Universidade do Minho

2ºSemestre 2019/20

(MIEI, 3ºAno)

# Modelos Estocásticos de Investigação Operacional

# Trabalho Prático

***Identificação do Grupo***

| *Número:* | *Nome completo:* | *Rubrica:* |
|---|---|---|
| a76936 | Luís Miguel Arrepia Ferreira | *Luís Ferreira* |
| a84610 | Gonçalo Carneiro Almeida | *Gonçalo Almeida* |
| a85776 | José Emanuel Silva Rodrigues | *José Rodrigues* |
| a86788 | Lázaro Donato Martins Pinheiro | *Lázaro Pinheiro* |

*Data de entrega*: 2020-05-11

# Índice

## Conteúdo

# Parte I

## 1. Introdução

O presente trabalho prático desenvolve-se no âmbito da Unidade Curricular Modelos Estocásticos de Investigação, lecionada no 2º semestre do 3º ano do curso de Engenharia Informática. Este é dividido em duas partes.

Na primeira parte, é apresentado um problema em que se pretende maximizar o lucro de um empresário que gere duas filiais de um grupo internacional de automóveis. O principal objetivo deste, é descobrir a política ótima de transferência diária de automóveis entre as duas filiais.

Numa segunda parte do trabalho, era proposto o resumo de um artigo científico relacionado com a aplicação de Processos *Markovianos* e ou Programação Dinâmica Estocástica no estudo de problemas reais. Posto isto, procedeu-se à análise do artigo intitulado "*A generative Markov model for bowling scores*".

## 2. Formulação do Problema

### 2.1. Objetivo

Este problema tem como objetivo maximizar o lucro da gestão das filiais de automóveis do empresário tendo em conta diversos fatores como os custos associados aos espaços de estacionamento e transferências de veículos entre filiais, bem como o lucro associado a alugueres.

### 2.2. Estágios

Neste modelo cada estágio corresponde a um dia, começando no início do dia e tendo fim após as decisões de transferência de veículos entre filiais, ao final do mesmo dia.

### 2.3. Estados

Os diferentes estados possíveis deste sistema correspondem a todas as combinações de (a,b) -> (c,d) em que o par (a,b) corresponde ao número de automóveis no início do dia nas filiais 1 e 2, e o par (c,d) corresponde ao número de automóveis no final do dia nas filiais 1 e 2, respetivamente. Cada uma destas variáveis varia entre 0 e 12 veículos.

## 2.4. Decisões

Ao final de cada dia o empresário pode tomar uma de 7 decisões de modo a gerir os custos dos espaços de estacionamento e evitar as situações de “perda de vendas”:

Não transferir nenhum automóvel entre filiais;
Transferir 1 automóvel da filial 1 para a filial 2;
Transferir 2 automóvel da filial 1 para a filial 2;
Transferir 3 automóvel da filial 1 para a filial 2;
Transferir 1 automóvel da filial 2 para a filial 1;
Transferir 2 automóvel da filial 2 para a filial 1;
Transferir 3 automóvel da filial 2 para a filial 1.

# 3. Descrição do Modelo

Neste capítulo iremos explicar as decisões enumeradas anteriormente.
O empresário pode decidir não efetuar nenhuma transferência de veículos entre filiais caso tenha ambas as filiais com limite máximo de stock, caso tenha ambas as filiais sem nenhum automóvel em stock e caso os custos de transferência não compensem os custos dos espaços de estacionamento.
Quanto às restantes decisões, estas são efetuadas sempre que seja possível (a filial origem tem stock suficiente e a filial destino tem espaço livre suficiente para realizar a transferência) e conveniente (minimizar os custos).

# 4. Análise dos Resultados Obtidos

Após realizar a modelação do problema em causa numa folha de cálculo do Excel (Anexo 2) passamos ao cálculo das matrizes necessárias à resolução com um programa em C.

Como este é um problema de decisão com número indeterminado de estágios n (em que n tende para infinito), procuramos uma política ótima independente de n. Para isto, começamos por calcular as matrizes de transição e de custo para cada decisão. Para cada iteração realizamos o cálculo da expressão $Q_n^k + P_n^k * F_{n-1}$ e calculamos as diferenças entre $F_n$ e $F_{n-1}$, sendo que as matrizes de transição e custo não se alteram. As iterações acabam quando as diferenças dos $F_n$ se igualarem entre si tendo em conta uma taxa de 0.0008. Quando isto acontecer podemos analisar esse estágio para definir uma política ótima.

```c
double f0[169];
double fxA[169];

for(int i = 0 ; i < 169 ; i++){
    f0[i] = 0;
}
maioresElems(q0,q1,q2,q3,q4,q5,q6,fxA);

double maior, menor;
maiorMenor(&maior,&menor,fxA);
printf("%lf %lf\n", maior,menor);


for (int i = 0 ; i < 10000; i++){

    double fx[169];
    double aux0[169], aux1[169], aux2[169], aux3[169], aux4[169], aux5[169], aux6[169];
    double dx[169];
    multiplicaMatrizVetor(matrix,fxA,aux0);
    multiplicaMatrizVetor(matrix1,fxA,aux1);
    multiplicaMatrizVetor(matrix2,fxA,aux2);
    multiplicaMatrizVetor(matrix3,fxA,aux3);
    multiplicaMatrizVetor(matrix4,fxA,aux4);
    multiplicaMatrizVetor(matrix5,fxA,aux5);
    multiplicaMatrizVetor(matrix6,fxA,aux6);
    somarVetores(q0,aux0);
    somarVetores(q1,aux1);
    somarVetores(q2,aux2);
    somarVetores(q3,aux3);
    somarVetores(q4,aux4);
    somarVetores(q5,aux5);
    somarVetores(q6,aux6);

    maioresElems(aux0,aux1,aux2,aux3,aux4,aux5,aux6,fx);
    subtrairVetores(fxA,fx,dx);
    maiorMenor(&maior,&menor,dx);

    for(int j = 0; j < 169 ; j++){
        fxA[j] = fx[j];
    }
    printf("%lf %lf\n", maior,menor);
    if(maior-menor <= 0.0008){
        printf("\n\n%d\n",i+2);
        break;
    }
}
```

*Figura 1 - Código em C das iterações*

```
113.447207 113.443752
113.447193 113.443738
113.447179 113.443724
113.447166 113.443711
113.447152 113.443697
113.447138 113.443683
113.447125 113.443669
113.447111 113.443656
113.447097 113.443999
113.447083 113.446858


7879
```

*Figura 2 - Últimas 10 iterações consideradas*

## Parte II

O presente artigo procura criar um modelo de Markov através de dados que geram pontuações de bowling, sendo um dos desportos que melhor permite estudar e compreender o efeito “hot hand”.
VanDerwerken e Kenter (2018), procuraram uma nova solução para modelar a distribuição das pontuações no bowling, usando dados de torneios Professional Bowlers Association (PBA) entre 2003 e 2014.
Dada a natureza repetitiva deste desporto, os autores modelaram o desempenho de um jogador em cada frame, atendendo ao efeito de “hot- or cold-handedness” com um modelo de Markov de 4ª ordem. Com recurso a esta investigação, o modelo pode ser usado não apenas para classificar os jogadores, mas também para fazer previsões no jogo em tempo real, permitindo uma análise detalhada da capacidade do jogador.
Uma vez que o número de amostras da distribuição empírica das pontuações de um jogo, tende para infinito, os autores desenvolveram um modelo capaz de capturar os dados sem comprometer a sua precisão. Além disso, este modelo explica a dependência entre os frames por vários investigadores do efeito “hot hand”.
Apesar de muitas investigações desenvolvidas acerca do assunto em análise, tal como comprovado através da profunda revisão da literatura explanada no artigo, o presente estudo procura fundir os dois corpos da literatura de análise de dados do bowling (“hot hand” versus modelação explícita das pontuações) para criar um modelo generativo para pontuações do bowling que incorpora formalmente a natureza dos dados de Markov.
O modelo foi elaborado recorrendo à análise de um conjunto de treinamento de apenas 100 jogos por pessoa. Ou seja, para cada um dos 53 jogadores que jogaram pelo menos 200 jogos, selecionaram 100 jogos aleatoriamente (o conjunto de treinamento) com o qual construíram a matriz de greve e não greve para cada jogador. Seguidamente, simularam100.000 jogos para cada jogador e, seguidamente, simularam as pontuações dos quadros. Foi obtida (do conjunto de treinamento) a distribuição para a pontuação do jogador k em outros cinco modelos: i) modelo generativo de Markov de quarta ordem para greves (sem encolhimento bayesiano) e matriz não batida sem batida; (ii) modelo gerador de independência para greves (sem encolhimento bayesiano) e matriz não-grosseira de não greve; (iii) modelo gerador de independência para greves com mudanças de regime e matriz reduzida de não greve; (iv) distribuição empírica; e (v) distribuição discreta de reconhecedores.
No modelo elaborado pelos autores do estudo, para cada jogador, foi criada uma matriz de transição 8 × 8 que fornece a probabilidade de transição de um dos estados recolhidos para outro. Quando recebe um strike ou não no seu próximo lançamento, há apenas duas entradas diferentes de zero por linha, e elas necessariamente são iguais a 1. Portanto, toda a matriz de transição para um determinado jogador é definida por oito parâmetros. Foi utilizado um modelo hierárquico bayesiano para obter a distribuição posterior para a matriz de transição de cada jogador. Um jogador inicia num dos oito estados acima, extraídos da distribuição estacionária definida pela sua matriz de transição. Isso permite descobrir que as probabilidades de sucesso podem diferir entre os jogos: se um jogador começa em um estado "hot", a probabilidade de ataque observada tende a ser maior do que se ele começar a "cold".

Foi também assumido empiricamente que os componentes de tk eram independentes do número de jogos disputados pelo jogador k.
Foram analisadas as probabilidades empíricas de ataque, dando informação acerca dos pontos fortes relativos de diferentes jogadores quando se trata de sua capacidade de lançar ataques. Importa salientar que sem modelo hierárquico bayesiano, algumas probabilidades parecem irreais, pelo que o modelo bayesiano corrige esse problema automaticamente, resultando em estimativas muito mais confiáveis. Além disso, através do mesmo é possível dar respostas probabilísticas sobre quem é melhor em fazer uma greve numa determinada situação. Foi analisada também a comparação de matrizes de não-ataque dos jogadores, adiantando a probabilidade de que o jogador X seja melhor que o jogador Y ao lançar qualquer non-strike em particular ou qualquer união destes.
As principais discussões do estudo referem que o modelo generativo é muito superior à modelagem com a distribuição empírica, mas oferece insights sobre como os jogadores diferem e até onde um jogador pode focar melhor sua atenção para melhorar seu jogo. O modelo também tem relevância na classificação dos jogadores. Poder-se-ia simular muitos jogos do modelo proposto para um determinado jogador e calcular a média, que, devido ao encolhimento bayesiano, deveria ser mais realista do que uma média empírica calculada a partir de uma pequena amostra. Outra vantagem deste modelo é que pode ser usado para fazer previsões no jogo, algo que é pouco direto ao resumir a capacidade do jogador via distribuição empírica ou resumo de um número, como a média.

# Bibliografia

VanDerwerken D, Kenter F. A generative Markov model for bowling scores. Journal of Quantitative Analysis in Sports. 2018;14(4):213.
doi: https://doi.org/10.1515/jqas-2017-0081

# Anexo 1

```
Grupo que inclui o Aluno com o Nº 86788
MEIO-TP1 - Tabelas de probabilidades de pedidos e entregas de automóveis

FILIAL 1
Número de clientes:    ;      0 ;      1 ;      2 ;      3 ;      4 ;      5 ;      6 ;      7 ;      8 ;      9 ;     10 ;     11 ;     12
Probabilidade (pedidos):  ; 0.0468 ; 0.1572 ; 0.2220 ; 0.2340 ; 0.1656 ; 0.0956 ; 0.0452 ; 0.0192 ; 0.0104 ; 0.0032 ; 0.0008 ; 0.0000 ; 0.0000
Probabilidade (entregas): ; 0.0456 ; 0.1428 ; 0.2112 ; 0.2192 ; 0.1756 ; 0.1060 ; 0.0568 ; 0.0256 ; 0.0116 ; 0.0024 ; 0.0012 ; 0.0012 ; 0.0008

FILIAL 2
Número de clientes:    ;      0 ;      1 ;      2 ;      3 ;      4 ;      5 ;      6 ;      7 ;      8 ;      9 ;     10 ;     11 ;     12
Probabilidade (pedidos):  ; 0.1056 ; 0.2276 ; 0.2560 ; 0.2128 ; 0.1176 ; 0.0584 ; 0.0164 ; 0.0036 ; 0.0012 ; 0.0008 ; 0.0000 ; 0.0000 ; 0.0000
Probabilidade (entregas): ; 0.0380 ; 0.0764 ; 0.1012 ; 0.1284 ; 0.1376 ; 0.1348 ; 0.1048 ; 0.0804 ; 0.0692 ; 0.0584 ; 0.0412 ; 0.0224 ; 0.0072
```

*Figura 3 - Página com os dados fornecidos pelo docente (meio_tp1_g36.dat)*

# Anexo 2

Filial 1

| Numero de clientes | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | TOTAL |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Prob. Pedidos | 0,0468 | 0,1572 | 0,222 | 0,234 | 0,1656 | 0,0956 | 0,0452 | 0,0192 | 0,0104 | 0,0032 | 0,0008 | 0 | 0 | 1 |
| Prob. Entregas | 0,0456 | 0,1428 | 0,2112 | 0,2192 | 0,1756 | 0,106 | 0,0568 | 0,0256 | 0,0116 | 0,0024 | 0,0012 | 0,0012 | 0,0008 | 1 |

Filial 2

| Numero de clientes | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | TOTAL |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Prob. Pedidos | 0,1056 | 0,2276 | 0,256 | 0,2128 | 0,1176 | 0,0584 | 0,0164 | 0,0036 | 0,0012 | 0,0008 | 0 | 0 | 0 | 1 |
| Prob. Entregas | 0,038 | 0,0764 | 0,1012 | 0,1284 | 0,1376 | 0,1348 | 0,1048 | 0,0804 | 0,0692 | 0,0584 | 0,0412 | 0,0224 | 0,0072 | 1 |

Matriz de Probabilidades da Filial 1

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | TOTAL |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0,0456 | 0,1428 | 0,2112 | 0,2192 | 0,1756 | 0,106 | 0,0568 | 0,0256 | 0,0116 | 0,0024 | 0,0012 | 0,0012 | 0,0008 | 1 |
| 1 | 0,0434659 | 0,138251 | 0,2079989 | 0,2188256 | 0,1776405 | 0,1092573 | 0,0591026 | 0,0270602 | 0,0122552 | 0,0028306 | 0,0012562 | 0,0012 | 0,0008562 | 1 |
| 2 | 0,0362976 | 0,1208371 | 0,1926974 | 0,2143669 | 0,18412 | 0,1222389 | 0,0700941 | 0,0342674 | 0,0159162 | 0,004932 | 0,0018754 | 0,0012562 | 0,001101 | 1 |
| 3 | 0,0261744 | 0,0920904 | 0,1600987 | 0,1972894 | 0,1893405 | 0,1441696 | 0,0939981 | 0,0521853 | 0,0262314 | 0,0106354 | 0,0042432 | 0,0018754 | 0,0016683 | 1 |
| 4 | 0,015504 | 0,0592224 | 0,1153464 | 0,1628187 | 0,1824654 | 0,1656765 | 0,1274416 | 0,0833901 | 0,0474253 | 0,0231034 | 0,0102274 | 0,0042432 | 0,0031357 | 1 |
| 5 | 0,0079526 | 0,0324557 | 0,0711514 | 0,1167416 | 0,1552149 | 0,1703272 | 0,157096 | 0,1220003 | 0,0809485 | 0,0458208 | 0,0228941 | 0,0102274 | 0,0071696 | 1 |
| 6 | 0,0035933 | 0,015612 | 0,0378456 | 0,0717818 | 0,1133059 | 0,1497304 | 0,1664502 | 0,1546374 | 0,1208971 | 0,0802235 | 0,0457262 | 0,0228941 | 0,0173024 | 1 |
| 7 | 0,0015322 | 0,0068592 | 0,0179102 | 0,0381144 | 0,0703168 | 0,1109674 | 0,1480773 | 0,1654019 | 0,154167 | 0,120588 | 0,0801832 | 0,0457262 | 0,0401562 | 1 |
| 8 | 0,0006566 | 0,0029318 | 0,0078442 | 0,0180254 | 0,0374866 | 0,0693146 | 0,1102589 | 0,147628 | 0,1652003 | 0,1540346 | 0,1205707 | 0,0801832 | 0,0858651 | 1 |
| 9 | 0,0001824 | 0,0010454 | 0,0032054 | 0,0078762 | 0,017851 | 0,0372082 | 0,0691178 | 0,1101341 | 0,147572 | 0,1651635 | 0,1540298 | 0,1205707 | 0,1660435 | 1 |
| 10 | 3,648E-05 | 0,0002602 | 0,0011002 | 0,0032118 | 0,0078413 | 0,0177954 | 0,0371688 | 0,0690928 | 0,1101229 | 0,1475646 | 0,1651626 | 0,1540298 | 0,2866133 | 1 |
| 11 | 0 | 3,648E-05 | 0,0002602 | 0,0011002 | 0,0032118 | 0,0078413 | 0,0177954 | 0,0371688 | 0,0690928 | 0,1101229 | 0,1475646 | 0,1651626 | 0,440643 | 1 |
| 12 | 0,0468 | 0,1572 | 0,222 | 0,234 | 0,1656 | 0,0956 | 0,0452 | 0,0192 | 0,0104 | 0,0032 | 0,0008 | 0 | 0 | 1 |

Matriz de Probabilidades da Filial 2

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | TOTAL |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0,038 | 0,0764 | 0,1012 | 0,1284 | 0,1376 | 0,1348 | 0,1048 | 0,0804 | 0,0692 | 0,0584 | 0,0412 | 0,0224 | 0,0072 | 1 |
| 1 | 0,0339872 | 0,072345 | 0,0985811 | 0,1255277 | 0,1366285 | 0,1350957 | 0,107968 | 0,0829766 | 0,0703827 | 0,0595405 | 0,0430163 | 0,0243853 | 0,0095654 | 1 |
| 2 | 0,0253384 | 0,0595923 | 0,0888816 | 0,1167181 | 0,1316622 | 0,1347614 | 0,1150917 | 0,0916981 | 0,0755085 | 0,0631813 | 0,0480715 | 0,0304805 | 0,0190144 | 1 |
| 3 | 0,0156104 | 0,0411131 | 0,0697802 | 0,1000554 | 0,1204974 | 0,130512 | 0,1224374 | 0,1050682 | 0,0870971 | 0,0710718 | 0,0561155 | 0,0403485 | 0,040293 | 1 |
| 4 | 0,007524 | 0,0232136 | 0,0460235 | 0,0751658 | 0,101877 | 0,119943 | 0,124572 | 0,1176062 | 0,1028506 | 0,0849587 | 0,0676662 | 0,0523931 | 0,0762062 | 1 |
| 5 | 0,0030552 | 0,0106114 | 0,0252075 | 0,0482104 | 0,0759054 | 0,1016518 | 0,117531 | 0,1226102 | 0,1167058 | 0,1019822 | 0,0835758 | 0,0661547 | 0,1267984 | 1 |
| 6 | 0,000836 | 0,0039 | 0,011157 | 0,0258059 | 0,0484128 | 0,0758438 | 0,1009918 | 0,1169942 | 0,1223638 | 0,1164682 | 0,1016038 | 0,0831622 | 0,1924603 | 1 |
| 7 | 0,0002128 | 0,001051 | 0,0040389 | 0,0113093 | 0,0258574 | 0,0483971 | 0,0756758 | 0,1008552 | 0,1169315 | 0,1223034 | 0,1163718 | 0,1014986 | 0,2754971 | 1 |
| 8 | 0,000076 | 0,0002896 | 0,0011006 | 0,0040933 | 0,0113277 | 0,0258518 | 0,0483371 | 0,075627 | 0,1008328 | 0,1169099 | 0,122269 | 0,1163342 | 0,3769509 | 1 |
| 9 | 0,0000304 | 0,0001067 | 0,0003094 | 0,0011224 | 0,0041006 | 0,0113254 | 0,0258278 | 0,0483176 | 0,0756181 | 0,1008242 | 0,1168962 | 0,1222539 | 0,4932672 | 1 |
| 10 | 0 | 0,0000304 | 0,0001067 | 0,0003094 | 0,0011224 | 0,0041006 | 0,0113254 | 0,0258278 | 0,0483176 | 0,0756181 | 0,1008242 | 0,1168962 | 0,6155211 | 1 |
| 11 | 0 | 0 | 0,0000304 | 0,0001067 | 0,0003094 | 0,0011224 | 0,0041006 | 0,0113254 | 0,0258278 | 0,0483176 | 0,0756181 | 0,1008242 | 0,7324173 | 1 |
| 12 | 0,1056 | 0,2276 | 0,256 | 0,2128 | 0,1176 | 0,0584 | 0,0164 | 0,0036 | 0,0012 | 0,0008 | 0 | 0 | 0 | 1 |

*Figura 4 - Matrizes de transição (Probabilidades) relativas a cada uma das filiais*