Proceeding Series of the Brazilian Society of Applied and Computational Mathematics, Vol. 3, N. 1, 2015.
Trabalho apresentado no XXXV CNMAC, Natal-RN, 2014.

Soluções Numéricas com Exatidão Máxima para as Funções Densidade de Probabilidade com Distribuições Exponencial e Pareto Intervalares

Alice F. Finger, **Aline B. Loreto**,
Programa de Pós-Graduação em Computação, Universidade Federal de Pelotas - UFPel,
96010-670, Pelotas, RS
E-mail: affinger@inf.ufpel.edu.br, aline.loreto@inf.ufpel.edu.br

Resumo: No estudo das variáveis aleatórias contínuas sobre o conjunto dos números reais, $\mathbb{R}$, um dos problemas é o cálculo de probabilidades, visto que é necessário resolver uma integral definida da função densidade que, na maioria das vezes, não possui primitiva explícita ou cuja primitiva não é simples de se obter. Considerando que integrais de funções densidade de probabilidade sejam resolvidas analiticamente, seu valor numérico é dado por aproximação e, portanto, afetado por erros de arredondamento ou truncamento. Utilizando-se intervalos para representação dos números reais, é possível controlar a propagação desses erros, pois resultados intervalares carregam consigo a segurança de sua qualidade. Para obter o valor numérico das funções densidade de probabilidade com distribuições Exponencial e Pareto se faz necessário o uso de integração numérica. O presente trabalho apresenta dois métodos para computar o intervalo solução das distribuições com entradas intervalares e analisa qual desses métodos retorna resultados mais exatos. Propõe-se as soluções destas distribuições através da primitiva da função com entradas intervalares e pelo método de Simpson Intervalar. Após a análise da qualidade dos intervalos encapsuladores para estas distribuições, verifica-se que a solução obtida através da primitiva da função com entradas intervalares apresentou exatidão máxima com controle da propagação de erros através da utilização da aritmética intervalar.

Palavras-chave: *Aritmética Intervalar, Estatística Intervalar, Qualidade do Intervalo*

1 Introdução

No processo de resolução de problemas podem ser constatadas fontes de erros, tais como: propagação dos erros nos dados iniciais, arredondamento e erros de truncamento, causados ao se truncar sequências infinitas de operações aritméticas, após um número finito de etapas. Neste contexto percebe-se a importância de técnicas intervalares. Ressalta-se que uma resposta intervalar carrega com ela a garantia de sua incerteza. Um valor pontual não carrega medidas de sua incerteza. Mesmo quando uma análise de sondagem do erro é executada, o número resultante é somente uma estimativa do erro que pode estar presente.

Os intervalos foram definidos com a finalidade de automatizar a análise do erro computacional. Através da utilização de intervalos, tem-se um controle automático de erros com limites confiáveis.

A aritmética intervalar utiliza intervalos reais para representar valores infinitos, valores desconhecidos ou para representar valores contínuos que podem ser conhecidos ou não. Os intervalos servem para representar dados inexatos, aproximações e erros de truncamento de procedimentos [8].

Na matemática intervalar, o valor real x é aproximado por um intervalo X, que possui como limites inferior ($\underline{x}$) e superior ($\bar{x}$) números de máquina de forma que o intervalo contenha x. Os

DOI: 10.5540/03.2015.003.01.0297

cálculos reais são substituídos por cálculos que utilizam a aritmética intervalar [9].

Segundo Kearfott [6], são muitas as aplicações de intervalos e nas mais diversas áreas, tais como: programação matemática, manipulação de equações, análise e projeto de circuitos elétricos, psicologia matemática, estatística, equações diferenciais, física e muitos outros.

Computar probabilidades em situações práticas envolve números e, consequentemente, problemas numéricos. Problemas numéricos na computação científica originam-se primordialmente da impossibilidade de se operar com números reais diretamente, pois tem-se que representar uma grandeza contínua (a reta real) de forma discreta (palavras de máquina) [2]. O sistema de ponto flutuante é uma aproximação prática dos números reais.

O presente trabalho apresenta dois métodos para computar o intervalo solução das distribuições com entradas intervalares e analisa qual desses métodos retorna resultados mais exatos. Propõe-se as soluções destas distribuições através da primitiva da função com entradas intervalares e pelo método de Simpson Intervalar. Após a análise da qualidade dos intervalos encapsuladores para estas distribuições, verifica-se que a solução obtida através da primitiva da função com entradas intervalares apresentou exatidão máxima com controle da propagação de erros através da utilização da aritmética intervalar.

2 Funções Densidade de Probabilidade com Distribuições Exponencial e Pareto

A função densidade de probabilidade com distribuição Exponencial é frequentemente utilizada para determinar a vida útil de equipamentos eletrônicos e do tempo entre ocorrências de eventos sucessivos, como por exemplo, o tempo entre chegadas de clientes a uma agência bancária [16].

Define-se a função densidade de probabilidade da distribuição Exponencial de parâmetro $\alpha \in \mathbb{R}$ por:

$$f(x) = \begin{cases} \alpha e^{-\alpha x}, & 0 \leq x \leq \infty, \\ 0, & x < 0. \end{cases} \tag{1}$$

com $\alpha > 0$.

A probabilidade de que um número $x \in$ (a, b) é:

$$P(a \leq x \leq b) = \alpha \int_a^b e^{-\alpha x} dx = e^{-a\alpha} - e^{-b\alpha} \tag{2}$$

Sendo a aplicação da integral a resolução de um número x estar compreendido no intervalo entre os valores a e b.

Seja $A = [a, b] \subset \mathbb{R}$. Para calcular $P(a \leq x \leq b)$, tem-se as seguintes situações possíveis:

i) Se $b \leq 0$, então $P(a < x \leq b) = 0$.

ii) Se 0 $\in$ (a, b), então $P(a < x \leq b) = P(0 \leq x \leq b) = \int_0^b \alpha e^{-\alpha x} dx$.

iii) Se $a \geq 0$, então $P(a \leq x \leq b) = \int_a^b \alpha e^{-\alpha x} dx$.

A função 1 tem primitiva na forma analítica, então, pelo Teorema Fundamental do Cálculo Integral, pode-se calcular a integral definida de $f(x)$ diretamente desta primitiva. O número e suas potências, de expoentes não nulos, são irracionais, sendo seus valores aproximados, portanto sujeitos a erros de arredondamento [13].

Para gerar o intervalo encapsulador da exponencial, Santos [13] utilizou o método de Simpson Intervalar, o qual é uma alternativa para encontrar um intervalo que aproxime o valor desta integral, com a finalidade de obter um intervalo com amplitude tão pequena quanto possível que contenha a integral. Para a aplicação do método é necessário determinar extensões intervalares para a densidade $f(x)$ e sua derivada de quarta ordem [13],

$$f^{(4)}(x) = \begin{cases} \alpha^5 e^{-\alpha x}, & 0 < x < \infty, \\ 0, & x < 0. \end{cases}$$

Como a probabilidade da variável aleatória x assumir valores negativos é 0 pode-se restringir, na aplicação do método de Simpson intervalar, o domínio de f e de $f^{(4)}$ ao conjunto $\mathbb{R}_+$. Sejam $f|\mathbb{R}_+$ e $f^{(4)}|\mathbb{R}_+$ as restrições de f e de $f^{(4)}$, isto é

$$f|\mathbb{R}_+(x) = \alpha e^{-\alpha x}, 0 \leq x < \infty, \text{ e } f^{(4)}|\mathbb{R}_+ = \alpha^5 e^{-\alpha x}, 0 < x < \infty,$$

são decrescentes e as funções intervalares

$F_E(X) = \alpha[e^{-\alpha\bar{x}}, e^{-\alpha\underline{x}}]$, $\underline{x} > 0$ e $G_E(X) = \alpha^5[e^{-\alpha\bar{x}}, e^{-\alpha\underline{x}}]$, como extensão intervalar para a função densidade $f_X|\mathbb{R}_+$ e a derivada $f_X^{(4)}|\mathbb{R}_+$.

O Princípio de Pareto ou a regra dos 80-20 que diz que, em muitos fenômenos, 80% dos efeitos vem de 20% das causas. Por exemplo, em um negócio de vendas, 80% das vendas provêm de 20% dos clientes [7].

Define-se a função de densidade de probabilidade da distribuição de Pareto, por [16]:

$$f(x) = \begin{cases} \frac{\alpha\beta^\alpha}{x^{\alpha+1}} & se \quad x \geq \beta \\ 0 & se \quad x < \beta \end{cases} \tag{3}$$

com $\alpha > 1$ e $\beta > 0$.

$$\int_{-\infty}^{+\infty} f(x)dx = \int_{\beta}^{+\infty} \frac{\alpha\beta^\alpha}{x^{\alpha+1}} dx$$

Pode-se, então, calcular a probabilidade de um valor distribuído conforme Pareto, através da integral apresentada acima.

Para computar o intervalo encapsulador da probabilidade intervalar para esta variável aleatória utiliza-se a seguinte fórmula [4]:

$$F_P(X) = \alpha \frac{c^\alpha}{X^{\alpha+1}} = \alpha \left(\left[\frac{c^\alpha}{\overline{x}^{\alpha+1}}, \frac{c^\alpha}{\underline{x}^{\alpha+1}} \right] \right), \ \ para \ \ 0 \notin X = [\underline{x}, \overline{x}], \tag{4}$$

onde α e $c \in \mathbb{R}$.

Define-se a função densidade aplicando a extensão intervalar, ou seja, para cada operação de $F_P(X)$, se é conhecido o intervalo X, computa-se a imagem através da aritmética intervalar definida por Moore [9] [10].

3 Métodos de Integração Numérica com Entradas Intervalares

Nesta seção descreve-se duas propostas para calcular as funções com distribuições Exponencial e Pareto com entradas intervalares. Primeiramente apresenta-se a solução através da primitiva da função e em seguida a solução através da aplicação do método de Simpson Intervalar.

3.1 Primitiva da função

Abaixo apresentam-se as fórmulas utilizadas como solução das duas funções através das suas primitivas.

Primitiva da função Exponencial:

$$\alpha \int_a^b e^{-\alpha X} dx = e^{-\alpha \times X} = e^{-\alpha \times \bar{x}} - e^{-\alpha \times \underline{x}}$$

DOI: 10.5540/03.2015.003.01.0297

Primitiva da função Pareto:

$$\int_{\beta}^{+\infty} \frac{\alpha\beta^{\alpha}}{X^{\alpha+1}} dx = \beta^{\alpha} \times (X^{-\alpha}) = \beta^{\alpha} \times (-\bar{x}^{-\alpha}) - \beta^{\alpha} \times (-\underline{x}^{-\alpha})$$

onde ($\underline{x}$) é o limite inferior e ($\bar{x}$) é o limite superior do intervalo X.

3.2 Método de Simpson Intervalar

A seguir, apresenta-se a definição intervalar para o método desenvolvida por Caprani *et al.* [3] e aplicada no presente trabalho.

Dada a função contínua $f : \mathbb{R} \to \mathbb{R}$ sobre o intervalo A = [a,b], o objetivo é calcular

$$S = \int_{a}^{b} f(x)dx. \tag{5}$$

Como f é contínua, f é integrável a Riemann. Se G é uma primitiva de f em [a,b] tem-se, pelo Teorema Fundamental do Cálculo Integral, que

$$S = G(b) - G(a)$$

$G(a)$ ou $G(b)$ podem não ter uma representação exata no computador, o que implica que seu valor deve ser aproximado por um número de máquina. Se não é possível encontrar uma expressão analítica para uma primitiva de f, métodos numéricos são essenciais para encontrar uma solução numérica do problema [13].

O método de Simpson Intervalar [3] é uma extensão intervalar do Método de Simpson real. O método na forma intervalar é fundamentado na propriedade aditiva da integral definida e no teorema do valor médio para integrais. Admite conhecidas as extensões intervalares inclusões monotônicas F e G de f e $f^{(4)}$, respectivamente.

Como $f(x) \in F(X)$, para todo $x \in (X \in \mathbb{R})$ então,

$$IS = \sum_{i=1}^{n} \int_{a_{i-1}}^{a_i} f(x)dx = \sum_{i=1}^{n} w(A_i)F(A_i).$$

Assim, tem-se:

$$IS_i = \int_{a_{i-1}}^{a_i} f(x)dx = \frac{w(A_i)}{6}(F(a_{i-1}) + 4F(m(A_i)) + F(a_i)) - \frac{(w(A_i))^5}{2880}G(A_i).$$

onde $F(a_{i-1})$, $F(m(A_i))$ e $F(a_i)$ são intervalos degenerados de números reais. Portanto,

$$S \subseteq IS = \sum_{i=1}^{n} IS_i. \tag{6}$$

Assumindo que G é uma extensão intervalar linear de $f^{(4)}$, tem-se

$$w(IS_i) = \frac{1}{2880} w(A_i)^5 w(G(A_i)) \leq \frac{1}{2880} w(A_i)^5 K_G w(A_i) = K_2 w(A_i)^6.$$

Se $w(A_i) = w(A)/n$ tem-se que $w(IS) = \sum_i w(IS_i) \leq K_3/n^5$, pode-se esperar que dobrando-se o valor de n reduz-se o erro de truncamento por um fator de 32. O intervalo IS depende das extensões intervalares usadas e do valor de n [13].

DOI: 10.5540/03.2015.003.01.0297

4 Análise da Qualidade dos Intervalos Encapsuladores

A linguagem escolhida para implementar os métodos foi Python [11], para a qual se utiliza o pacote intervalar IntPy [5]. Comparando-se os resultados obtidos no IntPy, IntLab, Maple Intervalar e C-XSC, o pacote desenvolvido na linguagem Python foi o que demonstrou os melhores resultados, retornando intervalos com qualidade e em tempos de processamento menores [1].

Segundo Ratschek e Rokne [12], a aritmética intervalar fornece uma ferramenta para estimar e controlar erros automaticamente. Os cálculos são executados usando intervalos ao invés de números reais e, consequentemente, a aritmética real é substituída pela aritmética intervalar. A computação com utilização de intervalos fornece as seguintes estimativas para o erro:

- *Erro Absoluto*: $|x - m(X)| < \frac{w(X)}{2}$,
  onde $m(X) = (\frac{\underline{x}+\overline{x}}{2})$ é o ponto médio do intervalo X;
- *Erro Relativo*: $\left|\frac{x-m(X)}{x}\right| \leq \frac{w(X)}{2min|X|}$ se $0 \notin X$.

Aplicam-se estas medidas de erros nos intervalos encapsuladores obtidos para as funções densidade de probabilidade com distribuições Exponencial e Pareto com o objetivo de verificar a qualidade do intervalo solução, obtido após o processamento de operações aritméticas intervalares.

Todos os resultados intervalares e reais que serão apresentados, utilizaram o sistema de ponto flutuante F(10, 14, -10, 10).

4.1 Exponencial

Para a Exponencial foram propostas duas implementações para a forma real e intervalar. A primeira utilizando a resolução pelo Método 1/3 de Simpson e Simpson Intervalar, e a segunda a partir da primitiva da função com entradas reais e intervalares.

Apresentam-se na Tabela 1 os valores obtidos em cada uma das soluções utilizadas para a forma real e para a forma intervalar, bem como o tempo (em segundos) de processamento para obter cada solução. **Exemplo:** Intervalo A = [20, 50], $\alpha = 0.01$, subintervalos = 50

Tabela 1: Aplicação com a primitiva da função e com o método de Simpson nas formas real e intervalar

	Resultado	**Tempo (s)**
Primitiva Real	0.21220009336535	0.00003
Primitiva Intervalar	[0.21220009336535, 0.21220009336536]	0.00036
Simpson Real	0.21220009336642	0.00006
Simpson Intervalar	[0.21220009336542, 0.21220009336543]	0.00238

Pelos valores apresentados na Tabela 1, nota-se que a resolução sem o método de Simpson apresentou um resultado melhor se analisarmos o tempo de processamento e o intervalo encapsulador.

Tabela 2: Qualidade dos intervalos encapsuladores para a função com distribuição Exponencial

	Diâmetro	**Erro Absoluto**	**Erro Relativo**
Primitiva	0.00000000000001	$4.99600361081\times10^{-15} < 4.99600361081\times10^{-15}$	$2.35438332358\times10^{-14} \leq 2.35438332358\times10^{-14}$
Simpson	0.00000000000001	$4.99600361081\times10^{-15} < 4.99600361081\times10^{-15}$	$2.35438332358\times10^{-14} \leq 2.35438332358\times10^{-14}$

No cálculo do erro absoluto, verifica-se que a diferença entre o valor exato e o ponto médio do intervalo encapsulador com a solução pela primitiva da função e pelo método de Simpson Intervalar apresenta variação na 16ª casa decimal. E o resultado da divisão do diâmetro do intervalo pela metade, também ocorre variação na 16ª casa decimal para ambas as soluções.

Para o erro relativo, as duas soluções apresentam diferença apenas na 15ª casa decimal. Assim, a qualidade do intervalo é garantida através das estimativas de erros calculados.

DOI: 10.5540/03.2015.003.01.0297

4.2 Pareto

A implementação da distribuição de Pareto foi realizada utilizando-se a primitiva da função e o Método 1/3 de Simpson. Assim, apresentam-se duas maneiras para computar o valor com entradas reais e duas para computar o intervalo solução para esta distribuição com exemplo aplicado em Varjão [15]. **Exemplo:** Intervalo = [1.0, 2.0], $\alpha = 0.25$, $c = 1.0$

Tabela 3: Aplicação com a primitiva da função e com o método de Simpson nas formas real e intervalar

	Resultado	**Tempo (s)**
Primitiva Real	0.15910358474629	0.00006
Primitiva Intervalar	[0.15910358474629, 0.15910358474630]	0.00039
Simpson Real	0.15910358666866	0.00007
Simpson Intervalar	[0.15910357486653, 0.15910359486654]	0.08407

A Tabela 3 mostra que o valor real está contido no intervalo solução gerado pelas duas implementações. Além disso, é possível verificar que o tempo de processamento para computar o intervalo encapsulador utilizando o Método 1/3 de Simpson é maior.

A partir desses dados, apresenta-se então a Tabela 4, a qual é uma análise da qualidade dos intervalos obtidos no presente trabalho.

Tabela 4: Qualidade dos intervalos encapsuladores para a função com distribuição Pareto

	Diâmetro	**Erro Absoluto**	**Erro Relativo**
Primitiva	0.00000000000001	$4.99600361081\times10^{-15} < 4.99600361081\times10^{-15}$	$3.14009493801\times10^{-14} \leq 3.14009493801\times10^{-14}$
Simpson	0.00000002000000	$1.80212500478\times10^{-9} < 1.00000049907\times10^{-8}$	$1.13267402861\times10^{-8} \leq 6.28521703496\times10^{-8}$

Verifica-se, através do cálculo do diâmetro, que o método de resolução utilizando a primitiva da função resulta em intervalos com menor comprimento.

A estimativa do erro absoluto é bem menor que na solução que utiliza o método de Simpson. Quanto ao erro relativo, pode-se afirmar que a mesma característica é mantida, comprovando que o resultado do intervalo encapsulador utilizando a primitiva da função como solução apresenta uma qualiddade melhor quanto a estimativa de erros.

Assim, foi possível concluir que a implementação utilizando a primitiva da função sempre se mostra de melhor qualidade em todas as estimativas.

5 Conclusões

Embora integrais de funções densidade de probabilidade como a Exponencial e a de Pareto, sejam resolvidas analiticamente, seu valor numérico no computador é dado por aproximação, e portanto afetado por erros de arredondamento ou truncamento.

Quando se trabalha com computação numérica, um dos fatores de maior importância é a exatidão da resposta desses cálculos. O que sempre se procura são resultados cada vez mais precisos e com um menor erro possível contido neles.

A matemática intervalar surge com o objetivo principal de realizar um controle automático de erros dos cálculos, retornando respostas com a maior exatidão possível.

A implementação das funções com entradas reais foi realizada na linguagem Python. Para a forma intervalar, também utilizou-se a mesma linguagem, a qual possui um pacote para programação com intervalos, chamado IntPy [5]. A escolha da linguagem se deu a partir de comparações realizadas entre alguns ambientes de programação intervalar [1].

Com o sistema de ponto flutuante F(10, 14, -10, 10) (ou com quatorze casas decimais) verificou-se, através das medidas de erros (absoluto e relativo) e diâmetro do intervalo, que todos os resultados para as probabilidades reais estão contidos nos intervalos encapsuladores.

O objetivo deste trabalho foi mostrar a importância e justificativa de se utilizar a matemática intervalar no cálculo de intervalos encapsuladores para as funções densidade de probabilidade

DOI: 10.5540/03.2015.003.01.0297

das variáveis aleatórias contínuas com distribuições Exponencial e Pareto. A partir da aplicação das soluções utilizando a primitiva da função e o método de Simpson Intervalar, foi possível concluir através da análise da qualidade dos intervalos que o método que utiliza a primitiva da função, resulta em intervalos encapsuladores bons e com qualidade, o que justifica o uso da aritmética intervalar para essas funções.

Referências

[1] Balboni, M. D. C.; Tortelli, L. M.; Lorini, M.; Furlan, V. S.; Finger, A. F.; Loreto, A. B, "Critérios para Análise e Escolha de Ambientes Intervalares". *Revista Jr de Iniciação Científica em Ciências Exatas e Engenharia*, Rio Grande - RS, n.7, 2014.

[2] Campos, M. A, "Uma Extensão Intervalar para a Probabilidade Real", Tese de Doutorado, Universidade Federal de Pernambuco, Recife, 1997.

[3] Caprani, O.; Madsen, K.; Nielsen, H. B, "Introduction to interval analysis". *IMM - Informatics and Mathematical Modelling*, 2002.

[4] Finger, A. F, "Extensão Intervalar para as Variáveis Aleatórias com Distribuições Uniforme, Normal, Gama, Exponencial e Pareto", Dissertação de Mestrado, Universidade Federal de Pelotas, 2014.

[5] IntPy, Disponível em: http://pypi.python.org/pypi/IntPy/0.1.3.

[6] Kearfott, R. B, "Interval computations: Introduction, uses and resources", *Euromath Bulletin*, v.2, pp.95-112, 1996.

[7] Lapponi, J. C, "Estatística Usando Excel", São Paulo: Treinamento Editora Ltda, 1995.

[8] Loreto, A. B, "Análise da Complexidade Computacional de Problemas de Estatística Descritiva com Entradas Intervalares", Tese de Doutorado, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2006.

[9] Moore, R. E, "Interval Analysis", Englewood Cliffs: Prentice Hall, 1966.

[10] Moore, R. E, "Methods and Applications of Interval Analysis", SIAM, 1979.

[11] Python, Disponível em: http://www.python.org.

[12] Ratschek, H.; Rokne, R, "New Computer Methods for Global Optimization", Ellis Horwood, 1988.

[13] Santos, M. G, "Probabilidades Autovalidáveis para as Variáveis Aleatórias Exponencial, Normal e Uniforme", Tese de Doutorado, Universidade Federal de Pernambuco, Recife, 2010.

[14] Sunaga, T, "Theory of an Interval Algebra and its Applications to Numerical Analysis", RAAG Memoirs, 1958.

[15] Varjão, F. R. G, "IntPy: Computação Científica Auto Validável em Python. Dissertação de Mestrado, Universidade Federal de Pernambuco, Recife, 2011.

[16] Vilches, M. A, "Cálculo para Economia e Administração", Disponível em: http://www.ime.urj.br/$\sim$ *calculo*.

