

# SIMULAÇÃO DE EXPERIMENTOS DE CAMPO

Autores : MOREIRA, F.R.; SERAPHIN, J.C.

Instituto de Matemática e Estatística

E-mail (Bolsista) - fernandoalunoufg@yahoo.com.br

E-mail (Orientador) - seraphin@mat.ufg.br

Palavras Chave - Simulação, Experimentos, Algoritmo.

## INTRODUÇÃO

Na maioria das áreas de estudo, são conduzidos experimentos com vários objetivos, entre eles obter informações sobre um processo particular ou para comparar efeitos de determinadas condições em particular sobre um fenômeno qualquer. Entretanto quando se deseja comparar metodologias ou avaliar suas eficiências do ponto de vista estatístico, há necessidade de um grande número de experimentos submetidos a diversas situações distintas, o que demandaria um período de tempo longo para se conseguir os dados através de experimentos formais, tornando muitas vezes inviável a realização do estudo. A simulação computacional dos dados utilizando-se de parâmetros que representam as diferentes realidades desejadas, nestes casos, torna viável tais estudos. Vários trabalhos são reportados na literatura usando simulação como meio de obtenção de dados para estudo, como por exemplo Bernhardson (1975), Conagin & Zimmermann(1990) e Silva *et al.*(1999).

O objetivo deste trabalho é apresentar a simulação de números aleatórios com distribuição uniforme  $U(0,1)$ , números aleatórios com distribuição normal e a descrição do processo e do algoritmo de geração de experimentos.

## METODOLOGIA

A simulação de variáveis aleatórias segundo uma distribuição uniforme  $U(0, 1)$  é mostrada através do modelo proposto por Lehmer (1951 *apud* GENTLE, 2003). Quanto a uma variável aleatória com distribuição normal  $N(0, 1)$  a simulação é mostrada a partir da utilização da transformação de Box-Muller (BOX & MULLER, 1958).

O algoritmo descrito para a simulação dos experimentos de campo baseou-se no modelo do delineamento inteiramente casualizado, especificado como

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad (1)$$

onde  $y_{ij}$  = é o valor simulado correspondente a combinação do  $i$ -ésimo tratamento na  $j$ -ésima repetição;  $\mu$  = média geral arbitrada;  $\tau_i$  = efeito do  $i$ -ésimo tratamento ( $i = 1, 2, \dots, I$ );  $\varepsilon_{ij}$ =erro experimental aleatório gerado, sendo que:  $\varepsilon_{ij} \cap N(0, \sigma^2)$ .

## RESULTADOS E DISCUSSÃO

Simulação da Uniforme  $U(0, 1)$

Uma variável aleatória que possui função densidade de probabilidade dada por

$$\begin{aligned} p(x) &= 1 \text{ se } x \in [0, 1] \\ &= 0 \text{ se } x \notin [0, 1] \end{aligned}$$

é definida como variável aleatória uniforme sobre o intervalo  $[0, 1]$

Lehmer (1951 *apud* GENTLE, 2003) propôs um gerador para a simulação de variáveis aleatórias que possuem distribuição uniforme  $U(0, 1)$ . Esse gerador tem por base a *congruência módulo  $m$*  e é dado por

$$x_i \equiv ax_{i-1} \text{mod}(m),$$

com  $0 < x_i < m$ , onde  $a$  e  $m$  são constantes dadas e é necessário oferecer uma semente  $x_0$  para iniciar a geração. A seqüência gerada é chamada seqüência de Lehmer e cada  $x_i$  é lançado no intervalo  $(0,1)$  pela divisão por  $m$ , isto é,

$$u_i = x_i/m.$$

Um número  $a$  é congruente a  $b$  módulo  $m$  ( $a \equiv b \pmod{m}$ ) se existe um  $k$  inteiro tal que  $a - b = k * m$ . O total de números distintos gerados é chamado de período do gerador.

O período do gerador com multiplicador  $a$  e módulo  $m$  é o menor  $k$  positivo que satisfaz a seguinte equação

$$a^k \equiv 1 \pmod{m}.$$

Por exemplo, supondo  $a = 7$  e  $m = 31$ , logo

$$x_i \equiv 7x_{i-1} \pmod{31}.$$

Tomando-se como semente  $x_0 = 19$ , a seqüência obtida é

$$9, 1, 7, 18, 2, 14, 5, 4, 28, 10, 8, 25, 20, 16, 19$$

e a partir do 19 começa a se repetir. O período nesse caso é 15 pois,

$$7^{15} \equiv 1 \pmod{31}$$

e para se obter os valores uniformes basta efetuar a divisão por 31.

Quanto maior for o valor de  $m$  maior será o período do gerador. Devido a capacidade de armazenamento de dados e de rapidez nos cálculos, os computadores nos permitem trabalhar até com  $m = 2^{31}$  gerando assim uma seqüência de Lehmer com muitos valores.

Simulação da Normal  $N(0, 1)$

Para a simulação da normal  $N(0, 1)$  faz-se necessário definir a transformação de Box-Muller (BOX & MULLER, 1958), que a partir de valores de uma variável uniforme  $U(0, 1)$  gera valores de uma normal  $N(0, 1)$ . Se  $U_1$  e  $U_2$  são  $U(0, 1)$  então  $X_1$  e  $X_2$  dados por

$$X_1 = \sqrt{-2 \ln(U_1)} \cos(2\pi U_2)$$

$$X_2 = \sqrt{-2 \ln(U_1)} \sin(2\pi U_2)$$

são independentes e com distribuição  $N(0, 1)$ . Essa transformação requer cálculos com raiz quadrada e função logarítmica, logo temos que esse processo é bastante lento. Uma maneira de aplicar a transformação mais eficientemente é através de um algoritmo descrito em Gentle (2003), apresentado a seguir

- 1 - Gera-se  $v_1$  e  $v_2$  independentes, de uma variável  $U(-1, 1)$ , e toma-se  $r^2 = v_1^2 + v_2^2$ ;
- 2 - Se  $r^2 \geq 1$ , então volta ao passo 1, caso contrário gera-se

$$x_1 = v_1 \sqrt{(-2 \ln r^2)/r^2}$$

$$x_2 = v_2 \sqrt{(-2 \ln r^2)/r^2}$$

O gerador apresentado acima e a transformação de Box-Muller são métodos de simulação de difícil implementação sem o uso de um computador. Entretanto vários softwares têm rotinas de simulação já implementadas.

A distribuição uniforme sobre um intervalo  $[a, b]$  com certeza é a distribuição que é mais simples de trabalhar pela simplicidade de seus cálculos, de uma importância enorme, pois a partir dela se "constrói" muitas distribuições através de transformações relativamente simples, como exemplo a simulação da Normal  $N(0, 1)$  citada acima.

Algoritmo para a geração dos Experimentos

A simulação de experimentos de acordo com o modelo descrito em (1) pode ser obtida a partir da implementação do algoritmo descrito a seguir

matriz com colunas  $y_1 - y_k$ ;  
 para  $i = 1$  até  $a$ ;  
      $\tau_i = f(i)$  (onde  $f$  para cada  $i$  associa o efeito do tratamento);  
 para  $j = 1$  até  $b$ ;  
     para  $k = 1$  até  $c$ ;  
          $x = s*\varepsilon_{ij}$  onde  $\varepsilon_{ij} \in N[0, 1]$  ( $\varepsilon_{ij}$  é o erro experimental)  
          $y = \mu + \tau_i + x$ ;  
     fim; (para o 1º laço)  
 fim; (para o 2º laço)  
 fim; (para o 3º laço)

onde as constantes  $a$ ,  $b$  e  $c \in \mathbb{N}$  são fixas e  $i$  representa o  $i$ -ésimo tratamento,  $j$  a  $j$ -ésima repetição,  $k$  é o  $k$ -ésimo experimento,  $s$  o desvio padrão,  $\tau_i$  o efeito do  $i$ -ésimo tratamento,  $\mu$  média geral arbitrada.

Este algoritmo pode ser adaptado às linguagens de programação atuais.

### CONCLUSÃO

A geração de variáveis que seguem uma certa distribuição conhecida é algo de extrema utilidade e importância no processo de pesquisa estatística e em muitas áreas do conhecimento.

### REFERÊNCIAS BIBLIOGRÁFICAS

- BERNHARDSON C.S. 1975. Type I error rates when multiple comparison procedures follow a significant F test of ANOVA. *Biometrics*, Washington, 31(1):337-340.
- BOX G.E.P. & MULLER M.E. 1958. A note on the Generation of Random Deviates *Ann. Math. Stat.*, 29:610-611.
- CONAGIN A. & ZIMMERMANN F.J.O. 1990. Seleção de materiais nos trabalhos de melhoramento de plantas. II. Poder discriminativo de diferentes testes estatísticos. *Pesq. Agropec. Bras.*, Brasília, 25(10):1415-1428.
- GENTLE J.E. 2003. *Random Number Generation and Monte Carlo Methods*. 2ª ed. New York, Springer, 381 p.
- SILVA E.C., FERREIRA D.F., BEARZOTI E. 1999. Avaliação do poder e taxas de erro tipo I do teste de Scott-Knott por meio do método de Monte Carlo. *Ciência Agrotec.*, Lavras, 23(3): 687-696.