



06 de novembro de 2022

Lista 1: geração de números pseudo-aleatórios

Prof. Guilherme Rodrigues

Métodos computacionais intensivos para mineração de dados

1. As questões deverão ser respondidas em um único relatório *PDF* ou *html*, produzido usando as funcionalidades do *Rmarkdown* ou outra ferramenta equivalente.
2. O aluno poderá consultar materiais relevantes disponíveis na internet, tais como livros, *blogs* e artigos.
3. O trabalho é individual. Suspeitas de plágio e compartilhamento de soluções serão tratadas com rigor.
4. Os códigos *R* utilizados devem ser disponibilizados na íntegra, seja no corpo do texto ou como anexo.
5. O aluno deverá enviar o trabalho até a data especificada na plataforma Microsoft Teams.
6. O trabalho será avaliado considerando o nível de qualidade do relatório, o que inclui a precisão das respostas, a pertinência das soluções encontradas, a formatação adotada, dentre outros aspectos correlatos.
7. Escreva seu código com esmero, evitando operações redundantes, visando eficiência computacional, otimizando o uso de memória, comentando os resultados e usando as melhores práticas em programação.

Questão 1

Simulando computacionalmente o gerador de Babel.

Todo seu destino, a cura do câncer e até o que vai acontecer no fim do mundo. Todas essas respostas já estão escritas na Biblioteca de Babel. Essa biblioteca proposta por Jorge Luís Borges é composta por um número infinito de galerias, contendo todos os livros possíveis.

” [...] Um (livro) constava das letras M C V malevolamente repetidas da primeira linha até a última. Outro é um simples labirinto de letras mas a penúltima página diz ‘ó tempo tuas pirâmides’.”

A maior parte dos livros não tem qualquer significado. Entretanto, embora improváveis, certos textos resultam em grandes obras, como o Bhagavad Gita. Considerando as afirmações acima e a lista de palavras existentes na língua portuguesa (disponível no arquivo “Dicionario.txt”), responda aos itens a seguir.

a) Estime via simulação computacional (*Monte Carlo*) a probabilidade de se gerar uma palavra *válida* (isso é, do dicionário) ao sortear ao acaso sequências de 5 letras (todas com a mesma probabilidade). Em seguida, calcule analiticamente tal probabilidade e faça um gráfico indicando se a estimativa obtida se aproxima do valor teórico conforme a amostra aumenta. **Atenção:** utilize somente as letras do alfabeto sem caracteres especiais.

b) Estime a probabilidade da sequência gerada ser um palíndromo (ou seja, pode ser lida, indiferentemente, da esquerda para direita ou da direita para esquerda). Compare o resultado com a probabilidade exata, calculada analiticamente.

c) Construa um gerador que alterne entre consoantes e vogais (se uma letra for uma vogal, a próxima será uma consoante e vice-versa). Qual a probabilidade de gerar uma palavra válida com este novo gerador?

d) Considere um processo gerador de sequências de 5 caracteres no qual cada letra é sorteada com probabilidade proporcional à sua respectiva frequência na língua portuguesa (veja essa página). Suponha que esse processo gerou uma sequência com ao menos um “a”. Neste caso, estime a probabilidade dessa sequência ser uma palavra válida. **Dica:** Use a função `sample` e edite o parâmetro `prob`. **Para pensar:** Você consegue calcular essa probabilidade analiticamente? (Não precisa responder.)

Questão 2

Gerando números pseudo-aleatórios.

a) Escreva uma função que gere, a partir do método da transformada integral, uma amostra aleatória de tamanho n da distribuição Cauchy para n e γ arbitrários. A densidade da Cauchy(γ) é dada por

$$f(x) = \frac{1}{\pi\gamma(1 + (x/\gamma)^2)}.$$

Dica: Veja essa página.

b) Uma variável aleatória discreta X tem função massa de probabilidade

$$\begin{aligned} p(2) &= 0.2 \\ p(3) &= 0.1 \\ p(5) &= 0.2 \\ p(7) &= 0.2 \\ p(9) &= 0.3 \end{aligned}$$

Use o método de transformação inversa para gerar uma amostra aleatória de tamanho 1000 a partir da distribuição de X . Construa uma tabela de frequência relativa e compare as probabilidades empíricas com as teóricas. Repita usando a função `sample` do R.

c) Escreva uma função que gere amostras da distribuição Normal padrão ($\mu = 0, \sigma = 1$) usando o método de aceitação e rejeição adotando como função geradora de candidatos, $g(x)$, a distribuição Cauchy padrão (isso é, com $\gamma = 1$).