

Ponto 9 (Teórica) - Análise de Séries Temporais

Ben Dêivide

6 de outubro de 2021

Grande parte dos modelos clássicos, pressupõe que as observações sejam independentes ou ao menos que a covariância entre as observações sejam iguais a zero. Contudo, quando estas observações são obtidas ao longo do tempo, a pressuposição de independência não é satisfeita, e assim não haverá garantias, por exemplo, que a previsão das observações com base no modelo clássico seja confiável.

Assim, precisamos de metodologias que levam em consideração a dependência entre as observações. Uma das metodologias baseadas na modelagem da correlação entre as observações é a análises de séries temporais.

Definição 1 (Série temporal). *Uma série temporal é uma coleção de observações ordenadas no tempo.*

Exemplos de séries temporais podem ser encontrados na Economia (preços diários de ações, taxa de desemprego), Meteorologia (Precipitação, temperatura), Medicina (Eletrocardiograma, Eletroencefalograma), entre outras áreas.

É com base nessa série que iremos modelar a dependência entre as observações ao longo do tempo. Vale ressaltar que o tempo pode ser substituído por outra variável como espaço, profundidade, etc.. Com isso, esse tipo de análise requer técnicas específicas.

Podemos classificar uma série temporal como discreta ou contínua.

Definição 2 (Série temporal contínua). *Uma série temporal é contínua quando as observações são feitas continuamente no tempo. Definindo o conjunto $T = \{t : t_1 < t < t_2\}$, então a série temporal será denotada por $\{Z(t) : t \in T\}$.*

Um exemplo de série temporal contínua é o registro de maré no porto.

Definição 3 (Série temporal discreta). *Uma série temporal é discreta quando as observações são feitas em tempos específicos. Definindo o conjunto $T = \{t_1, t_2, \dots, t_n\}$, então a série temporal será denotada por $\{Z_t : t \in T\}$.*

Um exemplo de série temporal discreta é o índice pluviométrico mensal em Pau dos Ferros/RN.

De um modo geral, dizemos que os objetivos de estudar séries temporais são:

- i) **DESCRIÇÃO:** Descrever propriedades da série, tendência, sazonalidade, observações discrepantes, etc..
- ii) **EXPLICAÇÃO:** Usar a variação de uma série para explicar a variação em outra série.

- iii) PREDIÇÃO: Predizer valores futuros com base em valores passados.
- iv) CONTROLE: Os valores da série temporal medem a “qualidade” de um processo de manufatura e o objetivo é o controle do processo. Um exemplo é o controle estatístico de qualidade aonde as observações são representadas em cartas de controle.

Uma segunda definição mais formal, dizemos

Definição 4 (Série temporal). *Uma série temporal é uma realização de um processo estocástico.*

Definição 5 (Processo estocástico). *Um processo estocástico é uma família $Z = \{Z(t), t \in T\}$ tal que para cada t , $Z(t)$ é uma variável aleatória.*

Mais rigorosamente, $Z(t)$ é uma função de dois argumentos, $Z(t, \omega)$, para $t \in T$ e $\omega \in \Omega$. Para cada $t \in T$ fixado, $Z(t, \omega) \equiv Z(\omega)$ é uma variável aleatória definida sobre o espaço amostral Ω . Para cada $\omega \in \Omega$ fixado, $Z(t, \omega) = Z(t)$, é uma função de t , ou seja, uma realização ou trajetória do processo, ou ainda, uma série temporal.

Como o processo estocástico $\{Z(t), t \in T\}$, é uma variável aleatória, então torna-se bastante interessante obter a distribuição conjunta de $\{Z(t_1), \dots, Z(t_k)\}$. O problema é que na prática só dispomos de uma única realização do processo. Assim, para que possamos obter a distribuição conjunta, certas restrições serão impostas.

No caso dos processos estocásticos, as restrições que devem ser impostas são de dois tipos:

- a) Heterogeneidade temporal; e
- b) Memória do processo.

Assumiremos que a distribuição conjunta é invariante por translações. Ou seja, dado um conjunto $T = \{t : 1 \leq t_1 \leq t_2 \leq \dots \leq t_k\}$, a restrição imposta à heterogeneidade temporal, em nosso caso, é que as distribuições conjuntas de $\{Z(t_{1+\tau}), Z(t_{2+\tau}), \dots, Z(t_{k+\tau})\}$ e $\{Z(t_1), Z(t_2), \dots, Z(t_k)\}$, são idênticas, para qualquer inteiro $\tau \geq 1$.

Com relação a memória do processo estocástico, a primeira ideia é fazer com que o processo não tenha memória, isto é, seja não correlacionado ou independente. A restrição de independência é muito forte e pouco plausível, uma vez que séries temporais, geralmente, apresentam algum tipo de dependência temporal. Assim, uma forma regular de abrandar essa suposição é fazer com que para instantes de tempo muito afastados não exista correlação. Dessa forma, o processo tem memória, mas tal memória vai diminuindo com o aumento dos intervalos entre os instantes de tempo.

Uma classe de processos estocásticos para o estudo de séries temporais é o processo estacionário.

Definição 6 (Processo Estacionário). *Um processo estocástico $\{Z(t), t \in T\}$, com segundo momento finito $E[Z(t)^2] < \infty$, diz-se estacionário (ou fracamente estacionário) se, e somente se,*

- i) $E[Z(t)] = \mu(t) = \mu, \forall t \in T$. (Média constante, não depende do tempo)
- ii) $Var[Z(t)] = E[(Z(t) - \mu)^2] = \sigma^2, \quad \forall t \in T$. (Variância constante)
- iii) $\gamma(\tau) = Cov[Z(t), Z(t + \tau)] = E[(Z(t) - \mu)(Z(t + \tau) - \mu)]$, para qualquer t e qualquer $\tau \geq 1$, depende somente de τ , não de t .

Chamamos a distância $(t + \tau) - t = \tau$ é chamado de defasagem ou “lag”, e na Definição 6, item (iii), chamamos $\gamma(\tau)$ de coeficiente de autocovariância (facv) na defasagem τ . As propriedades de facv seguem:

1. $\gamma(0) = \text{Var}(Z_t) > 0$;
2. $\gamma(-\tau) = \gamma(\tau)$;
3. $|\gamma(\tau)| \leq \gamma(0)$;
4. $\gamma(\tau)$ é positiva definida, no sentido que

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j \gamma(\tau_i - \tau_j) \geq 0, \quad a_1, \dots, a_n, \tau_1, \dots, \tau_n \in \mathbb{N}.$$

Note que o tamanho de $\gamma(\tau)$ depende da escala em que $Z(t)$ é medida. Portanto, para efeito de interpretação, é mais útil padronizar a função de autocovariância dando origem a uma função de autocorrelação

$$\rho(\tau) = \gamma(\tau) / \gamma(0), \quad (1)$$

que mede a correlação entre $Z(t)$ e $Z(t + \tau)$.

A função de autocorrelação (fac) satisfaz às seguintes propriedades:

- i) $\rho(0) = 1$;
- ii) $\rho(-\tau) = \rho(\tau)$;
- iii) $|\rho(\tau)| \leq \rho(0) = 1$;
- iv) $\rho(\tau)$ é positiva definida.

Por meio do coeficiente de autocorrelação teórico, $\rho(\tau)$, iremos identificar as propriedades de uma série temporal. Para isso, será utilizado o estimador de $\rho(\tau)$, denotado por r , que é o coeficiente de autocorrelação amostral. A ideia é similar ao coeficiente de correlação usual. Para n pares de observações das variáveis x e y o coeficiente de correlação amostral é dado por:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}. \quad (2)$$

Agora na série temporal, queremos medir a correlação entre as observações de uma mesma variável em diferentes horizontes de tempo, isto é, correlações defasadas 1, 2, ... períodos de tempo. Dessa forma, dadas n observações de uma série temporal discreta z_1, z_2, \dots, z_n , podemos formar os pares $(z_1, z_2), \dots, (z_{n-1}, z_n)$. Considerando z_1, \dots, z_n e z_2, \dots, z_n como duas variáveis, o coeficiente de correlação entre elas é dada por:

$$r_1 = \frac{\sum_{t=1}^{n-1} (z_t - \bar{z}_1)(z_{t+1} - \bar{z}_2)}{\sqrt{\sum_{t=1}^{n-1} (z_t - \bar{z}_1)^2 \sum_{t=1}^{n-1} (z_{t+1} - \bar{z}_2)^2}}, \quad (3)$$

sendo

$$\bar{z}_1 = \sum_{t=1}^{n-1} z_t / (n-1) \text{ e } \bar{z}_2 = \sum_{t=2}^n z_t / (n-1).$$

Como o coeficiente r_1 mede as correlações entre observações sucessivas, este é chamado de coeficiente de autocorrelação ou coeficiente de correlação serial.

É usual simplificar a equação (3) utilizando-se a média de todas as observações, isto é, $\bar{z} = \sum_{t=1}^n z_t / n$ já que $\bar{z}_1 \approx \bar{z}_2$, e assumindo variância constante. Assim, a versão simplificada de (3) fica

$$r_1 = \frac{\sum_{t=1}^{n-1} (z_t - \bar{z})(z_{t+1} - \bar{z})}{(n-1) \frac{\sum_{t=1}^n (z_t - \bar{z})^2}{n}}. \quad (4)$$

Alguns autores ainda retiram o termo $(n-1)/n$ que é próximo de 1 para n não muito pequeno, resultando em

$$r_1 = \frac{\sum_{t=1}^{n-1} (z_t - \bar{z})(z_{t+1} - \bar{z})}{\sum_{t=1}^n (z_t - \bar{z})^2}. \quad (5)$$

Para uma defasagem de τ períodos de tempo, podemos calcular o coeficiente de autocorrelação estimado por

$$r_\tau = \frac{\sum_{t=1}^{n-\tau} (z_t - \bar{z})(z_{t+\tau} - \bar{z})}{\sum_{t=1}^n (z_t - \bar{z})^2}. \quad (6)$$

Morettin usa a estimativa da autocovariância $\gamma(\tau)$ como

$$c_\tau = \frac{1}{n} \sum_{t=1}^{n-\tau} (z_t - \bar{z})(z_{t+\tau} - \bar{z}), \quad \tau = 1, 2, \dots, n-1. \quad (7)$$

A estimativa da autocorrelação $\rho(\tau)$ é dada por:

$$r_\tau = \frac{\gamma(\tau)}{\gamma(0)}. \quad (8)$$

Mais a frente, detalharemos a forma de como a função de autocorrelação irá indicar as propriedades de uma série temporal.

Um modelo clássico para séries temporais supõe que a série $\{Z_1, Z_2, \dots, Z_n\}$ pode ser escrita como:

$$Z_t = T_t + S_t + a_t, \quad t = 1, 2, \dots, n, \quad (9)$$

sendo Z_t uma observação no tempo t de uma série temporal discreta, T_t a componente tendência no instante t , S_t a componente sazonalidade, e a_t é a componente aleatória de

média zero, $E[a_t] = 0$, e variância constante $Var[a_t] = E[a_t^2] = \sigma_a^2$. Dizemos ainda que a_t é um ruído branco se $Cov(a_t, a_j) = E[a_t a_j] = 0$, para $t \neq j$.

O modelo (9) é dito aditivo, pois a componente sazonal é independente das outras componentes. Se as amplitudes sazonais variam de acordo com a tendência, o modelo mais adequado é o multiplicativo,

$$Z_t = T_t S_t a_t, \quad t = 1, 2, \dots, n. \quad (10)$$

Nesse caso, esse modelo é usado para descrever a dependência das amplitudes sazonais em relação à tendência.

Não temos uma definição precisa sobre tendência, mas a entendemos como o aumento ou diminuição gradual das observações ao longo do tempo. A sazonalidade mostra flutuações ocorridas em períodos menores ou iguais a dozes meses. E a componente aleatória é a parte não explicada, e espera-se que ela seja puramente aleatória.

Na prática, o que observamos nas séries temporais são características como as apresentadas nas componentes do modelo (9). Com essas características, dizemos que uma série desse tipo é não-estacionária. Uma série é estacionária se ela desenvolve no tempo ao longo de uma média constante, refletindo alguma forma de equilíbrio estável, isto é, a série apresenta as características de um processo estacionário, Definição 6. Assim, diríamos que uma série que não apresentasse tendência nem sazonalidade, por exemplo, seria uma série estacionária.

Alguns modelos como os de Box-Jenkins exigem que a série seja estacionária. Assim, precisamos modelar estas componentes e encontrar mecanismos para eliminá-la da série.

Pela limitação do tempo, suponha ausência de sazonalidade, e o modelo (9) pode ser reescrito como

$$Z_t = T_t + a_t, \quad t = 1, 2, \dots, n, \quad (11)$$

em que a_t é um ruído branco. Existem vários métodos para estimar a componente tendência, dentre eles, ajuste polinomial e suavização (filtragem).

Um primeiro filtro para eliminar a tendência é através de filtros lineares. Um filtro linear converte uma série $\{Z_t\}$ em outra $\{Y_t\}$ através da seguinte operação linear

$$\hat{T}_t = \sum_{j=-q}^s \alpha_j Z_{t+j}, \quad (12)$$

onde $\{\alpha_j\}$ é um conjunto de pesos. Além disso, como queremos estimar a média local os pesos devem ser tais que $\sum_{j=-q}^s \alpha_j = 1$, garantindo assim que $\min\{Z_t\} < \hat{T}_t < \max\{Z_t\}$. Neste caso a operação é chamado média móvel.

Um outro filtro muito utilizado nas séries temporais para eliminar a tendência, considerando o modelo (11), é tomar diferenças sucessivas na série original até que a série esteja livre do componente tendência. Seja a série Z_t , a primeira diferença é definida por:

$$\hat{T}_t = Z_t - Z_{t-1} = \Delta Z_t.$$

Note que isto nada mais é do que um filtro (assimétrico) com coeficientes 1 e -1. A segunda diferença é definida por:

$$\hat{T}_t = \Delta(Z_t - Z_{t-1}) = Z_t - 2Z_{t-1} + Z_{t-2} = \Delta^2 Z_t.$$

A n -ésima diferença de Z_t é

$$\hat{T}_t = \Delta[\Delta^{n-1}Z_t].$$

Ao final de qualquer uma das filtragens ou ajuste polinomial (apenas comentado), temos uma estimativa da tendência por meio de \hat{T}_t , e realizando o seguinte procedimento

$$Y_t = Z_t - \hat{T}_t,$$

temos uma série ajustada livre de tendência.

Para verificar se a série ajustada está livre de tendência, podemos inicialmente visualizar por meio de gráfico, e padrões característicos dão indícios que a série tem tendência. Contudo, a comprovação será por meio de teste de hipóteses. Existem diversos testes na literatura, tais como o teste de sequências (Wald-Wolfowitz), teste baseado no coeficiente de correlação de Spearman, teste do sinal (Cox-Stuart). Iremos apresentar este último.

Considerando as observações agrupadas aos pares $(Z_1, Z_{1+c}), (Z_2, Z_{2+c}), \dots, (Z_{n-c}, Z_n)$, em que $c = n/2$ se for par, e $c = (n+1)/2$ se for ímpar. A cada par (Z_i, Z_{i+c}) é atribuído o sinal “+” se $Z_i < Z_{i+c}$ e “-” se $Z_i > Z_{i+c}$. Os empates serão eliminados e n_t é o número de pares em que $Z_i \neq Z_{i+c}$. Testamos a hipótese:

$$\begin{cases} H_0 : P(Z_i < Z_{i+c}) = P(Z_i > Z_{i+c}), \forall i : & \text{Não existe tendência} \\ H_1 : P(Z_i < Z_{i+c}) \neq P(Z_i > Z_{i+c}), \forall i : & \text{Existe tendência} \end{cases} \quad (13)$$

Definindo a estatística do teste por W como sendo o número de sinais positivos, e n_t o número total de comparações, temos que W é uma variável aleatória com distribuição binomial. Sob H_0 , temos que os parâmetros de W é n_t e $p = 1/2$, isto é, para que não haja tendência na série, esperamos que a proporção de sinais positivos esteja próximo de $1/2$. Assim, avaliaremos a hipótese H_0 por meio do valor-p da estatística do teste, isto é,

$$\text{Valor-p} = P(W \geq w) = 2 \times \sum_{i=w}^{n_t} \binom{n_t}{i} (1/2)^i (1/2)^{n_t-i}. \quad (14)$$

O valor 2 é devido ao teste ser bilateral. Assim, ao nível de significância de α , rejeitaremos a hipótese H_0 se valor-p $< \alpha$, indicando que existe o componente tendência para a série temporal.

A sazonalidade (ou periodicidade), assim como a tendência, constitui uma forma de não-estacionaridade e deve ser estimada e também retirada da série. A componente sazonal capta características da série que ocorrem regularmente dentro do período de um ano, isto é, os fenômenos sazonais ocorrem regularmente em período de no máximo doze meses.

Os procedimentos mais comuns para se estimar a sazonalidade são o método de regressão (sazonalidade determinística) e o método de médias móveis (sazonalidade estocástica).

Estimada a sazonalidade \hat{S}_t , a série pode ser escrita livre da componente sazonal. Se o modelo da série for aditivo, tem-se

$$Y_t = Z_t - \hat{S}_t, \quad t = 1, 2, \dots, n.$$

Se o modelo for aditivo, tem-se

$$Y_t = Z_t / \hat{S}_t, \quad t = 1, 2, \dots, N.$$

A componente sazonal também pode ser constatada pela análise visual do gráfico da série. Para se confirmar a existência desta componente, aplica-se um teste de sazonalidade. Os testes mais conhecidos são: teste de Friedman, Kruskal-Wallis, Teste F para análise de variâncias e teste de Fisher. Iremos retratar este último teste.

O teste de Fisher se inicia com o gráfico do periodograma que testa a sazonalidade determinística. Esse periodograma é baseado numa função chamado de periodograma do processo estacionário (a_t) que é definido por

$$I_p(f_i) = \frac{2}{n} \left[\left(\sum_{t=1}^n a_t \cos \frac{2\pi i}{n} t \right)^2 + \left(\sum_{t=1}^n a_t \sin \frac{2\pi i}{n} t \right)^2 \right], \quad (15)$$

com $0 < f_i < 1/2$ e $t = 1, 2, \dots, n$, em que $I_p(f_i)$ indica a intensidade da frequência $f_i = i/n$, o que indica periodicidade de período $1/f_i$. Toda periodicidade acima de 12 meses é considerada ciclo.

No gráfico do periodograma, a frequência f_i é representada no eixo das ordenadas e a intensidade da frequência $I(f_i)$ no eixo das abscissas. Geralmente, o pico de maior intensidade é o componente periódico. Caso haja mais de um pico, aplica-se o teste de Fisher, para verificar se o pico é um componente periódico genuíno. O procedimento para aplicar o teste segue os seguintes passos:

1. Traça-se o periodograma;
2. Toma-se a maior periodicidade encontrada no periodograma, $\max(I_p)$ e calcula-se a estatística:

$$g = \frac{\max(I_p)}{\sum_{p=1}^{n/2} I_p};$$

3. Calcula-se a estatística do teste de Fisher, z_α , dada por:

$$z_\alpha = \left(\frac{\alpha}{n} \right)^{1/(n-1)} + 1,$$

em que α é o nível de significância adotado.

4. Avaliam-se as hipóteses:

$$\begin{cases} H_0 : \text{Não existe periodicidade} \\ H_1 : \text{Existe periodicidade,} \end{cases}$$

com a seguinte decisão: se $g > z_\alpha$, rejeita-se H_0 .

Dentre diversos tipos de modelos que podem ser utilizados na análise de séries temporais não estacionárias, nos restringiremos aos modelos de Box-Jenkins, que utilizam análise de modelos paramétricos. Esses modelos consistem em ajustar modelos autoregressivos integrados de médias móveis ARIMA a um conjunto de dados, sendo estes generalizados com a inclusão do componente sazonal, que são os modelos sazonais

auto-regressivos de médias móveis (SARIMA), do qual este último não será detalhado nessa dissertação devido ao tempo.

A estratégia para a construção do modelo é baseada em um ciclo iterativo, no qual a estrutura do modelo é baseado nos próprios dados. Os estágios do ciclo iterativo são:

1. Uma classe geral de modelos é considerada para análise (especificação);
2. Há identificação de um modelo, com base na análise de autocorrelações, autocorrelações parciais e outros critérios;
3. Posteriormente, a estimação, na qual os parâmetros dos modelos identificados são estimados;
4. E finalmente, a verificação do modelo ajustado, através de uma análise de resíduos, para saber se este modelo é adequado ou não.

Caso o modelo não seja adequado o ciclo é repetido, voltando-se à fase de identificação. Essa é a fase crítica do ciclo, uma vez que vários pesquisadores podem identificar modelos diferentes para uma mesma série temporal.

As vantagens desses modelos são por serem parcimoniosos, isto é, contém um número pequeno de parâmetros, e as previsões são bastante precisas, comparando-as com os demais métodos de previsão.

O modelo ARIMA considera a tendência da série temporal, de ordem (p, d, q) e pode ser representado por:

$$\phi(B)(1 - B)^d Z_t = \theta(B)a_t, \quad (16)$$

sendo:

- $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ o operador autorregressivo de ordem p ;
- $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ o operador de médias móveis de ordem q ;
- B é o operador de translação para o passado, tal que $B^m Z_t = Z_{t-m}$;
- d é o número de diferenças necessárias para retirar a tendência da série e transformá-la em estacionária, e a_t é o ruído branco.
- Assumimos que $Z_t = Z_t^* - \mu$, sendo Z_t^* é a série original e μ é o nível da série.

A modelagem baseada em (16) é usada para séries não-estacionárias cujo comportamento é não explosivo, e que, tomando um número finito de diferenças, tornam-se estacionárias, são chamadas de séries não-estacionárias homogêneas.

Quando a série é estacionária, assumimos $d = 0$ e a expressão (16), reduz-se ao modelo ARMA de ordem (p, q) , definido por:

$$\phi(B)Z_t = \theta(B)a_t. \quad (17)$$

Se $q = 0$ o modelo ARMA($p, 0$), reduz-se a um modelo autorregressivo AR(p), definido por:

$$\phi(B)Z_t = a_t. \quad (18)$$

As condições de estacionariedade e invertibilidade de (17) são, respectivamente, que as raízes dos polinômios $\phi(B) = 0$ e $\theta(B) = 0$ estejam fora do círculo unitário.

Se $p = 0$ o modelo $ARMA(0, q)$, reduz-se a um modelo de médias móveis $MA(q)$, definido por:

$$Z_t = \theta(B)a_t. \quad (19)$$

De forma geral, quando faz-se referência a modelos do tipo $ARIMA$, estes são ajustados à série original. Já ao fazer-se referência a modelos $ARMA$, considera-se uma série diferenciada ou livre de tendência.

Em (16), temos que $(1 - B)^d Z_t$ é estacionária. Tomando $W_t = (1 - B)^d Z_t$, e substituindo nessa mesma expressão, temos

$$\phi(B)W_t = \theta(B)a_t, \quad (20)$$

que pode ser representado por um modelo $ARMA(p, q)$.

Assim, a partir de agora em diante veremos que a diferença básica entre a regressão clássica e os modelos de séries temporais é que nos modelos de séries temporais $ARMA$ (ou $ARIMA$) não se pode assumir independência entre observações. Ao contrário, os modelos autoregressivos e de médias móveis vão modelar o grau de autocorrelação entre desvios e observações defasadas.

Expressando (16) sem a notação do operador de translação para o passado B , temos

$$Z_t = \phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}. \quad (21)$$

Multiplicando ambos os membros por $Z_{t-\tau}$ e tomando-se a esperança desses, obtém-se a função de autocovariância (facv) de lag τ , dada por

$$E[Z_t Z_{t-\tau}] = E[(\phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}) Z_{t-\tau}]$$

$$\gamma(\tau) = \phi_1 \gamma(\tau - 1) + \dots + \phi_p \gamma(\tau - p) + \gamma_{za}(\tau) - \theta_1 \gamma_{za}(\tau - 1) - \dots - \theta_q \gamma_{za}(\tau - q),$$

sendo que $\gamma(\tau)$ foi expresso na Definição 6, e γ_{za} é a covariância entre a_{t-i} , para $i = 0, 1, \dots, q$, e $Z_{t-\tau}$.

Para o caso

$$\gamma_{za}(\tau) = E[a_t Z_{t-\tau}] \begin{cases} = 0 & \text{se } \tau > 0, \\ \neq 0 & \text{se } \tau \leq 0. \end{cases}$$

Logo, para $\tau > q$,

$$\gamma(\tau) = \phi_1 \gamma(\tau - 1) + \dots + \phi_p \gamma(\tau - p), \quad (22)$$

se comportando como um modelo autorregressivo puro. Para as demais autocovariâncias entre a_{t-j} , para $j = 1, \dots, q$, e $Z_{t-\tau}$ quando $\tau < q$, as autocovariâncias serão afetadas diretamente pelos parâmetros de médias móveis.

Dividindo (22) por $\gamma(0)$, obtém-se a função de autocorrelação:

$$\rho(\tau) = \phi_1 \rho(\tau - 1) + \dots + \phi_p \rho(\tau - p). \quad (23)$$

As características da função de autocorrelação (fac) para os modelos $AR(p)$, $MA(q)$ e $ARMA(p, q)$ são:

- i) Um processo $AR(p)$ tem fac que decai de acordo com exponenciais e/ou senóides amortecidas, infinita em extensão;
- ii) Um processo $MA(q)$ tem fac finita, no sentido que ela apresenta um corte após o “lag” q ;
- iii) Um processo $ARMA(p, q)$ tem fac infinita em extensão, a qual decai de acordo com exponenciais e/ou senóides amortecidas após “lag” $q - p$.

Uma outra ferramenta para identificação de modelos ARIMA é a função de autocorrelação parcial (facp). Vamos denotar por ϕ_{kj} o j -ésimo coeficiente de um modelo $AR(k)$, de tal modo que ϕ_{kk} seja o último coeficiente. Sabemos que

$$\rho(j) = \phi_{k1}\rho(j-1) + \phi_{k2}\rho(j-2) + \dots + \phi_{kk}\rho(j-k), \quad j = 1, 2, \dots, k,$$

a partir das quais obtemos as equações de Yule-Walker

$$\begin{bmatrix} 1 & \rho(1) & \rho(2) & \dots & \rho(k-1) \\ \rho(1) & 1 & \rho(1) & \dots & \rho(k-2) \\ \vdots & & & & \\ \rho(k-1) & \rho(k-2) & \rho(k-3) & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(k) \end{bmatrix}.$$

Resolvendo estas equações sucessivamente para $k = 1, 2, \dots$, obtemos

$$\begin{aligned} \phi_{11} &= \phi(1) \\ \phi_{22} &= \frac{\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & \rho(2) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{vmatrix}} \\ \phi_{33} &= \frac{\begin{vmatrix} 1 & \rho(1) & \rho(1) \\ \rho(1) & 1 & \rho(2) \\ \rho(2) & \rho(1) & \rho(3) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) & \rho(2) \\ \rho(1) & 1 & \rho(1) \\ \rho(2) & \rho(1) & 1 \end{vmatrix}} \end{aligned}$$

e em geral,

$$\phi_{kk} = \frac{|\mathbf{P}_k^*|}{|\mathbf{P}_k|},$$

em que \mathbf{P}_k é a matriz de autocorrelações, e \mathbf{P}_k^* é a matriz \mathbf{P}_k com a última coluna substituída pelo vetor de autocorrelações. A quantidade ϕ_{kk} , encarada como função de k , é a chamada função de autocorrelação parcial. O estimador de ϕ_{kk} , denotado por $\hat{\phi}_{kk}$ é obtido substituindo a correlação $\rho(j)$ pela correlação amostral r_j expressas em (8).

Pode-se demonstrar que para os modelos ARMA, temos:

- i) Um processo $AR(p)$ tem facp $\phi_{kk} \neq 0$, para $k < p$ e $\phi_{kk} = 0$, para $k > p$;
- ii) Um processo $MA(q)$ tem facp que se comporta de maneira similar à fac de um processo $AR(p)$: é denominada por exponenciais e/ou senóides amortecidos;

iii) Um processo $ARMA(p, q)$ tem facp que se comporta como facp de um processo MA puro.

Na prática, sob um processo estacionário normal e sob a hipótese $\rho = 0$, dizemos que para o número de observações da série (n) muito grande, os valores de r_j tem distribuição normal com média zero e variância $1/n$. E sob a hipótese que o processo é $AR(p)$ e n suficientemente grande, $\hat{\phi}_{kk}$ também apresenta distribuição normal com média zero e variância $1/n$. Assim, dizemos que r_j é significativo se:

$$|r_j| > 2\frac{1}{\sqrt{n}}, \quad j > q$$

e que $\hat{\phi}_{kk}$ é significativo se:

$$|\hat{\phi}_{kk}| > 2\frac{1}{\sqrt{n}}, \quad k > p.$$

Com base nisso, a fase de identificação do modelo $ARIMA(p, d, q)$ consiste em identificar os valores de p , d e q , além de estimativas preliminares dos parâmetros a serem usadas no estágio de estimação.

A identificação consiste em três partes:

- (a) Verificar se existe a necessidade de transformação na série original, com o objetivo de estabilizar a variância. Podemos utilizar o gráfico da amplitude versus média para verificar a necessidade de transformação na série;
- (b) Tomar diferenças na série obtida em (a) tantas vezes necessária for para obter uma série estacionária, de modo que o processo $(1 - B)^d Z_t$ seja reduzido a um processo $ARMA(p, q)$;
- (c) Identificar o processo $ARMA(p, q)$ resultante, através da análise da função de autocorrelação e autocorrelação parcial.
- (d) Fazer a análise de resíduos para verificar se é ruído branco.

Depois de identificado o modelo $ARIMA(p, d, q)$, o passo seguinte é a estimação dos parâmetros do modelo. O vetor de parâmetros a ser estimado é $\xi = [\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \sigma_a^2]'$. Estamos supondo que a constante do modelo $\mu = 0$, quando $d > 0$, logo temos $p + q + 1$ parâmetros analisados sob a suposição que $a_t \sim N(0, \sigma_a^2)$. Caso contrário, μ sendo incluído, teríamos mais um parâmetro a ser estimado, sendo portanto $p + q + 2$ parâmetros.

Um dos métodos para estimar o vetor de parâmetros ξ é o método da máxima verossimilhança. Considerando n observações Z_1, Z_2, \dots, Z_n , temos a função de verossimilhança dada por $L(\xi, Z_1, \dots, Z_n)$ dada por $L(\xi|Z_1, \dots, Z_n)$ como função de ξ . Os estimadores de máxima verossimilhança (EMV) de ξ serão os valores que maximizam L ou $\ell = \log(L)$. Nestas condições, os EMV serão aproximadamente estimadores de mínimos quadrados (EQM).

Após a fase de estimação, considerando que o processo é estacionário normal, o modelo estimado de (20) é adequado se os resíduos são independentes e normalmente distribuídos (ruído branco), isto é $a_t \sim N(0, \sigma_a^2)$, em que o valor estimado do resíduo \hat{a} é dado por:

$$\hat{a}_t = \hat{\theta}^{-1}(B)\hat{\phi}(B)W_t. \quad (24)$$

Espera-se que \hat{a}_t esteja próximo do seu valor real a_t , e aproximadamente não correlacionado. Considerando $r_k(\hat{a}_t)$ a estimativa da autocorrelação, dada por:

$$r_k(\hat{a}_t) = \frac{\sum_{t=k+1}^n \hat{a}_t \hat{a}_{t-k}}{\sum_{t=1}^n \hat{a}_t^2}. \quad (25)$$

Sob a suposição de que o modelo é adequado, tem-se que $r_k(\hat{a}_t) \sim N(0, 1/n)$. Pode-se dizer que o modelo é adequado quando $r_k(\hat{a}_t)$ estiver dentro dos limites $\pm \frac{2}{\sqrt{n}}$, isto é, a_t pode ser considerado um ruído branco. Na prática, dizemos que a_t é um ruído branco se até 5% do número de “lags” do gráfico da função de autocorrelação estiverem fora do intervalo de confiança. Contudo se k é pequeno, dizemos que os limites subestimarão a significância de qualquer discrepância. Outros testes poderiam ser abordados para verificar a análise de resíduo, como o teste de Box-Pierce.

Cada pesquisador, poderia ter encontrado diversos modelos adequados, qual seria o melhor modelo? Poderia depender da finalidade do pesquisador, como por exemplo, previsão. Mas de uma forma geral, adota-se um modelo mais parcimonioso e que seja fácil de interpretar. Diversos critérios podem ser adotados para a seleção de modelos dependendo de sua finalidade, como critério de Akaike, critério Bayesiano, critério do erro quadrático médio, etc.

Depois de selecionado o modelo, em muitas situações práticas, como na área financeira, ocorre que a_t não é um ruído branco, devido a variância σ_a^2 não ser constante ao longo do tempo, é o que chamamos de volatilidade. Para esses tipos de séries, modelos por meio de modelos autorregressivos heterocedásticos, os modelos ARCH. Outras situações poderiam ocorrer como outras componentes influenciarem a série além da tendência, que poderia ser a sazonalidade. Daí teríamos modelos do tipo SARIMA, que além das diferenças sucessivas para eliminar a tendência, aplicaríamos diferenças sazonais, geralmente em períodos de 12 meses, para eliminar a componente sazonal, e o modelo reduzir a um ARMA, e o procedimento em diante seria equivalente. Outras abordagens, poderiam ainda ser implementadas nessa dissertação, como o estudo de intervenções, previsões, etc.

Portanto, dados coletados ao longo do tempo, geralmente não são independentes, e grande parte das metodologias de análises de dados tem como uma de suas pressuposições a independência das observações. Dessa forma, os resultados baseados nessas metodologias podem levar a erros. E assim, como forma de modelar a correlação entre essas observações, surge a metodologia das séries temporais, como metodologia de grande importância para esse tipo de dado.