

Trabalho No.3

28 de outubro de 2024

Redigir de maneira individual e entregar na área correspondente no sistema **Microsoft Teams** um relatório eletrônico até o dia **18 de novembro de 2024**.

- 1- A série de dados Deaths contém o número de mortes acidentais mensais nos EUA entre 1973 e 1978. Esses dados foram usados por diversos autores e podem ser encontrados no arquivo de dados R **USAccDeaths**. Aplique a técnica SSA a esse conjunto de dados para ilustrar a capacidade da técnica de extrair tendência, oscilação, ruído e previsão.

O comprimento da janela L é o único parâmetro no estágio de decomposição. A seleção do comprimento da janela adequado depende do problema em questão e de informações preliminares sobre a série temporal. Os resultados teóricos nos dizem que L deve ser grande o suficiente, mas não maior que $T/2$, T é o comprimento da série. Além disso, se sabemos que a série temporal pode ter um componente periódico com um período inteiro, por exemplo, se esse componente for um componente sazonal, então para obter melhor separabilidade desse componente periódico é aconselhável tomar o comprimento da janela proporcional a esse período. Usando essas recomendações, escolher $L = 24$. Então, com base nesse comprimento de janela e no SVD da matriz de trajetória (24×24), temos 24 autotriplos, ordenados por sua contribuição (participação) na decomposição.

A disponibilidade de informações auxiliares em muitas situações práticas aumenta a capacidade de construir o modelo adequado. Certamente, informações auxiliares sobre a série inicial sempre tornam a situação mais clara e ajudam na escolha dos parâmetros dos modelos. Essas informações não apenas podem nos ajudar a selecionar o grupo adequado, mas também são úteis para previsão e detecção de ponto de mudança com base na técnica SSA. Por exemplo, a suposição de que há uma periodicidade anual na série Death sugere que devemos prestar atenção à frequência $k/12$, $k = 1, \dots, 12$. Obviamente, podemos usar as informações auxiliares para selecionar o comprimento de janela adequado também.

- (a) Mostre o gráfico dos logaritmos dos 24 valores singulares da série Death.
Cinco pares evidentes com valores singulares principais quase iguais correspondem a cinco componentes (quase) harmônicos da série: os pares autotriplos 2-3, 4-5, 7-8, 9-10 e 11-12 devem estar relacionados a harmônicos com períodos específicos.
- (b) A mostre a tendência extraída no fundo da série original que é obtida do primeiro autotriplo.

Note que podemos construir uma aproximação mais complicada da tendência se usarmos alguns outros autotriplos, este seria o caso de utilizarmos o primeiro e sexto

autotriplos. No entanto, a precisão que ganharíamos seria muito pequena e o modelo da tendência se tornaria muito mais complicado.

Tendência é o componente de variação lenta de uma série temporal que não contém componentes oscilatórios. Suponha que a série temporal em si seja um componente desse tipo. A prática mostra que, neste caso, um ou mais dos autovetores principais também variarão lentamente. Sabemos que os autovetores têm (em geral) a mesma forma que os componentes correspondentes da série temporal inicial. Portanto, devemos encontrar autovetores de variação lenta. No nosso caso, o autovetor líder é definitivamente da forma necessária.

- (c) Identificação harmônica: mostre o gráfico da extração da oscilação obtida pelos autotriplos de 2 a 12.

Ao comparar a figura obtida desta maneira com os dados originais, fica claro que os autotriplos selecionados para identificar os componentes harmônicos foram feitos corretamente. Como um trabalho alternativo, mostre a figura da oscilação de nossa série obtida pelos autotriplos 2–5 e 7–12. Neste caso, consideramos o sexto autotriplo como um componente de tendência. Parece não haver grande discrepância entre a seleção do sexto autotriplo nos componentes de tendência ou oscilação, como aparece nas figuras aqui solicitadas.

O problema geral aqui é a identificação e separação dos componentes oscilatórios da série que não constituem partes da tendência. A declaração do problema em SSA é especificada principalmente pela natureza livre de modelo do método. A escolha $L = 24$ nos permite extrair simultaneamente todos os componentes sazonais: 12, 6, 4, 3 e 2.5 meses, bem como a tendência.

- (d) Previsão: Os valores para os primeiros seis meses de 1979 são 7798 7406 8363 8460 9217 9316.

Utilize os autotriplos identificados, ou seja, os autotriplos entre 1 e 12, para fazer a previsão dos seis primeiros meses de 1979. Mostre a série reconstruída, que é obtida dos autotriplos 1-12. Observe que a série original e a série reconstruída estão próximas, indicando que os valores previstos são razoavelmente precisos.

- (e) Utilize o erro absoluto médio (MAE), que é uma medida de erros entre observações pareadas que expressam o mesmo fenômeno. O MAE é calculado como a soma dos erros absolutos, ou seja, a distância de Manhattan dividida pelo tamanho da amostra:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^T |y_i - \hat{y}_i|,$$

onde y_i são os valores observados da série, \hat{y}_i os valores estimados e T é o número de pontos estimados.

Calcule também o erro percentual absoluto médio (MAPE), também conhecido como desvio percentual absoluto médio (MAPD), é uma medida de precisão de previsão de um método de previsão em estatística. Ele geralmente expressa a precisão como uma razão definida pela fórmula:

$$\text{MAPE} = 100 \times \frac{1}{n} \sum_{i=1}^T \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

onde y_i é o valor real da série e \hat{y}_i é o valor previsto. A diferença deles é dividida pelo valor real y_i . O valor absoluto dessa razão é somado para cada ponto previsto no tempo e dividido pelo número de pontos previstos T .

Análise espectral e filtragem

2- Exercício 9

3- Exercício 36

Observação: em ambos exercícios o conjunto de dados está disponível no pacote **astsa**.