Análise de Dados e Simulação

Márcia D'Elia Branco

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

Aplicações de Técnicas de Redução de Variância.

Primeira situação:
Clientes chegam segundo um PPH com taxa λ. O sistema tem capacidade limitada a N clientes. Aqueles que chegam quando sistema esta na sua capacidade máxima (ocupado) são descartados. Seja L o número de clientes descartados.
Interesse: $\theta = E[L]$
Construindo um estimador usando o técnica do condicionamento.

$$T = E[L \mid T_c] = \lambda T_c$$

com T_c o tempo total que o sistema ficou ocupado.

Na simulação basta controlar o tempo que o sistema esteve ocupado.

Segunda situação:
Similar a antes mas agora os clientes chegam segundo um PP Não Homogêneo. A taxa não é mais constante ao longo do tempo e não podemos obter o mesmo resultado pra esperança condicional.
Vamos estabelecer outras variáveis auxiliares:
n_c número de intervalos com sistema ocupado.
$I_1, I_2, I_3, \ldots, I_{n_c}$ intervalos de tempo com sistema ocupado.
Novo estimador

$$T = E[L \mid n_c, I_1, \ldots, I_{n_c}] = \sum_{i=1}^{n_c} \int_{I_i} \lambda(s) ds.$$

Observe que T é não-viesado para θ.

Sistema de Filas

Obtendo outro estimador não-viesado pra segunda situação. Se R é o número de clientes que chegam no sistema e M o número de clientes atendidos, então

$$R = M + L$$

Portanto,

$$\theta = E[L] = \int_0^t \lambda(s)ds - E[M]$$

Para obter um estimador não-viesado pra θ basta definir

$$W = \int_0^t \lambda(s)ds - M.$$

Qual dos dois é mais eficiente?

Vamos trabalhar com a combinação linear dois dois estimadores W e T que também é não-viesado e possui variância menor ou igual.

$$Z = p \sum_{i=1}^{n_c} \int_{I_i} \lambda(s)ds + (1-p)\left[\int_0^t \lambda(s)ds - M\right]$$

com n_c número de intervalos ocupados
M número de clientes atendidos
I_i i-ésimo intervalo ocupado.
Quanto vale p? Devemos obter p^* que minimize a vairância de Z.
Resulta em

$$p^* = \frac{Var(W) - Cov(T,W)}{Var(T) + Var(W) - 2Cov(T,W)}$$

Esses valores podem ser estimados via simulação, usando as medidas amostrais.

Células com Câncer

Um tecido doente inicialmente possue n células com câncer e m células normais. Este tecido é submetido a radioterapia com o objetivo de acabar com as células doentes.
Seja N o número de células normais vivas no instante em que todas as células doentes são eliminadas.
Interesse: $\theta = P(N \geq k)$ para algum $k \in \{1, 2, \ldots, m\}$
Seja S o conjunto de todas as células vivas num determinado instante de tempo t. Cada célula nesse conjunto terá probabilidade de morrer igual a

$$p_i = \frac{w_i}{\sum\limits_{j \in S} w_j}$$

Isso ocorre quando o tempo de vida de cada células, denotado por X_i, tem distribuição Exponencial com taxa w_i (inverso da média) independente um do outro, $i = 1, \ldots, n+m$.

Células com Câncer

Considere $T^{(k)}$ o primeiro instante de tempo em que menos de k células normais estão vivas.
Note que $T^{(1)} > T^{(2)} > \cdots > T^{(m+1)}$.
Se ordenarmos de forma decrescente os tempos de vidas das células normais, temos que $T^{(k)}$ será igual ao valor na k-ésima posição.
Além disso,

$$\theta = P(N \geq k) = P(\max_{i \leq n} X_i < T^{(k)})$$

Condicionando em $T^{(k)}$ temos

$$P(N \geq k \mid T^{(k)}) = P(\max_{i \leq n} X_i < T^{(k)} \mid T^{(k)}) =$$

$$\prod_{i=1}^{n} P(X_i \leq T^{(k)}) = \prod_{i=1}^{n} \left(1 - e^{-w_i T^{(k)}}\right).$$

Algoritmo de simulação:

(i) Simular X_i da $Exp(w_i)$ para $i = n+1, n+2, \ldots, n+m$.
(ii) Ordenamos os valores X_i´s de forma decrescente e obtemos na k-ésima posição o valor de $T^{(k)}$.
(iii) Calculamos $X = \prod_{i=1}^{n} (1 - e^{-w_i T^{(k)}})$

OBS: O algoritmo pode ainda ser melhorado considerando-se a técnica de variáveis negativamente correlacionada.
Como?

Exercícios Extras

1. Considere o exercício 8 da Lista 3. Desenvolva um estudo de simulação para estimar o número de médio de clientes perdidos.

2. Considere o exercício 2 da Aula 1 de Filas (entregue junto com a lista 3). Suponha que os trabalhos entram na fila somente se houver 10 ou menos trabalhos esperando. Desenvolva um estudo de simulação para estimar o número médio de trabalhos perdidos.

3. Implemente o algoritmo desenvolvido para "'Células com Câncer"' considerando $n = 10, m = 90, k = 45$ e $w_i = 1/2$ para $i = 1, \ldots, n + m$.