ACH2053 - Introdução à Estatística
Conteúdo Teórico:
12 - Simulação

Marcelo S. Lauretto

Referências:
Morris DeGroot, Mark Schervish. Probability and Statistics. 4th Ed. - 4o capítulo

Ilya M. Sobol. A Primer for the Monte Carlo Method. CRC Press, 1994.

www.each.usp.br/lauretto

Problema 1: Programa de Monty Hall I

- Você está participando de um programa de televisão, onde tem a chance de ganhar um bom prêmio
- São apresentadas três portas fechadas, sendo que em apenas uma delas está o prêmio (colocado de forma aleatória), sendo que o apresentador sabe onde ele está.
- O jogo:
  - Você escolhe uma porta inicial;
  - O apresentador abre uma porta sem o prêmio e lhe pergunta se você quer trocar ou não
  - Você deve decidir se troca de porta ou não
- Qual é a estratégia com maior chance de ganho?
- No Capítulo 2 calculamos analiticamente que a probabilidade de ganho com a estratégia de trocar de porta é de $2/3$. O raciocínio é apresentado a seguir.

Problema 1: Programa de Monty Hall II

- Variáveis aleatórias do experimento:
  - $X \in \{1,2,3\}$: porta onde está o prêmio
  - $Y \in \{1,2,3\}$: 1^a porta escolhida por você
  - Espaço amostral:
    $(1,1),(1,2),(1,3)$,
    $(2,1),(2,2),(2,3)$,
    $(3,1),(3,2),(3,3)$
  - Variável aleatória de interesse:

$$Z = \begin{cases} 1 \text{ se você ganhar o prêmio} \\ 0 \text{ caso contrário} \end{cases}$$

- Com a estratégia de trocar sempre de porta, $Z = 0$ somente se $X = Y$, ou seja, em $1/3$ dos pares possíveis.
  Se considerarmos que todos os 9 pares são equiprováveis, então $P(Z = 0) = 3/9 = 1/3 \Rightarrow P(Z = 1) = 2/3$.

Problema 1: Programa de Monty Hall III

- Note que Z é uma variável aleatória com distribuição de Bernoulli, e portanto $E(Z) = P(Z = 1)$ (ver o Capítulo 4). Logo, calcular $P(Z = 1)$ é o mesmo que calcular $\mu_Z = E(Z)$.
- Estimando $\mu_Z = E(Z)$ através de simulação:

1. Para $i = 1, ..., N$ (onde N é um número grande):
   - Gere independentemente valores para as variáveis categóricas X e Y, cada qual com distribuição uniforme nos inteiros $1, 2, 3$; denotamos esses dois valores $x_{(i)}$ e $y_{(i)}$ – onde (i) indexa a i-ésima iteração;
   - Verifique se $x_{(i)} \neq y_{(i)}$; em caso afirmativo, atribua $z_{(i)} = 1$; caso contrário, atribua $z_{(i)} = 0$.
2. Calcule $\hat{\mu}_Z = \frac{\sum_{i=}^{N} z_{(i)}}{N}$.

Problema 1: Programa de Monty Hall IV

- Cada iteração $i = 1, ..., N$ simula uma realização hipotética (ou tentativa) do jogo.
- O valor $\hat{\mu}_Z$, que é simplesmente a proporção de ganhos em relação ao total de tentativas, será uma aproximação para μ_Z e, portanto, para $P(Z = 1)$.
  (A notação $\hat{\mu}_Z$ com chapéu denota que estamos obtendo uma estimativa para μ_Z)

Problema 2: Sistema de atendimento I

- Considere um sistema simples para atendimento de requisições com n "linhas" (ou "canais").
- O sistema recebe requisições chegando em instantes aleatórios: $T_1 < T_2 < \ldots < T_k < \ldots$.
- Seja T_k o instante de chegada da k-ésima requisição. Se a linha 1 estiver livre em T_k, ela inicia o atendimento da requisição, o que consome t_h (t_h é o *tempo de espera* da linha). Se a linha 1 estiver ocupada, a requisição é imediatamente transferida para a linha 2, e assim sucessivamente...
- Se todas as n linhas estiverem ocupadas no instante T_k, o sistema rejeita a requisição.
- O problema é determinar quantas requisições (em média) o sistema conseguirá satisfazer durante um intervalo de tempo T, quantas serão rejeitadas e a proporção de requisições rejeitadas nesse intervalo.

Problema 2: Sistema de atendimento II

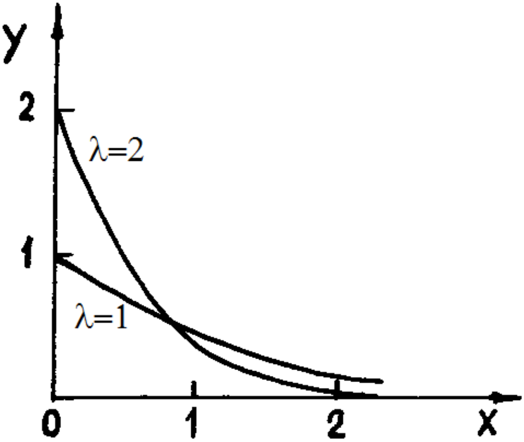
- A modelagem do fluxo de requisições é usualmente feita através de observações de sistemas similares sob longos períodos, sob várias condições.
- Por simplicidade, consideraremos, para este problema, que a sequência de requisições é um *processo de Poisson*, no qual o intervalo entre dois eventos (requisições) consecutivos é uma variável aleatória independente com função de densidade de probabilidade

$$f(x|\lambda) = \lambda \exp(-\lambda x),\ 0 < x < \infty\ .$$

- Essa densidade é denominada *distribuição exponencial* (exemplos no próximo slide); no Capítulo 3, apresentamos o método da transformação inversa para gerar variáveis aleatórias com distribuição exponencial.
- Apresentamos na sequência um método de simulação para estimação dos números esperados de requisições aceitas e rejeitadas.

Problema 2: Sistema de atendimento III

- Exemplos de distribuições exponenciais com $\lambda = 1$ e $\lambda = 2$.



Problema 2: Sistema de atendimento IV

- Parâmetros do programa:
  - n: número de linhas
  - λ: parâmetro da distribuição exponencial, modela a taxa de entrada de requisições
  - t_h: tempo de atendimento de cada linha
  - T: intervalo de tempo total sobre o qual se deseja calcular as médias de aceitações e rejeições
  - N: número de iterações da simulação.
- Variáveis do programa e condições iniciais:
  - T_r: instante da última requisição de entrada; inicialmente, $T_r = 0$.
  - t: vetor com n posições, onde $t[j] \geq 0$ denota o instante em que a linha j estará disponível. Inicialmente, todas as linhas estão disponíveis, e portanto $t[j] = 0, \ j = 1, \ldots, n$.
  - x: contador de requisições aceitas; inicialmente, $x = 0$.
  - y: contador de requisições rejeitadas; inicialmente, $y = 0$.
  - w: proporção de requisições rejeitadas: $w = y/(x + y)$

Problema 2: Sistema de atendimento V

- Simulação de uma única sequência de requisições no intervalo de tempo T:

  1. Gere o intervalo de tempo da próxima requisição: $z \sim \text{expon}(\lambda)$; se $T_r + z > T$, interrompa e retorne x, y e $w = y/(x+y)$.
  2. Atribua $T_r = T_r + z$
  3. Verifique se $t[1] \leq T_r$; em caso afirmativo, isso significa que a linha está livre no instante T_r e pode atender a requisição; nesse caso, atribua $t[1] = T_r + t_h$;
     caso contrário, teste $t[2] \leq T_r, t[3] \leq T_r$, e assim sucessivamente.
  4. Se ao menos uma das linhas estava disponível no passo anterior, incremente x; se nenhuma linha estava disponível, a requisição deve ser rejeitada e, portanto, y deve ser incrementada.
  5. volte ao passo 1.

Problema 2: Sistema de atendimento VI

- Simulação completa:
  1. Execute o procedimento descrito no slide anterior N vezes, obtendo, na i-ésima chamada, a tupla $(x_{(i)}, y_{(i)}, w_{(i)})$.
  2. Os valores médios de X, Y, W serão estimados, respectivamente, por

$$\hat{\mu}_X = \frac{\sum_{i=}^{N} x_{(i)}}{N}, \ \hat{\mu}_Y = \frac{\sum_{i=}^{N} y_{(i)}}{N}, \ \hat{\mu}_W = \frac{\sum_{i=}^{N} w_{(i)}}{N}.$$

Problema 3: Sistema de atendimento com clientes impacientes I

- Considere a seguinte variante do Problema 2:
- O sistema é um balcão de atendimento com n guichês (“linhas”), e os clientes (“requisições”) chegam de acordo com um processo de Poisson com uma taxa λ.
- Quando todos os guichês estão ocupados, forma-se uma fila única na qual o 1^o cliente da fila é atendido pelo 1^o guichê a ficar disponível disponível;
- Cada cliente que chega conta o comprimento r da fila e decide ir embora imediatamente (sem entrar na fila) com probabilidade $p_r = r/(r+n)$, para $r = 1, 2, \dots$
- Quando o cliente decide entrar na fila, ele aguarda em sua ordem de chegada até seu atendimento.
- O tempo de atendimento a cada cliente, depois que sua chegada no balcão, é uma variável aleatória exponencial com parâmetro μ.
- Todos os tempos de atendimento são independentes uns dos outros e também independentes dos tempos de chegada.

Problema 3: Sistema de atendimento com clientes impacientes II

- Perguntas:
  - Qual o número esperado de clientes atendidos até o instante T?
  - Qual o número esperado de clientes que foram embora até o instante T?
  - Qual a proporção de clientes que foram embora?
  - Qual o comprimento esperado da fila no instante T?
  - Considerando apenas os clientes que foram efetivamente atendidos, qual o valor esperado do tempo máximo de permanência dos clientes desde sua chegada até o término de seu atendimento?
- Parâmetros do programa de simulação:
  - n: número de guichês
  - λ: taxa de entrada de clientes
  - μ: taxa de atendimentos a clientes por cada guichê
  - T: intervalo de tempo total sobre o qual se deseja calcular as médias de aceitações e rejeições
  - N: número de iterações da simulação.

Problema 3: Sistema de atendimento com clientes impacientes III

- Variáveis do programa e condições iniciais:
  - T_c: instante de chegada do último cliente até o momento; inicialmente, $T_c = 0$.
  - gt_{disp}: vetor com n posições; $gt_{disp}[j] \geq 0$ denota o instante em que o guichê j estará disponível. Inicialmente, todos os guichês estão disponíveis, e portanto $gt_{disp}[j] = 0, \; j = 1, \ldots, n$.
  - k: contador de clientes que entraram na fila até o momento; inicialmente, $k = 0$.
  - ct_{cheg}: vetor de tamanho variável em que $ct_{cheg}[k] > 0$ denota o instante em que o k-ésimo cliente chegou.

Problema 3: Sistema de atendimento com clientes impacientes IV

- x: contador de clientes já atendidos; inicialmente, $x = 0$.
- y: contador de clientes que foram embora sem entrar na fila; inicialmente, $y = 0$.
- r: comprimento atual da fila; inicialmente, $r = 0$
- w: proporção de clientes que foram embora: $w = y/(x + y + r)$
- tm: tempo máximo de permanência dentre todos os clientes atendidos até o momento.

Problema 3: Sistema de atendimento com clientes impacientes V

- Simulação de uma única realização de atendimentos no intervalo de tempo T:

  1. Gere o intervalo de tempo de chegada do próximo cliente:
     $z \sim \text{expon}(\lambda)$;
     se $T_c + z > T$, interrompa e retorne x, y, r, w e tm.
  2. Atribua $T_c = T_c + z$; $k = k + 1$; $ct_{cheg}[k] = T_c$
  3. Enquanto $\min(gt_{disp}) \leq T_c$ e $x < k$:
     // existe algum guichê livre para atender o 1^o cliente da fila
     3.1 Atribua $x = x + 1$; $j = \arg\min(gt_{disp})$
     3.2 Gere o tempo de atendimento do guichê j: $a \sim \text{expon}(\mu)$
     3.3 Atribua $gt_{disp}[j] = \max\{gt_{disp}[j], ct_{cheg}[x]\} + a$
     3.4 Atribua $tm = \max\{tm, (gt_{disp}[j] - ct_{cheg}[x])\}$
  4. Atribua $r = \max\{0, (k-1) - x\}$ // não considera o último cliente
  5. Gere o indicador de que o cliente que acabou de chegar irá embora: $s \sim Ber(p_r)$ onde $p_r = r/(r+n)$
     Se $s = 1$, atribua $k = k - 1$; $y = y + 1$
  6. Atribua $r = k - x$; $w = y/(x + y + r)$
  7. volte ao passo 1.

Problema 3: Sistema de atendimento com clientes impacientes VI

- Comentários sobre o procedimento acima:
  - A condição de parada na linha 1 ocorre quando o instante em que o último cliente chega ($T_c + z$) é posterior ao intervalo de tempo total analisado (T).
  - Na linha 2, atualiza-se temporariamente o contador de clientes a entrar na fila, bem como o instante de chegada do cliente que acabou de chegar. (O contador k será posteriormente decrementado na linha 4 se esse cliente for embora.)
  - As linhas 3.1–3.4 processam o atendimento do primeiro cliente da fila, no intervalo de tempo transcorrido entre o penúltimo e o último clientes a chegarem.
  - A linha 5 simula o evento s do cliente que chegou por último ir embora, com probabilidade p_r proporcional ao comprimento da fila: $p_r = r/(r+n)$. Se $s = 1$, deve-se decrementar k.

Problema 3: Sistema de atendimento com clientes impacientes VII

- Simulação completa:
  1. Execute o procedimento descrito no slide anterior N vezes, obtendo, na i-ésima chamada, a tupla $(x_{(i)}, y_{(i)}, r_{(i)}, w_{(i)}, tm_{(i)})$.
  2. As médias
     $\hat{\mu}_X = \sum_{i=1}^{N} x_{(i)}/N$, $\hat{\mu}_Y = \sum_{i=1}^{N} y_{(i)}/N$,
     $\hat{\mu}_R = \sum_{i=1}^{N} r_{(i)}/N$, $\hat{\mu}_W = \sum_{i=1}^{N} w_{(i)}/N$
     $\hat{\mu}_{tm} = \sum_{i=1}^{N} tm_{(i)}/N$
     fornecem as estimativas necessárias.