



Variância Intervalar das funções densidade de probabilidade com distribuições Exponencial, Pareto e Uniforme

Interval variance of the probability density functions with exponential distributions, Pareto and Uniform

L. M. Tortelli^{1*}; M. D. C. Balboni¹; A. F. Finger²; A. B. Loreto¹

¹Centro de Desenvolvimento Tecnológico/ Universidade Federal de Pelotas, 96010-610, Pelotas-RS, Brasil

² Universidade Federal do Pampa/UNIPAMPA - Campus Alegrete, Alegrete-RS, Brasil

*lmtortelli@inf.ufpel.edu.br

(Recebido em 22 de outubro de 2014; aceito em 07 de março de 2015)

A computação científica demanda de grande exatidão e ao se utilizar o sistema de ponto flutuante os resultados podem ser afetados por erros. Estes inviabilizam a confiabilidade nos resultados alcançados, com isto surge a necessidade da utilização da matemática intervalar. Esta cria limites que são os extremos do intervalo onde o valor real x está contido, controlando a propagação desses erros uma vez que resultados intervalares carregam consigo a segurança de sua qualidade. O cálculo da variância necessita de exatidão devido ao grande processamento de valores reais. O presente trabalho tem como objetivo definir a variância intervalar das funções de densidade de probabilidade com distribuições Exponencial, Pareto e Uniforme a fim de obter um controle automático de erros com limites confiáveis.

Palavras-chave: Variância Intervalar, Aritmética Intervalar, Computação Científica

The scientific computing demand for high accuracy when using the system of floating point results can be affected by errors. These make it impossible to rely on results achieved with this comes the need for the use of interval mathematics. This creates boundaries that are the extremes of the range where the real value x is contained by controlling the spread of these errors since interval results carry with the safety of their quality. The calculation of the variance requires great accuracy due to the processing of actual values. This study aims to define the interval variance of the probability density functions with Exponential, Pareto and Uniform distributions order to get an automatic error control with trust boundaries.

Keywords: Interval Variance, Interval Arithmetic, Scientific Computing.

1. INTRODUÇÃO

A necessidade de se obter precisão e exatidão em cálculos numéricos é uma característica importante nas mais diversas áreas científicas. Computar probabilidades em situações práticas envolve números e, consequentemente, problemas numéricos. Os problemas numéricos na computação científica originam-se primordialmente da impossibilidade de se operar com números reais diretamente, pois é necessário representar uma grandeza contínua (a reta real) de forma discreta. A representação de valores reais no sistema de ponto flutuante é realizada pela aproximação de um subconjunto finito dos reais [4], ocorrendo assim erros numéricos difíceis de serem tratados ou reparados. Estes são originados através da aproximação (arredondamento e truncamento) feita pela máquina ao tentar representar valores reais com toda sua extensão e exatidão. Há também a possibilidade dos erros serem originados nos processos precedentes a computação, sendo eles a falha humana na obtenção dos dados e também falha nos instrumentos de medida utilizados. O erro total obtido será o acúmulo dos erros nos resultados intermediários [5].

A matemática intervalar proposta por Moore [3] surge para automatizar o cálculo do erro computacional científico com limites confiáveis. Nesta forma de representação numérica os valores pontuais não são aproximados, o que ocorre é a representação do valor real x em um intervalo X que o contenha. Os erros originados para tratamento dos valores como o arredondamento e truncamento, para adequar o valor pontual a ser representado pela máquina,

são tratados pela aritmética intervalar e pelos arredondamentos direcionados. O tamanho do intervalo é utilizado como medida de qualidade [5].

Dentro do ramo estatístico, comumente são utilizadas funções de densidade de probabilidade que descrevem a forma de organização de algumas populações, como distribuição Exponencial, Pareto, Uniforme, entre outras. O valor esperado serve como uma estimativa para estas distribuições [6]. Entretanto, apenas o valor esperado não fornece dados suficientes para descrever a população sendo necessário o cálculo da variância que é uma forma de medir a dispersão dos valores da distribuição em relação a sua média.

O presente trabalho tem como objetivo definir a variância intervalar das funções de densidade de probabilidade com distribuições Exponencial, Pareto e Uniforme a fim de obter um controle automático de erros com limites confiáveis. Adicionalmente, implementam-se as expressões intervalares de cada distribuição no ambiente IntPy[10], que consiste em um ambiente de programação intervalar com bons resultados de exatidão, cálculo de erros absoluto e relativo, fornecendo um intervalos solução com qualidade[7].

2. METODOLOGIA

Durante o processo de resolução de problemas podem ocorrer 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 certo número de processos. Deste modo os resultados atingidos demonstram pouca confiabilidade uma vez que sua exatidão foi perdida durante o ajuste do valor real fornecido, para o representado em máquina. Com o objetivo de diminuir essa propagação do erro e fornecer também uma garantia de sua certeza aplicam-se técnicas intervalares, as quais possuem medidas de erros que verificam a qualidade do intervalo.

Segundo FERSON et. al.[2], historicamente o primeiro método para computar o intervalo solução é o método chamado de extensão intervalar. Nele repete-se a computação formando o programa passo-a-passo, substituindo cada operação elementar de números reais pela correspondente operação da aritmética intervalar.

Funções densidade de probabilidade necessitam de grande precisão em seus cálculos, possuem como objetivo descrever a forma de organização dos indivíduos de uma população. Cada função de densidade de probabilidade com distribuição possui sua função analítica que representa o conjunto de dados de acordo com as características de uma população. A variância que é caracterizada como uma medida de dispersão denotada por $V(X)$ ou s^2 , visa calcular a variação de cada elemento da população em relação a média da distribuição em questão. Considerando que este cálculo fornece uma variação média de todos os dados a exatidão numérica e a confiabilidade nos resultados é de suma importância. Estas exigências podem ser satisfeitas através da utilização da aritmética intervalar, representando a variância através de intervalos confiáveis e com qualidade que encapsulem o valor real.

Cálculos numéricos em computadores devem ser realizados por meio das linguagens ou bibliotecas que tenham definidos o tipo intervalo e as operações sobre o tipo, usualmente denominadas de linguagens XSC (eXtended for Scientific Computation)[1], além de serem linguagens de grande precisão numérica voltados a utilização de cálculos científicos.

Na matemática intervalar, a qualidade dos intervalos solução pode ser verificada através das seguintes medidas de erro [4]:

- **Erro Absoluto** (E_A): conforme a Equação (1),

$$E_A = |x - m(X)| < \frac{w(X)}{2} \quad (1)$$

onde x é o valor real, $m(X)$ é o ponto médio do intervalo X e $w(X)$ é o diâmetro do intervalo X .

- **Erro Relativo** (E_R) identificado na Equação (2),

$$E_R = \left| \frac{x - m(X)}{x} \right| < \frac{w(X)}{2 \min|X|} \quad (2)$$

onde os parâmetros são exatamente os mesmos do Erro absoluto, com exceção da expressão $\min|X|$ que consiste no módulo limite inferior do intervalo.

No presente trabalho aplicam-se estas medidas de erros para verificar a qualidade dos intervalos solução das funções de densidade de probabilidade com distribuições Exponencial, Pareto e Uniforme.

A seguir descrevem-se as funções de densidade de probabilidade e os valor esperado das distribuições Exponencial, Pareto e Uniforme, as quais são fundamentais para realizar a definição intervalar das respectivas variâncias. Ressaltando que utilizando-se do mesmo princípio abordado neste trabalho para representar as funções de variância de cada distribuição, o mesmo foi utilizado para representar os valores esperados, como pode ser visto em Tortelli[11].

2.1. DISTRIBUIÇÃO EXPONENCIAL

Visa descrever a forma em que uma população se organiza quando há uma taxa de falhas constante ou também o tempo despendido para realizar uma determinada tarefa [8]. A distribuição Exponencial é de suma importância para áreas como microbiologia, com o estudo do tempo de vida de colônias de microrganismos, tempo de falhas de produtos e entre outras.

A partir da Figura 1 apresenta o gráfico de uma Distribuição Exponencial, observando atentamente a forma que sua curva varia no decorrer da reta real.

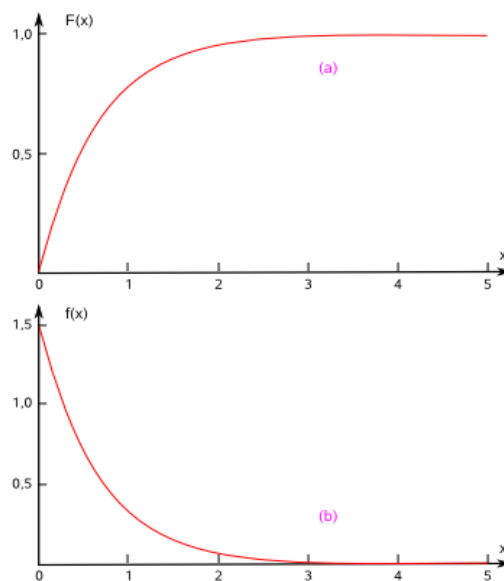


Figura 1: Gráfico das duas formas que uma distribuição Exponencial pode assumir

A distribuição Exponencial possui como principal parâmetro o λ , que consiste no tempo médio de vida da população estudada. Assim, através da Equação (3) pode-se verificar a função densidade de probabilidade da distribuição Exponencial

$$f(x) = \begin{cases} \frac{\lambda e^{-\lambda x}}{x^{\alpha+1}}, & \text{se } x \geq 0 \\ 0, & \text{se } x < 0 \end{cases} \quad (3)$$

Através da Equação(4) apresenta-se o valor esperado da distribuição exponencial, que contém como único parâmetro o λ .

$$E(x) = \frac{1}{\lambda} \quad (4)$$

A Equação (5) apresenta a função de variância real da distribuição Exponencial:

$$Var(x) = \frac{1}{\lambda^2} \quad (5)$$

Representando o parâmetro real λ pelo intervalo $A=[\underline{a}, \bar{a}]$, tem-se a Equação (6), que consiste na variância intervalar da distribuição Exponencial:

$$Var(X) = \left[\frac{1}{\bar{a}^2}, \frac{1}{\underline{a}^2} \right] \quad (6)$$

onde \underline{a} e \bar{a} são os limites inferior e superior do intervalo A.

2.2. DISTRIBUIÇÃO DE PARETO

Criada por Vilfredo Pareto em 1897 através do estudo de distribuição de renda da população, verificou que 80% das riquezas da cidade eram mantidas por 20% da população.[9] Esta suposição prega que a maioria dos resultados em qualquer situação é determinada por um pequeno número de causas e este princípio é aplicado em estudos relativos à Economia, Produtividade, Política, Desenvolvimento e entre outras áreas que apresentam o mesmo padrão.

Através da Figura 2 é possível observar o comportamento da distribuição de Pareto.

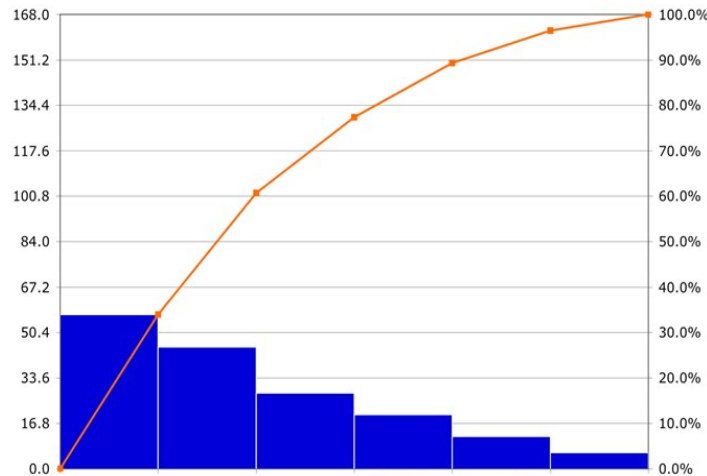


Figura 2: Forma de organização de uma população na distribuição de Pareto

A distribuição de Pareto contém a função de densidade de probabilidade, conforme apresentado na Equação (7):

$$f(x) = \begin{cases} \frac{\alpha \beta^\alpha}{x^{\alpha+1}}, & \text{se } x \geq \beta \\ 0, & \text{se } x < \beta \end{cases} \quad (7)$$

Sendo $x \in \mathbb{R}$, α o parâmetro de forma e β o de escala.

A partir desta fórmula, pode-se inferir como estimativa de concentração de dados da distribuição o seu valor esperado real, apresentado pela Equação (8):

$$E(x) = \frac{\alpha\beta}{\alpha - 1} \quad (8)$$

Utilizando exatamente os mesmos parâmetros de sua função densidade.

Para uma maior descrição do conjunto de dados analisados pela distribuição de Pareto, pode-se utilizar do cálculo da variância para determinar a variabilidade total dos elementos em relação ao seu valor esperado Equação (8). A variância real já é uma expressão conhecida, demonstrada através da Equação (9):

$$Var(x) = \frac{\alpha\beta^2}{(\alpha - 1)^2(\alpha - 2)} \quad (9)$$

Mantendo os mesmos parâmetros de sua função densidade de probabilidade.

Define-se a variância intervalar representando o parâmetro α por um intervalo $A=[\underline{a}, \bar{a}]$ e o parâmetro β pelo intervalo $B=[\underline{b}, \bar{b}]$, e cada operação aritmética real é substituída pela respectiva operação intervalar, conforme a matemática intervalar definida por Moore[3]. Sendo assim obtém-se a Equação (10) que consiste na mesma Equação (9), porém com todos os parâmetros intervalares:

$$Var(X) = \left[\frac{\underline{a}\underline{b}^2}{(\underline{a}-1)^2(\underline{a}-2)}; \frac{\bar{a}\bar{b}^2}{(\bar{a}-1)^2(\bar{a}-2)} \right] \quad (10)$$

onde \underline{a} e \bar{a} são os extremos inferior e superior do intervalo A, e \underline{b} e \bar{b} são os extremos inferior e superior do intervalo B.

2.3. DISTRIBUIÇÃO UNIFORME

Esta distribuição contém uma importante característica, em que um número finito de resultados tem chances iguais de acontecer.[8] O exemplo clássico adotado para relacionar-se com a distribuição Uniforme é o lançamento de um dado não viciado, a probabilidade de ocorrer qualquer um dos seis valores disponíveis no dado é sempre a mesma. Isto pode ser verificado através da Figura 3.



Figura 3: Gráfico da distribuição Uniforme

A função densidade de probabilidade com distribuição Uniforme é a que segue:

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{se } a \leq x \leq b \\ 0, & \text{caso contrario} \end{cases} \quad (11)$$

Onde a é o parâmetro de deslocamento e b corresponde ao parâmetro de dispersão. A partir da sua função de densidade apresentada pela Equação (11), pode-se compreender o valor esperado como apresentado pela Equação (12):

$$E(x) = \frac{a+b}{2} \quad (12)$$

Descreve-se a variância real da distribuição Uniforme conforme a equação (13):

$$Var(x) = \frac{(b-a)^2}{12} \quad (13)$$

Utilizando-se da extensão intervalar tem-se uma nova expressão para a variância da distribuição Uniforme expressa na Equação (14):

$$Var(X) = \left[\frac{(\underline{b} - \underline{a})^2}{12}; \frac{(\bar{b} - \bar{a})^2}{12} \right] \quad (14)$$

onde \underline{a} e \bar{a} são os extremos inferior e superior do intervalo A que encapsula o parâmetro a e \underline{b} e \bar{b} são os extremos inferior e superior do intervalo B que encapsula o parâmetro b .

3. RESULTADOS E DISCUSSÕES

As expressões definidas para a variância intervalar das distribuições Exponencial, Pareto e Uniforme, conforme as equações (6), (10) e (14) foram implementadas no ambiente de programação intervalar IntPy. Trata-se de uma biblioteca de programação intervalar da linguagem Python e em relação aos seus semelhantes obteve bons resultados de exatidão e qualidade do intervalo fornecido, como pode ser verificado em Balboni[4]. Esta seção está subdividida em exemplos comentados de cada distribuição, como também uma discussão sobre cada resultado obtido. Na sequência realiza-se uma análise das medidas de erros com o objetivo de verificar a qualidade do intervalo solução e, por consequência, garantir a máxima exatidão nos resultados alcançados.

3.1 EXEMPLO 1: DISTRIBUIÇÃO EXPONENCIAL

O tempo de vida (em horas) de um transistor é uma variável aleatória T com distribuição exponencial. O tempo médio de vida do transistor é de 500 horas. A partir destes dados encontre a variação deste tempo médio em relação a média da distribuição.

Solução: Pelas Equações (5) e (6) calcula-se a variância real e intervalar para a distribuição de Pareto, obtendo:

- I. Variância real = 4×10^{-6}
- II. Variância intervalar = $[3.999959840001604 \times 10^{-6}, 4.000040160001606 \times 10^{-6}]$

Nos resultados acima observa-se que o valor real está contido na solução intervalar. A seguir verifica-se a qualidade do intervalo solução através das medidas do Diâmetro e dos erros Absoluto e Relativo (Equações (1) e (2)):

- Diâmetro do intervalo = $8.03200000017 \times 10^{-11}$
- Erro Absoluto = $1.60512743505 \times 10^{-18} < 4.01600000008 \times 10^{-11}$
- Erro Relativo = $1.60512743505 \times 10^{-18} < 4.01600000008 \times 10^{-11}$

Considerando os resultados das medidas de qualidade, pode-se afirmar que a extensão intervalar definida para a variância da distribuição Exponencial possui intervalo solução com qualidade, o que garante uma boa representação e exatidão em relação ao valor real.

3.2 EXEMPLO 2: DISTRIBUIÇÃO DE PARETO

Numa população os ingressos são distribuídos segundo uma distribuição de Pareto, com $\delta = 9.42$ e $\beta = 1524$. Calcule a sua variância.

Solução: Pelas equações (9) e (10) calcula-se a variância real e intervalar para a distribuição Exponencial, obtendo:

- I. Variância real= 41590.3666032
- II. Variância intervalar = [41590.16706562708, 41590.56614185515]

A seguir pode verifica-se a qualidade do intervalo obtido acima, através do Diâmetro, Erro Absoluto (Equação(1)) e Relativo(Equação(2)):

- Diâmetro do intervalo = 0.399076228074
- Erro Absoluto= $5.51663106307 \times 10^{-7} < 0.199538114037$
- Erro Relativo = $1.32642039819 \times 10^{-11} \leq 4.79772331094 \times 10^{-6}$

Pelos resultados observa-se que o valor real obtido está contido no intervalo solução. Sendo este um resultado com qualidade, em que pode ser comprovado através das medidas de qualidades acima.

3.3 EXEMPLO 3: DISTRIBUIÇÃO UNIFORME

Seja X com uma distribuição contínua uniforme no intervalo [1,5; 5,5]. Determine sua variância.

Solução: A distribuição Uniforme utiliza como parâmetros de cálculo os extremos de seu intervalo como pode ser verificado nas Equações (14) e (15). Toma-se a = 1.5 e b = 5.5, obtendo assim os seguintes resultados.

- I. Variância real= 833.333333333
- II. Variância intervalar = [833.3323055564451, 833.3343611120011]

Observando o intervalo solução obtido, verifica-se que o valor real calculado está contido no intervalo. A fim de garantir que o intervalo encontrado tem boa qualidade, utiliza-se as medidas de qualidade a seguir:

- Diâmetro do intervalo = 0.00205555555601
- Erro Absoluto= $8.89826878847 \times 10^{-10} < 0.00102777777801$
- Erro Relativo = $1.06779225462 \times 10^{-12} \leq 1.23333485472 \times 10^{-6}$

De fato o intervalo, constata-se que o intervalo solução possui boa qualidade e também encapsula o valor real. Com isto verifica-se que a expressão intervalar da Variância da Distribuição Uniforme obtém resultados com exatidão máxima.

Todos os cálculos acima foram realizados em um computador com as seguintes configurações, processador Intel® Intel® Core™2 Quad CPU Q8300 @ 2.50GHz × 4, L1 Cache 4 x 32Kb, Cache L2 4Mb, Memória RAM de 4GB DDR3 1333MHz, armazenamento HD Sata 160 GB GB modelo ATA Samsung HD502HJ, placa gráfica onboard Intel® G33 x86/MMX/SSE2, sistema operacional openSUSE 13.1.

Observando o cálculo das medidas de qualidade dos exemplos de cada distribuição verifica-se que as mesmas apresentaram-se satisfatórias, uma vez que mantiveram o intervalo solução confiável. Por fim o controle e automatização de erros fornecidos pelo uso de intervalos oferecem garantias de exatidão a cada resultado alcançado.

4 CONCLUSÃO

Tendo como objetivo a maior exatidão em cálculos computacionais e também à necessidade de controlar a propagação dos erros durante todos os processos que envolvem o cálculo, o uso da aritmética intervalar torna-se importante na qualidade dos resultados atingidos, uma vez que os intervalos garantem a sua incerteza.

Dentro do estudo de funções de densidade de Probabilidade a variância, conjuntamente com outros indicadores estatísticos, serve para descrever com maiores detalhes a população estudada.

Sabendo que os cálculos sobre probabilidade normalmente não utilizam números inteiros, os erros gerados no decorrer do processo tornam-se preocupantes, uma vez que não há garantidas que o resultado alcançado é exato.

O objetivo deste trabalho foi mostrar a importância de trabalhar com intervalos na computação da variância das funções de densidade de probabilidade com distribuições Exponencial, Pareto e Uniforme. A partir dos métodos de extensão intervalar e das funções reais da variância de cada distribuição, foi viabilizado uma expressão que encapsule o valor real e que seja operada pelas propriedades da matemática intervalar. Através de todos os exemplos de cada distribuição de probabilidade constatou-se a qualidade dos intervalos obtidos através das medidas de qualidade.

Como trabalhos futuros pretende-se realizar o mesmo método para a variância das demais distribuições de probabilidade presentes da literatura, também das sequências no cálculo de valor esperado das distribuições. Realizando também a análise de qualidade de todos os resultados obtidos através das medidas de erro Absoluto e Relativo.

5. REFERÊNCIAS BIBLIOGRÁFICAS

1. Klatt R, Kulisch U, Wiehoff A, Lawo C, Rauch M. C-XSC - A C++ Class Library for Extended Scientific Computing. Springer-Verlag, Heidelberg; 1993.
2. Ferson SL, Ginzburg L, Kreinovich V. Absolute bounds on the mean of sum, product, etc.: A probabilistic extension of interval arithmetic, SIAM WORKSHOP ON VALIDATED COMPUTING, Toronto; 2002.
3. Moore RE. Interval Analysis. Englewood Cliffs: Prentice Hall; 1966.
4. Ratschek H, Rokne J. New Computer Methods for Global Optimization. Ellis Horwood; 1988.
5. Ruggiero MAG, Lopes VLR. Cálculo Numérico: aspectos Teóricos e Computacionais. 2. ed. São Paulo: Makron Books do Brasil; 2004.
6. Tortelli LM, Lorini M, Grehs EA, Balboni MDC, Finger AF, Loreto AB. Extensão Intervalar para a Esperança da Probabilidade. XXII Congresso de Iniciação Científica – UFPel; 2013.
7. Balboni MDC, Tortelli LM, Lorini M, Furlan VS, Finger AF, Loreto AB. Critérios para Análise e Escolha de Ambientes Intervalares, Revista Jr. de Iniciação Científica em Ciências Exatas e Engenharia, FURG, nº7, Rio-Grande; 2014.
8. Walpole RE, Myers RH, Myers SL, YE, K. Probability & Statistics for Engineers & Scientists. Pearson Education International. 2007, p.196.
9. Lorenz MO. Methods of measuring the concentration of wealth. Publications of the American Statistical Association 9; p.209-219, 1905.
10. Intpy. Interval Arithmetic package Online. Disponível em: <https://pypi.python.org/pypi/IntPy> Acessado em: 24/08/2014.
11. Tortelli LM, Balboni MDC, Finger AF, Loreto AB. Extensão Intervalar para a esperança da probabilidade das distribuições exponencial, gama e uniforme. XXVI Congresso Regional de Iniciação Científica e Tecnológico em Engenharia – CRICTE, UNIPAMPA, Alegrete. 2014