Como os computadores *sorteiam* números?

Geração de números pseudoaleatórios

ESTAT0090 – Estatística Computacional

Prof. Dr. Sadraque E. F. Lucena

sadraquelucena@academico.ufs.br

Cenário

Você está desenvolvendo um novo algoritmo para prever o comportamento do mercado financeiro ou o desempenho de uma nova vacina. Para ter certeza de que seu algoritmo é robusto e confiável, você precisa testá-lo com dados. No entanto, obter dados reais em grande volume pode ser demorado, caro ou inviável. Em vez de esperar por dados reais, você precisa de uma forma de simular esses dados. Isso permite que você:

- Crie bancos de dados sintéticos que imitam as características do mundo real.
- Avalie o desempenho do seu algoritmo sob diversas condições controladas.
- Implemente métodos estatísticos avançados (como Bootstrap ou Monte Carlo) que exigem muitas repetições ou amostras.
- Desenvolva e valide modelos complexos onde as distribuições de probabilidade não são facilmente observáveis ou manipuláveis.

Com o conhecimento sobre como gerar números pseudoaleatórios, você terá a ferramenta fundamental para criar seus próprios cenários de dados, testar suas ideias rapidamente e validar seus modelos de forma eficiente e reproduzível, sem depender exclusivamente da realidade.

Objetivos da aula

Na aula de hoje aprenderemos a:

- Distinguir números aleatórios de pseudoaleatórios.
- Compreender e aplicar os Métodos Congruencial Linear Multiplicativo e Congruencial Misto para gerar sequências numéricas.

Introdução

Há duas classes fundamentais de geração de número aleatórios:

1. **Números verdadeiramente aleatórios:** gerados usando algum fenômeno físico que é aleatório.
    - Exemplos clássicos incluem lançar uma moeda, jogar dados ou sortear números de uma urna.
    - Métodos modernos utilizam efeitos quânticos, ruído térmico em circuitos elétricos, o tempo de decaimento radioativo, entre outros.
2. **Números pseudoaleatórios:** gerados por algoritmos computacionais.
    - Embora esses métodos sejam normalmente rápidos e eficientes em recursos, um desafio com essa abordagem é que os programas de computador são intrinsecamente determinísticos e, portanto, não podem produzir uma saída realmente aleatória.

Neste curso focaremos na geração de números pseudoaletórios.

Número pseudoaleatórios

- Formalmente, definimos números pseudoaleatórios como uma sequência de valores gerados por um processo determinístico (ou seja, previsível e baseado em um conjunto de regras), mas que aparenta ser obtida de variáveis genuinamente aleatórias e independentes com distribuição uniforme entre 0 e 1.
- Essa sequência de números são obtidas por meio de expressões matemáticas aplicadas de forma recursiva, podendo ser utilizados diferentes métodos.

> **Nota**
>
> A partir de uma sequência de números com distribuição uniforme podemos gerar realizações de variáveis aleatórias de qualquer outra distribuição de probabilidade.

- Vejamos alguns métodos para a geração de números pseudoaleatórios.

Método Congruencial Linear Multiplicativo

Para esse método precisamos definir:

- um valor inicial x_0 (conhecido como **semente**);
- dois números interios positivos a e m.

Algoritmo

Passo 1: Calcule recursivamente o próximo valor $x_n, n > 1$, usando

$$x_n = ax_{n-1} \mod m.$$

Passo 2: Obtenha um valor entre 0 e 1 fazendo $y_n = x_n/m$.

- A função $\mod$ no Passo 1, calcula o resto da divisão de ax_{n-1} por m.
- O valor a é chamado de multiplicador e m é denominado de módulo.

Exemplo 9.1

Considere $x_0 = 11, a = 2$ e $m = 16$ Os três primeiros valores da sequência (y_1, y_2, y_3) gerada pelo Método Congruencial Linear Multiplicativo são:

- $x_1 = (2 \times 11) \bmod 16 = 6 \quad \Rightarrow \quad y_1 = 6/16 = 0.375$
- $x_2 = (2 \times 6) \bmod 16 = 12 \quad \Rightarrow \quad y_2 = 12/16 = 0.75$
- $x_3 = (2 \times 12) \bmod 16 = 8 \quad \Rightarrow \quad y_3 = 12/16 = 0.5$

Os três primeiros valores da sequência são $0.375, 0.75$ e 0.5.

- Agora calcule os próximos 3 números da sequência.

Método Congruencial Linear Multiplicativo

- Note que o Método Congruencial Linear Multiplicativo gera sempre valores de x_n entre 0 e $m - 1$.
- Isto implica que após um número finito de valores gerados, a sequência se repete (isto é chamado *período* de um gerador).
- Os valores de a e m devem então ser escolhidos de modo a gerar a maior sequência possível:
  - Em computadores de 32 bits $m = 2^{31} - 1$ e $a = 7^5$ resulta em uma boa sequência.
  - Em computadores com 36 bits uma boa escolha seria $m = 2^{35} - 31$ e $a = 5^5$.

Geradores Congruenciais Mistos

- Outros geradores de números pseudoaleatórios usam recursão do tipo

$$x_n = (ax_{n-1} + c) \mod m.$$

- Além do multiplicador a e do módulo m, chamamos c de incremento.
- O período de um gerador congruencial misto é no máximo m.
- Quando um gerador consegue gerar todos os valores de 0 a $m - 1$, dizemos que ele tem período completo. Esse gerador terá um período completo para todas as sementes se, e somente se,
  - o módulo m e o incremento c forem relativamente primos (números inteiros que não têm nenhum fator primo em comum, exceto o 1);
  - $a - 1$ for divisível por todos os fatores primos de m (todo número primo que divide m também divide $a - 1$);
  - $a - 1$ for divisível por 4 se m for divisível por 4.

Curiosidades

- Os *softwares* R e Python utilizam um gerador conhecido como *Mersenne Twister*.
  - Esse gerador foi desenvolvido por Makoto Matsumoto e Takuji Nishimura nos anos de 1996 e 1997 para eliminar as falhas dos diferentes geradores existentes.
  - Ele possui a vantagem de apresentar o maior período dentre os métodos implementados ($2^{19937} - 1 \approx 4{,}3154 \times 10^{6001}$).
  - É um dos mais rápidos geradores existentes, embora complexo, e faz uso de forma muito eficiente da memória.
- A ciência da computação possui uma área dedicada ao estudo de geradores de números aleatórios.
- No R, o gerador de número pseudo-aleatório e a semente podem ser alterados com a função `set.seed()` (verifique o *help* desta função).

Exemplo 9.2

Gere os seis primeiros valores de uma sequência de números pseudoaleatórios usando o Método Congruencial Misto considerando a semente $x_0 = 11$, o multiplicador $a = 7$, o incremento $c = 5$ e o módulo $m = 20$.

Atividade

Escreva uma função em R que gere uma sequência de n números aleatórios usando o método congruencial misto. Em seguida teste para $n = 10$, $x_0 = 11$, $a = 7$, $c = 5$ e $m = 20$.

Ganhos da aula

- Compreensão do que são números aleatórios e pseudoaleatórios.
- Conhecimento dos principais métodos para gerar sequências numéricas (Congruencial Linear Multiplicativo e Misto).

Atividade Extraclasse

Pratique o que aprendeu! Resolva os exercícios complementares que estão disponíveis junto com este material.

Fim

Esta aula foi baseada no capítulo 3 – *Random Numbers*, do livro *Simulation (Sixth Edition)*, de Sheldon M. Ross, 2023.