



**Universidade de Brasília
Departamento de Estatística**

Modelagem de dados de insolação por meio de cadeias de ordem variável

Aline de Almeida Ramos

Projeto apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

**Brasília
2023**

Aline de Almeida Ramos

Modelagem de dados de insolação por meio de árvores de contexto

Orientador: Prof. Dr. Lucas Moreira
Coorientador: Prof. Dr. Felipe Sousa Quintino

Projeto apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

Brasília
2023

Agradecimentos

Primeiramente, agradeço a Deus pelas maravilhas que Ele me concede diariamente. Sua presença e bênçãos têm sido fonte de inspiração e força.

À minha irmã, Amanda, por seu constante ensinamento, ajuda e amor.

Aos meus pais Luiz e Valneide, por todo o amor recebido e por todos os sacrifícios feitos em prol da educação das filhas.

Ao meu namorado, Daniel, por todo apoio e conforto durante essa jornada.

Aos meus amigos, por me motivarem nos momentos de desânimo. Um agradecimento especial às minhas amigas Luany, Juliana e Geovanna. Cada uma, à sua maneira, contribui para a minha felicidade.

Ao meus queridos orientadores, Prof. Dr. Lucas Moreira e Prof. Dr. Felipe Quintino, por toda a orientação e encorajamento que me proporcionaram ao longo da elaboração deste trabalho.

Resumo

O presente estudo tem como principal objetivo a estimação de árvores de contexto para dados de insolação utilizando Cadeias de Ordem Variável. As árvores foram estimadas para diferentes cidades do Brasil, localizadas em regiões distintas do país com dados do Instituto Nacional de Meteorologia (INMET). As cidades são, Brasília-DF, Boa Vista-RR, Natal-RN, Vitória-ES e Passo Fundo-RS. Um dos objetivos específicos do trabalho é permitir a comparação das árvores de contexto a fim de analisar o comportamento da insolação em diferentes partes do país. Outro objetivo estabelecido é a previsão do próximo dia, indicando se haverá pouca, média ou alta incidência de insolação, com base nas probabilidades de transição obtidas. A insolação como objeto de estudo pode permitir uma melhor compreensão das características climáticas e pode auxiliar na tomada de decisões em diversas áreas que dela dependem. Para tanto, a partir da coleta dos dados e tomando valores num alfabeto finito para categorizá-los, foi possível a realização dos objetivos destacados utilizando o Algoritmo Contexto, inicialmente proposto por Rissanen (1983). As cidades apresentam árvores distintas para um mesmo período de anos evidenciando as diferenças climáticas em que estão envolvidas. No caso específico de Brasília, as árvores foram estimadas para diferentes períodos, revelando também mudanças climáticas pelas quais a cidade passou. Dessa forma, constatamos que o maior número de contextos encontrados corresponde a treze enquanto o menor corresponde a três e se iguala a uma Cadeia de Markov de ordem um.

Palavras-chaves: Processos Estocásticos, Insolação, Cadeias de Ordem Variável, INMET.

Abstract

The present study aims to estimate context trees for insolation data using Variable Order Markov Chains. The trees were estimated for different cities in Brazil, located in distinct regions of the country, using data from the Instituto Nacional de Meteorologia (INMET). The cities include Brasília-DF, Boa Vista-RR, Natal-RN, Vitória-ES, and Passo Fundo-RS. One of the specific objectives of the study is to allow for the comparison of context trees in order to analyze the behavior of insolation in different parts of the country. Another established objective is the prediction of the next day, indicating whether there will be low, medium, or high insolation, based on the obtained transition probabilities. Studying insolation can provide a better understanding of climatic characteristics and assist in decision-making in various fields that depend on it. To achieve these objectives, data collection was conducted, and by categorizing the values into a finite alphabet, the highlighted objectives were accomplished using the Context Algorithm, initially proposed by Rissanen (1983). The cities exhibit distinct context trees for the same period, highlighting the climatic differences they are subjected to. In the specific case of Brasília, trees were estimated for different periods, revealing the climate changes the city has undergone. It was observed that the highest number of contexts found was thirteen, while the lowest corresponded to three, equivalent to a first-order Markov chain.

Keywords: Stochastic Processes, Insolation, Variable Order Chains, INMET.

Sumário

1	Fundamentação Teórica	11
1.1	Processos Estocásticos	11
1.2	Cadeias de Markov	11
1.2.1	Modelos Probabilísticos de Ordem k	12
1.3	Cadeias de Markov de Alcance Variável	12
1.4	Algoritmo Contexto.	16
2	Metodologia	19
2.1	Conjunto de Dados	19
2.2	Categorização dos Dados	20
2.3	Estimação das Árvores Probabilísticas de Contexto.	21
3	Resultados	22
3.1	Análise Exploratória	22
3.2	Árvores de Contexto Estimadas	24
3.3	Estimação do Próximo Símbolo	34
4	Conclusão	36

Lista de Figuras

2.1	Insolação Diária em Horas Considerando a Média Anual do Brasil, 2000. . .	20
3.1	Distribuição do Número de Horas de Insolação para Cinco Cidades no Brasil dos anos de 1980 a 2020.	23
3.2	Árvore Probabilística de Contextos Para a Cidade de Brasília dos anos de 1980 a 2000.	24
3.3	Árvore Probabilística de Contextos Para a Cidade de Brasília dos anos de 2001 a 2020.	26
3.4	Árvore Probabilística de Contextos Estimada Para a Cidade de Boa Vista dos anos 2001 a 2020.	28
3.5	Árvore Probabilística de Contextos Para a Cidade de Natal dos anos de 2001 a 2020.	30
3.6	Árvore Probabilística de Contextos Para a Cidade de Vitória dos anos de 2001 a 2020.	32
3.7	Árvore Probabilística de Contextos Para a Cidade de Passo Fundo dos anos de 2001 a 2020.	33

Lista de Tabelas

3.1	Medidas Descritivas da Variável Número de Horas de Insolação Para Cinco Cidades no Brasil, anos de 1980 a 2020	23
3.2	Probabilidades de Transição Estimadas. Brasília – DF, 1980–2000.	25
3.3	Probabilidades de Transição Estimadas. Brasília – DF, 2001–2020.	27
3.4	Probabilidades de Transição Estimadas. Boa Vista – RR, 2001–2020.	29
3.5	Probabilidades de Transição Estimadas. Natal – RN, 2001–2020.	31
3.6	Probabilidades de Transição Estimadas. Vitória – ES, 2001–2020.	32
3.7	Probabilidades de Transição Estimadas. Passo Fundo – RS, 2001–2020.	33
3.8	Proporção de Acertos da Estimação do Próximo Símbolo para Cidades no Brasil, anos de 1980 a 2020	34

Introdução

Medeiros, Holanda e França (2018) definem insolação como o período do dia em que há a presença de luz solar, ou seja, quando há visibilidade do disco solar para um observador na Terra. O fenômeno é capaz de auxiliar na identificação das estações do ano. Nesse sentido, no estudo de Saraiva (2004) foi constatado que a insolação para o município de Porto Alegre, no Estado do Rio Grande do Sul, é cerca de 66% maior no verão do que no inverno.

Além disso, devido à sua forma esférica, a insolação não é a mesma para todos os lugares do planeta. Dessa forma, é razoável pensar que a insolação é distinta entre um país sul americano e o Alasca, por exemplo. De forma específica, Saraiva (2004) afirma que a insolação é dada matematicamente pela razão entre a energia solar média que chega à superfície terrestre e a unidade de área.

Assim como a insolação, variáveis climáticas como a umidade do ar e o nível de chuvas também diferem entre estações em uma mesma cidade e entre locais variados. Nesse sentido, Borges (2018) apresenta um modelo de previsão da umidade relativa do ar para o dia seguinte em Brasília, o qual utiliza o conceito de Cadeias de Ordem Variável proposto por Rissanen (1983) dentro da Teoria de Processos Estocásticos. Esse arcabouço também foi utilizado por Cavalcante (2018) para prever o nível de chuva no Distrito Federal no dia seguinte dado o histórico realizado até o dia atual.

Rissanen (1983) definiu os modelos de Cadeias de Ordem Variável como aqueles em que a quantidade de informações do passado necessárias para prever a próxima ocorrência não é fixa, mas depende da sequência dos eventos anteriores. Essa parte relevante do passado recebe o nome contexto. Dessa forma, o conjunto de todos os contextos, pode ser representado por uma estrutura probabilística denominada árvore de contextos.

Dado o panorama acima, o presente trabalho estudará o conceito de insolação dentro da Teoria de Processos Estocásticos, buscando modelar os dados de insolação do Instituto Nacional de Meteorologia - (INMET, 2023) em cinco cidades brasileiras por meio de Cadeias de Ordem Variável. Especificamente, serão estimadas árvores probabilísticas de contexto para uma cidade em cada uma das regiões brasileiras. As cidades são, Brasília-DF, Boa Vista-RR, Natal-RN, Vitória-ES e Passo Fundo-RS. Para compreender o nível de insolação nessas localidades, serão estimadas árvores de contexto com base no referencial de Rissanen (1983).

A ideia da estimação por meio de árvores de contextos é que, para cada pas-

sado, somente uma porção dele é suficiente para prever a próxima ocorrência. Rissanen (1983) também introduziu o Algoritmo Contexto, desenvolvido para estimar árvores probabilísticas de contexto. Para estimar as árvores de probabilidade, será construído um alfabeto para os dados, o que implica em categorizá-los de acordo com algum critério ou comportamento de interesse.

O trabalho está estruturado em quatro capítulos. No Capítulo 1 é apresentada uma revisão de conceitos de Processos Estocásticos. No Capítulo 2, é apresentada a metodologia utilizada. Os resultados são apresentados no Capítulo 3. O Capítulo 4 contém a conclusão do trabalho, destacando suas principais contribuições.

Capítulo 1

Fundamentação Teórica

Este capítulo traz alguns conceitos que foram utilizados no desenvolvimento do trabalho. Assim, são explicitados tópicos dentro da teoria de Processos Estocásticos.

1.1 Processos Estocásticos

Ross et al. (1996) definem um processo estocástico $\underline{X} = \{X(t), t \in T\}$ como sendo uma coleção de variáveis aleatórias indexada a uma variável. Ou seja, para cada t , $X(t)$ é uma variável aleatória. Frequentemente, t é interpretado como tempo e os valores que $X(t)$ podem assumir são conhecidos como estados. T é chamado de espaço paramétrico. Se T for um conjunto finito ou contável, chamamos \underline{X} de processo estocástico discreto e, se T é não-contável, chamamos \underline{X} de processo em tempo contínuo.

Dentro da teoria de processos estocásticos, um símbolo se refere a um elemento ou valor discreto que compõe uma sequência ou trajetória de um processo estocástico.

1.2 Cadeias de Markov

Para definir uma Cadeia de Markov, Ross et al. (1996) consideram um processo estocástico X_n , $n = 0, 1, 2, 3, \dots$, com espaços paramétricos e de estados discretos. Quando $X_n = i$, lê-se como o processo está no estado i no tempo n e corresponde a uma cadeia de Markov.

Dentro de uma Cadeia de Markov, tem-se o que é conhecido como propriedade de Markov, onde a distribuição condicional para algum tempo no futuro dado o passado e o

presente, só depende do presente. Matematicamente, essa propriedade é dada por (1.2.1).

$$\begin{aligned} P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) &= P(X_{n+1} = j | X_n = i) \\ &= P(ij) \end{aligned} \quad (1.2.1)$$

onde $n+1$ representa o tempo futuro, X_0 representa o estado inicial, X_n denota o presente, e $i, j, i_0, \dots, i_{n-1} \in S$ para todo $n > 0$. Além disso, $P(ij)$ é definida como a probabilidade de transição do estado i para o estado j em um único passo.

Nesse sentido, a título de exemplo, seja um processo estocástico X_n que representa o estado de uma máquina no momento n . O espaço de estados seria dado por ‘ligado’ e ‘desligado’. Especificamente, seja $i = 0$ o estado em que a máquina se encontra ligada e $i = 1$ desligada, então $P(01)$ corresponde a probabilidade da máquina estar desligada em um tempo futuro dado que no presente ela se encontra ligada.

1.2.1 Modelos Probabilísticos de Ordem k

Conforme definição apresentada em Matta et al. (2018), o processo $\{X_t, t \in \mathbb{Z}\}$ é denominado cadeia de Markov de ordem k se existir um elemento $k \in \mathbb{N}$ tal que, para todo $t \in \mathbb{Z}$, temos

$$P(X_t = x_t | X_{t-1} = x_{t-1}, X_{t-2} = x_{t-2}, \dots) = P(X_t = x_t | X_{t-1} = x_{t-1}, \dots, X_{t-k} = x_{t-k}),$$

para toda sequência $x_t, x_{t-1}, x_{t-2}, \dots \in \mathcal{A}$, em que $\mathcal{A} = \{0, 1, 2, 3, \dots, N\}$ é um alfabeto, definido como qualquer conjunto finito formado por símbolos.

A cadeia é dita estacionária se, para todo $a \in \mathcal{S}$, tivermos $P(X_n = a) = \pi(a)$, onde π é a medida estacionária da cadeia. Além disso, no caso em que $k = \infty$, temos uma cadeia de ordem infinita, ou seja, não há limitação na ordem da cadeia.

1.3 Cadeias de Markov de Alcance Variável

Seja $\mathcal{A} = \{0, 1, 2, \dots, N-1\}$, onde $N \in \mathbb{N}$ e $N > 0$, um alfabeto finito de tamanho $|\mathcal{A}| = N$. Considere $m \leq n$ que geram a sequência (x_m, \dots, x_n) , que pode ser reescrita como x_m^n , representando uma sequência de símbolos em \mathcal{A} com comprimento $l(x_m^n) = n - m + 1$. A sequência vazia é denotada por \emptyset e possui comprimento 0.

uma sequência refere-se a uma série ordenada de símbolos que ocorrem em uma determinada ordem. Como definido em Csiszár e Talata (2005) é possível concatenar

sequências. Assim, se tivermos duas sequências finitas v e ω , a concatenação de ambas $v\omega$ gerará uma nova sequência de comprimento $l(v) + l(\omega)$. Também, no caso em que uma das sequências seja \emptyset , por exemplo, $v = \emptyset$, então $v\omega = \omega$.

O conjunto de todas as sequências semi-infinitas pode ser definido por

$$\mathcal{A}_{-\infty}^{-1} = \mathcal{A}^{\{\dots, -2, -1\}} \quad (1.3.1)$$

enquanto o conjunto de todas as sequências de símbolos de tamanho finito é definida por

$$\mathcal{A}^* = \bigcup_{j=0}^{\infty} \mathcal{A}_{-j}^{-1} \quad (1.3.2)$$

em que $j = 0$ representa o conjunto das sequências vazias.

Uma sequência u é dita ser um *sufixo* de ω se existir s , com $l(s) \geq 1$, tal que $\omega = su$ e será denotada por $u \prec \omega$. Caso $u \prec \omega$ ou $u = \omega$, será denotado por $u \preceq \omega$. Dada uma sequência finita ω denotamos por $\text{suf}(\omega)$ o maior sufixo de ω .

Por exemplo, sejam $\mathcal{A} = \{0, 1\}$, $v = 0111$ e $\omega = 00110001$.

- $v\omega = 011100110001$.
- $0001 \prec \omega$; $s = 0011$.
- $\text{suf}(v) = 111$.

Rissanen (1983) desenvolve a ideia de que o comprimento da memória necessária para prever o próximo símbolo não é fixo, mas é uma função do passado. Seu objetivo era construir um modelo em que fosse possível utilizar a parte relevante do passado, a qual ele chamou contexto.

O conjunto de todos os contextos satisfaz a propriedade do sufixo, e portanto, nenhum contexto é sufixo de outro contexto. Essa característica permite a construção de árvores probabilística com raiz e rótulos, a qual chamamos árvore de contexto do processo \mathbf{X} .

O trabalho considera o processo $\mathbf{X} = \{X_t, t \in \mathbb{Z}\}$ estacionário e ergódico sobre o alfabeto $\mathcal{A} = \{0, 1, 2, \dots, N - 1\}$. Assim, diremos que o processo \mathbf{X} é compatível com a probabilidade de transição $p_X(\cdot|\cdot)$, ou seja,

$$p_X(a|\omega) = P(X_0 = a | X_{-1} = \omega_{-1}, X_{-2} = \omega_{-2}, \dots), \quad (1.3.3)$$

para todo $\omega \in \mathcal{A}_{-\infty}^{-1}$ e para todo $a \in \mathcal{A}$. Para $\omega \in \mathcal{A}_{-j}^{-1}$ a probabilidade estacionária do cilindro definida por essa sequência será denotada por

$$\mu_X(\omega) = P(X_{-j}^{-1} = \omega). \quad (1.3.4)$$

Para que seja possível estimar a árvore de contextos de um processo \mathbf{X} , é necessário que \mathbf{X} satisfaça as seguintes definições.

Definição 1 Dizemos que um processo \mathbf{X} é não-nulo se satisfaz

$$\alpha_X = \inf\{p_X(a|\omega) : a \in \mathcal{A}, \omega \in \mathcal{A}_{-\infty}^{-1}\} > 0. \quad (1.3.5)$$

Definição 2 Dizemos que um processo \mathbf{X} possui taxa de continuidade somável se

$$\beta_X = \sum_{k \in \mathbb{N}} \beta_{k,X} < +\infty, \quad (1.3.6)$$

em que a sequência $\{\beta_{k,X}\}_{k \in \mathbb{N}}$ é definida por

$$\beta_{k,X} := \sup \left\{ \left| 1 - \frac{p_X(a|\omega)}{p_X(a|v)} \right| : a \in \mathcal{A}, v, \omega \in \mathcal{A}_{-\infty}^{-1} \text{ com } \omega_{-k}^{-1} = v_{-k}^{-1} \right\}. \quad (1.3.7)$$

A sequência $\{\beta_{k,X}\}_{k \in \mathbb{N}}$ é chamada *taxa de continuidade do processo \mathbf{X}* . Essa taxa é uma propriedade esperada para o processo \mathbf{X} , pois, desejamos que dois passados coincidindo nos últimos k símbolos tenham a mesma influência na predição do próximo símbolo da sequência, a medida que k cresce.

A seguir será melhor definido aspectos conceituais que caracterizam o contexto.

Definição 3 Dizemos que uma sequência $\omega \in \mathcal{A}_{-j}^{-1}$ é um contexto do processo se para toda sequência semi-infinita $x_{-\infty}^{-1} \in \mathcal{A}_{-\infty}^{-1}$ tendo ω como sufixo satisfazer

$$P(X_0 = a | X_{-\infty}^{-1} = x_{-\infty}^{-1}) = p_X(a|\omega), \quad \forall a \in \mathcal{A}, \quad (1.3.8)$$

e nenhum sufixo de ω satisfaz a equação (1.3.8).

Denotamos por $d(\mathcal{T})$ a profundidade da árvore, ou seja,

$$d(\mathcal{T}) := \max\{l(\omega) : \omega \in \mathcal{T}\}.$$

Para estimação da árvore probabilística de contextos não é necessário o conhecimento prévio da sua profundidade. Porém, é necessário estabelecer uma profundidade máxima de busca $D > d(\mathcal{T})$.

Uma árvore \mathcal{T} é dita *completa* se qualquer sequência em $\mathcal{A}_{-\infty}^{-1}$ pertence a \mathcal{T} ou tem sufixo que pertence a \mathcal{T} . Dizemos que a árvore de contextos é *limitada* se o comprimento do maior contexto é finito. Caso contrário, \mathcal{T} é dita *ilimitada*.

Dizemos que uma árvore é *irreduzível* se nenhuma sequência pode ser substituída por um sufixo sem violar a propriedade sufixo. Essa noção foi introduzida em Csiszár e Talata (2006) e generaliza o conceito de árvore completa.

A seguir definiremos de maneira mais formal uma *árvore probabilística de contextos* e uma Cadeia de Ordem Variável.

Definição 4 *Uma árvore probabilística de contextos em \mathcal{A} é um par ordenado (\mathcal{T}, \bar{p}) que satisfaz*

- (1) \mathcal{T} é uma árvore irreduzível.
- (2) $\bar{p} = \{\bar{p}(\cdot|\omega), \omega \in \mathcal{T}\}$ é uma família de probabilidades de transição sobre \mathcal{A} .

Definição 5 *Dizemos que o processo \mathbf{X} é compatível com a árvore probabilística de contextos (\mathcal{T}, \bar{p}) se satisfaz*

- (1) \mathcal{T} é a árvore de contextos do processo \mathbf{X} .
- (2) Para qualquer $w \in \mathcal{T}$ e $a \in \mathcal{A}$, $p_{\mathbf{X}}(a|\omega) = \bar{p}(a|\omega)$.

Se \mathbf{X} é compatível com a árvore probabilística de contextos (\mathcal{T}, \bar{p}) , dizemos que \mathbf{X} é uma *Cadeia de Ordem Variável* e denotamos a árvore de contextos de \mathbf{X} por $\mathcal{T}_{\mathbf{X}}$. Note que em (1.3.6) da Definição 2, se $d(\mathcal{T}_{\mathbf{X}}) < +\infty$, então $\beta_{k,X} = 0$ para $k \geq d(\mathcal{T}_{\mathbf{X}})$, ou seja,

$$\beta_X = \sum_{k=0}^{d(\mathcal{T}_{\mathbf{X}})-1} \beta_{k,X} < +\infty.$$

Em alguns casos é possível estar interessado não na árvore de contextos do processo \mathbf{X} mas na utilização desta árvore com uma restrição no tamanho da maior sequência. Seja K esta restrição. Neste caso, chamaremos de árvore truncada no nível $K \geq 1, K \in \mathbb{N}$. Dessa forma, se definirmos $K \geq d(\mathcal{T}_{\mathbf{X}})$, estaremos considerando a própria árvore de contextos do processo \mathbf{X} .

Definição 6 Dado um inteiro K , defina a árvore de contextos truncada no nível K por

$$\mathcal{T}_X|_K = \{\omega \in \mathcal{T}_X : l(\omega) \leq K\} \cup \{\omega : l(\omega) = K \text{ e } \omega \prec u, \text{ para algum } u \in \mathcal{T}_X\}.$$

Considere $\mathbf{Z} = \{Z_t, t \in \mathbb{Z}\}$ um processo tomando valores num alfabeto finito $\mathcal{A} = \{0, 1, \dots, N-1\}$. Seja Z_1, \dots, Z_n uma amostra aleatória do processo \mathbf{Z} . Para toda sequência finita ω , com $l(\omega) \leq n$, denotamos por $N_n(\omega)$ o número de vezes que observou-se a sequência ω na amostra, ou seja,

$$N_n(\omega) = \sum_{t=0}^{n-l(\omega)} \mathbf{1}_{\{Z_{t+1}^{t+l(\omega)} = \omega\}}. \quad (1.3.9)$$

Para todo elemento $a \in \mathcal{A}$ e para toda sequência finita ω , a probabilidade de transição empírica é dada por

$$\hat{p}_Z(a|\omega)_n = \frac{N_n(\omega a) + 1}{N_n(\omega) + |\mathcal{A}|}. \quad (1.3.10)$$

Observamos que a Definição 1.3.10 implica que $\hat{p}_Z(a|\omega)_n$ é assintoticamente equivalente ao Estimador de Máxima Verossimilhança de $p_Z(a|\omega)_n$ que é $\frac{N_n(\omega a)}{N_n(\omega)}$ e evita-se uma definição adicional no caso $N(\omega) = 0$.

1.4 Algoritmo Contexto

O Algoritmo Contexto proposto por Rissanen (1983) é um método de compressão de dados baseado na ideia de que a ocorrência de um símbolo em um determinado contexto é influenciada pelos símbolos anteriores e baseia-se, inicialmente, na maior árvore factível. O algoritmo busca identificar esses contextos e estimar as probabilidades de ocorrência dos diferentes símbolos com base neles.

Seja q um inteiro positivo. Chamamos de conjunto de todas as sequências $V_n = \left\{ \omega \in \bigcup_{j=0}^q A^j : N_n(\omega) \geq 1 \right\}$

Definição 7 Uma árvore \mathcal{T} é factível se satisfaz:

1. $\omega \in V_n$ para todo $\omega \in \mathcal{T}$;
2. Para cada sequência $\omega' \in V_n$, existe uma sequência $\omega \in \mathcal{T}$ tal que $\omega \preceq \omega'$ ou $\omega \prec \omega'$.

O conjunto de todas as árvores factíveis será denotado por \mathcal{F}_n . O objetivo é estimar a árvore de contextos \mathcal{T}_X a partir de uma amostra de \mathbf{X} .

Inicialmente, o Algoritmo Contexto baseia-se, na maior árvore factível. Em seguida, avalia as diferenças entre as probabilidades de transição observadas para o maior sufixo de um contexto e seus descendentes. Se a discrepância entre essas probabilidades for superior a um valor limite predefinido, o contexto é considerado relevante e mantido na árvore. Caso contrário, o contexto é podado, ou seja, removido da árvore. Esse processo é repetido até que não seja mais possível podar a árvore, resultando em uma árvore final que representa os contextos relevantes do processo.

Para medir a discrepância entre duas medidas de probabilidade em \mathcal{A} , utilizamos a divergência de Kullback-Leibler, definida a seguir.

Definição 8 *A divergência de Kullback-Leibler, definida para medidas de probabilidade P e Q em \mathcal{A} , é dada por*

$$D(P; Q) = \sum_{a \in \mathcal{A}} P(a) \log \frac{P(a)}{Q(a)}. \quad (1.4.1)$$

Por convenção, $P(a) \log \frac{P(a)}{Q(a)} = 0$ se $P(a) = 0$, e $P(a) \log \frac{P(a)}{Q(a)} = +\infty$ se $P(a) > Q(a) = 0$.

Para uma sequência $\omega \in V_n$, seja

$$\Lambda_n(\omega) = \sum_{b \in \mathcal{A}: b\omega \in V_n} N_n(b\omega) D(p_{bn}(\cdot|b\omega), p_{bn}(\cdot|\omega)). \quad (1.4.2)$$

Denotamos o limiar utilizado no Algoritmo Contexto por δ_n , em que $(\delta_n)_{n \in \mathbb{N}}$ é uma sequência de números reais tal que $\delta_n \rightarrow \infty$ e $\frac{\delta_n}{n} \rightarrow 0$ quando $n \rightarrow \infty$. Podemos descrever o Algoritmo Contexto em três passos:

Denotamos o limiar utilizado no Algoritmo Contexto por δ_n , onde $(\delta_n)_{n \in \mathbb{N}}$ é uma sequência de números reais que tende ao infinito ($\delta_n \rightarrow \infty$) e satisfaz a condição $\frac{\delta_n}{n} \rightarrow 0$ quando $n \rightarrow \infty$. Podemos descrever o Algoritmo Contexto em três passos da seguinte maneira:

Passo 1: Construção da maior árvore factível \mathcal{F}_n .

Passo 2: Medição da discrepância entre as probabilidades empíricas de transição do maior sufixo de um contexto e seus descendentes. Se a discrepância for maior que δ_n , o contexto é mantido; Caso contrário, é podado.

Passo 3: Repetição do Passo 2 até que não seja mais possível podar a árvore.

Para uma amostra X_1^n , podemos obter $\widehat{\mathcal{T}}_C(X_1^n)$ utilizando a função $C_\omega(X_n)$ definida para cada $\omega \in V_n$:

$$C_{\omega}(X_1^n) = \begin{cases} 0, & \text{se } N_n(\omega) \leq 1 \text{ ou } l(\omega) = d \\ \max\{I(\Lambda_n(\omega) \geq \delta_n), \max_{b \in A} C_{b\omega}(X_1^n)\}, & \text{se } N_n(\omega) > 1 \text{ e } l(\omega) < d \end{cases}$$

Definição 9 O estimador $\widehat{\mathcal{T}}_C(X_1^n)$ da árvore de contextos de \mathbf{X} é o conjunto dado por $\widehat{\mathcal{T}}_C(X_1^n) = \{\omega \in V_n : C_{\omega}(X_1^n) = 0 \text{ e } C_{\omega'}(X_1^n) = 1 \text{ para todo } \omega' \prec \omega\}$.

Capítulo 2

Metodologia

Neste capítulo será definida a metodologia utilizada no desenvolvimento do trabalho. O tema abordado está dentro da Teoria de Processos Estocásticos e todas as análises estatísticas são realizadas através do Software R.

2.1 Conjunto de Dados

Os dados que serão utilizados para as análises foram coletados no site do Instituto Nacional de Meteorologia (INMET, 2023). O conjunto de dados abrange o intervalo de tempo de quarenta anos, onde o início data de primeiro de janeiro de 1980 e o fim em 31 de dezembro de 2020. As cidades de interesse são: Brasília – DF, no centro-oeste do país, Boa Vista – RR na região norte, Natal – RN na região nordeste, Vitória – ES no sudeste brasileiro e por fim Passo Fundo – RS, na região sul do Brasil.

Nº	Nome
1	Data Medicao
2	INSOLACAO TOTAL, DIARIO(h) Boa Vista
3	INSOLACAO TOTAL, DIARIO(h) Brasília
4	INSOLACAO TOTAL, DIARIO(h) Passo Fundo
5	INSOLACAO TOTAL, DIARIO(h) Natal
6	INSOLACAO TOTAL, DIARIO(h) Vitória

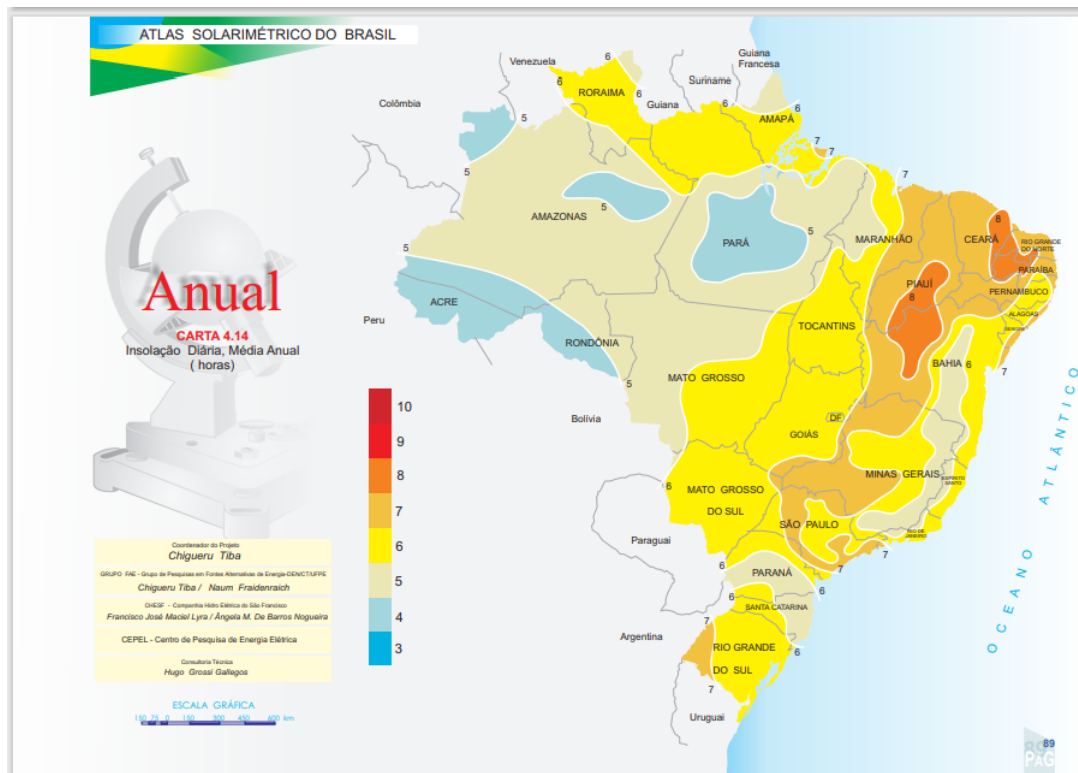
Quadro 2.1: Variáveis a Serem Analisadas no Desenvolvimento do Trabalho

2.2 Categorização dos Dados

Em 1994 foi submetido ao Centro de Pesquisas de Energia Elétrica – CEPTEL – o Projeto para a elaboração do Atlas Solarimétrico Nacional. À época acreditava-se que o Atlas seria capaz de por à disposição do público interessado uma importante base de dados que, na medida em que a tecnologia solar se difundisse no Brasil tornaria-se crescentemente importante. Dessa forma, foi elaborado o Atlas Solarimétrico do Brasil, no ano de 2000.

A Figura 2.1 retrata em uma escala de 3-10 horas, a insolação diária para o Brasil, considerando a média anual. Com base nessa escala e em outros estudos, como a análise exploratória dos dados, será definida uma primeira categorização para os dados.

Figura 2.1: Insolação Diária em Horas Considerando a Média Anual do Brasil, 2000.



Fonte: Atlas Solarimétrico do Brasil, Banco de Dados Terrestres, ano 2000.

2.3 Estimação das Árvores Probabilísticas de Contexto

Uma vez definido o alfabeto com a categorização dos dados, os passos seguintes consistem em estimar as árvores probabilísticas de contexto para os diferentes cenários de investigação, por meio do Algoritmo Contexto.

O código utilizado no trabalho para estimação das árvores foi desenvolvido por Bomfim (2015).

Capítulo 3

Resultados

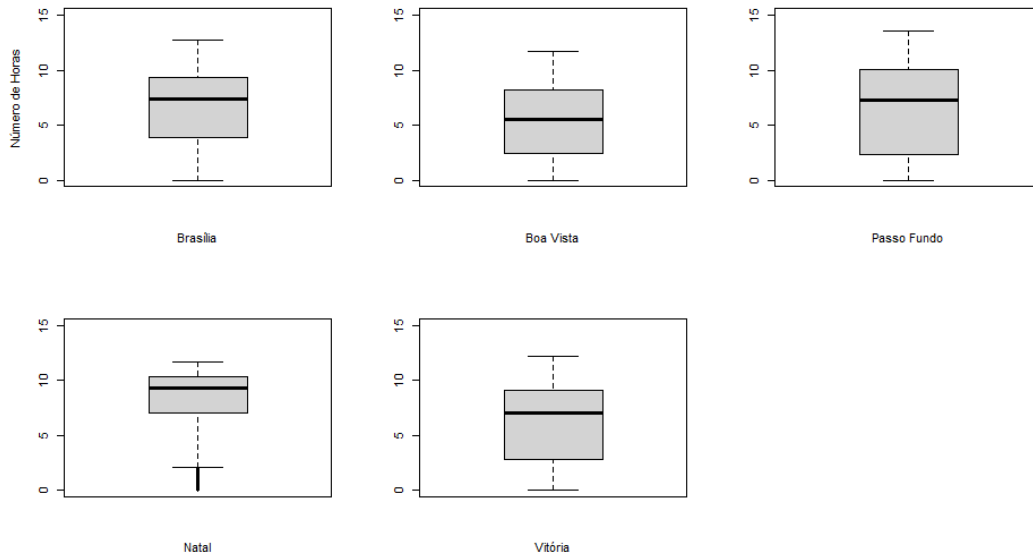
3.1 Análise Exploratória

Uma análise exploratória dos dados foi desenvolvida a fim de conhecer o comportamento das variáveis a serem estudadas.

A Figura 3.1 retrata o comportamento da variável que representa o número de horas de insolação para as cidades em estudo. Natal – RN, Brasília – DF e Passo Fundo – RS apresentam medianas maiores que Boa Vista – RR e Vitória – ES.

Para todas as cidades há certa variabilidade nos dados. A posição da linha da mediana é mais próxima ao terceiro quartil para os cinco gráficos, o que indica que os dados são levemente assimétricos negativos.

Figura 3.1: Distribuição do Número de Horas de Insolação para Cinco Cidades no Brasil dos anos de 1980 a 2020.



Os dados numéricos confirmam o que se visualiza nos gráficos. Há maior variabilidade na região sul e menor na região norte. A média dos dados se mantém entre 5 e 7 horas de insolação diárias.

Tabela 3.1: Medidas Descritivas da Variável Número de Horas de Insolação Para Cinco Cidades no Brasil, anos de 1980 a 2020

	Brasília	Boa Vista	Passo Fundo	Natal	Vitória
Média	6.5	5.3	6.4	8.2	6.0
Q1	3.9	2.4	2.3	7.0	2.8
Mediana	7.4	5.5	7.3	9.3	7.0
Q3	9.4	8.2	10.1	10.3	9.1
Variância	10.7	10.8	17.7	8.57	12.7
DP	3.3	3.8	4.2	2.93	3.6
CV	0.5	0.6	0.7	0.36	0.6

Nota-se que as cidades de Passo Fundo, Natal e Vitória têm variâncias maiores em comparação com Brasília e Boa Vista, indicando uma maior variabilidade nas horas de insolação nessas cidades.

De acordo com a escala média do número de horas de insolação diária, desenvol-

vida pelo Atlas Solarimétrico do Brasil e considerando também a distribuição dos dados nos boxplots, optou-se pela categorização dos dados em uma escala mínima de 3 horas diárias e máxima de 10. Dessa forma, o alfabeto foi estabelecido como na tabela a seguir:

Nº de horas	Alfabeto
3 - 5 horas e 59 minutos	0
6 - 7 horas e 59 minutos	1
8 - 10 horas	2

Quadro 3.1: Escala de Insolação Diária

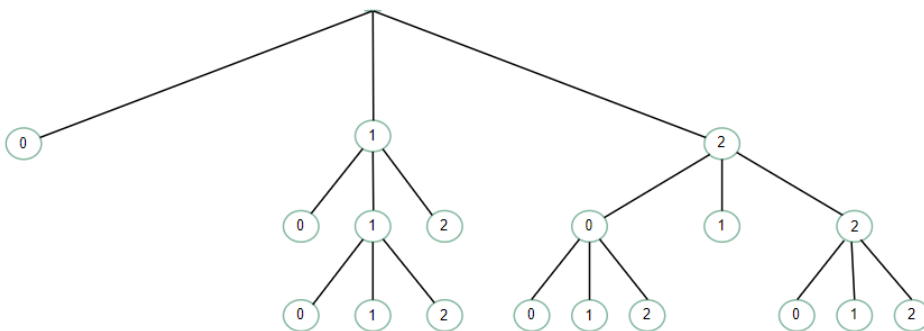
Com base na categorização estabelecida, procedeu-se à estimativa das árvores de contexto em diferentes cenários.

3.2 Árvores de Contexto Estimadas

As árvores são representações gráficas em que cada nó está representado um contexto específico e as probabilidades associadas a cada transição indicam a probabilidade de transitar de um contexto para outro.

O primeiro cenário considera as árvores, acompanhada das probabilidades de transição, ao longo de diferentes anos na cidade de Brasília.

Figura 3.2: Árvore Probabilística de Contextos Para a Cidade de Brasília dos anos de 1980 a 2000.



A Figura 3.2 elucida a presença de 13 contextos para os anos do século XX na cidade de Brasília. Para os primeiros nós, nota-se que existem descendentes apenas para os estados 1 e 2.

A Tabela 3.2 conta com as probabilidades de transição associadas a cada um dos 13 contextos encontrados.

Tabela 3.2: Probabilidades de Transição Estimadas. Brasília – DF, 1980–2000.

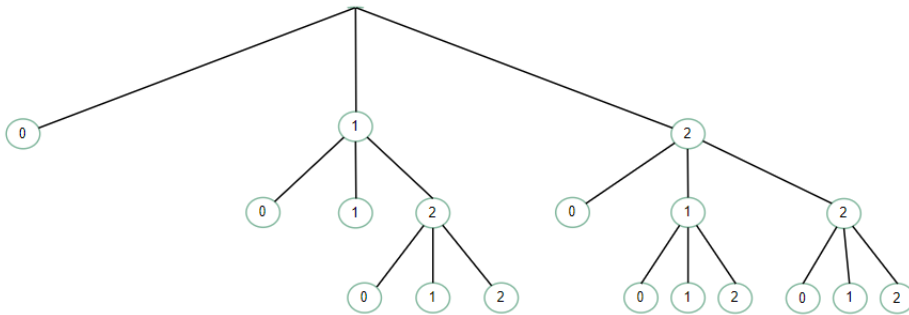
a	s	$\hat{p}(a s)_n$
0	0	0.2422
0	01	0.2116
0	21	0.1215
0	12	0.0608
0	011	0.2667
0	111	0.1850
0	211	0.1418
0	002	0.0882
0	102	0.1389
0	202	0.0500
0	022	0.0633
0	122	0.0521
0	222	0.0236
1	0	0.4727
1	01	0.4979
1	21	0.4338
1	12	0.2944
1	011	0.4250
1	111	0.4653
1	211	0.4752
1	002	0.5294
1	102	0.3472
1	202	0.2500
1	022	0.3038
1	122	0.2208
1	222	0.1644

Dessa forma, as maiores probabilidades de transição encontradas estão associadas a saída do contexto ‘222’ para estado ‘2’ (81.21%) seguida da saída do contexto ‘122’ para o estado ‘2’ (47.52%). Por outro lado, a menor probabilidade de transição está relacionada a saída do estado ‘222’ para o estado ‘0’ (2.36%).

Em outras palavras, para o contexto ‘222’, se ocorrem três dias de alta insolação, a chance de que no dia seguinte também ocorra insolação alta é de 81.21%. Em contrapartida, se ocorrem três dias de alta insolação, é pouco provável que no dia posterior haja baixa insolação (2.36%).

A seguir está disposta uma análise similar feita para os anos seguintes na cidade de Brasília.

Figura 3.3: Árvore Probabilística de Contextos Para a Cidade de Brasília dos anos de 2001 a 2020.



Para 2001-2020, a cidade manteve uma quantidade constante de contextos, embora eles fossem diferentes entre si.

Tabela 3.3: Probabilidades de Transição Estimadas. Brasília – DF, 2001–2020.

a	s	$\hat{p}(a s)_n$
0	0	0.2626
0	01	0.2366
0	11	0.1817
0	02	0.1032
0	021	0.3729
0	121	0.1622
0	221	0.1162
0	012	0.0845
0	112	0.0811
0	212	0.0356
0	022	0.1875
0	122	0.0485
0	222	0.0206
1	0	0.4358
1	01	0.4509
1	11	0.4483
1	02	0.3806
1	021	0.4407
1	121	0.3986
1	221	0.3662
1	012	0.2676
1	112	0.3288
1	212	0.1993
1	022	0.2375
1	122	0.2168
1	222	0.1470

De maneira análoga, ao observar a Tabela 3.3, nota-se que a maior probabilidade de transição estimada ocorre quando há a saída do contexto ‘222’ para o estado ‘2’ (83.24%).

É possível observar também, ainda para o contexto ‘222’, ou seja, em três dias consecutivos de alta insolação, que é mais provável que o dia posterior tenha média insolação (14.69%) do que baixa insolação (2.06%), mesmo que essas probabilidades sejam relativamente baixas.

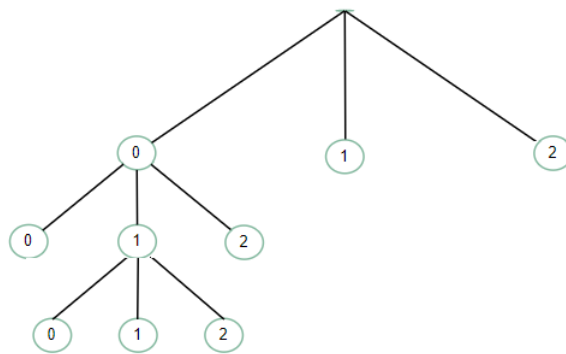
Quando a única informação obtida é sobre o nível de insolação atual, no caso,

baixa insolação as probabilidades se distribuem de forma mais equilibrada entre as possíveis opções para o próximo dia. Assim, existe uma probabilidade de 26.26% de que o próximo dia permaneça no mesmo estado, uma probabilidade de 43.58% de que o próximo dia tenha média insolação e uma probabilidade de 30.16% de que o dia seguinte tenha alta insolação.

Comparando as duas árvores estimadas para Brasília, é evidente que houve uma mudança climática ao longo do tempo, refletida nos diferentes contextos encontrados. Notamos que a tendência se concentra cada vez mais em dias com média ou alta insolação.

Estendendo a estimação das árvores para a região norte do país, a análise a seguir considera a capital de Roraima, Boa Vista.

Figura 3.4: Árvore Probabilística de Contextos Estimada Para a Cidade de Boa Vista dos anos 2001 a 2020.



É evidente, quando comparadas às árvores de Brasília, a presença de menos contextos nas árvores em questão, totalizando 7 contextos.

Tabela 3.4: Probabilidades de Transição Estimadas. Boa Vista – RR, 2001–2020.

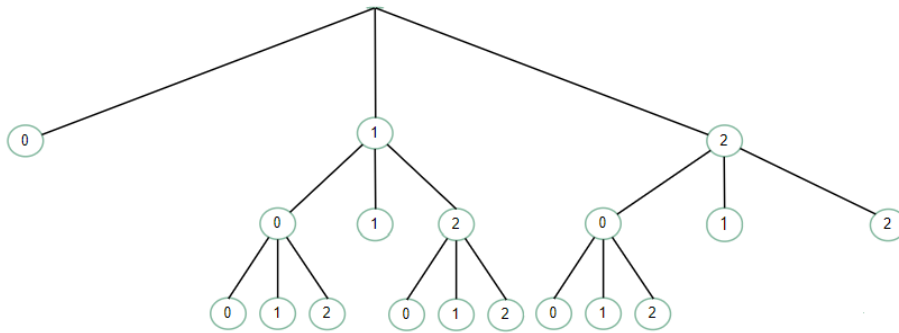
a	s	$\hat{p}(a s)_n$
0	1	0.2026
0	2	0.1325
0	00	0.1500
0	20	0.2258
0	010	0.5000
0	110	0.1538
0	210	0.4167
1	1	0.3791
1	2	0.3120
1	00	0.5000
1	20	0.1935
1	010	0.0000
1	110	0.4615
1	210	0.0833

É possível observar, pela Tabela 3.4, que há uma probabilidade significativa de transições entre os diferentes níveis de insolação. Por exemplo, em um dia de média insolação (estado ‘1’) a probabilidade de que o próximo dia tenha baixa insolação (estado ‘0’) é de aproximadamente 20.26%, enquanto a permanência no mesmo estado tem probabilidade estimada de 37.91%.

Uma característica da árvore apresentada na Figura 3.4 é que para a sequência de dias com a configuração baixa–média–baixa insolação as probabilidades se concentram apenas na chance de que o próximo dia tenha baixa ou alta insolação. Isso sugere que, após essa sequência específica, a tendência é que o padrão de insolação se mantenha em um extremo, seja baixa ou alta.

Ainda na parte superior do país, a árvore a seguir foi estimada para a cidade de Natal, capital do Rio Grande do Norte. Como o objetivo é comparar os contextos encontrados em diferentes partes do Brasil, o período de análise se manteve o mesmo.

Figura 3.5: Árvore Probabilística de Contextos Para a Cidade de Natal dos anos de 2001 a 2020.



Assim como as árvores estimadas para Brasília, conta com 13 contextos.

Tabela 3.5: Probabilidades de Transição Estimadas. Natal – RN, 2001–2020.

a	s	$\hat{p}(a s)_n$
0	0	0.1137
0	11	0.0772
0	12	0.0540
0	22	0.0382
0	001	0.2500
0	101	0.0000
0	201	0.0000
0	021	0.0833
0	121	0.0720
0	221	0.0728
0	002	0.0000
0	102	0.0833
0	202	0.1268
1	0	0.3128
1	11	0.3125
1	12	0.2178
1	22	0.1992
1	001	0.3750
1	101	0.4118
1	201	0.2683
1	021	0.3889
1	121	0.3840
1	221	0.2524
1	002	0.5714
1	102	0.1944
1	202	0.2958

As maiores probabilidades estão concentradas para quando o próximo símbolo é ‘2’, isto é, entre 8 e 10 horas diárias de insolação. Entre os 13 contextos encontrados, 11 deles possuem probabilidade de transição maior que 50% para o estado ‘2’.

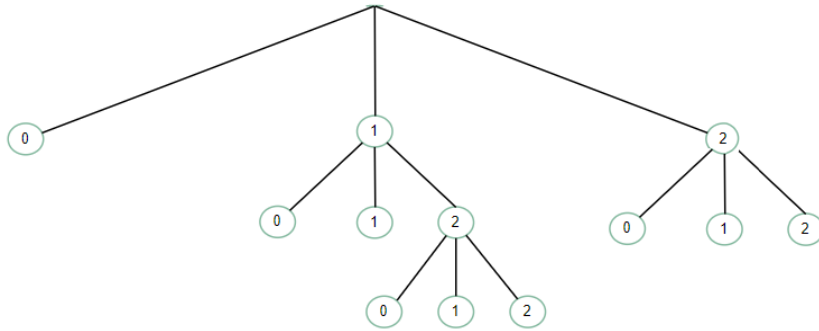
Quando há a configuração de dois dias consecutivos de alta insolação, a probabilidade de que o próximo dia se mantenha no mesmo estado é de 76.26%, o que corresponde também à maior probabilidade de transição estimada para a cidade.

Assim como na capital de Roraima, é possível observar a ocorrência de fenômenos com probabilidade nula. Por exemplo, quando os dias seguem a sequência de média-

baixa–média insolação ou média–baixa–alta insolação, a probabilidade de ocorrência de baixa insolação no próximo símbolo é zero.

Agora, para a região sudeste, na cidade de Vitória, foram encontrados 9 contextos dispostos na árvore apresentada na Figura 3.6.

Figura 3.6: Árvore Probabilística de Contextos Para a Cidade de Vitória dos anos de 2001 a 2020.



As probabilidades de transição associadas à esses contextos estão localizadas na Tabela 3.6.

Tabela 3.6: Probabilidades de Transição Estimadas. Vitória – ES, 2001–2020.

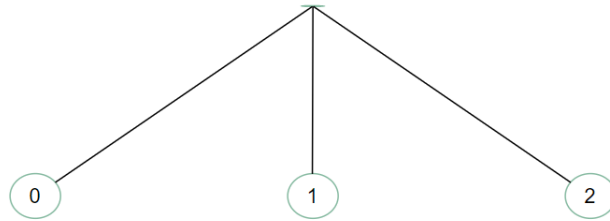
heighta	s	$\hat{p}(a s)_n$
0	0	0.1325
0	01	0.1032
0	11	0.0907
0	02	0.0782
0	12	0.0650
0	22	0.0595
0	021	0.1818
0	121	0.1422
0	221	0.1056
1	0	0.4026
1	01	0.3290
1	11	0.3385
1	02	0.2458
1	12	0.2808
1	22	0.2124
1	021	0.3864
1	121	0.3922
1	221	0.3254

Seguindo o mesmo padrão que a cidade de Natal, as menores probabilidades encontradas se referem ao estado ‘0’ como próximo símbolo, que corresponde ao menor nível de insolação da análise. Em contrapartida, e seguindo o mesmo padrão observado para todas as árvores anteriores, as maiores probabilidades estão relacionadas ao estado ‘2’ como próximo símbolo, chegando a atingir uma probabilidade de 72.81% de sair do contexto ‘22’ para o estado ‘2’.

As probabilidades associadas à insolação média como próximo símbolo estão situadas aproximadamente entre 20% e 41%. Ao observar o comportamento dos contextos para essa categoria, é possível notar que, a maior probabilidade está associada a informação de dia com baixa insolação (40.26%).

A última análise considera os anos de 2001-2020 na cidade de Passo Fundo, na região sul do Brasil.

Figura 3.7: Árvore Probabilística de Contextos Para a Cidade de Passo Fundo dos anos de 2001 a 2020.



A estimação da árvore consiste na forma mais concisa que abrange todos os símbolos do alfabeto. Assim, a árvore representa uma Cadeia de Markov de ordem 1, onde a transição entre os símbolos depende apenas do símbolo anterior.

Tabela 3.7: Probabilidades de Transição Estimadas. Passo Fundo – RS, 2001–2020.

a	s	$\hat{p}(a s)_n$
0	0	0.1383
0	1	0.1119
0	2	0.0902
1	0	0.3285
1	1	0.3677
1	2	0.3111

A transição com a maior probabilidade estimada ocorre quando há a saída do estado ‘2’ para o estado ‘2’, com uma probabilidade de 59.87%. É interessante observar que, para esse caso específico, ainda que alta, a probabilidade diminui em comparação às

probabilidades de alta insolação exploradas anteriormente.

É possível observar também que as menores probabilidades estão associadas ao estado ‘0’ como próximo símbolo. Isso significa que, independentemente das configurações anteriores de níveis de insolação, as chances de o próximo símbolo corresponder a baixa insolação são inferiores a 14%.

Assim, um comportamento frequente observado é que, para valores mais altos de horas de insolação, é mais provável que ocorra a permanência no mesmo estado do que a transição para um estado inferior.

3.3 Estimação do Próximo Símbolo

Após obter as probabilidades de transição dos contextos encontrados em cada cidade, torna-se possível estimar o próximo símbolo.

A seguir são listados os passos utilizados na previsão.

- Observar o passado mais recente x_0, x_{-1}, x_{-2} do processo;
- Com base no contexto estimado e na respectiva probabilidade de transição estimada, simular o próximo símbolo do processo;
- Comparar o símbolo simulado com o símbolo real observado no tempo $t=1$;
- Repetir os itens anteriores para 10 dias à frente dos dados usados na simulação das árvores de contextos.

Dessa forma, foram realizadas estimativas dos próximos dez símbolos para cada cidade, com base no ‘passado mais recente’ em que a proporção de acertos está disposta na Tabela 3.8.

Tabela 3.8: Proporção de Acertos da Estimação do Próximo Símbolo para Cidades no Brasil, anos de 1980 a 2020

	Vitória	Natal	Brasília	Passo Fundo
Proporção	50%	50%	60%	70%

Para a cidade de Boa Vista, não foi possível realizar a estimação do próximo símbolo devido à indisponibilidade de dados. Para as regiões Norte e Nordeste, há uma

maior quantidade de dados nulos, influenciando diretamente em estudos para menores períodos.

Capítulo 4

Conclusão

Neste trabalho modelamos dados de insolação para o desenvolvimento de árvores probabilísticas de contextos com uma determinada profundidade prefixada. Os dados originais passaram por tratamento e categorização, resultando em um alfabeto composto por três símbolos: $A = \{0,1,2\}$, em que 0, 1 e 2 representam baixa, média e alta insolação, respectivamente.

Foi utilizado o Algoritmo Contexto em R, disponível em Bomfim (2015), para estimar árvores em cada região do Brasil, representando diferentes cidades. O maior número de contextos encontrados corresponde a treze enquanto o menor corresponde a três e se iguala a uma Cadeia de Markov de ordem um.

Em seguida, com as probabilidades de transição obtidas, procedeu-se à estimação do próximo símbolo para cada uma das cidades estudadas.

Para valores mais altos de horas de insolação, foi observado que é mais provável que ocorra a permanência no mesmo estado do que a transição para um estado inferior.

Um resultado interessante foi obtido para a cidade de Boa Vista. Verificou-se que a informação de que a insolação é baixa não foi suficiente para prever o próximo símbolo. Isso indica que a sequência '0' não foi observada nesse contexto específico.

A árvore estimada para a cidade de Passo Fundo tem a configuração de uma Cadeia de Markov de ordem 1. Isso significa que a probabilidade de transição para o próximo estado depende apenas do estado atual e não é influenciada pelos estados anteriores.

Em relação a Brasília, foram identificados padrões distintos de insolação para os períodos analisados. Esses padrões evidenciaram níveis de insolação mais altos para o período de 2001-2020, sugerindo uma mudança nas condições climáticas nas últimas

décadas.

Durante o período de anos 2001-2020, foi observado que nenhuma das árvores estimadas para as diferentes cidades apresentava a mesma configuração. Essa constatação evidencia o comportamento distinto de insolação entre as regiões brasileiras, reforçando a ideia de que o clima varia significativamente de uma cidade para outra. Assim, conclui-se que cada cidade possui padrões únicos de insolação, com suas próprias características e tendências específicas.

Na estimação do próximo símbolo, a maior taxa de acerto na previsão, ou seja, a proporção de símbolos estimados corretamente, foi de 70% na cidade de Passo Fundo. Por outro lado, a estimação foi prejudicada por dados indisponíveis para Boa Vista.

Neste sentido, é importante destacar a qualidade dos dados como fator limitante. Para as regiões Norte e Nordeste, há uma maior quantidade de dados nulos, influenciando diretamente em estudos para menores períodos.

Logo, enfatiza-se a possibilidade de extensão deste trabalho para outras cidades, a fim de entender padrões de insolação e a relação com mudanças climáticas em diferentes partes do país.

Referências Bibliográficas

BOMFIM, A. B. A. Estudo de estimadores de árvores de contexto com aplicação em linguística. 2015.

BORGES, R. R. Estudo da umidade relativa do ar em Brasília: uma aplicação de cadeias de ordem variável. 2018.

CAVALCANTE, M. F. M. Modelagem de dados pluviométricos no df por meio de cadeias de ordem variável estocasticamente perturbadas. 2018.

CSISZÁR, I.; TALATA, Z. Context tree estimation for not necessarily finite memory processes, via bic and mdl. In: IEEE. *Proceedings. International Symposium on Information Theory, 2005. ISIT 2005*. [S.l.], 2005. p. 755–759.

INMET. *Instituto Nacional de Meteorologia*. 2023. Disponível em: <https://portal.inmet.gov.br/>.

MATTA, D. H. da et al. *Algoritmos de estimação para Cadeias de Markov de alcance variável: aplicações a detecção do ritmo em textos escritos*. Tese (Doutorado) — Universidade Estadual de Campinas (UNICAMP). Instituto de Matemática, 2018.

MEDEIROS, R. M. de; HOLANDA, R. M. de; FRANÇA, M. V. de. Interpolação da insolação média para o estado do piauí-brasil. *Revista de Geografia (Recife)*, v. 35, n. 5, 2018.

RISSANEN, J. A universal data compression system. *IEEE Transactions on information theory*, IEEE, v. 29, n. 5, p. 656–664, 1983.

ROSS, S. M. et al. *Stochastic processes*. [S.l.]: Wiley New York, 1996. v. 2.

SARAIVA, M. d. F. O. *Astronomia & Astrofísica*. [S.l.]: Editora Livraria da Física, 2004.