:

$f(t), t > 0$: função densidade;

$S(t) = P(T \geq t) = 1 - F(t)$: função de sobrevivência;

$h(t) = \lim_{h \rightarrow 0} \frac{P(t \leq T \leq t+h | T \geq t)}{h} = \frac{f(t)}{S(t)}$: função taxa de falha.

A função de sobrevivência pode ser interpretada como a probabilidade de uma observação não falhar até o tempo t .

A função de taxa de falha é interpretada como o risco de ocorrência do evento até o tempo t . O gráfico da função de taxa de falha pode apresentar algumas formas básicas conforme pode ser visualizado na figura 1, com diferentes interpretações práticas. Pode ser crescente ou decrescente conforme indicado por $h_1(t)$ e $h_2(t)$ respectivamente, constante representado por $h_3(t)$, ou alternar entre intervalos decrescentes, constantes e crescentes, como por exemplo $h_4(t)$ (forma de U ou banheira) e forma unimodal em $h_5(t)$, que inicia com um intervalo crescente até atingir um máximo e então decresce. (Santana, T. V. F., 2010)

Figura 1 – Exemplo de funções de taxa de falha

2.2 A Distribuição BXII

A distribuição Burr XII, também conhecida como BXII, foi construída em 1942 por Irving W. Burr. Existem diversas aplicações da distribuição BXII, uma vez que sua função densidade apresenta uma grande variedade de formas. As aplicações incluem simulação, resposta quântica, aproximação de distribuições e desenvolvimento de gráficos de controle não-normal. Rodriguez (1977) mostrou que a área de cobertura da BXII de um plano específico é ocupado por várias distribuições conhecidas, incluindo a normal, log-normal, gama e valor extremo tipo I. A distribuição BXII tem, ainda, como submodelos especiais a distribuição Weibull e a log-logística, que são distribuições muito comuns para modelar dados de tempo de vida e taxa de falha monótona. Entretanto essas distribuições não fornecem uma estimativa paramétrica razoável para modelar fenômenos com taxa de falha não monótona como a forma de banheira ou unimodal, que são comuns em estudos biológicos e confiabilidade. A taxa de falha unimodal pode ser observada no curso de uma doença que atinge um pico depois de um período finito e então decai gradualmente.

Outra vantagem dessa distribuição é que sua função acumulada e de sobrevivência tem forma fechada, o que simplifica o cálculo de percentis e da função de sobrevivência para dados censurados. Shao (2004) discutiu o estimador de máxima verossimilhança dos parâmetros da BXII e Shao et al. (2004) estudou modelos baseado na distribuição BXII com aplicação de dados relativos a enchentes. Silva et al. (2008) propôs o modelo de regressão log-BXII como uma alternativa ao modelo de regressão log-logística. A variabilidade e flexibilidade tornam essa distribuição atrativa como uma tentativa em modelar dados de tempo de falha (Silva, R. B. e Cordeiro, G. M., 2013)¹.

Se X segue uma variável aleatória (v.a.) com distribuição BXII, ou seja $X \sim \text{BXII}(s, k, c)$, então sua função densidade de probabilidade (fdp) é dada por

$$g(x; s, k, c) = ck \left[1 + \left(\frac{x}{s} \right)^c \right]^{-k-1} \frac{x^{c-1}}{s^c}, \quad x > 0$$

em que $k > 0$ e $c > 0$ são parâmetros de forma e $s > 0$ parâmetro de escala.

A função de distribuição acumulada é dada por:

$$G(x; s, k, c) = 1 - \left[1 + \left(\frac{x}{s} \right)^c \right]^{-k}.$$

O q -ésimo momento é dado por:

$$E[X^q] = s^q B \left[\frac{q}{c} + 1, k - \frac{q}{c} \right], \text{ e } \quad \text{se } ck > q$$

em que $B[a, b]$ é a função beta.

2.3 A transformação geométrica

Como exposto anteriormente, muitos modelos tem sido introduzidos na literatura para estender as distribuições exponencial, Weibull entre outras, tais como Weibull exponenciada, beta exponencial (Nadarajah, S. e Kotz, S., 2004) e beta Weibull modificada (Silva G.O., Ortega, E.M.M. e Cordeiro, G.M., 2010).

O objetivo deste trabalho é propor duas generalizações da distribuição BXII com o intuito de modelar dados de sobrevivência. Essas generalizações serão denominadas distribuição BXII geométrica tipo I (BXIIGI) e BXII geométrica tipo II (BXIIGII).

Essas distribuições são baseadas em uma dupla ativação, isto é, uma extensão da distribuição BXII. A seguir será apresentada a definição e a construção dessas novas distribuições

- Transformação geométrica do tipo I

A nova transformação que se propõe é uma mistura baseada em uma determinada distribuição, neste caso a distribuição BXII, e na distribuição geométrica. Assim seja $X_1, X_2, \dots, X_Z$ uma amostra aleatória de uma distribuição contínua qualquer com fdp $g(x; \boldsymbol{\psi})$ e função de distribuição acumulada $G(x; \boldsymbol{\psi})$ em que $\boldsymbol{\psi}$ indica o vetor de parâmetros e Z uma variável com distribuição geométrica dada por

$$P(z; p) = (1 - p)p^{z-1}, \quad (1)$$

em que $Z \in \mathbb{N}$ e $p \in [0,1]$. Assumindo independência entre $X_1, X_2, \dots, X_Z$ e Z , define-se a variável $T_1 = \min(X_1, X_2, \dots, X_Z)$.

Por meio da distribuição do mínimo para variáveis aleatórias contínuas independentes e identicamente distribuídas (iid) obtém-se a função de distribuição acumulada de T_1 .

Primeiramente obtém-se a função de distribuição acumulada condicional de T_1 dado Z .

$$\begin{aligned} F(t; \boldsymbol{\gamma}|z) &= P(T_1 \leq t|z) \\ &= P[\min(X_1, X_2, \dots, X_Z) \leq t_z] \\ &= 1 - P(X_1 > t, X_2 > t, \dots, X_Z > t) \\ &= 1 - \prod_{i=1}^z P(X_i > t), \end{aligned}$$

em que $\boldsymbol{\gamma}$ é o vetor de parâmetros da distribuição com fdp $f_1(t)$.

Devido $X_1, X_2, \dots, X_Z$ ser iid a última expressão pode ser reescrita como

$$\begin{aligned} F(t; \boldsymbol{\gamma}|z) &= 1 - [P(X > t)]^z \\ &= 1 - [1 - P(X \leq t)]^z \\ &= 1 - [1 - G(t; \boldsymbol{\psi})]^z, \end{aligned} \quad (2)$$

em que $\boldsymbol{\psi}$ é o vetor de parâmetros da distribuição com fdp $g(x; \boldsymbol{\psi})$.

Diferenciando a expressão (2) em relação a t , obtem-se a função densidade de probabilidade de T_1 condicionada a Z .

$$f(t; \boldsymbol{\gamma}|z) = z[1 - G(t; \boldsymbol{\psi})]^{(z-1)} g(t; \boldsymbol{\psi}).$$

Sabendo que

$$f(t; \boldsymbol{\gamma}|z) = \frac{f(t, z; \boldsymbol{\gamma}, p)}{f(z; p)}, \quad (3)$$

e considerando que T_1 e Z são independentes, obtem-se a função densidade de probabilidade conjunta de (T_1, Z)

$$\begin{aligned} f(t, z; \boldsymbol{\gamma}, p) &= f(t; \boldsymbol{\gamma}|z)P(z; p) \\ &= z[1 - G(t; \boldsymbol{\psi})]^{(z-1)} g(t; \boldsymbol{\psi})(1 - p)p^{(z-1)}. \end{aligned}$$

Assim obtem-se a função densidade de probabilidade marginal da variável aleatória T_1

$$\begin{aligned} f_1(t; \boldsymbol{\gamma}) &= \sum_{z=1}^{\infty} f(t, z; \boldsymbol{\gamma}, p) \\ &= (1 - p)g(t; \boldsymbol{\psi}) \sum_{z=1}^{\infty} z\{[1 - G(t; \boldsymbol{\psi})]p\}^{(z-1)}. \end{aligned}$$

Considerando a expressão $\sum_{z=1}^{\infty} z\{[1 - G(t; \boldsymbol{\psi})]p\}^{(z-1)}$ como uma expansão,

obtem-se finalmente a função densidade marginal de T_1

$$f_1(t; \gamma) = (1 - p)g(t; \psi)\{1 - p[1 - G(t; \psi)]\}^{-2}$$

a qual será denominada como a função distribuição de probabilidade para o mínimo, ou família geométrica tipo I. Neste caso a função de distribuição acumulada de T_1 é dada por

$$F_1(t; \gamma) = G(t; \psi)\{1 - p[1 - G(t; \psi)]\}^{-1}$$

e a função de sobrevivência

$$S_1(t; \gamma) = 1 - F_1(t; \gamma). \quad (4)$$

Note que $G(t; \psi)$ pode assumir qualquer distribuição, por exemplo se $G(t; \psi) = 1 - \exp\left[-\left(\frac{t}{\alpha}\right)^\eta\right]$, então obtemos a distribuição Weibull geométrica, proposto por Barreto-Souza, Morais, Cordeiro (2010).

- Transformação geométrica do tipo II

De maneira análoga a transformação geométrica do tipo I, pode-se obter a transformação geométrica do tipo II, que considera o máximo de uma variável aleatória.

Seja $X_1, X_2, \dots, X_Z$ uma amostra aleatória de uma distribuição contínua qualquer com fdp $g(x; \psi)$ e função de distribuição acumulada $G(x; \psi)$. Neste caso define-se a variável $T_2 = \max(X_1, X_2, \dots, X_Z)$.

Primeiramente obtem-se a função de distribuição acumulada condicional de T_2 dado Z .

$$\begin{aligned} F(t; \gamma|z) &= P(T_2 \leq t|z) \\ &= P[\max(X_1, X_2, \dots, X_Z) \leq t|z] \\ &= P(X_1 \leq t, X_2 \leq t, \dots, X_Z \leq t) \\ &= \prod_{i=1}^z P(X_i \leq t), \end{aligned}$$

em que γ é o vetor de parâmetros da distribuição com fdp $f_2(t)$.

Devido $X_1, X_2, \dots, X_Z$ ser iid a última expressão pode ser reescrita como

$$\begin{aligned} F(t; \gamma|z) &= [P(X \leq t)]^z \\ &= [G(t; \psi)]^z, \end{aligned} \quad (5)$$

em que ψ é o vetor de parâmetros da distribuição com fdp $g(x; \psi)$.

Diferenciando a expressão (5) em relação a t , obtem-se a função densidade de probabilidade de T_2 condicionada a Z .

$$f(t; \gamma|z) = z[G(t; \psi)]^{(z-1)}g(t; \psi).$$

Sabendo (3) e considerando que T_2 e Z são independentes, obtem-se a função densidade de probabilidade conjunta de (T_2, Z)

$$\begin{aligned} f(t, z; \gamma, p) &= f(t; \gamma|z)P(z; p) \\ &= z[G(t; \psi)]^{(z-1)}g(t; \psi)(1-p)p^{(z-1)}. \end{aligned}$$

Assim obtem-se a função densidade de probabilidade marginal da variável aleatória T_2

$$\begin{aligned} f_2(t; \gamma) &= \sum_{z=1}^{\infty} f(t, z; \gamma, p) \\ &= (1-p)g(t; \psi) \sum_{z=1}^{\infty} z\{[G(t; \psi)]p\}^{(z-1)}. \end{aligned}$$

Considerando a expressão $\sum_{z=1}^{\infty} z\{[G(t; \psi)]p\}^{(z-1)}$ como uma expansão de Taylor, obtem-se finalmente a função densidade marginal de T_2

$$f_2(t; \gamma) = (1-p)g(t; \psi)\{1 - p[G(t; \psi)]\}^{-2} \quad (6)$$

a qual será denominada como a função distribuição de probabilidade para o máximo, ou família geométrica tipo II. Neste caso a função distribuição acumulada de T_2 é dada por

$$F_2(t; \gamma) = (1 - p)G(t; \psi)\{1 - p[G(t; \psi)]\}^{-1}.$$

Note que $G(t; \psi)$ pode assumir qualquer distribuição, por exemplo se $G(t; \psi) = 1 - \exp\left[-\left(\frac{t}{\alpha}\right)\right]$, então obtemos a distribuição exponencial geométrica tipo II.

3 A DISTRIBUIÇÃO BXII GEOMÉTRICA

3.1 A distribuição BXIIGI

Uma vez que foram apresentadas a distribuição BXII(s, k, c) e a transformação geométrica do tipo I, é possível definir a fdp da BXIIGI como:

$$f_1(t; s, k, c, p) = \frac{(1-p)ck \left[1 + \left(\frac{t}{s}\right)^c\right]^{-k-1} \frac{t^{c-1}}{s^c}}{\{1 - p \left[1 + \left(\frac{t}{s}\right)^c\right]^{-k}\}^2}, \quad t > 0, \quad (7)$$

em que $s > 0$ é o parâmetro de escala, $k > 0$, $c > 0$ e $p \in [0, 1)$ são os parâmetros de forma. Neste caso dizemos que $T_1 \sim \text{BXIIGI}(s, k, c, p)$.

Essa distribuição tem como casos especiais a distribuição BXII quando $p=0$ e a distribuição log-logística quando $p=0$ e $k=1$.

Figura 2 – Gráficos da fdp da distribuição BXXIGI

A Figura 3 mostra as diversas formas que a fdp da distribuição BXIIGI para algumas combinações de parâmetros.

Calculando a função de sobrevivência, obtém-se:

$$S_1(t; s, k, c, p) = \frac{(1-p) \left[1 + \left(\frac{t}{s}\right)^c\right]^{-k}}{1 - p \left[1 + \left(\frac{t}{s}\right)^c\right]^{-k}}. \quad (8)$$

Figura 3 – Gráficos da função de sobrevivência da distribuição BXXIGI

A Figura 4 mostra as diversas formas da função de sobrevivência BXIIGI para algumas combinações de parâmetros.

Assim a função taxa de falha é dada por:

$$h_1(t; s, k, c, p) = \frac{ck \left[1 + \left(\frac{t}{s}\right)^c\right]^{-1} \frac{t^{c-1}}{s^c}}{1 - p \left[1 + \left(\frac{t}{s}\right)^c\right]^{-k}}.$$

Figura 4 – Gráficos da função taxa de falha da distribuição BXIIGI

A Figura 5 mostra os três tipos de formas da função taxa de falha BXIIGI, crescente, decrescente e unimodal.

3.2 A distribuição BXIIGII

Uma vez que foram apresentadas a distribuição BXII(s,c,k) e a transformação geométrica do tipo II, é possível definir a fdp da BXIIGII como:

$$f_2(t; s, k, c, p) = \frac{(1-p)ck \left[1 + \left(\frac{t}{s}\right)^c\right]^{-k-1} \frac{t^{c-1}}{s^c}}{\{1 - p + p \left[1 + \left(\frac{t}{s}\right)^c\right]^{-k}\}^2}, \quad t > 0, \quad (9)$$

em que $s > 0$ é o parâmetro de escala, $k > 0$, $c > 0$ e $p \in [0, 1)$ são os parâmetros de forma. Neste caso dizemos que $T_2 \sim \text{BXIIGII}(s, k, c, p)$.

Essa distribuição tem como casos especiais a distribuição BXII quando $p=0$ e a distribuição log-logística quando $p=0$ e $k=1$.

Figura 5 – Gráficos da fdp da distribuição BXIIGII

A Figura 6 mostra as diversas formas que a fdp da distribuição BXIIGII para algumas combinações de parâmetros.

Calculando a função de sobrevivência, obtem-se:

$$S_2(t; s, k, c, p) = \frac{\left[1 + \left(\frac{t}{s}\right)^c\right]^{-k}}{1 - p + p \left[1 + \left(\frac{t}{s}\right)^c\right]^{-k}}. \quad (10)$$

Figura 6 – Gráficos da função de sobrevivência da distribuição BXIIGII

A Figura 7 mostra as diversas formas da função de sobrevivência BXIIGI para algumas combinações de parâmetros.

Assim a função taxa de falha é dada por:

$$h_2(t; s, k, c, p) = \frac{(1-p)ck \left[1 + \left(\frac{t}{s}\right)^c\right]^{-1} \frac{t^{c-1}}{s^c}}{1-p + p \left[1 + \left(\frac{t}{s}\right)^c\right]^{-k}}.$$

Figura 7 – Gráficos da função taxa de falha da distribuição BXIIGII

A Figura 8 mostra os três tipo de formas da função taxa de falha BXIIGI, crescente, decrescente e unimodal.

3.3 Expansão da distribuição BXII Geométrica como combinação linear da distribuição BXII

A expansão da distribuição BXII Geométrica como combinação linear da distribuição BXII é útil para desenvolver propriedades importantes, tais como momentos, função geradora de momentos e desvio médio.

- Expansão da distribuição BXIIGI

Utilizando da expansão

$$(1 - z)^{-\rho} = \sum_{j=0}^{\infty} \frac{\Gamma(\rho + j)}{\Gamma(\rho)j!} z^j, \quad (11)$$

para $\rho > 0$ e $|z| < 1$ obtém-se a igualdade

$$\begin{aligned} f_1(t; s, k, c, p) &= \frac{(1-p)ck \left[1 + \left(\frac{t}{s}\right)^c\right]^{-k-1} \frac{t^{c-1}}{s^c}}{\left\{1 - p \left[1 + \left(\frac{t}{s}\right)^c\right]^{-k}\right\}^2} \\ &= (1-p)ck \left[1 + \left(\frac{t}{s}\right)^c\right]^{-k-1} \frac{t^{c-1}}{s^c} \sum_{j=0}^{\infty} \frac{\Gamma(2+j)}{\Gamma(2)j!} p^j \left[1 + \left(\frac{t}{s}\right)^c\right]^{-kj} \\ &= \sum_{j=0}^{\infty} (1-p)p^j ck(j+1) \frac{t^{c-1}}{s^c} \left[1 + \left(\frac{t}{s}\right)^c\right]^{-k(j+1)-1} \\ &= \sum_{j=0}^{\infty} w_j ck(j+1) \frac{t^{c-1}}{s^c} \left[1 + \left(\frac{t}{s}\right)^c\right]^{-k(j+1)-1} \\ &= \sum_{j=0}^{\infty} w_j f[t; s, k(j+1), c]. \end{aligned}$$

Assim foi mostrado que a distribuição BXIIGI(s,k,c,p) é uma combinação linear da distribuição BXII(s,k(j+1),c) com $w_j = (1-p)p^j$.

- Expansão da distribuição BXIIGII

Através da expansão

$$(1 - c(1 - z))^{-2} = \sum_{j=0}^{\infty} (j+1)c^j(1-c)^{-(j+2)}z^j(-1)^j, \quad (12)$$

para $0 < z < 1$ e $0 < c < 1$, obtém-se a igualdade

$$\begin{aligned} f_2(t; s, k, c, p) &= \frac{(1-p)ck \left[1 + \left(\frac{t}{s}\right)^c\right]^{-k-1} \frac{t^{c-1}}{s^c}}{\left[1 - p + p \left[1 + \left(\frac{t}{s}\right)^c\right]^{-k}\right]^2} \\ &= (1-p)ck \left[1 + \left(\frac{t}{s}\right)^c\right]^{-k-1} \frac{t^{c-1}}{s^c} \sum_{j=0}^{\infty} (j+1)p^j(1-p)^{-(j+2)}(-1)^j \\ &\quad \times \left[1 + \left(\frac{t}{s}\right)^c\right]^{-kj} \\ &= \sum_{j=0}^{\infty} (1-p)^{-(j+1)}p^j(-1)^j ck(j+1) \frac{t^{c-1}}{s^c} \left[1 + \left(\frac{t}{s}\right)^c\right]^{-k(j+1)-1} \\ &= \sum_{j=0}^{\infty} v_j ck(j+1) \frac{t^{c-1}}{s^c} \left[1 + \left(\frac{t}{s}\right)^c\right]^{-k(j+1)-1} \\ &= \sum_{j=0}^{\infty} v_j f[t; s, k(j+1), c]. \end{aligned}$$

Assim foi mostrado que a distribuição $BXIIGII(s, k, c, p)$ é uma combinação linear da distribuição $BXII(s, k(j+1), c)$ com $v_j = (1-p)^{-(j+1)}p^j(-1)^j$.

Esses resultados são muito importantes pois pode-se utilizar as propriedades da distribuição $BXII$ para calcular as propriedades da distribuições $BXIIGI$ e $BXIIGII$.

3.4 Momentos

É importante enfatizar a necessidade do cálculo dos momentos em qualquer análise estatística, especialmente em trabalhos aplicados. Algumas das mais importantes propriedades e características de uma distribuição pode ser estudado através dos momentos como tendência, dispersão, assimetria e curtose.

- Momentos da distribuição $BXIIGI$

Se T_1 segue uma distribuição BXIIGI então seus momentos podem ser escritos como

$$\begin{aligned}
 \mu'_r &= \int_0^\infty t^r \sum_{j=0}^\infty w_j f[t; s, k(j+1), c] dt \\
 &= \sum_{j=0}^\infty w_j \int_0^\infty t^r f[t; s, k(j+1), c] dt \\
 &= \sum_{j=0}^\infty w_j s^r k(j+1) B \left[\frac{r}{c} + 1, k(j+1) - \frac{r}{c} \right]
 \end{aligned}$$

se $ck(j+1) > r$ em que $B[.,.]$ é a função Beta.

- Momentos da distribuição BXIIGII

Se T_2 segue uma distribuição BXIIGII então seus momentos podem ser escritos como

$$\begin{aligned}
 \mu'_r &= \int_0^\infty t^r \sum_{j=0}^\infty v_j f[t; s, k(j+1), c] dt \\
 &= \sum_{j=0}^\infty v_j \int_0^\infty t^r f[t; s, k(j+1), c] dt \\
 &= \sum_{j=0}^\infty v_j s^r k(j+1) B \left[\frac{r}{c} + 1, k(j+1) - \frac{r}{c} \right]
 \end{aligned}$$

se $ck(j+1) > r$ em que $B[.,.]$ é a função Beta.

3.5 Assimetria e Curtose

Assim, utilizando os resultados apresentados na seção anterior, pode-se calcular as medidas de assimetria e curtose.

A assimetria mede o grau de afastamento que uma distribuição apresenta do seu eixo de simetria (média). Este afastamento pode acontecer do lado esquerdo ou do lado

direito da distribuição, chamado de assimetria negativa ou positiva respectivamente. Quando a moda, mediana e média de uma distribuição são iguais esta distribuição é dita simétrica. Segundo Nadarajh e Kotz (2004), essa medida pode ser definida por

$$A(T) = \frac{E[T^3] - 3E[T]E[T^2] + 2E^3[T]}{Var^{3/2}[T]}. \quad (13)$$

Já a curtose é o grau de achatamento da distribuição em relação à curva normal padrão. Quanto a curtose a distribuição pode ser classificada como: mesocúrtica, quando a curva da distribuição é semelhante a curva normal, nem achatada nem alongada; platicúrtica, quando a curva da distribuição é achatada em relação à curva normal e leptocúrtica, quando a curva da distribuição é alongada em relação à curva normal. Ainda Segundo Nadarajh e Kotz (2004), essa medida pode ser definida por

$$K(T) = \frac{E[T^4] - 4E[T]E[T^3] + 6E[T^2]E^2[T] - 3E^4[T]}{Var^2[T]}. \quad (14)$$

Portanto, as medidas de assimetria e curtose para as distribuições BXIIGI e BXIIGII podem ser calculadas a partir de (13) e (14).

Figura 8 – Gráficos de assimetria e curtose da distribuição BXIIGI (p variando de 0 a 1)

Figura 9 – Gráficos de assimetria e curtose da distribuição BXIIGII (p variando de 0 a 1)

As Figuras 9 e 10 mostram a assimetria e curtose das distribuições BXIIGI e BXIIGII respectivamente. Note que para p variando entre 0 e 1 as duas curvas se diferenciam relativamente.

3.6 Função Geradora de Momentos

A função geradora de momentos de uma variável aleatória T é dada por:

$$M_T(a) = E(e^{aT}),$$

desde que a esperança seja finita para a real em algum intervalo $-a_0 < a < a_0$ com $a_0 > 0$.

A função geradora de momentos tem esse nome porque a partir dela é possível encontrar todos os momentos da variável aleatória T , o qual é obtido por sucessivas derivadas, seguida da atribuição do valor de a igual a zero.

- Função Geradora de Momentos da distribuição BXIIGI

Através da expansão da função distribuição BXXIGI obtem-se:

$$\begin{aligned}
M_{T_1}(a) &= \int_0^\infty e^{at} \sum_{j=0}^\infty w_j f[t; s, k(j+1), c] dt \\
&= \sum_{j=0}^\infty w_j \int_0^\infty e^{at} f[t; s, k(j+1), c] dt \\
&= \sum_{j=0}^\infty w_j M_{(j+1)}(a),
\end{aligned}$$

onde $M_{j+1}(a)$ é a função geradora de momentos da distribuição BXII($s, k(j+1), c$) .

Por outro lado agora será calculado a função geradora de momentos da distribuição BXII que pode ser obtida com auxílio da função Meijer-G. Tem-se que

$$\begin{aligned}
M_{BXII}(\tau) &= \int_0^\infty e^{\tau x} ck \left[1 + \left(\frac{x}{s} \right)^c \right]^{-k-1} \frac{x^{c-1}}{s^c} dx \\
&= \int_0^\infty e^{\tau x} ck \left[1 + \left(\frac{x}{s} \right)^c \right]^{-k-1} \left(\frac{x}{s} \right)^{c-1} \frac{1}{s} dx.
\end{aligned}$$

Considere a transformação $y = \frac{x}{s}$, então $dy = \frac{1}{s} dx$.

$$\begin{aligned}
M_{BXII}(\tau) &= \int_0^\infty e^{\tau y s} ck (1 + y^c)^{-k-1} y^{c-1} dy \\
&= ck \int_0^\infty e^{\tau y s} (1 + y^c)^{-k-1} y^{c-1} dy
\end{aligned}$$

A função Meijer-G é dada por:

$$G_{p,q}^{m,n} \left(x \left| \begin{matrix} a_1 & \cdots & a_p \\ b_1 & \cdots & b_q \end{matrix} \right. \right) = \frac{1}{2\pi i} \int_L \frac{\prod_{j=1}^m \Gamma(b_j + \tau) \prod_{j=1}^n \Gamma(1 - a_j - \tau)}{\prod_{j=n+1}^p \Gamma(a_j + \tau) \prod_{j=m+1}^q \Gamma(1 - b_j - \tau)} x^{-\tau} d\tau$$

onde $i = \sqrt{-1}$ é a unidade complexa e L o limite de integração. Assumindo que $c = \frac{m}{k}$, onde m e k são inteiros e positivos. Essa condição não é restritiva, uma vez, que todo número real e positivo pode ser aproximado por um número racional.

$$I'(p, u, \frac{m}{k}, v) = \int_0^\infty e^{-px} x^u (1 + x^{\frac{m}{k}})^v dx$$

$$I'(p, u, \frac{m}{k}, v) = \frac{k^{-v} m^{u+\frac{1}{2}}}{2\pi^{\frac{(m-1)}{2}} \Gamma(-v) p^{u+1}} G_{k+m, k}^{k, k+m} \left(\frac{m^m}{p^m} \middle| \begin{matrix} \Delta(m, -u), & \Delta(k, v+1) \\ \Delta(k, 0) \end{matrix} \right)$$

Considere também $u > -1$, $p > 0$ e $\Delta(k, a) = \frac{a}{k}, \frac{a+1}{k}, \dots, \frac{a+k}{k}$.

Então a função geradora de momentos da distribuição BXII é dada por

$$M_{BXII}(\tau) = mI'(-s\tau, \frac{m}{k} - 1, \frac{m}{k}, -k - 1). \quad (15)$$

Dessa maneira a função geradora de momentos da distribuição BXIIGI é dada por

$$M_{T_1}(a) = \sum_{j=0}^{\infty} w_j mI' \left[-sa, \frac{m}{k(j+1)} - 1, \frac{m}{k(k+j)}, -k(j+1) - 1 \right].$$

- Função Geradora de Momentos da distribuição BXIIGII

Através da expansão da função distribuição BXIIGII obtem-se:

$$\begin{aligned} M_{T_2}(a) &= \int_0^\infty e^{at} \sum_{j=0}^{\infty} v_j f[t; s, k(j+1), c] dt \\ &= \sum_{j=0}^{\infty} v_j \int_0^\infty e^{at} f[t; s, k(j+1), c] dt \\ &= \sum_{j=0}^{\infty} v_j M_{(j+1)}(a), \end{aligned}$$

onde $M_{j+1}(a)$ é a função geradora de momentos da distribuição BXII $(s, k(j+1), c)$ que foi

obtida em (15)

Dessa maneira a função geradora de momentos da distribuição BXIIGII é dada por

$$M_{T_2}(a) = \sum_{j=0}^{\infty} v_j m I' \left[-sa, \frac{m}{k(j+1)} - 1, \frac{m}{k(k+j)}, -k(j+1) - 1 \right].$$

3.7 Desvio Médio

A dispersão em uma determinada população pode ser medida pelo desvio médio em relação a média e a mediana. Esses desvios são dados respectivamente por:

$$\delta_1(X) = \int_0^{\infty} |x - \mu| f(x) dx \quad \text{e} \quad \delta_2(X) = \int_0^{\infty} |x - M| f(x) dx$$

em que $\mu'_1 = E(X)$ e $M = \text{Mediana}(X)$ denota a mediana. Defina $T(q) = \int_q^{\infty} x f(x) dx$. Então tem-se as seguintes igualdades:

$$\delta_1(X) = 2\mu'_1 F(\mu'_1) - 2\mu'_1 + 2T(\mu'_1) \quad (16)$$

e

$$\delta_2(X) = 2T(\mu'_1) - \mu'_1, \quad (17)$$

em que $F(M)$ e $F(\mu'_1)$ são funções acumuladas no ponto M e μ'_1 respectivamente.

A igualdade (16) é obtida da seguinte maneira,

$$\begin{aligned} \delta_1(X) &= \int_0^{\infty} |x - \mu'_1| f(x) dx \\ &= \int_{\mu'_1}^{\infty} (x - \mu'_1) f(x) dx - \int_0^{\mu'_1} (x - \mu'_1) f(x) dx \end{aligned}$$

uma vez que $|x - \mu'_1| = (x - \mu'_1)$ se $x > \mu'_1$ e $|x - \mu'_1| = -(x - \mu'_1)$ se $x < \mu'_1$.

Então,

$$\delta_1(X) = \int_{\mu'_1}^{\infty} x f(x) dx - \mu'_1 \int_{\mu'_1}^{\infty} f(x) dx - \int_0^{\mu'_1} x f(x) dx + \mu'_1 \int_0^{\mu'_1} f(x) dx.$$

Considere

$$\int_0^{\mu'_1} x f(x) dx + \int_{\mu'_1}^{\infty} x f(x) dx = \mu'_1,$$

então

$$\int_0^{\mu'_1} x f(x) dx = \mu'_1 - T(\mu).$$

Dessa forma,

$$\begin{aligned} \delta_1(X) &= T(\mu'_1) - \mu'_1(1 - F(\mu'_1)) + T(\mu'_1) - \mu'_1 + \mu'_1 F(\mu'_1) \\ &= 2T(\mu'_1) - 2\mu'_1 + 2\mu'_1 F(\mu'_1). \end{aligned}$$

Já para $\delta_2(X)$,

$$\begin{aligned} \delta_2(X) &= \int_0^{\infty} |x - M| f(x) dx \\ &= \int_M^{\infty} (x - M) f(x) dx - \int_0^M (x - M) f(x) dx \end{aligned}$$

uma vez que $|x - M| = (x - M)$ se $x > M$ e $|x - M| = -(x - M)$ se $x < M$.

Então,

$$\delta_2(X) = \int_M^\infty x f(x) dx - M \int_M^\infty f(x) dx - \int_0^M x f(x) dx + M \int_0^M f(x) dx.$$

Utilizando a mesma consideração feita para $\delta_1(X)$, obtem-se

$$\begin{aligned} \delta_2(X) &= T(M) - M(1 - F(M)) + T(M) - \mu'_1 + MF(M) \\ &= 2T(M) - M + 2MF(M) - \mu'_1 \\ &= 2T(M) - M + 2M\frac{1}{2} - \mu'_1 = 2T(M) - \mu'_1. \end{aligned}$$

- Desvio Médio da distribuição BXIIGI

Considere os desvios médios dados em (16) e (17).

A função $T_1(q)$ é definida por:

$$\begin{aligned} T_1(q) &= \int_q^\infty t f_1(t) dt \\ &= \int_q^\infty t \sum_{j=0}^\infty w_j f(t; s, k(j+1); c) dt \\ &= \int_q^\infty t \sum_{j=0}^\infty w_j c k(j+1) \left[1 + \left(\frac{t}{s} \right)^c \right]^{-k(j+1)-1} \left(\frac{t}{s} \right)^{c-1} \frac{1}{s} dt. \end{aligned}$$

Utilizando a transformação $y = \frac{t}{s}$, $dy = \frac{1}{s} dt$, obtem-se

$$\begin{aligned} T_1(q) &= \int_{\frac{q}{s}}^\infty y s \sum_{j=0}^\infty w_j c k(j+1) (1 + y^c)^{-k(j+1)-1} y^{c-1} dy \\ &= c k s \sum_{j=0}^\infty w_j (j+1) \int_{\frac{q}{s}}^\infty y^c (1 + y^c)^{-k(j+1)-1} dy. \end{aligned} \tag{18}$$

Através da transformação $u = y^c$ $du = cy^{c-1}dy$, obtem-se a solução da integral (18).

$$T_1(q) = ks \sum_{j=0}^{\infty} w_j \int_{(\frac{q}{s})^c}^{\infty} u^{\frac{1}{c}} (1+u)^{-k(j+1)-1} du.$$

- Desvio Médio da distribuição BXIIGII

Considere os desvios médios dados em (16) e (17).

A função $T_2(q)$ é definida por:

$$\begin{aligned} T_2(q) &= \int_q^{\infty} t f(t) dt \\ &= \int_q^{\infty} t \sum_{j=0}^{\infty} v_j f[t; s, k(j+1); c] dt \\ &= \int_q^{\infty} t \sum_{j=0}^{\infty} v_j ck(j+1) \left[1 + \left(\frac{t}{s} \right)^c \right]^{-k(j+1)-1} \left(\frac{t}{s} \right)^{c-1} \frac{1}{s} dt. \end{aligned}$$

Utilizando a transformação $y = \frac{t}{s}$, $dy = \frac{1}{s} dt$, obtem-se

$$\begin{aligned} T_2(q) &= \int_{\frac{q}{s}}^{\infty} ys \sum_{j=0}^{\infty} v_j ck(j+1) (1+y^c)^{-k(j+1)-1} y^{c-1} dy \\ &= cks \sum_{j=0}^{\infty} v_j (j+1) \int_{\frac{q}{s}}^{\infty} y^c (1+y^c)^{-k(j+1)-1} dy. \end{aligned} \quad (19)$$

Através da transformação $u = y^c$ $du = cy^{c-1}dy$, obtem-se a solução da integral (19).

$$T_2(q) = ks \sum_{j=0}^{\infty} v_j \int_{(\frac{q}{s})^c}^{\infty} u^{\frac{1}{c}} (1+u)^{-k(j+1)-1} du.$$

3.8 Função quantílica

A função quantílica, também conhecida como cumulativa inversa, determina qual o valor máximo que a variável aleatória deve assumir tal que a probabilidade de ser menor ou igual a esse valor seja u . Deste modo a função quantílica é obtida isolando a variável aleatória X da função densidade acumulada

$$F(x; \boldsymbol{\lambda}) = u,$$

em que $F(x; \boldsymbol{\lambda})$ é a função densidade acumulada de X , $\boldsymbol{\lambda}$ é o vetor de parâmetros da distribuição de X e $u \in (0, 1)$ é a probabilidade de X ser menor ou igual a um dado valor.

A função quantílica é dada por

$$Q(u, \boldsymbol{\lambda}) = [F(x; \boldsymbol{\lambda})]^{-1}, \quad (20)$$

A função quantílica permite o cálculo de assimetria e curtose como demonstrado por Kenney e Keeping (1962) e Moors (1988), além de permitir o cálculo dos momentos.

Considerando $u = F(x)$

$$\begin{aligned} \eta'_s &= \int_{-\infty}^{\infty} x^s f(x) dx \\ &= \int_0^1 Q(u)^s du. \end{aligned}$$

A função quantílica também é importante para geração de dados de uma distribuição. Gerando valores aleatórios para u e substituindo em (20) é possível encontrar os valores simulados de x_i .

- Função quantílica da distribuição BXIIGI

A função quantílica da distribuição BXIIGI é definida por

$$Q_1(u) = F_1(u)^{-1} = s \left[\left(\frac{1-u}{1-pu} \right)^{\frac{-1}{k}} - 1 \right]^{\frac{1}{c}}.$$

Utilizando a expansão binomial generalizada para qualquer real a e $z \in (0, 1)$:

$$(1 - z)^a = \sum_{j=0}^{\infty} (-1)^j \binom{a}{j} z^j. \quad (21)$$

Definido $L_1 = \frac{1-u}{1-pu}$, que pertence ao intervalo $(0, 1)$, obtem-se a expansão de $L_1^{-1/k}$ usando (21),

$$L_1^{-1/k} = [1 - (1 - L_1)]^{-1/k} = \sum_{j=0}^{\infty} (-1)^j \binom{-1/k}{j} (1 - L_1)^j$$

e então

$$L_1^{-1/k} = \sum_{j=0}^{\infty} \sum_{r=0}^j (-1)^{j+r} \binom{-1/k}{j} \binom{j}{r} L_1^r.$$

Substituindo $\sum_{j=0}^{\infty} \sum_{r=0}^j$ por $\sum_{r=0}^{\infty} \sum_{j=r}^{\infty}$ obtem-se

$$L_1^{-1/k} = \sum_{r=0}^{\infty} \sum_{j=r}^{\infty} (-1)^{j+r} \binom{-1/k}{j} \binom{j}{r} L_1^r.$$

e então

$$L_1^{-1/k} = \sum_{r=0}^{\infty} s_r(-1/k) L_1^r,$$

em que

$$s_r(a) = \sum_{j=r}^{\infty} (-1)^{r+j} \binom{a}{j} \binom{j}{r}.$$

Então tem-se

$$\left(\frac{1-u}{1-pu} \right)^{-1/k} = \sum_{r=0}^{\infty} s_r(-1/k) \left(\frac{1-u}{1-pu} \right)^r.$$

Além disso, utilizando (21) e $(1 - u)^{-s} = \sum_{m=0}^{\infty} \binom{s+m-1}{m} u^m$, tem-se

$$\left(\frac{1-u}{1-pu}\right)^{-1/k} = \sum_{r,m=0}^{\infty} s_r(-1/k) \sum_{j=0}^r (-1)^j \binom{r}{j} \binom{r+m-1}{m} u^{j+m} p^m.$$

Substituindo $\sum_{r=0}^{\infty} \sum_{j=0}^r$ por $\sum_{j=0}^{\infty} \sum_{r=j}^{\infty}$, reescreve-se

$$\left(\frac{1-u}{1-pu}\right)^{-1/k} = \sum_{j,m=0}^{\infty} v_{1j,m} u^{j+m},$$

onde $v_{1j,m} = (-1)^j \sum_{r=j}^{\infty} s_r(-1/k) \binom{r}{j} \binom{r+m-1}{m} p^m$.

Para $k = 0, 1, \dots$, define-se o coeficiente w_{1k} por

$$w_{1k} = \sum_{\substack{m,j=0 \\ j+m=k}}^{\infty} v_{1j,m}$$

e então obtem-se

$$L_1^{-1/k} = \left(\frac{1-u}{1-pu}\right)^{-1/k} = \sum_{k=0}^{\infty} w_{1k} u^k.$$

Expandindo x^c e série de Taylor obtem-se

$$x^d = \sum_{k=0}^{\infty} \frac{(d)_k (x-1)^k}{k!} = \sum_{j=0}^{\infty} f_j x^j, \quad (22)$$

onde $f_j = \sum_{k=j}^{\infty} (-1)^{k-j} \binom{k}{j} (d)_k / k!$ e $(d)_k = d(d-1) \dots (d-k+1)$ é o fatorial descendente.

Obtem-se uma expansão para $(L_1^{-1/k} - 1)^{1/c}$ através de (22)

$$(L_1^{-1/k} - 1)^{1/c} = \sum_{j=0}^{\infty} g_j \sum_{m=0}^j (-1)^{j-m} \binom{j}{m} \left(\sum_{k=0}^{\infty} w_{1k} u^k \right)^m, \quad (23)$$

onde $g_j = \sum_{k=j}^{\infty} (-1)^{k-j} \binom{k}{j} (1/c)_k / k!$.

Agora, através de uma equação na seção 0.314 de Gradshteyn and Ryzhik (2007) para séries de potência para um inteiro positivo j :

$$\left(\sum_{k=0}^{\infty} w_{1k} u^k \right)^m = \sum_{i=0}^{\infty} c_{1m,k} u^k, \quad (24)$$

cujos coeficientes $c_{1m,k}$ (for $k = 1, 2, \dots$) são obtidos da equação de recorrência

$$c_{1m,k} = (k w_{10})^{-1} \sum_{\ell=1}^k [\ell(m+1) - k] w_{1\ell} c_{1m,k-\ell} \quad (25)$$

e $c_{1m,0} = w_{10}^j$. O coeficiente $c_{1m,k}$ vem diretamente de $c_{1m,0}, \dots, c_{1m,k-1}$ e, portanto de $w_{10}, \dots, w_{1k}$. De fato, $c_{1m,k}$ pode ser dado em termos dos coeficiente w_{1k} 's, embora isso não seja necessário para progamar numericamente a expansão em qualquer software algébrico ou numérico. Usando (23) e (24) segue que

$$Q_1(u) = s \sum_{k=0}^{\infty} h_{1k} u^k, \quad (26)$$

onde $h_{1k} = s \sum_{j=0}^{\infty} g_j \sum_{m=0}^j (-1)^{j-m} \binom{j}{m} c_{1m,k}$ e $c_{1m,k}$ é determinado através de (25). A equação (26) é o resultado principal desta seção.

- Função quantílica da distribuição BXIIGII

A função quantílica da distribuição BXIIGII é definida por

$$Q_2(u) = F_2(u)^{-1} = s \left\{ \left[\frac{(1-p)(1-u)}{1-p(1-u)} \right]^{\frac{-1}{k}} - 1 \right\}^{\frac{1}{c}}.$$

Definindo $L_2 = \frac{(1-p)(1-u)}{1-p(1-u)}$, que pertence ao intervalo $(0, 1)$, obtem-se a expansão de $L_2^{-1/k}$ através (21),

$$L_2^{-1/k} = [1 - (1 - L_2)]^{-1/k} = \sum_{j=0}^{\infty} (-1)^j \binom{-1/k}{j} (1 - L_2)^j$$

e então

$$L_2^{-1/k} = \sum_{j=0}^{\infty} \sum_{r=0}^j (-1)^{j+r} \binom{-1/k}{j} \binom{j}{r} L_2^r.$$

Substituindo $\sum_{j=0}^{\infty} \sum_{r=0}^j$ por $\sum_{r=0}^{\infty} \sum_{j=r}^{\infty}$ obtem-se

$$L_2^{-1/k} = \sum_{r=0}^{\infty} \sum_{j=r}^{\infty} (-1)^{j+r} \binom{-1/k}{j} \binom{j}{r} L_2^r.$$

e então

$$L_2^{-1/k} = \sum_{r=0}^{\infty} s_r(-1/k) L_2^r.$$

Então tem-se

$$\left[\frac{(1-p)(1-u)}{1-p(1-u)} \right]^{-1/k} = \sum_{r=0}^{\infty} s_r(-1/k) \left[\frac{(1-p)(1-u)}{1-p(1-u)} \right]^r.$$

Além disso, utilizando (21) e $(1-u)^{-s} = \sum_{m=0}^{\infty} \binom{s+m-1}{m} u^m$, tem-se

$$\begin{aligned} \left[\frac{(1-p)(1-u)}{1-p(1-u)} \right]^{-1/k} &= \sum_{r,m=0}^{\infty} s_r(-1/k) \sum_{j=0}^r (-1)^j \binom{r}{j} u^j (1-p)^r \sum_{m=0}^{\infty} \binom{r+m-1}{m} (1-u)^m p^m \\ &= \sum_{r,m=0}^{\infty} s_r(-1/k) \sum_{j=0}^r (-1)^j \binom{r}{j} u^j (1-p)^r \sum_{m=0}^{\infty} \binom{r+m-1}{m} p^m \sum_{n=0}^m (-1)^n \\ &\quad \times \binom{m}{n} u^n. \end{aligned}$$

Substituindo $\sum_{r=0}^{\infty} \sum_{j=0}^r$ por $\sum_{j=0}^{\infty} \sum_{r=j}^{\infty}$ e $\sum_{m=0}^{\infty} \sum_{n=0}^m$ por $\sum_{n=0}^{\infty} \sum_{m=n}^{\infty}$, reescreve-se

$$\left[\frac{(1-p)(1-u)}{1-p(1-u)} \right]^{-1/k} = \sum_{j,n=0}^{\infty} v_{2j,n} u^{j+n},$$

onde $v_{2j,n} = (-1)^{j+n} \sum_{r=j}^{\infty} s_r(-1/k) \binom{r}{j} (1-p)^r \sum_{m=n}^{\infty} \binom{r+m-1}{m} p^m \binom{m}{n}$.

Para $k = 0, 1, \dots$, define-se o coeficiente w_{2k} por

$$w_{2k} = \sum_{\underbrace{j,n=0}_{j+n=k}}^{\infty} v_{2j,n}$$

e então obtem-se

$$L_2^{-1/k} = \left[\frac{(1-p)(1-u)}{1-p(1-u)} \right]^{-1/k} = \sum_{k=0}^{\infty} w_{2k} u^k.$$

Utilizando (22), obtem-se uma expansão para $(L_2^{-1/k} - 1)^{1/c}$

$$(L_2^{-1/k} - 1)^{1/c} = \sum_{j=0}^{\infty} g_j \sum_{m=0}^j (-1)^{j-m} \binom{j}{m} \left(\sum_{k=0}^{\infty} w_{2k} u^k \right)^m, \quad (27)$$

onde $g_j = \sum_{k=j}^{\infty} (-1)^{k-j} \binom{k}{j} (1/c)_k / k!$.

Através de uma equação na seção 0.314 de Gradshteyn and Ryzhik (2007) para séries de potência para um inteiro positivo j :

$$\left(\sum_{k=0}^{\infty} w_{2k} u^k \right)^m = \sum_{i=0}^{\infty} c_{2m,k} u^k, \quad (28)$$

cujos coeficientes $c_{2m,k}$ (for $k = 1, 2, \dots$) são obtidos da equação de recorrência

$$c_{2m,k} = (k w_{20})^{-1} \sum_{\ell=1}^k [\ell(m+1) - k] w_{2\ell} c_{2m,k-\ell} \quad (29)$$

e $c_{2m,0} = w_{20}^j$. O coeficiente $c_{2m,k}$ vem diretamente de $c_{2m,0}, \dots, c_{2m,k-1}$ e, portanto de $w_{20}, \dots, w_{2k}$. Usando (27) e (28) segue que

$$Q_2(u) = s \sum_{k=0}^{\infty} h_{2k} u^k, \quad (30)$$

onde $h_{2k} = s \sum_{j=0}^{\infty} g_j \sum_{m=0}^j (-1)^{j-m} \binom{j}{m} c_{2m,k}$ e $c_{2m,k}$ é determinado através de (29). A equação (30) é o resultado principal desta seção.

3.9 Estimação dos parâmetros

O método de estimação mais utilizado em análise de sobrevivência é o método da máxima verossimilhança, pois permite que a informação dos tempos censurados seja

considerada no processo de estimação, ou seja, a função de verossimilhança leva em conta a contribuição dos indivíduos que foram falhas e os indivíduos que foram censurados (Lawless, J.F., 2003).

Considere $T_1, \dots, T_n$ uma amostra aleatória de uma distribuição de probabilidade, $E_1, \dots, E_n$ os tempos de censura aleatória, $t_i = \min(T_i, E_i)$ os dados observados e $\delta_i = I(T_i \leq E_i)$ um indicador de censura. Dessa forma, assumindo que as censuras são aleatórias e não-informativas, a função de verossimilhança é dada por

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n [f(t_i; \boldsymbol{\theta})]^{\delta_i} [S(t_i; \boldsymbol{\theta})]^{1-\delta_i},$$

em que $\boldsymbol{\theta}$ é o vetor de parâmetros desconhecidos, $f(t_i; \boldsymbol{\theta})$ é a função densidade de probabilidade e $S(t_i; \boldsymbol{\theta})$ é a função de sobrevivência.

Assim, o estimador de máxima verossimilhança $\hat{\boldsymbol{\theta}}$ do vetor de parâmetros $\boldsymbol{\theta}$ é calculado maximizando a função de verossimilhança ou, equivalentemente, maximizando o logaritmo da função de verossimilhança, $l(\boldsymbol{\theta})$. Portanto, esses estimadores são obtidos da solução do sistema de equações:

$$U(\boldsymbol{\theta}) = \frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 0.$$

Geralmente esse sistema é não linear, tornando-se necessário o uso de um algoritmo de otimização, como por exemplo, um algoritmo Newton-Raphson ou quase-Newton (Nocedal, J., Wright, S.J., 2006).

Quando o tamanho da amostra é grande e sob certas condições de regularidade para a função de verossimilhança, intervalos de confiança e testes de hipóteses podem ser obtidos levando em consideração o fato de que a distribuição assintótica para estimadores de máxima verossimilhança é a distribuição normal com média $\boldsymbol{\theta}$ e matriz de covariância dada pelo inverso da matriz de informação de Fisher (Sen, P.K.; Singer, J.M., 1993). Portanto, a matriz de covariância assintótica é dada por $I^{-1}(\boldsymbol{\theta})$ com $I(\boldsymbol{\theta}) = E[\ddot{L}(\boldsymbol{\theta})]$, em que $\ddot{L}(\boldsymbol{\theta}) = -\left\{ \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right\}$.

Devido a complexidade de calcular a matriz de informação de Fisher $I(\boldsymbol{\theta})$ por causa das informações censuradas, pode-se recorrer a matriz hessiana $[-\ddot{L}(\boldsymbol{\theta})]^{-1}$, avaliada em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, que é um estimador consistente para a matriz de covariância assintótica

(Mudholkar, G.S., Srivastava, D.K, Kollia, G.D., 1996). Assim a distribuição assintótica para $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ é dada por $\hat{\boldsymbol{\theta}}^T \sim N_{(q)}(\hat{\boldsymbol{\theta}}^T; \ddot{L}(\boldsymbol{\theta})^{-1})$, em que $\ddot{L}(\boldsymbol{\theta})$ é a matriz de informação observada e q é o número de parâmetros do modelo.

Para testar a adequabilidade de um modelo pode-se utilizar o teste da razão de verossimilhança. As hipóteses de interesse são:

$$H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0 \text{ versus } H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0,$$

em que $\boldsymbol{\theta}_0$ é o vetor de parâmetros a ser testado. O teste da razão de verossimilhança consiste em comparar os valores do logaritmo da função de verossimilhança do modelo geral e do modelo reduzido sob a hipótese nula H_0 . A estatística da razão de verossimilhança é dada por

$$RV = 2[l(\boldsymbol{\theta}) - l(\boldsymbol{\theta}_0)],$$

sendo $l(\boldsymbol{\theta})$ o logartmo da função de verossimilhança do modelo geral e $l(\boldsymbol{\theta}_0)$ o logaritmo da função de verossimilhança para o modelo reduzido. A estatística RV sob H_0 tem aproximadamente uma distribuição qui-quadrado com m graus de liberdade, que corresponde a diferença do número de parâmetros dos modelos a serem comparados. Rejeita-se a hipótese nula (H_0) a um nível $100\%\alpha$ de significância quando $RV > \chi_{1-\alpha}^2(m)$, que é o quantil $(1 - \alpha)$ da distribuição qui-quadrado com m graus de liberdade.

Por exemplo, para verificar a adequabilidade dos modelos utilizando as distribuições BXII e BXIIGI para um determinado conjunto de dados é equivalente a testar a hipótese $H_0 : p = 0$ versus $H_1 : p \neq 0$. Então a estatística da razão de verossimilhança reduz-se a

$$RV = 2[l(\hat{s}; \hat{k}; \hat{c}; \hat{p}) - l(\tilde{s}; \tilde{k}; \tilde{c}; 0)],$$

em que $\hat{s}, \hat{k}, \hat{c}$ e $\hat{p}$ são os estimadores de máxima verossimilhança do modelo geral e $\tilde{s}, \tilde{k}$ e $\tilde{c}$ são os estimadores de máxima verossimilhança do modelo reduzido calculado sob H_0 ; ou seja, com $p = 0$.

Outros critérios de seleção baseados no máximo da função verossimilhança

são o critério de informação de Akaike (AIC), o critério de informação bayesiano (BIC) e o critério de informação de Akaike corrigido (CAIC). Estas estatísticas consideram também q o número de parâmetros considerados no modelo.

$$\begin{aligned} AIC &= -2l(\hat{\boldsymbol{\theta}}) + 2q \\ BIC &= -2l(\hat{\boldsymbol{\theta}}) + q \log(n) \\ CAIC &= AIC + \frac{2q(q+1)}{n-q-1} \end{aligned}$$

Valores menores de AIC, BIC e CAIC indicam modelos mais apropriados.

Feita todas as considerações acima, seguem as funções de verossimilhança e as primeiras derivadas para as distribuições BXIIGI e BXIIGII.

- Estimador de Máxima Verossimilhança para a distribuição BXIIGI

Seja $(t_1, \delta_1), \dots, (t_n, \delta_n)$ um conjunto de dados de tamanho n , em que t_i é a variável resposta e δ_i é o indicador de censura, para $i = 1, \dots, n$. Então, o logartimo da função de verossimilhança para dados censurados é obtido considerando a fdp (7) e a função de sobrevivência (8)

$$\begin{aligned} l(\boldsymbol{\theta}) &= r^* \log[(1-p)ck] - (k+1) \sum_{i \in F} \log(u_i) + (c-1) \sum_{i \in F} \log(t_i) \\ &- r^* c \log(s) - 2 \sum_{i \in F} \log(1 - pu_i^{-k}) + (n - r^*) \log(1-p) \\ &- k \sum_{i \in E} \log(u_i) - \sum_{i \in E} \log(1 - pu_i^{-k}), \end{aligned}$$

onde r^* é o número de falhas, F indica o conjunto de falhas, E o conjunto de observações censuradas e $u_i = \left[1 + \left(\frac{t_i}{s}\right)^c\right]$.

Então o estimador de máxima verossimilhança da distribuição BXIIGI é a solução iterativa do seguinte sistema de equações não lineares

$$\frac{\partial l(\boldsymbol{\theta})}{\partial c} = \frac{\partial l(\boldsymbol{\theta})}{\partial k} = \frac{\partial l(\boldsymbol{\theta})}{\partial s} = \frac{\partial l(\boldsymbol{\theta})}{\partial p} = 0,$$

onde

$$\begin{aligned}
\frac{\partial l(\boldsymbol{\theta})}{\partial c} &= \frac{r^*}{c} - (k+1) \sum_{i \in F} \frac{\log(u_i - 1)(u_i - 1)^c}{u_i} + \sum_{i \in F} \log(t_i) - r^* \log(s) \\
&\quad - 2 \sum_{i \in F} \frac{k p u_i^{-k-1} \log(u_i - 1)(u_i - 1)^c}{1 - p u_i^{-k}} - k \sum_{i \in E} \frac{\log(u_i - 1)(u_i - 1)^c}{u_i} \\
&\quad - \sum_{i \in E} \frac{k p u_i^{-k-1} \log(u_i - 1)(u_i - 1)^c}{1 - p u_i^{-k}} \\
\frac{\partial l(\boldsymbol{\theta})}{\partial k} &= \frac{r^*}{k} - \sum_{i \in F} \log(u_i) + 2 \sum_{i \in F} \frac{p \log(u_i) u_i^{-k}}{1 - p u_i^{-k}} + \sum_{i \in E} \frac{p \log(u_i) u_i^{-k}}{1 - p u_i^{-k}} - \sum_{i \in E} \log(u_i) \\
\frac{\partial l(\boldsymbol{\theta})}{\partial s} &= -\frac{r^* c}{s} + (k+1) \sum_{i \in F} \frac{c t_i^c}{s^{c+1} u_i} + k \sum_{i \in E} \frac{c t_i^c}{s^{c+1} u_i} + 2 \sum_{i \in F} \frac{k p c u_i^{-k-1} t_i^c}{s^{c+1} (1 - p u_i^{-k})} \\
&\quad + \sum_{i \in E} \frac{k p c u_i^{-k-1} t_i^c}{s^{c+1} (1 - p u_i^{-k})} \\
\frac{\partial l(\boldsymbol{\theta})}{\partial p} &= -\frac{r^*}{(1-p)} + 2 \sum_{i \in F} \frac{u_i^{-k}}{1 - p u_i^{-k}} - \frac{n - r^*}{1-p} + \sum_{i \in E} \frac{u_i^{-k}}{1 - p u_i^{-k}}.
\end{aligned}$$

- Estimador de Máxima Verossimilhança para a distribuição BXIIGII

Analogamente obtém-se o logaritmo da função de verossimilhança da distribuição BXIIGII. Considere a fdp dada em (9) e a função de sobrevivência dada em (10).

$$\begin{aligned}
l(\boldsymbol{\theta}) &= r^* \log[(1-p)ck] - (k+1) \sum_{i \in F} \log(u_i) + (c-1) \sum_{i \in F} \log(t_i) \\
&\quad - r^* c \log(s) - 2 \sum_{i \in F} \log(1 - p + p u_i^{-k}) + (n - r^*) \log(1 - p) \\
&\quad - k \sum_{i \in E} \log(u_i) - \sum_{i \in E} \log(1 - p + p u_i^{-k}),
\end{aligned}$$

r , F e E foram definidas no logaritmo da função de verossimilhança da distribuição BXIIGI.

Então o estimador de máxima verossimilhança da distribuição BXIIGII é a solução iterativa do seguinte sistema de equações não lineares, onde

$$\begin{aligned}
\frac{\partial l(\boldsymbol{\theta})}{\partial c} &= \frac{r^*}{c} - (k+1) \sum_{i \in F} \frac{\log(u_i - 1)(u_i - 1)^c}{u_i} + \sum_{i \in F} \log(t_i) - r^* \log(s) \\
&+ 2 \sum_{i \in F} \frac{kp \log(u_i - 1)(u_i - 1)^c u_i^{-k-1}}{1 - p + pu_i^{-k}} - k \sum_{i \in E} \frac{\log(u_i - 1)(u_i - 1)^c}{u_i} \\
&+ \sum_{i \in E} \frac{kp \log(u_i - 1)(u_i - 1)^c u_i^{-k-1}}{1 - p + pu_i^{-k}} \\
\\
\frac{\partial l(\boldsymbol{\theta})}{\partial k} &= \frac{r^*}{k} - \sum_{i \in F} \log(u_i) - 2 \sum_{i \in F} \frac{p \log(u_i) u_i^{-k}}{1 - p + pu_i^{-k}} - \sum_{i \in E} \frac{p \log(u_i) u_i^{-k}}{1 - p + pu_i^{-k}} - \sum_{i \in E} \log(u_i) \\
\\
\frac{\partial l(\boldsymbol{\theta})}{\partial s} &= -\frac{r^* c}{s} + (k+1) \sum_{i \in F} \frac{ct_i^c}{s^{c+1} u_i} + k \sum_{i \in E} \frac{ct_i^c}{s^{c+1} u_i} - 2 \sum_{i \in F} \frac{kpcu_i^{-k-1} t_i^c}{s^{c+1} (1 - p + pu_i^{-k})} \\
&- \sum_{i \in E} \frac{kpcu_i^{-k-1} t_i^c}{s^{c+1} (1 - p + pu_i^{-k})} \\
\\
\frac{\partial l(\boldsymbol{\theta})}{\partial p} &= -\frac{r^*}{(1-p)} - 2 \sum_{i \in F} \frac{-1 + u_i^{-k}}{1 - p + pu_i^{-k}} - \frac{n - r^*}{1 - p} - \sum_{i \in E} \frac{-1 + u_i^{-k}}{1 - p + pu_i^{-k}}.
\end{aligned}$$

Neste trabalho foi utilizado o procedimento NLMixed do SAS para calcular os estimadores de máxima verossimilhança (env) e também considerou a distribuição assintótica dos env para calcular os intervalos de confiança e testes de hipóteses.

4 MODELOS DE REGRESSÃO PARAMÉTRICOS

Além dos tempos de sobrevivência e da variável indicadora de censura é comum em muitas situações serem observadas covariáveis que representam a heterogeneidade de uma população.

O efeito dessas covariáveis no tempo de sobrevivência pode ser estudado através de um modelo de regressão. Existem duas metodologias de modelos de regressão em análise de sobrevivência: os modelos de riscos proporcionais conhecidos também como modelo de regressão de Cox e os modelos de tempo de vida acelerado ou modelos de locação escala. Dentre a literatura sobre ambos os modelos destacam-se Cox (1972), Cox e Oakes (1984), Kalbfleisch e Prentice (1980), Lawless (2003).

Neste trabalho foi considerado unicamente o modelo de locação e escala que é descrito a seguir.

4.1 Modelos de Regressão locação-escala

Considere T uma variável aleatória representando o tempo até a ocorrência de um evento de interesse e seja $\mathbf{x}^T$ o vetor de covariáveis.

A classe de modelos de locação-escala se caracteriza pelo fato de $Y = \log(T)$ ter uma distribuição com parâmetro de locação μ e de um parâmetro de escala σ . Pode-se então escrever

$$Y = \mu + \sigma W,$$

em que $\mu \in \mathbb{R}$ é o parâmetro de locação, $\sigma > 0$ é o parâmetro de escala e W tem distribuição que não depende de $\mathbf{x}$. Uma característica especial deste modelo é que $Y|\mathbf{X}$ tem função de sobrevivência da forma $G\left(\frac{y-\mu}{\sigma}\right)$ em que $G(\cdot)$ é a função de sobrevivência de W .

Como $\mu \in \mathbb{R}$ pode-se considerar $\mu = \mathbf{x}^T \boldsymbol{\beta}$ onde $\mathbf{x}^T$ é o vetor de covariáveis e $\boldsymbol{\beta}$ é o vetor de parâmetros $(\beta_0, \beta_1, \beta_2, \dots, \beta_q)^T$, em que q é o número de covariáveis do modelo. Logo o modelo de locação-escala é dada por

$$Y = \mathbf{x}^T \beta + \sigma W.$$

- Distribuição log-BXIIGI

Seja T_1 uma variável aleatória com distribuição BXIIGI, com fdp dada por (7). Considere $Y_1 = \log(T_1)$, então utilizando o método jacobiano temos que

$$f_{1*}(y; \boldsymbol{\theta}) = \frac{(1-p)ck \left[1 + \left(\frac{e^y}{s}\right)^c\right]^{-k-1} \left(\frac{e^y}{s}\right)^c}{\left\{1 - p \left[1 + \left(\frac{e^y}{s}\right)^c\right]^{-k}\right\}^2}, \quad -\infty < y < \infty$$

onde $\boldsymbol{\theta} = (s, c, k, p)$ é o vetor de parâmetros da distribuição log-BXIIGI (LBXIIGI), em que $\mu \in \mathbb{R}$ é o parâmetro de locação, $\sigma > 0$ é o parâmetro de escala e $k > 0$ e $0 \leq p \leq 1$ são os parâmetros de forma.

Reparametrizando a equação acima, onde $s = e^\mu$ e $c = \frac{1}{\sigma}$, obtem-se a nova distribuição

$$\begin{aligned} f_{1*}(y; \boldsymbol{\theta}) &= \frac{(1-p)k}{\sigma} \left[1 + \exp\left(\frac{y-\mu}{\sigma}\right)\right]^{-k-1} \exp\left(\frac{y-\mu}{\sigma}\right) \\ &\times \left\{1 - p \left[1 + \exp\left(\frac{y-\mu}{\sigma}\right)\right]^{-k}\right\}^{-2}. \end{aligned} \quad (31)$$

Note que a equação (31) será denominada distribuição LBXIIGI, isto é $Y_1 \sim LBXIIGI(\mu, \sigma, k, p)$, ou seja, agora $\boldsymbol{\theta} = (\mu, \sigma, k, p)$.

Figura 10 – Gráficos da fdp da distribuição LBXIIGI

A Figura 11 mostra as diversas formas que a fdp da distribuição LBXIIGI pode assumir para algumas combinações de parâmetros.

- Distribuição log-BXIIGII

Analogamente, considerando T_2 uma variável aleatória com distribuição BXIIGII, com fdp dada por (9), obtém-se a distribuição LBXIIGII, em que $\mu \in \mathbb{R}$ é o parâmetro de locação, $\sigma > 0$ é o parâmetro de escala e $k > 0$ e $0 \leq p \leq 1$ são os parâmetros de forma.

$$f_{2^*}(y; \boldsymbol{\theta}) = \frac{(1-p)k}{\sigma} \left[1 + \exp\left(\frac{y-\mu}{\sigma}\right) \right]^{-k-1} \exp\left(\frac{y-\mu}{\sigma}\right) \times \left\{ 1 - p + p \left[1 + \exp\left(\frac{y-\mu}{\sigma}\right) \right]^{-k} \right\}^{-2},$$

em que $\boldsymbol{\theta} = (\mu, \sigma, k, p)$ é o vetor de parâmetros da distribuição LBXIIGII.

Figura 11 – Gráficos da fdp da distribuição LBXIIGII

A Figura 12 mostra as diversas formas que a fdp da distribuição LBXIIGII pode assumir para algumas combinações de parâmetros.

Essas novas distribuições podem ser utilizadas para analisar diversos conjuntos de dados pois o suporte é o conjunto dos reais.

- Função de sobrevivência LBXIIGI

A função da distribuição acumulada é dada por

$$F_{1*}(y; \boldsymbol{\theta}) = P(Y \leq y) = P[\log(T) \leq y] = P[T \leq \exp(y)] = F[\exp(y)].$$

Então, por sua vez a função de sobrevivência é

$$S_{1*}(y; \boldsymbol{\theta}) = 1 - F_{1*}(y; \boldsymbol{\theta}) = 1 - F_1[\exp(y)] = S_1[\exp(y)].$$

Considere a função de sobrevivência $S_1(t; s, c, k, p)$ dada na equação (7), então

$$S_{1*}(y; \boldsymbol{\theta}) = \frac{(1-p)[1 + \exp(\frac{y-\mu}{\sigma})]^{-k}}{1 - p[1 + \exp(\frac{y-\mu}{\sigma})]^{-k}}.$$

- Função de sobrevivência LBXIIGII

Seguindo o mesmo procedimento e considerando a função de sobrevivência $S_2(t; s, c, k, p)$ dada na equação (10), a função de sobrevivência da distribuição LBXIIGII é dada por

$$S_{2*}(y; \boldsymbol{\theta}) = \frac{[1 + \exp(\frac{y-\mu}{\sigma})]^{-k}}{1 - p + p[1 + \exp(\frac{y-\mu}{\sigma})]^{-k}}.$$

- Expansão do Modelo de Regressão LBXIIGI

Através da expansão (11) obtem-se a igualdade

$$\begin{aligned} f_{1*}(y; \boldsymbol{\theta}) &= \frac{(1-p)k}{\sigma} \left[1 + \exp\left(\frac{y-\mu}{\sigma}\right) \right]^{-k-1} \exp\left(\frac{y-\mu}{\sigma}\right) \left\{ 1 - p \left[1 + \exp\left(\frac{y-\mu}{\sigma}\right)^{-k} \right] \right\}^{-2} \\ &= \sum_{j=0}^{\infty} (1-p)p^j \frac{k(j+1)}{\sigma} \left[1 + \exp\left(\frac{y-\mu}{\sigma}\right) \right]^{-k(j+1)-1} \exp\left(\frac{y-\mu}{\sigma}\right) \\ &= \sum_{j=0}^{\infty} w_j f_*[y; k(j+1), \mu, \sigma]. \end{aligned}$$

Assim foi mostrado que a distribuição LBXIIGI(k, p, μ, σ) é uma combinação linear da distribuição LBXII($k(j+1), \mu, \sigma$) com fdp $f_*(y)$ e $w_j = (1-p)p^j$.

$$f_*(y; \boldsymbol{\theta}) = \frac{k}{\sigma} \left[1 + \exp\left(\frac{y-\mu}{\sigma}\right) \right]^{-k-1} \exp\left(\frac{y-\mu}{\sigma}\right).$$

- Expansão do Modelo de Regressão LBXIIGII

Através da expansão (12) obtem-se a igualdade

$$\begin{aligned}
f_{2*}(y; \boldsymbol{\theta}) &= \frac{(1-p)k}{\sigma} \left[1 + \exp\left(\frac{y-\mu}{\sigma}\right) \right]^{-k-1} \exp\left(\frac{y-\mu}{\sigma}\right) \left\{ 1 - p + p \left[1 + \exp\left(\frac{y-\mu}{\sigma}\right)^{-k} \right] \right\}^{-2} \\
&= \sum_{j=0}^{\infty} (1-p)^{-(j+1)} p^j (-1)^j \frac{k(j+1)}{\sigma} \left[1 + \exp\left(\frac{y-\mu}{\sigma}\right) \right]^{-k(j+1)-1} \exp\left(\frac{y-\mu}{\sigma}\right) \\
&= \sum_{j=0}^{\infty} v_j f_*[y; k(j+1), \mu, \sigma].
\end{aligned}$$

Assim foi mostrado que a distribuição LBXIIGII(k,p,μ,σ) é uma combinação linear da distribuição LBXII(k(j+1),μ,σ) com fdp $f_*(y)$ e $v_j = (1-p)^{-(j+1)} p^j (-1)^j$.

- Momentos do modelo de regressão LBXIIGI

Se Y_1 segue distribuição LBXIIGI e $w_j = (1-p)p^j$ então seus momentos podem ser escritos como

$$\begin{aligned}
\mu'_r &= \int_{-\infty}^{\infty} y^r \sum_{j=0}^{\infty} w_j f_*[y; k(j+1), \mu, \sigma] dy \\
&= \sum_{j=0}^{\infty} w_j \int_{-\infty}^{\infty} y^r \frac{k}{\sigma} \left[1 + \exp\left(\frac{y-\mu}{\sigma}\right) \right]^{-k(j+1)-1} \exp\left(\frac{y-\mu}{\sigma}\right) dy.
\end{aligned}$$

Considere a transformação $u = \exp\left(\frac{y-\mu}{\sigma}\right)$.

$$\mu'_r = \sum_{j=0}^{\infty} w_j \int_0^{\infty} k[\sigma \log(u) + \mu]^r (1+u)^{-k(j+1)-1} du.$$

Utilizando a expansão a expansão binomial de $[\sigma \log(u) + \mu]^r$ obtem-se

$$\mu'_r = k \sum_{j=0}^{\infty} \sum_{l=0}^r w_j \binom{r}{l} \mu^{r-l} \sigma^l \int_0^{\infty} [\log(u)]^l (1+u)^{-k(j+1)-1} du.$$

De acordo com Prudnikov, Brychkov, Marichev (1986, vol 1, seção 2.6.4, integral 6) pode-se reescrever

$$\int_0^\infty [\log(u)]^r (1+u)^{-k(j+1)-1} du = \left(\frac{\partial}{\partial \alpha} \right)^r B[\alpha, k(j+1) + 1 - \alpha] \Big|_{\alpha=1},$$

em que $B[.,.]$ é a função Beta.

Finalmente,

$$\mu'_r = k \sum_{j=0}^{\infty} \sum_{l=0}^r w_j \binom{r}{l} \mu^{r-l} \sigma^r \left(\frac{\partial}{\partial \alpha} \right)^r B[\alpha, k(j+1) + 1 - \alpha] \Big|_{\alpha=1}.$$

- Momentos do modelo de regressão LBXIIGII

Analogamente se Y_2 segue distribuição LBXIIGII e $v_j = (1-p)^{-(j+1)} p^j (-1)^j$ então seus momentos podem ser escritos como

$$\mu'_r = k \sum_{j=0}^{\infty} \sum_{l=0}^r v_j \binom{r}{l} \mu^{r-l} \sigma^r \left(\frac{\partial}{\partial \alpha} \right)^r B[\alpha, k(j+1) + 1 - \alpha] \Big|_{\alpha=1}.$$

- Distribuição associada ao erro: LBXIIGI

A distribuição associada ao erro é dada por

$$\Pi_{1*}(w) = (1-p)k[1 + \exp(w)]^{-k-1} \exp(w) \{1 - p(1 + \exp(w))^{-k}\}^{-2}.$$

- Distribuição associada ao erro: LBXIIGII

Da mesma forma obteve-se a distribuição associada ao erro para o modelo LBXIIGII.

$$\Pi_{2*}(w) = (1-p)k[1 + \exp(w)]^{-k-1} \exp(w) \{1 - p + p[1 + \exp(w)]^{-k}\}^{-2}.$$

- Estimação dos parâmetros da distribuição LBXIIGI

Considere uma amostra aleatória (y_i, δ_i, x_i^T) para todo $i = 1, 2, \dots, n$ em que $y_i = \log(t_i)$, δ_i é o indicador de censura e $\mathbf{x}_i^T$ é o vetor de covariáveis associado ao i -ésimo indivíduo. Considerando censura não informativa e aleatória tem-se que o logaritmo da função verossimilhança é dado por

$$\begin{aligned} l(\boldsymbol{\theta}) = & r^* \log \left[\frac{(1-p)k}{\sigma} \right] - (k+1) \sum_{i \in F} \log(b_i) + \sum_{i \in F} \log(b_i - 1) - 2 \sum_{i \in F} \log(1 - pb_i^{-k}) \\ & + (n - r^*) \log(1 - p) - k \sum_{i \in E} \log(b_i) - \sum_{i \in E} \log(1 - pb_i^{-k}), \end{aligned}$$

onde r^* é o número de falhas, F indica o conjunto de falhas, E o conjunto de observações censuradas e $b_i = \left[1 + \exp \left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma} \right) \right]$.

Então o estimador de máxima verossimilhança da distribuição LBXIIGI é a solução iterativa do sistema de equações não lineares dada a seguir.

$$\frac{\partial l(\boldsymbol{\theta})}{\partial p} = -\frac{r^* \sigma}{(1-p)} + 2 \sum_{i \in F} \frac{b_i^{-k}}{1 - pb_i^{-k}} - \frac{(n - r^*)}{(1-p)} + \sum_{i \in E} \frac{b_i^{-k}}{1 - pb_i^{-k}}$$

$$\frac{\partial l(\boldsymbol{\theta})}{\partial k} = \frac{r^* \sigma}{k} - \sum_{i \in F} \log(b_i) + 2 \sum_{i \in F} \frac{p \log(b_i) b_i^{-k}}{1 - pb_i^{-k}} - \sum_{i \in E} \log(b_i) + \sum_{i \in E} \frac{p \log(b_i) b_i^{-k}}{1 - pb_i^{-k}}$$

$$\begin{aligned} \frac{\partial l(\boldsymbol{\theta})}{\partial \sigma} = & -\frac{r^*}{\sigma} + (k+1) \sum_{i \in F} \frac{(b_i - 1) \log(b_i - 1)}{\sigma b_i} - \sum_{i \in F} \frac{\log(b_i - 1)}{\sigma} \\ & - 2 \sum_{i \in F} \frac{kpb_i^{-k-1} (b_i - 1) \log(b_i - 1)}{\sigma (1 - pb_i^{-k})} - \sum_{i \in E} \frac{kpb_i^{-k-1} (b_i - 1) \log(b_i - 1)}{\sigma (1 - pb_i^{-k})} \\ & - k \sum_{i \in E} \frac{(b_i - 1) \log(b_i - 1)}{\sigma b_i} \end{aligned}$$

Considere $j = 0, 1, 2, \dots, q$, sendo q o número de covariáveis.

$$\begin{aligned} \frac{\partial l(\boldsymbol{\theta})}{\partial \beta_j} &= (k+1) \sum_{i \in F} \frac{(b_i - 1)x_{ij}}{b_i \sigma} - \sum_{i \in F} \frac{x_{ij}}{\sigma} + 2 \sum_{i \in F} \frac{k p b_i^{-k-1} (b_i - 1)x_{ij}}{\sigma(1 - p b_i^{-k})} \\ &+ k \sum_{i \in E} \frac{(b_i - 1)x_{ij}}{\sigma b_i} + \sum_{i \in E} \frac{k p b_i^{-k-1} (b_i - 1)x_{ij}}{\sigma(1 - p b_i^{-k})}. \end{aligned}$$

- Estimação dos parâmetros da distribuição LBXIIGII

Analogamente, o logaritmo da verossimilhança da distribuição LBXIIGII é dado por

$$\begin{aligned} l(\boldsymbol{\theta}) &= r^* \log \left[\frac{(1-p)k}{\sigma} \right] - (k+1) \sum_{i \in F} \log(b_i) + \sum_{i \in F} \log(b_i - 1) - 2 \sum_{i \in F} \log(1 - p + p b_i^{-k}) \\ &+ (n - r^*) \log(1 - p) - k \sum_{i \in E} \log(b_i) - \sum_{i \in E} \log(1 - p + p b_i^{-k}) \end{aligned}$$

Então o estimador de máxima verossimilhança da distribuição LBXIIGII é a solução iterativa do sistema de equações não lineares dada a seguir.

$$\frac{\partial l(\boldsymbol{\theta})}{\partial p} = -\frac{r^* \sigma}{(1-p)} + 2 \sum_{i \in F} \frac{-1 + b_i^{-k}}{1 - p + p b_i^{-k}} - \frac{(n - r^*)}{(1-p)} + \sum_{i \in E} \frac{-1 + b_i^{-k}}{1 - p b_i^{-k}}$$

$$\frac{\partial l(\boldsymbol{\theta})}{\partial k} = \frac{r^* \sigma}{k} - \sum_{i \in F} \log(b_i) - 2 \sum_{i \in F} \frac{p \log(b_i) b_i^{-k}}{1 - p + p b_i^{-k}} - \sum_{i \in E} \log(b_i) - \sum_{i \in E} \frac{p \log(b_i) b_i^{-k}}{1 - p + p b_i^{-k}}$$

$$\begin{aligned} \frac{\partial l(\boldsymbol{\theta})}{\partial \sigma} &= -\frac{r^*}{\sigma} + (k+1) \sum_{i \in F} \frac{(b_i - 1) \log(b_i - 1)}{\sigma b_i} - \sum_{i \in F} \frac{\log(b_i - 1)}{\sigma} \\ &- 2 \sum_{i \in F} \frac{k p b_i^{-k-1} (b_i - 1) \log(b_i - 1)}{\sigma(1 - p + p b_i^{-k})} - \sum_{i \in E} \frac{k p b_i^{-k-1} (b_i - 1) \log(b_i - 1)}{\sigma(1 - p + p b_i^{-k})} \\ &- k \sum_{i \in E} \frac{(b_i - 1) \log(b_i - 1)}{\sigma b_i} \end{aligned}$$

Considere $j = 0, 1, 2, \dots, q$, sendo q o número de covariáveis.

$$\begin{aligned} \frac{\partial l(\boldsymbol{\theta})}{\partial \beta_j} &= (k+1) \sum_{i \in F} \frac{(b_i - 1)x_{ij}}{b_i \sigma} - \sum_{i \in F} \frac{x_{ij}}{\sigma} + 2 \sum_{i \in F} \frac{k p b_i^{-k-1} (b_i - 1)x_{ij}}{\sigma(1 - p + p b_i^{-k})} \\ &+ k \sum_{i \in E} \frac{(b_i - 1)x_{ij}}{\sigma b_i} + \sum_{i \in E} \frac{k p b_i^{-k-1} (b_i - 1)x_{ij}}{\sigma(1 - p + p b_i^{-k})}. \end{aligned}$$

Foi utilizado o proc NLMixed do SAS para calcular os emv, intervalos de confiança e testes de hipóteses.

5 MODELOS UNIFICADOS DE FRAÇÃO DE CURA

Pode ser de interesse do estudo modelar dados em que a cura seja uma possibilidade, e que pretende-se estimar a proporção de curados. Muitos pesquisadores consideram 5 anos uma medida de cura para dados de sobrevivência com longa duração.

A literatura relativa à distribuições que considera diferentes causas competitivas latentes é rica e está se desenvolvendo rapidamente. Neste contexto será proposto um novo modelo unificado de fração de cura que considera que o número de causas competitivas do evento de interesse segue a distribuição geométrica e o tempo até a ocorrência do evento segue a distribuição BXII.

Segundo Rodrigues, Cancho e Castro (2008) de uma forma geral a idéia básica do modelo unificado de fração de cura está baseada na noção de ocorrência do evento de interesse em um processo de dois estágios:

- Estágio de iniciação: seja M uma variável aleatória representando o número de causas ou riscos de ocorrência de um particular evento de interesse com a seguinte distribuição de probabilidade.

$$p_m = P[M = m], m = 0, 1, 2, \dots$$

O evento de interesse pode ser a morte de um paciente, a reincidência de um tipo de câncer ou, exemplos com dados financeiros como o desligamento de um cliente. A causa de ocorrência do evento é desconhecida, com distribuição de probabilidade p_m especificada.

- Estágio de maturação: dado $M = m$, sejam $Z_j, j = 1, \dots, m$, variáveis aleatórias contínuas (não-negativas) independentes com função distribuição acumulada $F(z) = 1 - S(z)$ e independentes de M , representando o tempo de ocorrência do evento de interesse devido à j -ésima causa ou risco. O tempo de ocorrência do evento de interesse é definido como

$$T = \min\{Z_0, Z_1, Z_2, \dots, Z_M\}$$

em que $P[Z_0 = \infty] = 1$. Esta suposição admite a possibilidade de uma proporção p_0 da população não apresentar a ocorrência do evento de interesse. As variáveis aleatórias Z_j e M são variáveis latentes e T é uma variável aleatória observável ou censurada.

Seja $a = a_m$ uma sequência de números reais. Se

$$A_a(s) = a_0 + a_1s + a_2s^2 + \dots$$

converge para valores de s no intervalo $[0, 1]$, então $A_a(s)$ é definida como a função geradora da sequência a .

Se $a = a_m = p_m = p$, temos que $A_p(s) = E[s^M]$. Observe que a função geradora de probabilidades é a função geradora de momentos da variável aleatória M calculada no ponto $\log(s)$, isto é, $A_p(s) = E[\exp\{\log(s)M\}]$.

A função de sobrevivência da variável aleatória T (função de sobrevivência da população) será indicada por

$$\begin{aligned} S_{pop}(t) &= P[M = 0] + P[Z_1 > t, Z_2 > t, \dots, Z_M > t, M \geq 1] \\ &= P[M = 0] + \sum_{m=1}^{\infty} P[M = m]P[Z_1 > t, Z_2 > t, \dots, Z_M > t] \\ &= p_0 + \sum_{m=1}^{\infty} p_m \{S(t)\}^m \\ &= A_p[S(t)], \end{aligned}$$

sendo que $A_p(\cdot)$ é a função geradora da sequência $p = p_m$. Ou seja, a função de sobrevivência da variável aleatória T , correspondente a um modelo de longa duração em dois estágios, é uma composição da função geradora de probabilidades e a função de sobrevivência.

A $S_{pop}(t)$, função de sobrevivência de longa duração em dois estágios, não é uma função de sobrevivência própria.

Figura 12 – Função de sobrevivência própria e imprópria

Observe que $\lim_{t \rightarrow 0} S(t) = 1$ e $\lim_{t \rightarrow \infty} S(t) = 0$ para a função de sobrevivência própria. Já para a função de sobrevivência imprópria observe que $\lim_{t \rightarrow 0} S_{pop}(t) = 1$ e $\lim_{t \rightarrow \infty} S_{pop}(t) = P[M = 0] = p_0$, como mostra figura 13. Considere p_0 como a proporção de não ocorrências do evento de interesse na população, ou seja a taxa de sobrevivência de longa duração ou fração de cura.

A derivada de $S_{pop}(t)$ satisfaz a relação

$$-\frac{dS_{pop}(t)}{dt} = f(t) \sum_{m=1}^{\infty} m p_m [S(t)]^{m-1},$$

notando que $f(t)$ é uma função densidade própria.

É possível expressar $S_{pop}(t)$ na forma do modelo de mistura. De fato,

$$S_{pop}(t) = p_0 + (1 - p_0)S^*(t),$$

em que

$$S^*(t) = \sum_{m=1}^{\infty} p_m^* [S(t)]^m$$

e

$$p_m^* = \frac{p_m}{1 - p_0},$$

sendo que $S^*(t) = P[T > t | M \geq 1]$ é uma função de sobrevivência própria, isto é, $\lim_{t \rightarrow \infty} S(t) = 0$. Considerando $A_p(S(t)) = \frac{1-p}{1-pS(t)}$, função geradora de probabilidades da distribuição geométrica, a função de sobrevivência própria correspondente é obtida através da transformação geométrica dada em (4).

Segue que a função densidade imprópria associada à função de sobrevivência de longa duração em dois estágios é dada por

$$f_{pop}(t) = f(t) \frac{dA_p(s)}{ds} \Big|_{s=S(t)} \quad (32)$$

Então de (32) a função de densidade (própria) da população em risco e sua correspondente função de risco são dadas por

$$f^*(t) = -\frac{dS^*(t)}{dt} = \frac{f(t)}{1 - p_0} \frac{dA_p(s)}{ds} \Big|_{s=S(t)}.$$

Algumas distribuições importantes, tais como Binomial, Poisson, Geométrica e Binomial Negativa contemplam diferentes modalidades de dispersão e possibilitam uma modelagem mais flexível do número de possíveis causas de ocorrência do evento de interesse (M). Essa flexibilização é fundamental para a formulação de modelos mais complexos para representar o número de causas.

Considerando $A_p(S(t)) = \frac{1-p}{1-pS(t)}$, função geradora de probabilidades da distribuição geométrica, e $f(t)$ a densidade da BXII(s,c,k) então obtem-se a distribuição BXIIGI dada em (7).

Considere $M \sim \text{Geométrica}(p)$ com fdp dada por (1), então a função geradora de probabilidade é dada por

$$\psi_M(x) = \frac{1-p}{1-px}.$$

Assim a função de sobrevivência do modelo unificado de cura é dada por:

$$S_{pop}(t) = \frac{1-p}{1-pS(t)},$$

em que $S(t)$ é a função de sobrevivência da distribuição BXII.

Dessa maneira, seja $S(t)$ a função de sobrevivência da distribuição BXII, então

$$S_{pop}(t) = \frac{1-p}{1-p[1+(\frac{t}{s})^c]^{-k}}$$

e a proporção de curados é dada por $\lim_{t \rightarrow \infty} S_{pop}(t) = 1-p$.

Figura 13 – Gráficos da função de sobrevivência da distribuição BXII considerando fração de cura

A figura 14 mostra as diversas formas que a função de sobrevivência do modelo unificado de fração de cura para algumas combinações de parâmetros.

Assim,

$$f_{pop}(t) = f(t) \frac{dS_{pop}(t)}{dS(t)} = \frac{ckp(1-p) \left[1 + \left(\frac{t}{s}\right)^c\right]^{-(k+1)} \frac{t^{(c-1)}}{s^c}}{\left\{1 - p \left[1 + \left(\frac{t}{s}\right)^c\right]^{-k}\right\}^2}.$$

- Estrutura de Regressão para modelo unificado de fração de cura

Caso seja de interesse estudar sobre covariáveis que influenciam a proporção de curados, a ligação logito pode ser utilizada na estrutura de regressão. Isto é, seja $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ o i -ésimo vetor de covariáveis associado ao i -ésimo indivíduo, então pode-se escrever

$$p(\mathbf{x}_i) = \frac{\exp(\mathbf{x}_i^T \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i^T \boldsymbol{\beta})}, \quad (33)$$

em que $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \dots, \beta_q)^T$ é o vetor de parâmetros.

De (33) obtem-se

$$\mathbf{x}_i^T \boldsymbol{\beta} = \underbrace{\log \left(\frac{p_i}{1 - p_i} \right)}_{\text{logito}}.$$

Neste caso o logaritmo da função de verossimilhança é dado por

$$\begin{aligned} l(\boldsymbol{\theta}) &= r^* \log(ck) + \sum_{i \in F} \log(p(\mathbf{x}_i)) + \sum_{i \in F} \log(1 - p(\mathbf{x}_i)) - (k + 1) \sum_{i \in F} \log(u_i) + (c - 1) \sum_{i \in F} \log(t_i) \\ &- r^* c \log(s) - 2 \sum_{i \in F} \log(1 - p(\mathbf{x}_i) u_i^{-k}) + \sum_{i \in E} \log(1 - p(\mathbf{x}_i)) - \sum_{i \in E} \log(1 - p(\mathbf{x}_i) u_i^{-k}), \end{aligned}$$

onde r^* é o número de falhas, F indica o conjunto de falhas, E o conjunto de observações censuradas e $u_i = \left[1 + \left(\frac{t_i}{s}\right)^c\right]$.

Então o estimador de máxima verossimilhança da distribuição do modelo unificado de fração de cura é a solução iterativa do seguinte sistema de equações não lineares.

$$\begin{aligned}\frac{\partial l(\boldsymbol{\theta})}{\partial c} &= \frac{r^*}{c} - (k+1) \sum_{i \in F} \frac{\log(u_i - 1)(u_i - 1)^c}{u_i} + \sum_{i \in F} \log(t_i) - r^* \log(s) \\ &- 2 \sum_{i \in F} \frac{kp(\mathbf{x}_i)u_i^{-k-1} \log(u_i - 1)(u_i - 1)^c}{1 - p(\mathbf{x}_i)u_i^{-k}} - \sum_{i \in E} \frac{kp(\mathbf{x}_i)u_i^{-k-1} \log(u_i - 1)(u_i - 1)^c}{1 - p(\mathbf{x}_i)u_i^{-k}}\end{aligned}$$

$$\frac{\partial l(\boldsymbol{\theta})}{\partial k} = \frac{r^*}{k} - \sum_{i \in F} \log(u_i) + 2 \sum_{i \in F} \frac{p(\mathbf{x}_i) \log(u_i) u_i^{-k}}{1 - p(\mathbf{x}_i) u_i^{-k}} + \sum_{i \in E} \frac{p(\mathbf{x}_i) \log(u_i) u_i^{-k}}{1 - p(\mathbf{x}_i) u_i^{-k}}$$

$$\frac{\partial l(\boldsymbol{\theta})}{\partial s} = -\frac{r^* c}{s} + (k+1) \sum_{i \in F} \frac{ct_i^c}{s^{c+1}u_i} + 2 \sum_{i \in F} \frac{kp(\mathbf{x}_i)cu_i^{-k-1}t_i^c}{s^{c+1}(1 - p(\mathbf{x}_i)u_i^{-k})} + \sum_{i \in E} \frac{kp(\mathbf{x}_i)cu_i^{-k-1}t_i^c}{s^{c+1}(1 - p(\mathbf{x}_i)u_i^{-k})}$$

Considere $j = 0, 1, 2, \dots, q$ sendo q o número de covariáveis.

$$\begin{aligned}\frac{\partial l(\boldsymbol{\theta})}{\partial \beta_j} &= \sum_{i \in F} x_{ij} - \sum_{i \in F} \frac{\exp(x_{ij}\beta_j)x_{ij}}{1 + \exp(x_{ij}\beta_j)} + \sum_{i \in F} [1 + \exp(x_{ij}\beta_j)] \exp(x_{ij}\beta_j) x_{ij} \\ &+ 2 \sum_{i \in F} \frac{u_i^{-k} \exp(x_{ij}\beta_j) x_{ij} [1 + \exp(x_{ij}\beta_j)] - \exp(x_{ij}\beta_j) x_{ij} \exp(x_{ij}\beta_j)}{\left[1 - \frac{u_i^{-k} \exp(x_{ij}\beta_j)}{1 + \exp(x_{ij}\beta_j)}\right] [1 + \exp(x_{ij}\beta_j)]^2} \\ &+ \sum_{i \in E} \frac{u_i^{-k} \exp(x_{ij}\beta_j) x_{ij} [1 + \exp(x_{ij}\beta_j)] - \exp(x_{ij}\beta_j) x_{ij} \exp(x_{ij}\beta_j)}{\left[1 - \frac{u_i^{-k} \exp(x_{ij}\beta_j)}{1 + \exp(x_{ij}\beta_j)}\right] [1 + \exp(x_{ij}\beta_j)]^2} \\ &+ \sum_{i \in E} [1 + \exp(x_{ij}\beta_j)] \exp(x_{ij}\beta_j) x_{ij}\end{aligned}$$

Seja $\hat{\boldsymbol{\beta}}$ o emv de $\boldsymbol{\beta}$, então

$$\hat{p}(\mathbf{x}_i) = \frac{\exp(x_i^T \hat{\boldsymbol{\beta}})}{1 + \exp(x_i^T \hat{\boldsymbol{\beta}})}$$

é o emv de p_i .

Dessa maneira $\hat{p}$ pode ser estimado pela média dos $\hat{p}_i$.

$$\hat{p} = \frac{\sum_{i=1}^n \hat{p}_i}{n}$$

Foi utilizado o proc NLMixed do SAS para calcular os emv, intervalos de confiança e testes de hipóteses.

6 APLICAÇÕES

Nesta seção será ilustrado o uso das distribuições BXIIGI, BXIIGII, LBXIIGII e do modelo com fração de cura considerando que o número de causas competitivas segue a distribuição geométrica e o tempo segue a distribuição BXII.

Os softwares utilizados para o cálculo das estimativas dos parâmetros das distribuições, foram o R *Development Core Team* (2009) e o SAS, *Statistical Analysis System*, 2004.

6.1 Churn Voluntário de clientes

O conjunto de dados estudado refere-se a clientes que encerram sua relação com uma determinada empresa de forma espontânea (churn voluntário). O tempo de vida é o dia em que o indivíduo se torna cliente até o dia em que ele encerra suas atividades com a empresa. Já a censura é caracterizada pelo fato de quando o estudo foi realizado o cliente ainda não havia cancelado os serviços com a empresa, ou seja, até o momento do estudo ele não apresentou o evento de interesse. Esses dados são sigilosos e não podem ser divulgados.

Primeiramente foram obtidas as estimativas de máxima verossimilhança (estimativas de m.v.) dos parâmetros com seu correspondente erro padrão entre parênteses e os valores das estatísticas AIC, CAIC e BIC, em que o menor valor deste critério determina o melhor ajuste (Tabela 1). Os parâmetros foram estimados por meio do método da máxima verossimilhança descritos na seção (3.9), utilizando o procedimento NLMixed do SAS. Esses resultados indicam que o modelo BXIIGI apresenta o menor valor em uma das três medidas estudadas entre todos os modelos ajustados, portanto pode ser escolhido como o modelo mais adequado.

Tabela 1 – Estimativas de m.v. do modelo para os dados do tempo de vida de clientes

Modelo	c	s	k	p	AIC	CAIC	BIC
Log-logística	1.725 (0.295)	28.803 (2.670)	- -	- -	4746	4746	4755.8
BXII	2.583 (0.295)	14.171 (3.212)	0.376 (0.024)	- -	4700.7	4700.7	4715.4
BXIIGI	2.560 (0.346)	19.239 (3.875)	0.207 (0.036)	0.7315 (0.004)	4696	4696	4715.6
BXIIGII	2.583 (0.299)	14.171 (3.215)	0.376 (0.026)	$(1E^{-7})$ $(1E^{-8})$	4702.7	4702.7	4722.3

Ainda como uma maneira de avaliar o melhor modelo foi realizado o teste da razão de verossimilhança (Tabela 2), que confirmou os resultados obtidos anteriormente.

Tabela 2 – Teste da razão de verossimilhança

Hipótese	RV	$p - \text{valor}$
Log-logística versus BXII ($H_0 : k = 1$ versus $H_1 : k \neq 1$)	47.3	<0.0001
BXII versus BXIIGI ($H_0 : p = 0$ versus $H_1 : p \neq 0$)	6.7	0.0096
Log-logística versus BXIIGI ($H_0 : k = 1$ e $p = 0$ versus $H_1 : k \neq 1$ e $p \neq 0$)	54	<0.0001

A figura 15 mostra o estimador de Kaplan Meier (EKM) para os dados do tempo de vida dos clientes, conjuntamente com a função de sobrevivência BXIIGI ajustada, mostrando o bom ajuste desta distribuição para os dados.

Figura 14 – EKM para o tempo do vida de clientes e sobrevivência estimada da distribuição BXIIGI

6.2 Dados de entomologia

As moscas-das-frutas, *Ceratitis capitata* e *Anastrepha fraterculus*, é um dos maiores obstáculos para a produção e comercialização de frutas no país, são consideradas as espécies mais prejudiciais, causando mais danos à agricultura do que qualquer outra.

O método mais utilizado nos dias de hoje é a aplicação de iscas tóxicas e pulverizações em cobertura (Sugayama, R. L., 2000). Embora efetivo, o controle químico acarreta diversos problemas, tais como o desequilíbrio biológico, problemas de saúde humana tanto para o aplicador quanto para os consumidores através dos resíduos tóxicos nos frutos, opinião pública contrária e o fato de atualmente os importadores de frutas *in natura* exigirem frutos com baixos níveis de agrotóxicos e sem mascas-da-fruta (Carvalho, R. S.; Nascimento, A. S., 2002).

A planta *Azadirachta indica* é mais conhecida internacionalmente por nim, sendo que no Brasil o nome foi aportuguesado para nim (Martinez, S., 2002). Essa planta tem origem do sul e o sudeste da Ásia, mas atualmente está estabelecida em áreas tropicais e subtropicais da África, América e Austrália. Nos últimos vinte anos, o nim foi introduzido em vários países principalmente para o reflorestamento e produção de madeira em áreas secas, mas também tem sido utilizado com outros propósitos, incluindo o uso para sombreamento e para produção de inseticidas naturais (Schmutterer, H., 1990).

No Brasil, o nim foi introduzido em 1984 e, atualmente, se encontra em quase todas as regiões do país, devido às condições climáticas adequadas para o plantio, principalmente nas regiões Centro-Oeste, Norte e Nordeste (Schmutterer, H., 1990).

A importância do nim como planta inseticida foi revelada para o mundo ocidental pelo professor Dr. Heinrich Schmutterer (Universidade de Giessen, Alemanha) na década de 50 do século passado, ao constatar que durante as migrações destruidoras dos gafanhotos (*Schistocerca gregária*) no Sudão, apenas as árvores de nim restavam intactas (Veitch, G. E., Boyer, A., Ley, S. V., 2008). Posteriormente, foram intensificadas as pesquisas sobre a composição química da planta em diferentes centros de pesquisa do mundo.

Recentemente tem sido estudado a utilização de derivados do nim para controle de moscas-das-frutas, com o objetivo principal de reduzir e ou substituir as aplicações de inseticidas no campo. A árvore de nim tem sido muito investigada quanto às suas propriedades e quanto ao seu potencial como inseticida, visto que seus derivados tem se revelado tão potentes quanto os inseticidas comerciais (Martinez, S., 2002).

Para avaliar a longevidade das moscas-das-frutas foram utilizadas gaiolas de plástico onde foi fixado um algodão inserido em um tubo plástico contendo 12.5 ml de um dos extratos (extrato aquoso de amêndoas (ESW), extrato de folhas em metanol (ELM) e extrato de ramos em diclorometano (EBD)) nas concentrações (39, 225 e 888 ppm (partículas por molécula)). Como controles foram utilizados água deionizada e acetona a 5% , resultando assim em 11 tratamentos. Em cada gaiola foram liberados dois casais e fornecida dieta artificial (proteína hidrolisada + açúcar). O experimento foi conduzido em delineamento inteiramente casualizado com quatro repetições, sob condições controladas de temperatura ($25 \pm 2^\circ$), umidade ($60 \pm 10\%$). O tempo para o fim do experimento foi de 51 dias e as moscas que não morreram foram consideradas como censura. Durante os procedimentos uma gaiola apresentou problema e foi descartada do estudo (parcela perdida), assim no total foram observadas o tempo de vida de 172 moscas. Este experimento

foi realizado na escola superior de agricultura “Luiz de Queiroz”.

De acordo com Silva, M. A., et al., 2013 os tratamentos são divididos em dois grupos, grupo 1 composto pelos tratamentos ELM (39 ppm), EBD (225 ppm) e EBD (888 ppm) e grupo 2 composto pelos tratamentos controles, ESW (39 ppm), ESW (225 ppm), ESW (888 ppm), ELM (225 ppm), ELM (888 ppm) e EBD (39 ppm). Os grupos foram assim divididos uma vez que não existe diferenças significativas dentre os grupos e existe diferença entre os grupos.

Primeiramente foram estimados os parâmetros dos modelos BXII, BXIIGI e BXIIGII através da maximização iterativa do logaritmo da função de verossimilhança. Foi calculado também as estatísticas AIC, BIC e CAIC que são úteis para a comparação entre os modelos. Essas informações estão na Tabela 3 e a medida entre parênteses é o desvio padrão. Essas informações foram obtidas utilizando a subrotina NLMixed no software SAS.

Os resultados mostram que o modelo BXIIGII apresenta os menores valores de AIC, BIC e CAIC dentre os modelos estudados, isto indica que este modelo pode ser o mais adequado para modelar este conjunto de dados.

Tabela 3 – Estimativas dos parâmetros do modelo BXIIGII ajustado para dados de entomologia, correspondentes desvios padrões em parênteses e estatísticas AIC, BIC e CAIC

Modelo	c	s	k	p	AIC	CAIC	BIC
BXII	2.2145 (0.2322)	34.4233 (10.3861)	2.6764 (1.2842)	- -	1282.7	1282.8	1292.1
BXIIGI	2.2089 (0.2337)	34.7298 (10.7228)	2.7138 (1.3303)	$1E^{-8}$ ($1E^{-9}$)	1284.7	1284.9	1284.3
BXIIGII	1.1643 (0.3890)	14.3538 (8.1751)	4.4140 (2.5322)	0.9818 (0.02114)	1270.1	1270.3	1282.7

Ainda como uma maneira de avaliar o melhor modelo foi realizado o teste da razão de verossimilhança (Tabela 4), que confirma os resultados obtidos anteriormente.

Tabela 4 – Teste da razão de verossimilhança

Hipótese	RV	$p - \text{valor}$
BXIIGI versus BXII ($H_0 : p = 0$ versus $H_1 : k \neq 0$)	1.20	0.2733
BXIIGII versus BXII ($H_0 : p = 0$ versus $H_1 : p \neq 0$)	14.60	0.00013

A fim de avaliar se o modelo é apropriado foi construído um gráfico (Figura 16) com a função de sobrevivência empírica e a função de sobrevivência estimada do modelo

BXIIGII, que mostra o bom ajuste para esses dados.

Figura 15 – (a) Densidade da distribuição BXII e BXIIGII ajustada para os dados de entomologia. (b) Função de sobrevivência da distribuição BXII e BXIIGII estimada e sobrevivência empírica para os dados de entomologia

Agora, considerando covariáveis, tem-se o modelo:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \sigma z_i$$

onde a variável aleatória Y_i segue a distribuição LBXIIGII ou LBXIIGI para $i = 1, \dots, n$, x_1 é a covariável sexo e x_2 é 1 quando o tratamento pertence ao grupo 1 e 0 quando o tratamento pertence ao grupo 2. As estimativas de máxima verossimilhança dos parâmetros foram calculadas utilizando a subrotina NLMixed do software SAS. A maximização iterativa o logaritmo da função de verossimilhança começa com valores iniciais para k , σ e β oriundos da estimativa obtida do modelo de regressão LBXII ($p = 0$).

As estimativas de máxima verossimilhança dos parâmetros dos modelos BXII, BXIIGI e BXIIGII e os erros padrões dessas estimativas são mostradas na tabela 6. Note que a covariável x_2 é significativa ao nível de 1% de significância, entretanto a outra covariável não é significativa em qualquer nível de significância usual.

Tabela 5 – Estimativas dos parâmetros do modelo de regressão LBXIIGII ajustado para dados de entomologia, correspondentes desvios padrões em parênteses e estatísticas AIC, BIC e CAIC

Modelo	β_0	β_1	β_2	σ	k	p	AIC	CAIC	BIC
LBXII	4.3085 (1.1349) [0.0002]	0.1104 (0.0963) [0.2528]	-0.4014 (0.0978) [<0.001]	0.4877 (0.0597)	9.4002 (8.6141)	- - -	346.1	346.4	361.8
LBXIIGI	4.3085 (1.1316) [0.0002]	0.1104 (0.0962) [0.2525]	-0.4014 (0.0978) [<0.001]	0.4877 (0.0596)	9.3993 (8.5578)	$1E^{-8}$ ($1E^{-9}$)	348.1	348.6	367.0
LBXIIGII	3.1649 (0.8847) [0.0005]	0.0354 (0.0803) [0.6605]	-0.3252 (0.0876) [0.003]	0.9107 (0.3379)	6.0541 (4.8403)	0.9798 (0.0247)	335.7	336.2	354.6

Os valores de AIC, BIC e CAIC apresentados na Tabela 5 indicam que o modelo de regressão LBXIIGII pode ser escolhido como o melhor dentre os modelos testados.

Ainda como uma maneira de avaliar o melhor modelo foi realizado o teste da razão de verossimilhança (Tabela 6), que confirma os resultados obtidos anteriormente, pois claramente rejeita a hipótese nula nos três testes em favor da distribuição LBXIIGII.

Tabela 6 – Teste da razão de verossimilhança

Hipótese	RV	$p - valor$
BXIIGI versus BXII ($H_0 : p = 0$ versus $H_1 : k \neq 0$)	1.60	0.2059
BXIIGII versus BXII ($H_0 : p = 0$ versus $H_1 : p \neq 0$)	12.40	0.0004

A fim de simplificar o modelo, foi retirado a variável x_1 como uma variável explicativa, então

$$y_i = \beta_0 + \beta_2 x_{i2} + \sigma z_i$$

As estimativas para o modelo de regressão LBXIIGII ajustado para os dados de entomologia são apresentadas na Tabela 7.

Tabela 7 – Estimativas dos parâmetros do modelo de regressão LBXIIGII ajustado para dados de entomologia, correspondentes desvios padrões em parênteses e estatísticas AIC, BIC e CAIC

Modelo	β_0	β_2	σ	k	p
LBXII	3.0993	-0.3238	0.8986	5.7188	0.9804
LBXIIGI	(0.8847)	(0.0875)	(0.3250)	(4.2283)	(0.0236)
LBXIIGII	[0.0002]	[0.0003]			

A fim de avaliar se o modelo é apropriado foi construído um gráfico (figura 17) com a função de sobrevivência empírica e a função de sobrevivência estimada do modelo de regressão LBXIIGII, que mostra o bom ajuste para esses dados.

Figura 16 – Função de sobrevivência do modelo LBXIIGII estimada e curva de sobrevivência empírica para os dados de entomologia

Conclui-se então que a cavariável sexo não influencia no tempo de vida e que as moscas-das-frutas submetidas aos tratamentos do grupo 1 tem o tempo de vida menor que as submetidas aos tratamentos do grupo2.

6.3 Dados de melanoma cutâneo

O melanoma cutâneo é um tipo de câncer de pele que tem origem nos melanócitos, células produtoras de melanina, substância que determina a cor da pele.

De acordo com Parker, S.L. et al. (1996), o melanoma cutâneo apresenta morbidez significativa e alta taxa de mortalidade, respondendo por quase 75% de todas as mortes devidas a neoplasmas malignos e quase 1 a 2% de todas as mortes por câncer nos Estados Unidos, assim representa um verdadeiro problema de saúde pública. Melanoma cutâneo está entre os mais sérios cânceres quando considerado o número de anos de vida perdidos, pois frequentemente atinge indivíduos jovens, diferente de outros tipos de câncer. Além disso, apesar de todos os esforços das terapias utilizadas em casos avançados como quimioterapia, radioterapia, terapia bioquímica e vacinas, não são

alcançados resultados satisfatórios perante à cura e taxa de sobrevivência (Ferrari Junior et al, 2008). O conhecimento das propriedades histológicas o melanoma cutâneo podem contribuir para uma melhor compreensão da doença. O estudo apresentado visa analisar dados histopatológicos de pacientes com melanoma cutâneo, levando em consediração a presença de ulceração. Isto pode facilitar ações médicas com o objetivo de identificar casos.

Considere x_1 como o status de ulceração (presente $x_1 = 1$; ausente $x_1 = 0$).

Primeiramente será considerado o moelo de regressão BXII com fração de cura em que

$$p(\mathbf{x}_i) = \frac{\exp(\beta_0 + \beta_1 x_{i1})}{1 + \exp(\beta_0 + \beta_1 x_{i1})} \quad i = 1, \dots, 215$$

Foi calculada as estimativas de máxima verossimilhança dos parâmetros utilizando a subrotina NLMixed do software SAS. A Tabela 8 mostra esses resultados.

Tabela 8 – Estimativas de máxima verossimilhança dos parâmetros do modelo de regressão BXIIG considerando fração de cura ajustado para os dados de melanoma cutâneo

Modelo	β_0	β_1	s	c	k
LBXIIGII	-1.4288 (0.3255) [<0.0001]	1.8048 (0.3255) [<0.0001]	351.24 (71.04)	1.8146 (0.2337)	1519.17 (210.43)

Então a proporção de curados é maior quando $x_1 = 0$, ou seja, quando há ausência de ulceração.

Em seguida, foi calculada a estimativa da proporção de curados. Note que

$$\hat{p} = \frac{1}{215} \sum_{i=1}^{215} \hat{p}(\mathbf{x}_i) = 0.3687,$$

onde

$$\hat{p}(\mathbf{x}_i) = \frac{\exp(-1.4288 + 1.8048x_{i1})}{1 + \exp(-1.4288 + 1.8048x_{i1})},$$

e então a proporção de curados é dada por

$$\hat{p}_0 = 1 - \hat{p} = 0.6312.$$

A fim de avaliar se o modelo é apropriado foi construído um gráfico figura 18 (a) com a função de sobrevivência empírica e a função de sobrevivência marginal estimada do modelo de regressão BXII com fração de cura. As estimativas da fração de cura dos pacientes estratificados pelo status de ulceração, 0 ou 1, são 0.8068 e 0.4071 respectivamente, e as curvas de sobrevivência estimada e empírica para os dois casos são apresentadas na figura 18 (b).

Assim o modelo de regressão BXII com fração de cura fornece um bom ajuste para esse conjunto de dados.

Figura 17 – Função de sobrevivência do modelo BXII com fração de cura estimada e curva de sobrevivência empírica para os dados de melanoma cutâneo (b) Função de sobrevivência do modelo BXII com fração de cura estimada e curva de sobrevivência empírica para os dados de melanoma cutâneo estratificado pelo status de ulceração

7 CONCLUSÕES

Neste trabalho foram propostas duas distribuições novas, baseadas na distribuição Burr XII seguindo a ideia de Adamidis e Loukas (1998), obtiveram-se as distribuições as Burr XII geométrica tipo I e Burr XII geométrica tipo II, distribuições importantes para a análise de tempos de vida. A principal vantagem dos novos modelos é a flexibilidade em acomodar diferentes formas de função de taxa de falha. Foi apresentada uma descrição detalhada de algumas propriedades matemáticas das distribuições novas, baseadas em simples expansões da função de densidade. Este resultado permite obter expressões para os momentos e para a função geradora de momentos. Também foram propostos dois modelos de regressão baseados nas distribuições novas para dados com censura.

Outra pesquisa importante foi propor o modelo de regressão Burr XII geométrico para incluir indivíduos de longa duração. A estrutura de riscos competitivos mostrou-se uma boa função de ligação para poder incluir covariáveis na proporção de curados do modelo de regressão Burr XII geométrico. O método da máxima verossimilhança foi usado para estimar os parâmetros dos modelos, obter a matriz de informação observada e realizar os testes de hipóteses para os parâmetros baseados na distribuição assintótica destas estimativas. Finalmente, foram demonstradas em três aplicações que os modelos propostos podem produzir melhores ajustes que seus sub-modelos.

- Trabalhos futuros

Possíveis trabalhos futuros:

1. Desenvolvimento e implementação computacional de outros métodos de estimação tais como: estimador bayesiano e a técnica bootstrap;
2. Realizar um estudo detalhado da análise residual para os modelos de regressão propostos;
3. Desenvolver testes de influência local que fornecem um diagnóstico global da qualidade de ajuste, mediante a modificação proposta por Zhu e Zhang (2004);
4. Utilizar a distribuição binomial negativa ao invés da distribuição geométrica para propor um novo modelo;
5. Considerar o modelo de regressão com fração de cura com efeito aleatório.

REFERÊNCIAS

AARSET, M.V. How to identify bathtub hazard rate. **Transactions on Reliability**, New York, v.36, p.106-108, 1987.

ADAMIDIS, K.; LOUKAS, S. A lifetime distribution with decreasing failure rate. **Statistics and Probability Letters**, Ioannina, v.39, p.35-42, 1998.

BARRETO-SOUZA, W.; MORAIS, A.L.; CORDEIRO, G.M. The Weibull-geometric distribution. **Journal of Statistical Computation and Simulation**, New York, v.81, p.645-657, 2010.

BARRETO, M.L.; SANTOS, L.M.P.; ASSIS, A.M.O.; ARAÚJO, M.P.N.; FARENZENA, G.G.; SANTOS, P.A.B.; FIACCONE R.L. Effect of vitamin A supplementation on diarrhoea and acute lower-respiratory-tract infections in young children in Brazil. **Lancet**, New York, v.344, p.228-231, 1994.

BASTOS, J.; ROCHA, C. Análise de Sobrevivência: Conceitos Básicos. **Arq Med Porto**, v. 20, n. 5-6, set. 2006 . Disponível em: <http://www.scielo.gpeari.mctes.pt/scielo.php?script=sci_arttext&pid=S0871-34132006000400007&lng=pt&nrm=iso>. acessos em: 10 nov. 2013.

BERKSON, J.; GAGE, R.P. Survival curves for cancer patients following treatment. **Journal of the American Statistical Association**, London, v.47, p.501-515, 1952.

BURR, I.W. Cumulative frequency functions. **Annals of Mathematical Statistics**, Baltimore, v. 13, p.215-232, 1942.

CARVALHO, R. S., NASCIMENTO, A. S. Criação e utilização de *Diachasmimorpha longicaudata* para controle biológico de moscas-das-frutas (Tephritidae). In: PARRA, J. R. P., BOTELHO, P. S. M., CORREA-FERREIRA, B. S. BENTO J. M. S. . **Controle Biológico no Brasil: parasitóides e predadores**. São Paulo: Manole, 2002. P. 165-179.

COLLET, D. **Modelling Survival Data in Medical Research**. 2nd Ed. London: Chapman and Hall, 2003. 408p.

COLOSIMO, E.A.; GIOLO, S.R. **Análise de sobrevivência aplicada**. São Paulo: Edgard Blücher, 2006. 369p.

COX, D.R. Regression models and life-tables (with discussion). **Journal of the Royal Statistical Society: Series B, Methodological**, London, v.34, n. 2, p.187-220, 1972.

COX, D.R.; OAKES, D. **Analysis of Survival Data**. London: Chapman and Hall, 1984. 201p.

FERRARI JR, N.M., MULLER H., RIBEIRO M., MAIA M., SANCHES JR J.A. Cutaneous melanoma: descriptive epidemiological. **São Paulo Medical Journal**, São Paulo, v.126, p.7-41, 2008.

FERREIRA, J.M. **Análise de Sobrevivência: uma Visão de Risco Comportamental na Utilização de Cartão de Crédito**. 2007. 129p. Dissertação (Mestrado em Biometria) - Instituto de Estatística e Informática da Universidade Federal Rural de Pernambuco, Universidade Federal Rural de Pernambuco, Recife, 2007.

GRADSHTEYN, I.S.; RYZHIK, I.M. **Table of Integrals, Series and Products**. 7th Ed. London: Academic Press, 2007. 345p.

IBRAHIM, J.G.; CHEN M.; SINHA D. **Bayesian Survival Analysis**. New York: Springer, 2000. 495p.

KALBFLEISCH, J.D.; PRENTICE, R.L. **The Statistical Analysis of Failure Time Data**. New York: Wiley, 1980. 321p.

KALBFLEISCH, J.D.; PRENTICE, R.L. **The Statistical Analysis of Failure Time Data**. 2nd Ed. New York: Wiley, 2002. 462p.

KENNEY, J.F.; KEEPING, E.S. **Mathematics os Statistics**. 3rd Ed. New Jersey: Van Nostrand, 1962, Pt1. 348p.

KLEIN, J.P.; MOESCHBERGER M.L. **Survival Analysis: Techniques for Censored and Truncated Data**. 2nd Ed. New York: Springer, 2003. 552p.

- KUS, C. A new lifetime distribution. **Computational Statistics and Data Analysis**, New York, v.52, p.4497-4509, 2007.
- LAWLESS, J.F. **Statistical Models and Methods for Life Time Data**. New York: John Wiley & Sons, 1982. 580p.
- LAWLESS, J.F. **Statistical Models and Methods for Life Time Data**. 2nd Ed. New York: John Wiley & Sons, 2003. 630p.
- LOPES, L.F.D. **Apostila Estatística**. Santa Maria: UFSM, 2003. 145p.
- LU, W.; SHI, D. A new compounding life distribution: the Weibull-Poisson distribution. **Journal of Applied Statistics**, London, DOI: 10.1080/02664763.2011.575126, 2011.
- MALLER, R.A.; ZHOU, S. **Survival Analysis with Long-Term Survivors**. New York: Wiley, 1996. 304p.
- MARTINEZ, S. **O nim *Azadirachta indica*: natureza, usos múltiplos, produção**. Londrina: IAPAR, 2002. 142p.
- MILLER, R.G. **Survival Analysis**. Chichester: Wiley, 1981. 238p.
- MOORS, J.J.A. A quantile alternative for kurtosis. **Journal of Royal Statistical Ser. D., The Statistician**, London, v.37, p.25-32, 1998.
- MUDHOLKAR, G.S; SRIVASTAVA, D.K.; KOLLIA, G.D. A generalization of the Weibull distribution with application to the analysis of survival data. **Journal of the American Statistical Association**, Boston, v.91, n.436, p.1575-1583, 1996.
- NADARAJAH, S.; KOTZ, S. The Beta Gumbel distribution. **Mathematical Problems in Engineering**, New York, v.10, p.323-332, 2004.
- NOCEDAL, J.; WRIGHT, S.J. **Numerical optimization**. 2nd Ed. New York: Springer, 2006. 664p.

- PAES, A.T. **Uso de modelos com fração de cura na análise de dados de sobrevivência com omissão de covariáveis**. 2007. 129p. Dissertação (Doutorado em Ciências) - Instituto de Matemática e Estatística da Universidade de São Paulo, Universidade de São Paulo, São Paulo, 2007.
- PARKER, S.L., TONG, T., BOLDEN S., WINGO P.A. Cancer statistics. **CA: A Cancer Journal for Clinicians**, Philadelphia, v.47, p.5-27, 1997.
- PRUDNIKOV, A.P.; BRYCHKOV, Y.A.; MARICHEV O.I. **Integrals and Series**. Amsterdam: Gordon and Breach Science Publishers, 1986. v.1, 800 p.
- RODRIGUES, C.; CORDEIRO, G.M.; DEMÉTRIO, C.G.B.; ORTEGA, E.M.M. The Weibull negative binomial distribution. **Advances and Applications in Statistics**, Allahabad, v.22, p.25-55, 2011.
- RODRIGUES, J.; CANCHO, V.G.; CASTRO M. **Teoria Unificada de Análise de Sobrevivência**. São Paulo: ABE, 2008. 94p.
- RODRIGUEZ, R.N. A guide to the Burr type XII distributions. **Biometrika**, Oxford, v.64, p.129-134, 1977.
- SANTANA, T.V.F. **As distribuições Kumaraswamy-log-logística e Kumaraswamy-logística**. 2010. 102p. Dissertação (Mestrado em Ciências Exatas) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2010.
- SCHMUTTERER, H. Properties and potential of natural pesticides from the neem tree. *Azadirachta indica* . **Annual Review of Entomology**, Palo Alto, v.35, p.271-297, 1990.
- SEN, P.K., SINGER, J.M. **Large sample methods in statistics: An introduction with applications**. New York: Chapman and Hall, 1993. 382p.
- SHAO, Q. Notes on maximum likelihood estimation for the three-parameter Burr XII distribution. **Computational Statistics and Data Analysis**, New York, v.45, p.675-687, 2004.

SHAO, Q.X.; WONG, H.; XIA, J.; IP, W.C. Models for extreme using the extended three-parameter Burr XII system with application to flood frequency analysis.

Hydrological Sciences Journal, Wallingford, v.49, p.685-702, 2004.

SILVA, A. S., BEZERRA-SILVA G. C. D., VENDRAMIM J. D., MASTRANGELO T.

Sublethal effect of nem extract on mediterranean fruit fly adults. **Revista Brasileira de Fruticultura**, Weinheim, v.35, n.1, p.93-101, 2013.

SILVA, G.O. **Modelos de regressão quando a função de taxa de falha não é monótona e o modelo probabilístico beta Weibull modificada**. 2008. 207p.

Dissertação (Doutorado em Ciências Exatas) - Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2008.

SILVA, G.O.; ORTEGA, E.M.M.; CANCHO, V.G.; BARRETO, M.L. Log-Burr XII

regression models with censored data. **Computational Statistics and Data Analysis**, New York, v.52, p.3820-3842, 2008.

SILVA, G.O.; ORTEGA, E.M.M.; CORDEIRO, G.M. The beta modified Weibull

distribution. **Lifetime Data Anal**, New York, v.16, p.409-430, 2010.

SUGAYAMA, R. L. ***Anastrepha fraterculus* (Wiedemann) (Diptera:**

Tephritidae) na região produtora de maçãs do Rio Grande do Sul: relação com seus inimigos naturais e potencial para o controle biológico. 2000. 117p. Tese (Doutorado em Ciências) - Instituto de Biociências, Universidade de São Paulo, 2000.

TAHMASBI, R; REZAEI, S. A two-parameter lifetime distribution with decreasing failure rate. **Computational Statistics and Data Analysis**, New York, v.52, p.3889-3901, 2008.

THERNEAU, T.M.; GRAMBSCH P.M. **Modeling Survival Data: Extending the Cox Model**. New York: Springer, 2000. 350p.

VEITCH, G. E., BOYER A., LEY S. V. The azadirachtin story. **Angewandte Chemie International Edition**, Weinheim, v.47, n.49 , p.9402-9429, 2008.