



Análisis de Series de Tiempo

Carrera de Especialización en Inteligencia Artificial

Clase 6

Ing. Magdalena Bouza, Ing. Carlos German Carreño Romano

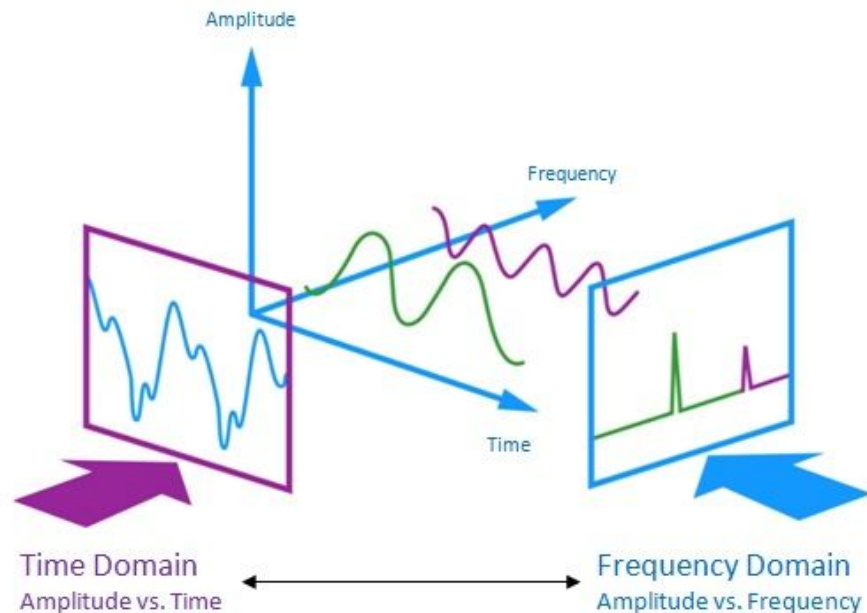
Contenido

- Análisis espectral
- Modelos de regresión
 - Cómo incorporar información externa en el modelado de las series de tiempo
 - Cómo realizar tratamiento de outliers

Introducción al análisis espectral

Introducción al análisis espectral

- Series periódicas
- Análisis en el dominio del tiempo
- Periodograma
- Densidad espectral de potencia
- Análisis en el dominio de frecuencias



Ref.: [1] *Time Series Analysis*, J.D. Cryer ; [2] *Introduction to Time Series Modeling*, G. Kitagawa

Series periódicas

Consideremos un proceso dado por la expresión

$$Y_t = R \cos (2\pi ft + \Phi)$$

Conceptos clave en este tipo de series son:

- Amplitud
- Frecuencia
- Fase

Usando trigonometría podemos pasar a una expresión distinta:

$$Y_t = R \cos (2\pi ft + \Phi) = A \cos (2\pi ft) + B \sin (2\pi ft)$$

Introducción al análisis espectral

$$Y_t = R \cos(2\pi ft + \Phi) = A \cos(2\pi ft) + B \sin(2\pi ft)$$



$$\begin{aligned} A &= R \cos(\Phi) & B &= -R \sin(\Phi) \\ R &= \sqrt{A^2 + B^2} & \Phi &= \arctan(-B/A) \end{aligned}$$

Si tenemos la frecuencia, entonces podemos ajustar los parámetros A y B a partir de, por ejemplo, cuadrados mínimos.

Introducción al análisis espectral

Una generalización de lo anterior nos permite definir un proceso Y a partir de una combinación lineal de senos y cosenos de distintas frecuencias, a partir de la expresión:

$$Y_t = A_0 + \sum_{j=1}^m (A_j \cos(2\pi f_j t) + B_j \sin(2\pi f_j t))$$

Nuevamente, podemos hacer estimaciones por cuadrados mínimos para encontrar los coeficientes A_j y B_j .

Fourier

Suponiendo que n es impar, esto es $n=2k+1$, se tiene que las frecuencias de la forma $1/n, 2/n, \dots, k/n$ se denominan frecuencias de Fourier. Las variables predictoras del seno y coseno a tales frecuencias son ortogonales, y la estimación por cuadrados mínimos resulta simplemente:

$$\begin{aligned}\hat{A}_0 &= \bar{Y} \\ \hat{A}_j &= \frac{2}{n} \sum_{t=1}^n Y_t \cos(2\pi t j / n) & \hat{B}_j &= \frac{2}{n} \sum_{t=1}^n Y_t \sin(2\pi t j / n)\end{aligned}$$

El resultado de la serie de suma de cosenos y senos es general. Esto es, cualquier serie de cualquier longitud n , sea determinística o estocástica y con o sin periodicidades verdaderas puede ajustarse perfectamente por este modelo eligiendo $m=n/2$ si n es par y $m=(n-1)/2$ si n es impar. Luego habrá n parámetros a estimar para ajustar la serie de largo n .

Periodograma

Para muestras de tamaño impar, el periodograma se define como:

$$I(j/n) = \frac{n}{2} (\hat{A}_j^2 + \hat{B}_j^2)$$

donde j/n es la frecuencia para $j=1, 2, \dots, k$. Si la muestra tiene tamaño par y $n=2k$, en la frecuencia de corte $f=k/n=1/2$ resulta:

$$I(1/2) = n\hat{A}_k^2$$

El periodograma es la suma de los cuadrados con dos grados de libertad asociados con el par de coeficientes de la frecuencia j/n cuando n es impar:

$$\sum_{j=1}^n (Y_j - \hat{Y})^2 = \sum_{j=1}^k I(j/n)$$

Introducción al análisis espectral

```
# example 1a
```

```
N=1000
```

```
f=50
```

```
T=1/f
```

```
t=np.arange(N)
```

```
Phi=np.random.normal(0,1,N)
```

```
X=sin(2*pi*f*t/N)
```

```
Y=sin(2*pi*f*t/N + Phi)
```

```
plt.plot(Phi)
```

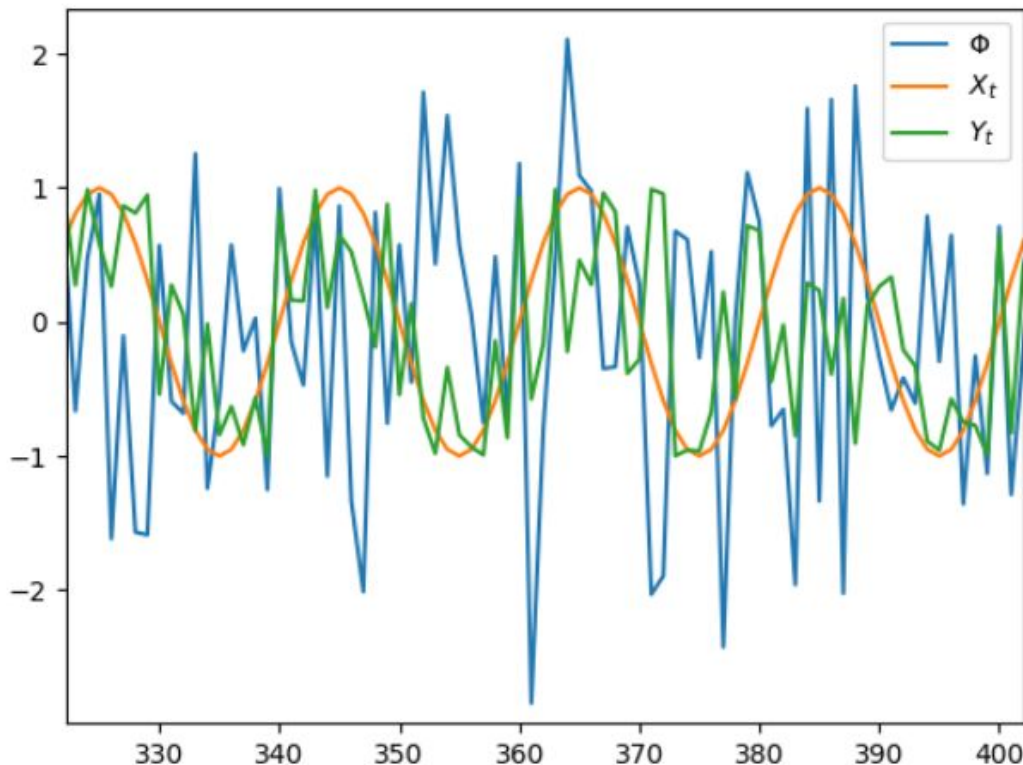
```
plt.plot(X)
```

```
plt.plot(Y)
```

```
plt.legend([' $\Phi$ ', ' $X_t$ ',
```

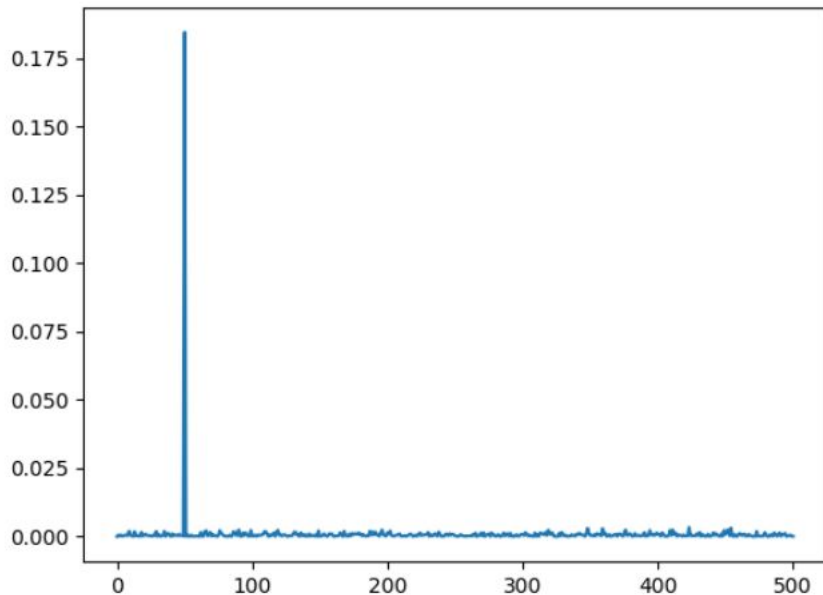
```
 $Y_t$ '],
```

```
plt.show())
```



Periodograma

```
G, Pyy_den = signal.periodogram(Y,N)  
plt.plot(G, Pyy_den)
```



Se puede observar el piso de ruido y una componente marcada en la frecuencia 50.

Ejemplos hands-on
(ver `spectral_intro.py`)

Densidad espectral de potencia

Si la autocovarianza decrece rápido a medida que aumentan los retardos y se satisface que tienen energía finita, esto es:

$$\sum_{k=-\infty}^{\infty} |C_k| < \infty$$

se puede definir la transformada de Fourier de la autocovarianza. La función definida entre $-\frac{1}{2} < f < \frac{1}{2}$ como sigue se denomina ***densidad espectral de potencia***:

$$p(f) = \sum_{k=-\infty}^{\infty} C_k e^{-2\pi j k f}$$

Densidad espectral de potencia

Se puede ver que:

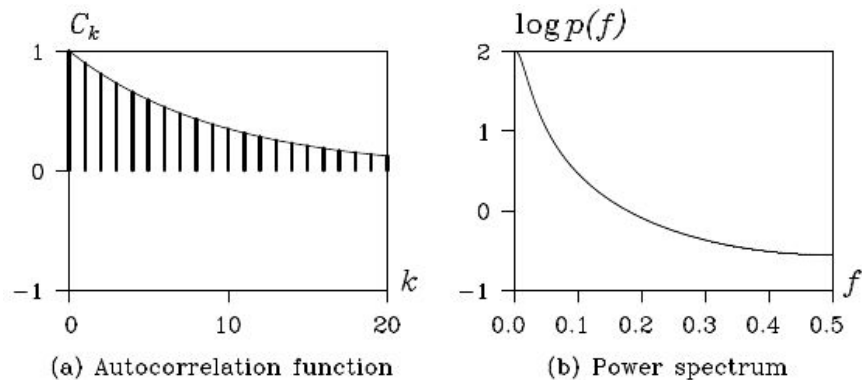
$$\begin{aligned} p(f) &= \sum_{k=-\infty}^{\infty} C_k e^{-2\pi j k f} \\ &= \sum_{k=-\infty}^{\infty} \cos(2\pi k f) - j \sum_{k=-\infty}^{\infty} \sin(2\pi k f) \\ &= \sum_{k=-\infty}^{\infty} \cos(2\pi k f) \\ &= C_0 + 2 \sum_{k=1}^{\infty} \cos(2\pi k f) \end{aligned}$$

Densidad espectral de potencia

Entonces una serie de tiempo se puede expresar en función de componentes trigonométricas y la densidad espectral de potencia es una representación en esos términos. Pero por otro lado, podemos partir de una densidad espectral y querer encontrar la función de autocovarianza, que se puede calcular a partir de la transformada inversa de Fourier:

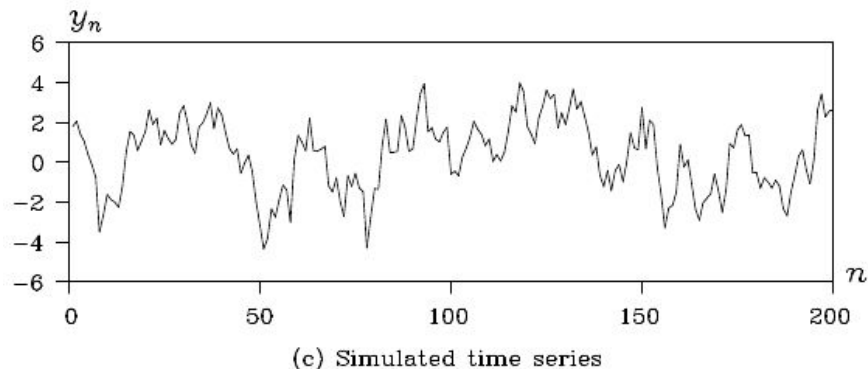
$$C_k = \int_{-1/2}^{1/2} p(f) e^{2\pi j k f} df = \int_{-1/2}^{1/2} p(f) \cos(2\pi k f) df$$

Ejemplo 1: densidad espectral del ruido blanco

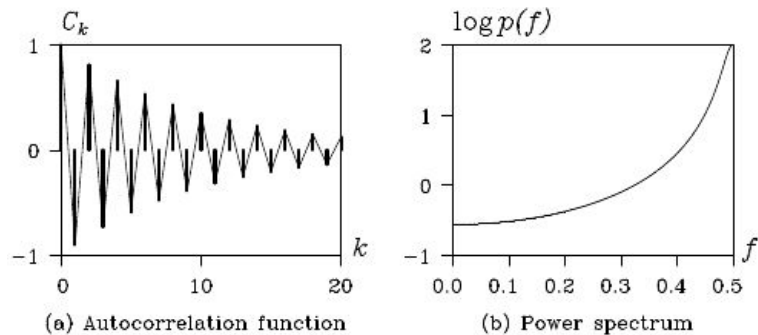


$$p(f) = \sum_{k=-\infty}^{\infty} \cos(2\pi k f) = C_0 = \sigma^2$$

$$C_0 = \sigma^2 \quad C_k = 0 \text{ para } k \neq 0$$



Ejemplo 2: densidad espectral de un proceso AR(1)

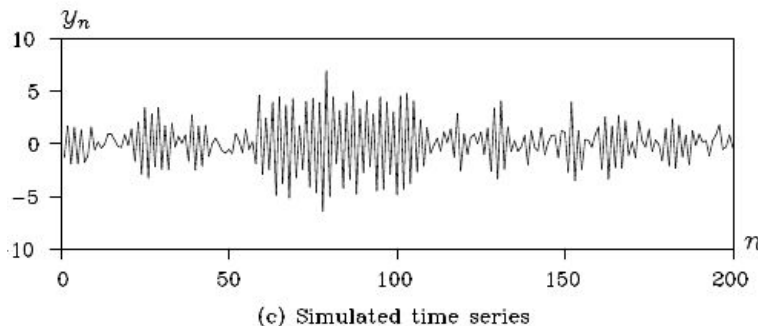


$$Y_t = a_1 Y_{t-1} + e_t$$

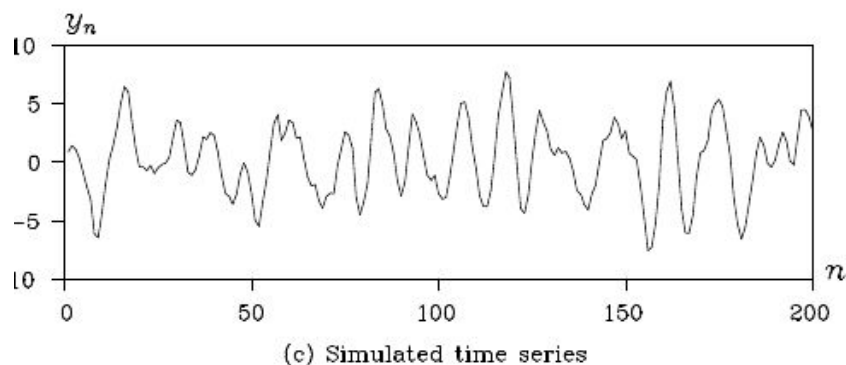
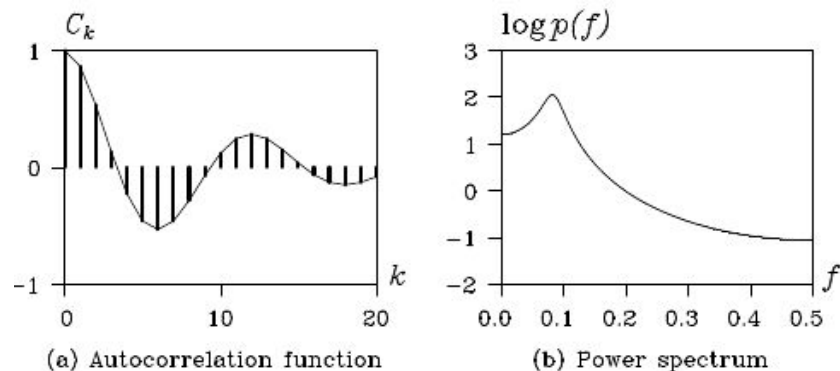
$$C_k = \sigma^2 \frac{a^{|k|}}{1 - a^2}$$

$$\begin{aligned} p(f) &= \frac{\sigma^2}{|1 - ae^{2\pi jf}|^2} \\ &= \frac{\sigma^2}{1 - 2a \cos 2\pi f + a^2} \end{aligned}$$

$$a = 0.9$$



Ejemplo 3: densidad espectral de un proceso AR(2)



$$Y_t = a_1 Y_{t-1} + a_2 Y_{t-2} + e_t$$

$$R_1 = \frac{a_1}{1 - a_2}$$

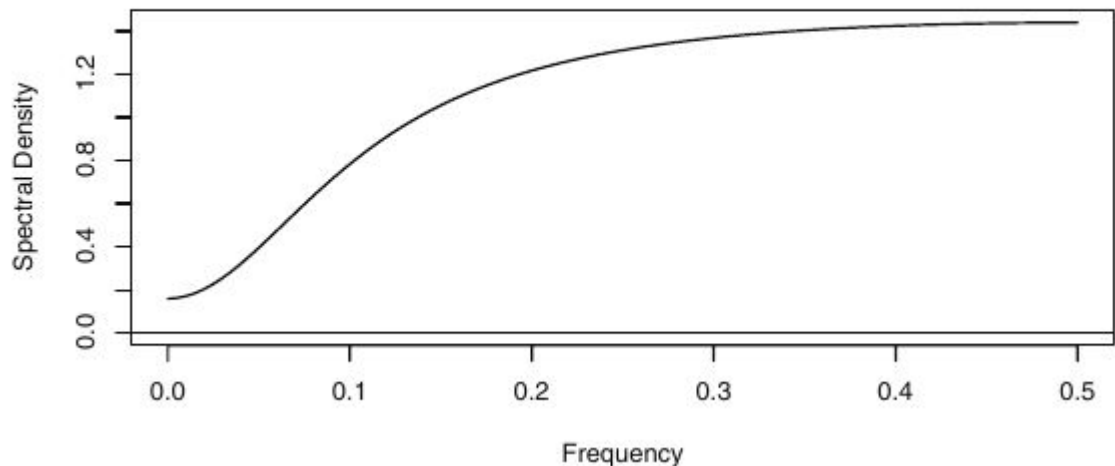
$$R_k = a_1 R_{k-1} + a_2 R_{k-2}$$

$$p(f) = \frac{\sigma^2}{|1 - a_1 e^{-2\pi j f} - a_2 e^{-4\pi j f}|^2}$$
$$= \frac{\sigma^2}{|1 - 2a_1(1 - a_2) \cos 2\pi f - 2a_2 \cos 4\pi f + a_1^2 + a_2^2|}$$

$$a_1 = 1.558$$

$$a_2 = -0.81$$

Ejemplo 3: densidad espectral de un ARMA(1,1)



$$Y_t = a_1 Y_{t-1} + b_1 e_{t-1}$$

$$p(f) = \frac{1 + b_1^2 - 2b_1 \cos(2\pi f)}{1 + a_1^2 - 2a_1 \cos(2\pi f)} \sigma_e^2$$

Usando la notación de operadores, el caso general se puede expresar como:

$$p(f) = \left| \frac{\theta(e^{-2\pi jf})}{\phi(e^{-2\pi jf})} \right|^2 \sigma_e^2$$

Análisis en el dominio espectral

Muchas veces la información en el dominio del tiempo para distintas series parece ser muy distinta y sin embargo en el dominio de frecuencias son similares. Cuando la representación espectral es continua, surge la necesidad de estimar el espectro.

Algunas técnicas de estimación espectral se llaman estimaciones no-paramétricas y asumen muy poco acerca de la densidad espectral “verdadera”.

La estimación paramétrica asume que un modelo autorregresivo de orden adecuado sirve para ajustar un modelo a la serie de tiempo. Luego la densidad espectral estimada va a tener las mismas propiedades que las del autorregresivo.

Análisis en el dominio espectral

- densidad espectral suavizada
- sesgo
- varianza
- fugas
- ancho de banda
- ensanchamiento

Resumen

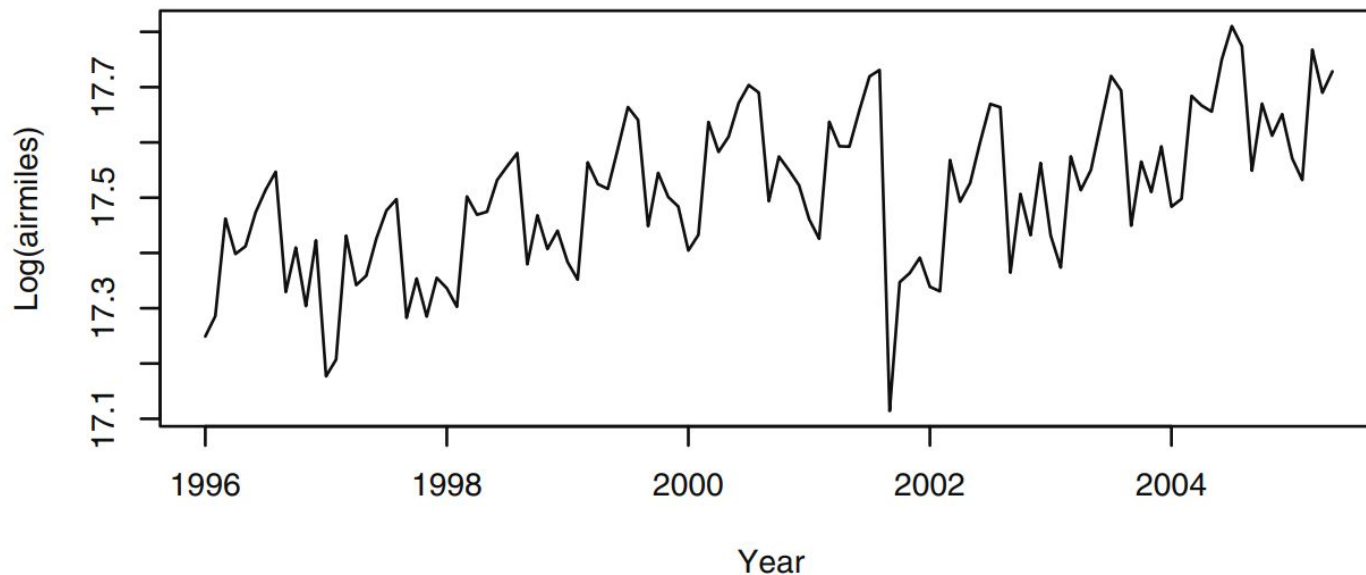
- Modelos de series de tiempo como combinaciones lineales de senos y cosenos, introducción al análisis espectral
- El Periodograma es una herramienta para encontrar la contribución de varias frecuencias en la representación espectral
- Se puede extender el modelado a rangos continuos de frecuencias
- Se exploran las densidades espectrales de procesos ARMA
- Algunas propiedades de la densidad espectral muestral sugieren que la densidad espectral no es un estimador consistente, por lo que hay que buscar más para estimar espectros...

Modelos de Regresión

Análisis de intervenciones

Análisis de intervenciones

Exhibit 11.1 Monthly U.S. Airline Miles: January 1996 through May 2005



Análisis de intervenciones

Consideremos el caso de una única intervención:

$$Y_t = \boxed{m_t} + \boxed{N_t}$$

Intervención:
cambio en la media

Proceso natural:
serie de tiempo que puede ser
modelada como un (S)ARIMA

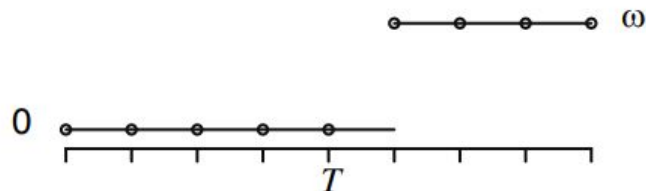
Si la intervención ocurre en un instante T , se asume que $m_t = 0$ para $t < T$ y a $\{Y_t\}_{t < T}$ se la conoce como *datos preintervención* y puede ser usada para modelar N_t .

La estimación de los parámetros de m_t se puede realizar por MV o el enfoque de modelo de estados ya presentados.

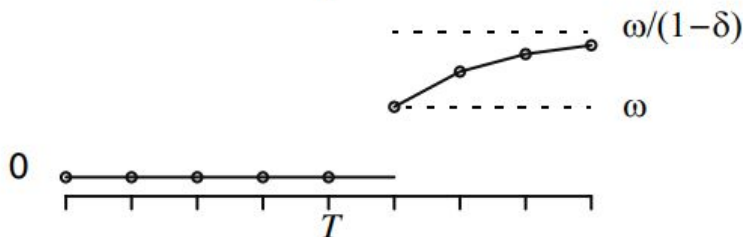
Análisis de intervenciones

Exhibit 11.3 Some Common Models for Step Response Interventions
(All are shown with a delay of 1 time unit)

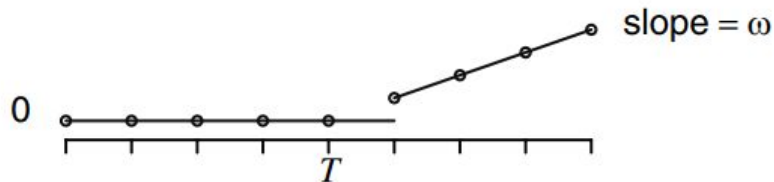
$$S_t^T \equiv \mathbf{I}\{t > T\} \quad \omega B S_t^{(T)}$$



$$(b) \quad \frac{\omega B}{1 - \delta B} S_t^{(T)}$$



$$(c) \quad \frac{\omega B}{1 - B} S_t^{(T)}$$



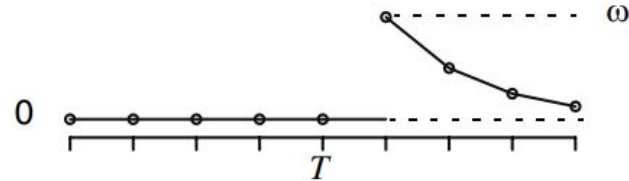
Análisis de intervenciones

Exhibit 11.4 Some Common Models for Pulse Response Interventions
(All are shown with a delay of 1 time unit)

$$[P_t^{(T)} \equiv \delta_T(t)]$$

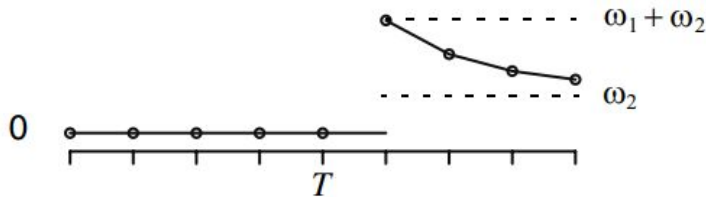
(a)

$$\frac{\omega B}{1 - \delta B} P_t^{(T)}$$



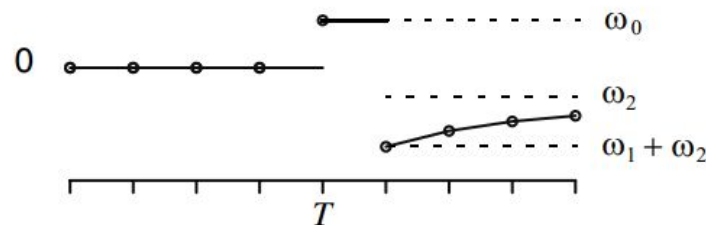
(b)

$$\left[\frac{\omega_1 B}{1 - \delta B} + \frac{\omega_2 B}{1 - B} \right] P_t^{(T)}$$



(c)

$$\left[\omega_0 + \frac{\omega_1 B}{1 - \delta B} + \frac{\omega_2 B}{1 - B} \right] P_t^{(T)}$$



Análisis de outliers

Análisis de outliers

Se corresponden con mediciones atípicas que aparecen por errores de medición o por cambios abruptos de corta duración en el proceso subyacente.

- **Outliers aditivos (AO):** $Y'_t = Y_t + \omega_A P_t^{(T)}$, $[P_t^{(T)} \equiv \delta_T(t)]$
- **Outliers innovativos (AI):** ocurre si el error (o innovación) e_t se ve perturbado, de forma que $e'_t = e_t + \omega_I P_t^T$. Al perturbarse el término de innovación, se ven perturbados también los valores de la serie en tiempos posteriores.

Usando la representación de proc. lineal general $Y_t = e_t + \psi_1 e_{t-1} + \psi_2 e_{t-2} + \dots$ vemos que

$$\begin{aligned} Y'_t &= e'_t + \psi_1 e_{t-1} + \psi_2 e_{t-2} + \dots \\ &= [e'_t + \psi_1 e_{t-1} + \psi_2 e_{t-2} + \dots] + \psi_{t-T} \omega_I = Y_t + \psi_{t-T} \omega_I \end{aligned}$$

Detección de outliers

En lo que sigue vamos a ver técnicas para detectar la presencia de ambos tipos de outliers a través de tests estadísticos.

Para ello es conveniente utilizar la representación $AR(\infty)$ de las series de tiempo para definir los residuos en términos de la serie con outliers:

$$a_t = Y'_t - \pi_1 Y'_{t-1} - \pi_2 Y'_{t-2} - \dots$$

Una vez que se encuentra un outlier se puede incorporar al modelo, y continuar el proceso de detección de outliers hasta que no se encuentren nuevos eventos.

Detección de IO

Si tenemos un único IO en $t=T$, $a_t = \omega_I \mathbf{I}\{T = t\} + e_t$ y un buen estimador para ω_I es $\tilde{\omega}_I = a_T$, con varianza σ_e^2 . Luego, se puede definir el estadístico $\lambda_{1,T} = \frac{a_T}{\sigma_e}$ para testear si hay un IO en $t=T$, donde

H₀: no hay outliers en $t=T$ vs **H₁**: hay un outlier en $T=t$.

Bajo H_0 , $\lambda_{1,T} \approx \mathcal{N}(0, 1)$.

Si conocemos el instante T , para un test de nivel significación α , rechazaremos el test cuando $|\lambda_{1,T}| > z_{1-\alpha/2}$.

Si no conocemos el valor de T debemos calcular el test para todas las observaciones. Luego rechazaremos H_0 cuando $\lambda_1 = \max_{1 \leq t \leq n} |\lambda_{1,t}| > z_{1-\alpha/(2n)}$

Corrección de Bonferroni

Detección de AO

Para detectar los AO, se puede mostrar que los residuos se pueden escribir como

$$a_t = -\omega_A \pi_{t-T} + e_t, \quad \pi_0 = -1, \quad \pi_j = 0 \text{ si } j < 0$$

Luego, se puede escribir de forma recursiva

$$a_T = \omega_A + e_T, \quad a_{T+1} = -\pi_1 \omega_A + e_{T+1}, \quad \dots, \quad a_n = -\pi_{n-T} \omega_A + e_n$$

y resolver el problema por cuadrados mínimos, obteniendo el estimador

$$\hat{\omega}_A = -\rho^2 \sum_{t=1}^n \pi_{t-T} a_t, \quad \rho = 1 + \pi_1^2 + \dots + \pi_n^2$$

con varianza $\rho^2 \sigma_e^2$. De esta forma, $\lambda_{2,T} = \frac{\hat{\omega}_A}{\rho^2 \sigma_e^2}$ es un estadístico para las test

H₀: la serie no presenta outliers vs. **H₁**: existe un AO en T.

Bajo H₀ $\lambda_{2,T} \approx \mathcal{N}(0, 1)$. Nuevamente, si desconocemos el valor de T aplicamos la corrección de Bonferroni.

Si $|\lambda_{1,T}| > |\lambda_{2,T}|$ diremos que es de tipo IO, y en caso contrario de tipo AO.