BIBLIOTECA DE PROGRAMAS

ANALISIS DE UN PROGRAMA PARA TRATAMIENTO DE SERIES TEMPORALES. EL PROGRAMA BMD 02T

Introducción

Se puede decir que una serie temporal es una sucesión de observaciones; cuyo índice de sucesión es el tiempo, aunque en algunos casos este índice puede variar de acuerdo con alguna otra dimensión. El hecho que distingue el análisis de series temporales de otros análisis estadísticos es el reconocimiento explícito de la importación del orden en que las observaciones son hechas. Mientras que en muchos problemas las observaciones son estadísticamente independientes, las observaciones sucesivas de las series temporales, pueden ser dependientes, y esta dependencia puede ser una función de las posiciones en la sucesión.

La gran importancia de las series temporales, reside en el hecho de que en casi todas las áreas del conocimiento existen fenómenos, cuyo desarrollo y variación con el paso del tiempo son de interés e importancia. Como ejemplo de fenómenos que se adapten al modelo matemático de series temporales podemos citar los siguientes:

a) *Características de una nación que afectan a numerosos individuos, tales como condiciones económicas y población, que evolucionan y fluctúan con el tiempo.*

b) *La variación de los distintos índices de coste de vida en los distintos países.*

c) *Las variaciones que experimenta la temperatura de una persona en el transcurso de una enfermedad.*

d) *Las variaciones de ozono en la atmósfera en el transcurso del tiempo, etc.*

Hay varias razones para el tratamiento de datos mediante - series temporales. Una de ellas puede ser la predicción de hechos futuros basados en el conocimiento del pasado. Otra puede ser el control del proceso que produce la serie temporal;- o bien la creación de un mecanismo que simule los hechos que- están sucediendo en la realidad. El estadístico está interesado en la inferencia estadística; sobre la base de una cantidad limitada de información; es decir, sobre el conocimiento- de la serie temporal en un intervalo acotado del tiempo; o en otras palabras sobre un número finito de observaciones; se -- trata de hacer inferencias acerca del mecanismo probabilístico que produce la serie, para analizar su estructura básica.

Conceptos fundamentales en análisis de series temporales

Sea un espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Un proceso estocástico es una aplicación

$$X : \Omega \times T \longrightarrow R$$

donde T es un intervalo del conjunto $(-\infty, \infty)$ o bien T es- un subconjunto de los números enteros y la aplicación X es -- tal que para cada $t \in T$, $X(\cdot, t)$ es una aplicación medible de-- $(\Omega, \mathcal{A})$ en (R,B) cuando $T \subset (-\infty, \infty)$ el proceso se denomina continuo y en otro caso discreto.

Claramente fijado un t,X(.,t) es una variable aleatoria; y por tanto llevará asociada una función de distribución $F_{X(.,t)}$.

Igualmente si fijamos $(t_1,t_2,\ldots,t_n)$; $(X(\cdot,t_1),X(\cdot,t_2),\ldots,X(\cdot,t_n))$ es una variable aleatoria n-dimensional llevando asociada una función de distribución n-dimensional.

El conjunto Ω nos representa el conjunto de individuos -- que están siendo medidos para obtener nuestros datos; y T nos representa los tiempos en los cuales efectuamos la medida.

Un proceso estocástico es estacionario en sentido fuerte- si para todo $(t_1,t_2,\ldots t_n)$ y todo h tenemos:

$$F_{X(.,t_1),\ldots X(.,t_n)}(x_1,x_2,\ldots,x_n) = F_{X(.,t_1+h),\ldots,X(.,t_n+h)}(x_1,x_2,\ldots,x_n)$$

donde las F son las funciones de distribución para las v.a $(X(.,t_1),X(.,t_2),\ldots,X(.,t_n))$ y

$$[X(\cdot,t_1+h),\ X(\cdot,t_2+h),\ \ldots,\ X(\cdot,t_n+h)]$$

respectivamente.

El proceso es débilmente estacionario si

$$\mu = E[X(\cdot,t)] = E[X(\cdot,t+h)]$$

para todo t y h dentro del dominio de definición de X y

$$E[(X(\cdot,t)-\mu)(X(\cdot,t+h)-\mu)] = \gamma(h)$$

Los conceptos siguientes se definen para procesos débilmente estacionarios; y los términos, serie temporal y proceso estocástico débilmente estacionario tendrán un mismo significado.

Se llama <u>función de autovarianza</u> de un proceso estocástico débilmente estacionario a la función

$$\gamma(h) = E\left[(X(\cdot,t)-\mu)(X(\cdot,t+h)-\mu) \right]$$

y función de autocorrelación a la función

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}$$

Al número real h se le denomina en inglés <u>lag</u>, en francés <u>decalage</u> y en español podría traducirse por desplazamiento o también por retardo.

Se denomina función de distribución espectral a una función $F: [-\pi,\pi] \rightarrow R$ ona no decreciente y tal que

$$\gamma(h) = \int_{-\pi}^{\pi} e^{i\lambda h} \, dF(\lambda)$$

Si $F(\lambda)$ es derivable, a la derivada $f(\lambda)=F'(\lambda)$ se la llama <u>función de densidad espectral</u> del proceso X(.,t). Cuando el proceso X(.,t) es discreto, si

$$\gamma(0) + 2\sum_{h=1}^{\infty} |\gamma(h)| = \sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$$

entonces $F(\lambda)$ es derivable, y la función de densidad espectral es:

$$f(\lambda) = \frac{1}{2\pi}\gamma(0) + \frac{1}{\pi}\sum_{h=1}^{\infty} \gamma(h)\cos\lambda h$$

$$\gamma(h) = \int_{-\pi}^{\pi} \cos\lambda h \, f(\lambda)\, d\lambda$$

El espectro del proceso X(.,t) viene dado por todos los puntos λ_0; tales que para todo $\varepsilon > 0$ se tiene:

$$F(\lambda_0+\varepsilon) - F(\lambda_0-\varepsilon) > 0$$

es decir λ_0 es un punto de crecimiento de F.

Veamos ahora estos mismos conceptos cuando se consideran dos procesos estocásticos débilmente estacionarios $X_1(.,t)$; $X_2(.,t)$; y supondremos que $E[X_1(\cdot,t)]=\mu_1$ $E[X_2(\cdot,t)]=\mu_2$

Se llaman funciones de <u>cross-covarianza o covarianza - cruzada</u> de los procesos X_1 y X_2 a las funciones

$$\gamma_{X_1X_2}(h) = E\left[(X_1(\cdot,t)-\mu_1)(X_2(\cdot,t+h)-\mu_2)\right]$$
$$\gamma_{X_2X_1}(h) = E\left[(X_1(\cdot,t+h)-\mu_1)(X_2(\cdot,t)-\mu_2)\right]$$

Se llaman funciones de cross-correlación o correlación cruzada a las siguientes:

$$\rho_{X_1X_2}(h) = \frac{\gamma_{X_1X_2}(h)}{\sqrt{\gamma_{X_1X_1}(0)\cdot\gamma_{X_2X_2}(0)}}$$

$$\rho_{X_2X_1}(h) = \frac{\gamma_{X_2X_1}(h)}{\sqrt{\gamma_{X_1X_1}(0)\cdot\gamma_{X_2X_2}(0)}}$$

Suponemos ahora que a partir de los procesos $X_1(.,t)$ y $X_2(.,t)$ formamos el siguiente:

$$Y(.,t)=X_1(.,t)+iX_2(.,t)$$

Entonces el proceso $Y(.,t)$ es un proceso a valores complejos . Su función de autocovarianza es

$$\gamma(h) = E\left[\left\{\overline{(X_1(\cdot,t)-\mu_1)+i\,(X_2(\cdot,t)-\mu_2)}\right\}\cdot\left\{(X_1(\cdot,t+h)-\mu_1)(X_2(\cdot,t+h)-\mu_2)\right\}\right]$$

$$= \gamma_{X_1X_1}(h) + \gamma_{X_2X_2}(h) + i\,(\gamma_{X_1X_2}(h) - \gamma_{X_2X_1}(h))$$

Si $\gamma_{X_1X_1}(h)=\gamma_{X_2X_2}(h)$ $\gamma_{X_1X_2}(h)=-\gamma_{X_2X_1}(h)$ entonces r(h) determina completamente las funciones de autocovarianza y covarianza cruzadas de los procesos $X_1(.,t)$ y $X_2(.,t)$.

Si definimos

$$I_T(\lambda)=\frac{1}{2\pi T}\left|\sum_{t=1}^{T} Y(\cdot,t)\,e^{-it\lambda}\right|=\frac{1}{2\pi T}\sum_{t=1}^{T}\sum_{s=1}^{T} Y(\cdot,t)\,\overline{Y}(\cdot,s)\,e^{-i\lambda(t-s)}$$

y tomamos esperanzas matemáticas tenemos;

$$f_T(\lambda)=E\,I_T(\lambda)=\frac{1}{2\pi T}\sum_{t=1}^{T}\sum_{s=1}^{T} r(t-s)\,e^{-i\lambda(t-s)}=\frac{1}{2\pi}\sum_{h=-(T-1)}^{T-1}\left(1-\frac{|h|}{T}\right)r(h)\,e^{-i\lambda h}$$

$$r(h)\left(1-\frac{|h|}{T}\right)=\int_{-\pi}^{\pi} e^{i\lambda h} f_T(\lambda)\,d\lambda \; ; \quad h=-(T-1),\dots,0,1,\dots,T-1$$

$$0=\int_{-\pi}^{\pi} e^{i\lambda h} f_T(\lambda)\,d\lambda \qquad h=\pm T,\pm(T+1),\pm(T+2),\dots$$

Entonces $F_T(\lambda)=\int_{-\pi}^{\lambda} f_T(\nu)\,d\nu$ es monótona no decreciente con $F_T(-\pi)=0$ y $F_T(\pi)=r(0)$.

Si tomamos límites cuando $T\to\infty$ obtenemos:

$\lim\limits_{T\to\infty} F_T(\lambda)=F(\lambda)$ con $F(-\pi)=0$ y $F(\pi)=r(0)$

con [2-1] $$\gamma_{YY}(h)=\int_{-\pi}^{\pi} e^{i\lambda h}\,dF(\lambda)$$

siendo por tanto $F(\lambda)$ la función de distribución espectral -- del proceso complejo Y(t)

Ahora consideraremos el proceso $Y_1(t)=X_1(t)+X_2(t)$ y sea $F_1(\lambda)$ su función de distribución espectral, tenemos entonces:

$$[2,2] \quad \gamma_{Y_1Y_1}(h) = \gamma_{X_1X_1}(h) + \gamma_{X_2X_2}(h) + \gamma_{X_1X_2}(h) + \gamma_{X_2X_1}(h) = \int_{-\pi}^{\pi} e^{i\lambda h}\, dF_1(\lambda)$$

Y llamando $F_2(\lambda)$ y $F_3(\lambda)$ a las funciones de distribución espectral de los procesos $X_1(t)$ y $X_2(t)$ obtenemos:

$$[2,3] \quad \gamma_{X_1X_1}(h) = \int_{-\pi}^{\pi} e^{i\lambda h}\, dF_2(\lambda) \qquad \gamma_{X_2X_2}(h) = \int_{-\pi}^{\pi} e^{i\lambda h}\, dF_3(\lambda)$$

La igualdad [2,1] la podemos poner como:

$$[2,4] \quad \gamma_{YY}(h) = \gamma_{X_1X_1}(h) + \gamma_{X_2X_2}(h) + i\left(\gamma_{X_1X_2}(h) - \gamma_{X_2X_1}(h)\right) = \int_{-\pi}^{\pi} e^{i\lambda h}\, dF(\lambda)$$

De 2-2 obtenemos:

$$\gamma_{X_1X_2}(h) + \gamma_{X_2X_1}(h) = \int_{-\pi}^{\pi} e^{i\lambda h}\, [dF_1(\lambda) - dF_2(\lambda) - dF_3(\lambda)]$$

De 2-4 obtenemos:

$$i\left(\gamma_{X_1X_2}(h) - \gamma_{X_2X_1}(h)\right) = \int_{-\pi}^{\pi} e^{i\lambda h}\, [dF(\lambda) - dF_2(\lambda) - dF_3(\lambda)]$$

Multiplicando por i obtenemos:

$$-\gamma_{X_1X_2}(h) + \gamma_{X_2X_1}(h) = i\int_{-\pi}^{\pi} e^{i\lambda h}\, [dF(\lambda) - dF_2(\lambda) - dF_3(\lambda)]$$

Sumando las dos últimas igualdades obtenemos:

$$\gamma_{X_2X_1}(h) = \frac{1}{2}\int_{-\pi}^{\pi} e^{i\lambda h}[dF_1(\lambda) - dF_2(\lambda) - dF_3(\lambda)] + \frac{i}{2}\int_{-\pi}^{\pi} e^{i\lambda h}[dF(\lambda) - dF_2(\lambda) - dF_3(\lambda)]$$

Si a su vez ponemos $\gamma_{X_2X_1}(h) = \int_{-\pi}^{\pi} e^{i\lambda h}\, dF_{21}(\lambda)$

obtenemos:

$$F_{21}(\lambda)=\frac{1}{2}\left[F_1(\lambda)-F_2(\lambda)-F_3(\lambda)\right]+\frac{i}{2}\left[F(\lambda)-F_2(\lambda)-F_3(\lambda)\right]$$

Si $F_{21}(\lambda)$ es derivable y llamamos $f_{21}(\lambda)$ a su derivada, entonces la parte real de $f_{21}(\lambda)$ se denomina la densidad del--coespectro o simplemente coespectro; y la parte imaginaria--cuadratura espectral.

Si colocamos $f_{21}(\lambda)=C_{21}(\lambda)+iQ_{21}(\lambda)$; se denomina amplitud del espectro a $AM_{21}(\lambda)=\sqrt{C^2_{21}(\lambda)+Q^2_{21}(\lambda)}$;

y fase del espectro a $FASE_{21}(\lambda)=\arg\left[C_{21}(\lambda)+iQ_{21}(\lambda)\right]$

y función de transferencia a

$$T_{21}(h)=\frac{f_{21}(h)}{f_2(h)}=TAM_{21}(h)\ e^{i\ FASE_{21}(h)}$$

donde $TAM_{21}(h)=\dfrac{AM_{21}(h)}{f_2(h)}$ y cuadrado de coherencia a

$$COSQ_{21}(h)=\left[AM_{21}(h)\right]^2/\ f_1(h)\ f_2(h)$$

donde $f_1(h)$ y $f_2(h)$ son las densidades espectrales de X_1 y X_2

PROGRAMA BMD 02T

a) Este programa computa la autocovarianza, potencia espec-- tral, covarianza cruzada, espectro cruzado, función de trans ferencia y función de coherencia de series temporales.

b) Limitaciones. El número de series temporales S debe ser -

$1 \leq S \leq 20$

El número de datos por cada serie temporal n ha de ser

$1 \leq n \leq 1000$

El número de desplazamientos, (lag, decalage) es $1 \leq m \leq 199$.

c) Entrada de datos. Los datos pueden ser perforados en tarje tas de las dos formas; que explicaremos a continuación:

LLamando Xij al dato i de la serie j tenemos::

Primera forma

$$\begin{array}{l}
X_{11}\ X_{21}\ X_{31} \cdots\cdots X_{k_1 1} \\
X_{k_1+1\,1}\ X_{k_1+2\,1} \cdots\cdots X_{2k_1\,1} \\
\vdots \\
X_{e_1k_1+1\,1}\ X_{e_1k_1+2\,1} \cdots X_{n1} \\
X_{12}\ X_{22} \cdots\cdots X_{k_2 2} \\
X_{k_2+1\,2}\ X_{k_2+2\,2} \cdots X_{2k_2\,2} \\
\vdots \\
X_{e_2k_2+1\,2}\ X_{e_2k_2+2\,2} \cdots X_{n2}
\end{array}$$

$$
\begin{array}{llll}
X_{1S} & X_{2S} & X_{3S} \; \cdots & X_{K_S S} \\
X_{K_S+1\,S} & X_{K_S+2\,S} & \cdots & X_{2K_S S} \\
\vdots & & & \\
X_{e_S K_S+1\,S} & X_{e_S K_S+2\,S} & \cdots & X_{nS}
\end{array}
$$

Es decir, se colocan primero los datos de la 1^{a} serie, a continuación los de la 2^{a} y así hasta acabar, teniendo en - cuenta que cada serie puede ser leída con formato diferente; siendo este formato de elección del usuario.

<u>Segunda forma</u>

Se colocan 1° los primeros datos de cada serie, a continuación los segundos y así hasta acabar; teniendo en cuenta que si hay más series que datos caben en una tarjeta se con tinúa perforando en una segunda tarjeta.

$$
\begin{array}{llll}
X_{11} & X_{12} & \cdots & X_{1K} \\
X_{1\,K+1} & \cdots & X_{1S} & \\
X_{21} & X_{22} & \cdots & X_{2K} \\
X_{2\,K+1} & \cdots & X_{2S} & \\
\vdots & \vdots & & \vdots \\
X_{n1} & X_{n2} & \cdots & X_{nK} \\
X_{n\,K+1} & \cdots & X_{nS} &
\end{array}
$$

Procedimiento computacional

Paso 1. Calcula las medias de cada serie; y resta a los valores de cada serie sus medias

$$m_{x'}=\frac{1}{n}\sum_{i=1}^{n} x'_i \qquad y \qquad x_i=x'_i-m_{x'}$$

Este paso se salta si la serie ha de ser desestacionarizada.

Paso 2. (Opcional). Se realiza un preblanqueo de la entrada de datos por la siguiente combinación lineal:

$$z_i=x_{i+1}-C\ x_i \qquad i=1,2, \quad , n-1$$

con $|C|<1$; siendo C un valor suministrado por el usuario

Paso 3. Se calcula la función de autocovarianza de cada serie.

$$R_x(p)=\frac{1}{n-p}\sum_{q=1}^{n-p} X_q X_{q+p} \quad \text{donde} \quad R_x(p)=R_x^*(p\Delta t)$$

donde Δt es el incremento de tiempo dado por el usuario.

Paso 4. (Opcional) Si la serie es desestacionarizada se realiza un método de ajuste por mínimos cuadrados mediante la fórmula:

$$A_x(p)=R_x(p)-\beta-\alpha i \quad \text{donde} \quad i=1,2,--, n-1$$

$$\alpha=\sum_{i=1}^{n} X_i(2i-n+1)/((C(n-1)n(n+1))/6)$$

$$\beta=\bar{X}-\alpha(n-1)/2$$

$\bar{X}$ es la media

Paso 5. Se estima la potencia espectral o función de densidad espectral por:

$$P_x(h) = \frac{2\Delta t}{m} \sum_{p=0}^{m} \varepsilon_p R_x(p) \cos\frac{h \cdot p\pi}{m} \quad h = 0,1,-,m$$

$$\varepsilon_p = \begin{cases} 1 & 0<p<m \\ 1/2 & p=0,m \end{cases} \qquad P_x(h) = P_x^*(w_h) = P_x^*\left(\frac{h\pi}{m\Delta t}\right)$$

PaPaso 6. Las estimaciones de la potencia espectral son allanadas por la ventana de Hamming

$$SP_x(0) = 0'54\ P_x(0) + 0'46\ P_x(1) \quad SP_x(m) = 0'54\ P_x(m) + 0'46\ P_x(m-1)$$

$$SP_x(h) = 0'23\ P_x(h-1) + 0'54\ P_x(h) + 0'23\ P_x(h+1)$$

$$0<h<m$$

P Paso 7. Un chequeo se realiza para comprobar la bondad de - las estimaciones

$$CHKSUM = \frac{\pi}{m\Delta t}\left[\frac{1}{2}\left(SP_x(0) + SP_x(m)\right) + \sum_{h=1}^{m-1} SP_x(h)\right]$$

debiendo ser igual este valor a Rx(0)

a Paso 8. Para compensar el preblanqueo del paso 2, el espectro allanado es recoloreado por

$$RSP_x(h) = \frac{SP_x(h)}{1 + C^2 - 2C\cos w_h \Delta t} \qquad h = 0,1,2,—,m$$

Paso 9. Las covarianzas cruzadas son computadas por las fórmulas :

$$R_{xy}(p) = \frac{1}{n-p}\sum_{q=1}^{n-p} X_q Y_{q+p} \qquad R_{xy}(-p) = \sum_{q=1}^{n-p} X_{q+p} Y_q$$

$$p = 0,1,2,--,m$$

Paso 10. Se hace un ajuste como en el paso 4. (Opcional)

$$A_{xy}(p) = R_{xy}(p) - \beta - \alpha i \quad \text{donde} \quad i = 0, 1, --, n-1$$

donde α y β son definidos como en el paso 4.

Paso 11. El espectro cruzado es dado por Pxy(h)
donde $P_{xy}(h) = C_{xy}(h) + i\, Q_{xy}(h)$

El coespectro

$$C_{xy}(h) = \frac{\Delta t}{M} \sum_{p=0}^{m} \varepsilon_p \left[R_{xy}(p) + R_{xy}(-p) \right] \cos \frac{hpM}{m}$$

La cuadratura del espectro

$$Q_{xy}(h) = \frac{\Delta t}{m} \sum_{p=0}^{m} \varepsilon_p \left[R_{xy}(p) - R_{xy}(-p) \right] \operatorname{sen} \frac{hpM}{m}$$

$$h = 0, 1, 2, --, m \qquad \varepsilon_p = \begin{cases} 1/2 & p = 0, m \\ 1 & 0 < p < m \end{cases}$$

Paso 12. Ambos espectros son allanados como en el paso 6 anterior.

Paso 13. La amplitud y la fase del espectro cruzado son- computados mediante:

$$AM_{xy}(h) = \sqrt{(SC_{xy}(h))^2 + (SQ_{xy}(h))^2}$$

$$FASE_{xy}(h) = \operatorname{Arg}\left(SC_{xy}(h) + i\, SQ_{xy}(h)\right)$$

El ángulo de fase computado es el salto de fase de la serie y respecto de la serie x.

Paso 14. Se computa la función de transferencia, de la serie x a la serie y , dada por

$$T_{xy}(h) = \frac{P_{xy}(h)}{P_x(h)} = TAM_{xy}(h)\, e^{i\, FASE_{xy}(h)}$$

$$TAM_{xy}(h) = AM_{xy}(h) / SP_x(h)$$

El salto de fase de x a y es el mismo que el del espectro cruzado. Son escritas las funciones de transferencia y amplitudes de x a y; y de y a x.

Paso 15. Finalmente el cuadrado de la coherencia entre x- e y es dado como:

$$COSQ_{xy}(h) = \frac{AM^2_{xy}(h)}{SP_x(h)SP_y(h)}$$

BIBLIOGRAFIA

1) ANDERSON.
The Statistical Analysis of Time Series.Wiley 1971.

2) JENKINS-WATTS.
Spectral Analysis and its Aplications.Holden-DAY 1968.

3) BMD.
Biomedical Computer Programs.University of California Press. 1973.

Miguel Sánchez García
Biblioteca de Programas