Tratamento Estatístico de Dados em Física Experimental (TEFE)

4300228 – Segundo Semestre de 2016

Quando temos uma função densidade de probabilidade que gostaríamos de estudar numéricamente, devemos ser capazes de reproduzí-la no computador. Isso é feito, em geral, de forma simples através do chamado "método da exclusão" ou "método da rejeição"[1]. Contudo, quando o domínio de nossa função densidade de probabilidade não é finito, torna-se impossível utilizar tal método. Uma das soluções para este problema é a utilização do chamado "método da inversão". Aqui comentaremos um pouco mais sobre estes dois métodos, dando exemplos concretos de como utilizá-los e alguns de seus resultados.

Considere a seguinte função densidade de probabilidade

$$f(x) = \begin{cases} Ax \quad , \ se\ 0 \leq x \leq L \\ \\ 0, \ caso\ contrário \end{cases} \tag{1}$$

Para que esta seja realmente uma função densidade de probabilidade, devemos normalizá-la, o que significa

$$\int_{-\infty}^{\infty} dx\, f(x) = 1 = A \int_{0}^{L} dx\, x = \frac{AL^2}{2} \Rightarrow \boxed{A = \frac{2}{L^2}} \tag{2}$$

Tomando $L = 1$ (por exemplo) e utilizando (2), podemos escrever a função de uma maneira que seja possível simulá-la no computador

$$\boxed{f(x) = \begin{cases} 2x \quad , \ se\ 0 \leq x \leq 1 \\ \\ 0, \ caso\ contrário \end{cases}} \tag{3}$$

O esboço desta função encontra-se na Figura 1. Gostaríamos de simular dados de tal forma que estes estejam sujeitos à esta distribuição de probabilidade, ou seja, após simularmos diversos dados, devemos obter um histograma muito próximo dela) quando fizermos um histograma dos dados simulados.

Descreveremos a seguir como obter esta densidade de probabilidade a partir dos métodos previamente mencionados.

[1] Um exemplo de aplicação deste método são as simulações do tipo Monte Carlo.

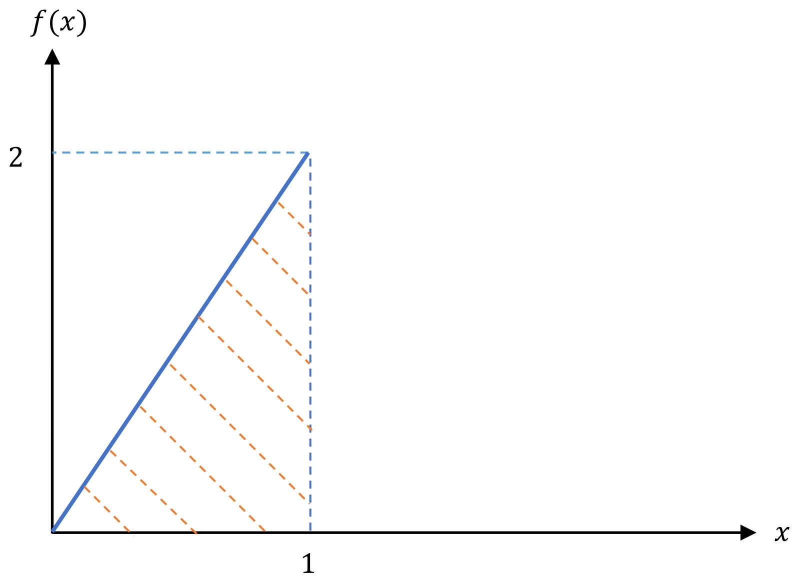

Figura 1 – Esboço da função densidade de probabilidade definida na equação (3).

O Método da Exclusão

O método da exclusão consiste em gerar conjuntos de pontos e submetê-los à algum teste, guardando apenas os pontos que passem no teste, descartando os outros.

Para o caso de nosso problema, devemos gerar pares (x, y) com $x \in [0, 1]$ e $x \in [0, 2]$, e sujeita-los ao teste para verificar se $y \leq 2$x. Caso essa condição seja satisfeita, armazenamos a coordenada x do ponto gerado, caso contrário o dado é excluído e um novo par é gerado. Na prática, estamos gerando pontos que preenchem a área haxurada da figura 1.

Um exemplo da eficácia deste método é fornecido na Figura 2, para uma simulação de 1×10^6 pares de pontos (x, y).

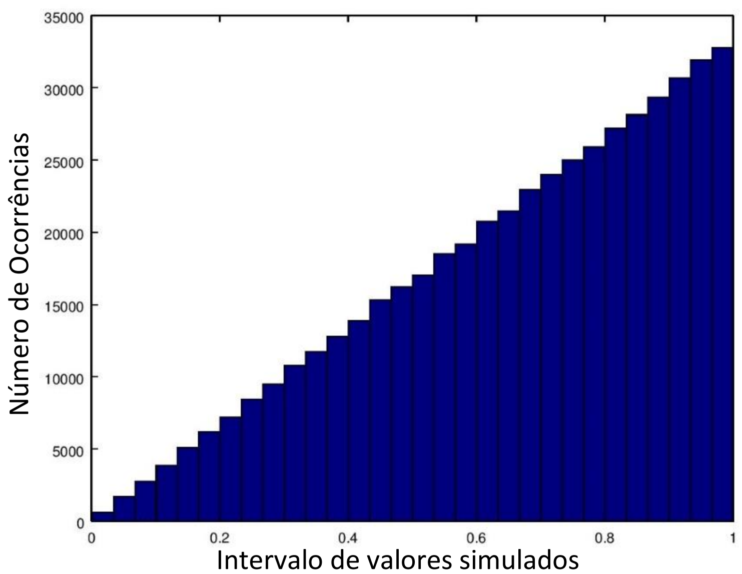

Figura 2: Histograma do número de ocorrência pelo intervalo de valores simulados, através do método da exclusão, para 1×10^6 pares de pontos (x, y) simulados.

O Método da Inversão

O método da inversão utiliza a função probabilidade acumulada, aqui definida como

$$g(x) = \int_{-\infty}^{x} dx' \, f(x') \tag{4}$$

Esta função nos dá a soma das probabilidades de obtermos os valores de x desde $-\infty$ até o valor de x que queremos. Note que, por definição, os valores possíveis para a função probabilidade acumulada estão obrigatóriamente entre 0 e 1 (compare a definição (4) com a condição que utilizamos em (2)). O método visa, então, obter a função inversa g^{-1} e utilizar esta função para gerar a função densidade de probabilidade que queiremos.

Apliquemos, agora, este método para o nosso caso específico. Temos

$$g(x) = \int_{0}^{x} dx' 2x' = x^2 \Rightarrow \boxed{g^{-1} = x = \sqrt{g(x)}} \tag{5}$$

Como sabemos, agora, que $0 \leq g(x) \leq 1$, podemos gerar números entre 0 e 1 e utilizar a equação (5) para gerar dados que sigam a função de probabilidade que queremos. O exemplo mostrado na Figura 3 corresponde a uma simulação de 1×10^6 pontos gerados.

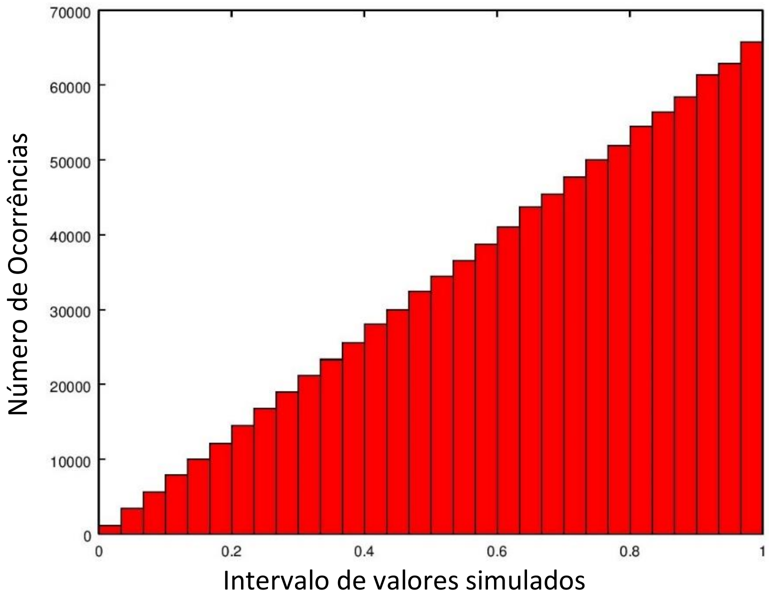

Figura 3: Histograma do número de ocorrência pelo intervalo de valores simulados, através do método da inversão, para 1×10^6 pontos simulados.