Introdução aos Números Pseudo-aleatórios

Profa. Dra. Soraia Raupp Musse

Conceito:

- Um **gerador de número pseudo-aleatório** é um algoritmo que gera uma seqüência de números, os quais são aproximadamente independentes um dos outros.
- A saída da maioria dos geradores de números aleatórios não é verdadeiramente aleatória; ela somente aproxima algumas das propriedades dos números aleatórios. John von Neumann enfatiza com este comentário "Qualquer um que considere métodos aritméticos para produzir dígitos está, certamente, cometendo um pecado".

Conceito:

- Enquanto números verdadeiramente aleatórios podem ser gerados usando hardware para geração de número aleatório, número pseudo-aleatórios são uma parte crítica da computação moderna, da criptografia até o método de Monte Carlo passando por sistemas de simulação. Uma cuidadosa análise matemática é necessária para assegurar que a geração dos números seja suficientemente "aleatória".

Conceito:

- Na computação, um **hardware gerador de número pseudo-aleatório** é um aparato que gera números aleatórios a partir de um processo físico. Estes dispositivos são normalmente baseados em fenômenos como ruído térmico, no efeito fotoelétrico ou outro fenômeno quântico. Estes processos são, em teoria, completamente imprevisíveis, e a afirmação de imprevisibilidade está sujeita a testes experimentais.

Para que serve?

- Números aleatórios são úteis em uma variedade de situações, como na simulação de fenômenos físicos, fumaças, nuvens
- Ainda na amostragem de populações, na programação de computadores, na tomada de decisões ou até mesmo em entretenimento (bingos, loterias ou jogos).

Para que serve?

- Na área de simulação, consideremos por exemplo, a modelagem do tempo de acesso de um disco rígido, num computador pessoal. Podemos determinar que a duração desse evento irá cair numa faixa conhecida, digamos de 0 a 200ms, de acordo com características físicas inerentes ao próprio disco rígido.
- Entretanto, o valor real desse evento vai depender de vários fatores, como a posição da cabeça de leitura quando a requisição é feita pelo sistema operacional, detalhes da implementação do suporte e até mesmo da temperatura e condições ambientais. Podemos considerar então que esse tempo de acesso é uma variável aleatória seguindo uma distribuição conveniente.
- Para fazermos essa simulação precisamos de números aleatórios que sigam uma dada distribuição, e para isso precisamos saber primeiro como gerar esses números aleatórios.

Fontes de números aleatórios:

- Algumas fontes de números aleatórios são o lançamento de dados, a retirada de bolas numeradas de uma urna (com reposição), o uso de uma roleta ou ainda ruído eletrônico cuja saída é quantizada periodicamente.
- Entretanto na esmagadora maioria das vezes usa-se o que foi convencionado chamar de números pseudo-aleatórios.

Características dos números pseudo-aleatórios:

- Possibilidade de repetição de seqüências, se desejado
- Seguem distribuição uniforme
- Geração rápida com baixo custo computacional

A geração

- O uso de um algoritmo para gerar um numero aleatório parece violar o princípio básico da aleatoriedade, por isso é que se convenciona chamar esses números de sintéticos ou pseudo-aleatórios.
- A geração começa sempre de um valor inicial chamado semente (seed)

Algoritmos e testes

- Algoritmos para geração de números aleatórios
- Testes de aleatoriedade: fornecem uma maneira de fazer avaliação quantitativa da aleatoriedade de uma dada sequência de números

Método do Quadrado do meio

- Esse método foi inventado por John Von Neumann. Começa-se com uma seed, esse numero é então elevado ao quadrado, e os dígitos do centro são usados como próximo elemento da seqüência.
- Caso o numero de dígitos que fique a esquerda seja maior que os que fiquem a direita não há problema, simplesmente fixamos para qual lado vamos fazer o corte.

Exemplo:

- Começando a partir de x0 = 44214 vamos gerar uma seqüência de números aleatórios de 5 dígitos:
  - x0 = 44214
  - $(44214)^2$ = 1954877796) x1 = 48777
  - $(48777)^2$ = 3011485129) x2 = 14851
  - $(14851)^2$ = 0131905225) x3 = 19052
  - $(19052)^2$ = 0362978704) x4 = 29787
  - $(29787)^2$ = 0887265369) x5 = 72653

Desvantagens:

- Sequências geradas se repetem
- Quando um zero é gerado, todos os outros da sequência são também zero

Exemplo:

$$
\begin{array}{rcll}
x_0 & = & 121 & \\
(121)^2 & = & 014641 & \Rightarrow \quad x_1 = 464 \\
(464)^2 & = & 215296 & \Rightarrow \quad x_2 = 529 \\
(529)^2 & = & 279841 & \Rightarrow \quad x_3 = 984 \\
(984)^2 & = & 968256 & \Rightarrow \quad x_4 = 825 \\
(825)^2 & = & 680625 & \Rightarrow \quad x_5 = 062 \\
(062)^2 & = & 003844 & \Rightarrow \quad x_6 = 384 \\
(384)^2 & = & 147456 & \Rightarrow \quad x_7 = 745 \\
(745)^2 & = & 555025 & \Rightarrow \quad x_8 = 502 \\
(502)^2 & = & 252004 & \Rightarrow \quad x_9 = 200 \\
(200)^2 & = & 040000 & \Rightarrow \quad x_{10} = 000 \\
(000)^2 & = & 000000 & \Rightarrow \quad x_{11} = 000
\end{array}
$$

Como avaliar a aleatoriedade?

- Por exemplo:
  - 1,2,3,4,5,... É aleatório?
- E se nós formos gerar números aleatórios?
  - Vamos criar seqüências com algum significado
  - Vamos evitar repetições e números parecidos, quando isso na verdade pode ocorrer

Teste de aleatoriedade

- Presume-se que a sequência não é aleatória
- Realiza-se testes de aleatoriedade

Dada uma variável X para a qual temos uma amostra de valores, desejamos verificar se essa varíavel se adequa ou não a uma dada distribuição. Dividimos esses valores em k categorias, montando a tabela de freqüência:

Categoria	1	2	3	....	k
Freq. Observada	o_1	o_2	o_3	...	o_k

A partir do modelo que supomos ser adequado, construímos a tabela de freqüências esperadas de acordo com as categorias acima:

Categoria	1	2	3	....	k
Freq. Esperada	e_1	e_2	e_3	...	e_k

Teste do Chi-quadrado

Sexo	F_0	F_e
Tipo 1	13	12
Tipo 2	11	12
Total	**24**	**24**

$$x_e^2 = \frac{\sum \left(\left| F_0 - F_e \right| - 0{,}5 \right)^2}{F_e}$$