# Sobre a geração de sequências numéricas de distribuição normal

Francisco Vaz

***Resumo* – Neste texto apresenta-se a demonstração de um método corrente de transformação utilizado para gerar sequências numéricas pseudo aletórias com distribuição normal partindo de sequências unifomemente distribuidas.**

***Abstract* – The text presents the proof of a current method to generate pseudo random numbers with a normal distribution from a tranformation of random sequences with an uniform distribution.**

***Palavras chave* – números aleatórios, geração de números aleatórios, distribuição normal.**

***Keywords* – random numbers, random number generation, normal distribuition**

## I. INTRODUÇÃO

A utilização computadores para a simulação de modelos aleatórios exige a geração automática de números que tenham uma distribuição estatística semelhante à gerada pelos sistemas que se pretende emular.

Estão disponíveis variados métodos para gerar diferentes tipos de propriedades estatísticas, particularmente a geração de números que obedeçam a uma distribuição gausseana ou normal. O método mais espalhado parte da geração de números com distribuição uniforme e através da aplicação de algumas transformações obtem-se a desejada transformação. Em geral este método aparece sem qualquer justificação e é por vezes interpretado como uma aproximação. No texto que se segue apresenta-se formalmente a validade da referida metodologia.

## II. GERAÇÃO DE SEQUÊNCIAS NÚMÉRICAS COM DISTRIBUIÇÃO UNIFORME

Habitualmente geram-se números "uniformemente" distribuidos no intervalo $[0, 1]$, significando que a frequência relativa com que são gerados é a mesma para todos.

Num computador, uma máquina de estados finita com capacidade limitada e finita de representação numérica, não será possível gerar números reais arbitrários em $[0, 1]$. Limitam-se pois os ojectivos à geração de um conjunto finito de números inteiros $\{0, 1, \ldots, M-1\}$ que divididos por $M$ darão $\{a_0, a_1, \ldots, a_n\}$, com $0 \leq a_i < 1$. O gerador deve garantir que todos devem aparecer com a mesma frequência, e o seu número $M$ será dependente das necessidades do problema a simular e das capacidades de representação numérica do computador. Quanto maior for $M$ mais "denso" é o conjunto dos números gerados.

O processo mais usado actualmente é o algoritmo recursivo conhecido como **método do resíduo** e que se resume na aplicação da seguinte fórmula recursiva:

$$x_k = a x_{k-1} \bmod M$$

isto é, para gerar uma sequência de números aleatórios basta uma multiplicação, sendo o novo número o resto da sua divisão por $M$, o que do ponto de vista computacional pode ser implementado de maneiras muito eficazes. A necessária divisão por $M$ também pode ser eficazmente implementada num computador.

Os problemas que se poêm têm a ver com a escolha de $a$ e de $M$. Demonstra-se que $M$ deve ser primo ou potência inteira de um primo e $a$ deve ser cuidadosamente escolhido no intervalo $[0, M]$. Um dos problemas na escolha, é que como a sequência gerada é periódica o seu período depende do valor de $a$. Está-se, pois, perante sequências não verdadeiramente aleatórias, mas periódicas, donde a designação de pseudo-aleatórias vulgarmente usada. Mas sendo o período da mesma ordem de $M$, se for muito maior que o comprimento necessário da sequência, o seu comportamento aproxima-se bastante do pretendido.

Considerando as capacidades de representação numérica dos actuais computadores pessoais, é prática generalizada o uso de $M = 2^{31} - 1$, o que dá uma sequência de comprimento 2147 483 647, número que é confortavelmente grande para grande parte das simulações que se fazem.

Para terminar, um breve comentário a $x_0$, o valor inicial para a construção da sequência. Em geral os programas existentes tentam gerá-lo aleatoriamente fazendo, por exemplo, uma leitura do relógio do computador no início da execução, ou pelo menos na primeira execução do programa numa dada sessão. No entanto, em geral permitem introduzir opcionalmente o seu valor (vulgarmente designado pela palavra inglesa *seed* -semente), facilitando a execução de sucessivas realizações da mesma sequência pseudo-aleatória.

## III. GERAÇÃO DE SEQÊNCIAS NUMÉRICAS COM DISTRIBUIÇÃO NORMAL

Seja $V$ uma variável aleatória uniformemente distribuida no intervalo $(0, 1]$ simbolicamente representável por $U(0, 1)$ que tem uma densidade de probabilidade

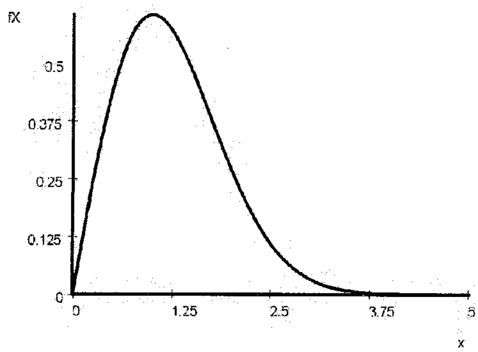
$$f_V(v) = \begin{cases} 1 & v \in (0, 1] \\ 0 & v \notin (0, 1] \end{cases}$$

Considere-se a transformação $x = \sqrt{-2 \ln v}, v \in (0, 1]$

Como $v = e^{-\frac{1}{2}x^2}, x \geq 0$ será $\left|\frac{dv}{dx}\right| = \left|\frac{d}{dx} e^{-\frac{1}{2}x^2}\right| = x e^{-\frac{1}{2}x^2}, x \geq 0$

A variável aleatória $X$ resultante desta transformação terá uma densidade de probabilidade que se obtem usando (3)

$$f_X(x) = \begin{cases} x e^{-\frac{1}{2}x^2} & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{1}$$



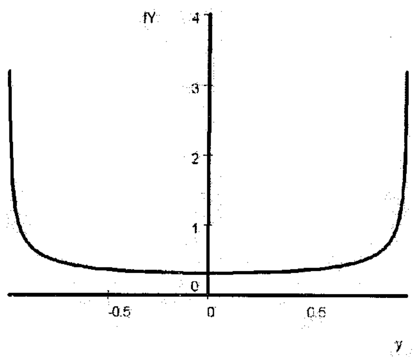

Pode-se concluir então que a aplicação de $x = \sqrt{-2\ln v}, v \in (0,1]$ transforma uma variável uniforme numa variável de Rayleigh de parâmetro 1.

Considere-se a variável aleatória $W \to U(0,1)$ e ainda a transformação $y = \cos(\pi w), w \in (0,1]$.

Como $w = \frac{1}{\pi}\arccos y$ , $|y| \leq 1$ será $\left|\frac{dw}{dy}\right| = \left|\frac{d}{dy}\frac{1}{\pi}\arccos y\right| = \frac{1}{\pi\sqrt{1-y^2}}, |y| \leq 1$

Analogamente a variável aleatória $Y$ resultante desta transformaçãoterá uma densidade de probabilidade que se obtem usando (3)

$$f_Y(y) = \begin{cases} \frac{1}{\pi\sqrt{1-y^2}}, & |y| \leq 1 \\ 0 & |y| > 1 \end{cases} \quad (2)$$



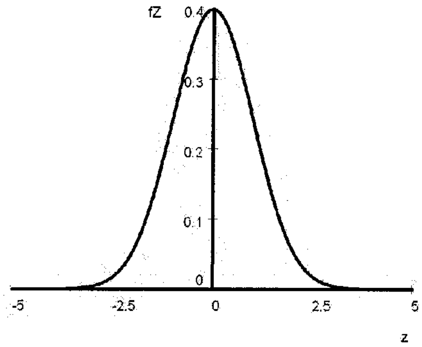

O mesmo resultado seria obtido se se usasse a transformação $y = \sin(\pi w)$ sobre a variável uniforme $W \to U(-0.5, 0.5)$.

Considere-se agora a variável aleatória $Z = XY$, produto das duas variáveis anteriormente consideradas, $Z = \sqrt{-2\ln V}\cos(\pi W)$ Admitindo que $V$ e $W$ são independentes e introduzindo os resultados 3 e 2 na fórmula 4 que caracteriza o produto de variáveis aleatórias obtem-se

$$\begin{aligned} f_Z(z) &= \int_0^{\infty} \frac{1}{y}\frac{1}{\sqrt{(1-z^2/y^2)}} y e^{y^2/2} dy = \\ &= \frac{1}{2\pi}\int_z^{\infty} \frac{2y}{\sqrt{(y^2-z^2)}} e^{-\frac{1}{2}\frac{z^2}{y^2}} dy \end{aligned}$$

Efectuando a transformação de variável $y^2 - z^2 = t$, e tendo em consideração que $\int \frac{1}{\sqrt{t}} e^{-\frac{1}{2}t} dt = \sqrt{2\pi}\left(\text{erf}(\sqrt{t/2}) - 1\right)$ com $\text{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-x^2} dx$

$$\begin{aligned} f_Z(z) &= \frac{1}{2\pi} e^{-\frac{1}{2}z^2} \int_0^{\infty} \frac{1}{\sqrt{t}} e^{-\frac{1}{2}t} dt = \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \left[\text{erf}\left(\sqrt{t/2}\right) - 1\right]_0^{\infty} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \end{aligned}$$




podendo, portanto, concluir-se que a variável aleatória $Z = XY$ é uma variável aleatoria gausseana de média nula e variancia unitária, $N(0,1)$.

## IV. NOTAS FINAIS

É possível demonstrar usando o mesmo tipo de métodos que as transformações $y = \sin(2\pi w)$ e $y = \cos(2\pi w)$ conduzem a resultados semelhantes. Basta recordar que as funções trignométricas inversas não são biunívocas, o que obriga a analisar a transformação em vários intervalos distintos, cada um dos quais equivalente às tranformações anteriores.

Muitos programas fornecem a geração de pares de variáveis normais gerados usando as transformações $x = \sqrt{-2\ln v_1}\sin 2\pi v_2$ e $y = \sqrt{-2\ln v_1}\cos 2\pi v_2$ permitindo gerar um par de variáveis normais $X, Y$ a partir de um par de variáveis $V_1, V_2$ uniformes em $(0,1]$ e independentes.Calculando a sua correlação

$$\begin{aligned} E[XY] &= E[-2\ln V_1 \sin 2\pi V_2 \cos 2\pi V_2] \\ &= E[-\ln V_1 \sin 4\pi V_2] \end{aligned}$$

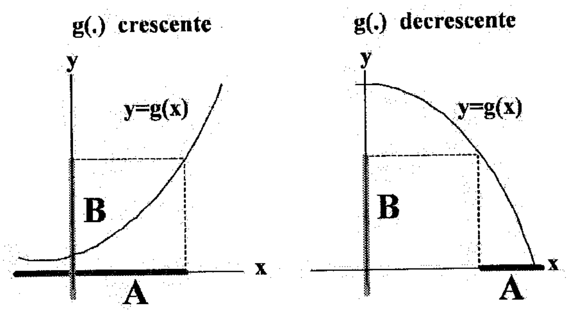
Como $E[\sin 4\pi V_2] = 0$, a correlação $E[XY] = 0$ as variáveis $X, Y$ são ortogonais e, portanto, descorrelacionadas.

Para se obter uma sequência normal com valor médio $\mu$ e desvio padrão $\sigma$, basta fazer uma transformação final $z \to \sigma z + \mu$, como é trivial provar-se.

## V. ANEXO

*Teorema 1:* Seja $g(x)$ uma função real de variável real contínua, com contradomínio em $(a,b) = (g(-\infty), g(+\infty))$ e com inversa $x = g^{-1}(y)$ derivável em todo o seu domínio $(a,b)$ .e $X$ uma variável aleatória real caracterizada por uma função de distribuição $F_X(x)$ e uma função de densidade de probabilidade $f_X(x)$. A variável aleatória $Y = g(X)$ será caracterizada por uma função de densidade de probabilidade $f_Y(y) = f_X(g^{-1}(y))\left|\frac{dx}{dy}\right|$

*Dem:*




Para determinar as relações procuradas vai-se recorrer ao conceito de acontecimentos equivalentes. Atendendo à definição de função de variável aleatória, os acontecimentos $A$ definido sobre $X$ e $B$ definido sobre $Y$ são a imagem de um mesmo acontecimento no espaço de amostragem $S$, sendo portanto equivalentes pelo que terão a mesma probabilidade.

1. Seja $g(x)$ crescente. Como $A = \{X \leq x\} = \{X \leq g^{-1}(y)\}$e $B = \{Y \leq y\}$são equivalentes, as suas probabilidades são iguais

$$P\left[X \leq g^{-1}(y)\right] = P\left[Y \leq y\right]$$

ou seja

$$F_Y(y) = F_X(g^{-1}(y)), y \in (a,b)$$

e calculando a sua derivada em ordem a $y$, $\frac{dF}{dy} = \frac{dF}{dx}\frac{dx}{dy}$ com $y = g(x)$ e $x = g^{-1}(y)$ obtem-se

$$f_Y(y) = f_X(g^{-1}(y))\frac{dx}{dy}, y \in (a,b)$$

2. Seja $g(x)$ decrescente. Analogamente, como $A = \{X \geq x\} = \{X > g^{-1}(y)\}$ e $B = \{Y \leq y\}$são equivalentes, as suas probabilidades são iguais

$$P\left[X > g^{-1}(y)\right] = 1 - P\left[X \leq g^{-1}(y)\right] = P\left[Y \leq y\right]$$

ou seja

$$F_Y(y) = 1 - F_X(g^{-1}(y)), y \in (a,b)$$

e derivando em ordem a $y$ :

$$f_Y(y) = -f_X(g^{-1}(y))\frac{dx}{dy}, y \in (a,b)$$

Como $\frac{dx}{dy} < 0$ quando $g(x)$ é decrescente, a densidade de probabilidade pode representar-se por

$$f_Y(y) = f_X(g^{-1}(y))\left|\frac{dx}{dy}\right|, y \in (a,b) \qquad (3)$$

■

*Teorema 2:* Dadas duas variáveis aleatórias independentes $X$ e $Y$, se $Y > 0$ a variável aleatória produto $Z = XY$ será caracterizada por

$$f_Z(z) = \int_0^\infty \frac{1}{y} f_X(z/y) f_Y(y) dy$$

*Dem:* se $Y > 0$

$$F_Z(z) = P\left[Z \leq z\right] = P\left[XY \leq z\right] = P\left[X \leq z/Y\right]$$

$$P\left[X \leq z/Y\right] = \int \int_{x \leq z/y} f_{XY}(x,y) dx dy =$$

$$= \int_0^\infty f_Y(y) \left[\int_{-\infty}^{z/y} f_X(x) dx\right] dy = \int_0^\infty F_X(z/y) f_Y(y) dy$$

e dentro das condições necessárias de integrabilidade e derivabilidade

$$f_Z(z) = \int_0^\infty \frac{1}{y} f_X(z/y) f_Y(y) dy \qquad (4)$$

■

## VI. BIBLIOGRAFIA

J.E. Gentle: *Random Number Generation and Monte Carlo Methods.* Statistics and Computing Series, Springer Verlag, 1998

A.Papoulis: *Probability & Statiscs*, Prentice Hall International Editions, 1984

F.Vaz: *Probabilidades e Processos Estocásticos para Engenharia Electrotécnica*, Universidade de Aveiro, 2002