Análise de Dados e Simulação

Márcia D'Elia Branco

Universidade de São Paulo
Instituto de Matemática e Estatística
http:www.ime.usp.br/ mbranco

Introdução a Simulação.

Números aleatórios

- Vamos assumir que temos um bom gerador de números (pseudo) aleatórios.
- O que denominados sequência de números aleatórios são realizações independentes de uma v.a. Uniforme no intervalo (0,1).
- O algoritmo computacional para simular esses números aleatórios é sempre determinístico. O bom algoritmo gera uma sequência que pareça aleatória (pseudo aleatória).
- A partir deste valor aleatório básico iremos simular diversos modelos probabilísticos discretos e contínuos.
- Nesta disciplina todos os programas de simulação deverão ser elaborados na linguagem **R** (http://www.r-project.org/).

Métodos de Monte Carlo

- Valores aproximados de integrais podem ser obtidos utilizando-se simulação. Esses métodos são conhecidos como Monte Carlo.
- Por exemplo, se $U \sim Unif(0,1)$ e temos interesse em calcular E[U], simulamos uma amostra grande de valores de U e utilizamos a média da amostra como uma aproximação do valor da esperança.
- O procedimento acima é justificado pela Lei dos Grandes Números.
- Procedimento similar pode ser utilizado para obter $E[g(X)]$, em que X é uma v.a. qualquer. Neste caso simulamos de X, calculamos $g(x)$ para cada valor simulado e obtemos a média dos $g(x)$ calculados.

Métodos de Monte Carlo

- Exemplos em sala de aula.
- Como determinar o valor de π (pi) via simulação de Monte Carlo? Desenvolvido em sala de aula.

Considere uma v.a. X tal que

$$P(X = x_j) = p_j \quad , \quad j = 0, 1, 2, \ldots \quad \text{e} \quad \sum_{j=1}^{\infty} p_j = 1.$$

Idéia geral do algoritmo:

- Simula $U \sim U_{(0,1)}$. Se $U < p_0$ faça $X = x_0$.
- Se $\sum_{j=0}^{k-1} p_j \leq U < \sum_{j=0}^{k} p_j$ faça $X = x_k$, $k = 1, \ldots$.

Observações:

- O algoritmo é mais eficiente de ordenamos de forma decrescente os p_i´s. Por que?
- Este método é denominado Método da Transformada Inversa. O nome se ajusta melhor ao uso do método em v.a. contínuas.

Simulação de V.A. Discreta

Caso especial: Uniforme em $\{1, 2 \ldots, n\}$

$$P(X = j) = \frac{1}{n} \ , \ j = 1, 2, \ldots, n.$$

Note que, usando a algoritmo anterior se $U \sim U_{(0,1)}$

$$X = j \ \Leftrightarrow \ j - 1 \leq nU < j$$

Podemos reescrever como

$$X = Int(nU) + 1$$

onde $Int(a)$ representa a parte inteira do número a.

Aplicação: Simulando Permutações Aleatórias (sala de aula).

1. V.A. Geométrica

$$P(X = i) = pq^{i-1} \quad , \; i \geq 1 \; , \; q = 1 - p.$$

Função distribuição acumulada
$F(x) = \sum_{i=1}^{x} P(X = i) = 1 - P(X > x) = 1 - q^x$
Usando o algoritmo da Transformada Inversa se $U \sim U_{(0,1)}$

$$X = j \;\Leftrightarrow\; 1 - q^{x-1} \leq U < 1 - q^x$$

Equivalente a
$X = min\{x : q^x < 1 - U\} = min\{x : x > \frac{log(1-U)}{logq}\}$
Portanto,

$$X = Int\left(\frac{logU}{logq}\right) + 1$$

2. V.A. Poisson

$$p_i = P(X = i) = \frac{\lambda^i}{i!} e^{-\lambda} \; i = 0, 1, \ldots \quad \lambda > 0.$$

Um modo eficiente de obter os valores da f.d.acumulada é considerar a seguinte relação recursiva

$$p_{i+1} = \frac{\lambda}{i+1} p_i \;, \; i \geq 0.$$

Algoritmo:
(i) Simula $U \sim U_{(0,1)}$
(ii) Faça $i = 0, p = e^{-\lambda}$ e $F = p$
(iii) Se $U < F$ faça $X = i$ e pare.
(iv) Faça $p = \frac{\lambda p}{(i+1)}, F = F + p$ e $i = i + 1$.
(v) Volte ao passo (iii).

3. V.A. Binomial

$$p_i = P(X = i) = \frac{n!}{i!(n-i)!} p^i (1-p)^{(n-i)} \quad i = 0, 1, \ldots, n \ .$$

Um modo eficiente de obter os valores da f.d.acumulada é considerar a seguinte relação recursiva

$$p_{i+1} = \frac{n-i}{i+1} \frac{p}{1-p} p_i.$$

Algoritmo:

(i) Simula $U \sim U_{(0,1)}$
(ii) Faça $c = \frac{p}{(1-p)}, i = 0, pr = (1-p)^n$ e $F = pr$
(iii) Se $U < F$ faça $X = i$ e pare.
(iv) Faça $pr = \frac{c(n-i)}{(i+1)} pr, F = F + pr$ e $i = i + 1$.
(v) Volte ao passo (iii).

Comentários

- Uma alternativa para simulação de $X \sim binomial(n, p)$ é usar a sua interpretação como soma de Bernoulli independentes.
- Como simular uma sequência de Bernoulli independentes? Desenvolvido em aula.

Método de Rejeição-Aceitação

- Suponha que exista um algoritmo eficiente para simular de uma v.a.discreta Y com probabilidades $q_0, q_1, \ldots$.
- Desejamos simular de uma outra v.a.discreta X com probabilidades $p_0, p_1, \ldots$.
- Seja c uma constante tal que $\frac{p_j}{q_j} \leq c$ para todo $p_j > 0$.
- Simulamos y de Y e um número aleatório U. Se $U < \frac{p_y}{cq_y}$ faça $X = y$ e pare. Caso contrário, repita o procedimento.
- $\frac{p_y}{cq_y}$ é denominada probabilidade de aceitação do valor y.
- $1/c$ é a probabilidade de aceitação do algoritmo (média).
- Por que funciona?

Método da Composição

Considere
$X = X_1$ com prob. α e $X = X_2$ com prob. $(1-\alpha)$.
Note que

$$P(X = j) = \alpha p_j^{(1)} + (1-\alpha) p_j^{(2)}$$

com $p_j^{(1)}$ associado a X_1 e $p_j^{(2)}$ a X_2.
Se $U < \alpha$ simulamos de X_1. Caso contrário, simulamos de X_2.

Exemplo:
$P(X = j) = 0.05$ para $j = 1, 2, 3, 4, 5$ e
$P(X = j) = 0.15$ para $j = 6, 7, 8, 9, 10$
Fazemos $p_j^{(1)} = 0.1,\ j = 1, 2, \ldots, 10$ e $p_j^{(2)} = 0.2,\ j = 6, 7, 8, 9, 10$.
Note que $P(X = j) = 0.5 p_j^{(1)} + 0.5 p_j^{(2)}$