



**FACULTAD
DE
CIENCIAS**

**INTRODUCCION A SERIES DE TIEMPO
DESDE UN PUNTO DE VISTA BAYESIANO**

T E S I S

QUE PARA OBTENER EL TITULO DE

M A T E M A T I C O

P R E S E N T A

ALEJANDRO ALEGRIA HERNANDEZ

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO



UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso

DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

I N D I C E

Introducción

1. PROCESOS ESTOCASTICOS

1.1. Conceptos generales	4
1.2. Procesos estacionarios	9
1.3. Propiedades de la función de autocorrelación y de la función de autocovarianzas	12
1.4. Proceso lineal general	14
1.5. Proceso autorregresivo	19
1.6. Proceso de promedios móviles	22
1.7. Proceso autorregresivo y de promedios móviles . . .	25
1.8. Procesos no estacionarios	29
1.9. Procesos estacionales	31

2. INFERENCIA

2.1. Conceptos generales	32
2.2. Proceso autorregresivo	38
2.3. Extensión a otros procesos	74
2.4. Proceso de promedios móviles	81
2.5. Proceso autorregresivo y de promedios móviles . . .	84
2.6. Distribución inicial informativa	85

3. TOPICOS RELACIONADOS

3.1. Condición de estacionariedad	96
3.2. Análisis usando otras distribuciones iniciales . .	100
3.3. Representaciones alternativas	104
3.4. Identificación de modelos	108
3.5. Decisiones estadísticas	114

4. CONCLUSIONES

Apéndice A	119
Apéndice B	125
Referencias	126

Introducción

La mayoría de los procedimientos estadísticos están diseñados para ser usados con información muestral consistente en un conjunto de observaciones independientes. Con este tipo de información, el orden en el cual se presenta cada dato es irrelevante para el estadístico. Sin embargo, no es difícil imaginar situaciones en donde los datos disponibles aparecen ordenados de acuerdo a algún parámetro (usualmente el tiempo), en cuyo caso se tendrá un conjunto de observaciones dependientes o serie de tiempo, siendo entonces el interés primordial el estudio de la dependencia presente.

Una serie de tiempo observada puede ser considerada como la realización de algún proceso teórico, el cual es llamado proceso estocástico. El objetivo inicial en el análisis de una serie de tiempo es hacer inferencias sobre las características básicas del proceso a partir de la información contenida en la serie observada. Para comprender más fácilmente el comportamiento de algún proceso estocástico, se propone un modelo que se espera tenga propiedades similares a las de éste, así que lo importante será el poder realizar inferencias sobre los parámetros que definen al modelo.

Dependiendo del interés que se tenga, una serie de tiempo se puede analizar en el dominio de tiempo o en el dominio de frecuencias. El primer tipo de análisis estudia las relaciones

existentes entre los valores observados de la serie, mientras que el segundo se enfoca, exclusivamente, a los movimientos cíclicos de la serie. En este trabajo sólo se hará el análisis en el dominio de tiempo, aunque algunos comentarios con respecto al análisis en el dominio de frecuencias se hacen en el capítulo 3.

En el capítulo 1 de este trabajo se presentan la definición y las principales características de una clase de procesos estocásticos llamados procesos estocásticos estacionarios. Al mismo tiempo se proponen algunos modelos pertenecientes al conjunto de los modelos conocidos como ARMA, los cuales han resultado ser bastante eficientes en la representación de procesos estacionarios. Las principales propiedades de cada modelo son establecidas con el fin de tener una caracterización adecuada del mismo.

El análisis estadístico, para tres de los modelos presentados en el primer capítulo, es realizado en el capítulo 2. Se decidió realizar el análisis estadístico con un enfoque bayesiano, pues se considera que es la única forma razonable y teóricamente bien sustentada de incorporar al análisis toda la información disponible. El problema de inferencia, desde el punto de vista bayesiano, es encontrar y estudiar la distribución final sobre los parámetros de interés. El capítulo 2 trata exclusivamente sobre este problema.

En el capítulo 3, se presentan algunos aspectos inferencia

les no tratados en el capítulo anterior, relacionados básicamente con la representación de la información inicial. Así mismo se plantea, en términos generales, la utilización de otros modelos para el análisis de series de tiempo, el problema de identificación, prueba de hipótesis, estimación y predicción puntual.

A manera de conclusión, en el capítulo 4 se plantean algunos de los problemas encontrados en el desarrollo del tema. Debe entenderse que en este trabajo no se pretendió, en ningún momento, comparar el procedimiento bayesiano de análisis con los métodos clásicos o frecuentistas.

Finalmente, se incluyen dos apéndices al final del trabajo. En el primero de ellos se definen las distribuciones que fueron utilizadas y en el segundo se dá la lista de las observaciones que fueron empleadas en el ejemplo de la sección 2.2.

CAPITULO 1

PROCESOS ESTOCASTICOS ESTACIONARIOS

En este capítulo se introduce el concepto de proceso estocástico estacionario y se dan algunas de sus propiedades más importantes. Asimismo se definen varios modelos, los cuales pretenden explicar el comportamiento de ciertos procesos. Cada modelo es caracterizado y al mismo tiempo se presenta su relación con los demás.

1.1. Conceptos fundamentales

Consideremos el espacio de probabilidad (Ω, \mathcal{A}, P) , en donde Ω es un conjunto de eventos elementales, \mathcal{A} es un σ -álgebra de subconjuntos de Ω y P es una medida de probabilidad definida en \mathcal{A} . Sea I un conjunto de índices, entonces un proceso estocástico con valores reales es una función $X : I \times \Omega \rightarrow \mathbb{R}$, tal que para todo $t \in I$ fijo, $X(t, \omega)$, $\omega \in \Omega$, es una variable aleatoria real definida en el espacio (Ω, \mathcal{A}, P) . Es común denotar a $X(t, \omega)$ como $X_t(\omega)$ ó bien X_t , y referirse a la colección de variables aleatorias $\{X_t : t \in I\}$ como una serie de tiempo.

Para un $\omega \in \Omega$ fijo, $X(t, \omega)$ es una función real que depende sólo de t , y a esta función se le conoce como una realización del proceso. La colección de todas las posibles

realizaciones es llamada el ensamble.

Dependiendo de si el conjunto de índices I es discreto o continuo, se tiene un proceso estocástico (con parámetro de tiempo) discreto o continuo, respectivamente.

El análisis de una serie de tiempo tiene como objetivo el estudiar la estructura probabilística del conjunto de variables aleatorias $\{X_t(\omega) : t \in I\}$, y con este propósito es común que la información muestral con que se cuente sea la dada por un conjunto finito de T datos, resultado de haber observado, dentro del conjunto $\{X_t(\omega) : t \in I\}$, T variables aleatorias consecutivas e igualmente espaciadas con respecto al índice t .

Un proceso estocástico, considerado como una sucesión infinita de variables aleatorias, induce una medida de probabilidad definida en un espacio de dimensión infinita. A los conjuntos de dicho espacio que tienen asignada una probabilidad se les llama conjuntos medibles. Ahora bien, como se mencionó en el párrafo anterior, en la práctica sólo se esta en posibilidad de estudiar un número finito de variables aleatorias. En particular, será de interés conocer la función de x_1, x_2, \dots, x_n , definida por $P\{X_{t_1}(\omega) \leq x_1 : i = 1, 2, \dots, n\}$, así como las probabilidades derivadas de la misma. De aquí la conveniencia de que la clase de conjuntos medibles incluya a los conjuntos de la forma $\{X_{t_1}(\omega) \leq x_1, X_{t_2}(\omega) \leq x_2, \dots, X_{t_n}(\omega) \leq x_n\}$, pues de esta manera se puede definir la función de distribución

de probabilidad dada por

$$F_{X_{t_1}, X_{t_2}, \dots, X_{t_n}}(x_1, x_2, \dots, x_n) = P\{X_{t_1}(\omega) \leq x_1, X_{t_2}(\omega) \leq x_2, \dots, X_{t_n}(\omega) \leq x_n\}.$$

Inversamente, si se dan las funciones de distribución de probabilidad para cada conjunto finito de índices, y éstas cumplen los supuestos dados por el teorema de consistencia de Kolmogorov (1950), se puede asegurar la existencia de una medida de probabilidad definida en el espacio de dimensión infinita, tal que la medida asignada a conjuntos de la forma

$\{X_{t_1}(\omega) \leq x_1, X_{t_2}(\omega) \leq x_2, \dots, X_{t_n}(\omega) \leq x_n\}$, coincide con la probabilidad calculada a partir de las funciones de distribución.

Ciertas cantidades resultan muy útiles en la caracterización del comportamiento de un proceso estocástico, algunas de éstas se darán enseguida.

La media y la varianza de un proceso estocástico al tiempo t se definen, respectivamente, como

$$\mu_t = E(X_t) \quad , \quad \sigma_t^2 = V(X_t) = E\{(X_t - \mu_t)^2\}$$

La función de autocovarianzas se define como

$$\sigma(X_s, X_t) = E\{(X_s - \mu_s)(X_t - \mu_t)\} \quad , \quad \forall s, t \in I,$$

en particular, es claro que $\sigma_t^2 = \sigma(X_t, X_t)$.

Finalmente, la función de autocorrelación está dada por

$$\rho(X_s, X_t) = \frac{\sigma(X_s, X_t)}{\sigma_s \sigma_t}, \quad \forall s, t \in I.$$

Si el proceso bajo estudio se supone gaussiano, lo cual significa que $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$ sigue una distribución normal n -variada (para toda $n \in \mathbb{N}$), entonces las medias y la función de autocovarianzas son suficientes para tener una caracterización completa de las propiedades distribucionales del proceso. Si no se supone normalidad pero el proceso es lineal, en el sentido de que X_t es combinación lineal de anteriores X_t 's y/o de valores pasados y presente de otro proceso, son las medias y las autocovarianzas las cantidades que permiten conocer la mayoría de las propiedades del proceso.

Como se dijo en un inicio, la información muestral que se pudiera tener para el estudio del comportamiento de un proceso estocástico, consiste de sólo una parte finita de alguna realización del proceso. Si con este tipo de información lo que se pretende es hacer inferencias, es necesario suponer que el proceso es ergódico, lo cuál significa intuitivamente que para tiempos suficientemente alejados hay independencia entre los valores que pueda tomar el proceso. Este supuesto permite asegurar que, conforme se tengan más observaciones, efectivamente se está aprendiendo más sobre el proceso (de hecho, se están cumpliendo ciertas convergencias en probabilidad). En este trabajo se su-

pondrá siempre que el proceso de interés es ergódico. Algunos teoremas ergódicos se presentan en Gnedenko (1969).

1.2. Procesos estacionarios.

Una clase interesante de procesos estocásticos es la formada por los procesos llamados estacionarios, que tienen como principal característica la de presentar cierto tipo de equilibrio en su comportamiento. Este equilibrio se puede observar a dos niveles, lo que dá lugar a los conceptos de estacionariedad estricta y estacionariedad débil.

Estacionariedad estricta.

Un proceso estocástico discreto* $\{X_t: t \in I\}$ se dice que es estrictamente estacionario, si la función de distribución conjunta de $\{X_{t_1}, X_{t_2}, \dots, X_{t_n}\}$ es la misma que la de $\{X_{t_1+k}, X_{t_2+k}, \dots, X_{t_n+k}\}$, para toda n y k , es decir, si

$$F_{X_{t_1}, X_{t_2}, \dots, X_{t_n}}(x_1, x_2, \dots, x_n) = F_{X_{t_1+k}, X_{t_2+k}, \dots, X_{t_n+k}}(x_1, x_2, \dots, x_n)$$

$$\forall x_n \in \mathbb{R}^n .$$

Entonces, las propiedades de un proceso estocástico estrictamente estacionario no se ven afectadas por un cambio de origen en el tiempo. En particular, F_{X_t} es la misma para todo t , por lo que el proceso tendrá media y varianzas constantes dadas respectivamente por

*En lo sucesivo se considerarán sólo procesos discretos, a menos que se especifique otra cosa. El caso continuo es tratado en forma similar. Ver ejemplo, Cox y Miller (1965), Gredenko (1969).

$$\mu = \int x_t dF_{X_t}(x_t) < \infty, \quad \sigma^2 = \int (x_t - \mu)^2 dF_{X_t}(x_t) < \infty.$$

La existencia de estas cantidades se puede asegurar si $E(X_t^2) < \infty$.

Así mismo, como la función de distribución de (X_s, X_t) es la misma para todos $s, t \in I$ tales que $|s - t|$ es una constante, la covarianza entre X_t y X_{t+k} sólo dependerá de la distancia k en el tiempo que hay entre X_t y X_{t+k} , pues

$$\sigma(X_t, X_{t+k}) = E\{(X_t - \mu)(X_{t+k} - \mu)\} = E(X_t, X_{t+k}) - \mu^2.$$

Entonces la función de autocovarianzas se denotará simplemente como $\sigma(k)$. En este caso también la función de autocorrelación dependerá exclusivamente de la distancia k en el tiempo, ya que

$$\rho(k) = \rho(X_t, X_{t+k}) = \frac{\sigma(X_t, X_{t+k})}{\sigma_t \sigma_{t+k}} = \frac{\sigma(k)}{\sigma^2}.$$

Estacionariedad débil.

Estacionariedad estricta implica trabajar con la distribución de probabilidad conjunta $X_{t_1}, X_{t_2}, \dots, X_{t_n}$ para cada n finita, lo cual puede resultar bastante complicado. Por esta razón se introduce el concepto de estacionariedad

débil, llamada también de segundo orden.

Un proceso estocástico $\{X_t: t \in I\}$ es llamado débilmente estacionario si $E(X_t)$ existe y es constante para toda t y además

$$\text{COV}(X_s, X_t) = \text{COV}(X_{s+k}, X_{t+k})$$

para todo s, t y k .

Obviamente todo proceso estrictamente estacionario con los dos primeros momentos finitos es débilmente estacionario, no cumpliéndose el resultado inverso excepto en el caso de trabajar con procesos gaussianos, pues la distribución normal queda totalmente especificada por su primero y segundo momentos.

1.3. Propiedades de $\sigma(k)$ y $\rho(k)$.

Algunas de las propiedades más importantes de las funciones de autocovarianza y de autocorrelación en procesos estacionarios son las siguientes:

P1. La función de autocovarianzas es una función par (con respecto a cero), es decir $\sigma(k) = \sigma(-k)$. (Es inmediato que también $\rho(k)$ es una función par).

P2. La matriz $\Gamma = (\sigma(t_i - t_j))$ de varianzas y covarianzas del conjunto de variables $\{X_{t_1}, X_{t_2}, \dots, X_{t_n} : t_i \in I\}$ es definida semipositiva.

P3. La función de autocorrelación es tal que

$$|\rho(k)| \leq 1, \quad \forall k.$$

P4. Si la función de autocorrelación es absolutamente sumable, es decir, si

$$\sum_{-\infty}^{\infty} |\rho(k)| < \infty,$$

entonces existe una única función $f(\omega) \geq 0$, tal que

$$a) \quad \rho(k) = \int_{-\pi}^{\pi} f(\omega) e^{i\omega k} d\omega$$

$$b) \quad \int_{-\pi}^{\pi} f(\omega) d\omega = 1$$

$$c) \quad f(\omega) = f(-\omega)$$

$f(\omega)$ es denominada la función de densidad espectral.

P5. La relación inversa entre $f(\omega)$ y $\rho(k)$ está dada por

$$f(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \rho(k) e^{-i\omega k} \quad , \quad -\pi < \omega < \pi \quad ,$$

que a su vez se reduce, al trabajar con procesos en \mathbb{R} , a la expresión

$$f(\omega) = \frac{1}{\pi} \left[1 + 2 \sum_{k=1}^{\infty} \rho(k) \cos \omega k \right] \quad , \quad 0 < \omega < \pi \quad .$$

Más información sobre las funciones de autocovarianza y de densidad espectral se pueden ver en Anderson (1971), Fuller (1976) y en las referencias allí citadas.

1.4. Proceso lineal general

Los modelos que más comunmente se usan en el análisis de series de tiempo forman parte de una clase más general de modelos lineales, en los cuales se parte del hecho de que una serie de tiempo $\{X_t : t \in \mathbb{Z}\}$ (de ahora en adelante y por facilidad de notación el conjunto de índices I será \mathbb{Z}), puede ser vista como generada por una serie de "choques" independientes representados por $\{Z_t : t \in \mathbb{Z}\}$. Se considera que estos "choques" son aleatorios y, por lo general, con media 0 y varianza común σ_Z^2 . A éste tipo de proceso se le conoce como ruido blanco.

Entonces, el valor del proceso al tiempo t se expresa como

$$X_t = \sum_{j=0}^{\infty} c_j Z_{t-j}, \quad c_0 = 1, \quad (1.4.1)$$

o equivalentemente,

$$X_t = [C(B)] Z_t,$$

en donde B es un operador de retraso definido como

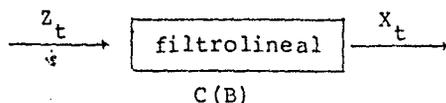
$$B^k(Z_t) = Z_{t-k}, \quad \forall k \in \mathbb{N} \quad \text{con} \quad B^0(Z_t) = Z_t,$$

y $C(Z)$ es el polinomio

$$C(Z) = \sum_{i=0}^{\infty} c_i Z^i, \quad c_0 = 1.$$

El operador $C(B)$ es la llamada función de transferencia

que transforma el ruido blanco en la serie $\{X_t : t \in \mathbb{Z}\}^*$,



Las condiciones bajo las cuales el proceso lineal definido anteriormente es estacionario (en todo lo que sigue debe entenderse débilmente estacionario), se establecerán a continuación.

De acuerdo a (1.4.1) se puede comprobar que

$$E(X_t) = E(Z_t) \sum_{j=0}^{\infty} c_j, \quad \sigma(k) = \sigma_Z^2 \sum_{j=0}^{\infty} c_j c_{j+k}.$$

En particular, la varianza del proceso será

$$\sigma_X^2 = \sigma(0) = \sigma_Z^2 \sum_{j=0}^{\infty} c_j^2.$$

La condición de estacionariedad se cumple si cada una de las siguientes sumas es convergente,

$$\sum_{j=0}^{\infty} c_j, \quad \sum_{j=0}^{\infty} c_j^2, \quad \sum_{j=0}^{\infty} c_j c_{j+k}, \quad \forall k.$$

Estas condiciones se pueden resumir en una sola, pidiendo que $C(z)$ convenga para todo $z \in \mathbb{E}$ tal que $||z|| \leq 1$.

(Ver Box y Jenkins (1970), Anderson (1971)).

* En realidad se puede considerar que $X_t = W_t - \mu$, en donde $\{W_t : t \in \mathbb{Z}\}$ es el proceso original y μ es una constante, que se toma como referencia o es la media del proceso si hay estacionariedad.

A partir del modelo (1.4.1), también es posible expresar X_t en términos de sus valores pasados,

$$X_t = \sum_{j=1}^{\infty} d_j X_{t-j} + z_t \quad (1.4.2)$$

o alternativamente,

$$[D(B)] X_t = z_t ,$$

en donde B es el operador de retraso definido anteriormente y $D(Z)$ es el polinomio

$$D(Z) = 1 - \sum_{i=1}^{\infty} d_i z^i .$$

Habiendo expresado el comportamiento de X_t en dos formas equivalentes, a saber

$$X_t = [C(B)] z_t , \quad [D(B)] X_t = z_t ,$$

es claro que

$$[D(B)] [C(B)] z_t = z_t ,$$

en donde

$$D(B) = C^{-1}(B) ,$$

por lo tanto la condición de estacionariedad, en términos del modelo (1.4.2), será también que $D(Z)$ converga para todo $Z \in \mathbb{C}$ tal que $||Z|| \leq 1$.

La función de autocovarianza del proceso (1.4.1) está dada por

$$\sigma(k) = E(X_t X_{t+k}) = \sigma_z^2 \sum_{j=0}^{\infty} c_j c_{j+k} ,$$

de donde la función de autocorrelación será

$$\rho(k) = \frac{\sum_{j=0}^{\infty} c_j c_{j+k}}{\sum_{j=0}^{\infty} c_j^2} .$$

Por último, la función de densidad espectral, de acuerdo a la propiedad P5 dada en la sección 1.3, es

$$f(\omega) = \frac{\sigma_z^2}{\sigma_x^2 2\pi} \sum_{k=-\infty}^{\infty} \sum_{j=0}^{\infty} c_j c_{j+k} e^{-i\omega k} .$$

Una forma alternativa de expresar $f(\omega)$ se obtiene al hacer $h = j + k$ y considerar que $c_h = 0$ para $h < 0$, con lo cual

$$\begin{aligned} f(\omega) &= \frac{\sigma_z^2}{\sigma_x^2 2\pi} \left[\sum_{h=0}^{\infty} c_h e^{i\omega h} \right] \left[\sum_{j=0}^{\infty} c_j e^{-i\omega j} \right] = \\ &= \frac{\sigma_z^2}{\sigma_x^2 2\pi} c(e^{i\omega}) c(e^{-i\omega}) = \frac{\sigma_z^2}{\sigma_x^2 2\pi} |c(e^{-i\omega})|^2 , \end{aligned}$$

$$-\pi < \omega < \pi . \quad (1.4.3)$$

En la práctica no es muy útil trabajar con representacioo

nes del proceso lineal tales como (1.4.1) o (1.4.2), pues contienen un número infinito de parámetros. Es por esta razón que se introducen modelos más sencillos, en términos del número de parámetros que se involucran, sin que por esto dejen de ser potencialmente útiles. Esto último se conoce, en la construcción de modelos, como parsimonia.

En las siguientes secciones de este capítulo se presentarán varios modelos y algunas de sus características más importantes.

1.5. Proceso autorregresivo.

Como un caso particular del modelo (1.4.2) se tiene el modelo autorregresivo de orden p , denotado por $AR(p)$ y el cual se escribe como

$$X_t = \sum_{j=1}^p \beta_j X_{t-j} + Z_t, \quad (1.5.1)$$

o alternativamente,

$$[A(B)]X_t = Z_t,$$

en donde $A(Z) = 1 - \sum_{j=1}^p \beta_j Z^j$.

La función de transferencia en este caso es $A^{-1}(B)$, y el proceso es estacionario, de acuerdo a lo visto en la sección anterior, si $A^{-1}(Z)$ converge para todo $Z \in \mathbb{C}$, tal que $\|Z\| \leq 1$ o, equivalentemente, si los ceros de $A(Z)$ están fuera del círculo unitario. $A(Z)$ es denominada la ecuación característica del proceso.

Supongamos que Z_i^{-1} , $i = 1, 2, \dots, p$, son las raíces de $A(Z)$, entonces

$$A^{-1}(Z) = \frac{1}{A(Z)} = \frac{1}{\prod_{i=1}^p (1 - Z_i Z)} = \prod_{i=1}^p \frac{1}{1 - Z_i Z} = \prod_{i=1}^p \sum_{k=0}^{\infty} (Z_i Z)^k, \quad (1.5.2)$$

de donde la convergencia de $A^{-1}(Z)$ para todo $Z \in \mathbb{C}$, con $\|Z\| \leq 1$, implica que $\|Z_i\| > 1$, es decir $\|Z_i^{-1}\| > 1$, $i = 1, 2, \dots, p$.

A partir de (1.5.2) se puede afirmar que la estacionariedad de un proceso AR(p), permite expresar a éste como la suma infinita

$$X_t = [A(B)^{-1}]z_t = \sum_{j=0}^{\infty} \delta_j z_{t-j}.$$

La función de autocorrelación en un proceso AR(p) es encontrada multiplicando ambos lados de (1.5.1) por X_{t-k} , tomando esperanza (sin pérdida de generalidad se supone que $E(X_t) = 0$) y dividiendo entre σ_X^2 , obteniendo así la ecuación en diferencias

$$\rho(k) = \beta_1 \rho(k-1) + \beta_2 \rho(k-2) + \dots + \beta_p \rho(k-p), \quad (1.5.3)$$

$$\forall k > 0.$$

Si en esta última ecuación se sustituye $k = 0, 1, \dots, p$ se tienen las conocidas ecuaciones de Yule-Walker.

La solución general a (1.5.3) está dada por

$$\rho(k) = E_1 z_1^k + E_2 z_2^k + \dots + E_p z_p^k, \quad \forall k > 0, \quad (1.5.4)$$

en donde z_i^{-1} , $i = 1, 2, \dots, p$ son las raíces de $A(z)$ y las constantes E_1, E_2, \dots, E_p son elegidas de tal forma que las restricciones impuestas por las ecuaciones de Yule-Walker se cumplan.

La conducta de la función de autocorrelación se puede estudiar a partir de (1.5.4), pues dependiendo del carácter real o complejo de las raíces de $A(z)$ se tendrán componentes en $\rho(k)$ que decaigan exponencialmente o en forma senoidal. Puesto que cada raíz de $A(z)$ es función de los parámetros $\beta_1, \beta_2, \dots, \beta_p$, también es posible analizar el comportamiento de $\rho(k)$ de acuerdo a éstos últimos.

De la expresión de la función de densidad espectral para el proceso lineal general dada en (1.4.3), se deduce que dicha función para el proceso AR(p) es

$$f(\omega) = \frac{\sigma_z^2}{2\pi\sigma_x^2} \left| A^{-1}(e^{-i\omega}) \right|^2 = \frac{\sigma_z^2}{2\pi\sigma_x^2} \frac{1}{|A(e^{-i\omega})|^2}, \quad -\pi < \omega < \pi.$$

1.6. Proceso de promedios móviles.

El modelo de promedios móviles de orden q , denotado por $MA(q)$, es un caso particular del modelo (1.4.1), que está dado por

$$X_t = - \sum_{j=1}^q \alpha_j Z_{t-j} + Z_t . \quad (1.6.1)$$

Si se define $M(Z) = 1 - \sum_{j=1}^q \alpha_j Z^j$, una forma alternativa de expresar (1.6.1) será,

$$X_t = [M(B)] Z_t .$$

En este caso la función de transferencia es $M(B)$, y al contener un número finito de términos, permite asegurar que el proceso es estacionario. No obstante si se desea una representación del tipo

$$[M(B)]^{-1} X_t = Z_t ,$$

es necesario pedir, de acuerdo a lo expuesto en la sección 1.5, que $M^{-1}(Z)$ converga en y dentro del círculo unitario, o lo que es equivalente, que los ceros de $M(Z)$ estén fuera del círculo unitario. Si esta última condición se cumple, se dice que el proceso es invertible. En forma análoga a lo hecho en la sección anterior, si se cumple la

condición de invertibilidad en un proceso MA(q), éste se puede expresar como la suma infinita,

$$z_t = [M(B)]^{-1} X_t = \sum_{j=0}^{\infty} \eta_j X_{t-j} .$$

Al pedir la condición de invertibilidad implícitamente se está pidiendo que el peso η_j asignado a cada X_{t-j} , $j = 0, 1, \dots$, sea menor conforme j aumenta. Esto es algo razonable pues se espera que cada observación sea menos importante conforme se está más alejado en el tiempo.

Asimismo, la condición de invertibilidad evita el tener más de un modelo que dé lugar a una misma función de autocovarianza, teniendo así cierto tipo de unicidad.

La función de autocovarianza de un proceso MA(q) (se supone que $E(X_t) = 0$), está dada por

$$\sigma(k) = E(X_t, X_{t-k}) = E \left[\left(- \sum_{j=1}^q \alpha_j z_{t-j} + z_t \right) \left(- \sum_{j=1}^q \alpha_j z_{t-k-j} + z_{t-k} \right) \right],$$

de donde

$$\sigma(k) = \begin{cases} \sigma_z (-\alpha_k + \alpha_1 \alpha_{k+1} + \alpha_2 \alpha_{k+2} + \dots + \alpha_{q-k} \alpha_q), & k = 1, 2, \dots, q \\ 0 & , k > q . \end{cases}$$

Entonces la función de autocorrelación es

$$\rho(k) = \begin{cases} \frac{-\alpha_k + \alpha_1\alpha_{k+1} + \alpha_2\alpha_{k+2} + \dots + \alpha_{q-k}\alpha_q}{1 + \alpha_1^2 + \dots + \alpha_q^2}, & k=1,2,\dots,q \\ 0 & , k > q . \end{cases} \quad (1.6.2)$$

A diferencia de las ecuaciones de Yule-Walker, que son lineales y por tanto fácilmente manipulables, las ecuaciones en (1.6.2) son no lineales.

Es importante notar que siempre la función de autocorrelación se anula para retrasos mayores que el orden del modelo.

Por último, y de acuerdo al resultado para el proceso lineal general, en (1.4.3), la función de densidad espectral de un proceso MA(q) es

$$f(\omega) = \frac{\sigma_z^2}{2\pi\sigma_x^2} \left| M(e^{-i\omega}) \right|^2, \quad -\pi < \omega < \pi.$$

1.7. Proceso autorregresivo y de promedios móviles.

Otro tipo de modelo en el cual se tienen una combinación de los procesos AR y MA es el modelo autorregresivo y de promedios móviles, el cual será denotado por ARMA(p,q) , y cuya expresión es la siguiente

$$X_t - \sum_{j=1}^p \beta_j X_{t-j} = z_t - \sum_{j=1}^q \alpha_j z_{t-j} . \quad (1.7.1.a)$$

En términos del operador de retraso B y de los polinomios A(Z) y M(Z) definidos en las dos secciones anteriores, el modelo (1.7.1.a) también se puede escribir como,

$$[A(B)]X_t = [M(B)]z_t . \quad (1.7.1.b)$$

La función que transforma el ruido blanco en el proceso original, es decir, la función de transferencia, es en este caso $[A(B)]^{-1}M(B)$.

La presencia de ambos tipos de términos, autorregresivos y de medios móviles, en un modelo ARMA , permite que se pueda hablar de las condiciones bajo las cuales el proceso es estacionario y/o invertible. La estacionariedad e invertibilidad en un proceso son dos características independientes, por lo que se pueden aplicar los resultados vistos para procesos AR y MA . Entonces, un proceso ARMA(p,q) modela-

do según (1.7.1) , será estacionario si los ceros de $A(Z)$ están fuera del círculo unitario (ó $A^{-1}(Z)$ converge para todo $Z \in \mathbb{C}$, tal que $\|Z\| \leq 1$) y será invertible si los ceros de $M(Z)$ están fuera del círculo unitario (ó $M^{-1}(Z)$ converge para todo $Z \in \mathbb{C}$, tal que $\|Z\| \leq 1$).

Si en (1.7.1.a) se multiplica por X_{t-k} en ambos lados y se toma esperanza (de nuevo se supone que $E(X_t) = 0$), se obtiene la función de autocovarianzas,

$$\sigma(k) = \sum_{j=1}^p \beta_j (k-j) + E(X_{t-k} Z_t) - \sum_{j=1}^q \alpha_j E(X_{t-k} Z_{t-j}). \quad (1.7.2)$$

Puesto que X_{t-k} es combinación de los choques ocurridos hasta el tiempo $t - k$, entonces $E(X_{t-k} Z_{t-j})$, $j = 0, 1, \dots, q$ se anularán si $k > q$, por lo que en (1.7.2) se tendría

$$\sigma(k) = \sum_{j=1}^p \beta_j \sigma(k-j) , \quad k > q , \quad (1.7.3)$$

o lo que es lo mismo,

$$[A(B)] \sigma(k) = 0 , \quad k > q ,$$

en donde el retraso se aplica a k .

Es inmediato que para la función de autocorrelación se

tiene algo análogo,

$$\rho(k) = \sum_{j=1}^p \beta_j \rho(k-j) \quad \delta \quad [A(B)]\rho(k) = 0, \quad k > q .$$

Las primeras q autocorrelaciones en un proceso ARMA(p, q) dependerán entonces tanto de los términos autorregresivos como de los de medias móviles. No así las autocorrelaciones $\rho(k)$ con $k > q$, que sólo dependen de la parte autorregresiva.

Para resolver la ecuación (1.7.3) se necesita contar con los valores iniciales

$$\sigma(q - p + 1), \sigma(q - p + 2), \dots, \sigma(q - 1), \sigma(q) ,$$

así que cuando $q - p < 0$, toda la función de autocorrelación estará determinada al resolver (1.7.3), y su comportamiento será el de una mezcla de funciones exponenciales y/o senoidales que decaen hacia cero, de acuerdo al carácter real o complejo de las raíces de $A(z)$, como se mencionó en la sección 1.5.

En caso de que $q - p \geq 0$, la conducta de $\rho(k)$ antes señalada se puede asegurar a partir de $k = q - p$.

La función de densidad espectral para el modelo

ARMA(p,q) especificado en (1.7.1) será, de acuerdo a (1.4.3) ,

$$f(\omega) = \frac{\sigma_z^2}{2\pi\sigma_x^2} \left| A^{-1}(e^{-i\omega})M(e^{-i\omega}) \right|^2 =$$

$$= \frac{\sigma_z^2}{2\pi\sigma_x^2} \frac{|M(e^{-i\omega})|^2}{|A(e^{-i\omega})|^2}, \quad -\pi < \omega < \pi .$$

Algo importante desde el punto de vista estadístico y que hace que valga la pena el considerar modelos ARMA es lo siguiente : la experiencia ha sugerido que en varias ocasiones el ajuste de un modelo ARMA(p,q) puede dar tan buenos resultados como el ajuste de un modelo AR(p') en donde $p + q < p'$, así que se podría ajustar un modelo con menos parámetros.

1.8. Procesos no estacionarios.

Consideremos el proceso $\{Z_t: t \in \mathbb{Z}\}$ compuesto de variables aleatorias independientes tales que $E(Z_t) = \mu$ y $V(Z_t) = \sigma^2$. El proceso $\{X_t: t \in \mathbb{Z}\}$ se dice que es una caminata aleatoria si

$$X_t = X_{t-1} + Z_t \quad .$$

Es usual que el proceso inicie en $t = 0$ con un valor de cero, así que $X_1 = Z_1$, y entonces por sustitución sucesiva,

$$X_t = \sum_{i=1}^t Z_i \quad .$$

El proceso $\{X_t: t \in \mathbb{Z}\}$ resulta ser no estacionario, pues

$$E(X_t) = t\mu \quad \text{y} \quad V(X_t) = t\sigma^2.$$

Si se transforma el proceso original a $W_t = X_t - X_{t-1}$, es claro que $W_t = Z_t$, con lo que se obtiene un proceso estacionario.

El procedimiento mostrado en el ejemplo anterior para lograr estacionariedad se puede generalizar. Sea $\{X_t: t \in \mathbb{Z}\}$ un proceso no estacionario y $\{W_t: t \in \mathbb{Z}\}$ otro proceso definido por $W_t = \nabla^d X_t$, en donde $\nabla^d = (1 - B)^d$,

(si $d = 1$, $W_t = X_t - X_{t-1}$). En la práctica se ha visto que esta transformación dá lugar a procesos aproximadamente estacionarios, siempre y cuando no se tengan en la serie de tiempo efectos estacionales, es decir, que la serie no presente conducta similar cada cierto período de tiempo.

El modelo que se propone entonces es el llamado autorregresivo integrado de medias móviles, denotado por ARIMA(p,d,q), y cuya expresión es

$$[A(B)] \nabla^d X_t = [M(B)] z_t ,$$

en donde $A(z)$ y $M(z)$ son los polinomios de orden p y q respectivamente, usados en las secciones anteriores. El análisis de este modelo es análogo al de un proceso ARMA(p,q), sólo que es aplicado a $W_t = \nabla^d X_t$.

Existen otro tipo de transformaciones con las cuales se puede lograr la estacionariedad de un proceso, no obstante el método de diferencias ha resultado ser muy útil, sobre todo cuando la no estacionariedad es en la media y/o la varianza, que es lo más común.

1.9. Procesos estacionales.

Una serie de tiempo es estacional si cada cierto periodo de tiempo la serie repite su comportamiento. Sea s el lapso de tiempo tal que X_t y X_{t+s} son semejantes para toda $t \in \mathbb{Z}$. La transformación que se propone para eliminar el efecto estacional de una serie con índice de estacionalidad igual a s es

$$U_t = [1 - B^s]X_t.$$

Si además $\{U_t: t \in \mathbb{Z}\}$ se considerara no estacionario, se usaría un modelo ARIMA, dando lugar al modelo

$$[A(B)]\nabla^d[1 - B^s]X_t = [M(B)]Z_t.$$

Más información sobre los modelos mencionados en este capítulo, aparte de las citas ya mencionadas, se pueden encontrar en Nelson (1973), Chatfield (1975), Granger y Newbold (1977), Kendall y Stuart (1966).

CAPITULO 2

INFERENCIA

En el capítulo anterior se presentaron algunos procesos estocásticos así como las propiedades más características de cada uno de ellos. Todo lo expuesto fué bajo la suposición de que los parámetros involucrados en la especificación del modelo eran conocidos. En la práctica, el conocimiento que se tiene sobre dichos parámetros no es total, es decir, existe incertidumbre sobre su posible valor. En este capítulo se verá cómo expresar dicha incertidumbre, y de ser posible, reducir ésta si es que se cuenta con información adicional.

2.1. Conceptos generales

Como se mencionó en la introducción del presente trabajo, se adoptará una postura bayesiana para el análisis estadístico de algunos modelos de series de tiempo. Entonces, es conveniente aclarar lo que se entenderá por inferencia desde el punto de vista bayesiano. De acuerdo a Box y Tiao (1973) y Bernardo (1981), la inferencia estadística comprende al conjunto de procedimientos que permiten incorporar, en un problema determinado, la información disponible, con objeto de aumentar el conocimiento del mundo real. Específicamente, supongamos que se está interesado en el valor de un parámetro $\theta \in \Theta$. El conocimiento que se tenga sobre dicha magnitud se debe expre-

sar en términos de una función de densidad de probabilidad* , $p(\theta)$, $\theta \in \Theta$, la cuál es llamada distribución inicial . En ocasiones, la información sobre θ contenida en $p(\theta)$ no se cree suficiente para los propósitos del interesado, así que se plantea un experimento para obtener información adicional. La realización de dicho experimento dá lugar a un conjunto z de observaciones las cuales se supone son valores observados de una variable aleatoria X , cuya densidad de probabilidad, $p(x|\theta)$, $\theta \in \Theta$, es conocida como la función de verosimilitud, y es la función a través de la cual los datos z modifican, como se verá más adelante, el conocimiento inicial que se tiene sobre θ .

Aplicando el teorema de Bayes (Kigman y Taylor (1966), pp. 361-362), es fácil obtener la siguiente relación

$$p(\theta|z) = \frac{p(z|\theta) p(\theta)}{p(z)} , \quad \theta \in \Theta ,$$

donde

$$p(z) = \int p(z|\theta) p(\theta) d\theta , \quad z \in X$$

La densidad $p(\theta|z)$, $\theta \in \Theta$, es llamada distribución final de θ , y es la que expresa el conocimiento sobre el

* Todas las densidades de probabilidad usadas en este trabajo son con respecto a una medida dominante σ -finita: la de Lebesgue para el caso continuo o la de conteo para el caso discreto.

parámetro θ una vez que se ha incorporado la información z proporcionada por el experimento. La función $p(z)$, $z \in X$, recibe el nombre de distribución predictiva, y describe probabilísticamente los resultados que se pueden obtener al observar la variable aleatoria X .

Entonces, y de acuerdo a lo anteriormente expuesto, el problema de inferencia quedará resuelto una vez que la información inicial y/o la información muestral se haya incorporado al análisis por medio de la aplicación del teorema de Bayes.

En la elección de la distribución inicial sobre la magnitud de interés se ha considerado conveniente el poder representar un conocimiento inicial que signifique "ignorancia" o "conocimiento vago", con lo cual se pretende hacer inferencias en las cuales la información obtenida experimentalmente es la que tiene un peso mayor si se compara con la información inicial representada por $p(\theta)$. El uso de este tipo de distribuciones iniciales es justificado por Box y Tiao (1973), quienes argumentan lo siguiente:

1. Una investigación por lo general no se lleva a cabo a menos de que se tenga cierta seguridad de que la información que se obtenga es más precisa que la información ya disponible. Es decir, una investigación no se realiza a menos de que se piense que el llevarla a cabo seguramente aumentará substancialmente nuestro conocimiento.

2. Aún cuando se tenga inicialmente un alto grado de conocimiento acerca del valor de un parámetro, al reportar los resultados de una investigación es sumamente apropiado, y más convincente para otras personas el analizar los datos utilizando una distribución inicial de referencia*, es decir, una distribución que dé más peso a la verosimilitud. De ésta forma, la distribución final obtenida representaría, aproximadamente, el conocimiento de una persona que en un inicio sabía "muy poco" y sólo considero la información contenida en los datos.

Varios procedimientos se han propuesto para determinar distribuciones iniciales de referencia o no informativas. Entre ellos, el método de Jeffreys propone elegir como distribución inicial de referencia para un parámetro $\underline{\theta} \in \mathbb{R}^k$, a una función $p(\underline{\theta})$ que es proporcional a la raíz cuadrada del determinante de la matriz de información de Fisher, es decir,

$$p(\underline{\theta}) \propto |I(\underline{\theta})|^{1/2}, \quad \theta \in \Theta,$$

donde $I(\underline{\theta}) = |m_{ij}|$, $m_{ij} = -E \left\{ \frac{\partial^2 \ln f(x|\underline{\theta})}{\partial \theta_i \partial \theta_j} \right\}$.

La distribución inicial $p(\underline{\theta})$ sugerida por Jeffreys posee ciertas propiedades que la hacen atractiva y una de ellas, muy importante, se presenta a continuación (la demostración puede verse en Zellner (1971)).

* llamada también no informativa.

Supongamos que una persona parametriza su modelo en términos de los elementos de $\underline{\tau}$, donde $\underline{\tau} = h(\underline{\theta})$ y h es una función diferenciable. Si esta persona elige como dist inicial de $\underline{\tau}$

$$p(\underline{\tau}) \propto |I(\underline{\tau})|^{1/2},$$

las distribuciones finales que se obtengan para $\underline{\tau}$ y para $\underline{\theta}$ serán consistentes en el sentido de que dada cualquiera de ellas y la transformación h , es posible obtener la otra, pues,

$$p(\underline{\theta}|z) d\underline{\theta} \propto p(\underline{\tau}|z) d\underline{\tau}.$$

En el desarrollo de este trabajo se hará uso de distribuciones iniciales de referencia obtenidos según la regla de Jeffreys. Otros procedimientos para encontrar este tipo de distribuciones serán brevemente mencionados en el capítulo 3.

Si además de incorporar en el análisis de un problema la evidencia experimental disponible, se desean tomar decisiones relacionadas con el valor parametral $\underline{\xi} \in \Theta$ ó con el valor $z \in X$, es necesario contar con una función de utilidad que cuantifique la consecuencia de haber tomado cada decisión. El criterio bayesiano de decisión (ver referencias dadas al final de la sección 3.5. del capítulo 3), consiste en elegir la decisión que maximice la utilidad esperada.

En este trabajo solamente se tratará el aspecto inferen-

cial, enfocado exclusivamente al análisis estadístico de series de tiempo.

Por último, cabe mencionar que Bernardo (1979) ha planteado el problema de inferencia como un problema de decisión, en donde hay que elegir la distribución que mejor refleje el conocimiento inicial que se tenga.

2.2. PROCESO AUTOREGRESIVO

Supongamos que se cuenta con una serie de observaciones $\underline{x}_n' = (x_1, x_2, \dots, x_n)$, las cuales fueron generadas por el modelo autorregresivo y estacionario de orden p .

$$x_t = \beta_1 x_{t-1} + \dots + \beta_p x_{t-p} + z_t \quad (2.2.1.)$$

Sin pérdida de generalidad se considera que $E(x_t) = 0$, $\forall t \in \mathbb{Z}$. Además se recuerda que $\{z_t : t \in \mathbb{Z}\}$ es un conjunto de variables aleatorias independientes e idénticamente distribuidas con $E(z_t) = 0$ y $V(z_t) = \sigma^2 < \infty$. En este caso los parámetros de interés son $\underline{\beta}' = (\beta_1, \beta_2, \dots, \beta_p)$ y σ .

Suponiendo que cada z_t tiene una distribución normal, lo cual implica que cada x_t también se comporta normalmente, la distribución de \underline{x}_n se puede expresar como

$$p(\underline{x}_n | \underline{\beta}, \sigma) = (2\pi\sigma^2)^{-n/2} |M_n^{(p,0)}|^{-1/2} \exp\left\{-\frac{1}{2\sigma^2} (\underline{x}_n' M_n^{(p,0)} \underline{x}_n)\right\},$$

$$\underline{x}_n \in \mathbb{R}^n \quad (2.2.2.)$$

De esta última expresión, es claro que

$$\sigma^{-2} M_n^{(p,0)} = \left[V(\underline{x}_n)\right]^{-1} = \Gamma_n^{-1},$$

en donde $\Gamma_n = V(X_{-n})$ es la matriz de autocovarianzas de X_{-n} , es decir,

$$\Gamma_n = \{\gamma_{|i-j|}\} = \begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{n-2} \\ \cdots & \cdots & \cdots & \cdots \\ \gamma_{n-1} & \gamma_{n-2} & \cdots & \gamma_0 \end{pmatrix}$$

En términos de la notación usada en el capítulo 1, $\gamma_k = \sigma(x_t, x_s)$, para todo t y s tales que $|t-s| = k$.

La función (2.2.2.), para un x_n dado, es la función de verosimilitud, y lo que enseguida se hará es expresarla en otra forma más conveniente. Para ésto, consideremos la siguiente relación,

$$p(x_n | \underline{\beta}, \sigma) = p(x_{p+1}, x_{p+2}, \dots, x_n | \underline{x}_p, \underline{\beta}, \sigma^2) p(\underline{x}_p | \underline{\beta}, \sigma^2) \quad (2.2.3.)$$

donde $\underline{x}_p = (x_1, x_2, \dots, x_p)$.

Por una parte se sabe que

$$p(\underline{x}_p | \underline{\beta}, \sigma) = (2\pi\sigma^2)^{-p/2} |M_p^{(p,0)}|^{1/2} \exp\left\{ \frac{-1}{2\sigma^2} \underline{x}_p' M_p^{(p,0)} \underline{x}_p \right\}, \quad (2.2.4.)$$

en donde, en forma análoga a (2.2.2.), se tiene que

$$\sigma^{-2} M_p^{(p,0)} = \Gamma_p^{-1}.$$

Ahora se encontrará $p(x_{p+1}, x_{p+2}, \dots, x_n | x_p, \underline{\beta}, \sigma)$.

Para un x_p dado se tiene que

$$z_{p+1} = x_{p+1} - \beta_1 x_p - \beta_2 x_{p-1} - \dots - \beta_p x_1$$

$$z_{p+2} = x_{p+2} - \beta_1 x_{p+1} - \beta_2 x_p - \dots - \beta_p x_2$$

.....

$$z_n = x_n - \beta_1 x_{n-1} - \beta_2 x_{n-2} - \dots - \beta_p x_{n-p}$$

Esta transformación tiene jacobiano igual a 1 pues

$$J = \begin{vmatrix} \frac{\partial z_{p+1}}{\partial x_{p+1}} & \frac{\partial z_{p+1}}{\partial x_{p+2}} & \dots & \frac{\partial z_{p+1}}{\partial x_n} \\ \frac{\partial z_{p+2}}{\partial x_{p+1}} & \frac{\partial z_{p+2}}{\partial x_{p+2}} & \dots & \frac{\partial z_{p+2}}{\partial x_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial z_n}{\partial x_{p+1}} & \frac{\partial z_n}{\partial x_{p+2}} & \dots & \frac{\partial z_n}{\partial x_n} \end{vmatrix} = \begin{vmatrix} 1 & 0 & \dots & 0 \\ -\beta_1 & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & -\beta_1 & 1 \end{vmatrix} = 1.$$

Como además se sabe que

$$P(z_{p+1}, z_{p+2}, \dots, z_n | \sigma) = (2\pi\sigma^2)^{-\frac{(n-p)}{2}} \exp\left\{ \frac{-1}{\sigma^2} \sum_{t=p+1}^n z_t^2 \right\},$$

es posible establecer el siguiente resultado,

$$P(x_{p+1}, x_{p+2}, \dots, x_n | x_p, \underline{\beta}, \sigma) = (2\pi\sigma^2)^{-\frac{(n-p)}{2}} \exp\left\{ \frac{-1}{\sigma^2} \sum_{t=p+1}^n (x_t - \beta_1 x_{t-1} - \dots - \beta_p x_{t-p})^2 \right\}. \quad (2.2.5)$$

Entonces es posible expresar la verosimilitud, de acuerdo a (2.2.3.), (2.2.4.) y (2.2.5.) como

$$p(\underline{x}_n | \underline{\beta}, \sigma) = (2\pi\sigma^2)^{-n/2} |M_p^{(p,0)}|^{1/2} \exp\left\{ \frac{-1}{2\sigma^2} S(\underline{\beta}) \right\} , \quad (2.2.6.)$$

donde

$$S(\underline{\beta}) = \underline{x}'_p M_p^{(p,0)} \underline{x}_p + \sum_{t=p+1}^n (x_t - \beta_1 x_{t-1} - \dots - \beta_p x_{t-p})^2 . \quad (2.2.7.)$$

Comparando (2.2.2.) con (2.2.6.), resulta evidente que $S(\underline{\beta})$ es igual a $\underline{x}'_n M_n^{(p,0)} \underline{x}_n$, y que además en un proceso AR(p), $|M_n^{(p,0)}| = |M_p^{(p,0)}|$.

Si $m_{ij}^{(p)}$ representa el elemento (i,j) de la matriz $M_p^{(p,0)}$, $S(\underline{\beta})$ se puede escribir alternativamente como

$$S(\underline{\beta}) = \sum_{i=1}^p \sum_{j=1}^p m_{ij}^{(p)} x_i x_j + \sum_{t=p+1}^n (x_t - \beta_1 x_{t-1} - \dots - \beta_p x_{t-p})^2 . \quad (2.2.8.)$$

Por un momento supongamos que $n = p+1$, entonces según (2.2.8.),

$$\begin{aligned} \underline{x}'_{p+1} M_{p+1}^{(p,0)} \underline{x}_{p+1} &= \sum_{i=1}^p \sum_{j=1}^p m_{ij}^{(p)} x_i x_j + (x_{p+1} - \beta_1 x_p - \dots - \beta_p x_1)^2 = \\ &= \underline{x}'_p M_p^{(p,0)} \underline{x}_p + (x_{p+1} - \beta_1 x_p - \dots - \beta_p x_1)^2 . \end{aligned}$$

El primer sumando se puede reescribir como

$$\underline{x}'_{p+1} \begin{pmatrix} (p,0) \\ M_p & \underline{0} \\ \underline{0}' & 0 \end{pmatrix} \underline{x}_{p+1}$$

mientras que el segundo sumando se puede expresar como

$$\underline{x}'_{p+1} \begin{pmatrix} -\beta_p \\ -\beta_{p-1} \\ \dots \\ -\beta_1 \\ 1 \end{pmatrix} (-\beta_p -\beta_{p-1} -\dots -\beta_1 \ 1) \underline{x}_{p+1} =$$

$$= \underline{x}'_{p+1} \begin{pmatrix} \beta_p^2 & \beta_p \beta_{p-1} & \dots & -\beta_p \\ \beta_p \beta_{p-1} & \beta_{p-1}^2 & \dots & -\beta_{p-1} \\ \dots & \dots & \dots & \dots \\ -\beta_p & -\beta_{p-1} & \dots & 1 \end{pmatrix} \underline{x}_{p+1} ,$$

por lo que

$$\underline{x}'_{p+1} M_{p+1} \begin{pmatrix} (p,0) \\ \underline{0}' \end{pmatrix} \underline{x}_{p+1} = \underline{x}'_{p+1} \left\{ \begin{pmatrix} (p,0) \\ M_p & \underline{0} \\ \underline{0}' & 0 \end{pmatrix} + \begin{pmatrix} \beta_p^2 & \beta_p \beta_{p-1} & \dots & -\beta_p \\ \beta_p \beta_{p-1} & \beta_{p-1}^2 & \dots & -\beta_{p-1} \\ \dots & \dots & \dots & \dots \\ -\beta_p & -\beta_{p-1} & \dots & 1 \end{pmatrix} \right\} \underline{x}_{p+1} ,$$

de donde

$$M_{p+1}^{(p,0)} = \begin{pmatrix} M_p^{(p,0)} & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} \beta_p^2 & \beta_p \beta_{p-1} & \dots & -\beta_p \\ \beta_p \beta_{p-1} & \beta_{p-1}^2 & \dots & -\beta_{p-1} \\ \dots & \dots & \dots & \dots \\ -\beta_p & -\beta_{p-1} & \dots & 1 \end{pmatrix}$$

Ejemplifiquemos este último resultado para los casos AR(1) y AR(2). Si $p = 1$,

$$M_2^{(1,0)} = \begin{pmatrix} M^{(1,0)} & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} \beta_1^2 & -\beta_1 \\ -\beta_1 & 1 \end{pmatrix} = \begin{pmatrix} m_{11}^{(1)} + \beta_1^2 & \\ -\beta_1 & 1 \end{pmatrix},$$

pero $M_n^{(p,0)}$ es simétrica respecto a sus dos diagonales principales así que

$$1 = m_{11}^{(1)} + \beta_1^2 \implies M_1^{(1,0)} = \{m_{11}^{(1)}\} = 1 - \beta_1^2,$$

y además

$$M_{(2)}^{(1,0)} = \begin{pmatrix} 1 & -\beta_1 \\ -\beta_1 & 1 \end{pmatrix}.$$

Como se dijo antes, es claro que $|M_1^{(1,0)}| = |M_2^{(1,0)}|$.

Si ahora $p = 2$,

$$M_3^{(2,0)} = \begin{pmatrix} M_2^{(2,0)} & 0 \\ 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} \beta_2^2 & \beta_2 \beta_1 & -\beta_2 \\ \beta_2 \beta_1 & \beta_1^2 & -\beta_1 \\ \beta_2 & \beta_1 & 1 \end{pmatrix} = \begin{pmatrix} m_{11}^{(2)} + \beta_2^2 & m_{12}^{(2)} + \beta_2 \beta_1 & -\beta_2 \\ m_{21}^{(2)} + \beta_2 \beta_1 & m_{22}^{(2)} + \beta_1^2 & -\beta_1 \\ -\beta_2 & -\beta_1 & 1 \end{pmatrix}$$

y por la doble simetría de $M_n^{(p,0)}$,

$$m_{11}^{(2)} + \beta_2^2 = 1, \quad m_{12}^{(2)} + \beta_2 \beta_1 = -\beta_1, \quad m_{21}^{(2)} + \beta_2 \beta_1 = -\beta_1,$$

de donde

$$m_{11}^{(2)} = 1 - \beta_2^2, \quad m_{12}^{(2)} = -\beta_1(1 + \beta_2), \quad m_{21}^{(2)} = -\beta_1(1 + \beta_2),$$

y además $m_{22}^{(2)} = 1 - \beta_2^2$, ya que $m_{11}^{(2)} = m_{22}^{(2)}$.

Entonces,

$$M_2^{(2,0)} = \begin{pmatrix} 1 - \beta_2^2 & -\beta_1(1 + \beta_2) \\ \beta_1(1 + \beta_2) & 1 - \beta_2^2 \end{pmatrix} \text{ y } M_3^{(2,0)} = \begin{pmatrix} 1 & -\beta_1 & -\beta_2 \\ -\beta_1 & 1 + \beta_1^2 - \beta_2^2 & -\beta_1 \\ -\beta_2 & \beta_1 & 1 \end{pmatrix}.$$

También los determinantes son iguales:

$$\begin{aligned} |M_2^{(2,0)}| &= (1 - \beta_2^2)^2 - \beta_1^2(1 + \beta_2)^2 = (1 - \beta_2^2)^2(1 + \beta_2)^2 - \beta_1^2(1 + \beta_2)^2 = \\ &= (1 + \beta_2)^2 |(1 - \beta_2^2)^2 - \beta_1^2|, \end{aligned}$$

$$\begin{aligned} |M_3^{(2,0)}| &= 1 + \beta_1^2 - \beta_2^2 - \beta_1^2 + \beta_1(-\beta_1 - \beta_1\beta_2) - \beta_2(\beta_1^2 + \beta_2 + \beta_1^2\beta_2 - \beta_2^3) = \\ &= 1 - \beta_2^2 - \beta_1^2 - \beta_1^2\beta_2 - \beta_2\beta_1^2 - \beta_2^2 - \beta_1^2\beta_2^2 + \beta_2^4 = \\ &= (1 - 2\beta_2^2 + \beta_2^4) - \beta_1^2(1 + 2\beta_2 + \beta_2^2) = (1 - \beta_2^2)^2 - \beta_1^2(1 + \beta_2)^2 = |M_2^{(2,0)}|. \end{aligned}$$

Finalmente, y en vista de lo anterior la verosimilitud sería, de acuerdo a (2.2.6.) y (2.2.7.), cuando $p = 1$,

$$p(\underline{x}_n | \beta_1, \sigma^2) = (2\pi\sigma^2)^{-n/2} (1-\beta_1^2)^{1/2} \exp\left\{ \frac{-1}{2\sigma^2} \left[x_1^2(1-\beta_1^2) + \sum_{t=2}^n (x_t - \beta_1 x_{t-1})^2 \right] \right\}$$

Si $p = 2$,

$$p(\underline{x}_n | \beta_1, \beta_2, \sigma^2) = (2\pi\sigma^2)^{-n/2} (1+\beta_2^2) \left[(1-\beta_2)^2 - \beta_1^2 \right]^{1/2} \exp\left\{ \frac{-1}{2\sigma^2} S(\beta_1, \beta_2) \right\},$$

donde

$$S(\beta_1, \beta_2) = (\underline{x}_1, \underline{x}_2) \begin{pmatrix} 1-\beta_2^2 & -\beta_1(1+\beta_2) \\ -\beta_1(1+\beta_2) & 1-\beta_2^2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \sum_{t=3}^n (x_t - \beta_1 x_{t-1} - \beta_2 x_{t-2})^2 =$$

$$= x_1^2 + x_2^2 - x_1^2 \beta_2^2 - x_2^2 \beta_2^2 - 2x_1 x_2 \beta_1 - 2x_1 x_2 \beta_1 \beta_2 + \sum_{t=3}^n (x_t - \beta_1 x_{t-1} - \beta_2 x_{t-2})^2.$$

De la expresión (2.2.8.) es claro que $S(\underline{\beta})$ es una forma cuadrática en \underline{x}_n , sin embargo también lo es en $\underline{\beta}$, lo cual es posible deducirlo por la forma en que se generan las $M_p^{(p,0)}$. De toda suerte que si se define $\underline{\beta}'_* = (1 \ \beta_1 \ \beta_2 \ \dots \ \beta_p)$, existe una matriz D de dimensiones $(p+1) \times (p+1)$, cuyos elementos son funciones de las x_i 's, y tal que

$$S(\underline{\beta}) = \underline{x}'_n M_n^{(p,0)} \underline{x}_n = \underline{\beta}'_* D \underline{\beta}_* \quad (2.2.9.)$$

La matriz D resulta ser, después de realizar las operaciones necesarias, de la siguiente forma,

$$D = \begin{pmatrix} d_{11} & -d_{12} & \dots & -d_{1,p+1} \\ -d_{21} & d_{22} & \dots & d_{2,p+1} \\ \dots & \dots & \dots & \dots \\ -d_{p+1,1} & d_{p+1,2} & \dots & d_{p+1,p+1} \end{pmatrix}, \quad (2.2.10.)$$

en donde

$$d_{ij} = d_{ji} = x_i x_j + x_{i+1} x_{j+1} + \dots + x_{n+1-j} x_{n+1-i}, \quad (2.2.11.)$$

y cada d_{ij} contiene $n-(i-1)-(j-1)$ sumandos.

Consideremos el caso $p = 2$. Como se vió anteriormente

$$S(\beta_1, \beta_2) = \underline{x}_2' M_2^{(2,0)} \underline{x}_2 + \sum_{t=3}^n (x_t - \beta_1 x_{t-1} - \beta_2 x_{t-2})^2,$$

donde

$$M_2^{(2,0)} = \begin{pmatrix} 1 - \beta_2^2 & -\beta_1(1 + \beta_2) \\ -\beta_1(1 + \beta_2) & 1 - \beta_2^2 \end{pmatrix}.$$

Al desarrollar el primer sumando en $S(\beta_1, \beta_2)$ y agrupar términos en β_1 , β_2^2 y $\beta_1\beta_2$ se tiene que

$$\underline{x}_2' M_2^{(2,0)} \underline{x}_2 = (1 \ \beta_1 \ \beta_2) \begin{pmatrix} x_1^2 + x_2^2 & -x_1 x_2 & 0 \\ -x_1 x_2 & 0 & -x_1 x_2 \\ 0 & -x_1 x_2 & -(x_1^2 + x_2^2) \end{pmatrix} \begin{pmatrix} 1 \\ \beta_1 \\ \beta_2 \end{pmatrix}$$

Por otra parte, es claro que

$$(x_t - \beta_1 x_{t-1} - \beta_2 x_{t-2})^2 = (1 \ \beta_1 \ \beta_2) \begin{pmatrix} x_t \\ x_{t-1} \\ x_{t-2} \end{pmatrix} (x_t \ x_{t-1} \ x_{t-2}) \begin{pmatrix} 1 \\ \beta_1 \\ \beta_2 \end{pmatrix},$$

lo cual permite expresar la suma de cuadrados en $S(\beta_1, \beta_2)$ como

$$\sum_{t=3}^n (x_t - \beta_1 x_{t-1} - \beta_2 x_{t-2})^2 = (1 \ \beta_1 \ \beta_2) \sum_{t=3}^n \begin{pmatrix} x_t^2 & -x_t x_{t-1} & -x_t x_{t-2} \\ -x_t x_{t-1} & x_{t-1}^2 & x_{t-1} x_{t-2} \\ -x_{t-2} x_t & x_{t-1} x_{t-2} & x_{t-2}^2 \end{pmatrix} \begin{pmatrix} 1 \\ \beta_1 \\ \beta_2 \end{pmatrix}$$

Siendo $\underline{\beta}_*^1 = (1 \ \beta_1 \ \beta_2)$, la sustitución de cada sumando en $S(\beta_1, \beta_2)$ por su expresión alternativa permite asegurar que $S(\beta_1, \beta_2) = \underline{\beta}_*^1 D \beta_*$, con D la matriz definida en (2.2.10.) y $p = 2$.

El hecho de haber podido expresar $S(\underline{\beta})$ como una forma cuadrática en las β_i 's, facilitará el trabajo de encontrar la matriz de información de Fisher, necesaria para obtener la distribución inicial de referencia de (β, σ) .

De acuerdo a (2.2.6.) y (2.2.9.) la verosimilitud es entonces,

$$p(x_n | \underline{\beta}, \sigma) = (2\pi\sigma^2)^{-n/2} |M_p^{(p,0)}|^{1/2} \exp \left\{ \frac{-1}{2\sigma^2} \underline{\beta}_*^1 D \beta_* \right\}, \quad (2.2.12.)$$

así que

$$\ell(\underline{\beta}, \sigma | \underline{x}_n) = \ln p(\underline{x}_n | \underline{\beta}, \sigma) = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 + \frac{1}{2} \ln |M_n^{(p,0)}| - \frac{1}{2\sigma^2} \underline{\beta}'_* D \underline{\beta}_* .$$

(2.2.13.)

De esta forma, las derivadas parciales de ℓ respecto a $\underline{\beta}$ y σ están dadas por

$$\frac{\partial \ell}{\partial \beta_j} = \frac{\partial}{\partial \beta_j} \left[\frac{1}{2} \ln |M_n^{(p,0)}| - \sigma^{-2} (-d_{1,j+1} + \beta_1 d_{2,j+1} + \dots + \beta_p d_{p+1,j+1}) \right] ,$$

$j = 1, 2, \dots, p$

(2.2.14.)

$$\frac{\partial \ell}{\partial \sigma} = -\frac{n}{\sigma} + \frac{S(\underline{\beta})}{\sigma^3} .$$

(2.2.15.)

Las derivadas en (2.2.14.) se pueden deducir del hecho de que la derivada de $\underline{\beta}'_* D \underline{\beta}_*$ con respecto a $\underline{\beta}_*$ es igual a $2D\underline{\beta}_*$, así que la suma entre paréntesis en (2.2.14.) es el elemento $(j+1)$ -ésimo del producto $D\underline{\beta}_*$, además se recuerda que $d_{ij} = d_{ji}$.

Las derivadas parciales de segundo orden serán, entonces,

$$\frac{\partial^2 \ell}{\partial \beta_j^2} = \frac{\partial^2}{\partial \beta_j^2} \left[\frac{1}{2} \ln |M_n^{(p,0)}| - \sigma^{-2} d_{j+1,j+1} \right] ,$$

(2.2.16.)

$$\frac{\partial^2 \ell}{\partial \beta_i \partial \beta_j} = \frac{\partial^2}{\partial \beta_i \partial \beta_j} \left[\frac{1}{2} \ln |M_n^{(p,0)}| - \sigma^{-2} d_{i+1,j+1} \right] ,$$

(2.2.17.)

$$\frac{\partial^2 \ell}{\partial \sigma \partial \beta_j} = \frac{\partial^2}{\partial \sigma \partial \beta_j} \frac{1}{2} \ln |M_n^{(p,0)}| + 2 \sigma^{-3} (-d_{1,j+1} + \beta_1 d_{2,j+1} + \dots + \beta_p d_{p+1,j+1}) , \quad (2.2.18.)$$

$$\frac{\partial^2 \ell}{\partial \sigma^2} = \frac{n}{\sigma^2} - 3 \frac{S(\beta)}{\sigma^4} . \quad (2.2.19.)$$

Antes de calcular el valor esperado de cada una de las cuatro derivadas anteriores, es conveniente expresarlas de otra forma. Con este fin recordemos que, bajo ciertas condiciones de regularidad, la esperanza del logaritmo natural de la función de verosimilitud es cero, lo cual quiere decir que,

$$E \left\{ \frac{\partial \ell}{\partial \beta_j} \right\} = 0 , \quad j = 1, 2, \dots, p , \quad (2.2.20.)$$

$$E \left\{ \frac{\partial \ell}{\partial \sigma} \right\} = 0$$

Por (2.2.14.), lo anterior implica que,

$$0 = E \left\{ \frac{\partial \ell}{\partial \beta_j} \right\} = \frac{\partial}{\partial \beta_j} \frac{1}{2} \ln |M_n^{(p,0)}| - \sigma^{-2} |E(-d_{1,j+1}) + \beta_1 E(d_{2,j+1}) + \dots + \beta_p E(d_{p+1,j+1})|$$

$$j = 1, 2, \dots, p .$$

La esperanza de cada d_{ij} , haciendo referencia a (2.2.11.) es igual a

$$[n - (i-1) - (j-1)] \gamma_{j-i} ,$$

por lo que

$$0 = \frac{\partial}{\partial \beta_j} \frac{1}{2} \ln |M_n^{(p,0)}| + \sigma^{-2} [(n-j)\gamma_j - (n-j-1)\beta_1\gamma_{j-1} - \dots - (n-p-j)\beta_p\gamma_{j-p}] \quad (2.2.21.)$$

En la sección 5 del capítulo 1 se dijo que la función de autocovarianza (o de autocorrelación) de todo proceso AR(p), satisface la ecuación en diferencias

$$\gamma_k - \beta_1\gamma_{k-1} - \beta_2\gamma_{k-2} - \dots - \beta_p\gamma_{k-p} = 0 \quad \forall k$$

Si en (2.2.21.) aplicamos este último resultado junto con el hecho de que

$$\begin{aligned} (n-j)\gamma_j - (n-j-1)\beta_1\gamma_{j-1} - \dots - (n-p-j)\beta_p\gamma_{j-p} &= \\ = n(\gamma_j - \beta_1\gamma_{j-1} - \dots - \beta_p\gamma_{j-p}) - (j\gamma_j - (j+1)\beta_1\gamma_{j-1} - \dots - (p+j)\beta_p\gamma_{j-p}) \end{aligned}$$

Se tiene que

$$\frac{\partial}{\partial \beta_j} \frac{1}{2} \ln |M_n^{(p,0)}| = \sigma^{-2} [j\gamma_j - (j+1)\beta_1\gamma_{j-1} - \dots - (j+p)\beta_p\gamma_{j-p}] \quad (2.2.22.)$$

La derivada (2.2.17.) se puede ahora reexpresar como

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \beta_i \partial \beta_j} &= \frac{\partial}{\partial \beta_i} \left(\frac{\partial}{\partial \beta_j} \frac{1}{2} \ln |M_n^{(p,0)}| \right) - \sigma^{-2} d_{i+1,j+1} = \\ &= -\sigma^{-2} (j+i) \gamma_{j-i} - \sigma^{-2} d_{i+1,j+1} \quad , \quad (2.2.23.) \end{aligned}$$

y la derivada (2.2.18.) como

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \sigma \partial \beta_j} &= \frac{\partial}{\partial \sigma} \left(\frac{\partial}{\partial \beta_j} \frac{1}{2} \ln |M_n^{(p,0)}| \right) + \sigma^{-3} (-d_{1,j+1} + \beta_1 d_{2,j+1} + \dots + \beta_p d_{p+1,j+1}) = \\ &= -2\sigma^{-3} [j\gamma_j - (j+1)\beta_1 \gamma_{j-1} - \dots - (j+p)\beta_p \gamma_{j-p}] - 2\sigma^{-3} (d_{1,j+1} - \beta_1 d_{2,j+1} - \dots - \beta_p d_{p+1,j+1}) = \\ &= -2\sigma^{-1} \{ \sigma^{-2} [j\gamma_j - (j+1)\beta_1 \gamma_{j-1} - \dots - (j+p)\beta_p \gamma_{j-p}] + \sigma^{-2} (d_{1,j+1} - \beta_1 d_{2,j+1} - \dots - \beta_p d_{p+1,j+1}) \} , \end{aligned}$$

lo cual, de acuerdo a (2.2.14.) y (2.2.22.) es igual a

$$\frac{\partial^2 \ell}{\partial \sigma \partial \beta_j} = -2\sigma^{-1} \frac{\partial \ell}{\partial \beta_j} \quad . \quad (2.2.24.)$$

Ahora estamos en posibilidad de calcular los valores esperados de las derivadas en (2.2.16.), (2.2.17.) y (2.2.18.) .

Por (2.2.11.) y (2.2.23.) ,

$$\begin{aligned} E \left\{ \frac{\partial^2 \ell}{\partial \beta_i \partial \beta_j} \right\} &= -\sigma^{-2} (j+i) \gamma_{j-i} - \sigma^{-2} E(d_{i+1,j+1}) = \\ &= -\sigma^{-2} (j+1) \gamma_{j-1} - \sigma^{-2} (n-i-j) \gamma_{j-i} = -n \sigma^{-2} \gamma_{j-i} \quad . \end{aligned}$$

También, por (2.2.20.) y (2.2.24.),

$$E \left\{ \frac{\partial^2 \ell}{\partial \sigma \partial \beta_j} \right\} = -2\sigma^{-1} E \left\{ \frac{\partial \ell}{\partial \beta_j} \right\} = 0 .$$

Por último, la suposición de normalidad implica

$$\frac{S(\underline{\beta})}{\sigma^2} \sim \chi^2(n) , \quad (2.2.25.)$$

por lo que, de acuerdo a (2.2.19.),

$$E \left\{ \frac{\partial^2 \ell}{\partial \sigma^2} \right\} = E \left\{ \frac{n}{\sigma^2} - 3 \frac{S(\underline{\beta})}{\sigma^4} \right\} = \frac{n}{\sigma^2} - \frac{3}{\sigma^2} n = \frac{-2n}{\sigma^2} .$$

De esta forma, la matriz de información de Fisher es

$$I(\underline{\beta}, \sigma) = \begin{pmatrix} \{ n\sigma^{-2} \gamma_{j-i} \} & \underline{0} \\ \underline{0}' & 2n/\sigma^2 \end{pmatrix} = \begin{pmatrix} n \left(M_p^{(p,0)} \right)^{-1} & \underline{0} \\ \underline{0}' & 2n/\sigma^2 \end{pmatrix} \quad (2.2.26.)$$

con determinante igual a

$$|I(\underline{\beta}, \sigma)| = \frac{2n}{\sigma^2} \left| n \left(M_p^{(p,0)} \right)^{-1} \right| = \frac{2n^{p+1}}{\sigma^2} \left| M_p^{(p,0)} \right|^{-1} .$$

El criterio de Jeffreys sugiere elegir la distribución inicial de referencia para $(\underline{\beta}, \sigma)$ como la distribución $p(\underline{\beta}, \sigma)$ tal que

$$p(\underline{\beta}, \sigma) \propto |I(\underline{\beta}, \sigma)|^{1/2},$$

es decir,

$$p(\underline{\beta}, \sigma) \propto \sigma^{-1} |M_p^{(p,0)}|^{-1/2}. \quad (2.2.27.)$$

De esta forma, la distribución final de referencia para $(\underline{\beta}, \sigma)$ está dada por

$$p(\underline{\beta}, \sigma | \underline{x}_n) \propto p(\underline{x}_n | \underline{\beta}, \sigma) p(\underline{\beta}, \sigma),$$

lo cual se puede expresar de acuerdo a (2.2.6.) y (2.2.27.)

como,

$$p(\underline{\beta}, \sigma | \underline{x}_n) \propto \sigma^{-n} |M_p^{(p,0)}|^{1/2} \exp\left\{\frac{-1}{2\sigma^2} S(\underline{\beta})\right\} \sigma^{-1} |M_p^{(p,0)}|^{-1/2},$$

es decir,

$$p(\underline{\beta}, \sigma | \underline{x}_n) \propto \sigma^{-(n+1)} \exp\left\{\frac{-1}{2\sigma^2} S(\underline{\beta})\right\}, \quad (2.2.28.)$$

Si uno desea hacer inferencias conjuntas de $\underline{\beta}$ y σ es ésta última distribución la que debe usarse. Sin embargo, es más usual desear hacer inferencias sobre el vector de parámetros $\underline{\beta}$, para lo cual es necesario obtener la distribución marginal $p(\underline{\beta} | \underline{x}_n)$. Es claro que dicha distribución se puede obtener de la distribución conjunta $p(\underline{\beta}, \sigma | \underline{x}_n)$, ya que

$$p(\underline{\beta} | \underline{x}_n) = \int_0^{\infty} p(\underline{\beta}, \sigma | \underline{x}_n) d\sigma.$$

Haciendo uso de la integral de una densidad gamma invertida (ver apéndice A), es inmediato que

$$\int_0^{\infty} \sigma^{-(n+1)} \exp\left\{\frac{-1}{2\sigma^2} S(\underline{\beta})\right\} d\sigma = \frac{\Gamma(n/2) [2 S^{-1}(\underline{\beta})]^{n/2}}{2},$$

de donde,

$$p(\underline{\beta} | \underline{x}_n) \propto |S(\underline{\beta})|^{-n/2}. \quad (2.2.29.)$$

Ahora bien, el valor esperado de $S(\underline{\beta}) = \underline{\beta}_* D \underline{\beta}_*$ es, por (2.2.25.), proporcional a n , mientras que $|M_p^{(p,0)}|$ es independiente de n . Generalmente se ha observado que con valores de p menores o iguales a tres se han obtenido modelos bastante aceptables, así que para tamaños de muestra no tan grandes, el término $S(\underline{\beta})$ será apreciablemente mayor que $|M_p^{(p,0)}|$. A partir de lo anterior la verosimilitud en (2.2.12.) será aproximadamente igual a

$$p(\underline{x}_n | \underline{\beta}, \sigma) = (2\pi\sigma^2)^{-n/2} \exp\left\{\frac{-1}{2\sigma^2} S(\underline{\beta})\right\},$$

de donde

$$\ell(\underline{\beta}, \sigma | \underline{x}_n) = \ln p(\underline{x}_n | \underline{\beta}, \sigma) = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} S(\underline{\beta}).$$

Si $\hat{\underline{\beta}}$ es el valor de $\underline{\beta}$ que maximiza $\ell(\underline{\beta}, \sigma | \underline{x}_n)$, (por lo que $\hat{\underline{\beta}}$ es aproximadamente el estimador máximo verosímil de $\underline{\beta}$), también $\hat{\underline{\beta}}$ es el valor de $\underline{\beta}$ que hace mínimo $S(\underline{\beta})$, entonces $\hat{\underline{\beta}}$ coincide con el estimador de $\underline{\beta}$ de mínimos cuadra

dos. Como $S(\underline{\beta}) = \underline{\beta}' D \underline{\beta}$, con D la matriz definida en (2.2.10), las derivadas con respecto a cada β_i , $i = 1, 2, \dots, p$, de $S(\underline{\beta})$, dan lugar a las siguientes ecuaciones (llamadas ecuaciones normales),

$$\begin{aligned} d_{1,2} &= \hat{\beta}_1 d_{2,2} + \hat{\beta}_2 d_{2,3} + \dots + \hat{\beta}_p d_{2,p+1} \\ d_{1,3} &= \hat{\beta}_1 d_{2,3} + \hat{\beta}_2 d_{3,3} + \dots + \hat{\beta}_p d_{3,p+1} \\ &\dots \dots \dots \\ d_{1,p+1} &= \hat{\beta}_1 d_{2,p+1} + \hat{\beta}_2 d_{3,p+1} + \dots + \hat{\beta}_p d_{p+1,p+1} . \end{aligned} \quad (2.2.31)$$

Sea ahora

$$\underline{d} = \begin{pmatrix} d_{1,2} \\ d_{1,3} \\ \dots \\ d_{1,p+1} \end{pmatrix} \quad D_p = \begin{pmatrix} d_{2,2} & d_{2,3} & \dots & d_{2,p+1} \\ d_{2,3} & d_{3,3} & \dots & d_{3,p+1} \\ \dots & \dots & \dots & \dots \\ d_{2,p+1} & d_{3,p+1} & \dots & d_{p+1,p+1} \end{pmatrix}$$

las ecuaciones (2.2.31) se pueden expresar entonces como

$$\underline{d} = D_p \underline{\beta} ,$$

y la solución está dada por

$$\hat{\underline{\beta}} = D_p^{-1} \underline{d} \quad (2.2.32)$$

Siendo $\hat{\underline{\beta}}$ el valor de $\underline{\beta}$ que minimiza $S(\underline{\beta})$, es posible demostrar que

$$S(\underline{\beta}) = vS^2 + (\underline{\beta} - \hat{\underline{\beta}})' D_p (\underline{\beta} - \hat{\underline{\beta}}), \quad (2.2.33)$$

donde

$$\begin{aligned} S^2 &= \frac{S(\hat{\underline{\beta}})}{v}, \quad S(\hat{\underline{\beta}}) = \hat{\underline{\beta}}' D_p \hat{\underline{\beta}} = d_{11} - \hat{\underline{\beta}}' D_p \hat{\underline{\beta}} = \\ &= d_{11} - \underline{d}' D_p^{-1} \underline{d}, \quad v = n - p. \end{aligned} \quad (2.2.34)$$

Lo hecho anteriormente permite expresar la distribución final de $\underline{\beta}$ dada en (2.2.29) como

$$p(\underline{\beta} | \underline{x}_n) \propto \left\{ 1 + \frac{(\underline{\beta} - \hat{\underline{\beta}})' D_p (\underline{\beta} - \hat{\underline{\beta}})}{vS^2} \right\} \quad (2.2.35)$$

Esta distribución de probabilidad resulta ser una distribución t de Student p-variada con v grados de libertad, $\hat{\underline{\beta}}$ el vector de localización y $S^{-2} D_p$ matriz de precisión ó $\frac{v}{v-2} S^2 D_p^{-1}$ matriz de covarianzas. (Ver anexo A).

Como caso particular, si $p = 1$, entonces $\underline{d} = (d_{12})$ y $D_p = (d_{22})$, así que

$$\hat{\beta}_1 = D_p^{-1} \underline{d} = \frac{d_{12}}{d_{22}}.$$

Además, utilizando (2.2.34),

$$S_{\hat{\beta}_1}^2 = \hat{S}^2 D_p^{-1} = \frac{1}{n-1} \left(d_{11} - \frac{d_{12}^2}{d_{22}} \right) d_{22}^{-1} = \frac{1}{n-1} \left(\frac{d_{11}}{d_{22}} - \hat{\beta}_1^2 \right).$$

Entonces es claro que

$$\frac{\beta_1 - \hat{\beta}_1}{S_{\hat{\beta}_1}}$$

tiene una distribución $t(0, 1, n-1)$.

Conviene hacer notar que para n grande $S_{\hat{\beta}_1}^2$ se aproxima

a

$$\frac{1}{n} (1 - \beta_1^2),$$

pues $\frac{d_{11}}{d_{22}} \rightarrow 1$ y $\hat{\beta}_1$ es consistente para β_1 .

Utilizando (2.2.26) y recordando que $M_j^{(1,0)} = 1 - \beta_1^2$, se tiene que

$$I(\beta_1) = n (1 - \beta_1^2)^{-1},$$

lo que concuerda con el resultado general de que el inverso de la matriz de información de $\underline{\beta}$ aproxima la varianza de $\hat{\underline{\beta}}$, el estimador máximo verosímil.

Si uno desea realizar inferencias sobre un subconjunto de las β_i 's, se tiene que obtener la distribución marginal correspondiente, que resulta ser también una distribución t de Student. (Ver Zellner (1971), apéndice B).

Es usual que en series de tiempo se cuente con tamaños de muestra grandes, por ejemplo $n > 50$, por lo que se podría recurrir a la aproximación normal para la distribución t , (Ver Zellner (1971), apéndice B), obteniéndose así que aproximadamente

$$\underline{\beta} \sim N(\hat{\underline{\beta}}, s^2 D_p^{-1}) .$$

Como parte de la información final que se tiene sobre $\underline{\beta}$, están las llamadas regiones de mayor densidad. Una γ -región de mayor densidad para $\underline{\beta}$ (en este caso contenida en \mathbb{R}^p), no es otra cosa que una región cuyo volumen es el más pequeño posible dentro del conjunto de regiones que con probabilidad γ contienen a $\underline{\beta}$.

Para construir estas regiones se utilizarán los siguientes resultados que dan lugar a dos distribuciones bien conocidas. Sea $\underline{\omega}' = (\omega_1, \omega_2, \dots, \omega_p)$, $\omega_i \in (-\infty, \infty)$, $i = 1, 2, \dots, p$. Consideremos a, b y m números reales positivos y $\underline{\omega}' Q \underline{\omega}$ una forma cuadrática definida positiva. Se puede demostrar, (Box y Jenkins (1970), que

$$\int_{\mathbb{R}^p} \left[a + \frac{\omega' Q \omega}{b} \right]^{-(m+p)/2} d\omega = \frac{(b\pi)^{p/2} \Gamma(m/2)}{a^{m/2} |Q|^{1/2} \Gamma\left(\frac{m+p}{2}\right)}$$

y además

$$\frac{\int_{\mathbb{R}^p} \left[1 + \frac{\omega' Q \omega}{m} \right]^{-(m+p)/2} d\omega}{\int_{\mathbb{R}^p} \left[1 + \frac{\omega' Q \omega}{m} \right]^{-(m+p)/2} d\omega} = \int_0^{F_0} p(r|p,m) dr,$$

donde $c_1 = \{\omega \in \mathbb{R}^p | \omega' Q \omega < pF_0\}$ y $p(r|p,m)$ es la distribución F con p y m grados de libertad (Ver apéndice A).

Si m tiende a infinito, entonces

$$\left[1 + \frac{\omega' Q \omega}{m} \right]^{-(m+p)/2} \longrightarrow e^{-\omega' Q \omega / 2}$$

Definiendo $s = pr$, también se tiene que

$$\frac{\int_{\mathbb{R}^p} e^{-\omega' Q \omega / 2} d\omega}{\int_{\mathbb{R}^p} e^{-\omega' Q \omega / 2} d\omega} = \int_0^{X_0^2} p(s|p) ds,$$

donde $c_2 = \{\omega \in \mathbb{R}^p | \omega' Q \omega < X_0^2\}$ y $p(s|p)$ es la distribución

χ^2 con p grados de libertad. (Ver apéndice A).

Al tener $\underline{\beta}$ una distribución final dada por (2.2.35), es decir una distribución t p -variada, la γ -región de mayor densidad para $\underline{\beta}$ estará dada por

$$R_\gamma = \{ \underline{\beta} \in \mathbb{R}^p \mid (\underline{\beta} - \hat{\underline{\beta}})' D_p (\underline{\beta} - \hat{\underline{\beta}}) < p S^2 F_{(p, \nu)}^{1-\alpha} \} , \quad (2.2.36)$$

donde $F_{(p, \nu)}^{1-\alpha}$ es el cuantil $1 - \alpha = \gamma$ de la distribución $F(p, \nu)$.

Si n es grande, $pF_{(p, \nu)} \simeq \chi^2_{(p)}$, y entonces la γ -región de mayor densidad para $\underline{\beta}$ sería

$$R_\gamma = \{ \underline{\beta} \in \mathbb{R}^p \mid (\underline{\beta} - \hat{\underline{\beta}})' D_p (\underline{\beta} - \hat{\underline{\beta}}) < S^2 \chi^2_{(p), 1-\alpha} \} , \quad (2.2.37)$$

donde $\chi^2_{(p), 1-\alpha}$ es el cuantil $1 - \alpha = \gamma$ de la distribución $\chi^2_{(p)}$. Esto último se puede deducir más directamente si recordamos que, aproximadamente

$$\underline{\beta} \sim N(\hat{\underline{\beta}}, S^2 D_p^{-1})$$

para n grande, de donde

$$(\underline{\beta} - \hat{\underline{\beta}})' D_p (\underline{\beta} - \hat{\underline{\beta}}) / S^2 \sim \chi^2_{(p)} .$$

En cuanto a la distribución final de σ , ésta se encuentra fácilmente integrando la distribución conjunta de $\underline{\beta}$ y σ con respecto a $\underline{\beta}$.

Como se recordará

$$p(\underline{\beta}, \sigma | \underline{x}_n) \propto \sigma^{-(n+1)} \exp\left\{\frac{-1}{2\sigma^2} S(\underline{\beta})\right\}, \quad \underline{\beta} \in \mathbb{R}^p \\ \sigma > 0$$

en donde

$$S(\underline{\beta}) = \underline{\beta}'_* D \underline{\beta}_* = v S^2 + (\underline{\beta} - \hat{\underline{\beta}})' D_p (\underline{\beta} - \hat{\underline{\beta}})$$

con $S^2 = S(\hat{\underline{\beta}})/v$ y $v = n-p$.

Entonces,

$$p(\sigma | \underline{x}_n) = \int_{\mathbb{R}^p} p(\underline{\beta}, \sigma | \underline{x}_n) d\underline{\beta} \propto \sigma^{-(n+1)} \int_{\mathbb{R}^p} \exp\left\{\frac{-1}{2\sigma^2} S(\underline{\beta})\right\} d\underline{\beta} = \\ = \sigma^{-(n+1)} \exp\left\{\frac{-vS^2}{2\sigma^2}\right\} \int_{\mathbb{R}^p} \exp\left\{\frac{-1}{2} (\underline{\beta} - \hat{\underline{\beta}})' (\sigma^2 D_p^{-1})^{-1} (\underline{\beta} - \hat{\underline{\beta}})\right\} d\underline{\beta} = \\ = \sigma^{-(n+1)} \exp\left\{\frac{-vS^2}{2\sigma^2}\right\} (2\pi)^{p/2} |\sigma^2 D_p^{-1}|^{1/2} \propto \sigma^{-(n+p-1)} \exp\left\{\frac{-vS^2}{2\sigma^2}\right\},$$

$\sigma > 0$.

Así que la distribución final de σ es una gamma invertida

con parámetros $\left(\frac{\nu}{2}, \frac{\nu S^2}{2} \right)$,

$$p(\sigma | \underline{x}_n) = 2 \left(\frac{\nu S^2}{2} \right)^{\nu/2} \frac{1}{\Gamma\left(\frac{\nu}{2}\right)} \sigma^{-(\nu+1)} \exp\left\{ \frac{-\nu S^2}{2\sigma^2} \right\}, \quad \sigma > 0. \quad (2.2.38)$$

Intervalos de alta densidad para σ se pueden calcular a partir de esta última distribución.

Distribución predictiva.

Uno de los objetivos más usuales en series de tiempo es la predicción de valores futuros. En estadística bayesiana, este problema es tratado haciendo uso de la llamada distribución predictiva, la cuál fué definida en la primera sección de este capítulo.

Supongamos que se desea hacer predicciones sobre los valores de X_t , $t = n+1, n+2, \dots, n+k$, $k \in \mathbb{N}$, y sea

$\underline{x}'_{(k)} = (x_{n+1}, x_{n+2}, \dots, x_{n+k})$. La distribución predictiva de $\underline{x}_{(k)}$ está dada por

$$p(\underline{x}_{(k)} | \underline{x}_n) = \int p(\underline{x}_{(k)} | \underline{\beta}, \sigma, \underline{x}_n) p(\underline{\beta}, \sigma | \underline{x}_n) d\underline{\beta} d\sigma, \quad \underline{x}_{(k)} \in \mathbb{R}^k,$$

pero como

$$p(\underline{x}_{(k)} | \underline{\beta}, \sigma, \underline{x}_n) = \frac{p(x_{n+k} | \underline{\beta}, \sigma)}{p(x_n | \underline{\beta}, \sigma)}$$

y

$$p(\underline{\beta}, \sigma | \underline{x}_n) = \frac{p(\underline{x}_n | \underline{\beta}, \sigma) p(\underline{\beta}, \sigma)}{p(\underline{x}_n)},$$

entonces

$$p(\underline{x}_{(k)} | \underline{x}_n) \propto \int p(\underline{x}_{n+k} | \underline{\beta}, \sigma) p(\underline{\beta}, \sigma) d\sigma d\underline{\beta}, \quad \underline{x}_{(k)} \in \mathbb{R}^k. \quad (2.2.39)$$

Consideremos primero la predicción de x_{n+1} , así que $k = 1$. Sea D_* la matriz $(p+1) \times (p+1)$ que se obtiene en forma análoga a la matriz D definida en (2.2.10), sólo que ahora con $n+1$ en lugar de n y $(x_1, x_2, \dots, x_{n+1})$ en lugar de (x_1, x_2, \dots, x_n) . Será necesario particionar a D_* como

$$D_* = \begin{pmatrix} d_{11} & -d_*' \\ -d_* & D \end{pmatrix} \quad (2.2.40)$$

en donde $d_{11} \in \mathbb{R}$. Además, sea $\underline{x}_*' = (x_n, x_{n-1}, \dots, x_{n-(p-1)})$, y \underline{h} el vector con i -ésimo elemento dado por $h_i = \sum_{j=1}^{n-i} x_j x_{j+i}$, $i = 1, 2, \dots, p$.

Análogamente a lo hecho en la obtención de (2.2.12), $p(\underline{x}_{n+1} | \underline{\beta}, \sigma)$ está dada por

$$p(\underline{x}_{n+1} | \underline{\beta}, \sigma) = (2\pi\sigma^2)^{-(n+1)/2} |M_p(p, 0)|^{1/2} \exp\left\{ \frac{-1}{2\sigma^2} \underline{\beta}_*' D_* \underline{\beta}_* \right\},$$

$\underline{x}_{n+1} \in \mathbb{R}^{n+1}$

siendo $M_p^{(p, \theta)}$ y $\underline{\beta}_*$ las mismas que en (2.2.12). Además se recuerda que

$$p(\underline{\beta}, \sigma) \propto \sigma^{-1} |M_p^{(p, \theta)}|^{-1/2}, \quad \underline{\beta} \in \mathbb{R}^p, \quad \sigma > 0,$$

entonces, de acuerdo a (2.2.39),

$$p(x_{n+1} | \underline{x}_n) \propto \int \sigma^{-(n+1)} \exp\left\{\frac{-1}{2\sigma^2} S_*(\underline{\beta})\right\} d\sigma d\underline{\beta}, \quad x_{n+1} \in \mathbb{R},$$

en donde $S_*(\underline{\beta}) = \underline{\beta}_*^! D_* \underline{\beta}_*$.

Al integrar con respecto a σ se obtiene

$$p(x_{n+1} | \underline{x}_n) \propto \int |S_*(\underline{\beta})|^{-(n+1)/2} x_{n+1} \in \mathbb{R}. \quad (2.2.41)$$

En forma semejante a lo expresado en (2.2.33) y (2.2.34), se puede escribir $S_*(\underline{\beta})$ como

$$S_*(\underline{\beta}) = v_* S_*^2 + (\underline{\beta} - \tilde{\underline{\beta}})' \mathcal{D} (\underline{\beta} - \tilde{\underline{\beta}}),$$

en donde $v_* = n+1-p$, $\tilde{\underline{\beta}} = \mathcal{D}^{-1} \underline{d}_*$, $v_* S_*^2 = S_*(\tilde{\underline{\beta}})$ y

$$S_*(\tilde{\underline{\beta}}) = \tilde{\underline{\beta}}_! D_* \tilde{\underline{\beta}}_* = d_{11} - \tilde{\underline{\beta}}' \mathcal{D} \tilde{\underline{\beta}} = d_{11} - \underline{d}_*^! \mathcal{D}^{-1} \underline{d}_*.$$

La sustitución de $S_*(\underline{\beta})$ en (2.2.41) dá como resultado

$$P(x_{n+1} | \underline{x}_n) \propto (v_* S_*^2)^{-(n+1)/2} \int \left[1 + \frac{(\underline{\beta} - \underline{\tilde{\beta}})' \mathcal{D} (\underline{\beta} - \underline{\tilde{\beta}})}{v_* S_*^2} \right]^{-v_* + p} / 2 \quad d\underline{\beta} ,$$

$$x_{n+1} \in \mathbb{R} .$$

Aparte de unas constantes, Ésta última integral es igual a $|S_*^2 \mathcal{D}|^{-1/2}$, pero

$$|S_*^{-2} \mathcal{D}|^{-1/2} = |(S_*^{-2})^p \mathcal{D}|^{-1/2} = (S_*^2)^{p/2} |\mathcal{D}|^{-1/2} ,$$

entonces

$$P(x_{n+1} | \underline{x}_n) \propto (v_* S_*^2)^{-(n+1)/2} (S_*^2)^{p/2} \propto (v_* S_*^2)^{-(n+1-p)/2} ,$$

$$x_{n+1} \in \mathbb{R} .$$

Como antes se dijo, $v_* S_*^2 = d_{11} - \underline{d}'_* \mathcal{D}^{-1} \underline{d}_*$, y es inmediato comprobar, a partir de la definición de D_* , que

$$d_{11} = x_{n+1}^2 + \frac{x_{n+1}' x_n}{n-n} , \quad \underline{d}_* = \underline{h} + x_{n+1} \underline{x}_* ,$$

siendo \underline{h} y \underline{x}_* los vectores definidos con anterioridad.

Finalmente, al sustituir d_{11} y \underline{d}_* en $v_* S_*^2$, y después de completar cuadrados en x_{n+1} , resulta que

$$P(x_{n+1} | \underline{x}_n) \propto \left[1 + \frac{(x_{n+1} - A)^2}{(n-p) E} \right]^{-n(n-p+1)/2} , \quad x_{n+1} \in \mathbb{R}, (2.2.42)$$

en donde

$$A = \underline{x}_*^t \underline{D}^{-1} \underline{h} (1 - \underline{x}_*^t \underline{D}^{-1} \underline{x}_*)^{-1} , .$$

$$E = (n-p)^{-1} \left[(\underline{x}_n^t \underline{x}_n - \underline{h}^t \underline{D}^{-1} \underline{h}) (\underline{x}_*^t \underline{D}^{-1} \underline{h})^{-1} A - A^2 \right] .$$

Entonces la distribución predictiva para x_{n+1} es una t-Student con $n - p$ grados de libertad, localización A y precisión E^{-1} (o varianza $\frac{n-p}{n-p-2} E$) .

Con n suficientemente grande, una buena aproximación a $p(x_{n+1} | \underline{x}_n)$ es la distribución normal con media A y varianza E .

Para $k > 1$, la distribución predictiva $p(\underline{x}_{(k)} | \underline{x}_n)$ no resulta tener, por ahora, una forma explícita. Esto puede verse si se expresa $p(\underline{x}_{(k)} | \underline{x}_n)$ como

$$p(\underline{x}_{(k)} | \underline{x}_n) = \prod_{i=1}^k p(x_{n+i} | \underline{x}_{n+i+1}) , \quad \underline{x}_{(k)} \in \mathbb{R}^k . \quad (2.2.43)$$

Ya se ha demostrado que la distribución predictiva para la observación que está un paso adelante en el tiempo, es una t-Student. Entonces, es claro, de la expresión anterior, que $p(\underline{x}_{(k)} | \underline{x}_n)$ resulta ser el producto de k distribuciones t-Student similares a la obtenida para x_{n+1} .

Métodos numéricos pueden ayudar en el cálculo de $p(\underline{x}_{(k)} | \underline{x}_n)$ y de sus distribuciones marginales.

Ejemplo.

Para ejemplificar lo desarrollado en esta sección, se tomó un conjunto de 70 observaciones (dadas en el apéndice B), las cuales forman parte de una serie de 200 observaciones generadas por Wold (1965) de acuerdo al siguiente modelo AR(2),

$$X_t = 0.7X_{t-1} - 0.49X_{t-2} + Z_t .$$

En este modelo $\beta_1 = 0.7$ y $\beta_2 = -0.49$, así que $\beta_1^2 = -\beta_2$. Además, se supone que $Z_t \sim N(0, (1+\beta_2^2)/1 - \beta_2)$, $t \in \mathbb{Z}$, en donde la elección de la varianza de Z_t se hizo con el fin de que $X_t \sim N(0,1)$, $t \in \mathbb{Z}$.

La serie generada comenzó bajo el supuesto de que $X_{-1} = Z_{-1}$ y $X_0 = 0.7X_{-1} + Z_0$.

La matriz D definida en (2.2.10) resulta ser

$$D = \begin{pmatrix} 75.3904 & -36.6128 & -(-6.4518) \\ -36.6128 & 71.924 & 34.2397 \\ -(-6.4518) & 34.2397 & 70.2819 \end{pmatrix} ,$$

por lo que

$$D_p = \begin{pmatrix} 71.924 & 34.2397 \\ 34.2397 & 70.2819 \end{pmatrix}, \underline{d} = \begin{pmatrix} 36.6128 \\ -6.4518 \end{pmatrix} \quad \text{y } d_{11} = 75.3904 .$$

Entonces, el vector de localización de la distribución final de (β_1, β_2) , dada en (2.2.35) tendrá $v = n - p = 68$ grados de libertad, vector de localización igual a

$$\hat{\underline{\beta}} = D_p^{-1} \underline{d} = \begin{pmatrix} 0.7197 \\ -0.4424 \end{pmatrix}$$

y la matriz de precisión

$$S^{-2} D_p = \begin{pmatrix} 105.8907 & 50.4096 \\ 50.4096 & 103.473 \end{pmatrix}$$

El valor S^2 , calculado según (2.2.34), dió

$$S^2 = \frac{S(\hat{\underline{\beta}})}{v} = \frac{d_{11} - \underline{d}' D_p^{-1} \underline{d}}{v} = \frac{46.1876}{68} = 0.6792 .$$

Una región de mayor densidad, que con una probabilidad de 0.99 contenga a (β_1, β_2) , fué construida de acuerdo a (2.2.36), dando como resultado la región contenida en la elipse de la fig. 2.1.

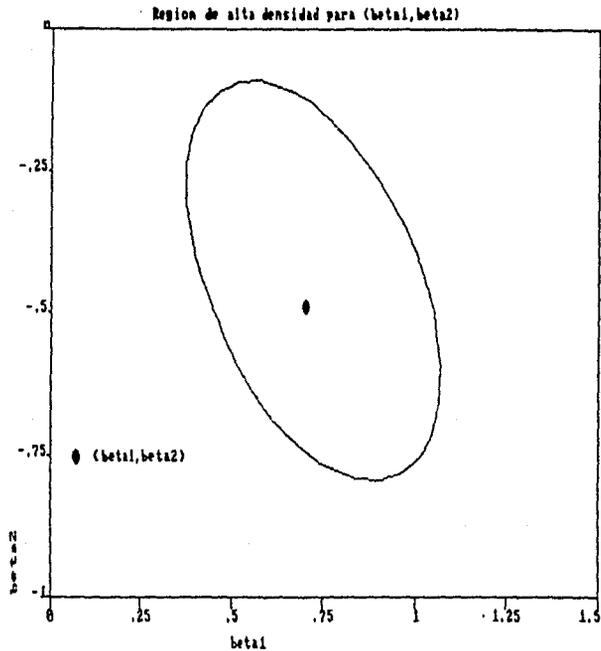


Figura 2.1.

Utilizando la aproximación normal dada por

$$\underline{\beta} \sim N(\hat{\underline{\beta}}, S^2 D^{-1}) ,$$

se tienen las siguientes distribuciones marginales, finales

$$\beta_1 \sim N(0.7197, 0.0123) ,$$

$$\beta_2 \sim N(-0.4424, 0.0126) .$$

La función de densidad correspondiente a cada una de estas dos últimas distribuciones, se presenta en la fig. 2.2.

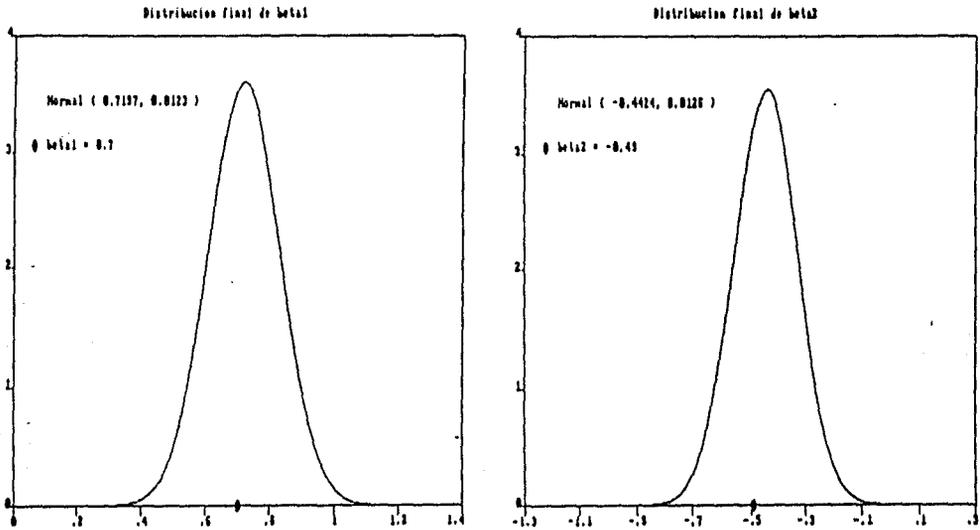


Figura 2.2.

Con respecto a la distribución final de σ , dada en (2.2.38), se llega a una distribución gamma invertida con parámetros $\nu/2 = 34$ y $\nu S^2/2 = 23.0938$. La función de densidad es entonces

$$p(\sigma | \underline{x}_{70}) = (5262392231) \sigma^{-69} \exp\left\{-\frac{23.0938}{\sigma^2}\right\}, \quad \sigma > 0$$

Algunas características de esta distribución son,

$$\text{Moda} = \left(\frac{\nu S^2}{\nu+1}\right)^{1/2} = 0.8182$$

$$E(\sigma) = \frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)} \left(\frac{\nu S^2}{2}\right)^{1/2} = 0.8242$$

$$V(\sigma) = \frac{vS^2}{v-2} - E^2(\sigma) = 0.0206$$

así que

$$\sqrt{V(\sigma)} = 0.1435$$

En la fig. 2.3. se presenta una gráfica de la densidad $p(\sigma | \underline{x}_{10})$ y se puede apreciar en donde está situado el verdadero valor de σ , que es igual a

$$\sigma = \left[(1-\beta_2)^3 / (1-\beta_2) \right]^{1/2} = 0.7695 .$$

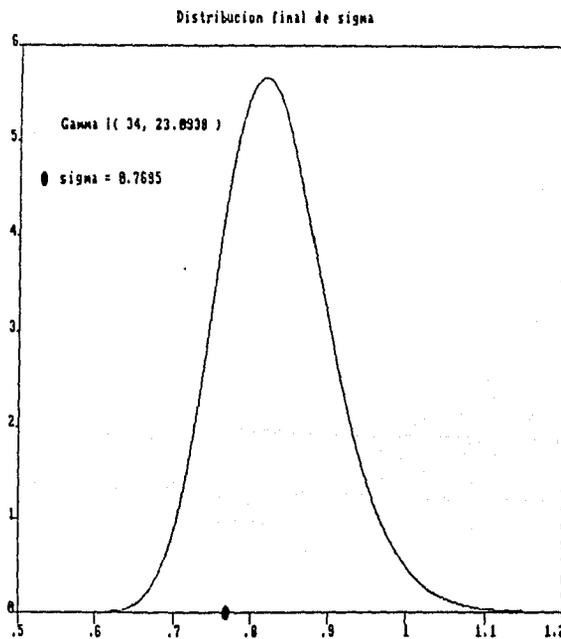


Figura 2.3.

La distribución predictiva para $x_{7,1}$ requiere el cálculo de la matriz D , definida en (2.2.40), que resulta ser

$$D = \begin{pmatrix} 73.9181 & 35.7263 \\ 35.7263 & 71.3903 \end{pmatrix}.$$

También son necesarios los vectores \underline{x}_* y \underline{h} , definidos inmediatamente después de la expresión (2.2.40), los cuales son igual a

$$\underline{x}_* = \begin{pmatrix} 1.0528 \\ 1.4121 \end{pmatrix} \quad \underline{h} = \begin{pmatrix} 36.6128 \\ -6.4518 \end{pmatrix}$$

Finalmente, y de acuerdo a (2.2.42), $p(x_{7,1} | \underline{x}_{7,0})$ es una distribución t de Student con $n - p = 68$ grados de libertad, localización igual a 0.1222 y precisión 0.7045. La aproximación normal a esta distribución esta dada por $N(0.1222, 1.4194)$.

En la fig. 2.4. se presenta la gráfica de ésta última densidad junto con el valor $x_{7,1} = -0.682$, correspondiente al valor generado.

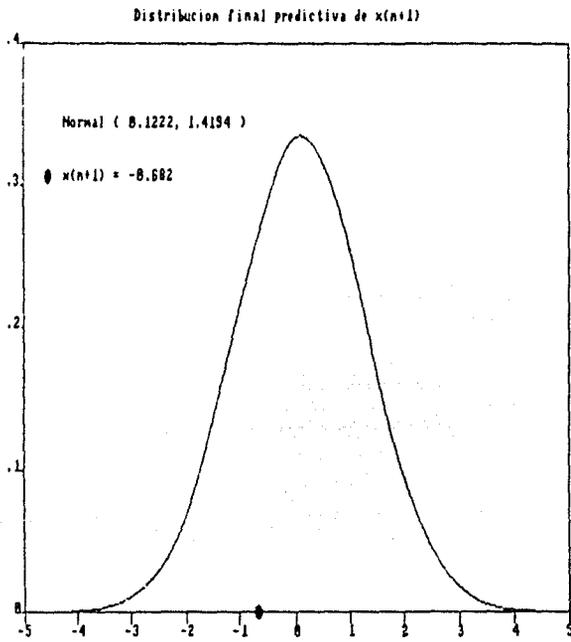


Figura 2.4.

2.3. EXTENSION A OTROS PROCESOS

Lo desarrollado en la sección anterior para procesos autoregresivos se puede aplicar a procesos como el MA(q) ó el ARMA(p,q). Para ver cómo es posible ésto, comencemos por considerar un proceso estacionario ARMA(p,q) y una serie de observaciones $\underline{x}'_n = (x_1 \ x_2 \ \dots \ x_n)$ generadas por dicho proceso, entonces,

$$X_t - \beta_1 X_{t-1} - \dots - \beta_p X_{t-p} = Z_t - \alpha_1 Z_{t-1} - \dots - \alpha_q Z_{t-q} \quad , \quad (2.3.1.)$$

El conjunto $\{Z_t : t \in Z\}$, es un conjunto de variables aleatorias independientes e idénticamente distribuidas con $E(Z_t) = 0$ y $V(Z_t) = \sigma^2 < \infty$. También se supone, sin pérdida de generalidad que $E(X_t) = 0$. Los parámetros de interés son $\underline{\beta}' = (\beta_1 \ \beta_2 \ \dots \ \beta_p)$, $\underline{\alpha}' = (\alpha_1 \ \alpha_2 \ \dots \ \alpha_q)$ y σ .

El modelo anterior se puede expresar en términos del operador de retraso B, como

$$[A(B)] X_t = [M(B)] Z_t \quad , \quad (2.3.2.)$$

donde $A(Z) = 1 - \sum_{i=1}^p \beta_i Z^i$, $M(Z) = 1 - \sum_{j=1}^q \alpha_j Z^j$.

Sean ahora $\{G_i^{-1}\}_{i=1}^p$ y $\{H_j^{-1}\}_{j=1}^q$ las soluciones respec

tivas de las ecuaciones

$$A(Z) = 0 \quad , \quad M(Z) = 0 \quad ,$$

entonces (2.3.2.) se puede escribir en la forma

$$\left[\begin{array}{c} p \\ \prod_{i=1} (1 - G_i B) \end{array} \right] X_t = \left[\begin{array}{c} q \\ \prod_{j=1} (1 - H_j B) \end{array} \right] Z_t \quad . \quad (2.3.3.)$$

Consideremos $\{H_j^0\}_{j=1}^q$ valores tal que $H_j = H_j^0 + \delta_j$, con δ_j una cantidad pequeña, $j = 1, 2, \dots, q$.

Hagamos la siguiente transformación

$$\left[\begin{array}{c} q \\ \prod_{j=1} (1 - H_j^0 B) \end{array} \right] W_t = X_t \quad (2.3.4.)$$

Con esta transformación es posible generar una serie de W_t 's dadas las $\{X_t\}_{t=1}^n$, pues si la transformación se denota por $T^{-1} \underline{W}_n = \underline{X}_n$, T^{-1} es una matriz triangular $n \times n$ con elementos en la diagonal iguales a la unidad, por lo que $|T^{-1}| = 1$, (de donde también $|T| = 1$).

Sean ahora

$$\underline{G}' = (G_1, G_2, \dots, G_p) \quad , \quad \underline{H}' = (H_1, H_2, \dots, H_q)$$

$$\underline{H}^0' = (H_1^0, H_2^0, \dots, H_q^0) \quad , \quad \underline{\delta}' = (\delta_1, \delta_2, \dots, \delta_q) \quad .$$

La verosimilitud asociada a los parámetros \underline{G} , \underline{H} y σ se denotará por

$$L(\underline{G}, \underline{H}, \sigma \mid \underline{x}_n, (p, q))$$

o alternativamente,

$$L(\underline{G}, \underline{H}^0 + \underline{\delta}, \sigma \mid \underline{x}_n, (p, q)) ,$$

pues $\underline{H} = \underline{H}^0 + \underline{\delta}$. Al considerar en la verosimilitud parámetros \underline{G} y $\underline{H} = \underline{H}^0 + \underline{\delta}$, se está pensando que el proceso que generó los datos \underline{x}_n es el ARMA(p, q)

$$\left[\prod_{i=1}^p (1 - G_i B) \right] X_t = \left[\prod_{j=1}^q (1 - H_j^0 + \delta_j) B \right] Z_t , \quad (2.3.5.)$$

y el proceso que generó las w_n está dado entonces por

$$\left[\prod_{i=1}^p (1 - G_i B) \prod_{j=1}^q (1 - H_j^0 B)^2 \right] W_t = \left[\prod_{j=1}^q (1 - H_j^0 + \delta_j) B \right] Z_t , \quad (2.3.6.)$$

lo cuál es equivalente, puesto que δ_j es considerado pequeño, $j = 1, 2, \dots, q$, a que las w_n esten generadas, aproximadamente por el modelo.

$$\left[\prod_{i=1}^p (1 - G_i B) \prod_{j=1}^q (1 - (H_j^0 - \delta_j) B) \right] W_t = Z_t . \quad (2.3.7.)$$

Lo importante es hacer notar que éste último modelo es un AR(p+q), siendo la verosimilitud correspondiente

$$L(\underline{G}, \underline{H}^0 - \underline{\delta}, \sigma \mid \underline{w}_n, (p+q, 0)) \dots$$

La aproximación que se está considerando en (2.3.7.) es tal que, aún para muestras no muy grandes, no presente efectos significativos, de tal forma que es posible afirmar que

$$L(\underline{G}, \underline{H}^0 + \underline{\delta}, \sigma \mid \underline{x}_n, (p, q)) \doteq L(\underline{G}, \underline{H}^0 - \underline{\delta}, \sigma \mid \underline{w}_n, (p+q, 0)) ,$$

sin olvidar que δ_j , $j = 1, 2, \dots, q$, es considerado pequeño.

Lo hecho anteriormente nos permite establecer lo siguiente, supongamos que para el modelo ARMA(p,q)

$$\left[\prod_{i=1}^p (1 - G_i B) \right] X_t = \left[\prod_{j=1}^q (1 - H_j B) \right] Z_t$$

los parámetros $(\underline{G}, \underline{H})$ tienen asociada una matriz de información $I\{\underline{G}, \underline{H} \mid (p, q)\}$. Del mismo modo sea $I\{\underline{G}, \underline{H} \mid (p+q, 0)\}$ la matriz de información que corresponde a los parámetros $(\underline{G}, \underline{H})$ del modelo

$$\left[\prod_{i=1}^p (1 - G_i B) \prod_{j=1}^q (1 - H_j B) \right] X_t = Z_t$$

Sea

$$I\{\underline{G}, \underline{H} \mid (p+q, 0)\} = \begin{pmatrix} I_{\underline{G}} & I_{\underline{G} \underline{H}} \\ I'_{\underline{G} \underline{H}} & I_{\underline{H}} \end{pmatrix} ,$$

entonces, como para muestras no pequeñas se tiene que

$$I(\underline{G}, \underline{H} \mid (p, q)) \approx I(\underline{G}, -\underline{H} \mid (p+q, 0)) = \begin{pmatrix} I_G & -I_{\underline{G} \underline{H}} \\ -I'_{\underline{G} \underline{H}} & I_H \end{pmatrix},$$

es claro que los determinantes de las matrices de información para un proceso ARMA(p,q) y un proceso AR(p+q) son aproximadamente iguales.

En términos de los parámetros originales este último resultado se traduce en lo siguiente: sea

$$\prod_{i=1}^p (1 - G_i B) \prod_{j=1}^q (1 - H_j B) = 1 - \beta_1^* B - \beta_2^* B^2 - \dots - \beta_{p+q}^* B^{p+q},$$

entonces, por lo visto en la sección anterior, la matriz de información asociada a los parámetros

$\underline{\beta}^* = (\beta_1^* \beta_2^* \dots \beta_{p+q}^*)$, $I(\underline{\beta}^*)$, es de la forma

$$I(\underline{\beta}^*) = n^{p+q} \begin{pmatrix} (p+q, 0) \\ M_{p+q} \end{pmatrix}^{-1} \doteq n^{p+q} \begin{pmatrix} (p, q) \\ M_n \end{pmatrix}, \quad (2.3.8.)$$

en donde la aproximación se justifica por lo demostrado en esta sección.

También y por propiedades de la matriz de información, se sabe que

$$|I(\underline{G}, \underline{H})|^{1/2} = |I(\underline{\beta}^*)|^{1/2} |J(\underline{\beta}^* | \underline{G}, \underline{H})| , \quad (2.3.9.)$$

donde

$$J(\underline{\beta}^* | \underline{G}, \underline{H}) = \frac{\partial \underline{\beta}^*}{\partial (\underline{G}, \underline{H})} ,$$

es el jacobiano de la transformación de $\underline{\beta}^*$ a $(\underline{G}, \underline{H})$.

Ahora bien, en general es posible demostrar que para un proceso ARMA(p,q), con el dado en (2.3.1.), la función de verosimilitud es de la forma,

$$p(\underline{x}_n | \underline{\beta}, \underline{\alpha}, \sigma) \propto \sigma^{-n} f(\underline{\beta}, \underline{\alpha}) \exp\left\{\frac{-1}{2\sigma^2} S(\underline{\beta}, \underline{\alpha})\right\} , \quad (2.3.10.)$$

donde,

$$S(\underline{\beta}, \underline{\alpha}) = \sum_{t=-\infty}^n E^2(z_t | \underline{x}_n, \underline{\beta}, \underline{\alpha}) \quad (2.3.11.)$$

Con una distribución inicial no informativa sobre $(\underline{\beta}, \underline{\alpha}, \sigma)$ dada por

$$p(\underline{\beta}, \underline{\alpha}, \sigma) \propto \sigma^{-1} |I(\underline{\beta}, \underline{\alpha})|^{1/2} , \quad (2.3.12.)$$

la distribución final estará dada por

$$p(\underline{\beta}, \underline{\alpha}, \sigma | \underline{x}_n) \propto \sigma^{-(n+1)} |I(\underline{\beta}, \underline{\alpha})|^{1/2} f(\underline{\beta}, \underline{\alpha}) \exp\left\{\frac{-1}{2\sigma^2} S(\underline{\beta}, \underline{\alpha})\right\} ,$$

de donde la marginal de $(\underline{\beta}, \underline{\alpha})$, obtenida de la misma forma que $p(\underline{\beta} | \underline{x}_n)$ en la sección anterior, será

$$p(\underline{\beta}, \underline{\alpha} | \underline{x}_n) \propto |I(\underline{\beta}, \underline{\alpha})|^{1/2} f(\underline{\beta}, \underline{\alpha}) |S(\underline{\beta}, \underline{\alpha})|^{-n/2} . \quad (2.3.13.)$$

En el caso de un proceso autoregresivo, los factores $|I(\underline{\beta})|^{1/2}$ y $f(\underline{\beta})$ se cancelan, salvo una constante, como ya se demostró anteriormente.

2.4. PROCESO DE PROMEDIOS MOVILES

El modelo para un proceso MA(q) , que puede ser visto como un proceso ARMA(0,q) , es de la forma

$$X_t = Z_t - \alpha_1 Z_{t-1} - \dots - \alpha_q Z_{t-q} , \quad (2.4.1.)$$

en donde las suposiciones distribucionales sobre $\{X_t : t \in \mathbf{Z}\}$ y $\{Z_t : t \in \mathbf{Z}\}$ son las mismas que se han manejado en todo este capítulo.

Con respecto a la verosimilitud y a la distribución inicial dada para el caso general en (2.3.10.) y (2.3.12.) , para el proceso MA(q) se tiene que

$$p(\underline{x}_n | \underline{\alpha}, \sigma) \propto \sigma^{-n} |M_n^{(0,q)}|^{1/2} \exp\left\{\frac{-1}{2\sigma^2} S(\underline{\alpha})\right\} ,$$

$$p(\underline{\alpha}, \sigma) \propto |I(\underline{\alpha}, \sigma)|^{1/2} = \sigma^{-1} |I(\underline{\alpha})|^{1/2} .$$

Pero, de acuerdo al resultado visto en la sección anterior, para n grandes,

$$|I(\underline{\alpha})| \doteq n^q |M_n^{(0,q)}|^{-1} ,$$

por lo que la distribución final de (α, σ) será

$$p(\underline{\alpha}, \sigma | \underline{x}_n) \propto \sigma^{-(n+1)} \exp \frac{-1}{2\sigma^2} S(\underline{\alpha}) \quad (2.4.2.)$$

Entonces,

$$p(\underline{\alpha} | \underline{x}_n) \propto \{S(\alpha)\}^{-n/2} \quad (2.4.3.)$$

A diferencia de un AR(p), ver (2.2.9.), la suma de cuadrados $S(\underline{\alpha})$ no es exactamente una forma cuadrática en $\underline{\alpha}$, lo cual se puede demostrar viendo que $E(Z_t | \underline{x}_n, \underline{\alpha})$, en (2.3.11.), no es lineal en las $\alpha_1, \alpha_2, \dots, \alpha_q$. Sin embargo, para n grande, se espera que la linealidad de $E(Z_t | \underline{x}_n, \underline{\alpha})$ se cumpla aproximadamente dentro de ciertos rangos, de donde es posible decir que aproximadamente

$$S(\underline{\alpha}) = v S^2 + \frac{1}{2} \sum_{i,j} S_{ij} (\alpha_i - \hat{\alpha}_i)(\alpha_j - \hat{\alpha}_j) \quad (2.4.4.)$$

donde, $S^2 = \frac{S(\hat{\alpha})}{v}$, $v = n - q$, $S_{ij} = \frac{\partial^2 S(\underline{\alpha})}{\partial \alpha_i \partial \alpha_j}$.

El porqué de esta última expresión se puede comprender si uno se dá cuenta que no es más que el análogo al caso AR(p) y que esta dado en (2.2.33.), en donde por cierto $S_{ij} = 2 d_{i+1, j+1}$.

Al substituir (2.4.4.) en (2.4.3.) se tiene entonces que

$$p(\underline{\alpha} | \underline{x}_n) \propto \left[1 + \sum_{i,j} \frac{S_{ij}}{2S^2} (\alpha_i - \hat{\alpha}_i)(\alpha_j - \hat{\alpha}_j) \right], \quad (2.4.5.)$$

lo cual, al estar considerando muestras grandes, nos dice que aproximadamente,

$$p(\underline{\alpha} \mid \underline{x}_n) \propto \exp(-(\underline{\alpha} - \hat{\underline{\alpha}})' \{S_{ij}\} (\underline{\alpha} - \hat{\underline{\alpha}}) / 2S^2) \quad , \quad (2.4.6.)$$

es decir,

$$p(\underline{\alpha} \mid \underline{x}_n) = N_{\underline{q}}(\hat{\underline{\alpha}} \mid 2S^2 \{S_{ij}\}^{-1}) \quad . \quad (2.4.7.)$$

Una γ -región de mayor densidad para $\underline{\alpha}$, es fácilmente construída si se recuerda que al cumplirse (2.4.7.) se tiene que

$$(\underline{\alpha} - \hat{\underline{\alpha}})' \{S_{ij}\} (\underline{\alpha} - \hat{\underline{\alpha}}) / 2S^2 \sim \chi^2_{(q)} \quad ,$$

por lo que la región buscada será

$$R_{\gamma} = \{ \underline{\alpha} \in \mathbb{R}^q \mid (\underline{\alpha} - \hat{\underline{\alpha}})' \{S_{ij}\} (\underline{\alpha} - \hat{\underline{\alpha}}) < 2S^2 \chi^2_{(q), 1-\alpha} \}$$

donde $\chi^2_{(q), 1-\alpha}$ es el cuantil $1-\alpha$ de la distribución $\chi^2_{(q)}$.

2.5. PROCESO AUTOREGRESIVO Y DE PROMEDIOS MOVILES

El proceso ARMA(p,q) se ha dado en (2.3.1.) con los supuestos que éste tiene. Consideremos el vector de parámetros $\underline{\beta}^*$ dado por la transformación

$$\left[1 - \sum_{i=1}^p \beta_i z^i \right] \left[1 - \sum_{j=1}^q \alpha_j z^j \right] = 1 - \sum_{k=1}^{p+q} \beta_k^* z^k .$$

De acuerdo a lo desarrollado en la sección 2.2. el proceso AR(p+q) definido por

$$\left[1 - \sum_{k=1}^{p+q} \beta_k^* z^k \right] x_t = z_t$$

asocia una matriz de información, $I(\underline{\beta}^*)$, tal que

$$|I(\underline{\beta}^*)| = n^{p+q} |M_{p+q}^{(p+q,0)}|^{-1}$$

También al analizar (2.2.6.) se deduce que la función $f(\underline{\beta}, \underline{\alpha})$, en este caso $f(\underline{\beta}^*)$, que aparece en (2.2.10.), es igual a

$$f(\underline{\beta}^*) = \left| M_{p+q}^{(p+q,0)} \right|^{1/2}$$

es decir

$$|I(\underline{\beta}^*)| = f^{-2}(\underline{\beta}^*) .$$

Aplicando (2.3.9.), es claro que

$$|I(\underline{\beta}, \underline{\alpha})|^{1/2} = |I(\underline{\beta}^*)|^{1/2} |J(\underline{\beta}^* | \underline{\beta}, \underline{\alpha})| ,$$

al sustituir $|I(\underline{\beta}^*)|$,

$$|I(\underline{\beta}, \underline{\alpha})|^{1/2} f(\underline{\beta}^*) \propto |J(\underline{\beta}^* | \underline{\beta}, \underline{\alpha})| ,$$

por lo que finalmente

$$|I(\underline{\beta}, \underline{\alpha})|^{1/2} f(\underline{\beta}, \underline{\alpha}) \propto |J(\underline{\beta}^* | \underline{\beta}, \underline{\alpha})| .$$

De esta forma la distribución final de $(\underline{\beta}, \underline{\alpha})$ que aparece en (2.3.13), está dada por

$$p(\underline{\beta}, \underline{\alpha} | \underline{x}_n) \propto |J(\underline{\beta}^* | \underline{\beta}, \underline{\alpha})| \left\{ S(\underline{\beta}, \underline{\alpha}) \right\}^{-n/2} .$$

2.6. DISTRIBUCION INICIAL INFORMATIVA

En las secciones anteriores el proceso de inferencia se llevó a cabo haciendo uso de la distribución inicial de referencia dada por la regla de Jeffreys. En esta sección se presentarán los resultados que se obtienen cuando se utiliza una distribución inicial informativa, la cual pertenece a una clase específica de distribuciones. No se especifica todo el desarrollo para llegar a dichos resultados pues éste se puede encontrar en Broemeling (1985).

Proceso AR.

Sea el modelo AR(p) dado por

$$X_t = \sum_{j=1}^p \beta_j X_{t-j} + Z_t$$

en donde $E(X_t) = 0$, $Z_t \sim N(0, \tau^{-1})$ para toda t y $\text{COV}(Z_t, Z_s) = 0$ para toda t, s .

La distribución inicial para $(\underline{\beta}, \tau)$ está dada por $p(\underline{\beta}, \tau) = p(\underline{\beta}|\tau)p(\tau)$, en donde $p(\underline{\beta}|\tau)$ es la distribución normal

$$p(\underline{\beta}|\tau) \propto \tau^{p/2} \exp \left\{ -\frac{\tau}{2} (\underline{\beta} - \underline{\mu})' P (\underline{\beta} - \underline{\mu}) \right\}, \quad \underline{\beta} \in \mathbb{R}^p$$

y $p(\tau)$ es la distribución gamma dada por

$$p(\tau) \propto \tau^{a-1} e^{-\tau b}, \quad \tau > 0$$

Esta distribución inicial pertenece a una familia conjugada, lo que significa que la distribución final también pertenecerá a la misma familia.

Sean $\underline{x}_n^1 = (x_1, x_2, \dots, x_n)$ las n observaciones del proceso y $x_0, x_{-1}, \dots, x_{-p}$ observaciones iniciales que se suponen constantes (si éstos últimos valores no se tienen, puede suponerse que todos son cero). La distribución final de

$(\underline{\beta}, \tau)$ resulta ser

$$p(\underline{\beta}, \tau | \underline{x}_n) \propto \tau^{(n+2a+p/2)-1} \exp\left\{-\frac{\tau}{2} | (\underline{\beta} - A^{-1}C)'A(\underline{\beta} - A^{-1}C) + D |\right\}$$

$$\underline{\beta} \in \mathbb{R}^p, \quad \tau > 0$$

en donde si G es la matriz $p \times p$ con elemento (i, j) dado

$$\text{por } g_{ij} = \sum_{k=1}^n x_{k-i} x_{k-j} \text{ y } \underline{h}' = (h_1, h_2, \dots, h_p) \text{ con}$$

$$h_j = \sum_{k=1}^n x_k x_{k-j}, \quad j = 1, 2, \dots, p, \text{ entonces}$$

$$A = P + G, \quad C = P\underline{\mu} + \underline{h} \quad \text{y}$$

$$D = 2b + \underline{\mu}'P\underline{\mu} + \sum_{k=1}^n x_k^2 - C'AC$$

Integrando la distribución conjunta de $\underline{\beta}$ y τ con respecto a τ , se obtiene

$$p(\underline{\beta} | \underline{x}_n) \propto \left[(\underline{\beta} - A^{-1}C)'A(\underline{\beta} - A^{-1}C) + D \right]^{-(n+2a+p)/2}, \quad \underline{\beta} \in \mathbb{R}^p$$

que a su vez se puede expresar como,

$$p(\underline{\beta} | \underline{x}_n) \propto \left[1 + \frac{(\underline{\beta} - A^{-1}C)'A(\underline{\beta} - A^{-1}C)}{D} \right]^{-(n+2a+p)/2}, \quad \underline{\beta} \in \mathbb{R}^p$$

Entonces la distribución marginal final de $\underline{\beta}$ resulta ser una distribución t de Student con $n+2a$ grados de libertad, vector de localización $A^{-1}C$ y matriz de precisión $(n+2a)D^{-1}A$.

La distribución correspondiente a τ esta dada por

$$p(\tau | \underline{x}_n) \propto \tau^{((n+2a)/2)-1} \exp\left\{-\frac{\tau}{2}\left[2b + \underline{\mu}'P\underline{\mu} + \sum_{k=1}^n x_k^2 - C'A^{-1}C\right]\right\}, \tau > 0$$

asi que $p(\tau | \underline{x}_n)$ es una distribución gamma con parámetros $(n+2a)/2$ y $\frac{1}{2}\left[2b + \underline{\mu}'P\underline{\mu} + \sum_{k=1}^n x_k^2 - C'A^{-1}C\right]$.

Por último, la distribución predictiva de $x = x_{n+1}$ resulta ser una distribución t de Student con $n+2a$ grados de libertad, vector de localización

$$\left[1 - F_*(A+E)^{-1}F_*\right]^{-1}F_*(A+E)^{-1}C$$

y precisión $(n+2) \left[1 - F_*(A+E)^{-1}F_*\right]x$

$$x\{D + C'A^{-1}C - C'(A+E)^{-1}F_*\left[1 - F_*(A+E)^{-1}F_*\right]^{-1}F_*(A+E)^{-1}C\}$$

donde E es la matriz $p \times p$ con elemento (i,j) igual a $x_{n-(i-1)}x_{n-(j-1)}$ y $F_* = (x_n, x_{n-1}, \dots, x_{n-(p-1)})$.

Al igual que en la sección 2.2., en donde se utilizó una distribución inicial informativa, la distribución predictiva de $(X_{n+1}, X_{n+2}, \dots, X_{n+k})$ está dada por el producto de distribuciones t-Student, cada una similar a la encontrada para X_{n+1} .

Proceso MA

Consideremos el proceso MA(1)

$$X_t = Z_t - \alpha_1 Z_{t-1}$$

en donde $Z_t \sim N(0, \tau^{-1})$ y $\text{COV}(Z_t, Z_s) = 0$ para todo t, s .

Sea $A(\alpha_1)$ la matriz $n \times (n+1)$ definida como

$$A(\alpha_1) = \begin{pmatrix} -\alpha_1 & 1 & 0 & \dots & 0 \\ 0 & -\alpha_1 & 1 & 0 & \dots & 0 \\ & & & \dots & & \\ 0 & 0 & \dots & -\alpha_1 & 0 \end{pmatrix}$$

La matriz $A(\alpha_1)A'(\alpha_1)$ resulta ser simétrica y definida positiva, así que existe una matriz ortogonal Q tal que $Q'A(\alpha_1)A'(\alpha_1)Q = D(\alpha_1)$, en donde $D(\alpha_1)$ es la matriz diagonal con elemento i dado por $d_i(\alpha) = 1 + \alpha^2 - 2\alpha \cos \frac{i\pi}{n+1}$.

Suponiendo que α_1 y τ son independientes y que la distribución inicial de τ es gamma con parámetros a y b , se obtiene la distribución final conjunta de (α_1, τ) ,

$$p(\alpha_1, \tau | \underline{x}_n) \propto \tau^{((n+2a)/2)-1} p(\alpha_1) \prod_{i=1}^n d_i^{-1/2}(\alpha_1) x$$

$$x \left\{ \exp \frac{-\tau}{2} \left[2b + \underline{x}'_n Q D^{-1}(\alpha_1) Q' \underline{x}_n \right] \right\}, \quad \alpha \in \mathbb{R}, \quad \tau > 0$$

y la distribución marginal de α_1 es

$$p(\alpha_1 | \underline{x}_n) \propto \frac{p(\alpha_1) \prod_{i=1}^n d_i^{1/2}(\alpha_1)}{\left[2b + \underline{x}'_n Q D^{-1}(\alpha_1) Q' \underline{x}_n \right]^{(n+2a)/2}}, \quad \alpha_1 > 0$$

Esta distribución se puede analizar numéricamente.

La distribución marginal de τ se debe determinar, también numéricamente, a partir de la distribución conjunta $p(\alpha_1, \tau | \underline{x}_n)$.

Dado α_1 , es claro que $p(\tau | \alpha_1, \underline{x}_n)$ resulta ser gamma con parámetros $(n+2\alpha)/2$ y $\frac{1}{2} \left[2b + \underline{x}'_n Q D^{-1}(\alpha_1) Q' \underline{x}_n \right]$.

Para encontrar la distribución predictiva de $x = x_{n+1}$, es necesario considerar las cantidades Q_* , $A_*(\alpha_1)$, $D_*(\alpha_1)$, que se definen en forma análoga a Q, A y $D(\alpha_1)$, pero ahora pensando que el vector de observaciones es $(x_1, x_2, \dots, x_n, x_{n+1})$. Además, sea $E(\alpha_1) = Q_* D_*^{-1}(\alpha_1) Q_*'$, la cual es particionada como

$$E(\alpha_1) = \begin{pmatrix} E_{11}(\alpha_1) & E_{12}(\alpha_1) \\ E_{21}(\alpha_1) & E_{22}(\alpha_1) \end{pmatrix}$$

siendo $E_{11}(\alpha_1)$ de $n \times n$ y $E_{22}(\alpha_1)$ escalar. Por último, si $A(\alpha_1) = E_{22}(\alpha_1)$, $B(\alpha_1) = \underline{x}'_n E_{12}(\alpha_1) \underline{x}_n$, y $C(\alpha_1) = \underline{x}'_n E_{11}(\alpha_1) \underline{x}_n + 2b$, la distribución predictiva de x_{n+1} , dado \underline{x}_n y α_1 , está dada por

$$p(x_{n+1} | \underline{x}_n, \alpha_1) \propto \left\{ \left[y_{n+1} - A^{-1}(\alpha_1) B(\alpha_1) \right]^2 A(\alpha_1) + \right. \\ \left. + C(\alpha_1) - B'(\alpha_1) A^{-1}(\alpha_1) B(\alpha_1) \right\}^{-(n+2\alpha+1)/2}, \quad x_{n+1} \in R$$

la cual es una distribución t-Student con $n+2$ grados de libertad, localización $A^{-1}(\alpha_1) B(\alpha_1)$ y precisión $(n+2\alpha) A(\alpha_1) \left[C(\alpha_1) - B'(\alpha_1) A^{-1}(\alpha_1) B(\alpha_1) \right]$.

Algunos momentos de la distribución predictiva $p(x_{n+1} | \underline{x}_n)$ se pueden calcular a partir de $p(x_{n+1} | \underline{x}_n, \alpha_1)$ y $p(\alpha_1 | \underline{x}_n)$, por ejemplo,

$$E(X_{n+1} | \underline{x}_n) = E_{\alpha_1 | \underline{x}_n} E(X_{n+1} | \underline{x}_n, \alpha_1)$$

$$V(X_{n+1} | \underline{x}_n) = E_{\alpha_1 | \underline{x}_n} V(X_{n+1} | \underline{x}_n, \alpha_1) + V_{\alpha_1 | \underline{x}_n} E(X_{n+1} | \underline{x}_n, \alpha_1)$$

El cálculo de $p(x_{n+1} | \underline{x}_n)$ es por métodos numéricos, pues

$$p(x_{n+1} | \underline{x}_n) = \int_R p(x_{n+1} | \underline{x}_n, \alpha_1) p(\alpha_1 | \underline{x}_n) d\alpha_1$$

En general, el modelo MA presenta algunos problemas en su análisis, pues aparecen distribuciones no conocidas. Para modelos de orden mayor a 1 se espera encontrar más dificultades.

Proceso ARMA

Sea el modelo ARMA(1,1) dado por

$$X_t = \beta_1 X_{t-1} - \alpha_1 Z_{t-1} + Z_t$$

en donde $Z_t \sim N(0, \tau^{-1})$, $E(X_t) = 0$ y $\text{COV}(Z_t, Z_s) = 0$ para todo t, s .

Broemeling obtiene la función de verosimilitud para este caso, y a partir de la forma de dicha función sugiere que la distribución inicial para $(\alpha_1, \beta_1, \tau)$ sea de la forma

$$p(\beta_1, \alpha_1, \tau) \propto p(\beta_1 | \tau) p(\alpha_1) p(\tau), \quad (\alpha_1, \beta_1) \in \mathbb{R}^2, \quad \tau > 0$$

donde

$$p(\beta_1 | \tau) \propto \exp\left\{-\frac{\tau t}{2}(\beta_1 - \mu)^2\right\}, \quad \beta_1 \in \mathbb{R}$$

$$p(\tau) \propto \tau^{a-1} e^{-\tau b}, \quad \tau > 0$$

y $p(\alpha_1)$ es una densidad propia en \mathbb{R} .

Siendo $D(\alpha_1), Q$ las mismas matrices utilizadas en el proceso MA, $\underline{x}'_n = (x_1, \dots, x_n)$ las observaciones que se tienen del proceso, y $\underline{x}'_* = (x_0, x_1, \dots, x_{n-1})$ con x_0 constante conocida, se tienen los siguientes resultados,

$$p(\alpha_1 | \underline{x}_n) \propto \frac{p(\alpha_1) \prod_{i=1}^n d_i^{1/2}(\alpha_1)}{A^{1/2}(\alpha_1) \{C(\alpha_1) - B^2(\alpha_1)A^{-1}(\alpha_1)\}^{(n+2a+1)/2}}, \alpha_1 \in \mathbb{R}$$

donde

$$A(\alpha_1) = \underline{x}_*^* D^{-1}(\alpha_1) \underline{x}_* + \xi$$

$$B(\alpha_1) = \underline{x}_n^* Q D^{-1}(\alpha_1) \underline{x}_* + \xi \mu$$

y

$$C(\alpha_1) = \underline{x}_n^* Q D^{-1}(\alpha_1) Q' \underline{x}_n + 2b + \xi \mu^2$$

Además

$$p(\beta_1 | \alpha_1, \underline{x}_n) \propto \left\{ \left[\beta_1 - A^{-1}(\alpha_1) B(\alpha_1) \right]^2 A(\alpha_1) + C(\alpha_1) - B^2(\alpha_1) A^{-1}(\alpha_1) \right\}^{-(n+2a+1)/2}, \beta_1 \in \mathbb{R}$$

por lo que la distribución condicional de β_1 dado α_1 es una t-Student con $n+2a$ grados de libertad, localización

$A^{-1}(\alpha_1) B(\alpha_1)$ y precisión $(n+2a) A(\alpha_1) \{C(\alpha_1) - B^2(\alpha_1) A^{-1}(\alpha_1)\}^{-1}$.

Para el análisis de $p(\alpha_1 | \underline{x}_n)$ y la obtención de las demás distribuciones marginales, se requiere de métodos numéricos.

Haciendo uso de la misma matriz $E = Q_* D_*^{-1}(\alpha_1) Q_*'$ que se definió en el caso del proceso MA y de la partición que se hizo de ella, es posible obtener la distribución conjunta condicional de $(\underline{x}_n, x_{n+1})$ dado α_1 .

$$p(\underline{x}_n, x_{n+1} | \alpha_1) \propto \frac{d_{n+1}^{-1/2}(\alpha_1) p(\alpha_1)}{\left\{ \left[x_{n+1} - A^{-1}(\alpha_1) B(\alpha_1) \right]^2 A(\alpha_1) + C(\alpha_1) - B^2(\alpha_1) A^{-1}(\alpha_1) \right\}^{(n+2a+1)/2}}$$

$$\underline{x}_n \in \mathbb{R}^n, \quad x_{n+1} \in \mathbb{R}$$

en donde

$$A(\alpha_1) = E_{22}(\alpha_1) - \left[E_{21}(\alpha_1) \underline{x}_* + E_{22}(\alpha_1) x_n \right]^2 \left[\xi + \underline{x}'^{(2)} E(\alpha_1) \underline{x}^{(2)} \right]$$

$$B(\alpha_1) = \left[\underline{x}_n' E_{11}(\alpha_1) \underline{x}_* + \underline{x}_n' E_{12}(\alpha_1) x_n + \xi \mu \right] x$$

$$x \left[E_{21}(\alpha_1) \underline{x}_* + E_{22}(\alpha_1) x_n \right] \left[\xi + \underline{x}'^{(2)} E(\alpha_1) \underline{x}^{(2)} \right]^{-1} - \underline{x}_n' E_{12}(\alpha_1)$$

$$C(\alpha_1) = \xi \mu^2 + 2b + \underline{x}_n' E_{11}(\alpha_1) \underline{x}_n -$$

$$- \left[\underline{x}_n' E_{11}(\alpha_1) \underline{x}_* + \underline{x}_n' E_{12}(\alpha_1) x_n + \xi \mu \right]^2 \left[\xi + \underline{x}'^{(2)} E(\alpha_1) \underline{x}^{(2)} \right]^{-1}$$

$$\text{con } \underline{x}^{(2)} = \begin{pmatrix} \underline{x}_* \\ x_n \end{pmatrix}.$$

De la distribución anterior se puede deducir que la distribución condicional de x_{n+1} dado \underline{x}_n y α_1 , $p(x_{n+1} | \alpha_1, \underline{x}_n)$, es una distribución t-Student con $n+2$ grados de libertad, localización $A^{-1}(\alpha_1) B(\alpha_1)$ y precisión $(n+2a) A(\alpha_1) \left[C(\alpha_1) - B^2(\alpha_1) A^{-1}(\alpha_1) \right]^{-1}$.

Algunas características de la distribución predictiva $p(x_{n+1} | \underline{x}_n)$ pueden encontrarse a partir de $p(x_{n+1} | \alpha_1, \underline{x}_n)$ y $p(\alpha_1 | \underline{x}_n)$, por ejemplo,

$$E(X_{n+1} | \underline{x}_n) = E_{\alpha_1 | \underline{x}_n} E(X_{n+1} | \underline{x}_n, \alpha_1)$$

$$V(X_{n+1} | \underline{x}_n) = E_{\alpha_1 | \underline{x}_n} V(X_{n+1} | \underline{x}_n, \alpha_1) + V_{\alpha_1 | \underline{x}_n} E(X_{n+1} | \underline{x}_n, \alpha_1) .$$

Al igual que el proceso MA , el cálculo de $p(x_{n+1} | \underline{x}_n)$ necesita de procedimientos numéricos.

Se espera que la distribución inicial informativa dada en cada uno de los tres modelos presentados en esta sección, sea la suficientemente flexible como para representar, en la mayoría de los casos, el conocimiento inicial que se tenga sobre los valores parametrales.

CAPITULO 3
TOPICOS RELACIONADOS

En este capítulo se presentan algunas otras formas de representar el conocimiento inicial y su implicación en el análisis final. También se hace mención de otro tipo de modelos aplicables a series de tiempo y se plantea el problema de identificación, dando algunas posibles soluciones. Por último, se comenta brevemente el problema de estimación, predicción y prueba de hipótesis.

3.1. CONDICION DE ESTACIONARIEDAD

Teniendo a nuestra disposición una distribución de probabilidad asociada a los parámetros de cierto modelo, es posible también calcular la probabilidad de que la solución dé lugar a un proceso estacionario y/o invertible, así como también de que sea oscilatorio*. Por ejemplo, si se está trabajando con un modelo ARMA(1,1), la probabilidad de que dicho proceso sea estacionario e invertible es igual a

$$\int_R p(\alpha, \beta | \underline{x}_n) d\alpha d\beta ,$$

siendo R la región que define los valores parametrales que dan lugar a un proceso con tales características. En este

*Se refiere al caracter oscilatorio de la función de autocorrelación.

caso es fácil comprobar, que

$$R = \{(\alpha, \beta) \in \mathbb{R}^2 : |\alpha| < 1, |\beta| < 1\} .$$

En un proceso AR(2), la región de estacionariedad está dada por

$$R_E = \{(\beta_1, \beta_2) : \beta_1 + \beta_2 < 1, \beta_2 - \beta_1 < 1, |\beta_2| < 1\}$$

y la región que dá lugar a una función de autocorrelación oscilatoria es

$$R_O = \{(\beta_1, \beta_2) : \beta_1^2 + 4\beta_2 < 0\} .$$

El volúmen determinado por cada región R_E , R_O y la superficie $p(\beta_1, \beta_2 | \underline{x}_n)$ será igual a la probabilidad de que se tenga una solución estacionaria o con características oscilatorias respectivamente.

En la figura 3.1 se presenta las regiones R_E , R_O y sus complementos respectivos R_{NE} y R_{NO} . Además se puede apreciar la región (elíptica) de alta densidad que con probabilidad .99 contiene a los parámetros (β_1, β_2) del ejemplo dado en 2.2. Se observa que la probabilidad de que se tenga un proceso estacionario con una función de autocorrelación que oscila es practicamente 1, lo cuál concuerdo con lo que se obtendría al estudiar el modelo teórico.

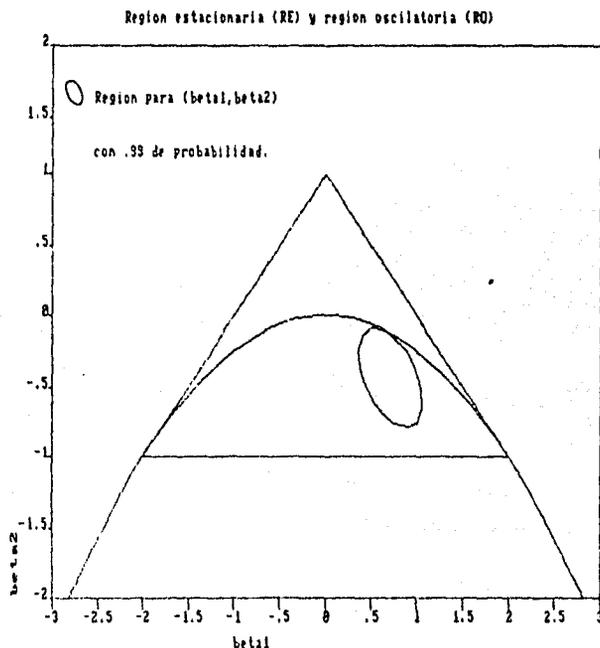


Figura 3.1.

Alternativamente la probabilidad de obtener una solución oscilatoria, $P\{(\beta_1, \beta_2) \in R_O\}$, se podría obtener si se cuenta con la distribución de $\beta_1^2 + 4\beta_2$. Zellner (1971), propone la transformación

$$h_1 = \beta_1 \quad , \quad h_2 = \beta_1^2 + 4\beta_2 \quad ,$$

de donde

$$p(h_1, h_2 | \underline{x}_n) \propto \left[v s^2 + (h_1 - \hat{\beta}_1)^2 d_{11} + \left(\frac{h_2 - h_1^2}{4} - \hat{\beta}_2 \right)^2 d_{22} + 2(h_1 - \hat{\beta}_1) \left(\frac{h_2 - h_1^2}{4} - \hat{\beta}_2 \right) d_{12} \right]^{-n/2} \quad , \quad (3.1.1.)$$

siendo d_{11} , d_{22} y d_{12} elementos de la matriz D dada en (2.2.10.) y $v s^2$, $\hat{\beta}$ cantidades definidas en la misma

sección 2.2.

Integrando numéricamente (3.1.1.) con respecto a h_1 se obtiene la distribución marginal (después de normalizar) de $h_2 = \beta_1^2 + 4 \beta_2$, $p(h_2 | \underline{x}_n)$, lo cual permitiría realizar inferencias con respecto a $\beta_1^2 + 4 \beta_2$, por ejemplo

$$P\{\text{solución oscilatoria}\} = P\{(\beta_1, \beta_2) \in R_0\} = P\{h_2 < 0\} .$$

3.2. ANALISIS USANDO OTRAS DISTRIBUCIONES INICIALES

La distribución inicial que se ha utilizado en este trabajo es, en el caso de una distribución no informativa, la propuesta por Jeffreys (1961), y en el caso de una distribución informativa, la basada en familias conjugadas, cuya teoría puede verse en Raiffa y Schlaifer (1961), De Groot (1970). Existen otros procedimientos que permiten asignar distribuciones iniciales de referencia o no informativas, como pueden ser el propuesto por Bernardo (1979), que está basado en el concepto de la información que se puede esperar de la realización de un experimento. Zellner (1977) utiliza la distribución inicial de "información mínima", haciendo uso de otra medida de información. Box y Tiao (1973), parten del hecho de que basta pedir a la distribución inicial "uniformidad local" en donde la verosimilitud tenga más importancia. Otro método es el de Novick y Hall (1965), Novick (1969), el cual hace uso de familias conjugadas.

En cualquiera de los procedimientos antes enunciados, así como en el de Jeffreys, la distribución inicial no informativa resulta impropia, no obstante, al combinarla con la verosimilitud por medio del teorema de Bayes, se obtienen distribuciones finales propias consideradas como informativas.

Broemeling (1985) hace mención de un procedimiento que

él considera muy prometedo y cuya referencia es Stigler (1982). La ventaja de este procedimiento, según Broemeling es que "se está en mejor disposición de pensar acerca de una variable aleatoria observable que hacerlo (directamente) sobre un parámetro θ no observable", y este método se basa precisamente en el uso de la distribución predictiva.

A pesar de las diferencias matemáticas entre los métodos anteriormente referidos, prácticamente los resultados finales son los mismos, más aún en el análisis de series de tiempo, en donde por lo general se cuenta con un número grande de observaciones.

Es importante hacer notar que lo desarrollado en el capítulo anterior sobre el modelo AR(p), no restringió los valores parametrales de tal forma que se impusiera alguna condición de estacionariedad. De hecho, se puede calcular la probabilidad de que la solución sea estacionaria o no, como se hizo en la sección 3.2. Si se sabe que el proceso es estacionario, entonces las raíces del polinomio $A(Z)$ definido en la sección 1.5 deben estar fuera del círculo unitario. Esto se traduce, en el caso de un proceso AR(1),

$$X_t = \beta_0 + \beta_1 X_{t-1} + \epsilon_t, \text{ en pedir que } |\beta_1| < 1, \text{ y en el}$$

caso de un AR(2), en que $\underline{\beta} = (\beta_1, \beta_2)$ esté en la región R_E definida en la sección anterior. Zellner (1971) presenta el desarrollo, siguiendo la regla de Jeffreys, para la obtención

de la distribución inicial no informativa de los parámetros del proceso AR(1), cuando se sabe que existe estacionariedad. La distribución que obtiene es

$$p(\beta, \theta, \sigma) \propto \frac{1-\beta}{(1-\beta^2)^{1/2}} \frac{1}{\sigma^2}, \quad |\beta| < 1,$$

en donde se usó la reparametrización

$$\beta = \beta_1, \quad \theta = \frac{\beta_0}{1-\beta_1}, \quad \sigma.$$

(Para (β, θ, σ) resulta que θ es independiente de β y de σ).

En términos de los parámetros originales, y haciendo uso de la propiedad de invarianza la cual fué brevemente explicada en la sección 2.1, se tiene que

$$p(\beta_0, \beta_1, \sigma) \propto |J|_+ p(\beta, \theta, \sigma),$$

siendo J el jacobiano de la transformación, de donde

$$p(\beta_0, \beta_1, \sigma) \propto \frac{1}{(1-\beta_1^2)^{1/2}} \frac{1}{\sigma^2}, \quad |\beta_1| < 1.$$

Suponiendo independencia entre σ , β_0 y β_1 se habría obtenido,

$$p(\beta_0, \beta_1, \sigma) \propto \frac{1}{(1-\beta_1^2)^{1/2}} \frac{1}{\sigma}, \quad |\beta_1| < 1.$$

Combinando esta última distribución inicial con la verosimilitud, Zellner obtiene la distribución final conjunta de $(\beta_0, \beta_1, \sigma)$ y la distribución final marginal de β_1 . Estas distribuciones no tienen alguna forma conocida y su análisis requiere métodos numéricos. Lahiff (1980) desarrolla un algoritmo que trata este problema. No obstante, las aproximaciones que se pueden realizar para n grande dan lugar a una distribución t-student para $p(\beta_1 | \underline{x}_n)$. Monohan (1981), ha estudiado en forma similar los modelos ARMA(p,q), para $p, q = 0, 1, 2$.

En el caso de querer establecer la condición de estacionariedad en una distribución inicial informativa, Broemeling (1985) propone el uso de familias conjugadas, en donde la asignación de los parámetros iniciales se realiza de tal forma que la distribución marginal de $\underline{\beta}$ asigna a la región de estacionariedad, R_E , la mayor densidad posible, es decir $p(\underline{\beta} \in R_E) \approx 1$. Refiriéndonos al modelo AR de la sección 2.6, el criterio anterior significa dar valores a $\underline{\mu}$, p , a y b , para los cuales la distribución marginal de $\underline{\beta}$, que es una t-student, se concentra lo más posible en la región R_E . Se debe hacer notar que esta elección de los parámetros iniciales también afecta a la distribución marginal de

$$\tau = \frac{1}{\sigma^2} .$$

3.3. Representaciones alternativas

El análisis bayesiano de series de tiempo se puede llevar a cabo utilizando modelos lineales dinámicos (MLD), los cuales permiten incluir cierto comportamiento de los parámetros a lo largo del tiempo. La formulación general de este tipo de modelos (caso discreto) es

Ecuación de observación: $\underline{X}(t) = F(t)\underline{\theta}(t) + \underline{v}(t)$, $\underline{v}(t) \sim N(\underline{0}, V(t))$

Ecuación del sistema: $\underline{\theta}(t) = G\underline{\theta}(t-1) + \underline{w}(t)$, $\underline{w}(t) \sim N(\underline{0}, W(t))$

En donde,

- $\underline{X}(t)$: vector $n \times 1$ de observaciones hechas al tiempo t .
- $\underline{\theta}(t)$: vector $r \times 1$ de parámetros del sistema al tiempo t .
- $F(t)$: matriz $n \times r$ de variables independientes, conocida al tiempo t .
- G : matriz $r \times r$ conocida que gobierna el movimiento del sistema.
- $\underline{v}(t)$: vector $n \times 1$ de errores en las observaciones al tiempo t .
- $\underline{w}(t)$: vector $r \times 1$ de errores en el sistema al tiempo t .

El modelo AR(p) analizado en el capítulo anterior se puede expresar en términos del MLD considerando

$$\underline{X}(t) = X_t \quad , \quad F(t) = (X_{t-1}, X_{t-2}, \dots, X_{t-p})$$

$$\underline{\theta}'(t) = (\beta_1, \beta_2, \dots, \beta_p) \quad , \quad \underline{y}(t) = z_t$$

$$G = I \quad , \quad \underline{w}(t) = \underline{0}$$

El análisis de los MLD se realiza a través del conocido Filtro de Kalman (que es por naturaleza bayesiano), en donde el vector $\underline{\theta}(t)$ es desconocido. Si además se desconocen, como es usual, las matrices de covarianzas $V(t)$ y $W(t)$, se utiliza filtraje adaptativo. En Maybeck (1979), Harrison and Stevens (1976) y Broemeling (1985) se puede encontrar un estudio de este tipo de modelos. La generalidad que representa un MLD (modelos econométricos, de regresión, de diseño de experimentos, de crecimiento, y otros más son casos particulares), es lo que ocasiona algunos problemas prácticos, como el tener distribuciones finales poly-t. Harrison y Ameen (1983) presentan un artículo en donde tratan de subsanar estas desventajas prácticas.

El estudio de series de tiempo también puede realizarse en el dominio de frecuencias lo cual significa, en términos generales, suponer que la variabilidad en una serie de tiempo está conformada por la variabilidad atribuída a cada uno de una serie de componentes ortogonales, senos y cosenos, en diferentes frecuencias.

Específicamente, si $\{X_t; t = 0, \pm 1, \pm 2, \dots\}$ es un proceso estocástico estacionario y σ es su función de covarianzas, existe una única función F real, no decreciente, continua por la derecha y definida en $[-\pi, \pi]$, tal que

$$\sigma(j) = \int_{-\pi}^{\pi} e^{i\omega j} dF(\omega), \quad j = 0, \pm 1, \pm 2, \dots$$

A la función F se le conoce como función de distribución espectral. Suponiendo que F es absolutamente continua con respecto a la medida de Lebesgue en $[-\pi, \pi]$, se puede encontrar una única (módulo conjuntos de medida cero) función f real, simétrica alrededor del cero, no negativa y definida en $[-\pi, \pi]$ tal que

$$\sigma(j) = \int_{-\pi}^{\pi} e^{i\omega j} f(\omega) d\omega, \quad j = 0, \pm 1, \pm 2, \dots$$

La función f es la función de densidad espectral.

Además, si se tiene una función con las características de f , se puede asegurar que ésta es la función de densidad espectral para alguna función de covarianza. Todos estos resultados y más se pueden encontrar en Anderson (1971), Loeve (1963).

El análisis espectral de una serie de tiempo se centra generalmente en estimar la función de densidad espectral, y en particular ciertas integrales de dicha densidad sobre subconjun-

tos de $[-\pi, \pi]$, (comunmente se introduce una constante de normalización de tal forma que

$$\int_{-\pi}^{\pi} f(\omega) d\omega = 1 ,$$

teniendo efectivamente una densidad), así como también las ordenadas de $f(\omega)$ para algún conjunto finito

$\{\omega_1, \omega_2, \dots, \omega_k\} \subset [-\pi, \pi]$. Shaman (1975) y Shore (1977) presentan un procedimiento bayesiano para estimar ciertas ordenadas $f(\omega_1), f(\omega_2), \dots, f(\omega_k)$, con $0 < \omega_1 < \omega_2 < \dots < \omega_k < \pi$. Ambos consideran a $\{f(\omega_j): j = 1, 2, \dots, k\}$ como un conjunto de parámetros, siendo el espacio parametral

$\{(f(\omega_1), f(\omega_2), \dots, f(\omega_k)): f(\omega_j) \geq 0, j = 1, 2, \dots, k\}$. La distribución inicial usada es conjugada. Dos de los principales problemas en este análisis son: tener una buena aproximación a la verosimilitud que sea de fácil manejo y considerar que un número finito de ordenadas no caracteriza en forma única a la función de densidad espectral.

3.4. IDENTIFICACION DE MODELOS

En términos generales el problema de identificación tratado bayesianamente lo presenta Zellner (1971) y éste consiste en lo siguiente (usando notación original): sean M_1 , M_2 dos modelos y θ_1 , θ_2 sus parámetros respectivos. Si ambos modelos junto con la información inicial asociada a cada uno, $p(\theta_i | M_i)$, $i = 1, 2$, dan lugar a la misma distribución marginal para un conjunto de observaciones, se dirá que M_1 con su información inicial y M_2 con su información inicial son observacionalmente equivalentes, es decir, M_1 no está identificado en relación a M_2 . Explícitamente, sea

$$p(y | M_i) = \int p(y | \theta_i, M_i) p(\theta_i | M_i) d\theta_i, \quad i = 1, 2, .$$

Si $p(y | M_1) = p(y | M_2)$, entonces los modelos con sus respectivas informaciones iniciales $p(\theta_i | M_i)$, $i = 1, 2$, son observacionalmente equivalentes, y no se puede "decidir" qué modelo explica la información muestral.

El incorporar información sobre las posibilidades que tiene cada modelo de ser el que mejor explique el proceso de interés, puede ayudar a resolver el problema de identificación. Esto significa el dar una probabilidad inicial, $p(M)$, para cada modelo M , que se desee considerar.

Harrison y Stevens (1976) proponen un método de identificación en donde se considera la clase de modelos lineales dinámicos, y proponen un criterio bayesiano para elegir, de un conjunto finito de tales modelos, aquél que sea el óptimo. En dicho artículo no se dice nada sobre la asignación de las probabilidades iniciales $p(M_i)$, $i = 1, 2, \dots, m$. Se hace uso de una función de utilidad basada a su vez en la utilidad que se tiene al realizar una predicción L pasos adelante. Las predicciones se realizan sobre las mismas observaciones que se tienen.

Procedimientos de identificación, pensados exclusivamente para series de tiempo, han sido desarrollados desde el punto de vista bayesiano por Akaike (1979), Díaz y Farah (1981) y Poskitt (1986). Este último autor trata el problema de identificación en modelos ARMA, mientras que los otros sólo trabajan con modelos AR.

Akaike basa su procedimiento de identificación en el uso de la estadística AIC (Akaike Information Criterion), la cuál surge del principio de minimización de entropía propuesto por él mismo en Akaike (1974), Akaike (1977). Específicamente se parte del hecho de que $\exp\{-\frac{1}{2} AIC(k)\}$ es aproximadamente igual a la verosimilitud $L(k)$ de un modelo AR(k), $k = 0, 1, 2, \dots, K$. Si la distribución inicial propuesta para el orden del modelo es $p(k)$, $k = 0, 1, 2, \dots, K$, la distri-

bución final correspondiente será entonces

$$p(k|\underline{x}_n) = \frac{L(k) p(k)}{\sum_{i=0}^k L(i)p(i)} , \quad k = 0, 1, \dots, K .$$

Akaike propone varias distribuciones iniciales $p(k)$, entre ellas una que permite considerar orden infinito, situación que sólo puede tener un interés teórico. Asimismo presenta un método de estimación de parámetros promediando sobre los modelos, usando pesos correspondientes según $p(k|\underline{x}_n)$.

Por otra parte, Díaz y Farah (1981) desarrollan un método de identificación para modelos AR en donde la distribución final $p(k|\underline{x}_n)$, $k = 1, 2, \dots, K$, para el orden del modelo, se obtiene a partir de la distribución final conjunta $p(\underline{\beta}^{(k)}, \sigma, k|\underline{x}_n)$ en donde $\underline{\beta}^{(k)}$ es el vector de parámetros del modelo AR(k), $k = 1, 2, \dots, K$ y σ es la desviación estandar del proceso.

Los supuestos que se hacen son que $p(\underline{\beta}^{(k)}|\sigma, k)$ sigue una distribución normal, σ y k son independientes, siendo $p(\sigma)$ una distribución gamma invertida y $p(k)$ uniforme en $\{1, 2, \dots, k\}$.

Habiendo encontrado $p(k|\underline{x}_n)$, $k = 1, 2, \dots, k$, la cuál está dada en forma exacta, se identifica el orden del modelo,

y las inferencias sobre $\underline{\beta}^{(k^*)}$ se pueden realizar a partir de $p(\underline{\beta}^{(k^*)} | k^*, \underline{x}_n)$, donde k^* es el valor identificado de k . Esta última distribución, según lo visto en el capítulo 2, es una distribución t-multivariada.

Se debe hacer notar que el anterior método de identificación implica el dar un vector de medias y una matriz de varianzas-covarianzas para cada distribución inicial sobre $\underline{\beta}^{(k)}$ dado σ , $k = 1, 2, \dots, K$, además se tienen que especificar los parámetros de la distribución gamma invertida para $p(\sigma)$. Broemeling (1985) cita un trabajo de Litterman (1980), en donde se propone una forma de asignar parámetros a las distribuciones anteriores.

Por último, el mismo Broemeling señala la posibilidad de aplicar el procedimiento de Díaz y Farah para la identificación de modelos MA ó ARMA, aclarando que se esperan encontrar problemas analíticos.

En el trabajo que presenta Poskitt (1986), el problema de identificación es resuelto dando un criterio que asintóticamente coincide con el criterio bayesiano de elegir la decisión con mayor utilidad esperada.

El desarrollo general del procedimiento se basa en el uso de la función de densidad espectral del proceso considera

do. Sean m los modelos considerados, cada uno dado por

$$M_i = \{f_i(\underline{\theta}_i, \omega) \in L^2, \quad \underline{\theta}_i \in \Theta_i\}, \quad i = 1, 2, \dots, m,$$

en donde f_i , $i = 1, 2, \dots, m$ son las funciones espectrales que definen cada modelo. Omitiendo los subíndices, si $m(\underline{\theta})$ es la decisión de elegir el modelo M , el criterio bayesiano de decisión escoge aquella decisión que maximice la utilidad esperada $U^*(m(\underline{\theta}))$, dada por

$$U^*\{m(\underline{\theta})\} = \int_{\Theta} U\{m(\underline{\theta})\} p(M, \underline{\theta} | x_n) d\underline{\theta}. \quad (3.4.1.)$$

La función de utilidad propuesta es

$$U\{m(\underline{\theta})\} = \exp\{-\eta\{m(\underline{\theta})\}\},$$

en donde

$$\eta\{m(\underline{\theta})\} = \int_{-\pi}^{\pi} \left[\frac{f^*(\omega)}{f(\underline{\theta}, \omega)} - 1 \right]^2 d\omega,$$

con $f^*(\omega)$ la verdadera función espectral del proceso que generó la información muestral x_n . El desconocimiento de f^* es resuelto dando una aproximación de dicha función por medio del periodograma ó función de densidad espectral muestral (en cualquiera de las referencias citadas en el capítulo puede verse la definición de tal función).

La distribución inicial de $(\underline{\theta}, M)$ esta dada por

$$p(\underline{\theta}, M) = p(\underline{\theta}|M) p(M) ,$$

en donde $p(\underline{\theta}|M)$ es la dada por la regla de Jeffreys y $p(M) \propto (2\pi)^{-d/2}$, con d la dimensión del espacio parametral correspondiente.

Para la clase de modelos ARMA, el criterio basado en (3.4.1.), resulta ser asintóticamente equivalente a elegir el modelo ARMA(p,q) que minimice la función $\Delta(p,q)$, dada por

$$\Delta(p,q) = \frac{n}{2} \left[\ln \hat{\sigma}^2 + (2n-1)^{-1} \sum_j \ln |\hat{K}(e^{-i\omega_j})|^2 \right] + \frac{1}{2} (p+q) \ln n$$

$$+ \frac{\pi}{2n-1} \sum_j \left[\frac{\hat{f}^*(\omega_j)}{\hat{\sigma}^2 |\hat{K}(e^{-i\omega_j})|^2} \right]^2 ,$$

en donde, siguiendo la notación de la sección 1.7 en el capítulo 1, $K(Z) = A^{-1}(Z)M(Z)$, $\sigma^2 = \sigma_Z^2 / \sigma_x^2$ y los estimadores se obtienen sustituyendo estimadores consistentes de los parámetros.

El análisis que realiza Poskitt de cada uno de los sumandos en $\Delta(p,q)$, permite decir que, en el contexto de los modelos ARMA, el criterio de selección basado en $\Delta(p,q)$, establece un compromiso entre las características predictivas, la estructura de autocorrelación de residuales y la dimensionalidad $d = p+q$ del modelo.

3.5. Decisiones estadísticas.

El realizar estimaciones y prueba de hipótesis sobre los parámetros de un modelo, así como predecir el valor futuro de una variable de interés, se consideran problemas de decisión en ambiente de incertidumbre.

Tal como se explicó en la primera sección del capítulo 2, el criterio bayesiano de decisión escoge aquella decisión que maximiza la utilidad esperada. Para el cálculo de la utilidad que se espera de cada decisión es necesario contar con una función de utilidad, por medio de la cual se cuantifica la consecuencia de tomar una decisión.

En el presente trabajo sólo se trató el aspecto inferencial, es decir, la obtención de la distribución final sobre los parámetros y de la distribución predictiva (final) sobre la variable de interés. Estas distribuciones finales son necesarias para el cálculo de la utilidad esperada.

La asignación de una función de utilidad que refleje, como se mencionó anteriormente, las consecuencias de tomar cada una de las posibles decisiones, no es un problema sencillo (lo mismo puede decirse sobre la asignación de una distribución inicial informativa sobre los parámetros, la cuál debe reflejar el conocimiento que se tiene sobre el verdadero valor de éstos).

La mayoría de las veces se han utilizado funciones de utilidad cuadráticas o lineales, las cuales son de fácil manejo, y que además permiten recuperar, si se utilizan distribuciones iniciales de referencia, algunos resultados que numéricamente coinciden con el procedimiento clásico.

Zellner (1985), analiza una clase de funciones de pérdida (en cuyo caso el criterio de decisión es minimizar la pérdida esperada) asimétricas, para el problema de estimación y de predicción. Lindley (1976) introduce el concepto de funciones de utilidad conjugadas.

Prueba de hipótesis, considerado como un problema de decisión, ha sido tratado por Bernardo (1982).

Más información sobre la teoría bayesiana de toma de decisiones se puede encontrar en Savage (1972), Raiffa y Schlaifer (1961), De Groot (1970), Berger (1980) y Aitchison y Dunsmore (1975).

CAPITULO 4

CONCLUSION

En el uso de los modelos ARMA para el estudio de series de tiempo, uno de los principales problemas es el manejo de la matriz de autocovarianzas, pues cada entrada de ésta es una función no lineal de los parámetros. Si a esto añadimos el hecho de que la función $S(\underline{\beta}, \underline{\alpha})$ definida en (2.3.11.), no es exactamente cuadrática en $(\underline{\beta}, \underline{\alpha})$, es fácil comprender el porqué del uso de aproximaciones a todo lo largo de este trabajo (lo mismo sucede al utilizar métodos clásicos de inferencial), y el porqué se presentaron, en la sección 2.6., algunos resultados en los modelos MA(q) y ARMA(p,q) solamente para $p = q = 1$. No obstante, es común que en series de tiempo se cuente con un número suficientemente grande de observaciones, con lo cuál se espera que las aproximaciones no sean malas.

En el análisis bayesiano de una serie de tiempo no es necesario el imponer restricciones de estacionariedad y/o invertibilidad, lo cuál es una ventaja. Sin embargo, se tiene la flexibilidad necesaria para imponer este tipo de restricciones, las cuales se incluirían como parte de la definición de la distribución inicial. Aún más, es posible calcular la probabilidad de que un proceso sea estacionario y/o invertible, o que presente alguna otra característica que dependa de los valores parametrales. Un ejemplo de esto fué dado en la

sección 3.1.

Generalmente el objetivo del análisis de una serie de tiempo es el de predecir valores futuros. La forma de la distribución predictiva para el valor un paso adelante sólo es reconocible para el proceso AR(p). Un problema abierto, de acuerdo a lo que dice Broemeling (1985), es caracterizar completamente la distribución predictiva para k observaciones futuras, la cual fué brevemente descrita en la sección 2.2 para un proceso AR(p).

De los métodos de identificación presentados, el de Poskitt (1986) es el más completo, pues aparte de poder utilizarlo en modelos ARMA, el orden es elegido de acuerdo al criterio bayesiano de decisión. Los otros métodos sólo calculan la distribución final del orden del modelo, y ésto solamente para modelos AR. Sería conveniente realizar un análisis que permitiera establecer comparaciones sobre la efectividad de cada uno de éstos métodos de identificación.

Las funciones de autocovarianza (γ_k) y de autocorrelación (ρ_k) para cualquiera de los procesos analizados en este trabajo, están determinados por los parámetros del modelo utilizado. Al tener como resultado del proceso inferencial una distribución final sobre los posibles valores parametrales es posible, en principio, obtener una distribución final

para γ_k y/o ρ_k , $k \in \mathbb{Z}$. La dificultad que surge a primera vista, es la dependencia no lineal sobre los parámetros que presentan tanto γ_k como ρ_k .

Mucho queda por hacer en el análisis bayesiano de series de tiempo (existen series de tiempo multivariadas). En este trabajo se intentó dar una visión general de lo que sería dicho tipo de análisis, siendo el problema de inferencia el tema más ampliamente desarrollado.

APENDICE A

A.1. Distribución normal univariada.

Una variable aleatoria $Y \in \mathbb{R}$ tiene una distribución normal sí y sólo si, su función de densidad de probabilidad está dada por

$$p(y|\theta, 1/\tau) = (2\pi)^{-1/2} \tau^{1/2} \exp\left\{-\frac{\tau}{2} (y - \theta)^2\right\}, \quad y \in \mathbb{R}.$$

Los parámetros $\theta \in \mathbb{R}$ y $\tau > 0$ son la media y la precisión respectivamente. La varianza es $1/\tau$.

Notación: $Y \sim N(\theta, 1/\tau)$.

A.2. Distribución de t de Student univariada.

Una variable aleatoria $Y \in \mathbb{R}$ tiene una distribución t de Student sí y sólo si, su función de densidad de probabilidad es de la forma

$$p(y|\theta, h, v) = \frac{\Gamma((v+1)/2)}{\Gamma(1/2)\Gamma(v/2)} \left(\frac{h}{v}\right)^{1/2} \left[1 + \frac{h}{v} (y-\theta)^2\right]^{-(v+1)/2}$$

$y \in \mathbb{R}.$

Los parámetros son $\theta \in \mathbb{R}$, $h > 0$ y $v > 0$. La media y la precisión son θ y h respectivamente. Al parámetro v

se le conoce como los grados de libertad. La media existe si $v > 1$ y la varianza existe si $v > 2$. Esta última es igual a $\frac{v}{v-2} h^{-1}$.

Notación: $Y \sim t(\theta, h, v)$.

A.3. Distribuciones gamma y χ^2 .

Una variable aleatoria $Y \in \mathbb{R}$ tiene una distribución gamma con parámetros $a > 0$ y $b > 0$ sí y sólo si, su función de densidad de probabilidad está dada por

$$p(y|a,b) = \frac{b^a}{\Gamma(a)} y^{a-1} e^{-by}, \quad y > 0.$$

La media y la varianza son, respectivamente a/b y a/b^2 .

La distribución χ^2 se obtiene de la distribución anterior si $a = v/2$ y $b = 1/2$, obteniéndose la siguiente función de densidad de probabilidad

$$p(y|v) = \frac{1}{2^{v/2} \Gamma(v/2)} y^{v/2-1} e^{-y/2}, \quad y > 0.$$

El único parámetro en esta última distribución es v , que son los grados de libertad.

Notación: $Y \sim \text{Gamma}(a,b)$, $Y \sim \chi^2(v)$.

A.4. Distribución gamma invertida.

Si X es una variable aleatoria tal que $X \sim \text{gamma}(a, b)$, entonces $Y = |X^{-1/2}|$ es una variable aleatoria con distribución gamma invertida, y cuya función de densidad de probabilidad está dada por

$$p(y|a, b) = \frac{2b^a}{\Gamma(a)y^{2a+1}} e^{-b/y^2}, \quad y > 0.$$

En particular, si $a = v/2$ y $b = vs^2/2$, se tiene que

$$p(y|v, s) = \frac{2}{\Gamma(v/2)} \left(\frac{vs}{2}\right)^{v/2} \frac{1}{y^{v+1}} \exp\left\{\frac{-vs^2}{2y^2}\right\}, \quad y > 0.$$

En términos de ésta última densidad se tiene que

$$\text{Moda} = S \left(\frac{v}{v+1}\right)^{1/2}$$

$$E(Y|v, s) = \frac{\Gamma((v-1)/2)}{\Gamma(v/2)} \left(\frac{v}{2}\right)^{1/2} S, \quad v > 1$$

$$V(Y|v, s) = \frac{v}{v-2} S^2 - E^2(Y|v, s), \quad v > 2.$$

Notación: $Y \sim \text{Gamma I}(v/2, vs^2/2)$.

A.5. Distribución F (Fisher-Snedecor).

La variable aleatoria Y tiene una distribución F con

$v_1 > 0$ y $v_2 > 0$ grados de libertad sí y sólo si, la función de densidad de probabilidad es de la forma

$$P(y|v_1, v_2) = \frac{1}{B\left(\frac{v_1}{2}, \frac{v_2}{2}\right)} \left(\frac{v_1}{v_2}\right)^{v_1/2} \frac{y^{\frac{v_1}{2}-1}}{\left[1 + (v_1/v_2)y\right]^{(v_1+v_2)/2}}, \quad y > 0$$

donde $B(\omega_1, \omega_2) = \frac{\Gamma(\omega_1)\Gamma(\omega_2)}{\Gamma(\omega_1+\omega_2)}$.

Notación: $Y \sim F_{(v_1, v_2)}$.

A.6. Distribución normal multivariada.

La variable aleatoria $\underline{Y} \in \mathbb{R}^m$ tiene una distribución normal m -variada con parámetros $\underline{\theta} \in \mathbb{R}^m$ y $\Sigma_{m \times m}$, sí y sólo si, la función de densidad correspondiente es

$$P(\underline{Y}|\underline{\theta}, \Sigma) = \frac{1}{(2\pi)^{m/2} |\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\underline{Y}-\underline{\theta})' \Sigma^{-1} (\underline{Y}-\underline{\theta})\right\}, \quad \underline{Y} \in \mathbb{R}^m.$$

El vector $\underline{\theta}$ es la media y la matriz $\Sigma_{m \times m}$, que es simétrica y definida positiva, es la matriz de covarianzas. ($T = \Sigma^{-1}$ es la matriz de precisión.

Notación: $Y \sim N(\underline{\theta}, \Sigma)$.

A.7. Distribución t de Student multivariada.

La variable aleatoria $\underline{Y} \in \mathbb{R}^m$ tiene una distribución t de

Student m-variada con parámetros $v > 0$, $\underline{\theta} \in \mathbb{R}^m$ y $T_{m \times m}$ sí, y sólo si, la función de densidad correspondiente es

$$p(\underline{y} | \underline{\theta}, T, v) = \frac{\Gamma((v+m)/2) |T|^{1/2}}{(v\pi)^{m/2} \Gamma(v/2)} \left[1 + \frac{1}{v} (\underline{y} - \underline{\theta})' T (\underline{y} - \underline{\theta}) \right]^{-(m+v)/2},$$

$\underline{y} \in \mathbb{R}^m$.

La media es $\underline{\theta}$, la matriz de precisión es T y v son los grados de libertad. La matriz de covarianzas es $\frac{v}{v-2} T^{-1}$. La matriz T es simétrica y definida positiva.

Notación: $\underline{y} \sim t(\underline{\theta}, T, v)$.

A.7. Distribución poly-t.

Sean $p_j(\underline{y} | \underline{\theta}_j, T_j, v_j)$, $j = 1, 2, \dots, k$, funciones de densidad correspondiente a k distribuciones t de Student m -variadas. Entonces, la variable aleatoria $\underline{y} \in \mathbb{R}^m$ tiene una distribución poly- t sí, y sólo si, su función de densidad está dada por

$$p(\underline{y} | \underline{\theta}, T, v) \propto \prod_{j=1}^k p_j(\underline{y} | \underline{\theta}_j, T_j, v_j), \quad \underline{y} \in \mathbb{R}^m.$$

Los parámetros de esta distribución son

$$\underline{\theta}' = (\underline{\theta}'_1, \underline{\theta}'_2, \dots, \underline{\theta}'_k), \quad T = (T_1, T_2, \dots, T_k) \quad y$$

$$\underline{v}' = (v_1, v_2, \dots, v_k).$$

Se sabe muy poco acerca de las características de esta distribución para $k \geq 2$. En general es multimodal y asimétrica por lo que las aproximaciones normales que se han propuesto (por ejemplo Broemeling (1985), cap. 6) no resultan muy convincentes. La constante de normalización es desconocida.

Más información sobre éstas y otras distribuciones se puede encontrar en Zellner (1971) y Johnson y Kotz (1972).

APENDICE B

A continuación se presentan los datos utilizados en el ejemplo de la sección 2.2.

t	x_t								
1	1.2134	15	-1.3703	29	-0.9831	43	-1.4436	57	-1.0581
2	0.7306	16	-1.1521	30	0.5643	44	-0.7501	58	-0.5495
3	0.0047	17	0.1012	31	0.4553	45	1.5647	59	2.1837
4	-0.6077	18	0.5038	32	-0.5585	46	1.6647	60	1.7427
5	-0.7256	19	-0.3252	33	-0.8012	47	1.2739	61	0.7170
6	0.5013	20	-0.6444	34	-0.1381	48	0.4376	62	-0.0817
7	1.9442	21	-1.0141	35	0.4880	49	0.1100	63	-2.9210
8	0.3980	22	-0.6302	36	2.4647	50	-1.3420	64	-1.8409
9	0.7894	23	-0.5947	37	1.0511	51	-0.7713	65	-0.8355
10	1.9233	24	-0.8493	38	-0.6219	52	0.7760	66	-1.0947
11	0.3883	25	-0.0111	39	-0.3329	53	1.4720	67	-0.4288
12	0.3789	26	0.3456	40	0.1428	54	1.4803	68	-0.3307
13	0.1451	27	0.4591	41	1.0182	55	-0.0246	69	1.0528
14	-0.4891	28	-0.8977	42	0.5887	56	-0.0096	70	1.4121

Los valores iniciales para iniciar la simulación fueron $x_{-1} = -0.0729$ y $x_0 = 0.4403$.

Hay que aclarar que estos datos no se obtuvieron del trabajo original de Wold (1965), sino de Anderson (1971), en donde se presentan éste y otros conjuntos de datos para su análisis.

REFERENCIAS

- Aitchinson, J.S. y Dunsmore, J.R. (1975). Estatistical Prediction Analysis. Londres: Cambridge University Press.
- Akaike, H. (1974). A new look at the statistical model identification. I.E.E.E. Trans. Auto. Control AC-19, 716-723.
- Akaike, H. (1977). On entropy maximization principle. En Applications of Statistics (P.R. Krishnaiah, ed.). Amsterdam: North-Holland.
- Akaike, H. (1979). A Bayesian extension of the minimum AIC procedure of autorregresive model fitting. Biometrika, 66,2, 237-242.
- Anderson, T.W. (1971). The Statistical Analysis of Time Series. Nueva York: John Wiley and Sons.
- Berger, J.O. (1985). Statistical Decision Theory, Foundations, Concepts, Methods. Nueva York: Springer-Verlag.
- Bernardo, J.M. (1979 b). Reference posterior distributions for Bayesian inference. J. Roy. Statist. Soc., B, 41, 113-147 (con discusión).
- Bernardo, J.M. (1981). Bioestadística, una perspectiva bayesiana. Barcelona: Vicens-Vives.
- Bernardo, J.M. (1982). Contraste de modelos probabilísticos desde una perspectiva bayesiana. Trabajos de Estadística y de Investigación Operativa, 33, 16-30.

- Box, G.E.P. y Jenkins, G.M. (1970). Time Series Analysis, Forecasting and Control. San Francisco: Holden-Day.
- Box, G.E.P. y Tiao, G.C. (1973). Bayesian Inference in Statistical Analysis. Reading, Mass.: Addison-Wesley.
- Broemeling, L.D. (1985). Bayesian Analysis of Linear Models. Nueva York: Marcel Dekker.
- Chatfield, C. (1975). The Analysis of Time Series: Theory and Practice. Londres: Chapman and Hall.
- Cox, D.R. y Miller, H.D. (1965). The Theory of Stochastic Processes. Londres: Chapman and Hall.
- DeGroot, M.H. (1970). Optimal Statistical Decisions. Nueva York: McGraw-Hill.
- Díaz, J. y Farah, J.L. (1981). Bayesian identification of autoregressive processes. 22nd NBER-NSF Seminar on Bayesian Inference in Econometrics.
- Fuller, W.A. (1976). Introduction to Statistical Time Series. Nueva York: John Wiley and Sons.
- Gnedenko, B.V. (1969). The Theory of Probability. Moscú : MIR.
- Granger, C.W. y Newbold, P. (1977). Forecasting Economic Time Series. Nueva York: Academic Press.
- Harrison, P.J. y Ameen, J.R.M. (1983). Normal discount Bayesian models. 2nd Valencia International Meeting on Bayesian Statistics.

Harrison, P.J. y Stevens, C.F. (1976). Bayesian forecasting.
J. Roy. Statist. Cos., B, 38, 205-247 (con discusión).

Jeffreys, H. (1939/1961). The Theory of Probability. Oxford:
Clarendon Press.

Johnson, N.L. y Kotz, S. (1969/1972). Distributions in Statistics
(4 vols.). Nueva York : Wiley.

Kendall, M.G. y Stuart, A. (1966). The Advanced Theory of Statistics
tics, vol. 3. Londres: Griffin.

Kingman, J.F.C. y Taylor, S.J. (1966). Measure and Probability.
Cambridge: University Press.

Kolmogorov, A.N. (1950). Foundations of the Theory of Probability.
Nueva York: Chelsea Press.

Lahiff, M. (1980). Time series forecasting with informative prior
distributions. Rep. Tec. No. 11, Depto. de Estadística,
Universidad de Chicago.

Lindley, D.V. (1976). A class of utility functions. Ann. Statisc.,
4, 1-10.

Litterman, R.B. (1980). A Bayesian procedure for forecasting with
vector autoregression. Artículo no publicado, MIT, Cambridge,
Mass.

Loeve, M. (1955/1977). Probability Theory (2 vols.). Berlín:
Springer.

Maybeck, P.S. (1979). Stochastic Models, Estimation and Control
(vols. 1 y 2). Nueva York: Academic Press.

- Monahan, J.F. (1981). Computations for Bayesian time series analysis. 141st Annual Meeting of the American Statistical Association.
- Nelson, C.R. (1973). Applied Time Series Analysis for Managerial Forecasting. San Francisco: Holden-Day.
- Novick, M.R. (1969). Multiparameter Bayesian indifference procedures. J. Roy. Statist. Soc., B, 31, 29-64 (con discusión).
- Novick, M.R. y Hall, W.J. (1965). A Bayesian indifference procedure. J. Amer. Statist. Assoc., 60, 1104-1117.
- Poskitt, D.S. (1986). A Bayes procedure for the identification of univariate time series models. Ann. Statist., 14, 502-516.
- Raiffa, H. y Schlaifer, R. (1961). Applied Statistical Decision Theory. Cambridge, Mass.: The MIT Press.
- Savage, L.J. (1972). The Foundations of Statistics. Nueva York: Dorer.
- Shaman, P. (1975). Some Bayesian considerations in spectral estimation. Rep. tec. No. 102, Depto. Estadística, Carnegie-Mellon University.
- Shore, R.W. (1977). A Bayesian approach to the spectral analysis of stationary time series. En Savage Dissertation Award Theses.
- Stigler, S.M. (1982). Thomas Bayes and Bayesian inference. J. Roy. Statist. Soc., A, 145, 250-258.
- Wold, H.O.A. (1965). Bibliografía on Time Series and Stochastic Processes. Edimburgo: Oliver and Boyd.

Zellner, A. (1971). An Introduction to Bayesian Inference in Econometrics. Nueva York: John Wiley and Sons.

Zellner, A. (1977). Maximal data information prior distributions. En New Developments in the Applications of Bayesian Methods. (Aykac y Brumat eds.). Amsterdam: North-Holland.

Zellner, A. (1985). Bayesian estimation and prediction using asymmetric loss functions. Rep. Tec., H.G.B. Alexander Research Foundation, Graduate School of Business, U. de Chicago.