Análise de Séries Temporais

Estimação dos modelos ARMA e ARIMA

José Augusto Fiorucci

Section 1

Estimador de Mínimos Quadrados

Calculo dos erros

Dada uma série observada, digamos $x_1, x_2, \ldots, x_n$, o termo de erro de um modelo ARMA ou ARIMA pode ser calculado recursivamente.

- Exemplo:
  1. Modelo ARMA(1,1): $x_t = \phi x_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$

logo, $\varepsilon_t = x_t - \phi x_{t-1} - \theta \varepsilon_{t-1}$

e portanto
$$\begin{aligned}
\varepsilon_1 &= 0 \\
\varepsilon_2 &= x_2 - \phi x_1 - \theta \varepsilon_1 \\
\varepsilon_3 &= x_3 - \phi x_2 - \theta \varepsilon_2 \\
&\vdots \\
\varepsilon_n &= x_n - \phi x_{n-1} - \theta \varepsilon_{n-1}
\end{aligned}$$

- Exemplo:

  2) Modelo ARIMA(1,1,1):

Se $\{x_t\}$ é um processo ARIMA(1,1,1), então o processo formado por

$$w_t = \nabla x_t = x_t - x_{t-1}, \quad t = 2, 3, \ldots, n$$

é ARMA(1,1). Desta forma,

$$w_t = \phi w_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$$

e portanto

$$\begin{aligned}
\varepsilon_1 &= 0 \\
\varepsilon_2 &= 0 \\
\varepsilon_3 &= w_3 - \phi w_2 - \theta \varepsilon_2 \\
&\vdots \\
\varepsilon_n &= w_n - \phi w_{n-1} - \theta \varepsilon_{n-1}
\end{aligned}$$

Soma do Erro Quadrático

Seja

$$\beta = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)'$$

o vetor de parametros do modelo ARIMA(p,d,q) e

$$w_t = \nabla^d x_t$$

a série formada pelas diferenças de ordem d.

Assim a soma dos erros quadráticos pode ser escrita como

$$S(\beta) = \sum_{t=p+d+1}^{n} \varepsilon_t(\beta)^2$$

em que

$$\varepsilon_t(\beta) = x_t - \sum_{i=1}^{p} \phi_i w_{t-1} - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}, \quad t = p+d+1, \dots, n$$

sendo $\varepsilon_1 = \varepsilon_2 = \cdots = \varepsilon_{p+d} = 0$.

Assim o estimador de mínimos quadrados (EMQ) é definido como

$$\widehat{\beta} = \operatorname{argmin}_{\beta} S(\beta)$$

OBS: Para a maior parte dos modelos da família ARIMA é impossível obter o EMQ de forma análitica e por conta disso, métodos númericos geralmente são empregados.

Exercício

- Exercício: Mostre que para o modelo AR(1),

$$x_t = \phi x_{t-1} + \varepsilon_t$$

o EMQ de ϕ é dado por

$$\widehat{\phi} = \frac{\sum_{t=2}^{n} x_t x_{t-1}}{\sum_{t=2}^{n} x_{t-1}^2}$$

Section 2

Estimador de máxima verossimilhança

Função de Verossimilhança

No que segue adotaremos a seguinte notação

- $x_{t|t-1}$ para representar a v.a. x_t condicionada em toda a informação prévia, no caso, $x_{t-1}, x_{t-2}, \ldots, x_1$
- $\mu_{t|t-1} = E[x_{t|t-1}] = E_{t-1}[x_t] = E[x_t|x_{t-1}, x_{t-2}, \ldots, x_1]$ para representar o valor esperado de $x_{t|t-1}$
- $\sigma^2_{t|t-1} = Var[x_{t|t-1}] = Var_{t-1}[x_t] = Var[x_t|x_{t-1}, x_{t-2}, \ldots, x_1]$ para representar a variância de $x_{t|t-1}$
- $f(x_{t|t-1}) = f_{t-1}(x_t) = f(x_t|x_{t-1}, x_{t-2}, \ldots, x_1)$ para representar função densidade de probabilidade de $x_{t|t-1}$.

Função de Verossimilhança Condicional

- As variáveis $x_1, \ldots, x_n$ são geralmente condicionalmente dependentes, no entanto a função densidade de probabilidade conjunta pode ser escrita como

$$\begin{aligned} f(x_1, \ldots, x_n) &= f(x_1) f(x_2|x_1) f(x_3|x_2, x_1) \ldots f(x_n|x_{n-1}, x_{n-2}, \ldots, x_1) \\ &= f(x_1) \prod_{t=2}^{n} f_{t-1}(x_t) \end{aligned}$$

- Se considerarmos que as primeras k observações são necessárias apenas para inicializar o modelo, será util escrever a equação anterior da seguinte forma

$$f(x_1, \ldots, x_n) = f(x_1, \ldots, x_k) \prod_{t=k+1}^{n} f_{t-1}(x_t)$$

- Como $f(x_1, \ldots, x_k)$ é desconhecido, em séries temporais tradicionalmente utilizamos a função de verossimilhança condicional, definida como

$$L(\beta) = f(x_{k-1}, \ldots, x_n | x_1, \ldots, x_k) = \prod_{t=k+1}^{n} f_{t-1}(x_t)$$

em que β é vetor com todos os parâmetros do modelo.

Verossimilhança do modelo ARMA(p,q)

Considere o modelo ARMA(p,q), dado por

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

em que $\{\varepsilon_t\}$ é um processo i.i.d $N(0, \sigma^2)$

Note que

$$\begin{aligned}
\mu_{t|t-1} &= E_{t-1}[x_t] = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \\
\sigma^2_{t|t-1} &= Var_{t-1}[x_t] = Var[\varepsilon_t] = \sigma^2
\end{aligned}$$

Assim

$$x_{t|t-1} \sim N(\mu_{t|t-1}, \sigma^2)$$

e portanto,

$$\begin{aligned}
L(\beta, \sigma^2) = \prod_{t=p+1}^{n} f_{t-1}(x_t) &= \prod_{t=p+1}^{n} (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2}(x_t - \mu_{t|t-1})^2\right\} \\
&= (2\pi\sigma^2)^{-(n-p-1)/2} \exp\left\{-\frac{1}{2\sigma^2} S(\beta)\right\}
\end{aligned}$$

Verossimilhança do modelo ARMA(p,q)

Assim o logaritmo da função de verossimilhança condicional é dado por

$$\ell(\beta, \sigma^2) \propto \frac{-(n-p-1)}{2} \log(\sigma^2) + -\frac{1}{2\sigma^2} S(\beta)$$

Donde obtemos que o Estimador de Máxima Verossimilhança para σ^2 é dado por,

$$\widehat{\sigma^2} = \frac{S(\beta)}{n-p-1}$$

Agora, substituindo σ^2 por $\widehat{\sigma^2}$ em $\ell(\beta, \sigma^2)$, obtemos

$$\ell(\beta) \propto \frac{-(n-p-1)}{2} \log(\, S(\beta)\,)$$

Conclusão: *sob a hipótese de normalidade, o Estimador de Máxima Verossimilhança para β corresponde ao Estimador de Mínimos Quadrados.*

OBS 1: *o resultado anterior pode ser extendido para qualquer modelo de espaço de estado linear, isto é, qualquer modelo que possa ser escrito no formato* $x_t = g_t + \varepsilon_t$, *em que* g_t *é uma funçao que depende apenas dos pontos anteriores ao tempo* t.

OBS 2: *em modelos ARIMA(p,d,q) o procedimento de estimação consiste em primeiro calcular a série com d-diferenças,* $w_t = \nabla^d x_t$, *e então obter o EMV de* $\{w_t\}$

OBS 3: *se o processo estacionário* $\{x_t\}$ *possui média diferente de zero, isto é,* $\mu = E[x_t] \neq 0$, *então*

$$\mu_{t|t-1} = \alpha + \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

em que $\alpha = (1 - \phi_1 - \cdots - \phi_p)\mu$ *é o intercepto.*

OBS 4: *em modelos ARMA(1,1) (ou mesmo ARIMA(1,d,1)) desconfie de redundância na parametrização quando* $\widehat{\phi} \approx -\widehat{\theta}$, *pois nesta situação,* $(1 - \widehat{\phi})x_t = (1 + \widehat{\theta})\varepsilon_t$ *e portanto* $x_t = \varepsilon_t$. *Em ordens maiores, o problema de cancelamento de raízes é difícel de ser identificado, de modo que, deve-se sempre tentar usar modelos ordens mais baixas.*