

ENCAPSULANDO PROBABILIDADES DE  
NORMAIS REDUZIDAS**Maria das Graças dos Santos**Doutorado em Matemática Computacional – UFPE,  
Rua Prof<sup>o</sup> Luiz Freire, s/n, Cidade Universitária,  
50740-540, Recife – Pe,  
E-mail: [batgal60@yahoo.com.br](mailto:batgal60@yahoo.com.br).**Marcília Andrade Campos,**Centro de Informática UFPE,  
Rua Prof<sup>o</sup> Luiz Freire, s/n, Cidade Universitária,  
50740-540, Recife – Pe,  
E-mail: [mac@cin.ufpe.br](mailto:mac@cin.ufpe.br)

**Resumo:** *No estudo das variáveis aleatórias contínuas sobre o conjunto dos números reais,  $R$ , 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 obter. Funções densidade como a Normal, por exemplo, não possuem primitivas na forma analítica, sendo necessário o uso de integração numérica onde erros de arredondamento e truncamento são propagados devido às operações aritméticas realizadas no computador.*

*O objetivo deste trabalho é utilizar o método de Simpson Intervalar definido por Caprani para calcular um intervalo encapsulador da probabilidade real de uma variável aleatória, com distribuição Normal Reduzida. Os algoritmos foram implementados no Matlab, usando Intlab. Os resultados mostraram que as probabilidades intervalares encapsularam as probabilidades reais.*

**Palavras-chave:** Matemática Intervalar, Variável Aleatória, intervalo encapsulador.

## 1. Introdução

No estudo das variáveis aleatórias contínuas sobre o conjunto dos números reais  $R$  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 obter. Algumas funções densidade de probabilidade como a normal ou gama, por exemplo, 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 no computador. Existem na literatura [1] vários métodos de integração numérica: os métodos conhecidos como fórmulas de Newton-Cotes, Quadratura Gaussiana, entre outros. Como a integral definida é aproximada se faz necessário o cálculo do erro nessa aproximação e que também é um valor aproximado.

A Matemática Intervalar fornece uma alternativa para resolver este problema: encontra um intervalo que encapsula o valor da integral. Extensões intervalares de vários métodos de integração numérica, tais como Fórmula de Newton-Cotes, Quadratura Gaussiana, são citados em Caprani [3] e em Moore [9].

O objetivo deste trabalho é utilizar o método de Simpson Intervalar definido em Caprani [3] para calcular a probabilidade de uma variável aleatória, com distribuição Normal Reduzida, no intervalo  $[a, b]$ . Os algoritmos foram implementados no Matlab [5], usando Intlab [4].

Este artigo está organizado da seguinte forma: a Seção 2 apresenta conceitos básicos na análise intervalar. A Seção 3 contém os resultados reais e intervalares sobre integração numérica. A Seção 4 é a grande contribuição deste trabalho; nele está descrito um método para se obter um intervalo encapsulador para uma variável aleatória com distribuição Normal Reduzida e resultados numéricos decorrentes da

aplicação do método. As conclusões estão na Seção 5 e por último as referências bibliográficas.

## 2. Conceitos Básicos na Análise Intervalar

Seja  $IR$  o conjunto de todos os intervalos fechados de números reais,

$$IR = \{ [x_1, x_2] / x_1, x_2 \in \mathbb{R}, x_1 \leq x_2 \}.$$

Associando-se a cada intervalo  $[x, y] \in IR$  um ponto  $(x, y) \in \mathbb{R}^2$ , obtém-se uma representação geométrica Moore [8] para  $IR$ , como mostra a Figura 1, obtida no Maple® [7] através do procedimento

**geoIR (x, y, n1, n2);**

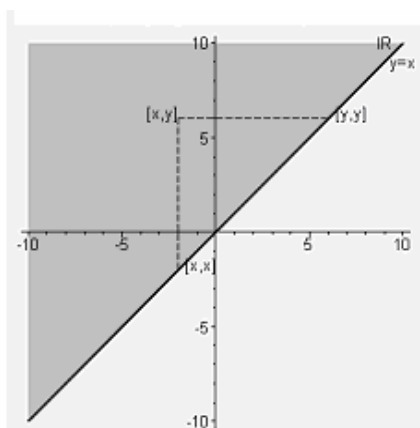


Figura 1: Interpretação geométrica do conjunto  $IR$ .

Os pontos na diagonal  $y = x$  entre os pontos  $(x, x)$  e  $(y, y)$  representam os números reais contidos nos intervalos  $[x, y]$  e o intervalo é representado pelo vértice  $(x, y)$  do triângulo.

A aritmética intervalar e a topologia no espaço dos intervalos são as definidas em Moore [8] [9].

**Definição 1** Uma função intervalar  $F$  de uma variável intervalar  $X$  é **inclusão monotônica** se

$$Y \subseteq X \Rightarrow F(Y) \subseteq F(X).$$

**Definição 2** Seja  $f$  uma função real de uma variável real. A função intervalar  $F$  de uma variável intervalar  $X$  é uma **extensão intervalar de  $f$**  se

$$F(x) = f(x), \quad x \in X.$$

Existem várias extensões intervalares de uma função real  $f$ . Isto acontece porque expressões racionais que são equivalentes na aritmética real podem não o ser na aritmética intervalar.

**Exemplo 1.** Seja

$$f(x) = x(1 - x) = x - x.x,$$

e as extensões intervalares de  $f$

$$F_1(X) = X(1 - X)$$

e

$$F_2(X) = X - X.X.$$

Se  $X = [0, 1]$ , então

$$F_1(X) = [0, 1]$$

e

$$F_2(X) = [-1, 1].$$

Portanto,

$$F_1(X) \neq F_2(X).$$

## 3. O Método de Simpson Intervalar para Integração Numérica

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

$$S = \int_a^b f(x) dx.$$

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

$S = G(b) - G(a)$ .  $G(a)$  ou  $G(b)$  podem não ser representados no computador e seu valor será aproximado. Se não é possível encontrar uma expressão analítica para uma primitiva de  $f$ , métodos numéricos são essenciais para encontrar uma solução numérica do problema.

O método de Simpson Intervalar Caprani [3] é uma extensão intervalar do método de Simpson real e é fundamentado na propriedade aditiva da integral definida e no teorema do valor médio para integrais [1]. Supondo que  $f$  é quatro vezes continuamente derivável em  $A = [a, b]$ , e sendo conhecidas extensões intervalares para  $f$  e  $f^{(4)}$  derivada de  $f$  de quarta ordem, o método encontra um intervalo que encapsula a integral definida, isto é, um intervalo que contém o valor da integral definida, como visto a seguir.

### 3.1 O Método de Simpson nos Reais

Dada uma partição  $P = \{a_0, a_1, \dots, a_n\}$  de  $A = [a, b]$  tal que  $a_i - a_{i-1} = (b - a) / n$ . Então

$$S = \sum_{i=1}^n \int_{a_{i-1}}^{a_i} f(x) dx = \sum_{i=1}^n w(A_i) f(\varepsilon_i), \quad (1)$$

onde  $A_i = [a_{i-1}, a_i]$ ,  $\varepsilon_i \in A_i$  e  $w(A_i)$  é o diâmetro de  $A_i$ .

Como  $f$  é quatro vezes continuamente diferenciável tem-se

$$\begin{aligned} S_i &= \int_{a_{i-1}}^{a_i} f(x) dx \\ &= \frac{w(A_i)}{6} \left( f(a_{i-1}) + 4f(m(A_i)) + f(a_i) \right) - \\ &\quad \frac{w(A_i)^5}{2880} f^{(4)}(\varepsilon_i), \end{aligned} \quad (2)$$

onde  $\varepsilon_i \in A_i$  e  $m(A_i)$  é o ponto médio de  $A_i$ .  
Portanto,

$$S \cong \sum_{i=1}^n S_i. \quad (3)$$

### 3.2 O método de Simpson intervalar

O método de Simpson Intervalar em Caprani [3] admite conhecidas extensões intervalares  $F$  e  $G$  de  $f$  e  $f^{(4)}$ . Como  $f(x) \in F(X)$ , para todo  $x \in (X \in IR)$  então, pela Equação (1),

$$IS = \sum_{i=1}^n \int_{a_{i-1}}^{a_i} f(x) dx = \sum_{i=1}^n w(A_i) F(A_i), \quad (4)$$

e pela Equação (2),

$$\begin{aligned} IS_i &= \int_{a_{i-1}}^{a_i} f(x) dx \\ &= \frac{w(A_i)}{6} \left( F(a_{i-1}) + 4F(m(A_i)) + F(a_i) \right) - \\ &\quad \frac{w(A_i)^5}{2880} G(A_i), \end{aligned} \quad (5)$$

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

Portanto,

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

A precisão deste intervalo depende das extensões intervalares usadas e do valor de  $n$ .

## 4. Um intervalo encapsulador para a probabilidade de uma variável aleatória contínua

A distribuição de probabilidade de uma variável aleatória contínua [6] é caracterizada por sua função densidade de probabilidade a qual satisfaz às propriedades:

- a)  $f(x) \geq 0$ ;
- b)  $\int_a^b f(x) dx = P(a < X \leq b)$ ,  $a < b$ ;
- c)  $\int_{-\infty}^{\infty} f(x) dx = 1$ .

O item b) indica que a probabilidade da variável aleatória assumir valor em um intervalo é dada pela integral da função nesse intervalo. As funções densidade de probabilidade mais frequentes nas aplicações práticas, como por exemplo a distribuição normal, 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 no computador. A seguir será apresentado um intervalo encapsulador, usando o método de Simpson Intervalar, para a probabilidade de variáveis aleatórias com distribuição Normal Reduzida.

### 4.1 Distribuição Normal Reduzida

**Definição 3** Uma variável aleatória  $X$  tem uma **distribuição normal** se sua função densidade de probabilidade (fdp) for da forma

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)}, \quad -\infty < x < \infty \quad (7)$$

onde os parâmetros  $\mu$  e  $\sigma$  devem satisfazer às condições  $-\infty < \mu < \infty$ ,  $\sigma > 0$ . Prova-se que  $\mu$  e  $\sigma$  são respectivamente a média e o desvio-padrão de  $X$ .

Notação:  $X$  terá distribuição  $N(\mu, \sigma^2)$  se, e somente se sua distribuição de probabilidade for dada pela Equação (7).

**Teorema 1** Se  $X$  tiver distribuição  $N(\mu, \sigma^2)$ , e se  $Y = cX + d$ , então  $Y$  terá distribuição  $N(c\mu + d, c^2\sigma^2)$ .

Diz-se que  $X$  tem uma distribuição normal reduzida quando  $\mu = 0$  e  $\sigma = 1$ .

**Corolário 1** Se  $X$  tiver distribuição  $N(\mu, \sigma^2)$ , e se  $Y = \frac{X - \mu}{\sigma}$ , então  $Y$  terá distribuição  $N(0, 1)$ .

Suponha que  $X$  tenha distribuição  $N(0, 1)$ . Nesse caso, se  $A = [a, b] \subset R$ ,

$$P(a \leq x \leq b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}x^2} dx.$$

A função  $f$  não tem uma primitiva na forma analítica, e não é possível aplicar o Teorema Fundamental do Cálculo Integral. Contudo, métodos de integração numérica podem ser empregados para calcular integrais da forma acima, e no caso da  $N(\mu, \sigma^2)$ , gerar tabelas de probabilidades.

O método de Simpson intervalar é uma alternativa para encontrar um intervalo que melhor aproxime o valor desta integral, no sentido de se obter um intervalo com amplitude tão pequena quanto possível que a contenha. Para aplicá-lo é necessário determinar extensões intervalares para  $f$  e  $f^{(4)}$ . O método proposto neste trabalho consiste em utilizar na Equação (5) extensões intervalares  $F$  e  $G$  para  $f$  e  $f^{(4)}$  que, quando computados, produzem intervalos encapsuladores para a probabilidade procurada.

A derivada de  $f$  de quarta ordem é dada por

$$f^{(4)}(x) = \frac{-x^2}{\sqrt{2\pi}} (x^4 - 6x^2 + 3).$$

Observa-se que as funções  $f$  e  $f^{(4)}$  não são monótonas.

Sendo  $X = [x_1, x_2]$ , as funções intervalares

$$F(X) = \frac{1}{\sqrt{2\pi}} \begin{cases} \left[ \frac{-x_2^2}{e^{\frac{x_2^2}{2}}}, \frac{-x_1^2}{e^{\frac{x_1^2}{2}}} \right], & \text{se } x_1 > 0, \\ \left[ \frac{-x_1^2}{e^{\frac{x_1^2}{2}}}, \frac{-x_2^2}{e^{\frac{x_2^2}{2}}} \right], & \text{se } x_2 < 0, \\ \left[ \frac{-\max\left\{\frac{x_1^2}{2}, \frac{x_2^2}{2}\right\}}{e^{\frac{\max\{x_1^2, x_2^2\}}{2}}}, 1 \right], & \text{se } 0 \in X \end{cases}$$

e

$$G(X) = \frac{-X^2}{\sqrt{2\pi}} (X^4 - 6X^2 + 3),$$

são extensões intervalares para  $f$  e  $f^{(4)}$ , respectivamente.

O método foi implementado no Matlab [5], como arquivo M de funções, usando a biblioteca Intlab[4]. Sua chamada é dada por

**>> normal1(a, b)**

onde:

**a:** limite inferior do intervalo  $A$ .

**b:** limite superior do intervalo  $A$ .

Assume-se que  $F(a_{i-1})$ ,  $F(m(A_i))$  e  $F(a_i)$

são intervalos degenerados.

Os exemplos abaixo ilustram aplicações do método.

**Exemplo 2.** A duração de um certo tipo de pneu, em quilômetros rodados, é uma variável normal com duração média 60000km e desvio-padrão de 10 000km.

- Qual a probabilidade de um pneu durar entre 63 500km e 70 000km?
- Qual a probabilidade de um pneu durar entre 50 000km e 70 000km?
- Qual a probabilidade de um pneu durar exatamente 65 555km?
- Qual a probabilidade de um pneu escolhido ao acaso durar mais de 75 000km?

As soluções a seguir foram calculadas no Matlab [5], usando a biblioteca IntLab [4] e a função normal1, nos formatos short (precisão simples) e long (precisão dupla).

Observe que a variável aleatória não tem distribuição normal reduzida, sendo necessário o uso do Corolário 1. Denotando-se por  $X$  a variável do problema então,

$$Y = \frac{X - 60\,000}{10\,000}$$

tem distribuição  $N(0, 1)$ .

a) Para a probabilidade de um pneu durar entre 63 500km e 70 000km, usando a transformação  $Y$ , tem-se,

$$y(63\,500) = \frac{63\,500 - 60\,000}{10\,000} = 0.35;$$

$$y(70\,000) = \frac{70\,000 - 60\,000}{10\,000} = 1.00.$$

#### Probabilidade real:

Precisão simples:

0.2045,

Precisão dupla:

0.20451409489292.

#### Probabilidade encapsulada:

>> normal1(0.35, 1.00)

Precisão simples:

[0.2045, 0.2046],

Precisão dupla:

[0.20451409489283, 0.20451409489302].

b) Para a probabilidade de um pneu durar entre 50 000km e 70 000km, usando a transformação  $Y$ , tem-se,

$$y(50\,000) = \frac{50\,000 - 60\,000}{10\,000} = -1.00;$$

$$y(70\,000) = \frac{70\,000 - 60\,000}{10\,000} = 1.00.$$

#### Probabilidade real:

Precisão simples:

0.6827,

Precisão dupla:

0.68268949213709.

#### Probabilidade encapsulada:

>> normal1(-1.00, 1.00)

Precisão simples:

[0.6826, 0.6827],

Precisão dupla:

[0.68268949207779, 0.68268949219732].

c) Como em qualquer tipo de variável aleatória contínua, a probabilidade de a variável tomar exatamente um determinado valor é zero.

$$y(65\,555) = \frac{65\,555 - 60\,000}{10\,000} = 0.5555;$$

#### Probabilidade real:

Precisão simples:

0.0000

Precisão dupla:

0

#### Probabilidade encapsulada:

>>normal1(0.5555, 0.5555)

Precisão simples:

[0.0000, 0.0000],

Precisão dupla:

[0.00000000000000, 0.00000000000000].

d) Para a probabilidade de um pneu durar mais que 75 000km, usando a transformação  $Y$ , tem-se,

$$y(75\,000) = \frac{75\,000 - 60\,000}{10\,000} = 1.5.$$

E, portanto,

$$P(X > 75000) = P(Y > 1.5) = 0.5 - P(0 \leq Y \leq 1.5)$$

#### Probabilidade real:

Precisão simples:

0.0668,

Precisão dupla:

0.06680720126886.

#### Probabilidade encapsulada:

>> 0.5 - normal1(0, 1.5)

Precisão simples:

[0.0668, 0.0669],

Precisão dupla:

[0.06680720126054, 0.06680720127720].

**Exemplo 3** Suponha que  $X$  tenha a distribuição  $N(2, 0.16)$ . Calcule as seguintes probabilidades:

a)  $P(X \geq 2.3)$ ;

b)  $P(1.8 \leq X \leq 2.1)$ .

As soluções a seguir foram calculadas no Matlab [5], usando a biblioteca IntLab [4] e a função normal1, nos formatos short (precisão simples) e long (precisão dupla).

Pelo Corolário 1 tem-se,

$$Y = \frac{X - 2}{0.4}$$

com distribuição  $N(0, 1)$ .

a) Usando a transformação  $Y$  acima tem-se,

$$y(2.3) = \frac{2.3 - 2}{0.4} = 0.75.$$

E, portanto,

$$P(X > 2.3) = P(Y > 0.75) = 0.5 - P(0 \leq Y \leq 0.75).$$

**Probabilidade real:**

Precisão simples:

0.2266,

Precisão dupla:

0.22662735237687.

**Probabilidade encapsulada:**

`>> 0.5 - normal1(0, 0.75)`

Precisão simples:

[0.2266, 0.2267],

Precisão dupla:

[0.22662735237686, 0.22662735237688].

b) Usando a transformação  $Y$  tem-se,

$$y(1.8) = \frac{1.8 - 2.0}{0.4} = -0.5;$$

$$y(2.1) = \frac{2.1 - 2.0}{0.4} = 0.25.$$

**Probabilidade real:**

Precisão simples:

0.2902,

Precisão dupla:

0.29016878695694.

**Probabilidade encapsulada:**

`>> 0.5 - normal1(-0.5, 0.25)`

Precisão simples:

[0.2901, 0.2902],

Precisão dupla:

[0.29016878695693, 0.29016878695695].

Assim, observa-se que as probabilidades reais pertencem às probabilidades intervalares ou, as probabilidades intervalares Campos [2] encapsulam as probabilidades reais.

## 5. Conclusão

Neste trabalho foi desenvolvida uma versão do método de Simpson Intervalar para variáveis aleatórias contínuas com função densidade de probabilidade Normal Reduzida, denominado normal1, no Matlab [5]. Como visto as probabilidades reais foram encapsuladas pelas probabilidades intervalares [2] e, portanto probabilidades intervalares constituem uma alternativa numérica para validar os erros no cálculo de probabilidades para variáveis aleatórias Normais Reduzidas.

## Referências

- [1] Burden, R. L., Faires, J. D., Análise Numérica. São Paulo, Pioneira Thompson Learning, 2003.
- [2] Campos, M. A., Uma Extensão Intervalar para a Probabilidade Real. Ph. D. Thesis, Departamento de Informática – UFPE, 1997.
- [3] Caprani O., Madsen K., Nielsen H. B., Introduction to Interval Analysis, IMM, 2002
- [4] INTLAB – INTerval LABoratory. Disponível em [www.ti3.harburg.de/rump/intlab](http://www.ti3.harburg.de/rump/intlab), acessada em julho/2007.
- [5] MATLAB 7.0. User's Guide. MathWorks Inc.
- [6] Meyer, P. L., Probabilidade Aplicações à Estatística. Rio de Janeiro, LTC, 1983.
- [7] Monagan, M. B., Geddes, K. O. Maple 7 Programming Guide. Waterloo Maple Inc. – Canadá.
- [8] Moore, R. E. Interval Analysis. N. J. Prentice Hall Inc., Englewood Cliffs, 1966.
- [9] Moore, R. E., Methods and Applications for Interval Analysis. Philadelphia: SIAM, 1979.