Universidade Federal de Juiz de Fora

Pós-Graduação em Modelagem Computacional

Introdução à Modelagem Matemática

Eduardo Santos de Oliveira Marques
Thiago Esterci Fernandes

Avaliação - Módulo IV

Juiz de Fora

2023

LISTA DE SÍMBOLOS

λ	Taxa de distribuição
μ	Média
σ	Desvio padrão
C	Valor crítico
E	Erro
I	Intervalo de confiança
N	Número total
s	Distância
t	Tempo
V	Variância
v	Velocidade

Questão 1. Existem diversos tipos de variáveis aleatórias, contaminando todos os tipos de sistemas em todos os experimentos. Sobre este tema, descreva:

(a) Uma Função Densidade de Probabilidade (*Probability Density Function*, PDF) representa qual método de identificação? Quais suas condições de existência? Responda com exemplos.

Resolução: [1, 2, 3]

Como o próprio nome diz, a PDF é um método utilizado para descrever a distribuição de probabilidade de variáveis contínuas, representando a probabilidade relativa de ocorrência de diferentes valores dentro de um intervalo específico. As suas condições de existência são:

1. O domínio da variável deve ser contínuo;
2. A função deve ser não negativa em todo o domínio;
3. A integral da função deve ser igual a 1 em todo o domínio.

Dois exemplos muitos famosos de PDF são a distribuição normal (ou gaussiana), e a distribuição exponencial. Suas equações são dadas respectivamente por:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad ; \quad f(x) = \lambda \cdot e^{-\lambda x}$$

onde:

- $f(x) =$ PDF;
- $x =$ Variável contínua;
- $\sigma =$ Desvio padrão;
- $\mu =$ Média da distribuição;
- $\lambda =$ Parâmetro de taxa de distribuição.

A distribuição normal é simétrica em torno da média, possuindo uma forma semelhante a um sino. A distribuição exponencial é usada, por exemplo, para modelar o tempo de espera entre eventos em um processo de Poisson.

(b) Como interpretar corretamente uma PDF? Responda com exemplos.

Resolução: [1, 2, 3]

A interpretação correta da PDF depende muito do contexto no qual está inserido, além da forma da distribuição (como a gaussiana e a exponencial, por exemplo). Tendo essas considerações, a interpretação envolve a compreensão da probabilidade relativa dos diferentes valores em uma distribuição contínua.

A PDF não fornece probabilidades diretas de eventos específicos, mas sim a **distribuição de probabilidade** de todo o evento. Para se obter a probabilidade de um evento específico, basta integrar a PDF no intervalo referido.

A sua análise correta permite obter informações sobre a sua forma, dispersão e tendência. Abaixo são mostradas interpretações para os dois exemplos dados na *letra a)*:

- Distribuição Normal:
  - A PDF é simétrica em torno da média;
  - O formato da curva da PDF indica a dispersão dos dados;
    - Uma curva mais "achatada" indica uma maior dispersão;
    - Uma curva mais "fina" indica uma menor dispersão.

Exemplo: Uma PDF com uma distribuição normal com média 0 e desvio padrão 1, indica que valores próximos de 0 possuem uma maior probabilidade de acontecer, enquanto valores mais distantes de 0 possuem menor probabilidade.

- Distribuição Exponencial:
  - A PDF possui uma taxa de decaimento exponencial;
  - A probabilidade de ocorrência diminui à medida que o valor aumenta.

Exemplo: Uma PDF com uma distribuição exponencial com taxa $\lambda = 0.5$, indica que valores menores possuem uma maior probabilidade de ocorrência.

Questão 2. Explique em detalhes o que são, quais os principais e a importância dos momentos estatísticos. Dê ênfase aos primeiro e segundo momentos e explique seus funcionamentos.

Resolução: [4, 5]

Os momentos estatísticos são medidas utilizadas para descrever diferentes características de uma distribuição de dados, fornecendo informações importantes sobre a sua forma, tendência, dispersão, assimetria, achatamento, entre outros. Os dois momentos mais utilizados são o primeiro e o segundo.

- Primeiro momento:
  - Média dos valores da distribuição;
  - Representa o centro de gravidade;
  - Fornece uma medida de tendência central;
  - Calculado somando todos os valores e dividindo pelo número total.

$$\mu_1 = \int_{-\infty}^{\infty} x f(x) dx$$

Através desse momento é possível resumir um conjunto de dados em um único valor de forma representativa, além de fornecer uma estimativa do valor central da distribuição. Ele é amplamente utilizado em análises estatísticas e no cálculo de outros indicadores, tais como o desvio padrão.

- Segundo momento (dispersão):
  - Representa a variância dos valores em uma dispersão;
  - Fornece o espalhamento dos dados em relação à média;
  - Calculado como a média dos quadrados das diferenças dos valores com a média.

$$\mu_2 = \int_{-\infty}^{\infty} (x - \mu_1)^2 f(x) dx$$

Esse momento fornece informações sobre a dispersão dos dados em torno da média; onde valores mais altos indicam uma dispersão maior, e valores menores indicam uma maior concentração dos dados em torno da média. A raiz quadrada da variância também é chamada de desvio padrão, sendo uma medida amplamente utilizada na estatística.

Além destes momentos, existem os de ordem superior, fornecendo informações adicionais, tais como assimetria, curtose, etc. Esses momentos são menos utilizados, porém são relevantes em análises mais avançadas.

Questão 3. Explique em detalhes o que são, quais os principais e qual a importância dos intervalos de confiança. Dê ênfase aos níveis de 1σ, 90% e 5σ.

Resolução: [6, 7]

Os intervalos de confiança são estimativas que fornecem uma faixa de valores dentro de um intervalo, onde se espera que um dado desconhecido de um domínio esteja contido através de um certo nível de confiança. Eles são utilizados para quantificar incertezas associadas a uma estimativa. Eles são calculados com base em uma amostra de dados dentro de uma população, onde seu objetivo é fornecer uma estimativa do possível intervalo onde o parâmetro populacional esteja contido. O nível de confiança determina a probabilidade de que o intervalo contenha o valor real do parâmetro. Os principais intervalos de confiança utilizados são:

- Nível 1σ (ou intervalo de confiança de 68.3%):
  - Aproximadamente uma vez o desvio padrão da média para ambos os lados;
  - Existe 68.3% de probabilidade que o valor real esteja contido neste intervalo.
- Nível de 90% (ou 1.64σ):
  - Aproximadamente 1.64 o desvio padrão da média para ambos os lados;
  - Existe 90.0% de probabilidade que o valor real esteja contido neste intervalo.
- Nível 5σ (ou intervalo de confiança de 99.99994%):
  - Aproximadamente cinco vezes o desvio padrão da média para ambos os lados;
  - Existe 99.99994% de probabilidade que o valor real esteja contido neste intervalo.

É importante ressaltar que todos os níveis de confiança seguem uma distribuição gaussiana onde sua forma é simétrica, onde maior o valor de σ, maior o seu intervalo de confiança. Eles fornecem uma medida quantitativa da incerteza em torno das estimativas amostrais, sendo possível observar a precisão e a variabilidade associadas aos resultados, auxiliando assim nas tomadas de decisões e no estabelecimento de limites estatísticos. A escolha do nível de confiança adequado depende do contexto e das necessidades da análise a ser requerida. Níveis de confiança mais altos fornecem intervalos mais esparsos, gerando uma maior incerteza; e mais baixos de confiança produzem intervalos mais estreitos, indicando uma maior precisão, porém com uma menor confiança do valor real.

Questão 4. Explique em detalhes o Teorema Central do Limite (TCL) e sua importância. Dê um exemplo de como utilizar o TCL para identificar os erros de um determinado sensor.

Resolução: [8, 9]

TCL é uma afirmação de que, sob certas condições, a média de uma amostra grande de observações tende a se aproximar de uma distribuição normal, independentemente da sua forma original, sejam elas observações independentes de uma variável aleatória. Essa afirmação permite realizar inferências estatísticas e estimativas se baseando em distribuições gaussianas (de forma aproximada), mesmo quando os dados não seguem essa distribuição. Ou seja, quando mais dados tiver, mais "gaussiana" será a distribuição.

Um exemplo de como utilizar o TCL pode ser apresentado em um sensor de temperatura. Supondo que o sensor esteja sujeito a erros aleatórios que podem variar os valores reais medidos, uma forma de avaliar a sua precisão seria coletar inúmeras amostras (ou leituras) da temperatura, calculando a média das observações.

Através do TCL, espera-se que a distribuição das médias das amostras siga uma distribuição aproximadamente normal, independendo das variações individuais das leituras. Com base nisso, é possível calcular um intervalo de confiança para a média calculada, considerando a variância e do tamanho dos dados.

Caso as leituras estejam se afastando muito do intervalo de confiança, isso pode indicar que o sensor esteja apresentando imprecisões nas suas medidas. O TCL fornece uma boa base para análise estatísticas, permitindo tomada de decisões em distribuições normais mesmo quando a distribuição dos dados não é conhecida.

Questão 5. Demonstre que $E(X+Y) = E(X) + E(Y)$, sabendo que X e Y são variáveis aleatórias contínuas e independentes, $E(\cdot)$ é o operador Esperança Estatística e $f(X, Y)$ é a função densidade de probabilidade conjunta (bidimensional) de ambas variáveis.

Resolução: [10, 11, 12]

Para essa demonstração, pode-se utilizar a definição de esperança e propriedades básicas de integração, tal como a propriedade de independência.

A esperança estatística de uma função de uma variável aleatória é definida como:

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x) dx \quad ; \quad \text{onde } f_X(x) \text{ é a PDF de } X$$

Calculando $E(X+Y)$ usando a definição acima:

$$E(X+Y) = \int_{-\infty}^{\infty} (x+y) f_{X+Y}(x+y) d(x+y)$$

Como X e Y foram definidas como variáveis independentes no enunciado, a função densidade de probabilidade (PDF) conjunta $f_{X+Y}(x+y)$ pode ser expressa como o produto das funções densidades de probabilidade de X e Y:

$$f_{X+Y}(x+y) = f_X(x) f_Y(y)$$

Substituindo essa expressão na integral, têm-se:

$$E(X+Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x+y) f_X(x) f_Y(y) dx dy$$

Reorganizando a equação e separando a integral em duas partes:

$$E(X+Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_X(x) f_Y(y) dx dy + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f_X(x) f_Y(y) dx dy$$

As integrais internas representam a esperança de X e Y, respectivamente. Com isso, é possível escrever:

$$E(X+Y) = \int_{-\infty}^{\infty} x f_X(x) dx \int_{-\infty}^{\infty} f_Y(y) dy + \int_{-\infty}^{\infty} y f_X(x) dx \int_{-\infty}^{\infty} f_Y(y) dy$$

Essas duas integrais são exatamente as esperanças de X e Y:

$$E(X+Y) = E(X) \int_{-\infty}^{\infty} f_Y(y) dy + E(Y) \int_{-\infty}^{\infty} f_X(x) dx$$

Como $f_X(x)$ e $f_Y(y)$ são funções densidades de probabilidade, suas integrais em relação a x e y são iguais a 1. Portanto:

$$E(X+Y) = E(X) \cdot 1 + E(Y) \cdot 1 = E(X) + E(Y)$$

Finalmente conclui-se que $E(X+Y) = E(X) + E(Y)$ para variáveis aleatórias contínuas e independentes X e Y.

Questão 6. A empresa *Waymo* está desenvolvendo mais um carro autônomo para navegação em grandes centros como táxi. Durante os testes de distância percorrida (x), os seguintes dados foram obtidos: média de 112 km, variância de 30.25 km^2, tempo médio de testes de $\Delta t = 1h52min$. Considerando a variável x como sendo Normal, faça:

(a) Obtenha a distância média percorrida pelo carro em km.

Resolução:

A distância média percorrida foi dada pelo próprio enunciado, sendo 112 km.

(b) Calcule o intervalo de confiança de 95% para a distância percorrida.

Resolução: [6, 7, 13]

Utilizando a média e a variância fornecidas no enunciado e considerando que a distância percorrida segue uma distribuição normal, o intervalo de confiança pode ser calculado através da seguinte fórmula:

$$I = \mu \pm E \quad ; \quad \text{onde:} \begin{cases} I = & \text{Intervalo de confiança;} \\ \mu = & \text{Média;} \\ E = & \text{Margem de erro.} \end{cases}$$

A margem de erro é calculada pelo desvio padrão da média, que este por sua vez é calculado dividindo o desvio padrão da população pelo tamanho da amostra, e o multiplicando pelo valor crítico associado ao nível de confiança desejado, sendo neste caso 95%. Para este nível de confiança, o valor crítico é aproximadamente 1.96.

Com a média e o desvio padrão dados, o desvio padrão da média pode ser calculado dividindo a variância pelo tamanho da amostra. Como o tamanho da amostra não foi fornecido no enunciado, o seu valor será dado como 30. Calculando o seu valor:

$$\sigma_\mu = \sqrt{\frac{V}{N}} = \sqrt{\frac{30.25 \text{ km}^2}{30}} \approx 1.7416 \text{ km} \quad ; \quad \text{onde:} \begin{cases} \sigma_\mu = & \text{Desvio padrão da média;} \\ V = & \text{Variância;} \\ N = & \text{Tamanho da amostra.} \end{cases}$$

Com isso, a margem de erro é calculada como:

$$E = \sigma_{mu} \cdot C = 1.7416 \text{ km} \times 1.96 \approx 3.4112 \text{ km} \quad ; \quad \text{onde: } C = \text{Valor crítico}$$

Finalmente, podemos calcular o intervalo de confiança:

$$I = \mu \pm E = 112 \text{ km} \pm 3.4112 \text{ km}$$

Portanto, o intervalo de confiança de 95% para a distância percorrida está aproximadamente entre 108.5888 km e 115.4112 km.

(c) Obtenha a velocidade média desenvolvida pelo carro em km/h.

Resolução: [14]

Utilizando a distância média percorrida e o tempo médio de testes, a velocidade média desenvolvida pode ser calculada através da seguinte fórmula:

$$v = \frac{\Delta s}{\Delta t} \quad ; \quad \text{onde:} \begin{cases} v = & \text{Velocidade média;} \\ \Delta s = & \text{Distância média percorrida;} \\ \Delta t = & \text{Tempo médio.} \end{cases}$$

Como o tempo também foi dado em minutos, é preciso fazer a conversão, sendo ela realizada abaixo.

$$\Delta t = 1 + \frac{52}{60} = 1.87\ h$$

Agora, calculando a velocidade média:

$$v = \frac{\Delta s}{\Delta t} = \frac{112 \text{ km}}{1.87\ h} \approx 59.89 \text{ km/h}$$

Por fim, a velocidade média desenvolvida pelo carro é por volta de 59.89 km/h.

(d) Calcule o intervalo de velocidades desenvolvidas em 95% de intervalo de confiança.

Resolução: [6, 7, 13]

Utilizando o intervalo de confiança da distância percorrida e o tempo médio de testes, o intervalo de velocidades desenvolvidas pode ser calculado através da faixa de valores entre os dois intervalos. Porém, o enunciado não forneceu a variância do tempo.

Seu valor será definido como a multiplicação do tempo médio e a razão entre a média da distância e a sua variância, sendo dada como:

$$V_t = \Delta t \cdot \left(\frac{V}{\Delta s}\right) = (1.87\ h) \cdot \left(\frac{30.25\ km^2}{112\ km}\right) \approx 0.5051\ h^2$$

OBS: Essa razão na realidade é uma regra de 3 para determinar a porcentagem da variância com relação a média, sendo 112 km dados como 100% e o parâmetro X como uma proporção de 30.25 km^2 como mostrado abaixo.

$$\begin{array}{rcc} 112\ km & - & 100\% \\ 30.25\ km^2 & - & X \end{array}$$

OBS: Mesmo sabendo que essa não é maneira correta de calcular a variância, essa foi a forma encontrada pela falta do dado no enunciado.

O desvio padrão do tempo médio é dado como:

$$\sigma_t = \sqrt{V_t} = \sqrt{0.5051\ h^2} \approx 0.7109\ h$$

Considerando que os intervalo de confiança da distância percorrida sejam 108.5888 km e 115.4112 km (calculados anteriormente) e que o intervalo de confiança do tempo médio sejam $1.87 - \sigma_t$ horas e $1.87 + \sigma_t$ horas, o intervalo de confiança da velocidade é obtido através da velocidade máxima e mínima, sendo suas fórmulas dadas por:

$$v_{min} = \frac{Is_{min}}{It_{max}} = \frac{108.5888\ \text{km}}{(1.87 + 0.7109)\ h} \approx 40.7886 km/h$$

$$v_{max} = \frac{Is_{max}}{It_{min}} = \frac{115.4112\ \text{km}}{(1.87 - 0.7109)\ h} \approx 70.1815 km/h$$

Portanto, o intervalo de confiança de 95% para a velocidade desenvolvida pelo carro é de aproximadamente 40.7886 km/h a 70.1815 km/h, considerando a variância do tempo como 0.5051.

(e) Faça o esboço da curva que representa os experimentos e marque os principais pontos da mesma (pontos de inflexão e tendência central) com seus respectivos valores nos eixos x e y.

Resolução:

Houve uma dificuldade na interpretação da pergunta, ela não será resolvida, mas abaixo são dispostas as dificuldades encontradas:

- Quais experimentos seriam? As médias e os intervalos de confiança?
- O objetivo é fazer uma PDF e considerar a distribuição gaussiana?
- O eixo x e y seriam a distância e o tempo, respectivamente?

Questão 7. Demonstre em detalhes o Teorema de Bayes: $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$.

Resolução: [15, 16]

A demonstração do Teorema de Bayes é obtida através das definições de probabilidade condicional e total. Como mostrado no enunciado, a fórmula desse teorema:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

onde:

- $P(A|B)$ = Probabilidade de o evento A ocorrer dado que o evento B ocorreu;
- $P(B|A)$ = Probabilidade de o evento B ocorrer dado que o evento A ocorreu;
- $P(A)$ = Probabilidade do evento A ocorrer;
- $P(B)$ = Probabilidade do evento B ocorrer.

A probabilidade condicional do evento A ocorrer dado que o evento B ocorreu é:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad ; \quad P(B|A) = \frac{P(A \cap B)}{P(A)}$$

A probabilidade total do evento B ocorrer pode ser escrita como:

$$P(B) = P(A \cap B) + P(\bar{A} \cap B)$$

onde $\bar{A}$ representa o evento complementar de A.

Substituindo essas definições na fórmula da probabilidade condicional, têm-se:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A \cap B)}{P(A \cap B) + P(\bar{A} \cap B)}$$

Reescrevendo o numerador como $P(A \cap B) = P(B|A)P(A)$, o denominador como $P(A \cap B) + P(\bar{A} \cap B) = P(B)$, aplicando a definição de probabilidade condicional $P(B|A)$ e a definição de probabilidade total, têm-se:

$$P(A|B) = \frac{P(B|A)P(A)}{P(A \cap B) + P(\bar{A} \cap B)} = \frac{P(B|A)P(A)}{P(A \cap B) + P(\bar{A} \cap B)}$$

Portanto, o Teorema de Bayes for demonstrado.

Questão 8. A Samsung possui duas fábricas no Brasil, uma em Manaus-AM e outra em Campinas-SP. A linha de celulares Galaxy é produzida em ambas para que a logística de distribuição no país seja facilitada. A fábrica de Manaus é responsável por 35% da produção enquanto o restante da linha Galaxy é feita em Campinas. Por questões técnicas, eventualmente existem produtos defeituosos na linha. A ocorrência de produtos defeituosos em Manaus é de 1.5% enquanto em Campinas é de 1%. Considere um determinado lote que possui produtos fabricados em ambas as fábricas. Responda:

(a) Qual a probabilidade de uma equipe de inspeção encontrar um produto defeituoso?

Resolução: [17, 18]

Para se obter a probabilidade da equipe de inspeção encontrar um produto defeituoso, é preciso considerar a probabilidade do produto ter sido fabricado em cada fábrica e a probabilidade do produto ser defeituoso em cada uma.

Denota-se por M o evento de um produto ser fabricado em Manaus, C o evento de um produto ser fabricado em Campinas, e D o evento de um produto ser defeituoso. A probabilidade de um produto ser fabricado em Manaus é 35%, e ser fabricado em Campinas é 65%. A probabilidade de um produto ser defeituoso em Manaus é 1.5%, e a probabilidade de ter defeito em Campinas é 1%. Equacionando essas informações:

$$P(M) = 0.35 \quad ; \quad P(C) = 0.65 \quad ; \quad P(D|A) = 0.015 \quad ; \quad P(D|B) = 0.01$$

Usando o Teorema da Probabilidade Total, é possível calcular probabilidade de um produto ser defeituoso, sendo:

$$P(D) = P(D|M) \cdot P(M) + P(D|C) \cdot P(C) = (0.015 \cdot 0.35) + (0.01 \cdot 0.65) = 0.01175$$

Portanto, a probabilidade de uma equipe de inspeção encontrar um produto defeituoso é de 1.175% ou 0.01175.

(b) Dado que a equipe de inspeção encontrou um produto defeituoso no lote, qual a probabilidade de que ele tenha vindo da fábrica de Campinas?

Resolução: [15, 16, 18]

Para calcular a probabilidade do produto defeituoso ter vindo da fábrica de Campinas, é possível aplicar o Teorema de Bayes.

Fazendo as mesmas considerações de letras e valores do exercício anterior (D para defeituoso e C para Campinas), têm-se através do teorema:

$$P(C|D) = (P(D|C) * P(C))/P(D) = \frac{(0.01 \cdot 0.65)}{0.01175} \approx 0.5532$$

Portanto, a probabilidade do produto ter vindo defeituoso dado que veio da fábrica de Campinas é aproximadamente 55.32% ou 0.5532.

Questão 9. Um investidor *day trade* do mercado de ações de curtíssimo prazo pode estar em quatro estados diferentes: **Rico, Classe média, Pobre e Devendo**. Todo dia ele tem chance de alterar seu estado com base em suas decisões. As mudanças podem ocorrer com as seguintes probabilidades:

- Se o investidor é Rico, poderá ser Classe média com chance de 65%, Pobre com chance de 20% e poderá estar Devendo com chance de 10%.
- Se o investidor é Classe média, poderá ser Rico com chance de 5%, Classe média com chance de 25%, poderá estar Devendo com chance de 40%.
- Se o investidor é Pobre, poderá ser Classe média com chance de 30%, Pobre com chance de 35%, poderá estar Devendo com chance de 25%.
- Se o investidor está Devendo, poderá ser Classe média com chance de 10%, Pobre com chance de 30%, e poderá estar Devendo com chance de 55%.

Modele este investidor de acordo com as Cadeias de Markov.

(a) Monte a máquina de estados estocástica do problema e a matriz de Markov (M).

Resolução: [19, 20, 21]

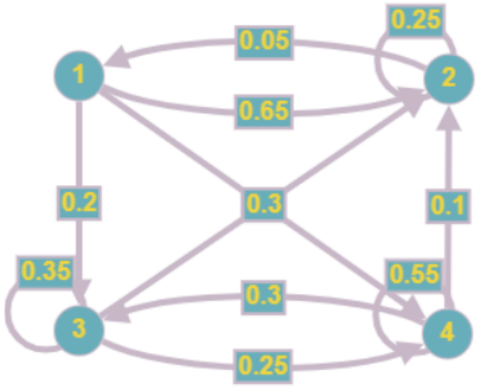
Modelando o investidor de acordo com a Cadeia de Markov, onde os possíveis estados são: Rico (R), Classe média (C), Pobre (P) e Devendo (D).

A matriz de transição de Markov (M) é uma matriz quadrada onde cada elemento representa a probabilidade de transição de um estado para outro. É possível montar essa matriz com base nas probabilidades de transição fornecidas no enunciado:

$$
M = \begin{array}{cl}
 & \quad R \qquad C \qquad P \qquad D \\
R & [[0.05, 0.65, 0.20, 0.10], \\
C & [0.05, 0.25, 0.30, 0.40], \\
P & [0.10, 0.30, 0.35, 0.25], \\
D & [0.05, 0.10, 0.30, 0.55]]
\end{array}
$$

As linhas da matriz representam os estados iniciais e as colunas os estados finais. Por exemplo, o elemento M[0][2] representa a probabilidade de transição do estado Rico para o estado Pobre, sendo seu valor igual a 0.30. A soma das probabilidades de cada linha deve ser igual a 1, pois o investidor deve sempre fazer uma transição para algum estado.

A máquina de estados estocástica pode ser visualizada como um grafo direcionado, onde os estados são representados por nós e as probabilidades de transição são representadas por arestas. O grafo foi feito através do site *Graph Online*.



Os números 1, 2, 3 e 4 representam os estados Rico, Classe média, Pobre e Devendo, respectivamente. Cada estado possui uma seta direcionada para outro estado com uma probabilidade associada. Por exemplo, a seta do estado Rico para o estado Classe média possui uma probabilidade de 0.65.

(b) Calcule todas as probabilidades deste *daytradder* 6 dias a partir de hoje. Qual é o estado com maior probabilidade ao final deste intervalo de tempo?

Resolução:

As probabilidades do *daytradder* variam de acordo com a sua condição inicial, podendo ser quatro casos: Rico, Classe média, Pobre e Devendo.

Este exercício exige simulações computacionais, sendo feitas em *Python* com o auxílio do *Online GDB*, utilizando a biblioteca *NumPy*.

Para iniciar as simulações, é preciso definir as condições iniciais, sendo elas os quatro estados possíveis. No código eles serão representados por vetores indicando o estado atual com o valor igual a um, sendo os demais iguais a zero. Além disso, é preciso definir a matriz de transição de Markov, sendo ela idêntica a matriz definida na *letra a*.

Para calcular as probabilidades após 6 dias, multiplica-se o vetores de probabilidades iniciais pela matriz de transição, repetindo esse processo 6 vezes. Para evitar repetições, foi feita uma função da probabilidade levando em consideração a matriz de Markov e as condições iniciais. Chamando a função para cada condição é possível obter os resultados, sendo o código mostrado abaixo.

```python
import numpy as np

# Função de probabilidade
def Prob(M, i):
    P_dias = np.linalg.matrix_power(M, 6) @ i
    prob_Rico = P_dias[0]
    prob_ClasseMedia = P_dias[1]
    prob_Pobre = P_dias[2]
    prob_Devendo = P_dias[3]

    print("- Probabilidade de estar no estado Rico após 6 dias:", prob_Rico)
    print("- Probabilidade de estar no estado Classe média após 6 dias:", prob_ClasseMedia)
    print("- Probabilidade de estar no estado Pobre após 6 dias:", prob_Pobre)
    print("- Probabilidade de estar no estado Devendo após 6 dias:", prob_Devendo)
    print()

# Matriz de transição de Markov
M = [[0.05, 0.65, 0.20, 0.10],
    [0.05, 0.25, 0.30, 0.40],
    [0.10, 0.30, 0.35, 0.25],
    [0.05, 0.10, 0.30, 0.55]]

# Condições inciais
i_R = np.array([1, 0, 0, 0])
i_C = np.array([0, 1, 0, 0])
i_P = np.array([0, 0, 1, 0])
i_D = np.array([0, 0, 0, 1])

# Chamada da função para cada caso inicial
print("Probabilidade inicial de estar no estado Rico")
P_iR = Prob(M, i_R)
print("Probabilidade inicial de estar no estado Classe Média")
P_iC = Prob(M, i_C)
print("Probabilidade inicial de estar no estado Pobre")
P_iP = Prob(M, i_P)
print("Probabilidade inicial de estar no estado Devendo")
P_iD = Prob(M, i_D)
```

Os resultados são mostrados abaixo.

```
Probabilidade inicial de estar no estado Rico
- Probabilidade de estar no estado Rico após 6 dias: 0.06544503125000001
- Probabilidade de estar no estado Classe média após 6 dias: 0.06544499999999999
- Probabilidade de estar no estado Pobre após 6 dias: 0.065445078125
- Probabilidade de estar no estado Devendo após 6 dias: 0.065445

Probabilidade inicial de estar no estado Classe Média
- Probabilidade de estar no estado Rico após 6 dias: 0.23268268750000004
- Probabilidade de estar no estado Classe média após 6 dias: 0.23267375000000004
- Probabilidade de estar no estado Pobre após 6 dias: 0.23269465625000002
- Probabilidade de estar no estado Devendo após 6 dias: 0.23266235937500002

Probabilidade inicial de estar no estado Pobre
- Probabilidade de estar no estado Rico após 6 dias: 0.30890034375
- Probabilidade de estar no estado Classe média após 6 dias: 0.3089005
- Probabilidade de estar no estado Pobre após 6 dias: 0.30890060937499997
- Probabilidade de estar no estado Devendo após 6 dias: 0.3089005

Probabilidade inicial de estar no estado Devendo
- Probabilidade de estar no estado Rico após 6 dias: 0.39297193750000003
- Probabilidade de estar no estado Classe média após 6 dias: 0.39298075000000005
- Probabilidade de estar no estado Pobre após 6 dias: 0.39295965625000007
- Probabilidade de estar no estado Devendo após 6 dias: 0.39299214062500004
```

Portanto é possível criar a seguinte tabela que define o estado com maior probabilidade dados diferentes condições iniciais ao final do intervalo de tempo de seis dias.

Condição inicial	**Maior probabilidade**	**Valor**
Rico	Pobre	0.65445070125
Classe média	Pobre	0.23269465625000002
Pobre	Pobre	0.30890060937499997
Devendo	Devendo	0.39299214062500004

Questão 10. Seja um sistema linear x unidimensional que é invariante no tempo (seus parâmetros são sempre os mesmos, mesmo com o tempo evoluindo), corrompido por ruído Gaussiano, dado por $x_k = Ax_{k-1} + \zeta k$, em que $\zeta\ N(0, Q)$. Os parâmetros do sistema são: $A = 1.1$ e variância $Q = \sigma_x^2 = 0.1$. Para obter o estado x deste sistema, tem-se um sensor y igualmente linear, invariante no tempo e corrompido por ruído Gaussiano: $y_k = Cx_k + \theta k$, em que $\theta\ N(0, R)$. Os parâmetros do sensor são: $C = 1$ e variância $R = \sigma_y^2 = 0.6$. Foram feitas 10 medidas consecutivas de sensor y, dadas a seguir:

k	1	2	3	4	5	6	7	8	9	10
y_k	2.60	2.51	3.08	4.52	5.07	4.44	5.92	6.39	6.64	7.26

Não há informações sobre o estado inicial do sistema x (no tempo $k = 1$), portanto, faça $\hat{x}1 = y_1$ e $P_1 = Q$.

Deseja-se obter estimativas melhores do estado do sistema do que as dadas pelo sensor. Utilize as equações recursivas do Filtro de Kalman linear unidimensional para calcular uma estimativa melhor que o sensor para o estado x do sistema. Mostre todos os cálculos para todos os instantes de tempo.

Resolução: [22, 23, 24]

Para se obter estimativas melhores do estado do sistema usando o Filtro de Kalman linear unidimensional, pode-se seguir as etapas de um algoritmo recursivo, envolvendo a previsão do estado e a atualização da estimativa com base na medida do sensor. Dado um sistema linear unidimensional com ruído Gaussiano, as equações recursivas do Filtro de Kalman para a previsão e atualização são:

Previsão:

$$\hat{x}_{\bar{k}} = A\hat{x}_{k-1} \quad ; \quad P_{\bar{k}} = AP_{k-1}A^T + Q$$

Atualização:

$$K_k = \frac{P_{\bar{k}}C^T}{CP_{\bar{k}}C^T + R} \quad ; \quad \hat{x}_k = \hat{x}_{\bar{k}} + K_k(y_k - C\hat{x}_{\bar{k}}) \quad ; \quad P_k = (I - K_kC)P_{\bar{k}}$$

Os valores iniciais são $\hat{x}_1 = y_1$ e $P_1 = Q$. A partir desses valores é possível calcular as estimativas do estado $\hat{x}_k$ e as matrizes de covariância P_k para cada instante de tempo, sendo feitos de forma iterativa. Abaixo são mostrados os cálculos para cada instante de tempo, utilizando as equações do Filtro de Kalman:

Para $k = 1$:

$$\hat{x}_1 = y_1 = 2.60$$
$$P_1 = Q = 0.1$$

Para $k = 2$:

$$\hat{x}_{\bar{2}} = A\hat{x}_1 = 1.1 \times 2.60 = 2.86$$
$$P_{(\bar{2})} = AP_1A^T + Q = 1.1 \times 0.1 \times 1.1 + 0.1 = 0.121$$
$$K_2 = \frac{P_{\bar{2}}C^T}{CP_{\bar{2}}C^T + R} = \frac{0.121 \times 1}{1 \times 0.121 \times 1 + 0.6} = 0.176$$
$$\hat{x}_2 = \hat{x}_{\bar{2}} + K_2(y_2 - C\hat{x}_{\bar{2}}) = 2.86 + 0.176(2.51 - 1 \times 2.86) = 2.75$$
$$P_2 = (1 - K_2C)P_{\bar{2}} = (1 - 0.176 \times 1) \times 0.121 = 0.1003$$

Para $k = 3$:

$$\hat{x}_{\bar{3}} = A\hat{x}_2 = 1.1 \times 2.75 = 3.025$$
$$P_{\bar{3}} = AP_2A^T + Q = 1.1 \times 0.1003 \times 1.1 + 0.1 = 0.1114$$
$$K_3 = \frac{P_{\bar{3}}C^T}{CP_{\bar{3}}C^T + R} = \frac{0.1114 \times 1}{1 \times 0.1114 \times 1 + 0.6} = 0.1671$$
$$\hat{x}_3 = \hat{x}_{\bar{3}} + K_3(y_3 - C\hat{x}_{\bar{3}}) = 3.025 + 0.1671(3.08 - 1 \times 3.025) = 3.029$$
$$P_3 = (1 - K_3C)P_{\bar{3}} = (1 - 0.1671 \times 1) \times 0.1114 = 0.0929$$

Para $k = 4$:

$$\hat{x}_{\bar{4}} = A\hat{x}_3 = 1.1 \times 3.029 = 3.332$$
$$P_{\bar{4}} = AP_{\bar{3}}A^T + Q = 1.1 \times 0.0929 \times 1.1 + 0.1 = 0.1139$$
$$K_4 = \frac{P_{\bar{4}}C^T}{CP_{\bar{4}}C^T + R} = \frac{0.1139 \times 1}{1 \times 0.1139 \times 1 + 0.6} = 0.1588$$
$$\hat{x}_{\bar{4}} = \hat{x}_{\bar{4}} + K_4(y_4 - C\hat{x}_{\bar{4}}) = 3.332 + 0.1588(4.52 - 1 \times 3.332) = 4.050$$
$$P_4 = (1 - K_4C)P_{\bar{4}} = (1 - 0.1588 \times 1) \times 0.1139 = 0.0966$$

Continuando com o mesmo procedimento, é possível calcular as estimativas e as matrizes de covariância para os instantes de tempo restantes, sendo seus valores:

$$\begin{aligned}
\text{Para } k = 5: &\quad \hat{x}_5 = 4.908, \quad P_5 = 0.0893 \\
\text{Para } k = 6: &\quad \hat{x}_6 = 4.931, \quad P_6 = 0.0862 \\
\text{Para } k = 7: &\quad \hat{x}_7 = 5.299, \quad P_7 = 0.0709 \\
\text{Para } k = 8: &\quad \hat{x}_8 = 5.527, \quad P_8 = 0.0597 \\
\text{Para } k = 9: &\quad \hat{x}_9 = 5.659, \quad P_9 = 0.0526 \\
\text{Para } k = 10: &\quad \hat{x}_{10} = 5.760, \quad P_{10} = 0.0476
\end{aligned}$$

REFERÊNCIAS

[1] PIRES, J. F. *Cálculo das Probabilidades e Estatística I*. 2023. Último acesso: 09/06/2023. Disponível em: <http://www.de.ufpb.br/~juliana/Calculo%20das%20Probabilidades%20e%20Estatistica%20I/Aula3.pdf>.

[2] KRIGAGEM, G. *Distribuição normal: O que é e sua grande importância na estatística*. 2021. Último acesso: 09/06/2023. Disponível em: <https://geokrigagem.com.br/distribuicao-normal-o-que-e-e-sua-grande-importancia-na-estatistica/>.

[3] BRAULINO, L. *DISTRIBUIÇÃO EXPONENCIAL (P2/P3)*. 2016. Último acesso: 09/06/2023. Disponível em: <http://resumoslegal.blogspot.com/2016/09/distribuicao-exponencial.html>.

[4] WIKIPEDIA. *Momento (estatística)*. 2023. Último acesso: 09/06/2023. Disponível em: <https://pt.wikipedia.org/wiki/Momento_(estat%C3%ADstica)>.

[5] UFPR. *Capítulo 2: Momentos amostrais e suas distribuições*. 2023. Último acesso: 09/06/2023. Disponível em: <http://leg.ufpr.br/~lucambio/CE050/20211S/Cap02.pdf>.

[6] LEG/UFPR. *5 Intervalos de Confiança*. 2023. Último acesso: 09/06/2023. Disponível em: <http://leg.ufpr.br/~paulojus/ce003/ce003/node5.html>.

[7] EDUCAçãO, I. a. *Química: Analítica 2. Tratamento e Avaliação Estatística de Dados (Capítulo 7, Skoog; Fundamentos de Química Analítica) - População, amostra, exatidão, precisão, média, desvio padrão, teste z, teste t, teste F, teste Q*. 2018. Último acesso: 09/06/2023. Disponível em: <http://incentivandoaeducacaonodf.blogspot.com/2018/09/quimica-analitica-2-tratamento-e.html>.

[8] SILVEIRA, F. L. da. *Teorema Central do Limite: exemplificações utilizando o Método de Monte Carlo*. 2023. Último acesso: 09/06/2023. Disponível em: <http://incentivandoaeducacaonodf.blogspot.com/2018/09/quimica-analitica-2-tratamento-e.html>.

[9] WIKIPEDIA. *Teorema central do limite*. 2022. Último acesso: 09/06/2023. Disponível em: <https://pt.wikipedia.org/wiki/Teorema_central_do_limite>.

[10] COSTA, P. A. *Propriedades da Integral Definida*. 2023. Último acesso: 09/06/2023. Disponível em: <https://www.professoraluizio.com.br/Blog/1340/propriedades-da-integral-definida/>.

[11] JúNIOR Éliton Fontana e F. C. M. *MATERIAL DE APOIO – Integrais*. 2023. Último acesso: 09/06/2023. Disponível em: <https://fontana.paginas.ufsc.br/files/2017/02/Material-de-Apoio-Integrais-Metodos.pdf>.

[12] ACADEMY, K. *Revisão das propriedades das integrais definidas*. 2023. Último acesso: 09/06/2023. Disponível em: <https://pt.khanacademy.org/math/ap-calculus-ab/ab-integration-new/ab-6-6/a/definite-integrals-properties-review>.

[13] MATéRIA, T. *Variância e desvio padrão*. 2023. Último acesso: 09/06/2023. Disponível em: <https://www.todamateria.com.br/variancia-e-desvio-padrao/#:~:

text=A%20vari%C3%A2ncia%20(V)%20%C3%A9%20%C3%BAtil,a%20raiz%20quadrada%20desse%20par%C3%A2metro.>.

[14] EDUCAçãO, M. *Velocidade média*. 2023. Último acesso: 09/06/2023. Disponível em: <https://mundoeducacao.uol.com.br/fisica/velocidade-media.htm>.

[15] WIKIPEDIA. *Teorema de Bayes*. 2023. Último acesso: 09/06/2023. Disponível em: <https://pt.wikipedia.org/wiki/Teorema_de_Bayes>.

[16] SUNO. *Teorema de Bayes: Entenda o que é e como calcular*. 2023. Último acesso: 09/06/2023. Disponível em: <https://www.suno.com.br/artigos/teorema-de-bayes/>.

[17] PRATES, M. O. *Probabilidade - aula III*. 2012. Último acesso: 09/06/2023. Disponível em: <https://www.est.ufmg.br/~marcosop/est031/aulas/Capitulo_2_3.pdf>.

[18] CEDERJ, U. *Aula 9 – Teorema da probabilidade total e teorema de Bayes*. 2023. Último acesso: 09/06/2023. Disponível em: <https://cesad.ufs.br/ORBI/public/uploadCatalago/10161710102012Probabilidade_e_Estatistica_aula_9.pdf>.

[19] PéREZ, F. L. *I.1 Cadeias de Markov*. 2021. Último acesso: 09/06/2023. Disponível em: <http://leg.ufpr.br/~lucambio/CM/CMI.html>.

[20] WIKIPEDIA. *Matriz de transição*. 2019. Último acesso: 09/06/2023. Disponível em: <https://pt.wikipedia.org/wiki/Matriz_de_transi%C3%A7%C3%A3o>.

[21] ONLINE, G. *Graph Online*. 2023. Último acesso: 09/06/2023. Disponível em: <https://graphonline.ru/pt/>.

[22] JUNIOR, J. R. F. *Filtro de Kalman*. 2019. Último acesso: 09/06/2023. Disponível em: <https://medium.com/@web2ajax/filtro-de-kalman-6e84f82993fc#>.

[23] KALMANFILTER.NET. *Tutorial sobre o Filtro de Kalman*. 2023. Último acesso: 09/06/2023. Disponível em: <https://www.kalmanfilter.net/PT/default_pt.aspx>.

[24] WIKIPEDIA. *Filtro de Kalman*. 2023. Último acesso: 09/06/2023. Disponível em: <https://pt.wikipedia.org/wiki/Filtro_de_Kalman>.