ACOTAÇÃO DE ERROS NO MÉTODO DE MONTECARLO

Campoy Vázquez, Carlos

Universidade da Corunha
Área de Electromagnetismo - Departamento de Física
Escola Universitária de Arquitectura Técnica - Campus da Zapateira
Corunha. e-mail: campoy@edu.xumta.es

RESUMO

Quando se utilizam outros métodos de integração aproximada (rectângulos, trapézios, Simpson etc.) o valor do resultado acompanha-se do correspondente intervalo de incerteza dentro do qual, *com segurança total*, estará o valor exacto desconhecido. Para o método de Montecarlo não é possível a construção de tal intervalo pois a segurança total perde-se como consequência da natureza probabilística do método. Porém há a possibilidade de fornecer um intervalo de incerteza junto com um determinado coeficiente de confiança como se faz na apresentação de medições físicas, onde os erros aleatórios são impossíveis de evitar ou de quantificar com certeza absoluta.

1. Intervalo de incerteza e nível de confiança

Existem procedimentos de cálculo aproximado que nos permitem acotar o erro com um 100% de confiança, por exemplo no cálculo aproximado da raiz de uma equação pelo método da bissecção, ou no cálculo aproximado da integral definida de uma função contínua num intervalo fechado pelo método das sumas inferiores e superiores de Riemann. Nesses casos o resultado pode exprimir-se:

$$a < L < b \quad \text{ou} \quad M - \varepsilon < L < M + \varepsilon \quad \text{sendo} \quad \varepsilon = \frac{|b-a|}{2} \quad \text{e} \quad M = \frac{|b+a|}{2}$$

Diferente é o caso em que apresentamos o resultado da medição de uma grandeza física. Uma vez refinado o procedimento para conseguir a ausência total de erros sistemáticos ou enviesamentos de qualquer tipo, ficam inevitavelmente os erros aleatórios, cujas causas são impossíveis de controlar pelo experimentador. Como consequência deles ao repetir a mesma medição iremos achar valores ligeiramente diferentes. Há poderosas razões para modelizar a população dos infinitos resultados possíveis da mesma medição como se seguissem a distribuição normal. Então se os resultados são:

$$X_1, X_2, \cdots, X_N \qquad \text{calculamos} \qquad M = \frac{\sum X_i}{N} \qquad S^2 = \frac{\sum (X_i - M)^2}{N-1}$$

e podemos determinar a probabilidade de que uma medição posterior, seja X_{N+1}, fique entre os valores arbitrários a e b:

$$p(a < X_{N+1} < b) = \frac{1}{S\sqrt{2\pi}} \cdot \int_a^b e^{-\frac{1}{2}\left(\frac{x-M}{S}\right)^2} \cdot dx$$

Quer dizer, uma distribuição normal $N(M,S)$ (Em realidade a distribuição adequada seria a t de Student, mas no que segue suporemos N bastante grande para que os valores de uma ou outra distribuição sejam praticamente idênticos)

O importante é constatar que qualquer resultado é possível, de modo que não pode acotar-se um intervalo de incerteza dentro do qual teríamos a segurança absoluta de achar o verdadeiro valor da grandeza cuja medida se está a procurar.

O que acontece é que estamos a ser excessivamente exigentes por pedir um nível de confiança do 100%. Veremos como se acha um intervalo de incerteza com uma confiança não tão desmedida de que o valor verdadeiro se ache lá.

2. Erros aleatórios e intervalos de confiança

Se os infinitos resultados da medição se distribuem como a normal $N(M,S)$, os valores de M para infinitas séries diferentes $X_1, X_2, \cdots, X_N$ vão-se distribuir como a normal $N(M,\frac{S}{\sqrt{N}})$, e o valor verdadeiro desconhecido que estamos a procurar, L, estará, com uma certa probabilidade, no intervalo:

$$p\left(M - 2{,}58 \cdot \frac{S}{\sqrt{N}} < L < M + 2{,}58 \cdot \frac{S}{\sqrt{N}}\right) = 0{,}99$$

$$p\left(M - 1{,}96 \cdot \frac{S}{\sqrt{N}} < L < M + 1{,}96 \cdot \frac{S}{\sqrt{N}}\right) = 0{,}95$$

$$p\left(M - 1{,}64 \cdot \frac{S}{\sqrt{N}} < L < M + 1{,}64 \cdot \frac{S}{\sqrt{N}}\right) = 0{,}90$$

$$p\left(M - \frac{S}{\sqrt{N}} < L < M + \frac{S}{\sqrt{N}}\right) = 0{,}68$$

Os coeficientes que precedem ao desvio padrão calculam-se facilmente utilizando as tabelas da distribuição normal padronizada. Na literatura científica apresenta-se apenas o último dos intervalos na forma:

$$L = M \pm \frac{S}{\sqrt{N}}$$

e o leitor já sabe o que tem que fazer se quiser uma confiança superior ao 68%

3. O método de Montecarlo para as quadraturas

.

Para descrever o método utilizarei uma função tão simples que não requeriria método nenhum. Pretendo assim ganhar em brevidade. Procuro o valor aproximado de

$$\int_0^1 x \cdot dx$$

cujo valor exacto sabemos que é 1/2. No quadrado $\{(x, y) | 0 < x, y < 1\}$, o gráfico da função sub-integral $f(x) = x$ é a diagonal ascendente e o valor que estou a procurar é a área do triângulo entre o lado inferior, a diagonal e o lado da direita.

Para o cálculo utilizo uma série de N pares de números aleatórios x_i, y_i, cada um deles compreendido entre 0 e 1. Se dos N pontos correspondentes, T estão dentro do triângulo, a proporção $\rho = T/N$ será o valor aproximado da área do triângulo que fornece o método de Montecarlo.

4. Intervalo de confiança para a proporção

Para a proporção dos indivíduos de uma população que cumprem determinado requisito, a teoria dos intervalos de confiança fornece resultados análogos aos dados acima para o valor médio. Neste caso seria (L é o valor desconhecido da integral):

$$p\left(\rho - 2{,}58\sqrt{\frac{\rho(1-\rho)}{N}} < L < \rho + 2{,}58\sqrt{\frac{\rho(1-\rho)}{N}}\right) = 0{,}99$$

$$p\left(\rho - 1{,}96\sqrt{\frac{\rho(1-\rho)}{N}} < L < \rho + 1{,}96\sqrt{\frac{\rho(1-\rho)}{N}}\right) = 0{,}95$$

$$p\left(\rho - 1{,}64\sqrt{\frac{\rho(1-\rho)}{N}} < L < \rho + 1{,}64\sqrt{\frac{\rho(1-\rho)}{N}}\right) = 0{,}90$$

$$p\left(\rho - \sqrt{\frac{\rho(1-\rho)}{N}} < L < \rho + \sqrt{\frac{\rho(1-\rho)}{N}}\right) = 0{,}68$$

e do mesmo modo que anteriormente, apresentaremos o resultado assim:

$$\int_0^1 x \cdot dx = \rho \pm \sqrt{\frac{\rho(1-\rho)}{N}}$$

5. Conclusão

O método de Montecarlo pode-se explicar na cadeira *Métodos Estatísticos e Numéricos* de 2º de bacharelato, e seguramente resultará muito mais divertido para os alunos utilizar um par de dados em lugar de recorrer a uma tabela de números aleatórios, pelo menos na primeira vez. O erro, mesmo quando se utiliza o coeficiente 2,58 , não garante com total segurança que o valor procurado vai pertencer ao intervalo conseguido, mas a natureza do método não permite mais. Porém é formativo para o aluno compreender este modo de proceder que é o próprio das ciências experimentais.

Bibliografia

[1] Rodríguez Alvarez, Francisco e Garcia Suárez, Xenaro *Métodos Estatísticos e Numéricos*. Xerais, Vigo 1997

[2] Campos Guimarães, Rui e Sarsfield Cabral, José A., *Estatística*, Ed. McGraw-Hill de Portugal, Lda., Lisboa 1997