Algoritmos de Monte Carlo e Cadeias de Markov – CPS 767
2019/1

Prof. Daniel R. Figueiredo

Terceira Lista de Exercícios

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

Questão 1: Calculando $\sqrt{2}$
Vimos em aula um algoritmo de Monte Carlo para calcular o valor de π utilizando a relação entre áreas. Inspirado nesta mesma ideia, construa um algoritmo de Monte Carlo para calcular o valor de $\sqrt{2}$.

1. Descreva a variável aleatória cujo valor esperado está relacionado com $\sqrt{2}$. Obtenha analiticamente o valor esperado da sua variável aleatória.

2. Calcule analiticamente a variância dessa variável aleatória.

3. Implemente o método de Monte Carlo para gerar amostras da sua variável aleatória, calculando a média amostral M_n e utilizando-a para estimar $\sqrt{2}$. Sua função deve usar o algoritmo de Mersenne Twister para geração de números pseudo-aleatórios uniformes entre 0 e 1 (usar este algoritmo em todos os problemas da lista).

4. Seja $\hat{e}_n$ o valor do estimador após n amostras. Trace um gráfico do erro relativo do estimador, ou seja $|\hat{e}_n - \sqrt{2}|/\sqrt{2}$ em função de n, para $n = 1, \ldots, 10^6$ (utilize escala $\log - \log$ no gráfico). O que você pode concluir?

Questão 2: Transformada inversa
Utilize o método da transformada inversa para gerar amostras de uma v.a. 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 3: 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. Obtenha analiticamente o valor esperado da sua variável aleatória.

2. Calcule a variância dessa variável aleatória.

3. Implemente o método de Monte Carlo para gerar amostras da sua variável aleatória. Ou seja, você deve consultar o domínio gerado para determinar se o mesmo existe (utilize uma biblioteca para isto).

4. 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 4: Gerando amostras Normais
Seja Z uma variável aleatória com distribuição Normal com média 0 e variância 1. Em particular, a função densidade de Z é dada por $f_Z(x) = 1/(\sqrt{2\pi})e^{-x^2/2}$, com $-\infty < x < \infty$. Repare que Z assume valores positivos e negativos, mas com caudas que possuem a mesma probabilidade. Ou seja, $P[Z \geq z] = P[Z \leq -z]$, para todo $z \geq 0$. Construa um gerador de números aleatórios para Z. Dica: Utilize o método de amostragem por rejeição e a distribuição exponencial!

Questão 5: Estimando somas
Considere o problema visto em aula, de aplicar o método de Monte Carlo para estimar o valor de $G_N = \sum_{i=1}^{N} i \log(i)$. Use sua intuição para encontrar uma função de probabilidade proponente, $h(i)$, que tenha variância inferior ao melhor estimador visto em aula.

1. Assuma que $N = 1000$. Calcule numericamente o segundo momento do seu estimador.
2. Implemente o método de Monte Carlo para estimar o valor de G_N. Trace um gráfico do erro relativo do estimador, em função de $n = 1, \dots, 10^6$ (calcule o valor exato da soma para determinar o erro relativo).

Questão 6: Integração de Monte Carlo
Considere a função $f(x) = x^\alpha$ com $\alpha > 0$. Defina $g(\alpha, a, b) = \int_a^b f(x)dx$ com $0 \leq a < b$, como sendo a integral de $f(x)$ no intervalo $[a, b]$. Iremos calcular g usando Monte Carlo.

1. Determine analiticamente o valor de $g(\alpha, a, b)$. Dica: relembre Cálculo I.
2. Descreva a variável aleatória cujo valor esperado está relacionado com $g(\alpha, a, b)$. Obtenha analiticamente o valor esperado da sua variável aleatória.
3. Implemente o método de Monte Carlo para gerar amostras da sua variável aleatória, calculando a média amostral M_n e utilizando-a para estimar $g(\alpha, a, b)$. Repare que α, a, b são parâmetros do seu programa.
4. Seja $\hat{g}_n$ o valor do estimador após n amostras. Trace um gráfico do erro relativo do estimador, ou seja $|\hat{g}_n - g(\alpha, a, b)|/g(\alpha, a, b)$ em função de n, para $n = 1, \dots, 10^6$ (utilize escala $\log - \log$ no gráfico). Utilize os seguintes valores para os parâmetros: $\alpha = \{1, 2, 3\}$, $a = 0$, $b = \{1, 2, 4\}$. O que você pode concluir em relação ao erro e os parâmetros?

Questão 7: Gerando subconjuntos
Considere $S_{k,n}$ um espaço amostral dado por todos os subconjunto de tamanho k dentre n objetos. Assuma que cada elemento deste espaço amostral tem a mesma probabilidade, dada por $1/|S_{k,n}|$. Descreva um algoritmo eficiente para gerar amostras deste espaço. Dica: pense em permutação!

Questão 8: Estimando probabilidade de caudas
Considere o problema de calcular a probabilidade de cauda de uma determinada distribuição, $P[X \geq x]$.

1. Descreva um algoritmo de Monte Carlo simples para estimar $P[X \geq x]$ para um determinado $x > 0$ fixo.
2. Calcule a variância deste estimador.
3. Para x grande, $P[X \geq x]$ pode ser bem pequena. Descreva uma abordagem utilizando *importance sampling* para reduzir a variância do estimador acima.
4. Calcule a variância deste novo estimador.

5. Assumindo que X possui distribuição Normal padrão, implemente os dois algoritmos acima e calcule o valor do estimador para $x = 5$. Trace um gráfico do valor estimado por cada algoritmo em função de $n = 1, \ldots, 10^6$.