Algoritmos de Monte Carlo e Cadeias de Markov – CPS767 2023/1

Prof. Daniel R. Figueiredo

Segunda Lista de Exercícios

Dica: Para ajudar no processo de aprendizado responda às perguntas mostrando o desenvolvimento das respostas.

Questão 1: Cauda do dado em ação
Considere um icosaedro (um sólido Platônico de 20 faces) honesto, tal que a probabilidade associada a cada face é 1/20. Considere que o dado será lançado até que um número primo seja observado, e seja Z a variável aleatória que denota o número de vezes que o dado é lançado. Responda às perguntas abaixo:

1. Determine a distribuição de Z, ou seja $P[Z = k], k = 1, 2, \ldots$. Que distribuição é esta?
2. Utilize a desigualdade de Markov para calcular um limitante para $P[Z \geq 10]$.
3. Utilize a desigualdade de Chebyshev para calcular um limitante para $P[Z \geq 10]$.
4. Calcule o valor exato de $P[Z \geq 10]$ (dica: use probabilidade complementar). Compare os valores obtidos.

Questão 2: Moedas
Você tem duas moedas: uma honesta e outra enviesada que produz cara com probabilidade 3/4. Uma das duas moedas é escolhida aleatoriamente e lançada n vezes. Seja S_n o número total de caras observadas nas n jogadas. Responda às perguntas abaixo:

1. Determine o valor esperado da fração de caras que será observada.
2. Podemos determinar com certeza a moeda escolhida, depois da mesma ser lançada n vezes?
3. Determine o valor de n tal que tenhamos 95% de chance de acertar qual moeda foi escolhida.

Questão 3: Sanduíches
Você convidou 64 pessoas para uma festa e agora precisa preparar sanduíches para os convidados. Você acredita que cada convidado irá comer 0, 1 ou 2 sanduíches com probabilidades 1/4, 1/2 e 1/4, respectivamente. Assuma que o número de sanduíches que cada convidado irá comer é independente de qualquer outro convidado. Quantos sanduíches você deve preparar para ter uma confiança de 95% de que não vai faltar sanduíches para os convidados?

Questão 4: Graus improváveis
Considere o modelo de grafo aleatório de Erdös-Rényi (também conhecido por $G(n, p)$), onde cada possível aresta de um grafo rotulado com n vértices ocorre com probabilidade p, independentemente. Responda às perguntas abaixo:

1. Determine a distribuição do grau do vértice 1 (em função de n e p).
2. Determine o valor γ (em função de n e p) tal que com alta probabilidade $(1 - 1/n)$ o grau observado no vértice 1 é menor ou igual a γ.

Questão 5: Calculando uma importante constante
Seja X_i uma sequência i.i.d. de v.a. contínuas uniformes em $[0, 1]$. Seja V o menor número k tal que a soma das primeiras k variáveis seja maior do que 1. Ou seja, $V = \min\{k \mid X_1 + \cdots + X_k \geq 1\}$.

1. Escreva e implemente um algoritmo para gerar uma amostra de V.
2. Escreva e implemente um algoritmo de Monte Carlo para estimar o valor esperado de V.
3. Trace um gráfico do valor estimado em função do número de amostras. Para qual valor seu estimador está convergindo?

Questão 6: Transformada inversa
Mostre como o método da transformada inversa pode ser usado para gerar amostras de uma v.a. contínua X com as seguintes densidades:

1. Distribuição exponencial com parâmetro $\lambda > 0$, cuja função densidade é dada por $f_X(x) = \lambda e^{-\lambda x}$, para $x \geq 0$.
2. Distribuição de Pareto com parâmetros $x_0 > 0$ e $\alpha > 0$, cuja função densidade é dada por $f_X(x) = \frac{\alpha x_0^\alpha}{x^{\alpha+1}}$, para $x \geq x_0$.

Questão 7: Contando domínios na Web
Quantos domínios web existem dentro da UFRJ? Mais precisamente, quantos domínios existem dentro do padrão de nomes *http://www.[a-z](k).ufrj.br*, onde *[a-z](k)* é qualquer sequência de caracteres de comprimento k ou menor? Construa um algoritmo de Monte Carlo para estimar este número.

1. Descreva a variável aleatória cujo valor esperado está relacionado com a medida de interesse. Relacione analiticamente o valor esperado com a medida de interesse.
2. Implemente o método de Monte Carlo para gerar amostras e estimar a medida de interesse. Para determinar o valor de uma amostra, você deve consultar o domínio gerado para determinar se o mesmo existe (utilize uma biblioteca www para isto).
3. Assuma que $k = 4$. Seja $\hat{w}_n$ o valor do estimador do número de domínios após n amostras. Trace um gráfico de $\hat{w}_n$ em função de n para $n = 1, \ldots, 10^4$ (ou mais, se conseguir). O que você pode dizer sobre a convergência de $\hat{w}_n$?

Questão 8: Rejection Sampling
Considere o problema de gerar amostras de uma v.a. $X \sim \text{Binomial}(1000, 0.2)$.

1. Descreva uma proposta simples de função de probabilidade para gerar amostras de X usando Rejection Sampling. Calcule a eficência dessa proposta.
2. Lembrando que a distribuição Binomial tem a forma de sino, centrada em sua média, proponha outra função de probabilidade para gerar amostras de X usando Rejection Sampling. Calcule a eficiência dessa proposta e compare com a eficiência acima. O que você pode concluir?

Questão 9: Integração de Monte Carlo e Importance Sampling
Considere a função $g(x) = e^{-x^2}$ e a integral de $g(x)$ no intervalo [0,1].

1. Implemente um método de Monte Carlo simples para estimar o valor da integral.

2. Intuitivamente, muitas amostras de $g(x)$ vão ter valores muito baixos. Dessa forma, utilize Importance Sampling para melhorar a qualidade do estimador do valor da integral. Em particular, utilize a função de densidade $h(x) = Ae^{-x}$ definida em $[0, 1]$ onde A é o valor da constante de normalização. Mostre como gerar amostras de $h(x)$.

3. Compare os dois métodos. Trace um gráfico do erro relativo de cada um dos estimadores em função do número de amostras. Ou seja $|\hat{I}_n - I|/I$ onde I é o valor exato da integral e $\hat{I}_n$ é o valor do estimador com n amostras, para $n = 10^1, 10^2, \ldots, 10^6$.