



Universidade de Brasília
IE - Departamento de Estatística
Trabalho de Conclusão de Curso

**Uma versão truncada da distribuição
Birnbaum-Saunders generalizada aplicada a
análise de risco**

Eduardo Barreto Sulz Gonsalves

Relatório Final

Orientador: Prof.º Helton Saulo Bezerra dos Santos

Brasília

Junho de 2018

Resumo

O ajuste de modelos estatísticos é cada vez mais utilizado por bancos que desejam obter boas estimativas de algumas medidas de risco sobre seu capital como VaR e TVaR. Uma distribuição que tem sido frequentemente utilizada nesse ramo é a distribuição Birnbaum-Saunders generalizada. Neste trabalho, é proposta uma nova distribuição de probabilidades para modelar dados financeiros, a distribuição Birnbaum-Saunders generalizada truncada. Inicialmente será feita uma revisão bibliográfica e, então, serão apresentadas algumas funções do novo modelo como a função densidade de probabilidades (FDP), função de distribuição acumulada (FDA), estimadores de máxima verossimilhança e algumas medidas de risco como VaR e TVaR. Por fim, o novo modelo será aplicado a dados reais a fim de verificar o ajuste de cada caso especial do novo modelo abordados neste trabalho.

Abstract

The adjustment of statistical models is increasingly used by banks that wish to obtain good estimates of some risk measures as VaR and TVaR on their capital . One distribution that has been frequently used in this field is the generalized Birnbaum-Saunders distribution. In this work, a new distribution of probabilities is proposed to model financial data, the truncated generalized Birnbaum-Saunders distribution. Initially a bibliographic review will be done and then some functions of the new model will be presented as PDF, CDF, maximum likelihood estimators and some risk measures such as VaR and TVaR. Finally, the new model will be applied to real data in order to verify the adjustment of each special case of the new model addressed in this work.

Sumário

1. Introdução	4
2. Distribuições e Análise de Risco	8
2.1 Distribuições Truncadas	8
2.2 Valor em Risco e Valor em Risco na Cauda	10
2.3 Distribuição Birnbaum-Saunders	11
2.4 Distribuição Birnbaum-Saunders Truncada	13
2.5 Distribuição Birnbaum-Saunders Generalizada	16
3. Distribuição Birnbaum-Saunders Generalizada Truncada	20
3.1 FDP e FDA da TGBS	20
3.2 Função quantil da TGBS	22
3.3 Estimador de Máxima Verossimilhança	22
4. Aplicação do novo modelo	24
4.1 Análise exploratória dos dados	24
4.2 Ajuste dos dados aos modelos	27
4.3 Análise de risco	29
4. Conclusão	31

1. Introdução

Muitas instituições financeiras possuem clientes que contraem empréstimos de pequeno valor e não os pagam. Embora os valores dos referidos empréstimos sejam inexpressivos, o volume de operações dessa espécie acarreta prejuízos consideráveis para os agentes financeiros. Além disso, o custo para ajuizar ação judicial objetivando o recebimento do crédito nem sempre vale a dívida que será recuperada. Geralmente, as instituições financeiras costumam definir os valores mínimos para serem cobrados judicialmente com o propósito de recuperar o crédito devido. Para isso são utilizados modelos estatísticos que se adequem bem aos dados fornecendo boas estimativas. Entretanto, se escolhido de forma ineficiente, o modelo selecionado pode fornecer previsões erradas dos mesmos.

Os dados de perda monetária estão diretamente ligados aos modelos de tempo de vida. Na maioria dos casos, a distribuição normal não se ajusta bem aos dados, pois esses conjuntos de dados costumam possuir assimetria a direita, unimodalidade e suporte positivo; ver Ahmed et al. (2010). Dessa forma, deve ser escolhida uma das duas alternativas: transformar os dados ou utilizar uma distribuição que se ajuste bem a eles. A transformação dos dados pode levar a interpretações errôneas dos resultados, além do fato de que, se uma transformação for feita de forma inadequada,

o poder do estudo diminui.

A partir dos estudos de Davis (1952) a distribuição exponencial se mostrou um ótimo modelo para descrever dados de tempo de vida. Uma de suas principais propriedades é a da perda de memória, ou seja, apresenta uma taxa de falha constante. Isto permite simplificação dos problemas e da inferência associados a esse modelo. Entretanto, por causa dessa propriedade a sua aplicação a dados reais é limitada. Com a intensiva aplicação da distribuição exponencial a dados de tempo de vida, a distribuição Gamma se mostrou uma boa opção para modelar somas de tempo de vida. Uma série de estudos sobre o tema apontaram outras distribuições de vida que ajustam dados de tempo de vida tão bem quanto à exponencial. Kaon (1959) sugeriu a Weibull como uma boa distribuição para ajustar dados de falhas. Além de ser matematicamente tratável, a distribuição Weibull foi muito utilizada por conseguir descrever situações com diferentes riscos. Seu desenvolvimento por Weibull (1951) se deu com o objetivo de descrever a força necessária para a ruptura de materiais.

Motivados por problemas de fadiga dos materiais em aeronaves comerciais, Birnbaum-Saunders (1969) propuseram a distribuição Birnbaum-Saunders (BS), a qual descreve o tempo de vida de equipamentos expostos à fadiga causada por estresse ou por um ciclo de tensão. A distribuição BS se baseia no fato de que a exposição de equipamentos à fadiga geram falhas ou rachaduras no material em decorrência ao acúmulo de dano ao mesmo. Para a construção do modelo foram adotados alguns pressupostos. Mais detalhes podem ser vistos em Leiva (2016). A BS é uma transformação monótona da distribuição normal padrão. Além disso, por

possuir algumas propriedades como suporte positivo, assimetria à direita e unimodalidade, a BS vem sendo aplicada nas mais diversas áreas do conhecimento como biologia, engenharia e medicina. Como citado em Ahmed et al. (2010), acredita-se, por esses motivos, que essa distribuição pode ser amplamente aplicada nos campos de finanças, economia e gerenciamento de risco.

Em Ahmed et al. (2010), é proposta uma aplicação da versão truncada da distribuição Birnbaum-Saunders (TBS), visto que o truncamento desta distribuição fornece melhores previsões quando se trata de risco financeiro. O truncamento da BS em κ é obtido a partir da aplicação de uma distribuição normal truncada à esquerda em η na transformação monótona da BS. Outra forma de se obter a TBS é a partir do truncamento de uma BS em κ .

Uma classe de distribuições que vem sendo frequentemente utilizada é a Birnbaum-Saunders generalizada (GBS), proposto por Díaz-García e Leiva (2005). Ao invés de ser obtida através de uma transformação da distribuição normal padrão, a GBS é obtida a partir da transformação linear de uma distribuição simétrica. Diferente dos modelos BS citados anteriormente, a GBS possui um terceiro parâmetro ou vetor de parâmetros que controla, por exemplo, a curtose. Diversas propriedades da BS também são válidas para a GBS como a assimetria positiva, suporte positivo, entre outras. Por esses motivos, essa classe de modelos é amplamente utilizada, além do fato de poderem ter caudas mais leves ou mais pesadas do que a distribuição BS tradicional.

A partir das distribuições BS, TBS e GBS é possível obter algumas medidas de

risco como a função de risco. Porém, essas medidas de risco não são suficientes para organizações financeiras. Essas organizações precisam de meios mais eficientes de medir o risco a fim de minimizar as perdas com o capital empregado nas operações, visto que ele está em alto nível de exposição. Uma medida muito utilizada é o valor em risco (VaR), que indica o quanto as ações e capital da empresa estão expostos aos riscos. Para isso o VaR verifica a perda associada ao risco e fornece um único valor que quantifica essa exposição.

Portanto, o principal objetivo desse trabalho é propor um novo modelo, ou seja, uma versão truncada da distribuição Birnbaum-Saunders Generalizada (TGBS) e desenvolver um pacote computacional do novo modelo proposto no software R. Além disso, aplicar o novo modelo a dados financeiros reais a fim de comparar os diferentes ajustes. Por fim, será realizada uma breve análise de risco para os diferentes modelos ajustados.

2. Distribuições e Análise de Risco

Será feita uma revisão bibliográfica sobre as distribuições BS, BS truncada, GBS e alguns outros conceitos como truncamento e medidas de risco, como VaR e TVaR, a fim de fornecer insumos para o desenvolvimento da TGBS e as correspondentes medidas de risco.

2.1 Distribuições Truncadas

Uma distribuição truncada é aquela em que o seu suporte foi alterado em função do truncamento. Como consequência de um truncamento de uma distribuição, sua função densidade de probabilidades (FDP) é alterada. Seja X uma variável aleatória com a seguinte distribuição, $X \sim S(f)$ em que $S(f)$ é uma distribuição de probabilidades conhecida.

Sendo assim, a FDP de uma distribuição truncada é o resultado da divisão da FDP da distribuição não truncada pela função de distribuição acumulada (FDA) da distribuição não truncada, no intervalo de truncamento. Isto é:

$$f_X(x; a, b) = \frac{f(x)}{F(b) - F(a)}, a < x < b,$$

em que b é o ponto de truncamento à direita e a é o ponto de truncamento à esquerda.

Além disso, f é a FDP e F é a FDA.

Para o caso da distribuição normal, $f = \phi$ e $F = \Phi$. ϕ e Φ são a FDP e FDA da normal padrão $N(0, 1)$.

No caso de distribuições simétricas, o truncamento a esquerda é dado por:

$$f_x(x; a) = \frac{f(x)}{1 - F(a)}, a < x < \infty, a > 0.$$

A Figura 1 abaixo representa graficamente o truncamento distribuição normal em diferentes pontos.

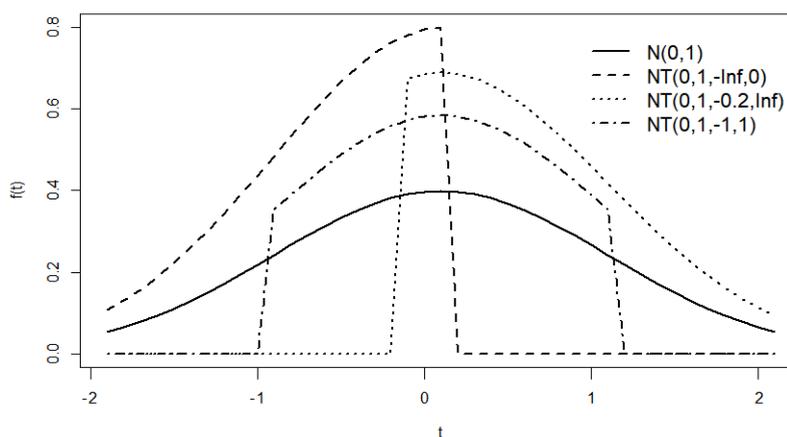


Figura 1: FDP da normal padrão e normal padrão truncada

Na Figura 1, a linha contínua representa a distribuição normal padrão. Note que, uma distribuição pode ser truncada à direita, representada na linha tracejada; à esquerda, indicada pela linha pontilhada; ou por ambos os lados, representada pela linha tracejada e pontilhada.

2.2 Valor em Risco e Valor em Risco na Cauda

Valor em risco, ou VaR, é uma medida de risco. Essa medida estabelece a maior perda esperada de um investimento sob uma determinada probabilidade p em um período de tempo d . Considere a variável aleatória $X \sim S(f)$ em que $S(f)$ é uma distribuição de probabilidades conhecida. Então, de forma matemática, o VaR é o quantil q da distribuição dos dados.

$$VaR_q[X] = t_q.$$

em que t_q é o q -ésimo quantil da distribuição dos dados.

Além do Valor em Risco, há o Valor em Risco na Cauda (TVaR). Essa medida de risco informa qual é maior perda esperada de um investimento sob um determinado nível de confiança, em um período de tempo d , dado que a perda foi maior que o q -ésimo quantil, ou seja, a perda foi maior que o VaR. Sendo assim, sua fórmula é dada por:

$$TVaR_q[X] = VaR_q[X] + e(t_q),$$

em que $e(t_q)$ é a função de vida média residual dada por $e(t_q) = \int_{t_q}^{\infty} S_X(x)dt/S_X(t_q)$.

A função $S_X(\cdot)$ é a função de sobrevivência no ponto especificado.

2.3 Distribuição Birnbaum-Saunders

Seja X uma variável aleatória que segue uma distribuição BS com parâmetros de forma e escala α e β , respectivamente, ou seja, $X \sim BS(\alpha, \beta)$, então tem-se a seguinte representação estocástica:

$$X = \beta \left(\frac{\alpha Z}{2} + \sqrt{\left(\frac{\alpha Z}{2} \right)^2 + 1} \right)^2 \sim BS(\alpha, \beta),$$

em que $\alpha > 0$, $\beta > 0$ e $Z \sim N(0,1)$. Note que Z pode ser obtido de:

$$Z = \frac{1}{\alpha} \left(\sqrt{\frac{X}{\beta}} - \sqrt{\frac{\beta}{X}} \right) \sim N(0, 1).$$

A FDP associada a X é dada por:

$$f_x(x) = \phi \left(\frac{1}{\alpha} \xi \left(\frac{x}{\beta} \right) \right) \frac{1}{\alpha \beta} \xi' \left(\frac{x}{\beta} \right), x > 0, \quad (1)$$

em que ϕ representa a FDP da normal padrão, $\xi(x) = [x^{1/2} - x^{-1/2}]$ e $\xi'(x) = d\xi(x)/dx = [x^{-1/2} + x^{-3/2}]/2$.

A correspondente FDA é dada por:

$$F_x(x) = \Phi \left(\frac{1}{\alpha} \xi \left(\frac{x}{\beta} \right) \right), x > 0, \quad (2)$$

em que Φ representa a FDA da normal padrão.

O q -ésimo quantil de $X \sim \text{BS}(\alpha, \beta)$ é dado por $x_q = [\beta/4][\alpha z_q + \sqrt{\alpha^2 z_q^2 + 4}]^2$, em que $0 < q < 1$ e z_q é o q -ésimo quantil da distribuição normal padrão. A maneira mais formal de apresentar sua função quantílica é apresentada abaixo:

$$t(q; \alpha, \beta) = F_X^{-1}(x; \alpha, \beta) = \beta \left(\frac{\alpha z_q}{2} + \sqrt{\left(\frac{\alpha z_q}{2}\right)^2 + 1} \right)^2, 0 < q < 1,$$

em que F_X^{-1} é o inverso de F_T e $z_q = \Phi^{-1}(q)$ é o $q \times 100$ -ésimo quantil da normal padrão.

Note que Φ^{-1} é a inversa de Φ , ou seja, a FDA da normal padrão. Utilizando $q = 0.5$, temos que $z(0.5) = 0$, o que corresponde à média, moda e mediana da normal padrão. A partir disso, é possível verificar que $t(0.5; \alpha, \beta) = \beta$, o que comprova que o parâmetro de escala β é, de fato, a mediana da BS.

Algumas das propriedades da distribuição BS são: sua distribuição é contínua, unimodal, assimétrica e possui suporte positivo. Além disso, a partir da Figura 2, é possível observar que, para diversas distribuições BS com $\beta = 1$, a medida que o valor do parâmetro α tende a 0, a distribuição tende a ficar centralizada de forma mais simétrica ao redor de β , ou seja, sua mediana. Essa propriedade não vale apenas para $\beta = 1$ mas sim para qualquer β . Além disso, sua variabilidade diminui.

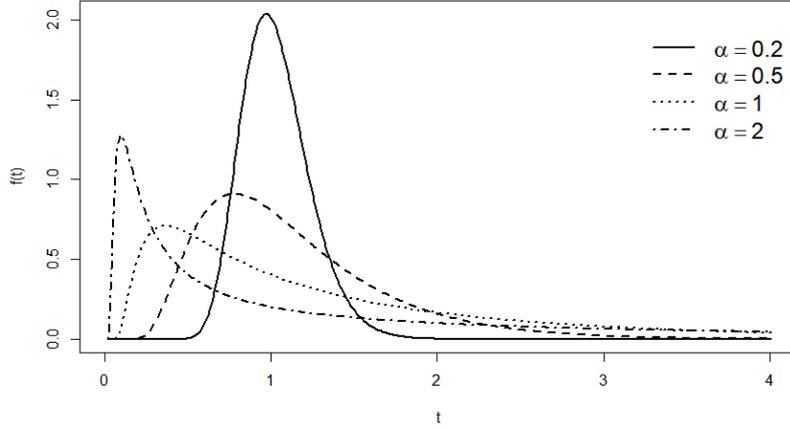


Figura 2: FDP da BS para diferentes α e $\beta=1$

Por fim, tem-se que $cX \sim BS(\alpha, c\beta)$ se $c > 0$ e $1/X \sim BS(\alpha, 1/\beta)$.

Outras duas funções úteis para a análise de tempo de vida são as de sobrevivência e de risco, as quais, associadas a variável aleatória $X \sim BS(\alpha, \beta)$, correspondem respectivamente a:

$$S_X(x; \alpha, \beta) = 1 - F_X(x; \alpha, \beta) = \Phi \left(-\frac{1}{\alpha} \xi \left(\frac{x}{\beta} \right) \right), x > 0,$$

$$h_X(x; \alpha, \beta) = \frac{f_X(x; \alpha, \beta)}{R_X(x; \alpha, \beta)} = \frac{\phi \left(\frac{1}{\alpha} \xi \left(\frac{x}{\beta} \right) \right) \frac{1}{\alpha \beta} \xi' \left(\frac{x}{\beta} \right)}{\Phi \left(-\frac{1}{\alpha} \xi \left(\frac{x}{\beta} \right) \right)}, x > 0.$$

A função de risco da BS é unimodal para qualquer α , crescente em $x < x_c$ e decrescente se $x > x_c$, em que x_c é o ponto de máximo da função de risco.

2.4 Distribuição Birnbaum-Saunders Truncada

Ao se truncar $X \sim BS(\alpha, \beta)$ à esquerda em um valor positivo igual a κ , obtem-

se $T \sim TBS_{\kappa}(\alpha, \beta)$. A partir das Equações (1) e (2), é possível obter a FDP da distribuição BS truncada em κ , isto é,

$$f_X(x) = \frac{\phi\left(\frac{1}{\alpha}\xi\left(\frac{t}{\beta}\right)\right)\xi'\left(\frac{t}{\beta}\right)}{\alpha\beta\Phi\left(-\frac{1}{\alpha}\xi\left(\frac{\kappa}{\beta}\right)\right)}, t \geq \kappa > 0.$$

Se $T \sim TBS_{\kappa}(\alpha, \beta)$, então a variável aleatória $Z = [[T/\beta]^{1/2} - [\beta/T]^{1/2}]/\alpha$ segue uma distribuição normal padrão truncada em η , que pode ser representada por $Z \sim TN_{\eta}(0,1)$; ver Cohen (1991, p.9). Além disso, $T = [\beta/4][\alpha Z + \sqrt{\alpha^2 Z^2 + 1}]^2 \sim TBS_{\kappa}(\alpha, \beta)$ com o ponto de truncamento $\kappa = [\beta/4][\alpha\eta + \sqrt{\alpha^2\eta^2 + 4}]^2$. Com isso, pode-se obter a TBS de duas maneiras distintas: truncando a BS à esquerda em κ ou gerando uma BS a partir de uma distribuição normal truncada à esquerda em η , em que $\eta = [\kappa/\beta]^{1/2} - [\beta/\kappa]^{1/2}/\alpha$.

A FDA da TBS, além de ser utilizada para observar probabilidades, também pode ser usada para obter a função quantílica da mesma, visto que, se F_T é a FDA da BS, então a função quantílica será o inverso da acumulada, ou seja, $t_q = F_T^{-1}(q)$. Com isso, sendo $T \sim TBS_{\kappa}(\alpha, \beta)$, temos as seguintes funções de distribuição acumulada e quantílica, respectivamente,

$$F_T(t) = \frac{\Phi\left(\frac{1}{\alpha}\xi\left(\frac{t}{\beta}\right)\right) - \Phi\left(\frac{1}{\alpha}\xi\left(\frac{\kappa}{\beta}\right)\right)}{\Phi\left(-\frac{1}{\alpha}\xi\left(\frac{\kappa}{\beta}\right)\right)}, t \geq \kappa > 0,$$

$$t_q = F_T^{-1}(q) = \frac{\beta}{4} \left[\alpha z_q + \sqrt{\alpha^2 z_q^2 + 4} \right]^2,$$

em que, z_q é o q -ésimo quantil da distribuição normal padrão truncada em η , ou seja, $Z \sim TN_\eta(0, 1)$, com $\eta = [[\kappa/\beta]^{1/2} - [\beta/\kappa]^{1/2}]/\alpha$.

Considere uma variável aleatória $T \sim TBS_\kappa(\alpha, \beta)$, então, algumas das suas propriedades são:

- $cT \sim TBS_{c\kappa}(\alpha, c\beta)$, se $c > 0$;
- $1/T$ segue uma distribuição BS com parâmetros α e $1/\beta$ truncada à esquerda em $1/\kappa$.

As funções de sobrevivência e de risco da TBS podem ser obtidas por meio da FDA da TBS e das funções densidade e de sobrevivência da mesma. Considerando $T \sim TBS_\kappa(\alpha, \beta)$, as funções de sobrevivência e risco são dadas a seguir, respectivamente,

$$S_T(t) = \frac{\Phi\left(-\frac{1}{\alpha}\xi(t/\beta)\right)}{\Phi\left(-\frac{1}{\alpha}\xi(\kappa/\beta)\right)}, t > \kappa,$$

$$h_T(t) = \frac{\phi\left(\frac{1}{\alpha}\xi(t/\beta)\right)\xi'(t/\beta)}{\alpha\beta\Phi\left(-\frac{1}{\alpha}\xi(t/\beta)\right)}, t > \kappa.$$

O comportamento do limite da taxa de risco de uma distribuição é importante pois, após atingir o ponto de mudança, algumas taxas de risco decrescem até o zero enquanto outras se estabilizam em constantes positivas não nulas. Se $T \sim TBS_\kappa(\alpha, \beta)$, então $\lim_{t \rightarrow \infty} h_T(t) = 1/[2\alpha^2\beta]$.

Seja T uma variável aleatória, então, o VaR de T para o quantil q será denotado por $\text{VaR}_q(T)$ ou t_q . Como foi apresentado nas seções anteriores, o VaR é uma medida de risco de uma variável aleatória, que pode ser obtido a partir do quantil da distribuição de perda. Portanto, para $T \sim TBS_\kappa(\alpha, \beta)$, tem-se:

$$\text{VaR}_q(T) = \frac{\beta}{4} \left[\alpha z_q + \sqrt{\alpha^2 z_q^2 + 4} \right]^2,$$

em que z_q é o q -ésimo quantil de $Z \sim TN_\eta(0, 1)$ com $\eta = [[\kappa/\beta]^{1/2} - [\beta/\kappa]^{1/2}]/\alpha$.

2.5 Distribuição Birnbaum-Saunders Generalizada

A distribuição GBS é obtida a partir de qualquer distribuição simétrica, podendo apresentar caudas mais leves ou pesadas que a BS tradicional, dependendo da distribuição utilizada. Dessa forma, as variáveis aleatórias,

$$T = \frac{\beta}{4} \left(\alpha Z + \sqrt{\alpha^2 Z^2 + 4} \right)^2$$

e

$$Z = \alpha^{-1} (\sqrt{T/\beta} - \sqrt{\beta/T})$$

seguem, respectivamente, uma distribuição GBS, denotada por $T \sim GBS(\alpha, \beta, f)$, e $Z \sim S(f)$, em que $S(f)$ é qualquer distribuição simétrica do caso esférico como visto em Leiva (2008). Na GBS, α é o parâmetro de forma, β é o parâmetro de

escala e $S(f)$ é a função densidade de Z , com parâmetros de posição e escala dados por $\mu = 0$ e $\sigma = 1$, respectivamente. Algumas das propriedades da GBS são listadas a seguir:

- $cT \sim GBS(\alpha, c\beta; f)$, com $c > 0$;
- $T^{-1} \sim GBS(\alpha, \beta^{-1}; f)$.

Dessa forma, a FDP e a FDA generalizadas para GBS são obtidas por meio de:

$$f_T(t) = f(a_t(\alpha, \beta))A_t(\alpha, \beta)$$

e

$$F_T(t) = F(a_t(\alpha, \beta)),$$

em que, $a_t(\alpha, \beta) = \frac{1}{\alpha}(\sqrt{t/\beta} - \sqrt{\beta/t})$, $A_t(\alpha, \beta) = \frac{d}{dt}a_t(\alpha, \beta) = \frac{t^{-\frac{3}{2}}(t+\beta)}{2\alpha\sqrt{\beta}}$ e Z segue uma distribuição simétrica $S(f)$.

Alguns casos particulares da GBS são a GBS-Laplace, GBS-Logística, BS e GBS-t(v). Então, se $T \sim GBS\text{-Laplace}(\alpha, \beta)$, $T \sim GBS\text{-Logística}(\alpha, \beta)$, $T \sim BS(\alpha, \beta)$ e $T \sim GBS\text{-t}(\alpha, \beta)$, tem-se, respectivamente, as seguintes FDPs na Tabela 1 abaixo.

Como apresentado em Díaz-Garcia e Leiva (2005), com base na Figura 3, observa-se que todas as distribuições possuem curtose maior que a distribuição

Tabela 1: FDP da GBS para diferentes núcleos

Distribuição	FDP
GBS-Laplace	$f_T(t) = \frac{1}{2} \exp\left(-\left \frac{1}{\alpha} \left[\frac{\sqrt{t}}{\sqrt{\beta}} - \frac{\sqrt{\beta}}{\sqrt{t}}\right]\right \right) \frac{t^{-\frac{3}{2}}(t+\beta)}{2\alpha\sqrt{\beta}}, t > 0$
GBS-Logística	$f_T(t) = \frac{\exp\left(\frac{1}{\alpha} \left[\frac{\sqrt{t}}{\sqrt{\beta}} - \frac{\sqrt{\beta}}{\sqrt{t}}\right]\right)}{\left[1 + \exp\left(\frac{1}{\alpha} \left[\frac{\sqrt{t}}{\sqrt{\beta}} - \frac{\sqrt{\beta}}{\sqrt{t}}\right]\right)\right]^2} \frac{t^{-\frac{3}{2}}(t+\beta)}{2\alpha\sqrt{\beta}}, t > 0$
BS	$f_T(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2\alpha^2} \left[\frac{t}{\beta} + \frac{\beta}{t} - 2\right]\right) \frac{t^{-\frac{3}{2}}(t+\beta)}{2\alpha\sqrt{\beta}}, t > 0$
GBS-t(v)	$f_T(t) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{(v\pi)^{\frac{1}{2}}\Gamma\left(\frac{v}{2}\right)} \left(1 + \frac{1}{r\alpha^2} \left[\frac{t}{\beta} + \frac{\beta}{t} - 2\right]\right)^{-\left(\frac{v-1}{2}\right)} \frac{t^{-\frac{3}{2}}(t+\beta)}{2\alpha\sqrt{\beta}}, t > 0, v > 0$

normal. Ou seja, essas distribuições possuem caudas mais pesadas, o que significa que seus dados estão mais dispersos do que na distribuição normal. Além disso, nota-se que, apesar de todas as distribuições terem curtose inflexível, a distribuição t-Student apresenta curtose flexível. Isto é, à medida que os graus de liberdade dessa distribuição aumentam, sua curtose diminui.

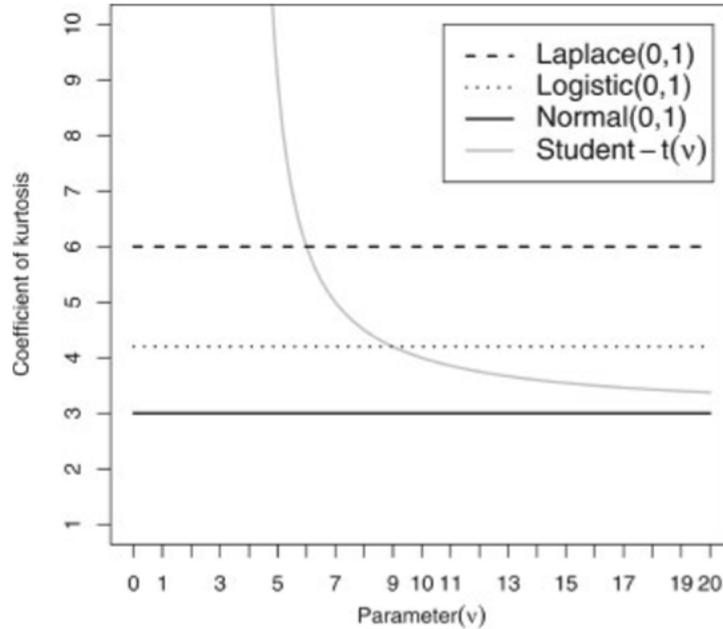


Figura 3: Curtose vs. graus de liberdade para cada distribuição.

Desse modo, a distribuição GBS apresenta grande flexibilidade e variabilidade

de formas, tanto em relação à curtose quanto à simetria e à bimodalidade. Além disso, é interessante ressaltar que o parâmetro α também afeta a curtose do modelo e, sempre que existir assimetria no modelo, ele será assimétrico à direita. Figura retirada do artigo de Leiva et al. (2007).

3. Distribuição Birnbaum-Saunders Generalizada Truncada

Nesse capítulo é apresentada a nova distribuição de probabilidade proposta, a TGBS. Serão apresentadas algumas funções como FDP, FDA, estimador de máxima verossimilhança e medidas de risco.

Foi desenvolvido um pacote computacional para o software R, o `tgbs`, que estará disponível mediante solicitação aos autores.

3.1 FDP e FDA da TGBS

Para o desenvolvimento do novo modelo proposto foram aplicadas as metodologias e conceitos apresentados acima. Seja T uma variável aleatória que segue a distribuição GBS, ou seja, $T \sim GBS(\alpha, \beta, f)$, onde f é o núcleo de uma distribuição simétrica do caso esférico. Foi realizado o truncamento à esquerda da GBS para obter a função densidade de probabilidade da TGBS. Dessa forma, foi aplicada a seguinte equação:

$$f_T(t) = \frac{f(a_t(\alpha, \beta))A_t(\alpha, \beta)}{1 - F(a_t(\alpha, \beta))}, 0 < \kappa < t,$$

Em que, como visto anteriormente, $a_t(\alpha, \beta) = \frac{1}{\alpha}(\sqrt{t/\beta} - \sqrt{\beta/t})$, $A_t(\alpha, \beta) = \frac{d}{dt}a_t(\alpha, \beta) = \frac{t^{-\frac{3}{2}}(t+\beta)}{2\alpha\sqrt{\beta}}$ e Z , que é uma variável aleatória presente no cálculo da variável aleatória T , segue uma distribuição simétrica $S(f)$ específica. F e f são a FDA e FDP, respectivamente. Neste estudo foram trabalhados os kernels normal, logístico, Laplace e t-Student. Sua FDA é dada por:

$$F_T(t) = \int_{\kappa}^t \frac{f(a_t(\alpha, \beta))A_t(\alpha, \beta)}{1 - F(a_{\kappa}(\alpha, \beta))} dt, \forall t > \kappa$$

$$F_T(t) = \frac{F(a_t(\alpha, \beta)) - F(a_{\kappa}(\alpha, \beta))}{1 - F(a_{\kappa}(\alpha, \beta))}, \forall t > \kappa$$

em que κ é o ponto de truncamento estabelecido.

Na Figura 4 é possível observar a FDP de cada kernel estudado nesse trabalho da TGBS para ambos α e β iguais a 1.

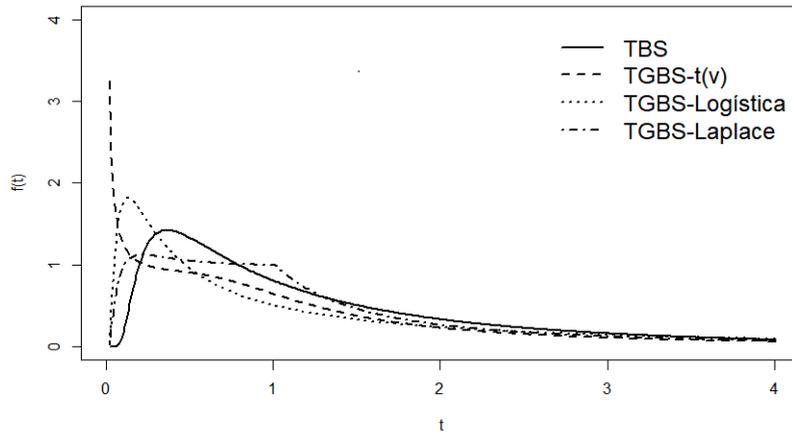


Figura 4: FDP das diferentes TGBS com α e β iguais a 1

3.2 Função quantil da TGBS

Considere a variável aleatória X que segue uma TGBS, ou seja, $X \sim TGBS_{\kappa}(\alpha, \beta, f)$, onde κ é o ponto de truncamento. Sua função quantil é obtida por:

$$t_q = F_X^{-1}(q) = \frac{\beta}{4} \left[\alpha z_q + \sqrt{\alpha^2 z_q^2 + 4} \right]^2,$$

em que z_q é o q -ésimo quantil de $Z \sim S_{\eta}(f)$, com $\eta = [[\kappa/\beta]^{1/2} - [\beta/\kappa]^{1/2}]/\alpha$, e $S_{\eta}(f)$ é uma distribuição simétrica truncada em η . No caso em que $S_{\eta}(f)$ é a normal truncada, então $z_q = F^{-1}(1 - (1 - q)[1 - F(\eta)])$, $\forall q \in (0, 1)$.

3.3 Estimador de Máxima Verossimilhança

O procedimento para o cálculo do estimador de máxima verossimilhança da TGBS é o mesmo que para o cálculo do estimador de máxima verossimilhança de qualquer distribuição.

Seja uma amostra aleatória $T_1, T_2, T_3, \dots, T_n$ em que $T \sim TGBS_{\kappa}(\alpha, \beta, f)$, então, sua verossimilhança é dada por:

$$L(\alpha, \beta) = \prod_{i=1}^n f_T(t_i),$$

$$L(\alpha, \beta) = \prod_{i=1}^n \frac{f(a_{t_i}(\alpha, \beta)) A_{t_i}(\alpha, \beta)}{1 - F(a_{t_i}(\alpha, \beta))},$$

em que, $a_t(\alpha, \beta) = \frac{1}{\alpha}(\sqrt{t/\beta} - \sqrt{\beta/t})$, $A_t(\alpha, \beta) = \frac{d}{dt}a_t(\alpha, \beta) = \frac{t^{-\frac{3}{2}}(t+\beta)}{2\alpha\sqrt{\beta}}$, f e F são a FDP e FDA de $Z \sim S(f)$, respectivamente.

Dessa forma, a log-verossimilhança da TGBS é dada por:

$$l = \ln L(\alpha, \beta)$$

$$l = \sum_{i=0}^n (\ln(f(a_{t_i}(\alpha, \beta))A_{t_i}(\alpha, \beta)) - \ln(1 - F(a_{t_i}(\alpha, \beta))))$$

Entretanto, pela complexidade dos cálculos, os estimadores de máxima verossimilhança não serão apresentados neste documento. Todo o cálculo da aplicação foram realizados por meio do software estatístico R.

4. Aplicação do novo modelo

Os dados trabalhados nesse estudo são referentes aos empréstimos não pagos a um banco de varejo mexicano. Um dos objetivos de um banco de varejo é calcular algumas medidas de risco, como VaR e TVaR, a fim de obter embasamento estatístico ao fornecer crédito a um cliente. Os créditos não pagos ao banco requerem ações judiciais para garantir seu pagamento. Entretanto, nem todo crédito devido ao banco é passível de cobrança judicial, visto que, em determinados casos, os custos necessários para viabilizar o recebimento podem acarretar prejuízo ainda maior. Desta forma, o truncamento em um determinado ponto que o banco defina como limite para ajuizar ação fornecerá um valor mais preciso dessas medidas de risco.

Para a análise e implementação do novo modelo, será utilizado o software R. O pacote `tbs`, desenvolvido pelos autores de Ahmed et al. (2010), e o pacote `gbs` foram usados como base para o desenvolvimento do pacote `tgbs`. Esse pacote contém a FDP, FDA, funções quantil, sobrevivência, risco, gráfico QQ, estimadores de máxima verossimilhança, teste de Kolmogorov-Smirnov, VaR, TVaR, entre outros.

4.1 Análise exploratória dos dados

Inicialmente, foi observada a distribuição dos dados. Note que as quantidades monetárias estão em Pesos Mexicanos vezes 100 para simplificação dos cálculos.

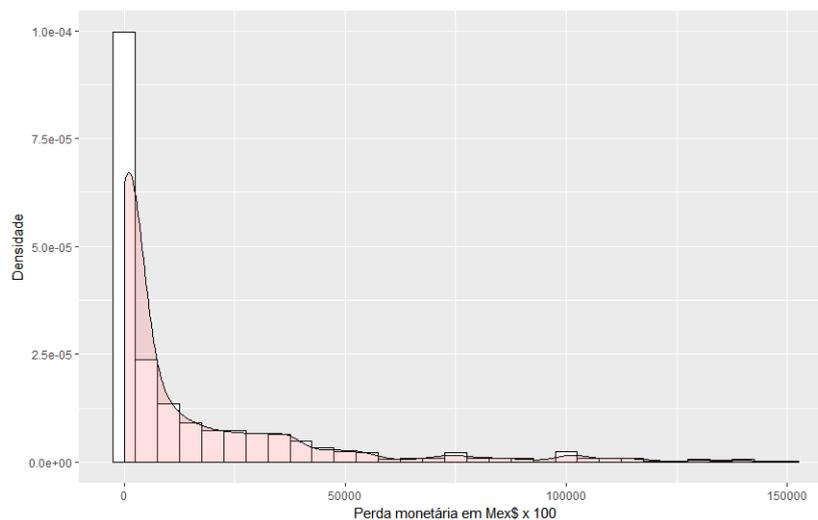


Figura 5: Distribuição dos dados não truncados

Tabela 2: Medidas descritivas dos dados

Distribuição	Min.	Med.	Média	Max.	Amp.	Des. Pad.	Coef. Var.	Assimetria	Curtose	n
Não truncada	50,22	2.531,27	15.385,01	147.656,1	147.605,9	26.177,97	1,70	2,48	9,42	1040
Truncada	264,12	6.515,37	19.196,91	147.656,1	147.392	28.001,29	1,46	2,17	7,68	832

Com base na Figura 5 e na Tabela 2, nota-se que os valores mais frequentes são créditos de baixo valor. Os dados referentes a este trabalho representam 1.040 créditos não pagos à instituição financeira em questão. Entretanto, nota-se a presença de uma forte assimetria à direita e cauda pesada (assimetria = 2,48). Além disso, sua distribuição é leptocúrtica por apresentar coeficiente de curtose igual a 9,42. Por fim, é interessante observar que 50% dos valores não pagos são inferiores a Mex\$ 2.531,27.

Diversas instituições financeiras se deparam com muitos créditos não pagos de baixa quantia e, muitas vezes, o valor gasto do banco para receber esses créditos em ações judiciais acaba sendo mais alto do que o próprio valor do crédito. Usualmente, as instituições financeiras estabelecem um limite no qual se torna efetiva a ação

judicial. Neste trabalho, foi estabelecido um limite de $\kappa = 259.55$ de pesos mexicanos pelo fato dos banco em questão analisado julgar esse valor como o limite mínimo para ajuizar uma ação contra o devedor.

Dessa forma, o total de créditos não pagos foi de 832. Na Tabela 2, também é possível observar algumas medidas descritivas dos dados truncados e não truncados. Além disso, é possível observar na Figura 6 o histograma dos dados truncados. O coeficiente de assimetria e curtose para os dados truncados foram 2,17 e 7,74, respectivamente. O desvio padrão, as medidas de dispersão e os valores extremos indicam que distribuições com caudas pesadas devem produzir um bom ajuste à distribuição desses dados.

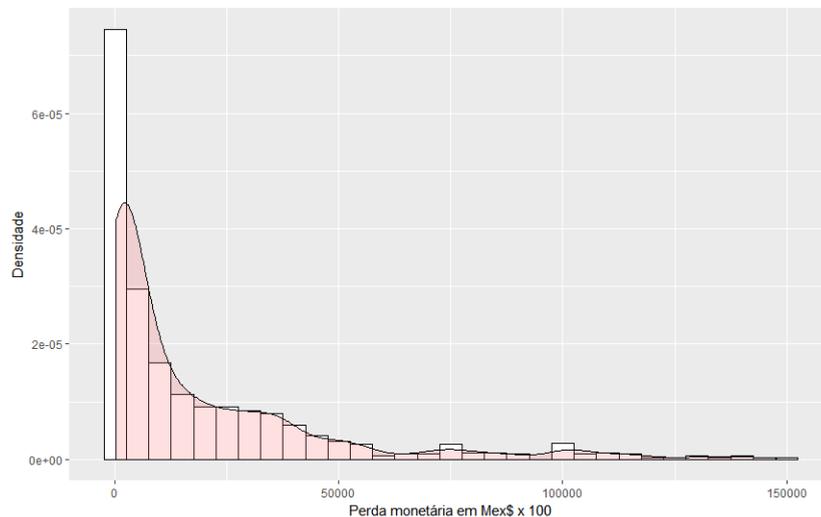


Figura 6: Distribuição dos dados truncados

Um dos objetivos neste trabalho é ajustar diferentes modelos de GBS truncadas, ou seja, a TGBS aos dados a fim de obter um melhor ajuste, visto que a curtose da TGBS é flexível.

4.2 Ajuste dos dados aos modelos

Foi aplicada a metodologia do estimador de máxima verossimilhança desenvolvido no pacote computacional da TGBS. Sendo assim, as estimativas de máxima verossimilhança para α , β e v (graus de liberdade da distribuição t-Student) são apresentadas na Tabela 3.

Tabela 3: Estimativas de máxima verossimilhança para cada modelo.

Kernel	α	β	v	logLikelihood	κ
Normal	3,6463	1849,192	-	-8706,105	259,55
t-Student	2,4066	4858,806	12	-8727,316	259,55
Logística	2,3841	1407,211	-	-8718,394	259,55
Laplace	2,2427	5371,939	-	-8788,182	259,55

Para verificar o ajuste de cada um dos quatro modelos propostos, foram verificados os gráficos QQ, realizados testes de Kolmogorov-Smirnov (KS), que compara o modelo ajustado com a distribuição teórica e foram analisados os valores de AIC e BIC, que foram obtidos da seguinte forma:

$$AIC = 2k - \ln(L(\alpha, \beta)),$$

$$BIC = \ln(n)k - \ln(L(\alpha, \beta)),$$

em que k é o número de parâmetros do modelo, n é o tamanho da amostra e $L(\alpha, \beta)$ é a verossimilhança do modelo.

Sendo assim, abaixo estão os gráficos quantil-quantil para cada modelo ajustado.

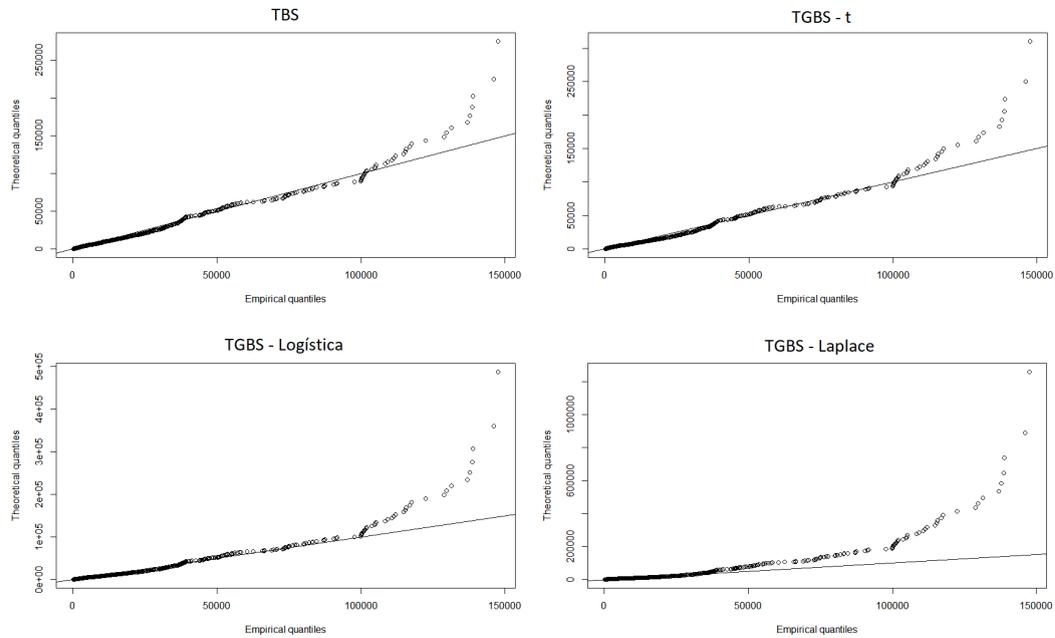


Figura 7: Gráficos QQ dos modelos ajustados.

Com base nos gráficos QQ, observa-se que o modelo ajustado com a TBS parece ter o melhor ajuste. Além disso, a tabela a seguir apresenta o AIC, BIC, estatística do KS teste e o p-valor do teste para cada modelo ajustado. O teste foi realizado sob as seguintes hipóteses:

H_0 : O modelo ajustado segue a distribuição teórica.

H_1 : O modelo ajustado não segue a distribuição teórica.

Tabela 4: AIC, BIC e KS teste para os modelos ajustados

Kernel	AIC	BIC	Estatística do teste	P-Valor
Normal	17416,21	17425,66	0,0358	0,2371
t-Student	17460,63	17474,80	0,0596	0,0054
Logística	17440,79	17450,24	0,0446	0,0727
Laplace	17580,36	17589,81	0,1061	< 0,0001

Baseado no gráficos QQ da Figura 7, nos menores valores do AIC e BIC e no p-valor do KS teste na Tabela 4, nota-se que a TGBS com núcleo normal, ou seja, a TBS produziu o melhor ajuste dos dados. Por outro lado, nota-se que, dos quatro kernels abordados neste trabalho, o kernel Laplace foi o que produziu o pior ajuste do modelo, apresentando os maiores valores de AIC, BIC e menor p-valor para o teste de KS. Considerando um nível de significância de 5% para o teste KS, a distribuição TGBS-Logística ajustou-se aos dados. Nota-se que todos os núcleos utilizados não se ajustaram bem aos dados mais extremos. Para melhorar esse ajuste, acredita-se que seria necessário um núcleo que fornecesse uma curtose mais flexível.

4.3 Análise de risco

A modelagem de dados de perda monetária é fundamental para estabelecer adequadamente indicadores de risco como VaR e TVaR. Um modelo inadequado, possivelmente levará a indicadores de risco não condizentes com a realidade e podem interferir na saúde financeira da instituição.

Como visto anteriormente, com base nos testes de ajuste do modelo, AIC, BIC e gráficos QQ, a TBS fornece o melhor ajuste. Com base na Tabela 5, nota-se que essa distribuição também forneceu o menor valor de VaR e TVaR.

Tabela 5: VaR e TVaR de 99% para cada modelo ajustado

Kernel	VaR	TVaR
Normal	149.990,3	165.259
t-Student	216.717	234.987,1
Logística	200.408,4	218.651,9
Laplace	424.170,3	442.433,5

Isso indica que as demais distribuições TGBS estão produzindo estimativas inadequadas para o VaR e TVaR pois estão superestimando seu real valor.

4. Conclusão

Neste trabalho, foi apresentada uma nova distribuição de probabilidades, a TGBS. Primeiramente, foi apresentada uma revisão teórica de alguns modelos existentes para o desenvolvimento da nova distribuição de probabilidades. Em seguida, foi proposto o modelo TGBS a partir dos modelos GBS e TBS. No novo modelo proposto, foram apresentadas algumas funções como FDP, FDA e estimador de máxima verossimilhança. Foram apresentadas algumas propriedades dos modelos conhecidos que se estendem para a TGBS. Foi desenvolvido um pacote computacional no software R para o novo modelo, o `tgbs`, que está disponível sob solicitação aos autores. Após o desenvolvimento do pacote, o modelo TGBS foi aplicado a dados reais e os ajustes para os diferentes núcleos foram comparado e analisado. Observou-se que de todos os ajustes realizados, o mais adequado foi com o núcleo da normal, entretanto, todos os modelos ajustados apresentavam falta de ajuste na cauda superior. Por fim, realizou-se a análise de risco de cada modelo proposto e verificou-se que os valores mais realistas para o VaR e TVaR são os valores do ajuste com o núcleo normal. Entretanto, como para estimar o VaR de forma precisa é necessário um bom ajuste na cauda da distribuição, provavelmente os valores apresentados para o VaR e TVaR de 99% não são boas estimativas. Porém, apesar

de não ter apresentado um bom ajuste aos dados, a TGBS pode ser amplamente utilizada em diversos casos onde o truncamento é necessário a fim de obter o melhor ajuste para o conjunto de dados.

Referências Bibliográficas

- [1] S. Ahmed, C. Castro-Kuriss, V. Leiva, and A. Flores, E.and Sanhueza. Truncated version of the birnbaum-saunders distribution with an application in financial risk. *Pakistan Journal of Statistics*, 26:293–311, 2010.
- [2] Z. W. Birnbaum and S. C. Saunders. A probabilistic interpretation of miner’s rule. *SIAM Journal of Applied Mathematics*, 16:637–652, 1968.
- [3] D.J. Davis. An analysis of some failure data. *Journal of the American Statistical Association*, 47:113–150, 1952.
- [4] J.A. Díaz-García and V. Leiva. A new family of life distributions based on elliptically contoured distributions. *Journal of Statistical Planning and Inference*, 128:445–457, 2005.
- [5] J.H.K. Kao. A graphical estimation of mixed weibull parameters in life-testing of electron tubes. *Technometrics*, 1:389–407, 1959.
- [6] V. Leiva. *The Birnbaum-Saunders Distribution*. Academic Press, New York, US, 2016.
- [7] V. Leiva, M. Barros, and G.A. Paula. *Generalized Birnbaum-Saunders Models using R*. Brazilian Statistical Association, Sao Paulo, Brazil, 2009.
- [8] V. Leiva, M. Barros, G.A. Paula, and A. Sanhueza. Generalized birnbaum-saunders distribution applied to air pollutant concentration. *Environmetrics*, 19:235–249, 2008.
- [9] W. Weibull. A statistical distribution function of wide applicability. *Journal of Applied Mechanics*, 1:293–297, 1951.