Intervalo Encapsulador para Probabilidades Reais de Variáveis Aleatórias Contínuas Unidimensionais

Maria das Graças dos Santos
Doutorado em Matemática Computacional – UFPE,
Rua Profº Luiz Freire, s/n, Cidade Universitária,
50740-540, Recife – Pe,
E-mail: batgal60@yahoo.com.br,

Marcília Andrade Campos,
Centro de Informática UFPE,
Rua Profº Luiz Freire, s/n, Cidade Universitária,
50740-540, Recife – Pe,
E-mail: 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. Embora funções densidade de probabilidade como a Exponencial e a Uniforme sejam resolvidas analiticamente seu valor numérico no computador é dado por aproximação, e, portanto afetado por erros de arredondamento ou truncamento.*

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 Exponencial. 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

Um dos problemas no estudo das variáveis aleatórias contínuas unidimensionais sobre o conjunto dos números reais, *R,* 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. Embora integrais de funções densidade de probabilidade como a Exponencial e a Uniforme sejam resolvidas analiticamente, computacionalmente seu valor numérico é aproximado (representação dos reais por meio dos números de ponto flutuante), e, portanto afetado por erros de arredondamento ou truncamento. Outras funções densidade como a Normal ou a 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. Existe na literatura [1][10] 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 o que também é um valor aproximado.

A Matemática Intervalar [8][9] fornece uma alternativa para resolver este problema, ou seja, o controle automático do erro numérico, pois 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 probabilidades de uma variável aleatória, com distribuição Exponencial. Os algoritmos para tal foram implementados no Matlab [5], usando o Intlab [4].

Este artigo está organizado da seguinte forma: a Seção 2 apresenta conceitos básicos da matemática intervalar, incluindo as operações aritméticas e topologia; aproveitando a relação entre o espaço dos intervalos, IR, e o plano R^2, os conceitos apresentados são ilustrados geometricamente através do Maple [7]. 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 densidade exponencial e resultados numéricos decorrentes da aplicação do método. As conclusões estão na Seção 5 e por último as referências bibliográficas.

2. Conceitos Básicos na Matemática Intervalar

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

$$\mathbf{IR} = \{[x_1, x_2] / x_1, x_2 \in \mathbf{R},\ x_1 \le x_2\}.$$

Associando-se a cada intervalo [x, y] $\in$ IR um ponto (x, y) $\in$ R^2, obtém-se uma representação geométrica Moore[9] para IR, como mostra a Figura 1.

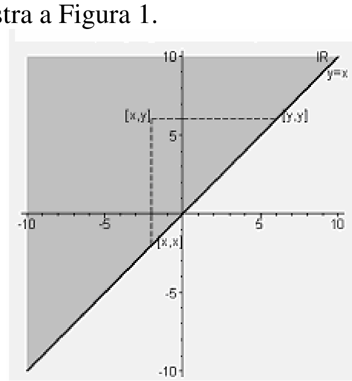

Figura 1: Interpretação geométrica do conjunto IR.

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.

2.1 A Aritmética Intervalar

A aritmética intervalar é a definida em Moore [8] [9].

Seja $* \in \{+, -, \cdot, /\}$ as operações binárias sobre o conjunto dos números reais R. Se A, B $\in$ IR, então

$$\mathbf{A * B = \{z = a * b \mid a \in A, b \in B\}},$$

define uma operação binária sobre IR. É assumido que $0 \notin$ B no caso da divisão.

As operações aritméticas sobre os intervalos A = [a, b] e B = [c, d] podem ser calculadas explicitamente por:

2.1.1 Adição

$$\mathbf{A + B = [a + c,\ b + d].}$$

Exemplo 1. Se A = [-3, 4.9] e B = [1.5, 3.7] então A + B = [-1.5, 8.6]. Vide Figura 2.

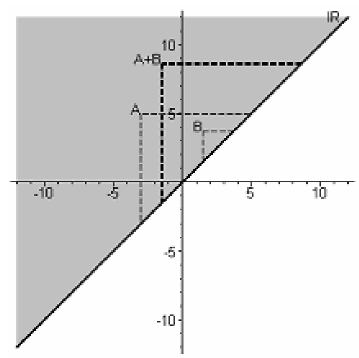

Figura 2: Interpretação geométrica da Adição.

2. 1.2 Subtração

$$\mathbf{A - B = [a - d,\ b - c].}$$

Exemplo 2. Se A = [-1.4, 2.3] e B = [-3, 6] então A - B = [-7.4, 5.3]. Vide Figura 3.

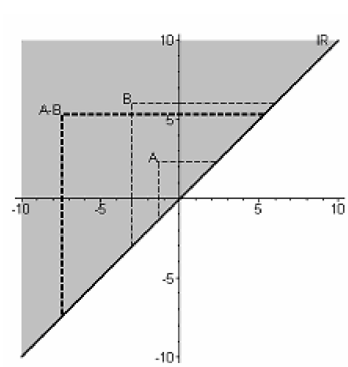

Figura 3: Interpretação geométrica da Subtração.

2.1.3 Multiplicação

$$\mathbf{A \cdot B = [\min(ac, ad, bc, bd), \max(ac, ad, bc, bd)].}$$

Exemplo 3. Se A = [-2, 3] e B = [-4.2, 4] então. $A \cdot B = [-12.6,\ 12]$. Vide Figura 4.

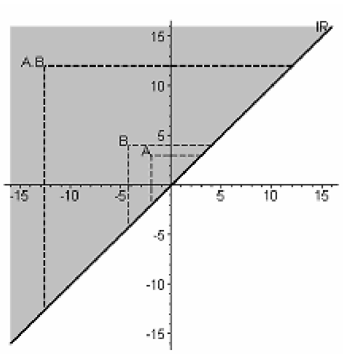

Figura 4: Interpretação geométrica da Multiplicação.

2.1.4 Divisão

$$\mathbf{A / B = [\min(\frac{a}{c}, \frac{a}{d}, \frac{b}{c}, \frac{b}{d}), \max(\frac{a}{c}, \frac{a}{d}, \frac{b}{c}, \frac{b}{d})]}$$

desde que $0 \notin B$.

Exemplos 4. Se A = [-4, 5] e B = [1.5, 2.6] então $C = A / B = [-2.667, 3.333]$. Vide Figura 5.

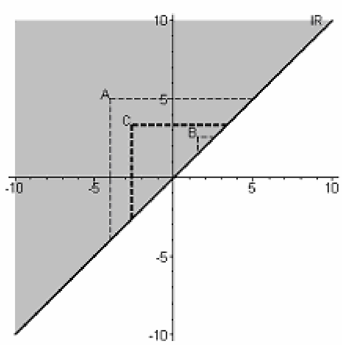

Figura 5: Interpretação geométrica da Divisão.

Além das operações binárias, têm-se as unárias:

2.1.5 Pseudo - Oposto Aditivo

O Pseudo-Oposto Aditivo do intervalo A, denotado por -A, é o intervalo definido por

$$\mathbf{-A = [-b, -a].}$$

Exemplo 5. Se A = [$-\sqrt{2}$,5] então –A = [-5, $\sqrt{2}$]. Vide Figura 6.

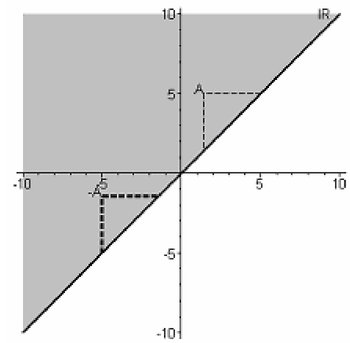

Figura 6: Interpretação geométrica do Pseudo-Oposto Aditivo.

2.1.6 Pseudo - Inverso Multiplicativo

Se $0 \notin$ A, o Pseudo - Inverso Multiplicativo de A, denotado por 1/A, é o intervalo definido por

$$\mathbf{1/A = [1/b, 1/a].}$$

Exemplo 6. Se A = [1.5, 4] então 1/A = [0.250, 0.667]. Vide Figura 7.

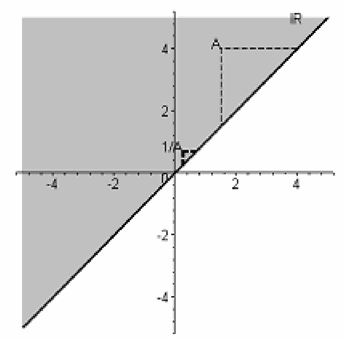

Figura 7: Interpretação geométrica do Pseudo-Inverso Multiplicativo.

2.2 Topologia

Sejam A = [a, b] e B=[c, d] em IR. Definimos a distância entre A e B como sendo o número real não negativo dado por

$$\mathbf{d(A, B) = \max\{|a - c|, |b - d|\}.}$$

Teorema 1. O conjunto IR dos intervalos de reais munidos da função distância d(A, B), é um espaço métrico completo.

Como a função distância é uma métrica sobre IR, podemos definir algumas noções geométricas básicas sobre o conjunto IR, que apresentaremos a seguir.

Definição 1. Define-se **Bola Aberta** de centro em P∈ IR e raio $r > 0$, como o conjunto dos pontos X ∈ IR cuja distância ao ponto P é menor do que r. Notação $B_r(P)$.

$$\mathbf{B_r(P) = \{X \in IR \mid d(X,P) < r\}.}$$

Exemplo 7. Sejam P = [-5, 6] e r = 1 então $B_r(P)$ é o conjunto de todos os X=[x, y] ∈ IR tais que

$$\mathbf{d\,(X,P) = max\{|x+5|,\ |y-6|\} < 1.}$$

Vide Figura 8.

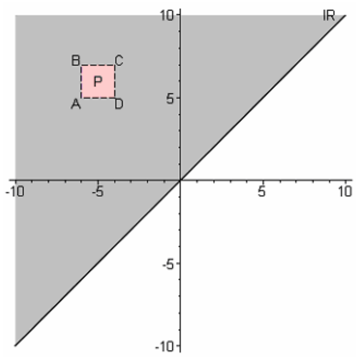

Figura 8: Bola Aberta de centro P e raio r.

Definição 2. Define-se **Bola Fechada** de centro em P∈ IR e raio $r > 0$, como o conjunto dos pontos X ∈ IR cuja distância ao ponto P é menor ou igual a r. Notação $BF_r(P)$.

$$\mathbf{BF_r(P) = \{X \in IR \mid d(X,P) \le r\}.}$$

Exemplo 8. Sejam P = [-5, 6] e r = 1 então $BF_r(P)$ é o conjunto de todos os X = [x, y] ∈ IR tais que

$$\mathbf{d\,(X,P) = max\{|x+5|,\ |y-6|\} \le 1.}$$

Vide Figura 9.

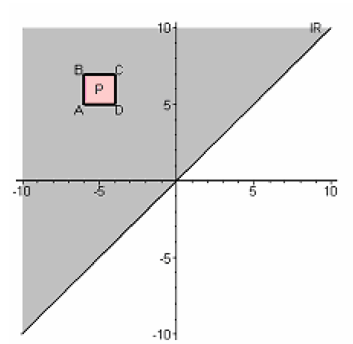

Figura 9: Bola Fechada de centro P e raio r.

Definição 3. Define-se **Esfera** de centro em P∈ IR e raio $r > 0$, como o conjunto dos pontos X ∈ IR cuja distância ao ponto P igual a r. Notação $E_r(P)$.

$$\mathbf{E_r(P) = \{X \in IR \mid d(X,P) = r\}.}$$

Exemplo 9. Sejam P = [-5, 6] e r =1 então $E_r(P)$ é o conjunto de todos os X = [x, y] ∈ IR tais que

$$\mathbf{d\,(X,P) = max\{|x+5|,\ |y-6|\} = 1.}$$

Vide Figura 10.

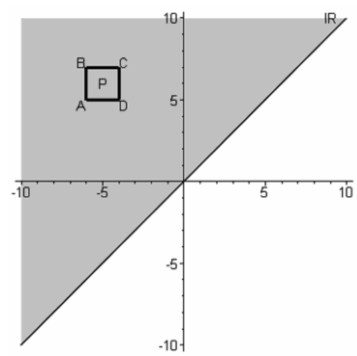

Figura 10: Esfera de centro P e raio r.

Definição 4. Uma função intervalar *F* de uma variável intervalar *X* é **Inclusão Monotônica** se

$$\mathbf{Y \subseteq X \Rightarrow F(Y) \subseteq F(X).}$$

Definição 5. 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

$$\mathbf{F(x) = f(x),\ \ x \in X.}$$

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 10. 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 um a função contínua $f : R \to R$ 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] tem-se, 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 se obter uma solução numérica do problema.

O método de Simpson intervalar Caprani [3] é uma extensão intervalar do método de Simpson real [1][10]. O método de Simpson intervalar 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 são conhecidas extensões intervalares para f e $f^{(4)}$ derivada de f de quarta ordem, o método encontra 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 $P = \{a_0, a_1, ..., a_n\}$ de A = [a, b] tal que $a_i - a_{i-1} = (b-a)/n$.

Então

$$S = \sum_{i=1}^{n} \int_{a_{i-1}}^{a_i} f(x)dx = \sum_{i=1}^{n} w(A_i) f(\varepsilon_i), \qquad (3.1)$$

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,

$$\begin{aligned} S_i &= \int_{a_{i-1}}^{a_i} f(x)dx \\ &= \frac{w(A_i)}{6}\left(f(a_{i-1}) + 4f(m(A_i)) + f(a_i)\right) \\ &\quad - \frac{w(A_i)^5}{2880} f^{(4)}(\varepsilon_i), \end{aligned} \qquad (3.2)$$

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

Portanto,

$$S \cong \sum_{i=1}^{n} S_i. \qquad (3.3)$$

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 $f^{(4)}$ respectivamente. Como $f(x) \in F(X)$, para todo $x \in (X \in IR)$ então, pela Equação (3.1),

$$IS = \sum_{i=1}^{n} \int_{a_{i-1}}^{a_i} f(x)dx = \sum_{i=1}^{n} w(A_i) F(A_i), \qquad (3.4)$$

e pela Equação (3.2)

$$\begin{aligned} IS_i &= \int_{a_{i-1}}^{a_i} f(x)dx \\ &= \frac{w(A_i)}{6}\left(F(a_{i-1}) + 4F(m(A_i)) + F(a_i)\right) \\ &\quad - \frac{w(A_i)^5}{2880} G(A_i), \end{aligned} \qquad (3.5)$$

onde $F(a_{i-1})$, $F(m(A_i))$ e $F(a_i)$ são intervalos degenerados de números reais, mas como o integrando está sujeito a erros, faz-se necessário trocar os intervalos degenerados por outros de amplitude tão pequena quanto possível que os contenham.

Portanto,

$$S \subseteq IS = \sum_{i=1}^{n} IS_i. \qquad (3.6)$$

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

4. Um intervalo Encapsulador para a Probabilidade Real 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:

$$\mathbf{a)}\ f(x) \geq 0;$$

$$\mathbf{b)} \int_a^b f(x)dx = P(a < X \leq b), \quad a < b;$$

$$\mathbf{c)} \int_{-\infty}^{\infty} f(x)dx = 1.$$

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 freqüentes nas aplicações práticas, como por exemplo a distribuição normal, não possuem primitivas na forma analítica. A distribuição Exponencial possui função densidade com primitiva na forma analítica, mas o valor da probabilidade, em geral, não é racional, portanto não é um número exatamente representável no computador. A seguir será apresentado um intervalo encapsulador, usando o método de Simpson Intervalar, para calcular probabilidade de variáveis aleatórias com distribuição Exponencial.

4.1 Distribuição Exponencial

Uma variável aleatória contínua com função densidade dada por

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

com $\alpha > 0$, é uma variável aleatória Exponencial.

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

a) Se $b \leq 0$, então

$$P(a < x \leq b) = 0;$$

b) Se $0 \in (a, b)$, então

$$P(a < x \leq b) = P(0 \leq x \leq b) = \int_0^b \alpha e^{-\alpha x}\, dx;$$

c) Se $a \geq 0$, então

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

A função f tem uma primitiva na forma analítica, e pelo Teorema Fundamental do Cálculo Integral pode-se calcular a integral definida diretamente desta primitiva. O número "e" e suas potências não exatas são irracionais, sendo seus valores aproximados, portanto sujeitos a erros de arredondamento.

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 (3.5) extensões intervalares F e G para f e $f^{(4)}$ que, quando computados, produzem intervalos de amplitude tão pequena quanto possível.

Como f e $f^{(4)}$ são decrescentes em $[0, \infty)$ as funções intervalares

$$F(X) = \alpha\left[e^{-\alpha \overline{x}},\ e^{-\alpha \underline{x}}\right]$$

e

$$G(X) = \alpha^5\left[e^{-\alpha \overline{x}},\ e^{-\alpha \underline{x}}\right]$$

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

$$\mathbf{dexp\ (a, b, c, p)} \qquad \mathbf{(4.1)}$$

onde:

a: limite inferior do intervalo A.
b: limite superior do intervalo A.
c: parâmetro da função densidade.
p: número de subintervalos determinado pela partição de A.

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

Exemplo 11. Se os defeitos de um tecido seguem uma lei de Poisson com média de um defeito a cada 500m, qual a probabilidade que o intervalo entre dois defeitos consecutivos sejam:

a) no mínimo de 1250m
b) entre 1000 e 1250m
c) menor do que 1000m.

As soluções a seguir foram calculadas no Matlab [5], usando a biblioteca Intlab [4] e a função (4.1), nos formatos short (precisão simples) e long (precisão dupla).

a) **Probabilidade real:**

Precisão simples:

$$0.0821, \tag{4.2}$$

Precisão dupla:

0.08208499862390.

Intervalo encapsulador para a probabilidade real:

```
>> dexp(1250, inf, 1/500, 100)
```

Precisão simples:

$$[0.0820, 0.0821], \tag{4.3}$$

Precisão dupla:

[0.08208499862389, 0.08208499862390].

b) **Probabilidade real:**

Precisão simples:

$$0.0532 \tag{4.4}$$

Precisão dupla:

0.05325028461271.

Intervalo encapsulador para a probabilidade real:

```
>> dexp(1000, 1250, 1/500, 800)
```

Precisão simples:

$$[0.0532, 0.0533], \tag{4.5}$$

Precisão dupla:

[0.05325028461271, 0.05325028461272].

c) **Probabilidade real:**

Precisão simples

$$0.8647 \tag{4.6}$$

Precisão dupla

0.86466471676339.

Intervalo encapsulador para a probabilidade real:

```
>> dexp(-inf, 1000, 1/500, 100)
```

Precisão simples:

$$[0.8646, 0.8647], \tag{4.7}$$

Precisão dupla:

[0.86466471676338, 0.86466471676339].

Assim de (4.2), (4.3), (4.4), (4.5), (4.6) e (4.7), 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 Exponencial, denominado dexp, 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 Exponenciais.

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, acessada em julho/2007.

[5] MATLAB 7.0. The MathWorks, Inc. 2004

[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.

[10] Santos, J. D., Silva. Z. C., Métodos Numéricos. Recife. Editora Universitária da UFPE, 2006.