Tratamento Estatístico de Dados em Física Experimental

Prof. Zwinglio Guimarães

2º semestre de 2019

Tópico 1 – 2ª semana:
Simulação de experimentos por Monte Carlo

O Método de Monte Carlo - revisão

- Método estatístico baseado amostragens aleatórias para se obter resultados numéricos
- Possibilita obter soluções numéricas para problemas cuja abordagem analítica é muito difícil ou até mesmo impossível
  - Usados em diversas áreas do conhecimento
  - **Pode ser usado para propagação de incertezas e estimativa da distribuição de probabilidades de resultados de experimentos ("Toy Monte Carlo")**

Propagação de incerteza por "Toy Monte Carlo" (TMC) - revisão

$\boldsymbol{w(x) = 5x^2}$, com $\boldsymbol{x = 10,0 \pm 0,2}$. Qual a incerteza de $\boldsymbol{w}$?

Possíveis valores de x são gerados por:

$$\boldsymbol{x = 10,0 + 0,2 * e1}$$

Onde $\boldsymbol{e1}$ é uma variável aleatória de valor esperado 0 e desvio-padrão 1, com distribuição igual aos erros de $\boldsymbol{x}$.

A ideia do TMC é sortear possíveis valores para $\boldsymbol{x}$ e calcular os valores de $\boldsymbol{w}$ correspondentes.

O desvio-padrão de $\boldsymbol{w}$ é a incerteza de $\boldsymbol{w}$ por causa da incerteza de $\boldsymbol{x}$.

Considerando dados gaussianos no Octave:

```
x0=10;
sx=0.2;
N=1000;
w = zeros( N,1 );
for i = 1:N
 x = x0 + sx*randn;
 w(i) = 5*(x^2);
end
sw = std( w )
```

TMC para simular erros aleatórios e sistemáticos - revisão

Suponha um experimento em que foram medidos $N = 25$ dados de valor verdadeiro $x_0 = 50,0$ sujeitos a erros aleatórios de desvio-padrão $\sigma_a = 2,0$ e a um erro sistemático residual de desvio-padrão $\sigma_s = 0,5$, ambos com distribuição gaussiana.

No Octave, um conjunto de dados desse tipo é simulado por:

```
x0 = 50;
sA = 2.0;
sS = 0.5;
N = 25;
x = zeros(N,1);
erroSistematico = sS*randn;
for i = 1:N
   x(i) = x0 + erroSistematico + sA*randn;
end
```

Simulação de experimentos completos incluindo a análise de dados

- O Método de Monte Carlo pode ser usado para simular experimentos de maneira mais completa
- Para isso, é preciso considerar desde a relação dos dados com os parâmetros do modelo até os procedimentos de análise de dados a serem utilizados
  - Prever a incerteza final de um experimento
    - Para comparar com estimativas obtidas por outros métodos
  - Avaliar se o procedimento de análise não é tendencioso
    - Para corrigir eventuais tendenciosidades
    - Para decidir entre procedimentos de análise distintos

Exemplo A – Incerteza no resultado de experimentos sujeitos a erros aleatórios e sistemáticos

Qual é a incerteza no resultado obtido em um experimento com N dados sujeitos a erros aleatórios de desvio-padrão σ_a e a erros sistemáticos de desvio-padrão σ_s (ambos com distribuição gaussiana)?

1ª Etapa: escrever uma rotina para simular os N dados medidos em uma realização do experimento e devolver o resultado obtido (o valor médio dos N dados)

no Octave:

```
function [ xf  ] = simExpA( N, sA, sS, x0)

  x = zeros(N,1);
  erroSist = sS * randn;
  for i = 1:N
      erroAleat = sA * randn;
      x(i) = x0 + erroSist + erroAleat;
  end
  xf = mean(x);
```

2ª Etapa: Usar a rotina para simular M repetições do experimento. A incerteza final é dada pelo desvio-padrão amostral desses M possíveis resultados:

no Octave, considerando os valores do exemplo 4 da revisão:
($N = 25$, $s_A = 2$, $s_S = 0.5$, $x_0 = 50$)

```
M = 1e4;
X = zeros( M, 1 );
for j = 1:M
   X(j) = simExpA( 25, 2, 0.5, 50);
end
sigma_final = std( X )
```

Exemplo B – Incerteza provocada por arredondamentos

Qual é a incerteza provocada pelo fato dos dados medidos estarem arredondados até a casa da unidade?

O efeito do arredondamento pode ser avaliado gerando-se possíveis valores verdadeiros com distribuição uniforme e os correspondentes valores medidos considerando o arredondamento.

Os erros (diferença entre os valores medidos e os gerados) podem ser usados para avaliar a incerteza associada ao arredondamento, pois o desvio-padrão dos erros reflete a incerteza provocada apenas pelo fato dos valores terem sido arredondados.

no Octave:

Valores verdadeiros gerados usando ***rand*** (distribuição uniforme entre 0 e 1)

```
N = 1e4;
e = zeros( N, 1 );
for i = 1:N
    x0 = 1000*rand;
    x = round( x0 );
    e(i) = x – x0;
end
sigma_arredond = std( e )
```

Exemplo C – Incerteza (e tendenciosidade) provocada por <u>truncamentos</u>

Qual é a incerteza e a tendenciosidade provocada pelo fato dos dados medidos serem **truncados** na casa da unidade?

Como no caso anterior, o efeito do truncamento pode ser avaliado gerando-se possíveis valores verdadeiros com distribuição uniforme e os correspondentes valores medidos considerando o truncamento.

Os erros (diferença entre os valores medidos e os gerados) podem ser usados para estimar a incerteza (desvio-padrão dos erros) e a tendenciosidade (valor médio dos erros) associada apenas ao fato dos valores terem sido truncados.

no Octave:

```
N = 1e4;
e = zeros( N, 1 );
for i = 1:N
    x0 = 1000*rand;
    x = fix( x0 );
    e(i) = x – x0;
end
sigma_truncamento = std( e )
tendenciosidade = mean( e )
inc_tend = std( e ) / sqrt( N )
```

Exemplo D – E se desconsiderarmos o maior e o menor dos valores de uma série de N dados?

Qual é o efeito (incerteza e tendenciosidade) no resultado obtido em um experimento com N dados se desconsiderarmos o maior e o menor valor medido para calcular o valor médio? (só erros aleatórios)

1ª Etapa: escrever uma rotina para simular os N dados medidos em uma realização do experimento e devolver o resultado obtido usando o procedimento proposto

no Octave:
(supondo erros gaussianos)

```
function [ xf  ] = simExpD( N, sA, x0)

  x = zeros(N,1);
  for i = 1:N
      x(i) = x0 + sA * randn;
  end
  xf = ( sum(x) - max(x) - min(x) ) / (N-2);
```

2ª Etapa: Usar a rotina para simular M repetições do experimento. A incerteza é o desvio-padrão e a tendenciosidade é a diferença entre a média e o valor verdadeiro:

no Octave: ($N = 10$, $s_A = 1$)

```
M = 1e4;
X = zeros( M, 1 );
for j = 1:M
  X(j) = simExpD( 10, 1, 20);
end
sigma_final = std( X )
tendenciosidade = mean( X ) – 20
inc_tendenc = std( X ) / sqrt( M )
```