Modelos Box-Jenkins

Procesos ARMA: Selección del modelo

Las etapas que hemos seguido para proponer un modelo ARMA como posible generador de una serie de tiempo han sido:

1. Asesorarnos (gráficamente) acerca de la "estacionariedad de la serie". Si es estacionaria, pasar a la etapa 2.
2. Identificar los órdenes p y q del ARMA: estudio de sus fas y fap muestrales.
3. Estimar el modelo cuyos órdenes se identificaron en la etapa 2: mín. cuadr. condicionados, máx. verosimilitud.
4. Chequear el modelo estimado: análisis de residuos.

Si los residuos pueden ser considerados como procedentes de un proceso de ruido blanco (preferiblemente gaussiano), el modelo estimado es propuesto como posible generador de la serie de tiempo.

Procesos ARMA: Selección del modelo

Como acabamos de recordar, hemos basado la etapa correspondiente a la identificación de los órdenes p y q del ARMA en el estudio de las fas y fap muestrales. Sin embargo, esta metodología tiene distintos problemas:

- Las fas y fap muestrales son una herramienta útil para identificar procesos AR o MA puros, pero son difíciles de interpretar en la identificación de procesos ARMA mixtos.
- Las fas y fap muestrales pueden llevarnos a identificar distintos procesos como posibles generadores de la serie de tiempo. En ese caso, necesitamos un criterio que nos permita determinar cuál de ellos es preferible.

Resulta, pues, evidente que la etapa correspondiente a la identificación de los órdenes p y q debe ser revisada.

Procesos ARMA: Selección del modelo

Denotemos por k a la cantidad de coeficientes de un modelo ARMA (esto es, $k = p + q + 1$ o $k = p + q$ para un ARMA(p,q) con o sin constante, respectivamente), y por φ_{k+1} al vector formado por dichos coeficientes y por σ_a^2.

Se propone seleccionar aquel modelo ARMA que minimice el valor de una de las funciones siguientes:

- $AIC = -2ln\left(L(\varphi_{k+1})\right) + 2\,(k+1)$ (Criterio de Información de Akaike).
- $AICC = -2ln\left(L(\varphi_{k+1})\right) + 2\,(k+1)\;T/\,(T-k-2)$ (Criterio de Información de Akaike corregido).
- $BIC = -2ln\left(L(\varphi_{k+1})\right) + (k+1)\,ln\,(T)$ (Criterio de Información Bayesiano).

Procesos ARMA: Selección del modelo

La estructura de las funciones AIC, AICC y BIC es similar:

- Primer sumando: mide tanto la calidad del ajuste como la credibilidad que le da a la serie de tiempo. Cuanto menor es su valor mejor es el ajuste y mayor la credibilidad que le da a la serie; su valor disminuye al aumentar k.
- Segundo sumando: penaliza el aumento en la cantidad de coeficientes del ARMA; su valor disminuye al disminuir k.

El modelo que minimiza a una de estas 3 funciones consigue un equilibrio entre ambos sumandos (ambos serán “pequeños”); esto es, un buen ajuste sin demasiados parámetros (que darían problemas tanto a la hora de estimar como de predecir).
Las estimaciones de los parámetros del modelo seleccionado por este proseso son estimaciones de máxima verosimilitud.

Procesos ARMA: Selección del modelo

Comparación entre criterios

- El criterio BIC es consistente: Si realmente la serie ha sido generada por un ARMA, el BIC selecciona los órdenes correctos con probabilidad 1 (esto no ocurre con los criterios AIC y AICC).
- Los criterios AIC y AICC son asintóticamente eficientes: Si realmente la serie ha sido generada por un AR (posiblemente de orden ∞), el AIC y el AICC seleccionan el modelo que da lugar al menor error de predicción esperado (esto no ocurre con el criterio BIC).

Procesos ARMA: Selección del modelo

Continuando con el análisis de la serie (un proceso AR(2) resultaba adecuado para haberla generado), minimizamos el valor de la función AIC para distintos procesos ARMA(p,q) (consideramos p y q en $\{0, 1, 2, 3, 4, 5\}$).

Los 3 valores más pequeños fueron alcanzados por el ARMA(3,3) ($AIC(3, 3) = 264.6635$), por el ARMA(1,1) ($AIC(1, 1) = 264.9980$) y por el AR(2) ($AIC(2, 0) = 265.9452$).

Puesto que el AR(2) es un modelo mucho más simple que el ARMA(3,3) y más fácil de interpretar que el ARMA(1,1), y la diferencia entre los valores de sus AICs es pequeña (menor de 2 unidades), el modelo AR(2) es el que finalmente seleccionamos.

Modelos Box-Jenkins

Procesos ARMA: Predicción

Supongamos que la serie de tiempo $x_1, \ldots, x_T$ ha sido generada por un proceso ARMA $\{X_t\}_t$ cuyos parámetros son conocidos (recuérdese que en la clase ARMA se incluyen los ARMA regulares (ARMA(p,q)), los ARMA estacionales (ARMA$(P,Q)_s$) y los ARMA estacionales multiplicativos (ARMA$(p,q)\times(P,Q)_s$)).

El siguiente objetivo de este tema es predecir, a partir de la serie de tiempo observada, el valor futuro del proceso dentro de k instantes de tiempo; esto es, predecir el valor de X_{T+k}.

Dicha predicción se denomina predicción con origen en T y horizonte k, y la denotaremos por $\widehat{x}_T\left(k\right)$.

Predicción: Procesos AR(1)

Supongamos que la serie ha sido generada por el proceso

$$X_t = 7.5 + 0.8X_{t-1} + a_t.$$

Para predecir con origen en T y horizonte 1, escribimos

$$X_{T+1} = 7.5 + 0.8X_T + a_{T+1}.$$

Basándonos en la relación anterior, $\widehat{x}_T(1)$ será obtenida una vez que dispongamos de valores "apropiados" para X_T y a_{T+1}.

- El valor que ha tomado la v.a. X_T es conocido (x_T).
- El valor de la v.a. a_{T+1} no lo tenemos. Su predicción a partir de la serie $x_1, \ldots, x_T$ es su media ($E(a_{T+1}) = 0$), pues la serie no contiene información acerca de a_{T+1}.

Predicción: Procesos AR(1)

Ahora, para obtener la predicción $\widehat{x}_T(1)$, basta con sustituir X_T por x_T y a_{T+1} por 0 en la ecuación

$$X_{T+1} = 7.5 + 0.8X_T + a_{T+1}.$$

Dicha sustitución da lugar a la ecuación de predicción del AR(1) a un horizonte 1:

$$\widehat{x}_T(1) = 7.5 + 0.8x_T.$$

Así, si el valor de x_T fuese 1.5, la predicción del futuro valor de X_{T+1} sería

$$\widehat{x}_T(1) = 7.5 + 0.8 \times 1.5 = 8.7$$

Predicción: Procesos AR(1)

Para predecir con origen en T y horizonte 2, escribimos

$$X_{T+2} = 7.5 + 0.8X_{T+1} + a_{T+2}.$$

Basándonos en la relación anterior, $\widehat{x}_T(2)$ será obtenida una vez que dispongamos de valores "apropiados" para X_{T+1} y a_{T+2}.

- El valor de la v.a. X_{T+1} no lo tenemos. Su predicción a partir de la serie $x_1, \ldots, x_T$ ha sido obtenida en la transparencia anterior ($\widehat{x}_T(1)$).
- El valor de la v.a. a_{T+2} no lo tenemos. Su predicción a partir de la serie $x_1, \ldots, x_T$ es su media ($E(a_{T+2}) = 0$).

Predicción: Procesos AR(1)

Ahora, para obtener la predicción $\widehat{x}_T(2)$, basta con sustituir X_{T+1} por $\widehat{x}_T(1)$ y a_{T+2} por 0 en la ecuación

$$X_{T+2} = 7.5 + 0.8X_{T+1} + a_{T+2}.$$

Dicha sustitución da lugar a la ecuación de predicción del AR(1) a un horizonte 2:

$$\widehat{x}_T(2) = 7.5 + 0.8\widehat{x}_T(1) = 7.5(1 + 0.8) + 0.8^2 x_T$$

Así, si el valor de x_T fuese 1.5, la predicción del futuro valor de X_{T+2} sería

$$\widehat{x}_T(2) = 14.46$$

Predicción: Procesos A(p)

El procedimiento anterior, diseñado para predecir valores de un AR(1) a horizontes $k = 1$ y $k = 2$, puede generalizarse fácilmente para:

- Cualquier horizonte $k > 0$.
- Cualquier proceso autorregresivo de orden finito $p > 0$.

Puede demostrarse que la predicción a largo plazo de futuros valores de un proceso AR(p) es la media del proceso; esto es,

$$\widehat{x}_T(k) \to \mu \text{ si } k \to \infty.$$

Predicción: Procesos MA(1)

Supongamos que la serie ha sido generada por el proceso

$$X_t = 7.5 + a_t + 0.8a_{t-1}.$$

Para predecir con origen en T y horizonte 1, escribimos

$$X_{T+1} = 7.5 + a_{T+1} + 0.8a_T.$$

Basándonos en la relación anterior, $\widehat{x}_T(1)$ será obtenida una vez que dispongamos de valores "apropiados" para a_{T+1} y a_T.

- El valor de la v.a. a_{T+1} no lo tenemos. Su predicción a partir de la serie $x_1, \ldots, x_T$ es su media ($E(a_{T+1}) = 0$).
- El valor de la v.a. a_T no lo tenemos. Su predicción a partir de la serie $x_1, \ldots, x_T$ no es inmediata, pues dicha serie contiene información acerca de a_T.

Predicción: Procesos MA(1)

Supongamos que disponemos de una predicción del valor de a_T obtenida a partir de la serie $x_1, \ldots, x_T$ (en las próximas transparencias veremos cómo obtenerla). Denotémosla por $\widehat{a}_T(0)$.

Ahora, para obtener la predicción $\widehat{x}_T(1)$, basta con sustituir a_{T+1} por 0 y a_T por $\widehat{a}_T(0)$ en la ecuación

$$X_{T+1} = 7.5 + a_{T+1} + 0.8a_T.$$

Dicha sustitución da lugar a la ecuación de predicción del MA(1) a un horizonte 1:

$$\widehat{x}_T(1) = 7.5 + 0.8\widehat{a}_T(0).$$

Predicción: Procesos MA(1)

Para predecir con origen en T y horizonte 2, escribimos

$$X_{T+2} = 7.5 + a_{T+2} + 0.8a_{T+1}.$$

Basándonos en la relación anterior, $\widehat{x}_T(2)$ será obtenida una vez que dispongamos de valores "apropiados" para a_{T+2} y a_{T+1}.

- Los valores de las v.a. a_{T+2} y a_{T+1} no los tenemos. Sus predicciones a partir de la serie $x_1, \ldots, x_T$ son sus medias ($E(a_{T+2}) = E(a_{T+1}) = 0$).

Sustituyendo entonces a_{T+2} y a_{T+1} por 0 en la ecuación anterior, se obtiene la ecuación de predicción del MA(1) a un horizonte 2: $\widehat{x}_T(2) = 7.5$.

Predicción: Procesos MA(1)

Existen distintas técnicas para predecir el valor de a_T a partir de la serie $x_1, \ldots, x_T$:

- Utilizar la propiedad de invertibilidad del proceso:
$$X_t = c + a_t + \pi_1 X_{t-1} + \pi_2 X_{t-2} + \cdots$$

  Cambiando t por T y despejando a_T, podemos escribir
$$a_T = -c + X_T - \pi_1 X_{T-1} - \cdots - \pi_{T-1} X_1 - \pi_T X_0 - \cdots$$
  Considerando que $X_t = x_t$ si $1 \leq t \leq T$, y fijando $X_t = \mu$ si $t \leq 0$, se obtiene una predicción del valor de a_T a partir de la serie $x_1, \ldots, x_T$.
  Cuando T es "grande", es esperable que la modificación realizada a los valores de X_t $(t \leq 0)$ tenga poco impacto sobre los resultados finales.

Modelos Box-Jenkins

Predicción: Procesos MA(1)

- Fijar un valor inicial para a_1 (en general, su media 0).

  Suponiendo que el MA(1) es
  $$X_t = 7.5 + a_t + 0.8a_{t-1},$$
  se tiene que
  $$a_t = X_t - 7.5 - 0.8a_{t-1}.$$
  Para obtener a_T a partir de la expresión anterior, necesitamos conocer un a_t con $1 \leq t < T$. Así, fijando el valor de a_1, podemos predecir recursivamente los valores de a_t $(t = 2, \ldots, T)$ a partir de la serie $x_1, \ldots, x_T$. Cuando T es "grande", es esperable que la condición inicial (generalmente, $a_1 = 0$) tenga poco impacto sobre los resultados finales.

Predicción: Procesos MA(1)

- Predecir a_T a través de una ("la mejor") combinación lineal de $x_1, \ldots, x_T$.

  Los coeficientes de la combinación lineal de $x_1, \ldots, x_T$ que mejor predice a a_T se pueden obtener a partir de la media y las autocovarianzas de $X_1, \ldots, X_T$, y de las covarianzas entre cada una de estas variables y a_T.

  Este procedimiento es el que sigue el paquete R.

Predicción: Procesos MA(q)

El procedimiento anterior, diseñado para predecir valores de un MA(1) a horizontes $k = 1$ y $k = 2$, puede generalizarse fácilmente para:

- Cualquier horizonte $k > 0$.
- Cualquier proceso de medias móviles de orden finito $q > 0$.

Puede demostrarse que la predicción a horizontes mayores que q de futuros valores de un proceso MA(q) es la media del proceso; esto es,

$$\widehat{x}_T(k) = \mu \text{ si } k > q.$$

Predicción: Procesos ARMA(p,q)

Los procedimientos utilizados para predecir valores futuros de procesos AR(p) y MA(q) pueden ser combinados fácilmente para predecir valores futuros de procesos ARMA(p,q).

Puede demostrarse que la predicción a largo plazo de futuros valores de un proceso ARMA(p,q) es la media del proceso; esto es,

$$\widehat{x}_T(k) \rightarrow \mu \text{ si } k \rightarrow \infty.$$

Procesos ARMA: Intervalos de predicción

Para finalizar, construimos un intervalo de predicción utilizando la distribución muestral del error de predicción:

$$e_T(k) = X_{T+k} - \widehat{X}_T(k).$$

Se verifica que, si "T es grande" y $\{a_t\}_t$ es gaussiano, entonces

$$e_T(k) \approx N\left(0, \sigma_a^2\left(1 + \psi_1^2 + \cdots + \psi_{k-1}^2\right)\right),$$

donde ψ_i son los coeficientes de la representación

$$X_t = c + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots$$

Un intervalo de predicción (al 95%) para el valor de X_{T+k} será

$$\left(\widehat{X}_T(k) \pm 1.96\sqrt{\sigma_a^2\left(1 + \psi_1^2 + \cdots + \psi_{k-1}^2\right)}\right).$$

Procesos ARMA: Intervalos de predicción

Observación

En el cálculo de la varianza asintótica del error de predicción,

$$Var\left(X_{T+k} - \widehat{X}_T\left(k\right)\right) \approx \sigma_a^2\left(1 + \psi_1^2 + \cdots + \psi_{k-1}^2\right),$$

las estimaciones de los parámetros incluidos en la predicción $\widehat{X}_T\left(k\right)$ fueron tratadas como fijas (esto es, como si no dependiesen de los valores del proceso estocástico $\{X_t\}$).

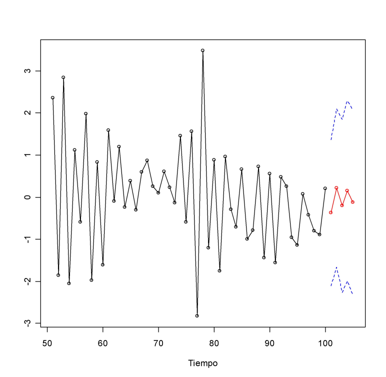
Sin embargo, puesto que dichas estimaciones dependen de la serie $x_1, \ldots, x_T$, también dependen de los valores de $\{X_t\}$.

Esto provoca un cambio en la varianza del error de predicción, y por tanto en los intervalos de predicción. Este cambio se puede considerar despreciable si el tamaño de la serie es "grande".

Modelos Box-Jenkins

Para finalizar el estudio de la serie, realizamos predicciones basadas en el modelo AR(2) que habíamos seleccionado.

Horizonte de predicción: k=5



Horizonte de predicción: k=30