Vem Ser Tech - Dados

Séries Temporais

1. Introdução

De maneira geral, chamamos de **séries temporais** uma **sequência de valores indexados pelo tempo**. O conteúdo dos valores varia muito de acordo com o contexto específico de cada série. Alguns exemplos são: preços de ativos financeiros (série de preços); quantidade de clientes que visitam determinada loja; temperatura medida em um local durante um dado intervalo de tempo, etc.

A área de **modelagem de séries temporais** tem como objetivo desenvolver modelos matemáticos que descrevam o comportamento de uma variável aleatório ao longo do tempo. Em geral, os modelos utilizados para descrever séries temporais são **processos estocásticos**, onde o processo estocástico em si compõe uma família de variáveis aleatórias e uma realização ou trajetória do processo estocástico pode ser visto como uma **série temporal**.

Os processos estocásticos podem ser **discretos**, por exemplo, o número de chamadas telefônicas que chega a uma central em cinco horas, ou **contínuo**, por exemplo, o preço de uma ação na bolsa de valores.

Pode-se descrever um processo estocástico por meio da distribuição de probabilidade conjunta de $Z(t_1), \ldots, Z(t_k)$, mas aqui será abordado o assunto de uma forma um pouco mais simples. Portanto utilizando as funções para o caso contínuo, tem-se que:

Média: $\mu(t) = E(Z(t))$

Variância: $\sigma^2(t) = Var(Z(t))$

Alguns exemplos de processos estocásticos serão definidos nos tópicos a seguir, sendo eles:

Sequência Aleatória;
Ruído branco (*White-Noise*);
Passeio Aleatório (*Random Walk*).

1.1 Sequência Aleatória

Considere $\{X_n, n = 1, 2, \ldots\}$ uma sequência de variável aleatória definidas no mesmo espaço amostral Ω. Aqui, $\mathcal{T} = \{1, 2, \ldots\}$ e tem-se um processo com parâmetro discreto, ou uma sequência aleatória. Para todo $n \geq 1$, podemos escrever $P(X_1 = a_1, \ldots, X_n = a_n) = P(X_1 = a_1)P(X_2 = a_2|X_1 = a_1)\ldots P(X_n = a_n|X_1 = a_1, \ldots, X_{n-1} = a_{n-1})$

1.2 Ruído Branco (*White-Noise*)

Pode-se definir um ruído branco como $\{\epsilon_t, t \in \mathbb{Z}\}$ se as variáveis aleatórias ϵ_t são não-correlacionadas, isto é, se $Cov\{\epsilon_t, \epsilon_s\} = 0$ para $t \neq s$.

Esse processo será estacionário (ou seja, a média e a variância não se alteram ao longo do tempo) se $E(\epsilon_t) = \mu_\epsilon$ e $Var(\epsilon_t) = \sigma^2_\epsilon$ para todo t.

1.3 Passeio Aleatório (*Random Walk*)

Considere uma sequência aleatória $\{\epsilon_t, t \geq 1\}$ de v.a. i.i.d (variáveis aleatórias independentes e identicamente distribuídas) $(\mu_\epsilon, \sigma^2_\epsilon)$.

Defina a sequência $X_t = \epsilon_1 + \ldots + \epsilon_t$.

Segue que $E(X_t) = t\mu_\epsilon$ e $Var(X_t) = t\sigma^2_\epsilon$, ou seja, ambos dependem de t.

Esse processo é chamado de passeio aleatório e é claramente não estacionário (média e variância se alteram ao longo do tempo).

2. Definição de Séries Temporais

Uma série temporal pode ser considerado como uma sequência $Z(t_1), Z(t_2), \ldots, Z(t_n)$ observada nos instantes $t_1, \ldots, t_n$, ou seja, uma realização ou trajetória de um processo estocástico. Ao utilizar séries temporais têm-se diversos objetivos como por exemplo investigar o fenômeno gerador da série, descrever o comportamento da série, procurar periodicidades relevantes nos dados ou menos fazer previsões de valores futuros.

As séries temporais podem ser **contínuas**, quando as observações são feitas continuamente no tempo. Nesse caso, $\{Z(t) : t \in T\}, T = \{t : t_1 < t < t_2\}$. Ou podem ser **discretas**, quando as observações são feitas em tempos específicos $\{Z(t) : t \in T\}, T = \{t_1, ..., t_n\}$.

3. Componentes da Série Temporal

Uma série temporal pode ser decomposta nas **componentes** descritas a seguir:

> **Tendência (T)**: indica o seu comportamento ao longo do tempo, ou seja, se é **crescente**, **decrescente** ou **estável**. Além disso, a tendência indica também a velocidade destas mudanças.
>
> **Ciclos (C)**: são **oscilações de subida e de queda nas séries**, de forma **suave e repetida**, ao longo da componente de tendência. Os movimentos cíclicos tendem a ser irregulares.
>
> **Sazonalidade (S)**: são **oscilações de subida e de queda que sempre ocorrem em um determinado período do ano, do mês, da semana ou do dia**. Estes movimentos são facilmente previsíveis, ocorrendo em **intervalos regulares de tempo**.
>
> **Ruído Aleatório (ϵ)**: ou erro no período **t** são variações irregulares ou flutuações inexplicáveis, resultado de fatos fortuitos e inesperados.

Para implementar em *Python* a decomposição de séries temporais, utiliza-se da função *seasonal_decompose* oriunda da biblioteca *statsmodels* (uma das principais bibliotecas a respeito de ferramentas estatísticas):

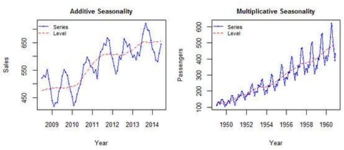
```
# Carregando a função seasonal_decompose
from statsmodels.tsa.seasonal import seasonal_decompose
```

```python
# Aplicando a função seasonal_decompose
results = seasonal_decompose(df,                       # conjunto de dados da série temporal
                             model = 'additive',   # tipo de série (aditiva ou multiplicativa)
                             period = 7)           # período de tempo da série

# Gera os gráficos da decomposição
results.plot()

# Mostrar o gráfico
plt.show()
```

Note que para a implementação da função *seasonal_decompose* é necessário conhecer o **período da série temporal**, ou seja o intervalo de tempo entre duas observações (pode ser segundos, minutos, horas, dias e etc.). Além disso, é importante conhecer o tipo de série a ser analisada, podendo ser classificada em 2 tipos principais:

> **Série Aditiva:** Uma série é considerada **aditiva** a amplitude da sazonalidade mantém-se constante ao longo do tempo;

> **Série Multiplicativa:** Já a série considerada **multiplicativa**, a amplitude da sazonalidade aumenta ao longo do tempo.



Fonte: Medium

Materiais Complementares

Documentação do statsmodels;

Artigo O que são séries temporais e como aplicar em Machine Learning escrito por Paulo Vasconcellos;

Referências

Pedro A. Morettin, Wilton O. Bussab, Estatística Básica, 8ª edição

Peter Bruce, Andrew Bruce & Peter Gedeck, Practical Statistics for Data Scientists, 50+ Essential Concepts Using R and Python, 2ª edition

Ron Larson & Betsy Farber, Estatística Aplicada, 6ª edição.