Tema
Tendências em Matemática Aplicada e Computacional, **19**, N. 1 (2018), 33-42

www.scielo.br/tema
doi: 10.5540/tema.2018.019.01.0033

Análise do Esforço Computacional das Funções Densidade de Probabilidade com Diferentes Distribuições[†]

A. FINGER[1*] e A. LORETO[2]

Recebido em Novembro 20, 2016 / Aceito em Outubro 23, 2017

RESUMO. Quando se trabalha com números de ponto flutuante o resultado é apenas uma aproximação de um valor real e erros gerados por arredondamentos ou por instabilidade dos algoritmos podem levar a resultados incorretos. Não se pode afirmar a exatidão da resposta estimada sem o auxílio de uma análise de erro. 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 das variáveis aleatórias contínuas com distribuições Uniforme, Exponencial, Normal, Gama e Pareto se faz necessário o uso de integração numérica, uma vez que a primitiva da função nem sempre é simples de se obter. Além disso, o resultado é obtido por aproximação e, portanto, afetado por erros de arredondamento ou truncamento. Neste contexto, o presente trabalho possui como objetivo analisar a complexidade computacional para computar as funções densidade de probabilidade com distribuições Uniforme, Exponencial, Normal, Gama e Pareto nas formas real e intervalar. Assim, certifica-se que ao utilizar aritmética intervalar para o cálculo da função densidade de probabilidade das variáveis aleatórias com distribuições, é possível obter um controle automático de erros com limites confiáveis, e, no mínimo, manter o esforço computacional existente nos cálculos que utilizam a aritmética real.

Palavras-chave: Aritmética Intervalar, complexidade computacional, probabilidade.

1 INTRODUÇÃO

Na matemática intervalar, o valor real x é aproximado por um intervalo X, que possui $\underline{x}$ e $\bar{x}$ como limites inferior e superior, de forma que o intervalo contenha x. O tamanho deste intervalo pode ser usado como medida para avaliar a qualidade de aproximação [11]. Os cálculos reais são substituídos por cálculos que utilizam a aritmética intervalar [7].

[†]Trabalho apresentado no XXXVI Congresso Nacional de Matemática Aplicada e Computacional.

*Autor correspondente: Alice Finger – E-mail: alicefinger@unipampa.edu.br.

[1]Unipampa - Universidade Federal do Pampa, Campus Alegrete, 97546-550, Alegrete, RS, Brasil. Programa de Pós-Graduação em Computação, UFPel - Universidade Federal de Pelotas, Pelotas, RS, Brasil.

[2]Universidade Federal de Santa Maria, Campus Cachoeira do Sul, 96506-322, Cachoeira do Sul, RS, Brasil. E-mail: aline.loreto@ufsm.br

A análise intervalar surgiu com o objetivo inicial de controlar a propagação de erros numéricos em procedimentos computacionais. Mas, aparentemente, a matemática intervalar duplica o problema de representação dos números reais em processadores numéricos, uma vez que ao invés de operar com um número real, operam-se com dois. Entretanto, sua realização é feita por meio de números de ponto flutuante, isto é, os extremos do intervalo X são números de máquina $\underline{x}_{pf}$ e $\bar{x}_{pf}$ [9, 8].

Intervalos automatizam a análise do erro computacional, ou seja, através de sua utilização tem-se um controle automático de erros com limites confiáveis.

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 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.

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.

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) [1]. O sistema de ponto flutuante é uma aproximação prática dos números reais.

Considerando que o controle do erro numérico é realizado através do uso de intervalos ao invés de números reais, Kulisch [5] e Kulisch e Miranker [6] propuseram que a implementação da aritmética intervalar seja realizada através da chamada aritmética de exatidão máxima, o que significa a busca para que resultados numéricos ou sejam números de pontos flutuantes ou estejam entre dois números de pontos flutuantes consecutivos.

Dentre a diversas soluções ou implementações para determinado problema, muitas vezes é preciso escolher a mais eficiente. Para isso, existe a análise de algoritmos, a qual tem como objetivo melhorar, se possível, seu desempenho e escolher entre os algoritmos disponíveis o melhor. Existem vários critérios de avaliação de um algoritmo como: quantidade de trabalho requerido, quantidade de espaço requerido, simplicidade, exatidão de resposta e otimalidade [13].

Preocupados se a quantidade de trabalho dispendido pelo algoritmo aumenta ao utilizar intervalos para controlar a propagação de erros, o presente trabalho realiza a análise da complexidade para

computar as funções densidade de probabilidade das variáveis aleatórias com distribuições Uniforme, Exponencial, Normal, Pareto e Gama, nas formas real e intervalar. Através desta análise, justifica-se que ao utilizar intervalos para calcular tais funções é possível ter um controle automático de erros com limites confiáveis e manter o esforço computacional (ou quantidade de trabalho requerido) quando calculado com as funções na sua forma real.

2 DISTRIBUIÇÕES E MÉTODOS DE RESOLUÇÃO

Como no cálculo de probabilidade das variáveis aleatórias contínuas é preciso resolver uma integral e nem sempre a primitiva é simples de se obter, torna-se necessário utilizar outros métodos de resolução de integral.

As funções com entradas reais, nas quais a primitiva da função era simples de se obter, foram utilizadas duas soluções, onde uma calcula o resultado através da primitiva da função e a outra realiza o cálculo através do método 1/3 de Simpson [12]. Já naquelas onde a primitiva não era explícita utiliza-se somente o método 1/3 de Simpson. Para as funções com entradas intervalares em que a primitiva era simples de se obter utilizou-se a primitiva da função na forma intervalar e o método de Simpson Intervalar definido por Caprani *et al* [2] para obter o intervalo solução. Já aquelas funções em que a primitiva não é explícita, os intervalos encapsuladores foram calculados através do método de Simpson Intervalar.

Abaixo, é listada cada função com suas respectivas formas de resolução, primeiro para entradas reais e depois para as intervalares.

- Distribuição Uniforme: a=0 e b=1
  - Real
    - Primitiva da função:
$$\int_a^b f(x)dx = \frac{1}{b-a}$$
  - Intervalar
    - Primitiva da função:
$$\int_a^b f(X)dx = \bar{x} \times \frac{1}{(b-a)} - \underline{x} \times \frac{1}{(b-a)}$$
- Distribuição Exponencial: $\alpha > 0$
  - Real
    - Primitiva da função:
$$\alpha \int_a^b e^{-\alpha x}dx = e^{-\alpha \times x} = e^{-\alpha \times a} - e^{-\alpha \times b}$$

        * 1/3 de Simpson
    - Intervalar
        * Primitiva da função:

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

        * Simpson Intervalar

- Distribuição Normal padronizada: $f_x(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$, com $\mu = 0$ e $\sigma = 1$
    - Real
        * 1/3 de Simpson
    - Intervalar
        * Simpson Intervalar
- Distribuição de Pareto: $\alpha > 1$ e $\beta > 0$
    - Real
        * Primitiva da função:

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

        * 1/3 de Simpson
    - Intervalar
        * Primitiva da função:

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

        * Simpson Intervalar
- Distribuição Gama:

$$f(x) = \begin{cases} \frac{\lambda^\nu}{\Gamma(\nu)} e^{-\lambda x} x^{\nu-1}, & se \;\; x > 0 \\ 0, & outro \;\; caso. \end{cases}$$

    - Real
        * 1/3 de Simpson
    - Intervalar
        * Simpson Intervalar

A análise da complexidade é realizada considerando os métodos 1/3 de Simpson e Simpson Intervalar, os quais são aplicados em quase todas as distribuições, com exceção da Uniforme.

3 ANÁLISE DO ESFORÇO COMPUTACIONAL

O termo complexidade, no contexto de algoritmos, refere-se aos requerimentos de recursos necessários para que um algoritmo possa resolver um problema sob o ponto de vista computacional, ou seja, à quantidade de trabalho despendido pelo algoritmo [13]. Quando o recurso é o tempo, são escolhidas uma ou mais operações fundamentais e então são contados os números de execuções desta operação fundamental na execução do algoritmo. Segundo Toscani [13] a escolha de uma operação como operação fundamental é aceitável se o número de operações executadas pelo algoritmo é proporcional ao número de execuções da operação fundamental.

A complexidade também pode ser vista como uma propriedade do problema, o que significa dar uma medida independente do tratamento dado ao problema, independente do caminho percorrido na busca da solução, portanto independente de algoritmos [13].

Questões relativas à complexidade de um algoritmo em termos do tempo de computação e espaço de memória são determinantes para o julgamento da eficiência do mesmo [13]. Um algoritmo, para ser razoável ou não, vai depender de quantos passos computacionais ele necessita para chegar a solução de um problema. Um algoritmo é considerado razoável quando obtém a solução de um problema em tempo polinomial [4].

Sob um ponto de vista computacional, Garey e Johnson [3] descrevem informalmente um problema como uma questão genérica a ser respondida, geralmente possuindo vários parâmetros, ou variáveis livres.

Um problema é chamado de problema computável se existir um procedimento efetivo que o resolva em um número finito de passos, ou seja, se existe um algoritmo que leve à sua solução. Observa-se, contudo, que um problema considerado “em princípio” computável pode não ser tratável na prática, devido às limitações dos recursos computacionais para executar o algoritmo implementado [13].

Se existe um algoritmo de tempo polinomial que resolve todas as instâncias de um problema, este problema é tratável, caso contrário diz-se que é intratável [13].

Nas subseções a seguir, apresentam-se as ordens de complexidade encontradas para cada uma das funções de densidade de probabilidade utilizadas no presente trabalho. Primeiramente, realiza-se a análise para os algoritmos com entradas reais, uma vez que não se encontrou na literatura esse tipo de análise. Em seguida, apresenta-se a análise dos algoritmos intervalares que utilizam a primitiva da função, e, posteriormente, a análise da complexidade de cada distribuição que utiliza o método de Simpson Intervalar.

3.1 Distribuições com Entradas Reais

Nesta subseção apresenta-se a análise da complexidade computacional realizada para as funções densidade de probabilidade com entradas reais, utilizando a primitiva da função e o método 1/3 de Simpson para resolução da integral.

Tabela 1: Análise de complexidade computacional para funções com entradas reais.

Distribuição	Primitiva Real	1/3 de Simpson
Uniforme	É preciso fornecer os intervalos necessários para o cáculo da distribuição. A resolução requer operações aritméticas, como subtração e divisão. A complexidade computacional de resolver tais operações é de ordem constante, logo $O(1)$.	-
Exponencial	Parâmetro α e o intervalo em que a probabilidade do valor a ser calculado deve estar contido são passados como entrada. Realiza operações aritméticas para resolução da integral. Assim, a complexidade é de ordem $O(1)$.	Parâmetro α e o intervalo em que a probabilidade do valor a ser calculado deve estar contido são fornecidos, além do número de subdivisões n. Executa operações aritméticas e operações em um laço de repetição que deve ser executado n vezes. Assim, a complexidade é de ordem $O(n)$.
Normal	-	Entrada de dados e valor das subdivisões no cálculo do método. Utiliza operações aritméticas e um laço que é repetido o número de subdiviões definidos. Assim, sua complexidade depende do número de subdivisões n, obtendo ordem de complexidade linear, $O(n)$.
Pareto	Parâmetros α e β, bem como o intervalo em que se quer a probabilidade são fornecidos. Realizado o cálculo de uma integral, na qual são resolvidas operações aritméticas como soma, multiplicação, divisão e exponenciação. Assim, a complexidade é de ordem constante, $O(1)$.	Parâmetros como entrada, bem como o número de subdivisões a ser utilizado no método. Resolvidas operações aritméticas como soma, multiplicação, divisão e exponenciação, seguida de um laço, o qual é repetido em função do número de subdivisões. Assim, a complexidade é de ordem linear, $O(n)$.
Gama		Parâmetros λ e v, além de uma variável n a qual armazena o número de subdivisões a ser utilizada no cálculo do método. Realizadas operações aritméticas de subtração, multiplicação e exponenciação e um laço, o qual é repetido n vezes. Portanto, a ordem de complexidade do algoritmo é linear, $O(n)$.

Para o método de 1/3 Simpson é preciso fornecer como dados de entrada os parâmetros bem como o número de subdivisões a ser utilizado no método. Após, são resolvidas operações aritméticas seguida de um laço, o qual é repetido em função do número de subdivisões. A complexidade computacional quando se utiliza o método 1/3 de Simpson não depende da entrada, pois esta é sempre a mesma, mas depende das subdivisões do método. A eficiência do algoritmo está relacionada com o tamanho dessas subdivisões que deve ser passado como parâmetro na execução do método, portanto a complexidade é de ordem $O(2^n)$.

3.2 Entradas Intervalares

A seguir, na Tabela 2 são apresentadas as análises de complexidade dos algoritmos desenvolvidos para as funções com entradas intervalares.

Tabela 2: Análise de complexidade computacional para funções com entradas intervalares.

Distribuição	Primitiva Intervalar	Simpson Intervalar
Uniforme	Utiliza as mesmas entradas do algoritmo real. Operações aritméticas realizadas na forma real foram substituídas pelas operações intervalares. A complexidade de resolver tais operações é constante. Dessa maneira, a complexidade computacional permanece constante, $O(1)$.	-
Exponencial	É utilizado o parâmetro α e as entradas reais são substituídas por intervalares. Operações aritméticas reais substituídas pelas intervalares. Complexidade é de ordem $O(1)$.	Realiza operações aritméticas intervalares e parâmetros de entrada que são utilizados também pela forma intervalar sem Simpson. A diferença é que é criada uma lista com todas as n subdivisões, o que acarreta em um laço executado n vezes. Por fim, o algoritmo apresenta mais um laço, no qual são realizados os cálculos do método. Assim, a complexidade computacional do algoritmo é de ordem linear, $O(n)$.
Normal	-	Utiliza os parâmetros da distribuição, além do número de subdivisões do método. É criada uma lista com as n distribuições, após é realizado um laço para calcular o resultado com o método de Simpson. Portanto, a complexidade é $O(n)$.
Pareto	Parâmetros α e β, bem como o intervalo em que se quer a probabilidade como entrada. Após, operações aritméticas foram substituídas pelas intervalarares. Assim, a complexidade computacional é de ordem constante, $O(1)$.	Mesmos dados de entrada sem o método, além do número de subdivisões que serão realizadas. Assim, também existem dois laços, um para criar a lista de subdivisões e o outro para realizar o cálculo do método. A complexidade é de ordem linear, $O(n)$.
Gama	-	Dados de entrada da distribuição e um laço de repetição, o qual cria uma lista de acordo com o número de subdivisões selecionadas. A finalização do algoritmo se dá com um laço de repetição que realiza os cálculos do método de acordo com o número n de subdivisões. Complexida é $O(n)$.

A forma intervalar de Simpson realiza operações aritméticas intervalares e recebe parâmetros de entrada que são utilizados também pela forma intervalar sem Simpson. A diferença em utilizar o método é que ele precisa criar uma lista com todas as n subdivisões, o que acarreta em um laço o qual é executado n vezes. Por fim, o algoritmo apresenta mais um laço, no qual são realizados os cálculos do método.

Assim como na complexidade do método 1/3 de Simpson, a complexidade computacional quando se utiliza o método de Simpson Intervalar também não depende da entrada, e sim do número de subdivisões para o método. Portanto, a eficiência do algoritmo está relacionada com o tamanho dessas subdivisões, sendo classificado como $O(2^n)$.

A Tabela 3 apresenta, resumidamente, a ordem de complexidade encontrada para todas as distribuições exploradas no presente trabalho.

Tabela 3: Esforço computacional das Distribuições.

Distribuição	**Complexidade Real**		**Complexidade Intervalar**	
	Primitiva da Função	Simpson	Primitiva da Função	Simpson Intervalar
Uniforme	$O(1)$	-	$O(1)$	-
Exponencial	$O(1)$	$O(2^n)$	$O(1)$	$O(2^n)$
Normal	-	$O(2^n)$	-	$O(2^n)$
Gama	-	$O(2^n)$	-	$O(2^n)$
Pareto	$O(1)$	$O(2^n)$	$O(1)$	$O(2^n)$

A partir da análise da complexidade computacional dos algoritmos propostos para computar as funções com entradas reais, é possível afirmar que utilizando a primitiva da função como solução, os algoritmos são executados em uma complexidade menor, ou seja, menos trabalho para computar o resultado. Com as funções na forma intervalar o método de Simpson Intervalar gera resultados utilizando um maior esforço computacional do que os obtidos a partir da primitiva da função. Assim, conclui-se que em ambas as formas, real e intervalar, as melhores soluções são obtidas através da aplicação da primitiva da função.

4 CONCLUSÃO

Embora integrais de funções densidade de probabilidade como a Uniforme, 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. Outras funções densidade como a Normal ou Gama não possuem primitivas na forma analítica, sendo necessário o uso de integração numérica onde erros de arredondamentos e truncamentos são propagados devido às operações aritméticas realizadas no computador.

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 exatos 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.

Verificou-se ainda se, ao utilizar intervalos para calcular a função densidade de probabilidade das variáveis aleatórias contínuas, a quantidade de trabalho despendido pelo algoritmo aumenta em relação a forma real. Após a análise, constata-se que o esforço computacional é o mesmo, tanto

na forma real quanto na intervalar. Resultado importante, o qual justifica o uso da matemática intervalar na resolução das funções. A aplicação de intervalos proporciona o controle de erros e exatidão dos resultados para estas variáveis.

Como resultado foi possível analisar se, ao utilizar intervalos para calcular a função densidade de probabilidade das variáveis aleatórias contínuas, a quantidade de trabalho despendido pelo algoritmo aumenta em relação a forma real. Após a análise, constatou-se que o esforço computacional é o mesmo, tanto na forma real quanto na intervalar. Resultado importante, o qual justifica o uso da matemática intervalar na resolução das funções. A aplicação de intervalos proporciona o controle de erros e exatidão dos resultados para estas variáveis.

ABSTRACT. When working with floating point numbers the result is only an approximation of a real value and errors generated by rounding or by instability of the algorithms can lead to incorrect results. We can't affirm the accuracy of the estimated answer without the contribution of an error analysis. Using intervals for the representation of real numbers, it is possible to control this error propagation, because intervals results carry with them the security of their quality. To obtain the numerical value of the probability density functions of continuous random variables with distributions Uniform, Exponential, Normal, Gamma and Pareto is necessary to use numerical integration, once the primitive of the integral do not always is simple to obtain. Moreover, the result is obtained by approximation and therefore affected by truncation or rounding errors. Moreover, the result is obtained by approximation and therefore affected by truncation or rounding errors. In this context, this paper has aims to analyze the computational complexity to compute the probability density functions with Uniform, Exponential, Normal, Gamma and Pareto distributions in the real and interval forms. Thus, make sure that by using interval arithmetic to calculate the probability density function of the random variables with distributions, it is possible to have an automatic error control with reliables boundaries, and, at least, keep the existing computational effort in the calculation using the real arithmetic.

Keywords: Interval arithmetic, computational complexity, probability.

REFERÊNCIAS

[1] M.A. Campos. *Uma Extensão Intervalar para a Probabilidade Real*. Tese de doutorado, Universidade Federal de Pernambuco, Recife, (1997).

[2] O. Caprani, K. Madsen & H.B. Nielsen. Introduction to interval analysis. *IMM - Informatics and Mathematical Modelling*, Dinamarca, (2002).

[3] M.R. Garey & D.S. Johnson. *Computers and intractability: a guide to the theory of NP-completeness*. Freeman, San Francisco, (1979).

[4] V. Kreinovich, A. Lakeyev, J. Rohn & P. Kahl. *Computational Complexity and Feasibility of Data Processing and Interval Computations*. Dordrecht, Kluwer, (1998).

[5] U.W. Kulisch. Complete interval arithmetic and its implementation on the computer. In *Numerical Validation in Current Hardware Architectures*, Springer, **5492** (2008).

[6] U. Kulisch & L. Miranker. *Computer Arithmetic in Theory and Practice*. Academic Press, New York, (1981).

[7] R.E. Moore. *Interval Analysis*. Prentice Hall, Englewood Cliffs, NJ, (1966).

[8] R.E. Moore, M. Kearfott & J. Cloud. *Introduction to Interval Analysis*. Studies in Applied and Numerical Mathematics (SIAM), Philadelphia, (2009).

[9] R.E. Moore. *Methods and Applications of Interval Analysis*. Studies in Applied and Numerical Mathematics (SIAM), Madison, Wisconsin, (1979).

[10] M. Naghettin & E. Pinto. *Hidrologia Estatística*. CPRM Serviço Geológico do Brasil, Belo Horizonte, (2007).

[11] H. Ratschek & J. Rokne. *New Computer Methods for Global Optimization*. Limited, Chichester, United Kingdom, (1988).

[12] M.A.G. Ruggiero & V.L.R. Lopes. *Cálculo numérico: aspectos teóricos e computacionais*. Pearson Makron Books, São Paulo, (1996).

[13] L. Toscani & P. Veloso. *Complexidade de Algoritmos: análise, projetos e métodos*. Sagra-Luzzato, Porto Alegre, (2001).