Estatística e Probabilidade - Projeto 02 - 2023/02

Prof. Hugo Carvalho

29/10/2022

– INSTRUÇÕES – LEIAM ATENTAMENTE! –

- A data limite de entrega do projeto é domingo 03/12/2023 às 23h59. Avaliações entregues após esse prazo serão desconsideradas.
- O projeto pode ser feito *até* em dupla.
- A entrega deve ser feita exclusivamente através do Google Classroom, e o projeto deve ser redigido na forma de um relatório incluindo os códigos. Para isso, use a linguagem Python dentro de um notebook do Google Colaboratory, a ser anexado na atividade da seguinte forma:
  - Dentro da turma no Google Classroom, clique em "Atividades"
  - No tema "Projetos", clique em "Projeto 02" e em seguida em "Ver atividade"
  - No canto superior direito, clique em "+ Adicionar ou criar" e em seguida selecione em "Arquivo" ou "Google Drive", se optar fazer upload do notebook ou selecionar um arquivo de seu Google Drive, respectivamente.
  - O título do notebook deverá ser, obrigatoriamente, "SEU NOME - Projeto 02" ou "NOMES DA DUPLA - Projeto 02", caso tenha optado por fazer o projeto em dupla. Nesse último caso, apenas uma das pessoas da dupla deve fazer a entrega.
  - Para entregar o projeto, abra a atividade que o contém (conforme procedimento explicado no ponto acima), e no canto superior direito clique em "ENTREGAR" (o texto estará em uma caixa cinza, mas ele é clicável).
  - Na tela a seguir clique em "Entregar" para confirmar a entrega do trabalho.
  - Se você se arrependeu e quiser fazer outra entrega, basta abrir a atividade, no canto superior direito clique em "Cancelar envio" e confirme clicando novamente em "Cancelar envio" na janela que irá abrir. Você pode fazer isso quantas vezes quiser até o prazo determinado.
  - As explicações que você precisará fazer devem ser feitas em células de texto em seu notebook. Somente comentários pequenos devem ser feitos no código.
  - Caso seja necessário, equações podem ser introduzidas em LaTeX nas células de texto.
  - Figuras também podem ser incluídas em células de texto, tanto através de upload quanto simplesmente arrastando-a para alguma célula de texto. Nesse caso, sempre informe a fonte de onde a figura foi retirada.
- A troca de conhecimento na realização da avaliação é permitida e encorajada: ciência se faz com colaboração, e devemos seguir esse espírito aqui. Porém, cada aluno (ou dupla) deverá ter a sua própria resolução, e cópias ou outras ilegalidades serão severamente punidas com a anulação da avaliação para o(s) aluno(s) suspeito(s).
- Todos os passos de sua resolução devem ser devidamente justificados.
- Ao entregar essa avaliação, você afirma ter lido e estar de acordo com essas regras, comprometendo-se a cumpri-las.

Questão 1: Disserte sobre a *distribuição de Zipf* (ou *lei de Zipf*, como também é chamada). Seu texto deve falar sobre: motivação e surgimento da distribuição, algumas de suas diversas aplicações, trazer discussões/interpretações de seus sumários numéricos (média, variância, etc.) e tratar de outros aspectos probabilísticos da distribuição que você julgue importante. Finalmente, diga como que a distribuição de Zipf se aplica em compressão de arquivos. Faça um texto o mais detalhado possível, evitando frases "misteriosas"; coloque referências bibliográficas para afirmações que você não irá se aprofundar.

Questão 2: O objetivo desta questão é ilustrar que certas probabilidades podem ser calculadas malandramente de modo levemente indireto.

a) Seja Z uma variável aleatória com distribuição normal padrão, ou seja, normal de média zero e variância unitária. Argumente que $\mathbb{P}(Z > 20)$ não é uma quantidade nula mas que é impossível obtê-la "ingenuamente", ou seja, através de algum dos seguintes métodos: consultando uma tabela, utilizando as funções elementares de algum *software* ou pacote (p. ex., `import scipy.stats` e em seguida `1 - scipy.stats.norm(0, 1).cdf(20)` no Python), ou através de simulação estocástica (p. ex., calculando a proporção de observações acima de 20 em uma grande quantidade de simulações de tal variável aleatória).

b) Escreva a integral que calcula a quantidade $\mathbb{P}(Z > 20)$, faça uma mudança de variáveis esperta e conclua que tal probabilidade pode ser expressada como $\mathbb{E}[g(Y)]$, onde Y é uma variável aleatória com distribuição uniforme no intervalo $[0, 1/20]$ e a função g é dada por

$$g(y) = \frac{1}{20y^2\sqrt{2\pi}} e^{-1/2y^2}.$$

c) Usando o resultado do item b) encontre uma aproximação para $\mathbb{P}(Z > 20)$ através de simulações da variável aleatória Y.

d) Ao se repetir o processo do item c) diversas vezes, você obterá várias estimativas para $\mathbb{P}(Z > 20)$. Estime a incerteza em tais estimativas, justificando o seu procedimento. Discuta o que acontece com tal incerteza à medida que a quantidade de simulações de Y aumenta.

Questão 3: Um inseto se move aleatoriamente em um tabuleiro de tamanho 5×5 de acordo com a seguinte regra: a cada movimento ele tem igual probabilidade de saltar para alguma das casas adjacentes àquela ocupada, onde **não** considera-se casas na diagonal como adjacentes; porém, na casa do canto superior esquerdo há uma armadilha que imobiliza o inseto; na casa do canto inferior direito há um portal que transporta o inseto para qualquer outra casa do tabuleiro, com igual probabilidade; finalmente, nas casas dos cantos superior direito e inferior esquerdo há portais que transportam o inseto entre essas casas somente. Com base nessa descrição, faça o que se pede abaixo.

a) Justifique que uma cadeia de Markov é um modelo adequado para descrever tal situação, e construa a matriz de probabilidades de transição de tal processo.

b) Construa um algoritmo para simular a posição do inseto, onde é possível selecionar a casa do tabuleiro onde o inseto localiza-se inicialmente.

c) Utilizando o algoritmo do item b), encontre uma aproximação para a probabilidade do inseto ser capturado pela armadilha do tabuleiro. Note que tal quantidade será função da casa inicialmente ocupada pelo inseto. Justifique intuitivamente o resultado obtido.

d) Também usando o algoritmo do item b), encontre uma aproximação para o número médio de saltos que o inseto dá antes de ser capturado pela armadilha. Analogamente ao item c), isso também será uma função da casa inicialmente ocupada no tabuleiro. Qual casa inicial maximiza o número médio de saltos do inseto antes de ser capturado?

e) Assumindo que a posição inicial do inseto é a casa central, quantas outras visitas ele faz, em média, à casa central, antes de ser capturado pela armadilha?