

UNIVERSIDADE FEDERAL DO PARANÁ

BRUNO HENRIQUE ABREU

FABIO HENRIQUE SCHROEDER

**Aplicação de Cadeias de Markov para determinação de probabilidades de
estados climáticos na cidade de Curitiba**

CURITIBA

2017

SUMÁRIO

1. RESUMO	03
2. INTRODUÇÃO	04
3. VARIAÇÃO CLIMÁTICA	05
3.1 MODELOS DE PREVISÃO NÚMERICA	05
4. PROCESSOS ESTOCÁSTICOS	06
4.1 CADEIAS DE MARKOV	06
5. ESTUDO DE CASO	08
5.1 METODOLOGIA.....	08
5.2 RESULTADOS.....	09
6. CONCLUSÃO	14
7. REFERÊNCIAS.....	15
8. ANEXOS	16

1. RESUMO

Prever o tempo tem se tornado fundamental não apenas na relação do ser humano com o meio ambiente, mas também para o desenvolvimento de diversas atividades econômicas. Hoje em dia, utilizam-se modelos de previsão numérica que tem como base a massiva coleta de dados e uso de alta tecnologia para acompanhamento do estado da atmosfera terrestre em tempo real. Sendo assim, o trabalho sugere um modelamento a partir do processo estocástico de Cadeias de Markov para determinação de probabilidades de ocorrência de estados climáticos na cidade de Curitiba para o mês de janeiro. O objetivo é saber se eu um dado dia de janeiro se choveu ou não choveu, qual a probabilidade de chover ou não chover no dia seguinte. Para verificar se esse tipo de modelagem é eficiente, vamos comparar com dados de janeiro de 2017.

2. INTRODUÇÃO

Prever o tempo, é de fundamental importância, não apenas para atividades econômicas, mas também para a sobrevivência dos seres humanos. Prever a quantidade de chuva que cairá nas áreas serranas para evitar deslizamento, o quanto de chuva que cairá nas represas de abastecimento, para evitar o racionamento de água, e vários outros exemplos práticos.

Esta ciência vem sendo aprimorada por meio do uso de alta tecnologia para permitir melhor e mais precisa análise das variáveis do clima. Atualmente, a previsão é feita a partir da análise de dados captados em todo o mundo por uma rede internacional. As previsões meteorológicas são realizadas a partir de uma massiva coleta de dados sobre o estado atual da atmosfera terrestre, e por meio da compreensão científica dos processos atmosféricos para projetar como o clima irá evoluir.

Para tornar possível a análise da pressão atmosférica e as causas de suas variações foram criados modelos meteorológicos capazes de acompanhar o movimento das massas de ar com diferentes pressões atmosféricas e suas relações.

No entanto, com a atual tecnologia não é possível prever todos os desdobramentos da atmosfera. Isto se deve ao fato de que esta apresenta um comportamento caótico, ou seja, um pequeno fator, que pode ser menor do que a margem de erro dos dados numéricos, pode desencadear em eventos imprevisíveis.

Sendo assim, esse trabalho tem o objetivo de modelar, via cadeias de Markov, a probabilidade de chover ou não chover em um determinado dia de janeiro na cidade de Curitiba. A partir da análise dos dados históricos e identificação das probabilidades de ocorrência dos estados, o desenvolvimento de um modelo que apresente de maneira prática essa transição de um dia para o outro, realizando uma validação no mês de janeiro de 2017.

3. VARIAÇÃO CLIMÁTICA

O clima no mundo está em constante mudança, e a ação humana está interferindo de modo que essa variação seja mais intensa. Pesquisas comprovam que de fato a Terra está aquecendo e as consequências são observadas no derretimento das calotas polares, que está ocasionando o aumento do nível dos mares.

Outro impacto é o fenômeno El niño, que se caracteriza por um aquecimento das águas do oceano pacífico. Esse aquecimento, decorrente da poluição, acarreta na mudança dos padrões climáticos, afetando os regimes de chuva em regiões de latitudes médias.

3.1 MODELOS DE PREVISÃO NUMÉRICA

Em 1950 concretizou-se a primeira previsão numérica de tempo bem-sucedida no computador ENIAC, localizado na U.S. Army Aberdeen Proving Ground, em Maryland. Segundo Moura (1996), este trabalho é considerado como marco da meteorologia moderna. Hoje em dia, a previsão numérica é o tipo de previsão do tempo com cujo formato está-se mais familiarizado onde as observações meteorológicas servem como entrada para os modelos numéricos de computador por meio de um processo conhecido como assimilação de dados para produzir saídas de temperatura, precipitação, e centenas de outros elementos meteorológicos dos oceanos para o topo da atmosfera.

Logo abaixo, segue alguns exemplos de sistemas de previsão:

a) Global Data Assimilation System (GDAS): apresenta, para um modelo em 3D, observações de superfície, os dados de balão, dados dos ventos, relatórios de aeronaves, observações das boias, observações de radar e as observações de satélite.

b) Global Ensemble Forecast System (GEFS): modelo de previsão do tempo composta de 21 previsões separadas. Este conjunto de previsões é gerado perturbando-se minimamente e de formas diferentes as observações originais, isto porque diferenças imperceptíveis entre a realidade e o que é realmente medido podem, ao longo do tempo, conduzir a diferenças notáveis entre o que é um modelo de previsão do tempo prevê e própria realidade.

c) North American Mesoscale (NAM): é um dos principais modelos de tempo, executado pelos centros nacionais de previsão ambiental para a produção de previsões meteorológicas. Gera múltiplas redes de previsões de tempo sobre o continente norte-americano em várias resoluções horizontais. Previsões de alta resolução são geradas dentro do NAM com modelos meteorológicos numéricos adicionais. Estas janelas de previsão de alta resolução são geradas em regiões fixas e são ocasionalmente utilizadas para acompanhar os acontecimentos meteorológicos significativos.

4. PROCESSOS ESTOCÁSTICOS

Processo estocástico é uma coleção de variáveis aleatórias (sobre um espaço de probabilidade (Ω, F, P)), indexados por um conjunto de índices T . O conjunto T é chamado de espaço paramétrico. Os valores assumidos por $X(t)$ são chamados de estados, e o conjunto de todos os possíveis estados é chamado de espaço de estados do processo estocástico.

4.1 CADEIAS DE MARKOV

Um processo estocástico ser considerado uma Cadeia de Markov, seu comportamento dinâmico deve respeitar uma característica, chamada propriedade markoviana. Essa característica demonstra a probabilidade condicional que em um evento futuro, dado qualquer evento passado, no estado presente $X_t = i$, depende somente do estado presente do processo, não importando como o processo chegou a tal estado.

Para ser considerado uma Cadeia de Markov, um processo estocástico $\{X_t, t = 0, 1, 2, \dots\}$ com espaço de estado $S = \{1, 2, \dots, s\}$ deve satisfazer a propriedade de Markov, para todo n pertencente aos números naturais e para todo i pertencente ao conjunto S :

$$P(X_n = i_n | X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = P(X_n = i_n | X_{n-1} = i_{n-1})$$

Para a cadeia ser considerada homogênea ou estacionária no tempo, a probabilidade de ir de um estado a outro deve independe do tempo em que o passo é dado. Ou seja, para quaisquer estados $i, j \in S$, temos:

$$P(X_n = j | X_{n-1} = i) = P(X_{n+k} = j | X_{n+k-1} = i), \text{ para } k = -(n-1), -(n-2), \dots, -1, 0, 1, 2, \dots$$

Num determinado conjunto de estados, o processo começa em um desses estados e move-se sucessivamente de um estado para outro. Se a cadeia está atualmente no estado s_i , então ela se move para o estado s_j no próximo passo com uma probabilidade denotada por p_{ij} , e essa probabilidade não depende dos estados ocorridos nos passos anteriores, apenas do estado atual. Este termo é chamado de probabilidade de transição.

Realizar uma matriz é mais conveniente de resumir as probabilidades de transição. Assim, a matriz P define a cadeia de Markov, sendo todas as probabilidades de transição fixas e independentes ao longo do tempo. Para determinar a matriz, eleva-se a matriz P na n -ésima potência. A matriz P , demonstrada abaixo, expressa as probabilidades de transição de uma Cadeia de Markov

$$P = \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1j} & \dots \\ p_{21} & p_{22} & \dots & p_{2j} & \dots \\ \vdots & \vdots & \ddots & \vdots & \ddots \\ p_{i1} & p_{i2} & \dots & p_{ij} & \dots \\ \vdots & \vdots & \ddots & \vdots & \ddots \end{pmatrix}$$

Sendo X_n , com $n \geq 0$ uma Cadeia de Markov com espaço de estados S e função de transição P . Seja um vetor de números $\pi(x)$, com $x \in S$, não negativo e cujo a soma resulte em 1 e se $\sum \pi(x)P(x,y) = \pi(y)$, $y \in S$, então π será chamado de distribuição estacionária.

Supondo que essa distribuição existe e que $\lim_{n \rightarrow \infty} P^n = \pi(y)$, $y \in S$. Então, independentemente da distribuição inicial da cadeia, a distribuição de X_n irá se aproximar de π conforme n vai para infinito.

A condição de distribuição estacionária pode ser escrita na forma matricial como $\pi \cdot P = \pi$.

5. ESTUDO DE CASO

5.1 METODOLOGIA

A metodologia aplicada para desenvolvimento do trabalho foi dividida em quatro etapas:

Etapa 1 – Coleta de dados: os dados utilizados para este trabalho foram retirados de uma base de informações coletadas sobre o clima da cidade de Curitiba, no período de 2002 a 2017, apenas dos dias de janeiro, fornecida pelo INPE (Instituto Nacional de Pesquisas Espaciais);

Etapa 2 – Identificação das probabilidades: realizou-se a separação dos dados em faixas de precipitação, ou seja, até 5 milímetros de precipitação, marcou-se que não choveu no dia, caso contrário, choveu. Sendo assim, realizou-se a contagem do número de transições ocorridas entre as faixas e desta forma, obteve-se a frequência absoluta das transições entre as faixas no intervalo de tempo considerado, e com estes dados construiu-se a tabela de transição;

Etapa 3 – Montagem da cadeia de Markov: Após a montagem da tabela de transição foi obtida a matriz de transição probabilística. As entradas dessa matriz contêm as probabilidades associadas à transição entre estado.

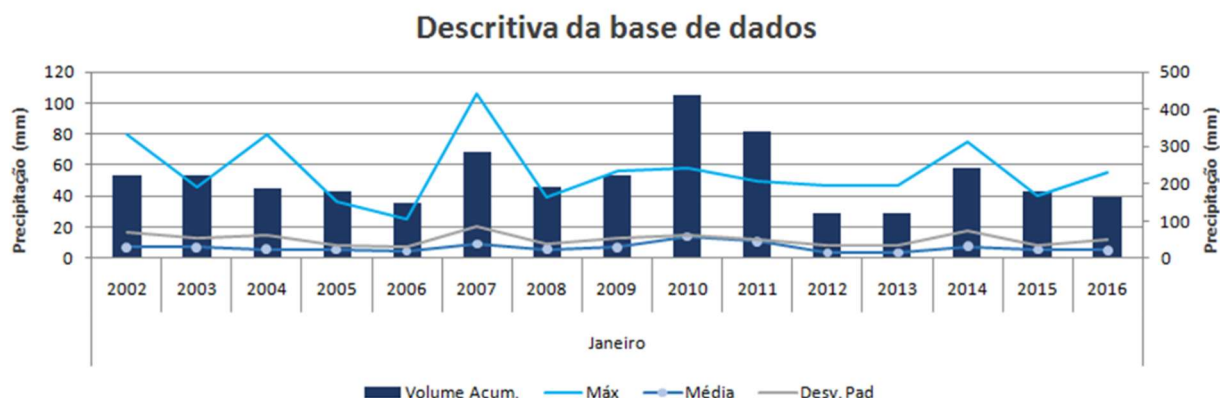
Etapa 4 – Validação do método: Nessa etapa, buscou-se validar a previsão gerada por meio da Cadeia de Markov, com os dados retirados de janeiro de 2017.

5.2 RESULTADOS

Antes da descrição das etapas realizadas, foi realizado uma análise descritiva dos dados, afim de verificar como os dados estão distribuídos.

No gráfico abaixo, é possível verificar que houve um aumento no acumulado de precipitação nos meses de janeiro dos anos de 2010 e 2011, isso se deve pelo aumento da temperatura registrado na época.

Houve também um pico na quantidade máxima de precipitação, observado em 2007, isso se dá, pois no dia 20/01/2007, foi o 5º dia com maior precipitação registrado em 24 horas desde 1961, quando começou a se registrar os dados.



Na etapa um, foram utilizados dos dados fornecidos pelo INPE, para o período de 2002 a 2016 (apenas o mês de janeiro), foram totalizados 465 dias. Na etapa dois, realizou-se a separação dos dados em duas categorias (choveu e não choveu), com base na precipitação, acima de 5 milímetros choveu, caso contrário, não choveu.

Após, foi realizada a contagem do número de transições das faixas correspondentes (choveu e não choveu). Desta forma, obteve-se a frequência absoluta das transições entre as faixas no intervalo de tempo considerado, e com estes dados construiu-se a tabela de transição, conforme abaixo:

		D + 1	
		Choveu	Não Choveu
D	Choveu	p	p
	Não Choveu	p	p

Na etapa três foi realizada a montagem da Cadeia de Markov. Foram calculadas as probabilidades de transição de uma faixa inicial para cada uma das possibilidades de faixas finais.

Para realizar a matriz de transição, foi utilizado o recurso da tabela dinâmica do excel.

A matriz de transição resultante do dia 1º é apresentada logo abaixo:

		2	
		Choveu	Não Choveu
1	Choveu	0	1
	Não Choveu	0,5	0,5

Para os demais dias, as matrizes de transição, encontram-se no anexo 1.

Para a realização da etapa quatro, foi retirado o mês de janeiro de 2017 para a análise comparativa entre a transição ocorrida e a probabilidade da mesma segundo a matriz resultante das etapas anteriores.

Nas tabelas abaixo, é possível verificar tal comparação, tomando como base o exemplo da transição do dia 19 para o dia 20:

		20	
		Choveu	Não Choveu
19	Choveu	0,75	0,25
	Não Choveu	0,36	0,64

		Realizada em 2017		Probabilidade de transição	
		19-Jan	20-Jan	Choveu	Não Choveu
Precipitação	Não Choveu	Não Choveu	Não Choveu	0,36	0,64

A partir da matriz de transição, observamos a probabilidade de ocorrência do dia 19 de janeiro (não choveu) para o dia 20 de janeiro (não choveu) é de 0,64.

Com base nesse exemplo, foi possível iterar a matriz de transição, para verificar a probabilidade de chover ou não chover para os próximos dias, como descrito na tabela abaixo:

	Realizada em 2017	Probabilidade de transição	
		Choveu	Não Choveu
21-jan	Não Choveu	0,36	0,64
22-jan	Não Choveu	0,51	0,50
23-jan	Não Choveu	0,56	0,44
24-jan	Não Choveu	0,58	0,42
25-jan	Não Choveu	0,59	0,41
26-jan	Não Choveu	0,59	0,41
27-jan	Não Choveu	0,59	0,41
28-jan	Não Choveu	0,59	0,41
29-jan	Não Choveu	0,59	0,41
30-jan	Não Choveu	0,59	0,41
31-jan	Não Choveu	0,59	0,41
Estacionaria	Não Choveu	0,59	0,41

É possível verificar que a partir da sexta iteração (dia 25 de janeiro), as probabilidades de transição, permanecem inalteradas, demonstrando que existe uma distribuição estacionária.

Para encontrar a distribuição estacionaria, utilizou-se o software R, com o pacote expm, e a função `%^%`, conforme abaixo:

```
>library(expm)
>excel_sheets("chuva.xlsx")
>chuva <- read_excel("chuva.xlsx",
  col_names = FALSE,
  skip = 0)
>class(chuva)
>chuva2<- as.matrix(chuva)
>class(chuva2)
>distr_esta <- chuva2%^%1000
```

	x0	x1
[1,]	0.5925926	0.4074074
[2,]	0.5925926	0.4074074

Para as demais matrizes de transição, também verificou se que a partir da sexta iteração, existe a distribuição estacionária.

Sendo assim, a distribuição estacionária, para esse exemplo é:

Distribuição Estacionaria		
	Choveu	Não Choveu
Choveu	0,59	0,41
Não Choveu	0,59	0,41

Seguindo com a validação dos dados, a tabela abaixo, demonstra a probabilidade de acerto de todas as transições ocorridas e os dados retirados no INPE do mês de janeiro de 2017:

	Realizada em 2017	Probabilidade	
		Acerto	> 0,8
1-jan	Não Choveu	0,50	< 0,80
2-jan	Choveu	0,80	=> 0,80
3-jan	Não Choveu	0,33	< 0,80
4-jan	Choveu	0,50	< 0,80
5-jan	Não Choveu	0,88	=> 0,80
6-jan	Não Choveu	0,88	=> 0,80
7-jan	Não Choveu	0,77	< 0,80
8-jan	Não Choveu	0,46	< 0,80
9-jan	Choveu	0,50	< 0,80
10-jan	Não Choveu	0,50	< 0,80
11-jan	Não Choveu	0,44	< 0,80
12-jan	Não Choveu	0,12	< 0,80
13-jan	Choveu	0,60	< 0,80
14-jan	Choveu	0,42	< 0,80
15-jan	Não Choveu	0,22	< 0,80
16-jan	Choveu	1,00	=> 0,80
17-jan	Choveu	1,00	=> 0,80
18-jan	Não Choveu	0,72	< 0,80
19-jan	Não Choveu	0,63	< 0,80
20-jan	Não Choveu	0,75	< 0,80
21-jan	Não Choveu	0,80	=> 0,80
22-jan	Não Choveu	0,81	=> 0,80
23-jan	Não Choveu	0,81	=> 0,80
24-jan	Não Choveu	0,63	< 0,80
25-jan	Não Choveu	0,70	< 0,80
26-jan	Não Choveu	0,63	< 0,80
27-jan	Não Choveu	0,88	=> 0,80
28-jan	Não Choveu	0,72	< 0,80
29-jan	Não Choveu	0,81	=> 0,80
30-jan	Não Choveu	0,90	=> 0,80

Com base na tabela, para melhor visualização dos acertos, foi realizado um

ponto de corte, onde abaixo de 0,80, seria classificado previsão errada, caso contrário, como previsão correta.

Realizando a contagem dessa classificação, verifica-se que houve 37 % de acertos e 63 % de erros.

6. CONCLUSÃO

A partir de algumas simulações de previsão do tempo, através da utilização das matrizes de transição em diferentes momentos, observou-se que as matrizes não possuem boa assertividade para as previsões de $D \rightarrow D+1$, piorando quando utilizamos essas matrizes para fazer previsões mais longas.

Os erros associados a essas previsões devem-se à fato da atmosfera apresentar um comportamento caótico, sendo influenciada por diversas variáveis.

Para uma maior assertividade devem ser considerados demais variáveis que possuem influência (Ex: Temperatura, Umidade etc.), com métodos capazes de comportar esse conjunto de variáveis explicativas, como modelos de regressão entre outras, sendo assim, com apenas uma variável (precipitação) fica inviável a previsão de estados climáticos.

7. REFERÊNCIAS:

AGUILERA, B. O. **Processos Estocásticos Aplicados** – Anotações em Aula. Universidade Federal do Paraná.

HOEL, PORT & STONE. **Introduction to Stochastic Processes**. Houghton Mifflin Co.

XXXVI ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO **Contribuições da Engenharia de Produção para Melhores Práticas de Gestão e Modernização do Brasil** João Pessoa/PB, Brasil, de 03 a 06 de outubro de 2016.

INPE. Disponível em: <http://enos.cptec.inpe.br/> Acessado em 25 de junho de 2017.

ALLEN, Linda J. S., **An Introduction to Stochastic Processes with Applications to Biology**, 2nd Edition, Chapman and Hall, 2010.

MOURA, Antônio Divino. **Von Neumann e a previsão numérica de tempo e clima**. Revista Estudos Avançados, vol.10 no.26 São Paulo Jan./Apr. 1996. Disponível em: <http://www.scielo.br/scielo.php?pid=S0103-0141996000100021&script=sci_arttext>. Acesso em: 3 junho 2015.

8. ANEXO 1

Matrizes de transição

		2	
		Choveu	Não Choveu
1	Choveu	0	1
	Não Choveu	0,5	0,5
		4	
		Choveu	Não Choveu
3	Choveu	0,7	0,3
	Não Choveu	0,3	0,7
		6	
		Choveu	Não Choveu
5	Choveu	0,8	0,2
	Não Choveu	0,1	0,9
		8	
		Choveu	Não Choveu
7	Choveu	0,2	0,8
	Não Choveu	0,2	0,8
		10	
		Choveu	Não Choveu
9	Choveu	0,5	0,5
	Não Choveu	0,5	0,5
		12	
		Choveu	Não Choveu
11	Choveu	0,3	0,7
	Não Choveu	0,6	0,4
		14	
		Choveu	Não Choveu
13	Choveu	0,6	0,4
	Não Choveu	0,4	0,6
		16	
		Choveu	Não Choveu
15	Choveu	0,2	0,8
	Não Choveu	0,2	0,8
		18	
		Choveu	Não Choveu
17	Choveu	0,0	1,0
	Não Choveu	0,4	0,6
		20	
		Choveu	Não Choveu
19	Choveu	0,8	0,3
	Não Choveu	0,4	0,6
		22	
		Choveu	Não Choveu
21	Choveu	0,4	0,6
	Não Choveu	0,2	0,8
		24	
		Choveu	Não Choveu
23	Choveu	0,5	0,5
	Não Choveu	0,2	0,8

		3	
		Choveu	Não Choveu
2	Choveu	0,2	0,8
	Não Choveu	0,2	0,8
		5	
		Choveu	Não Choveu
4	Choveu	0,5	0,5
	Não Choveu	0,3	0,7
		7	
		Choveu	Não Choveu
6	Choveu	0,8	0,2
	Não Choveu	0,1	0,9
		9	
		Choveu	Não Choveu
8	Choveu	0,0	1
	Não Choveu	0,2	0,8
		11	
		Choveu	Não Choveu
10	Choveu	0,3	0,7
	Não Choveu	0,5	0,5
		13	
		Choveu	Não Choveu
12	Choveu	0,6	0,4
	Não Choveu	0,1	0,9
		15	
		Choveu	Não Choveu
14	Choveu	0,6	0,4
	Não Choveu	0,3	0,8
		17	
		Choveu	Não Choveu
16	Choveu	1,0	0,0
	Não Choveu	0,1	0,9
		19	
		Choveu	Não Choveu
18	Choveu	0,3	0,8
	Não Choveu	0,3	0,7
		21	
		Choveu	Não Choveu
20	Choveu	0,4	0,6
	Não Choveu	0,3	0,8
		23	
		Choveu	Não Choveu
22	Choveu	0,5	0,5
	Não Choveu	0,2	0,8
		25	
		Choveu	Não Choveu
24	Choveu	0,3	0,8
	Não Choveu	0,4	0,6

		26	
		Choveu	Não Choveu
25	Choveu	0,2	0,8
	Não Choveu	0,3	0,7
		28	
		Choveu	Não Choveu
27	Choveu	0,5	0,5
	Não Choveu	0,1	0,9
		30	
		Choveu	Não Choveu
29	Choveu	0,8	0,3
	Não Choveu	0,2	0,8

		27	
		Choveu	Não Choveu
26	Choveu	0,5	0,5
	Não Choveu	0,4	0,6
		29	
		Choveu	Não Choveu
28	Choveu	0,3	0,8
	Não Choveu	0,3	0,7