UNIVERSIDAD CENTROCCIDENTAL
"LISANDRO ALVARADO"

Decanato de Ciencias y Tecnología
Licenciatura en Ciencias Matemáticas

"Introducción a los Modelos Lineales Dinámicos: Modelo Polinomial de Primer Orden"

Trabajo Especial de Grado presentado por

Br. Yajaira R. Hernández A.

como requisito final
para obtener el título de Licenciado
en Ciencias Matemáticas
Área de Conocimiento: Probabilidad y Estadistica.
Tutor: MSc Luz Rodríguez

Barquisimeto, Venezuela. Junio de 2010

Universidad Centroccidental
"Lisandro Alvarado"
Decanato de Ciencias y Tecnología
Licenciatura en Ciencias Matemáticas

ACTA
TRABAJO ESPECIAL DE GRADO

Los suscritos miembros del Jurado designado por el Jefe del Departamento de Matemáticas del Decanato de Ciencias y Tecnología de la Universidad Centroccidental "Lisandro Alvarado", para examinar y dictar el veredicto sobre el Trabajo Especial de Grado titulado:

"Introducción a los Modelos Lineales Dinámicos: Modelo Polinomial de Primer Orden"

Presentado por la ciudadana Br. Yajaira R. Hernández A. titular de la Cédula de Identidad Nº 16.646.731. Con el propósito de cumplir con el requisito académico final para el otorgamiento del título de Licenciado en Ciencias Matemáticas.

Luego de realizada la Defensa y en los términos que imponen los Lineamientos para el Trabajo Especial de Grado de la Licenciatura en Ciencias Matemáticas, se procedió a discutirlo con el interesado habiéndose emitido el veredicto que a continuación se expresa:

[1] ____________________

Con una calificación de ____________________ puntos.

En fe de lo expuesto firmamos la presente Acta en la Ciudad de Barquisimeto a los _____ días del mes de ____________________ de __________.

TUTOR	FIRMA
PRINCIPAL	FIRMA
PRINCIPAL	FIRMA

OBSERVACIONES:

____

____

____

[1] Aprobado ó Reprobado

A mi madre por ser ejemplo digno de lucha y perseverancia

"Te Amo Dionisia"

Agradecimientos

Fueron diferentes los obstáculos he imprevistos que se presentaron, aún así pense que me detendrían, pero mi fuerza, voluntad y el apoyo incondicional de mi familia y amigos me motivaron a seguir adelante, es por todos ustedes que logre alcanzar mi gran meta, por eso agradezco:

A Dios, por permitirme ser parte de su creación he impulsarme a luchar por cada uno de mis sueños.

A Mis padres "Dionisia" por creer en mí y apoyarme en todo momento, eres ejemplo de lucha y perseverancia juntas lo hemos logrado, a ti "Secundino +" que donde quieras que te encuentres estas celebrando conmigo este gran logro.

A Mi Padrino "Francisco" por ser unos de los pilares y apoyo condicional a lo largo de mi carrera, eres mi segundo padre.

A Mis Hermanas "Nery, Nelly y Nancy" a ustedes por brindarme su amor, apoyo, comprensión y estar allí a mi lado motivándome para continuar adelante.

A Mis Sobrinos (José, Isbel, Ander, Crisbel, Javier, Neilimar, Yuletsi, Yunetsi) que este logro sea de estimulo para ustedes los quiero mucho.

A ti amor "Abigail" que en tan poco tiempo me has dado gran apoyo para seguir adelante, eres de gran importancia en mi vida, doy gracias a Dios por haberte colocado en mi camino. Te quiero cielo.

A mis amigos (Janet, Victor, Dairilus, Verónica, Mariela, Carolina, Eliecer, Elizabet, Elifer, Manuel, Mediomundo, Maxiel, Dicmar, Luigi, Jorge, Yuber, Gabriel, Yolimar, Anuar) por todo esos momentos inolvidables vividos dentro y fuera de la universidad.

A el "Grupo de Excursionismo Roraima" por dejarme formar parte de su familia, y por enseñarme muchas cosas buenas.

A las familias que me acogieron con amor y apoyo incondicional, me hicieron sentir parte de ustedes en esta hermosa tierra crepuscular. Flia. Cañizalez Escobar y Villanueva.

A mi tutora "Luz Rodriguez" por su paciencia y dedicación.

A mis profesores (Ismael Huerta, Mario Rodriguez, Eibar Hernández, Rómulo Castillo, Isabel Márquez) por transmitirme gran parte de sus conocimientos durante mi

formación académica.

A todos aquellos que de una u otra manera, creyeron y confiaron en mí, se que de corazón celebran a mi lado este hermoso sueño.

“Introducción a los Modelos Lineales Dinámicos: Modelo Polinomial de Primer Orden”

Resumen

Un punto de vista básico del modelaje científico es que un modelo es alguna descripción simplificada de un sistema que ayuda a realizar cálculos y predicciones, (Diccionario de Inglés Oxford).

En sentido más amplio, un modelo es algún esquema de descripción y explicación que organiza información y experiencias siempre y cuando signifique aprender y predecir. La primera razón para el modelaje es proveer procesos de aprendizaje los cuales aumentarán el conocimiento y permitirán decisiones sabias.

Las herramientas para las decisiones tecnológicas tales como los modelos matemáticos han sido aplicados a una amplia gama de situaciones en la toma de decisiones dentro de diversas áreas. En la toma consciente de decisiones bajo incertidumbre, siempre realizamos pronósticos o predicciones. Podríamos pensar que no estamos pronosticando, pero nuestras opciones estarían dirigidas por la anticipación de resultados de nuestras acciones o inacciones.

El Modelo Polinomial de Primer Orden será desarrollado en detalles como objetivo de este trabajo a modo de introducir los modelos lineales dinámicos. La teoría básica concerniente a dichos modelos, ayudan a un mejor manejo de la incertidumbre mediante el uso de técnicas de predicción y pronóstico efectivas.

El producto de este trabajo puede servir como material de apoyo para trabajos futuros relacionados con el área de probabilidad y estadística, específicamente con los modelos lineales dinámicos.

Índice

Introducción

En los últimos años, el análisis de modelos estadísticos ha sido un tema de mucho interés y que se ha desarrollado ampliamente dada la importancia que tiene en la resolución de problemas dentro de las aplicaciones en diversos campos científicos. Por esta razón, muchos esfuerzos se han realizado para proponer y construir nuevos modelos estadísticos útiles dada la necesidad de resolver problemas con distintas estructuras.

Las aproximaciones para el modelaje y predicción sintetiza conceptos, modelos y métodos cuyo desenvolvimiento ha sido influenciado por trabajos en muchos campos. Conjuntamente con los conceptos fundamentales de métodos científicos, se identifican las aproximaciones bayesianas como estructuras de rutina de aprendizaje y organización de conocimientos inciertos dentro de sistemas de predicción completos.

En los últimos 50 años se ha observado un rápido crecimiento de las aproximaciones bayesianas para el aprendizaje científico y la toma de decisiones. El uso de modelos matemáticos ha sido incrementado para interpretar y predecir las dinámicas y controles en la toma de decisiones en determinada situación; el caso del modelaje estadístico de los procesos de series de tiempo, por ejemplo, se basa sobre una clase de Modelos Dinámicos, ya que el término dinámico relaciona los cambios debido al paso del tiempo como un motivo de fuerza fundamental. Principalmente la idea es modelar y predecir solo series de tiempos de observaciones derivadas de un sistema o proceso de interés.

Capítulo 1
Preliminares

§1.1. Variables Aleatorias

La relación entre los sucesos del espacio muestral y el valor numérico que se les asigna se establece a través de variables aleatorias.

Definición **1.1.1.** Una variable aleatoria es una función que asigna un valor numérico a cada suceso elemental del espacio muestral.

Es decir, una variable aleatoria es una variable cuyo valor numérico está determinado por el resultado del experimento aleatorio. La variable aleatoria la denotaremos con letras mayúsculas X, Y, ... y con las letras minúscula x, y, ... sus valores.

La v.a. puede tomar un número numerable o no numerable de valores, dando lugar a dos tipos de v.a.: discretas y continuas.

Definición **1.1.2.** Se dice que una variable aleatoria X es discreta si puede tomar un número finito o infinito, pero numerable, de posibles valores.

Definición **1.1.3.** Se dice que una variable aleatoria X es continua si puede tomar un número infinito (no numerable) de valores, o bien, si puede tomar un número infinito de valores correspondientes a los puntos de uno o más intervalos de la recta real.

§1.2. Distribución Condicional

La distribución condicional es de interés, y ocurre cuando un grupo de variables aleatorias están siendo estudiadas mientras un segundo grupo se mantiene fijo.
Sean A y B dos eventos que pueden ocurrir en un espacio de 2-dimensiones, entonces por definición, la probabilidad condicional de B dado A ésta dada por:

$$P(B|A) = \frac{P(AB)}{P(A)}$$

si $P(A) \neq O$. Si A es un evento donde la variable aleatoria X está en el intervalo $a \leq X \leq b$, y B es un evento donde la variable aleatoria Y esta en el intervalo $c \leq Y \leq d$, entonces

$$P\{c \leq Y \leq d | a \leq X \leq b\} = \frac{P\{a \leq X \leq b, c \leq Y \leq d\}}{P\{a \leq X \leq b\}}$$

y en el caso continuo

$$P\{c \leq Y \leq d | a \leq X \leq b\} = \frac{\int_c^d \int_a^b f(x,y)dxdy}{\int_a^b g(x)dx}$$

donde $f(x,y)$ es la densidad conjunta de X,Y y $g(x)$ es la densidad marginal de X. La densidad condicional de Y dado $X = x$ está definida como:

$$h(y|x) = \frac{f(x,y)}{g(x)}.$$

Así,

$$P\{c \leq Y \leq d | X = x\} = \int_c^d h(y|x)dy.$$

Generalizado a una dimensión p, sean $X' = (X_1, ..., X_p), Y' = (X_1, ..., X_k)$ y $Z' = (X_{k+1}, ..., X_p)$ los vectores aleatorios y con letra minúscula denotaremos los valores observados. La densidad condicional de Y dado Z está dada por:

$$g(Y|Z) = \frac{f(y,z)}{h(z)} = \frac{f(x)}{h(z)}$$

donde $f(x)$ denota la densidad del vector aleatorio X, y $h(z)$ denota la densidad marginal del vector Z.

§1.3. Independencia

Dos vectores aleatorios, Y, Z; se dicen que son independiente si una de las siguientes aplicaciones se cumplen:

$$f(y,z) = g(y)h(z)$$

ó

$$F(y,z) = G(y)H(z)$$

ó

$$P\{Y|Z\} = g(y)$$

donde $f(y,z), g(y), h(z)$ son la densidad de $X = (Y,Z), Y, Z$ respectivamente; F,G,H son la respectivas Fda, y $P(y|z)$ es la densidad condicional de $Y|Z$.

§1.4. Esperanza

Sea $X : p \times 1$ un vector columna con $X_i, i = 1, ..., p$ componentes aleatorias; donde $f(X) = f(x_1, ..., x_p)$ es la función de densidad conjunta
Cuando ésta existe, la esperanza de un vector X está definido como:

$$E(X) = \begin{bmatrix} E(X_1) \\ \vdots \\ E(X_p) \end{bmatrix}$$

Análogamente, si $V : p \times n$, $E(V) = (E(V_{ij}))$, donde $V = (V_{ij})$

§1.4.1. Momento de segundo orden

La covarianza entre dos variables aleatorias Y y Z con momento de segundo orden finito está definido como:

$$Cov(Y,Z) = E[(Y - E(Y))(Z - E(Z))]$$

esto cuantitativamente puede ser positivo, negativo o cero; la covarianza matricial de un vector X está dada por la siguiente definición X:

$$\Sigma = (\sigma_{ij}) = E[(X - E(X))(X - E(X))']$$

para $i, j = 1, ..., p$. Un elemento típico de Σ es $\sigma_{ij} = E[(X_i - E(X_i))(X_j - E(X_j))]$, $i, j = 1, .., p$ cuando $j = i$ los elementos están ubicados a lo largo de la diagonal de Σ y es llamada la varianza de X.
Recordemos que

$$Var(X_i) = E(X_i - E(X_i))^2$$

si $i \neq j$, σ_{ij} es la covarianza de X_i y X_j. El coeficiente de correlación entre dos variables aleatorias escalares Y y Z con momento de segundo orden finito está definido por

$$\rho = corr(Y,Z) = \frac{Cov(Y,Z)}{[Var(Y)Var(Z)]^{1/2}}.$$

Esta es una medida de causa y efecto asociada con Y y Z. En general, $-1 \leq \rho \leq 1$, aunque en algunos casos, ρ es restringido a un intervalo más pequeño.

Una matriz de correlación es una matriz de coeficiente de correlación $R = (\rho_{ij})$ $i, j = 1, ..., p$. Esta matriz es útil para estudiar todas las asociaciones entre las componentes de un vector de variables aleatorias simultáneamente. La matriz de correlación es calculada en muchos modelos usados en análisis de datos multivariado ya que la matriz R frecuentemente provee un rápido entendimiento dentro de muchas relaciones insospechadas.

Los elementos de la diagonal, ρ_{jj}, de una matriz de correlación deberían ser todos uno, y los elementos fuera de la diagonal dados por:

$$\rho_{ij} = Corr(x_i, x_j) = \frac{cov(x_i, x_j)}{[Var(x_i)Var(x_j)]^{1/2}}, i \neq j$$

además, los ρ_{ij} deberán también satisfacer siempre la inecuación $-1 \leq \rho_{ij} \leq 1$ para $i, j = 1, ..., p$

§1.5. Teorema de Bayes, distribuciones a priori y a posteriori

Sean X, Θ vectores aleatorios de dimensión p y k respectivamente que están distribuidos conjuntamente con densidad condicional de X dado Θ denotada por $f(x|\theta)$, y la densidad marginal de Θ dada por $g(\theta)$; los datos han sido generados al observar X, para algún Θ fijo no-observado. Nos gustaría hacer inferencia sobre Θ considerando tanto nuestro prejuicio (creencia priori), como también las observaciones de $(X|\Theta)$ que indirectamente relacionan a ésta. El teorema de bayes proporciona un mecanismo formal para llevar a cabo esto. En terminología bayesiana $g(\theta)$ es llamada la priori de Θ o densidad a priori Θ ya que ésta es la densidad de Θ priori a los datos observados.

Teorema **1.5.1.** *(Bayes)*

Sean $g_1(\theta)$ la densidad a priori de Θ, y $f(x|\theta)$ la densidad condicional de X dado θ, entonces la densidad de Θ dado X=x, es dada por:

$$h(\theta|x) = \frac{f(x|\theta)g_1(\theta)}{\int_{-\infty}^{+\infty} f(x|\theta)g_1(\theta)d\theta}. \tag{1.1}$$

Demostración. Sean z y g_2 las funciones de densidad de (X, θ) y de X respectivamente. Por la definición de la densidad condicional

$$f(x|\theta) = \frac{z(x, \theta)}{g_1(\theta)} \tag{1.2}$$

$$h(\theta|x) = \frac{z(x, \theta)}{g_2(x)} \tag{1.3}$$

despejando $z(x, \theta)$ de (1.2) y reemplazando en (1.3) , obtenemos:

$$h(\theta|x) = \frac{f(x|\theta)g_1(\theta)}{g_2(x)} \tag{1.4}$$

por otra parte tenemos que:

$$\begin{aligned} g_2(x) &= \int_{-\infty}^{\infty} z(x,\theta)d\theta \\ &= \int_{-\infty}^{\infty} f(x|\theta)g_1(\theta)d\theta \end{aligned}$$

luego, sustituyendo $g_2(x)$ en (1.4) se tiene que:

$$h(\theta|x) = \frac{f(x|\theta)g_1(\theta)}{\int_{-\infty}^{+\infty} f(x|\theta)g_1(\theta)} d\theta.$$

□

Observemos que en (1.1) la integral depende solo de x y éste es fijo y conocido, así dicha integral es constante. Por lo tanto:

$$h(\theta|x) \propto f(x|\theta)g_1(\theta).$$

Una distribución para $\Theta : k \times 1$ posee todas la propiedades habituales de distribuciones de variables aleatorias observadas, excepto que Θ no es observada. Este tipo de distribuciones son llamadas distribuciones de probabilidad subjetiva.

EJEMPLO **1.5.1.** Se desea estimar la probabilidad, θ, de un evento, a partir del resultado de una sucesión de n *ensayos Bernoulli*, esto es, datos $x_1, x_2, ..., x_n$ que son iguales a uno si ocurre el evento (éxito) y cero si no ocurre. Sea x el número total de éxitos en la muestra de n ensayos. En este caso, el modelo muestral establece que:

$$f(x|\theta) = Bin(y|n,\theta) = \begin{pmatrix} n \\ x \end{pmatrix} \theta^x (1-\theta)^{n-x}$$

Suponiendo que $g_1(\theta)$ es uniforme en el intervalo [0,1], se tiene que:

$$h(\theta|x) \propto \theta^x (1-\theta)^{n-x}$$

o lo que es lo mismo, la distribución no normalizada tiene un kernel equivalente a la distribución *Beta*, es decir:

$$\theta|x \sim Beta(x+1, n-x+1)$$

§1.6. Densidad general de la distribución Normal multivariada

Sea $X : p \times 1$ un vector aleatorio con función de densidad $f(x)$. X tiene una distribución Normal multivariada (p variada) no-singular con vector de media $\theta : p\times 1$ y matriz covarianza $\Sigma : p \times p$ si

$$f(x) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} exp\left[-\frac{1}{2}(x-\theta)'\Sigma^{-1}(x-\theta)\right] \tag{1.5}$$

para $\Sigma > 0$. Si $|\Sigma| = 0$, la distribución de X es llamada singular o normal degenerada y la densidad no existe.
Denotaremos esta distribución de X por:

$$\mathcal{L}(X) = N(\theta, \Sigma).$$

§1.6.1. Distribución Normal Bivariada

Sea $X : 2 \times 1$ un vector aleatorio bivariado con $\mathcal{L}(X) = N(\theta, \Sigma)$ y $\Sigma > 0$. Sean $\theta = (\theta_i)$y $\Sigma = (\sigma_{ij})$ i,j=1,2. Para simplificar tomemos $\sigma_{11} = \sigma_1^2, \sigma_{22} = \sigma_2^2$ y $\sigma_{12} = \rho\sigma_1\sigma_2$ donde ρ es el coeficiente de correlación entre X_1 y X_2. Si escribimos (1.5) para $p = 2$ tenemos que la densidad bivariada es dada por:

$$\begin{aligned} f(x) &= f(x_1, x_2) \\ &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times \\ & exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\theta_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\theta_1}{\sigma_1}\right)\left(\frac{x_2-\theta_2}{\sigma_2}\right) + \left(\frac{x_2-\theta_2}{\sigma_2}\right)^2\right]\right\}. \end{aligned} \tag{1.6}$$

Acá

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\theta_2 & \sigma_2^2 \end{pmatrix} \qquad \Sigma^{-1} = \begin{pmatrix} \frac{1}{\sigma_1^2(1-\rho^2)} & -\frac{\rho}{\sigma_1\sigma_2(1-\rho^2)} \\ -\frac{\rho}{\sigma_1\sigma_2(1-\rho^2)} & \frac{1}{\sigma_2^2(1-\rho^2)} \end{pmatrix}.$$

La expresión entre corchetes de (1.6) controla la varianza de $f(x)$, es decir; si la expresión entre corchetes es constante entonces $f(x)$ es constante y viceversa.

§1.6.2. Independencia

Sea $\mathcal{L}(X) = N(\theta, \Sigma)$ y $X : 2\times 1$. Entonces, si $\rho = 0$ en (1.6), X_1 y X_2 no sólo están no-correlacionados, también son independientes. Es fácil ver que al sustituir $\rho = 0$ en (1.6) $f(x_1, x_2)$ se reduce al producto de una función de x_1 y una función de x_2. Por supuesto, lo contrario es también cierto, es decir; sí X_1 y X_2 son independientes, entonces X_1 y X_2 son no-correlacionados; en este sentido, el resultado se cumple para toda distribución bivariada (mientras que en el otro sentido, la falta de correlación generalmente no implica la independencia, aunque si para la distribución normal)).

§1.7. Modelos Lineales

§1.7.1. Modelo Lineal Simple

Un modelo de la forma

$$Y_i = \beta_0 + \beta_1 X_1 + \varepsilon_i, \quad i = 1, 2, ..., n \tag{1.7}$$

donde Y_i es la i-ésima observación de la variable de repuesta, la cual corresponde al i-ésimo valor X_i de la variable de predicción, ε_i es el error aleatorio no observable asociado con Y_i; y β_0 y β_1 son los parámetros desconocidos que representan la intersección y la pendiente, respectivamente. La expresión (1.7) se conoce como modelo lineal simple, debido a que es lineal en los parámetros y se tiene sólo una variable de predicción. Cada observación Y_i es una variable aleatoria que es la suma de dos componentes; el término no aleatorio $\beta_0 + \beta_1 X_1$, y la componente aleatoria ε_i.

§1.7.2. Modelo Lineal General

Sean $X_1, X_2, ..., X_k$ k variables de predicción, las cuales pueden tener alguna influencia sobre una repuesta Y, y supóngase que el modelo tiene la forma

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_k X_{ik} + \varepsilon_i \, , \quad i = 1, 2, ..., n \tag{1.8}$$

donde Y_i es la i-ésima observación de la repuesta para un conjunto de valores fijo $X_{i1}, X_{i2}, ..., X_{ik}$ de las variables de predicción, ε_i es el error aleatorio no observable asociado con Y_i, y $\beta_0, \beta_1, ..., \beta_k$ son $m = k + 1$ parámetros lineales desconocidos. La ecuación (1.8) recibe el nombre de modelo lineal general y da origen a lo que se conoce como una regresión lineal múltiple.

Si se supone el caso de la teoría basada en el modelo normal, las observaciones Y_i son variables aleatorias independientes, normalmente distribuidas con

$$E(Y_i) = \beta_0 + \beta_1 X_{i1} + ... + \beta_k X_{ik}$$

$$Var(Y_i) = \sigma^2 \, , \;\; i = 1, 2, ..., n \, .$$

La única restricción funcional que se impone al modelo lineal general es que sea lineal en los parámetros desconocidos; el modelo no tiene ninguna restricción con respecto a la naturaleza de las variables de predicción.

Cuando en (1.8) se tiene $X_{ij} = X_i^j$; $i = 1, 2, ..., n$ y $j = 1, 2, ..., n$ entonces el modelo lineal general toma la forma

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + ... + \beta_k X_i^k + \varepsilon_i \tag{1.9}$$

la cual se conoce como modelo polinomial.

Regresaremos al modelo general dado en (1.7) para obtener los estimadores de mínimos cuadrados de los parámetros y para desarrollar técnicas de regresión para este modelo. Se empleará el álgebra de matrices, ya que ésta simplifica en gran medida la presentación.

Dada una muestra aleatoria de observaciones $Y_1, Y_2, ..., Y_n$ en los puntos de observación $X_{11}, X_{12}, ..., X_{1k}, X_{21}, X_{22}, ..., X_{2k}, ..., X_{n1}, X_{n2}, ..., X_{nk}$ respectivamente, con base en el modelo lineal general, se tiene las n ecuaciones siguientes:

$$\begin{aligned} Y_1 &= \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} + ... + \beta_k X_{1k} + \varepsilon_1 \\ Y_2 &= \beta_0 + \beta_1 X_{21} + \beta_2 X_{22} + ... + \beta_k X_{2k} + \varepsilon_2 \\ &. \\ &. \\ &. \\ Y_n &= \beta_0 + \beta_1 X_{n1} + \beta_2 X_{n2} + ... + \beta_k X_{nk} + \varepsilon_n \end{aligned}$$

Como resultado el modelo general también puede expresarse en forma matricial como:

$$Y = X\beta + \varepsilon$$

donde X es una matriz $n \times m$ para las variables de predicción, y β es un vector de parámetros desconocidos $m \times 1$, $(m = k + 1)$ mientras que Y y ε siguen siendo vectores $n \times 1$, los que contienen las observaciones de la variable de repuesta y los errores aleatorios asociados con éstas, respectivamente

$$Y = \begin{bmatrix} Y_1 \\ . \\ . \\ . \\ Y_n \end{bmatrix}_{n \times 1}, X = \begin{bmatrix} 1 & X_{11} & . & . & . & X_{1k} \\ 1 & X_{21} & . & . & . & X_{2k} \\ . & & & & & \\ . & & & & & \\ . & & & & & \\ 1 & X_{n1} & . & . & . & X_{nk} \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ . \\ . \\ . \\ \beta_k \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ . \\ \varepsilon_n \end{bmatrix} .$$

Bajo el caso de la teoría normal

$$Y \sim N(X\beta, \sigma^2 I)$$

$$\varepsilon \sim N(0, \sigma^2 I)$$

donde $Var(Y) = Var(\varepsilon) = \sigma^2 I$, de esta manera Y y ε son vectores de variables aleatorias independientes distribuidos normalmente.

§1.8. Modelo Bayesiano y de Predicción

Importantes influencias sobre el pensamiento y práctica bayesiana pueden ser encontrados, en particular, en los textos de Box y Tiao (1973), De Groot (1971), Jeffreys (1961) y Lindley (1965). Una notable y reciente contribución es la de Berger (1985), el cual incluye una compresiva bibliografía. En cuanto a las aproximaciones para la práctica de predicción, poco se había hecho hasta que las computadoras comenzaron aparecer en el año 1950. Por ello, a finales de 1960, los modelos básicos ya habían sido extendidos y generalizados a una completa clase de modelos lineales dinámicos, y a modelos multi-procesos.

Definición **1.8.1.** Una predicción es una hipótesis, conjetura, o especulación hecha acerca de algo futuro.

Consideremos que una variable de salida Y está relacionada con una variable de entrada X mediante una parametrización de la forma

$$Y = X\theta + \varepsilon$$

donde θ es un parámetro desconocido y ε es un término de error aleatorio. Además, supongamos que las creencias de un predictor acerca del parámetro θ son expresadas a través de una distribución de probabilidad $P(\theta)$. Esto incluye una forma de modelos de incertidumbre. La dinámica natural de procesos y sistemas requiere también que se reconozca incertidumbre debido al paso del tiempo. Así tomamos esa posición, y es necesario pensar que θ está cambiando lentamente con el tiempo.

§1.8.1. Series de Tiempo

Definición **1.8.2.** Una serie de tiempo es una colección de observaciones tomadas a lo largo del tiempo cuyo objetivo principal es describir, explicar, predecir y controlar algún proceso. Las series de tiempo se utilizan para predecir lo que ocurrirá con una variable en el futuro a partir del comportamiento de esa variable en el pasado.

En la modelización de una serie, los valores reales de las cantidades observadas en el tiempo se denotan por Y, de modo que Y_t denota el valor de la observación en el tiempo t . Los modelos matemáticos y estadísticos de procesos de series de tiempo se basan en clases de Modelos Dinámicos, el término dinámico se relaciona a los cambios que se producen en estos procesos debido al paso del tiempo como una fuerza motriz fundamental. El más conocido y utilizado es el Modelo Lineal Dinámico Normal, que se refiere simplemente como Modelos Lineales Dinámicos, cuando se tiene la normalidad. Esta clase de modelos es la base para el desarrollo de éste trabajo.

Capítulo 2

Modelo Lineal Bayesiano

En inferencia Bayesiana, una distribución a priori para los parámetros debería ser especificada, además de la función de verosimilitud. En primer lugar, los resultados involucran prioris propias (por ejemplo la conjugada natural). Asumamos que el vector de parámetro β tiene una distribución a priori condicional $N(\mu_0, \phi^{-1}C_0^{-1})$ donde $\phi = \sigma^{-2}$ y que $n_0\sigma_0^2\phi \sim \chi^2_{n_0}$. Así, la distribución a priori es totalmente especificada con densidad dada por

$$\begin{aligned}
p(\beta, \phi) &= (2\pi)^{-p/2}|\phi C_0|^{p/2} \exp\left\{-\frac{\phi}{2}(\beta - \mu_0)' C_0(\beta - \mu_0)\right\} \times \\
&\quad \frac{(n_0\sigma_0^2/2)^{n_0/2}}{\Gamma(n_0/2)}\phi^{(n_0/2)-1} \exp\left\{-\frac{n_0\sigma_0^2}{2}\phi\right\} \\
&\propto \phi^{[(n_0+p)/2]-1} \exp\left\{-\frac{\phi}{2}[n_0\sigma_0^2 + (\beta - \mu_0)' C_0(\beta - \mu_0)]\right\} .
\end{aligned}$$

Entonces, como en el caso univariado, la distribución condicional de $\phi|\beta$ puede ser obtenida de la distribución a priori conjunta de β y ϕ coleccionando solo los términos que involucran a ϕ. Esto está dado por $[n_0\sigma_0^2 + (\beta - \mu_0)' C_0(\beta - \mu_0)]\phi|\beta \sim \chi^2_{n_0+p}$. La distribución conjunta marginal de β puede ser obtenida dividiendo $p(\beta, \phi)$ por $p(\phi|\beta)$, o integrando la distribución con respecto a ϕ. Está densidad está dada por:

$$p(\beta) \propto [n_0\sigma_0^2 + (\beta - \mu_0)' C_0(\beta - \mu_0)]^{-(n_0+p)/2} .$$

En efecto,

$$\begin{aligned}
p(\beta) &= \int p(\beta, \phi) d\phi \\
&= \int (2\pi)^{-p/2}|\phi C_0|^{p/2} \exp\left\{-\frac{\phi}{2}(\beta - \mu_0)' C_0(\beta - \mu_0)\right\} \times \\
&\times \frac{(n_0\sigma_0^2/2)^{n_0/2}}{\Gamma(n_0/2)}\phi^{(n_0/2)-1} \exp\left\{-\frac{n_0\sigma_0^2}{2}\phi\right\} d\phi \\
&= |C_0|^{p/2}(2\pi)^{-p/2}\frac{(n_0\sigma_0^2/2)^{n_0/2}}{\Gamma(n_0/2)} \int \phi^{(\frac{p}{2}+\frac{n_0}{2})-1} \exp\left\{-\frac{\phi}{2}[n_0\sigma_0^2 + (\beta - \mu_0)' C_0(\beta - \mu_0)]\right\}
\end{aligned}$$

ϕ tiene una distribución gamma con parámetros α y λ, denotado por $\phi \sim G(\alpha, \lambda)$, la función de densidad esta dada por

$$p(\phi|\alpha, \lambda) = k\phi^{\alpha-1}e^{-\lambda\phi}$$

$\phi > 0$ para $\alpha, \lambda > 0$. La constante k es dada por $k = \lambda^{\alpha}/\Gamma(\alpha)$. Esta distribución tiene media α/β y varianza α/β^2.

Luego se tiene para

$$\lambda = \frac{1}{2}[n_0\sigma_0^2 + (\beta - \mu_0)^{'}C_0(\beta - \mu_0)]$$

y

$$\alpha = \frac{p + n_0}{2}\,, \;\; k = \frac{\left\{\frac{1}{2}[n_0\sigma_0^2 + (\beta - \mu_0)^{'}C_0(\beta - \mu_0)]\right\}^{\frac{p+n_0}{2}}}{\Gamma(p + n_0/2)}\,.$$

Haciendo $C = |C_0|^{p/2}(2\pi)^{-p/2}\frac{(n_0\sigma_0^2/2)^{n_0/2}}{\Gamma(n_0/2)}$ se tiene que

$$\begin{aligned} p(\beta) &= C\,\frac{\frac{1}{2}[n_0\sigma_0^2 + (\beta - \mu_0)^{'}C_0(\beta - \mu_0)]^{(p+n_0/2)}/\Gamma(p + n_0/2)}{\frac{1}{2}[n_0\sigma_0^2 + (\beta - \mu_0)^{'}C_0(\beta - \mu_0)]^{(p+n_0)/2}/\Gamma(p + n_0/2)}\int \phi^{\alpha-1}\exp\{-\lambda\phi\}d\phi \\ &= C\,\frac{2\Gamma(\frac{p+n_0}{2})}{[n_0\sigma_0^2 + (\beta - \mu_0)^{'}C_0(\beta - \mu_0)]^{\frac{p+n_0}{2}}} \end{aligned}$$

por tanto

$$p(\beta) \propto [n_0\sigma_0^2 + (\beta - \mu_0)^{'}C_0(\beta - \mu_0)]^{-\frac{(p+n_0)}{2}}\,,$$

el cual corresponde a la densidad de $t_{n_0}(\mu_0, \sigma_0^2C_0^{-1})$. La constante de normalización es

$$\frac{\Gamma[(n_0 + p)/2]}{\Gamma(\frac{n_0}{2})n_0^{p/2}}(n_0\sigma_0^2)^{n_0/2}|C_0|^{1/2}\,.$$

En otras palabras, la verosimilitud de β y ϕ es dada por

$$\phi^{n/2}\exp\{-\frac{\phi}{2}[S_e + (\beta - \hat{\beta})^{'}X^{'}X(\beta - \hat{\beta})]\}$$

y tiene la misma forma que la densidad a priori. Por lo tanto, la posteriori esta dada por

$$p(\beta, \phi|y) \propto \phi^{((n+n_0+p)/2)-1}\times\exp\{-\frac{\phi}{2}[n_0\sigma_0^2 + S_e + (\beta - \mu_0)^{'}C_0(\beta - \mu_0) + (\beta - \hat{\beta})^{'}X^{'}X(\beta - \hat{\beta})]\}$$

ahora mostremos que:

$$(\beta - \mu_0)^{'}C_0(\beta - \mu_0) + (\beta - \hat{\beta})^{'}X^{'}X(\beta - \hat{\beta}) = (\beta - \mu_1)^{'}C_1(\beta - \mu_1) + \mu_0^{'}C_0\mu_0 + \hat{\beta}^{'}X^{'}X\hat{\beta} - \mu_1^{'}C_1\mu_1$$

donde $\mu_1 = C_1^{-1}(C_0\mu_0 + X^{'}y)$ y $C_1 = C_0 + X^{'}X$.
En efecto, sea $G = (\beta-\mu_0)^{'}C_0(\beta-\mu_0) + (\beta-\hat{\beta})^{'}X^{'}X(\beta-\hat{\beta})$

$$\begin{aligned}
G &= (\beta^{'}-\mu_0^{'})(C_0\beta - C_0\mu_0) + (\beta^{'}-\hat{\beta}^{'})X^{'}X(\beta-\hat{\beta})\\
&= (\beta^{'}C_0\beta - \beta^{'}C_0\mu_0 - \mu_0^{'}C_0\beta + \mu_0^{'}C_0\mu_0) + (\beta^{'} - [(X^{'}X)^{-1}X^{'}Y]^{'})X^{'}X(\beta-\hat{\beta})
\end{aligned}$$

ya que $\hat{\beta} = (X^{'}X)^{-1}X^{'}Y$

$$\begin{aligned}
G &= (\beta^{'}C_0\beta - \beta^{'}C_0\mu_0 - \mu_0^{'}C_0\beta + \mu_0^{'}C_0\mu_0) + \beta^{'}X^{'}X\beta - \beta^{'}X^{'}X\hat{\beta}\\
&\quad -((X^{'}X)^{-1}X^{'}Y)^{'}(X^{'}X\beta) + ((X^{'}X)^{-1}X^{'}Y)^{'}X^{'}X\hat{\beta}\\
&= (\beta^{'}C_0\beta - \beta^{'}C_0\mu_0 - \mu_0^{'}C_0\beta + \mu_0^{'}C_0\mu_0) + \beta^{'}X^{'}X\beta - \beta^{'}X^{'}X\hat{\beta}\\
&\quad -(X^{'}Y)^{'}(X^{'}X)^{-1}(X^{'}X)\beta + \hat{\beta}^{'}X^{'}X\hat{\beta}\\
&= \beta^{'}C_0\beta - \beta^{'}C_0\mu_0 - \mu_0^{'}C_0\beta + \mu_0^{'}C_0\mu_0 + \beta^{'}X^{'}X\beta - \beta^{'}X^{'}X\hat{\beta} - Y^{'}X\beta\\
&\quad +\hat{\beta}^{'}X^{'}X\hat{\beta}\\
&= \beta^{'}C_0\beta - \beta^{'}C_0\mu_0 - \mu_0^{'}C_0\beta + \mu_0^{'}C_0\mu_0 + \beta^{'}X^{'}X\beta - \beta^{'}X^{'}X((X^{'}X)^{-1}X^{'}Y)\\
&\quad -Y^{'}X\beta + \hat{\beta}^{'}X^{'}X\hat{\beta}\\
&= \beta^{'}C_0\beta - \beta^{'}C_0\mu_0 - \mu_0^{'}C_0\beta + \mu_0^{'}C_0\mu_0 + \beta^{'}X^{'}X\beta - \beta^{'}X^{'}Y - Y^{'}X\beta + \hat{\beta}^{'}X^{'}X\hat{\beta}\\
&= \beta^{'}(C_0 + X^{'}X)\beta - (\mu_0^{'}C_0 + Y^{'}X)\beta - \beta^{'}(C_0\mu_0 + X^{'}Y) + \mu_0^{'}C_0\mu_0 + \hat{\beta}^{'}X^{'}X\hat{\beta}
\end{aligned}$$

sumando y restando $\mu_1^{'}C_1\beta$ y notando que $C_1 = C_0\mu_0 + X^{'}X$ se tiene

$$\begin{aligned}
G &= \beta^{'}C_1\beta - (\mu_0^{'}C_0 + Y^{'}X)\beta - \beta^{'}(C_0\mu_0 + X^{'}Y) + \mu_0^{'}C_0\mu_0 + \hat{\beta}^{'}X^{'}X\hat{\beta} + \mu_1^{'}C_1\beta\\
&\quad -\mu_1^{'}C_1\beta
\end{aligned}$$

Ahora sumando y restando $\mu_1^{'}C_1\mu_1$ se tiene

$$\begin{aligned}
G &= \beta^{'}C_1\beta - \beta^{'}(C_0\mu_0 + X^{'}Y) + \mu_1^{'}C_1\beta - \mu_1^{'}C_1\beta + \mu_0^{'}C_0\mu_0 - (\mu_0^{'}C_0 + Y^{'}X)\beta\\
&\quad +\hat{\beta}^{'}X^{'}X\hat{\beta} + \mu_1^{'}C_1\mu_1 - \mu_1^{'}C_1\mu_1\\
&= (\beta-\mu_1)^{'}C_1(\beta-\mu_1) + \mu_0^{'}C_0\mu_0 + \hat{\beta}^{'}X^{'}X\hat{\beta} - \mu_1^{'}C_1\mu_1 + \mu_1^{'}C_1\beta\\
&\quad -(\mu_0^{'}C_0 + Y^{'}X)\beta
\end{aligned}$$

donde

$$\begin{aligned}
\beta^{'}C_1\mu_1 &= \beta^{'}C_1C_1^{-1}(C_0\mu_0 + X^{'}Y)\\
&= \beta^{'}(C_0\mu_0 + X^{'}Y)\,.
\end{aligned}$$

Por otro lado, como C_1 es simétrica se tiene que $C_1 = C_1^{'}$ así,

$$\begin{aligned} \mu_1^{'} C_1 \beta &= (C_0\mu_0 + X^{'}Y)^{'} C_1^{-1^{'}} (C_1\beta) \\ &= (\mu_0^{'} C_0^{'} + Y^{'}X)\beta \\ &= (\mu_0^{'} C_0 + Y^{'}X)\beta \quad pues \;\; C_0 = C_0^{'} \ . \end{aligned}$$

Así se tiene

$$\begin{aligned} G &= (\beta-\mu_1)^{'} C_1(\beta-\mu_1) + \mu_0^{'} C_0 \mu_0 + \hat{\beta}^{'} X^{'} X \hat{\beta} - \mu_1^{'} C_1 \mu_1 + (\mu_0^{'} C_0 + Y^{'} X)\beta \\ &\quad -(\mu_0^{'} C_0 + Y^{'} X)\beta \\ &= (\beta-\mu_1)^{'} C_1(\beta-\mu_1) + \mu_0^{'} C_0 \mu_0 + \hat{\beta}^{'} X^{'} X \hat{\beta} - \mu_1^{'} C_1 \mu_1 \end{aligned}$$

por tanto

$$(\beta-\mu_0)^{'} C_0(\beta-\mu_0) + (\beta-\hat{\beta})^{'} X^{'} X(\beta-\hat{\beta}) = (\beta-\mu_1)^{'} C_1(\beta-\mu_1) + \mu_0^{'} C_0 \mu_0 + \hat{\beta}^{'} X^{'} X \hat{\beta} - \mu_1^{'} C_1 \mu_1 \ .$$

Notemos que aún cuando X no tiene rango completo, C_1 tendrá inversa y siempre puede ser invertible. Reescribiendo los términos

$$\begin{aligned} S_e + \mu_0^{'} C_0 \mu_0 + \hat{\beta}^{'} X^{'} X \hat{\beta} - \mu_1^{'} C_1 \mu_1 &= y^{'} y - \hat{\beta}^{'} X^{'} X \hat{\beta} + \mu_0^{'} C_0 \mu_0 + \hat{\beta}^{'} X^{'} X \hat{\beta} \\ &\quad -\mu_1^{'} C_1 (C_1^{-1}(C_0\mu_0 + X^{'} y)) \\ &= y^{'} y - \hat{\beta}^{'} X^{'} X \hat{\beta} + \mu_0^{'} C_0 \mu_0 + \hat{\beta}^{'} X^{'} X \hat{\beta} - \mu_1^{'} C_0 \mu_0 \\ &\quad -\mu_1^{'} X^{'} y \\ &= (y^{'} - \mu_1^{'} X^{'}) y + (\mu_0^{'} - \mu_1^{'}) C_0 \mu_0 \\ &= (y - X\mu_1)^{'} y + (\mu_0 - \mu_1)^{'} C_0 \mu_0 \ ; \end{aligned}$$

la densidad posterior de β y ϕ puede escribirse como

$$p(\beta,\phi|y) \propto \phi^{((n+n_0+p)/2)-1} \times \exp\{-\frac{\phi}{2}[n_0\sigma_0^2 + S_e + (\beta-\mu_1)^{'} C_1(\beta-\mu_1) + \mu_0^{'} C_0 \mu_0 + \hat{\beta}^{'} X^{'} X \hat{\beta} - \mu_1^{'} C_1 \mu_1]\}$$

$$p(\beta,\phi|y) \propto \phi^{p/2} \exp\{-\frac{\phi}{2}(\beta-\mu_1)^{'} C_1(\beta-\mu_1)\} \phi^{(\frac{n+n_0}{2})-1} \exp\{-\frac{\phi}{2}(n_0\sigma_0^2 + (y-\mu_1)^{'} y + (\mu_0-\mu_1)^{'} C_0 \mu_0)\} \ .$$

Haciendo $n_1 = n + n_0$ y $n_1\sigma_1^2 = n_0\sigma_0^2 + (y - X\mu_1)^{'} y + (\mu_0 - \mu_1)^{'} C_0 \mu_0$ se tiene que

$$p(\beta,\phi|y) \propto \phi^{p/2} \exp\{-\frac{\phi}{2}(\beta-\mu_1)^{'} C_1(\beta-\mu_1)\} \phi^{(\frac{n_1}{2})-1} \exp\{-\frac{\phi}{2} n_1 \sigma_1^2\} \ .$$

Esta densidad tiene la misma forma de la priori y, también, es conjugada de los modelos lineales normales. En particular, $\beta|y \sim t_{n_1}(\mu_1, \sigma_1^2 C_1^{-1})$ y $\beta_j|y \sim t_{n_1}(\mu_{1j}, \sigma_1^2 (C_1^{-1})_{jj})$, $j = 1, ..., p$. La media posteriori y matriz de varianza-covarianza de β están dadas respectivamente, por

$$\mu_1 \text{ y } \tfrac{n_1}{n_1-2}\sigma_1^2 C_1^{-1}\,, \quad n_1 > 2\ .$$

La distribución posteriori de ϕ es $n_1\sigma_1^2\phi|y \sim \chi^2_{n_1}$ con media σ_1^{-2}. Los estimadores puntuales de β y ϕ están dados por μ_1 y σ_1^{-2}, respectivamente. Intervalos de confianza para β_j y ϕ son obtenidos de los percentiles de las distribuciones t_{n_1} y $\chi^2_{n_1}$ respectivamente. También es posible hacer inferencia sobre la distribución conjunta de β basada en el hecho de que $(\beta-\mu_1)'C_1(\beta-\mu_1)/\sigma_1^2|y \sim F(p, n_1-p)$.
La priori no informativa puede ser usada para representar, en algún sentido, la ausencia de información inicial. Como β es esencialmente un parámetro de localización(multivariado) y ϕ es un parámetro de escala, se deduce que la distribución a priori conjunta no-informativa está dada por $p(\beta,\phi) \propto \phi^{-1}$. Esta priori es un caso particular, degenerado de la conjugada a priori natural. Haciendo las sustituciones convenientes la densidad a posteriori es

$$\begin{aligned} p(\beta,\phi|y) &\propto \phi^{(n/2)-1}\exp\left\{-\frac{\phi}{2}[S_e + (\beta-\hat\beta)'X'X(\beta-\hat\beta)]\right\} \\ &\propto \phi^{p/2}\exp\left\{-\frac{\phi}{2}(\beta-\hat\beta)'X'X(\beta-\hat\beta)\right\}\phi^{((n-p)/2)-1}\exp\left\{-\frac{\phi}{2}(n-p)s^2\right\}\ . \end{aligned}$$

Por lo tanto, la posteriori seguirá perteneciendo a la misma clase cambiando solamente los valores de los hiperparámetros de las distribuciones relevantes. Así, $\beta|y \sim t_{n-p}(\hat\beta, s^2(X'X)^{-1})$ y $(n-p)s^2\phi|y \sim \chi^2_{n-p}$ y la forma cuadrática en β se reducirá a $(\beta-\hat\beta)X'X(\beta-\hat\beta)/s^2$ con distribución a posteriori $F(p, n-p)$. Estas distribuciones proporcionan resultados Bayesianos paralelos a los que se obtienen usando la aproximación clásica.

Las distribuciones de predicción son necesarias para realizar las pruebas de hipótesis y predicción desde el punto de vista Bayesiano. Estas distribuciones se pueden obtener mediante la integración de la distribución muestral con respecto a la distribución de los parámetros. En el caso normal, este trabajo es simplificado por la estructura lineal de la distribución. Consideremos el modelo

$$Y = X\beta + e,\ \ e \sim N(0, \phi^{-1}I_n)$$

donde $\phi = 1/\sigma^2$ y supongamos que la distribución condicional a priori de β es

$$\beta|\phi \sim NM(\mu, \phi^{-1}C^{-1})\,.$$

Combinando estos resultados, tenemos

$$\begin{aligned} Y|\phi &\sim XN(\mu, \phi^{-1}C^{-1}) + N(0, \phi^{-1}I_n) \\ &\sim N[X\mu, \phi^{-1}(I_n + XC^{-1}X^{'})] \end{aligned}$$

ya que las distribuciones de arriba son normal independientes.
Suponiendo ahora que $\nu\sigma_0^2\phi \sim \chi_\nu^2$ obtenemos que

$$Y \sim t_\nu[X\mu, \sigma_0^2(I_n + XC^{-1}X^{'})]$$

La distribución marginal de Y_i es $t_\nu[x_i\mu_i, \sigma_0^2(1 + x_iC^{-1}x_i^{'})]$ $i = 1, ..., n$

Para comparar dos hipótesis H_0 y H_1, la evaluación de la distribución de $(Y|H_l)$, $l = 1, 0$ es requerida. La validación del test del modelo se basa en el factor de Bayes

$$BF(H_0; H_1) = \frac{p(Y|H_0)}{P(Y|H_1)}$$

donde el denominador es la densidad de distribución dada por

$$\frac{\Gamma[(n_0+n)/2]}{\Gamma(n_0/2)n_0^{n/2}}(n_0\sigma_0^2)^{n/2}|I_n+XC_0^{-1}X^{'}|^{-1/2}\times[n_0\sigma_0^2+(y-X\mu_0)^{'}(I_n+XC_0^{-1}X^{'})^{-1}(y-X\mu_0)]^{-\frac{n_0+n}{2}}.$$

Si deseamos probar las hipótesis de validación del modelo, construimos $H_0 : \beta_2 = ... = \beta_p = 0$. Bajo H_0, el modelo simplifica a $Y = 1_n\beta_1 + e$ y $\beta_1|\phi \sim N[\mu_{00}, (c_{00}\phi)^{-1}]$ y por tanto $Y|\phi \sim N[1_n\mu_{00}, \phi^{-1}(I_n + c_{00}^{-1}1_n1_n^{'})]$.
Entonces, la densidad $p(y|H_0)$ es dada por

$$\frac{\Gamma[(n_0+n)/2]}{\Gamma(n_0/2)n_0^{n/2}}(n_0\sigma_0^2)^{n/2}|I_n+c_{00}^{-1}1_n1_n^{'}|^{-1/2}\times[n_0\sigma_0^2+(y-1_n\mu_{00})^{'}(I_n+c_{00}^{-1}1_n1_n^{'})^{-1}(y-1_n\mu_{00})]^{-\frac{n_0+n}{2}}.$$

El factor de Bayes está, entonces dado por

$$\left\{\frac{|I_n + XC_0^{-1}X^{'}|}{|I_n + c_{00}^{-1}1_n1_n^{'}|}\right\}^{1/2} \times \left\{\frac{|n_0\sigma_0^2 + (y - X\mu_0)^{'}(I_n + XC_0^{-1}X^{'})^{-1}(y - X\mu_0)|}{|n_0\sigma_0^2 + (y - 1_n\mu_{00})^{'}(I_n + c_{00}^{-1}1_n1_n^{'})^{-1}(y - 1_n\mu_{00})|}\right\}^{(\frac{n_0+n}{2})}.$$

En el caso de la predicción de un vector de dimensión m de las observaciones futuras Y^* con la matriz x^* de variable explicativa, el mismo resultado usado anteriormente se puede aplicar en cuanto la predicción sea basada en la distribución predictiva de $Y^*|y$ con densidad dada por:

$$\begin{aligned} p(y^*|y) &= \int\int p(y^*|\beta;\phi, y)p(\beta, \phi|y)d\beta d\phi \\ &= \int\left\{\int p(y^*|\beta, \phi)p(\beta|\phi, y)d\beta\right\}p(\phi|y)d\phi \end{aligned}$$

y el cálculo es similar a los involucrados en la evaluación del factor de Bayes. Una diferencia importante es que las marginalizaciones son con respecto a la distribución posteriori de los parámetros, mientras que el factor de Bayes es obtenido usando marginalizaciones con respecto a la distribución a priori. Entonces, utilizando la notación adoptada se sigue que

$$Y^*|y \sim t_{n_1}(x^*\mu_1, \sigma_1^2(I_m + x^*C_1^{-1}(x^*)^{'}))$$

y así los puntos de predicción y los intervalos de confianza para Y^* pueden obtenerse fácilmente. El análisis usando prioris no-informativas conduce a $\mu_1 \longrightarrow \hat{\beta}$, $C_1 \longrightarrow XX^{'}$, $n_1 \longrightarrow n$ y $\sigma_1^2 \longrightarrow s^2$ y la distribución predictiva de Y^* se reduce a

$$Y^*|y \sim t_{n-p}(x^*\hat{\beta}, s^2(I_m + x^*(X^{'}X)^{-1}(x^*)^{'}))\,.$$

Esta distribución coincide con la distribución predictiva de la aproximación clásica.

Capítulo 3

Modelos Lineales Dinámicos

Muchos conceptos fundamentales y características analíticas de los modelos lineales dinámicos están de manifiesto en el caso más simple y completo más usado, El Modelo Polinomial de Primer Orden. El modelo polinomial de primer orden es el más simple, aunque no trivial, modelo de serie de tiempo en el cual la serie de observación Y_t es representada como $Y_t = \mu_t + \nu_t$, donde μ_t es el nivel común de la serie en un tiempo t , y $\nu_t \sim N[0, V_t]$ es el error de observación o el término de ruido. La evolución del tiempo del nivel de la serie es un simple paseo aleatorio $\mu_t = \mu_{t-1} + \omega_t$, con el error de evolución $\omega_t \sim N[0, Wt]$. Esta última ecuación describe lo que a menudo se refiere como un modelo localmente constante. Asumimos que las dos condiciones de error, de observación y errores de evolución, son distribuidas normalmente para cada t. Además suponemos que las secuencias de error son independientes con el tiempo y mutuamente independientes. Así, para todo t y todo s con $t \neq s$, ν_t y ν_s son independiente, ω_t y ω_s son independientes, y ν_t y ω_s son independientes.

Definición **3.0.3.** Para cada t, El Modelo Lineal Dinámico esta definido por:

- La ecuación Observacional: $Y_t = \mu_t + \nu_t$, donde $\nu_t \sim N[0, V_t]$
- La ecuación del sistema: $\mu_t = \mu_{t-1} + \omega_t$, donde $\omega_t \sim N[0, W_t]$,
- La Información Inicial: $(\mu_0|D_0) \sim N[m_0, C_0]$, donde la media m_0 es una estimación del nivel y la varianza C_0 una medida de incertidumbre acerca de la media.

La sucesión de errores ν_t y ω_t son independientes. También serán independientes de $(\mu_0|D_0)$.

La tercera componente es la representación probabilística de la creencia e información del predictor sobre el nivel en el tiempo $t = 0$ dada la información en el tiempo, D_0. La media m_0 es una estimación del nivel y la varianza C_0 una medida de incertidumbre sobre la media. De acuerdo a la información establecida, asumimos que D_{t-1} comprende D_0, los valores de las varianzas V_t y W_t para todo t, y los valores de las observaciones $Y_{t-1}, Y_{t-2}, ...Y_1$. Así, la única nueva información disponible en cualquier tiempo es el valor de las series de tiempo observacionales tal que $D_t = \{Y_t, D_t - 1\}$.

TEOREMA **3.0.1.** *Para cada t, tenemos la siguiente predicción de un paso hacia adelante y las distribuciones posteriores*

(a) Posteriori para μ_{t-1}: $(\mu_{t-1}|D_{t-1}) \sim N[m_{t-1}, C_{t-1}]$ para alguna media m_{t-1} y varianza C_{t-1};

(b) Priori para μ_t: $(\mu_t|D_{t-1}) \sim N[m_{t-1}, R_t]$ donde $R_t = C_{t-1} + W_t$;

(c) Predicción 1-paso: $(Y_t|D_{t-1}) \sim N[f_t, Q_t]$, donde $f_t = m_{t-1}$ y $Q_t = R_t + V_t$;

(d) Posteriori para μ_t: $(\mu_t|D_t) \sim N[m_t, C_t]$, con $m_t = m_{t-1} + A_t e_t$ y $C_t = A_t V_t$, donde $A_t = \frac{R_t}{Q_t}$, y $e_t = Y_t - f_t$

Demostración. (a) Nótese que la única información disponible en cualquier tiempo es el valor de la serie de tiempo observacional tal que $D_t = \{Y_t, D_{t-1}\}$ con D_0 describiendo la información inicial.

Así por definición se tiene que la distribución posteriori para $t-1$ está dada por

$$(\mu_{t-1}|D_{t-1}) \sim N[m_{t-1}, C_{t-1}]$$

para alguna media m_{t-1} y varianza C_{t-1}.

(b) Por definición se tiene que $\mu_t = \mu_{t-1} + \omega_t$, donde $(\mu_{t-1}|D_{t-1}) \sim N[m_{t-1}, C_{t-1}]$ y $\omega_t \sim N[0, W_t]$, así se tiene la suma de dos distribuciones normales, por tanto μ_t es normal. Además la media y la varianza se obtiene sumando las medias y las varianzas de μ_{t-1} y ω_t , ya que son variables independientes, esto es;

$$\begin{aligned} E(\mu_t|D_{t-1}) &= E(\mu_{t-1}+\omega_t|D_{t-1}) \\ &= E(\mu_{t-1}|D_{t-1})+E(\omega_t|D_{t-1}) \\ &= m_{t-1}+0 \\ &= m_{t-1} \end{aligned}$$

y

$$\begin{aligned} V(\mu_t|D_{t-1}) &= V(\mu_{t-1}+\omega_t|D_{t-1}) \\ &= V(\mu_{t-1}|D_{t-1})+V(\omega_t|D_{t-1}) \\ &= C_{t-1}+W_t\ . \end{aligned}$$

Así,

$$(\mu_t|D_{t-1}) \sim N[m_{t-1}, C_{t-1}+W_t].$$

Luego definiendo $R_t = C_{t-1}+W_t$, se tiene que la distribución a priori para μ_t está dada por

$$(\mu_t|D_{t-1}) \sim N[m_{t-1}, R_t].$$

(c) Procediendo similarmente como en la parte anterior y usando la condicional sobre D_{t-1}, $Y_t = \mu_t + \nu_t$, es la suma de dos normales independientes, por tanto Y_t es normal. La media y la varianza se obtiene sumando las medias y las varianzas de μ_t y ν_t, Luego

$$(Y_t|D_{t-1}) \sim N[m_{t-1}, R_t+V_t]$$

haciendo $f_t = m_{t-1}$ y $Q_t = R_t + V_t$, se tiene

$$(Y_t|D_{t-1}) \sim N[f_t, Q_t].$$

Otra forma de obtener la media y la varianza es:

$$\begin{aligned} f_t &= E[E(Y_t|\mu_t, D_{t-1})] \\ &= E(\mu_t|D_{t-1}) \\ &= m_{t-1} \end{aligned}$$

$$\begin{aligned} Q_t &= V[E(Y_t|\mu_t, D_{t-1})] + E[V(Y_t|\mu_t, D_{t-1})] \\ &= V(\mu_t|D_{t-1}) + V_t \\ &= R_t + V_t \ . \end{aligned}$$

Por tanto la predicción de un paso es

$$(Y_t|D_{t-1}) \sim N[f_t, Q_t]$$

donde $f_t = m_{t-1}$ y $Q_t = R_t + V_t$.

(d) Para la demostración utilizaremos dos técnicas:

1- Actualización a través del Teorema de Bayes

El método Bayesiano es, por supuesto, general, aplicado a todos los modelos sin importar que distribuciones están involucradas. En este caso, el modelo muestral proviene de la distribución observacional con densidad

$$p(Y_t|\mu_t, D_{t-1}) = \frac{\exp\left[\frac{-(Y_t-\mu_t)^2}{(2V_t)}\right]}{(2\pi V_t)^{\frac{1}{2}}}$$

La priori para μ_t dado D_{t-1}, usando (b) , tiene densidad

$$p(\mu_t|D_{t-1}) = \frac{\exp\left[\frac{-(\mu_t-m_{t-1})^2}{(2R_t)}\right]}{(2\pi R_t)^{\frac{1}{2}}}.$$

Observando Y_t, la verosimilitud para μ_t es obtenida proporcional a la densidad observacional vista como una función de μ_t, y, directamente del Teorema de Bayes, la posteriori para μ_t está simplemente dada por

$$\begin{aligned} p(\mu_t|D_t) &= p(\mu_t|D_{t-1}, Y_t) \\ &= \frac{p(\mu_t|D_{t-1})p(Y_t|\mu_t, D_{t-1})}{p(Y_t|D_{t-1})} \\ &= \frac{\frac{\exp\left[\frac{-(\mu_t-m_{t-1})^2}{(2R_t)}\right]}{(2\pi R_t)^{\frac{1}{2}}} \times \frac{\exp\left[\frac{-(Y_t-\mu_t)^2}{(2V_t)}\right]}{(2\pi V_t)^{\frac{1}{2}}}}{\frac{\exp\left[\frac{-(Y_t-f_t)^2}{(2Q_t)}\right]}{(2\pi Q_t)^{\frac{1}{2}}}} \end{aligned}$$

donde el denominador es la densidad de predicción de un paso hacia delante evaluado en la observación actual. Esta posteriori es de una forma Bayesiana estándar derivada de un modelo normal observacional con una priori normal, y el resultado es bien conocido. El cálculo es simplificado concentrandose en la forma de una función que depende sólo de μ_t, ignorando los factores multiplicativos que dependen de Y_t y otras constantes. Haciendo esto nos da la forma proporcional de Teorema de Bayes

$$\begin{aligned}
p(\mu_t|D_t) &\propto p(\mu_t|D_{t-1})p(Y_t|\mu_t, D_{t-1}) \\
&= \frac{\exp\left[\frac{-(Y_t-\mu_t)^2}{(2V_t)} - \frac{(\mu_t-m_{t-1})^2}{(2R_t)}\right]}{(2\pi V_t)^{\frac{1}{2}}(2\pi R_t)^{\frac{1}{2}}} \\
&\propto \exp\left[\frac{-(Y_t-\mu_t)^2}{(2V_t)} - \frac{(\mu_t-m_{t-1})^2}{(2R_t)}\right].
\end{aligned}$$

Como en la mayoría de las aplicaciones del teorema de Bayes, el logaritmo natural proporciona una expresión aditiva más simple

$$\ln[p(\mu_t|D_t)] = K_1 + \ln\left[\exp\left(\frac{-(Y_t-\mu_t)^2}{2V_t} - \frac{(\mu_t-m_{t-1})^2}{2R_t}\right)\right]$$

donde $K_1 = \frac{(Y_t-f_t)^2}{(2Q_t)} + \ln(2\pi Q_t)^{\frac{1}{2}}$, luego,

$$\begin{aligned}
\ln[p(\mu_t|D_t)] &= K_1 - \left[\frac{(Y_t-\mu_t)^2}{2V_t} + \frac{(\mu_t-m_{t-1})^2}{2R_t}\right] \\
&= K_1 - 0{,}5\left[\frac{(Y_t-\mu_t)^2}{V_t} + \frac{(\mu_t-m_{t-1})^2}{R_t}\right] \\
&= K_1 - 0{,}5\left[\frac{(Y_t^2 - 2Y_t\mu_t + \mu_t^2)R_t + (\mu_t^2 - 2\mu_t m_{t-1} + m_{t-1}^2)V_t}{V_tR_t}\right] \\
&= K_1 - 0{,}5\left[\frac{Y_t^2R_t - 2\mu_tY_tR_t + \mu_t^2R_t + \mu_t^2V_t - 2\mu_t m_{t-1}V_t + m_{t-1}^2V_t}{V_tR_t}\right] \\
&= K_1 - 0{,}5\left[\frac{\mu_t^2(R_t+V_t) - 2\mu_t(Y_tR_t + m_{t-1}V_t) + Y_t^2R_t + m_{t-1}^2V_t}{V_tR_t}\right]
\end{aligned}$$

haciendo $Q_t = R_t + V_t$ se tiene

$$\begin{aligned}
\ln[p(\mu_t|D_t)] &= K_1 - 0{,}5\left[\frac{\mu_t^2Q_t}{V_tR_t} - 2\mu_t\left(\frac{Y_tR_t + m_{t-1}V_t}{V_tR_t}\right) + \frac{Y_t^2}{V_t} + \frac{\mu_{t-1}^2}{R_t}\right] \\
&= K_1 - 0{,}5\left[\frac{\mu_t^2}{(\frac{V_tR_t}{Q_t})} - 2\mu_t\left[\frac{Y_tR_t + m_{t-1}V_t}{Q_t(\frac{V_tR_t}{Q_t})}\right] + \frac{Y_t^2}{V_t} + \frac{m_{t-1}^2}{R_t}\right]
\end{aligned}$$

haciendo $C_t = \frac{V_t R_t}{Q_t}$ y $A_t = \frac{R_t}{Q_t}$ se tiene

$$\begin{aligned}
\ln[p(\mu_t|D_t)] &= K_1 - 0{,}5\left[\frac{\mu_t^2}{C_t} - 2\frac{\mu_t}{C_t}\left[Y_tA_t + m_{t-1}\frac{V_t}{Q_t}\right] + \frac{Y_t^2}{V_t} + \frac{m_{t-1}^2}{R_t}\right] \\
&= K_1 - 0{,}5\left[\frac{\mu_t^2 - 2\mu_t\left[Y_tA_t + m_{t-1}\frac{V_t}{Q_t}\right]}{C_t} + \frac{Y_t^2}{V_t} + \frac{m_{t-1}^2}{R_t}\right] \\
&= K_1 - 0{,}5\left[\frac{\mu_t^2 - 2\mu_t\left[Y_tA_t + m_{t-1}\frac{V_t}{Q_t}\right] + \left[Y_tA_t + m_{t-1}\frac{V_t}{Q_t}\right]^2 - \left[Y_tA_t + m_{t-1}\frac{V_t}{Q_t}\right]^2}{C_t}\right] \\
&\quad -0{,}5\left[\frac{Y_t^2}{V_t} + \frac{m_{t-1}^2}{R_t}\right] \\
&= K_1 - 0{,}5\left[\frac{\left[\mu_t - \left(Y_tA_t + m_{t-1}\frac{V_t}{Q_t}\right)\right]^2}{C_t} - \frac{\left[Y_tA_t + m_{t-1}\frac{V_t}{Q_t}\right]^2}{C_t} + \frac{Y_t^2}{V_t} + \frac{m_{t-1}^2}{R_t}\right] \\
&= K_1 - 0{,}5\left[\frac{\left[\mu_t - \left(Y_tA_t + m_{t-1}\left(\frac{R_t+V_t-R_t}{Q_t}\right)\right)\right]^2}{C_t} + K_2\right]
\end{aligned}$$

donde $K_2 = -\frac{\left[Y_tA_t + m_{t-1}\frac{V_t}{Q_t}\right]^2}{C_t} + \frac{Y_t^2}{V_t} + \frac{m_{t-1}^2}{R_t}$

$$\begin{aligned}
\ln[p(\mu_t|D_t)] &= K_1 - 0{,}5\left[\frac{\left[\mu_t - \left(Y_tA_t + m_{t-1}\left(\frac{Q_t-R_t}{Q_t}\right)\right)\right]^2}{C_t} + K_2\right] \\
&= K_1 - 0{,}5\left[\frac{\left[\mu_t - \left(Y_tA_t + m_{t-1}\left(1 - \frac{R_t}{Q_t}\right)\right)\right]^2}{C_t} + K_2\right] \\
&= K_1 - 0{,}5\left[\frac{\left[\mu_t - (Y_tA_t + m_{t-1}(1 - A_t))\right]^2}{C_t} + K_2\right] \\
&= K_1 - 0{,}5\left[\frac{\left[\mu_t - (Y_tA_t + m_{t-1} - m_{t-1}A_t)\right]^2}{C_t} + K_2\right] \\
&= K_1 - 0{,}5\left[\frac{\left[\mu_t - (m_{t-1} + A_t\left\{Y_t - m_{t-1}\right\})\right]^2}{C_t} + K_2\right]
\end{aligned}$$

haciendo $e_t = Y_t - m_{t-1}$

$$\begin{aligned}\ln[p(\mu_t|D_t)] &= K_1 - 0{,}5\left[\frac{[\mu_t - (m_{t-1} + A_t e_t)]^2}{C_t} + K_2\right] \\ &= K - 0{,}5\left[\frac{[\mu_t - m_t]^2}{C_t}\right]\end{aligned}$$

donde $m_= m_{t-1} + A_t e_t$, y $K = K_1 - (0,5)K_2$.

Así,

$$p(\mu_t|D_t) \propto \exp\left[\frac{-(\mu_t - m_t)^2}{2C_t}\right]$$

y la renormalización de esta densidad conduce a una constante de normalización

$$\frac{1}{[2\pi C_t]^{\frac{1}{2}}}$$

2- Prueba basada en la Teoría Normal Estándar

Dentro de este caso normal, especial, de estructura lineal, una prueba más específica basada en la distribución normal bivariada se deriva de la siguiente manera:

(i) Se calcula la distribución conjunta de Y_t y μ_t dado D_{t-1};

(ii) Se deduce directamente la distribución condicional para μ_t dado Y_t.

Cualquier función lineal de Y_t y μ_t será una combinación lineal de las cantidades normales independientes ν_t, ω_t y μ_{t-1}, y así, la condicional sobre D_{t-1}, será distribuida normalmente. Entonces por definición, $(Y_t, \mu_t|D_{t-1})$ es normal bivariada. Para identificar el vector de media y la matriz de varianza, notemos que tenemos las distribuciones marginales en (b) y (c), $(Y_t|D_{t-1}) \sim N[f_t, Q_t]$ y $(\mu_t|D_{t-1}) \sim N[m_{t-1}, R_t]$ y la covarianza es simplemente

$$\begin{aligned}C[Y_t, \mu_t|D_{t-1}] &= C[\mu_t + \nu_t, \mu_t|D_{t-1}] \quad ya\ \ que\ \ Y_t = \mu_t + \nu_t \\ &= V[\mu_t|D_{t-1}] \\ &= R_t,\end{aligned}$$

por aditividad y usando la independencia de μ_t y ν_t. Luego, la matriz de varianza y covarianza de (Y_t, μ_t) está dada por

$$\begin{aligned} C &= \begin{bmatrix} cov(Y_t, Y_t) & cov(Y_t, \mu_t) \\ cov(\mu_t, Y_t) & cov(\mu_t, \mu_t) \end{bmatrix} \\ &= \begin{bmatrix} var(Y_t) & cov(Y_t, \mu_t) \\ cov(Y_t, \mu_t) & var(\mu_t) \end{bmatrix} \\ &= \begin{pmatrix} Q_t & R_t \\ R_t & R_t \end{pmatrix} \end{aligned}$$

Así la distribución conjunta está dada por

$$\left(\begin{array}{c|c} Y_t & D_{t-1} \\ \mu_t & \end{array} \right) \sim N \left[\begin{pmatrix} m_{t-1} \\ m_{t-1} \end{pmatrix}, \begin{pmatrix} Q_t & R_t \\ R_t & R_t \end{pmatrix} \right].$$

Apliquemos la distribución normal bivariada, donde el coeficiente de correlación es $\rho_t = \frac{R_t}{\sqrt{Q_t R_t}} > 0$ y $\rho_t^2 = \left(\frac{R_t}{\sqrt{Q_t R_t}}\right)^2 = \frac{R_t^2}{Q_t R_t} = \frac{R_t}{Q_t} = A_t$. Así la normal bivariada esta dada por:

$p(Y_t, \mu_t) = \frac{1}{2\pi\sqrt{Q_t R_t (1-A_t)}} \times \exp\left[\left(\frac{Y_t - m_{t-1}}{\sqrt{Q_t}}\right)^2 + \left(\frac{\mu_t - m_{t-1}}{\sqrt{R_t}}\right)^2 - 2A_t \left(\frac{Y_t - m_{t-1}}{\sqrt{Q_t}}\right)\left(\frac{\mu_t - m_{t-1}}{\sqrt{R_t}}\right)\right]$
y la distribución condicional de $(\mu_t | Y_t, D_{t-1})$ es normal con media

$$m_t = m_{t-1} + A_t(Y_t - m_{t-1})$$

y varianza

$$\begin{aligned} (1-\rho^2)R_t &= (1-A_t)R_t \\ &= \left(1 - \frac{R_t}{Q_t}\right) R_t \\ &= \left(\frac{(Q_t - R_t)}{Q_t}\right) R_t \\ &= \frac{R_t V_t}{Q_t} \\ &= A_t V_t \\ &= C_t \ . \end{aligned}$$

□

Alguna discusión de varios elementos de las distribuciones está en orden. e_t es el error de predicción un paso hacía adelante, la diferencia entre el valor observado

Y_t y la esperanza f_t. A_t es el coeficiente de regresión a priori de μ_t sobre Y_t y, en este caso particular, es el cuadrado de sus coeficientes de correlación; claramente $0 \leq A_t \leq 1$. Los resultados son computacionalmente simples y elegantes debido al uso de la distribución normal para cada componente del modelo. Usando estos resultados, la secuencia de actualización y revisión de predicción es directa. Cabe señalar que una alternativa para la representación de m_t es

$$m_t = A_t Y_t + (1 - A_t) m_{t-1} \, ,$$

mostrando que m_t es una media ponderada de la estimación a priori del nivel m_{t-1} y la observación sobre el nivel Y_t. El valor, o coeficiente de adaptación, A_t, definiendo esta combinación entre 0 y 1, estando más cerca de 0 cuando la distribución a priori es más concentrada que la verosimilitud con $R_t < V_t$, y cerca a 1 cuando la priori es más difusa, o menos informativa, que la verosimilitud. Además, la posteriori es menos difusa que la priori ya que $C_t < R_t$, representando un crecimiento en la información sobre μ_t debido a la observación adicional Y_t.

§3.1. Distribuciones de Predicción

Para el tiempo t, las dos distribuciones principales requeridas por el predictor son las marginales de $(Y_{t+k}|D_t)$ y $(X_t(k)|D_t)$ donde $X_t(k) = Y_{t+1} + Y_{t+2} + ... + Y_{t+k}$ para $k > 0$. La forma es de la predicción k pasos hacia adelante. Estas distribuciones están dadas en el siguiente teorema.

Teorema **3.1.1.** *Para $k > 0$, las siguientes distribuciones existen:*

(a) k pasos hacia adelante $(Y_{t+k}|D_t) \sim N[m_t, Q_t(k)]$

(b) k pasos de tiempo adelantado $(X_{t+k}|D_t) \sim N[km_t, L_t(k)]$
donde

$$Q_t(k) = C_t + \sum_{j=1}^{k} W_{t+j} + V_{t+k}$$

y

$$L_t = k^2 C_t + \sum_{j=1}^{k} V_{t+j} + \sum_{j=1}^{k} j^2 W_{t+k+1-j} \ .$$

Demostración. (a) Por definición de la ecuación de evolución de μ_t, $\mu_t = \mu_{t-1} + \omega_t$, para cada t, con $\omega_t \sim N[0, W_t]$, tenemos:
Para $t+1$,

$$\mu_{t+1} = \mu_t + \omega_{t+1}, \;\; donde \;\; \omega_{t+1} \sim N[0, W_{t+1}]$$

Para $t+2$,

$$\begin{aligned}
\mu_{t+2} &= \mu_{(t+2)-1} + \omega_{t+2}, \quad donde \;\; \omega_{t+2} \sim N[0, W_{t+2}] \\
&= \mu_{t+1} + \omega_{t+2} \\
&= \mu_t + \omega_{t+1} + \omega_{t+2} \\
&= \mu_t + \sum_{j=1}^{2} \omega_{t+j}, \;\; donde \; \sum_{j=1}^{2} \omega_{t+j} \sim N[0, W_{t+1} + W_{t+2}]
\end{aligned}$$

Para $t+3$,

$$\begin{aligned}
\mu_{t+3} &= \mu_{t+2} + \omega_{t+3}, \quad donde \; \omega_{t+3} \sim N[0, W_{t+3}] \\
&= \mu_t + \sum_{j=1}^{2} \omega_{t+j} + \omega_{t+3} \\
&= \mu_t + \sum_{j=1}^{3} \omega_{t+j}, \; donde \;\; \sum_{j=1}^{3} \omega_{t+j} \sim N\left[0, \sum_{j=1}^{3} W_{t+j}\right] \; .
\end{aligned}$$

Así para cada $k \geq 1$, se tiene

$$\mu_{t+k} = \mu_t + \sum_{j=1}^{k} \omega_{t+j} \;\; .$$

Ahora por definición de la ecuación de observación tenemos $Y_t = \mu_t + \nu_t$, donde $\nu_t \sim N[0, V_t]$ y $\omega_t \sim N[0, W_t]$
Para $t+1$

$$\begin{aligned}
Y_{t+1} &= \mu_{t+1} + \nu_{t+1} \\
&= \mu_t + \omega_{t+1} + \nu_{t+1}
\end{aligned}$$

Para $t+2$

$$\begin{aligned}
Y_{t+2} &= \mu_{t+2} + \nu_{t+2} \\
&= \mu_t + \sum_{j=1}^{2} \omega_{t+j} + \nu_{t+2}
\end{aligned}$$

.

.

.

Para $t+k$

$$Y_{t+k} = \mu_t + \sum_{j=1}^{k} \omega_{t+j} + \nu_{t+k}$$

ya que todos los términos son normal y mutuamente independientes, $(Y_{t+k}|D_t)$ es normal y la media y la varianza se obtienen directamente. Es decir,

$$\begin{aligned} V(Y_{t+k}|D_t) &= V(\mu_t|D_t) + V\left(\sum_{j=1}^{k} \omega_{t+j}|D_t\right) + V(\nu_{t+k}|D_t) \\ &= C_t + \sum_{j=1}^{k} W_{t+j} + V_{t+k} \end{aligned}$$

y

$$\begin{aligned} E(Y_{t+k}|D_t) &= E(\mu_t|D_t) \\ &= m_t. \end{aligned}$$

Por tanto

$$Q_t(k) = C_t + \sum_{j=1}^{k} W_{t+j} + V_{t+k}\ .$$

(b) Sabemos que $X_t(k) = Y_{t+1} + Y_{t+2} + ... + Y_{t+k}$, para $k > 0$, así tenemos que

$$\begin{aligned} X_t(k) &= (\mu_t + \omega_{t+1} + \nu_{t+1}) + ... + (\mu_t + \omega_{t+1} + \omega_{t+2} + ... + \\ &\quad +\omega_{t+(k-1)} + \nu_{t+(k-1)}) + (\mu_t + \omega_{t+1} + ... + \omega_{t+k} + \nu_{t+k}) \\ &= \sum_{j=1}^{k} \mu_t + (\omega_{t+1} + \sum_{j=1}^{2} \omega_{t+j} + ... + \sum_{j=1}^{k} \omega_{t+j}) + (\nu_{t+1} + \nu_{t+2} + ... + \nu_{t+k}) \\ &= k\mu_t + (\omega_{t+1} + \sum_{j=1}^{2} \omega_{t+j} + ... + \sum_{j=1}^{k} \omega_{t+j}) + \sum_{j=1}^{k} \nu_{t+j}\ . \end{aligned}$$

Por otro lado

$$
\begin{aligned}
(\omega_{t+1}+\sum_{j=1}^{2}\omega_{t+j}+...+\sum_{j=1}^{k-1}\omega_{t+j}+\sum_{j=1}^{k}\omega_{t+j}) &= \omega_{t+1}+(\omega_{t+1}+\omega_{t+2})+...+(\omega_{t+1}+\omega_{t+2}+ \\
&\quad +...+\omega_{t+(k-1)})+(\omega_{t+1}+\omega_{t+2}+...+ \\
&\quad +\omega_{t+(k-1)}+\omega_{t+k}) \\
&= k\omega_{t+1}+(k-1)\omega_{t+2}+(k-2)\omega_{t+3}+...+ \\
&\quad +(k-(k-2))\omega_{t+(k-1)}+(k-(k-1))\omega_{t+k} \\
&= k\omega_{t+1}+(k-1)\omega_{t+2}+(k-2)\omega_{t+3}+...+ \\
&\quad +2\omega_{t+(k-1)}+\omega_{t+k}\,.
\end{aligned}
\tag{3.1}
$$

Además

$$
\begin{aligned}
\sum_{j=1}^{k} j\omega_{t+k+1-j} &= \omega_{t+k}+2\omega_{t+(k-1)}+3\omega_{t+(k-2)}+...+(k-1)\omega_{t+k+1-(k-1)}+k\omega_{t+k+1-k} \\
&= \omega_{t+k}+2\omega_{t+(k-1)}+3\omega_{t+(k-2)}+...+(k-1)\omega_{t+2}+k\omega_{t+1}\;\;.
\end{aligned}
\tag{3.2}
$$

Igualando (3.1) y (3.2)

$$
\omega_{t+1}+\sum_{j=1}^{2}\omega_{t+j}+...+\sum_{j=1}^{k}\omega_{t+j} \;=\; \sum_{j=1}^{k} j\omega_{t+k+1-j}\,;
$$

obtenemos

$$
X_t(k)=k\mu_t+\sum_{j=1}^{k} j\omega_{t+k+1-j}+\sum_{j=1}^{k}\nu_{t+j}\;\;,
$$

obviamente, $X_t(k)$ es normal ya que todos los términos son normales, y su media es km_t.

Ahora usando la independencia de la estructura de los términos de error, se tiene

$$
\begin{aligned}
V[X_t(k)|D_t] &= V(k\mu_t|D_t)+V(\sum_{j=1}^{k} j\omega_{t+k+1-j}|D_t)+V(\sum_{j=1}^{k}\nu_{t+j}|D_t) \\
&= k^2C_t+\sum_{j=1}^{k} j^2W_{t+k+1-j}+\sum_{j=1}^{k}V_{t+j}
\end{aligned}
$$

Por tanto

$$L_t(k) = k^2 C_t + \sum_{j=1}^{k} j^2 W_{t+k+1-j} + \sum_{j=1}^{k} V_{t+j} \ .$$

$\square$

§3.2. El Modelo Constante

Un caso especial de un modelo en el cual las varianzas observacional y de evolución son constante en el tiempo es llamado modelo constante. Además, cuando el conjunto de información $D_t = \{Y_t, D_{t-1}\}$ para cada $t \geq 1$ no contiene información externa a la serie de tiempo, el modelo es llamado cerrado. El modelo polinomial de primer orden cerrado, constante, es prácticamente usado, aunque con algunas restricciones, caso que es de interés ya que permite derivar resultados límites importantes que dan a conocer sobre la estructura más general de los DLMs, algunas de las cuales están relacionados con modelos clásicos de series de tiempo y métodos de predicción puntual de uso común.

Los modelos lineales dinámicos constantes, cerrados tiene a V y W ambas positivas, finitas y constantes, una priori en el tiempo de origen $t = 0$ dada por $(\mu_0|D_0) \sim N(m_0, C_0)$, y para $t \geq 1$, D_t=$\{Y_t, D_{t-1}\}$. Una constante importante es la razón $r = W/V$, la cual desempeña un papel relacionado con el concepto de la razón señal a ruido en la terminología de ingeniería, midiendo la variación relativa del estado o ecuación del sistema de evolución a la ecuación observacional. Cuando V y W son dados, el modelo siempre puede ser reparametrizado en términos de V y r tomando $W = rV$, y esto es adoptado por el modelo constante.

§3.2.1. Convergencia y Comportamiento Límite

En el modelo cerrado la tasa de adaptación para una nueva data, medida por el coeficiente de adaptación A_t, converge rápidamente a un valor constante. Esto se prueba en el siguiente teorema.

Teorema **3.2.1.** *Sea A_t el coeficiente de adaptación. Si $t \longrightarrow \infty$ entonces $A_t \longrightarrow A$ y $C_t \longrightarrow C = AV$ donde*

$$A = \frac{r\left(\sqrt{1 + (\frac{4}{r})} - 1\right)}{2}$$

Demostración. Tenemos que $C_t = A_t V$ donde $0 < A_t < 1$, así $0 < C_t \leq V$ para todo t. Entonces, usando la recursión $C_t^{-1} = R_t^{-1} + V^{-1}$ y $R_t = C_{t-1} + W$, se tiene que

$$\begin{aligned}
C_t^{-1} - C_{t-1}^{-1} &= R_t^{-1} - R_{t-1}^{-1} \\
&= \frac{1}{C_{t-1} + W} - \frac{1}{C_{t-2} + W} \\
&= \frac{C_{t-2} - C_{t-1} + W - W}{(C_{t-1} + W)(C_{t-2} + W)} \\
&= \frac{C_{t-2} - C_{t-1}}{(C_{t-1} + W)(C_{t-2} + W)} \cdot \frac{C_{t-1}C_{t-2}}{C_{t-1}C_{t-2}} \\
&= \frac{C_{t-1}C_{t-2}}{(C_{t-1} + W)(C_{t-2} + W)} \left(\frac{1}{C_{t-1}} - \frac{1}{C_{t-2}} \right) \\
&= \frac{C_{t-1}C_{t-2}}{R_t R_{t-1}} (C_{t-1}^{-1} - C_{t-2}^{-1}) \\
&= K_t (C_{t-1}^{-1} - C_{t-2}^{-1})
\end{aligned}$$

donde $K_t = \frac{C_{t-1}C_{t-2}}{R_t R_{t-1}} > 0$.

Así C_t es una sucesión monótona y acotada con límite C, y $R_t = C_{t-1} + W$ implica que R_t converge a $R = C + W$. Usando $C_t = \frac{R_t V}{(R_t + V)}$ se obtiene que

$$\begin{aligned}
C = \frac{RV}{(R+V)} &\implies C(R+V) = RV \\
&\implies CR + CV - RV = 0 \\
&\implies C(C+W) + CV - (C+W)V = 0 \\
&\implies C^2 + CW + CV - CV - WV = 0 \\
&\implies C^2 + CW - VW = 0\ .
\end{aligned}$$

Esta ecuación cuadrática tiene una raíz positiva dada por:

$$\begin{aligned} C &= \frac{-W+\sqrt{W^2-4(-VW)}}{2} \\ &= \frac{-W+\sqrt{W^2+4VW}}{2} \\ &= \frac{-W+\sqrt{W^2(1+4\frac{V}{W})}}{2} \\ &= \frac{-W+W\sqrt{(1+4\frac{V}{W})}}{2} \\ &= \frac{W\left(\sqrt{1+4\frac{V}{W}}-1\right)}{2} \\ &= \frac{\frac{W}{V}V\left(\sqrt{1+4\frac{V}{W}}-1\right)}{2}\,. \end{aligned}$$

Haciendo $r=\frac{W}{V}$ se tiene

$$C=\frac{rV\left(\sqrt{1+\frac{4}{r}}-1\right)}{2}\,.$$

Además $A_t=\frac{C_t}{V}$, entonces A_t converge a $A=\frac{C}{V}$ la cual es una función de r

$$A=\frac{\frac{rV(\sqrt{1+\frac{4}{r}}-1)}{2}}{V}=\frac{r\left(\sqrt{1+\frac{4}{r}}-1\right)}{2}\,.$$

□

Resumiendo los resultados anteriores, se deducen algunas consecuencias:

(i) $A_t \longrightarrow A=\frac{r(\sqrt{1+\frac{4}{r}}-1)}{2}$ con $r=\frac{A^2}{(1-A)}$

(ii) $C_t \longrightarrow C=AV$

(iii) $R_t \longrightarrow R=\frac{C}{(1-A)}=\frac{AV}{(1-A)}$

(iv) $Q_t \longrightarrow Q=\frac{V}{(1-A)}$ y $V=(1-A)Q$

(v) $W=A^2Q$.

En efecto,

(i)

$$\begin{aligned}
A &= \frac{r\left(\sqrt{1+\frac{4}{r}}-1\right)}{2}\\
\Longrightarrow 4A^2 &= r^2\left(\sqrt{1+\frac{4}{r}}-1\right)^2\\
\Longrightarrow 4A^2 &= r^2\left(1+\frac{4}{r}-2\sqrt{1+\frac{4}{r}}+1\right)\\
\Longrightarrow 4A^2 &= r^2+4r-2r^2\sqrt{1+\frac{4}{r}}+r^2\\
\Longrightarrow 4A^2 &= 2r^2+4r-2r^2\sqrt{1+\frac{4}{r}}\\
\Longrightarrow 4A^2 &= \left(2-2\sqrt{1+\frac{4}{r}}\right)r^2+4r
\end{aligned}$$

$$\Longrightarrow 2\left(1-\sqrt{1+\frac{4}{r}}\right)r^2+4r-4A^2=0$$

$$\Longrightarrow -4Ar+4r-4A^2=0$$

$$\Longrightarrow (-4A+4)r=4A^2$$

$$\begin{aligned}
\Longrightarrow r &= \frac{4A^2}{(-4A+4)}\\
\Longrightarrow r &= \frac{A^2}{(1-A)}\ .
\end{aligned}$$

(ii) Como $C_t = A_tV$ y sabemos que $A_t \longrightarrow A$, por propiedades de límite se tiene que $C_t \longrightarrow AV = C$.

(iii) Sabemos que $R_t = C_{t-1} + W$ y tiende a $R = C + W$ ya que $C_t \longrightarrow C$ cuando

$t \longrightarrow \infty$, así

$$\begin{aligned} R &= C + W \\ &= C + rV \\ &= C + \frac{A^2}{(1-A)}V \\ &= \frac{C - CA + A^2V}{(1-A)} \\ &= \frac{C - CA + CA}{(1-A)} \\ &= \frac{C}{(1-A)} \\ &= \frac{AV}{(1-A)} \ . \end{aligned}$$

(iv) $Q_t = \frac{R_t}{A_t}$, $R_t \longrightarrow R$, $A_t \longrightarrow A$, por tanto

$$\begin{aligned} Q_t \longrightarrow Q &= \frac{R}{A} \\ &= \frac{\frac{AV}{(1-A)}}{A} \\ &= \frac{V}{(1-A)} \end{aligned}$$

luego, $Q = \frac{V}{(1-A)} \Longrightarrow V = (1-A)Q$.

(v)

$$\begin{aligned} R = C + W \Longrightarrow W &= R - C \\ &= \frac{AV}{(1-A)} - AV \\ &= \frac{AV - AV(1-A)}{(1-A)} \\ &= \frac{AV - AV + A^2V}{(1-A)} \\ &= \frac{A^2V}{(1-A)} \\ &= A^2Q \ . \end{aligned}$$

Sobre la forma de convergencia, C_t es monótona creciente/decreciente dependiendo si la varianza inicial C_0 es menor/mayor que el valor límite C. Similar se aplica para la sucesión A_t.

Para obtener una mayor comprensión sobre el comportamiento de esta última sucesión, una expresión exacta en términos del valor límite A y el valor inicial A_1 será obtenida. Definamos $\delta = 1 - A$, claramente $0 < \delta < 1$ ya que $0 < A < 1$ y para t grande, $m_t \approx \delta Y_t + (1-\delta)m_{t-1}$. Dado C_0, el coeficiente de adaptación inicial es $A_1 = \frac{(C_0+W)}{(C_0+W+V)}$. En el modelo constante, cerrado se tiene $C_t = A_t V$, $R_t = A_{t-1}V + W$ y $Q_t = R_t + V$, esto implica que

$$\begin{aligned} A_{t+1} &= \frac{R_{t+1}}{Q_{t+1}} \\ &= \frac{A_t V + W}{R_{t+1} + V} \\ &= \frac{C_t + W}{A_t V + W + V} \\ &= \frac{C_t + W}{C_t + W + V} \\ &= \frac{\frac{C_t+W}{V}}{\frac{C_t+W+V}{V}} \\ &= \frac{A_t + r}{A_t + r + 1} \end{aligned}$$

donde $r = \frac{W}{V}$.

Tomando límite se tiene $A = \frac{A+r}{A+r+1}$ ya que $A_t \longrightarrow A$ cuando t tiende a infinito. Ahora definiendo

$$u_t = \frac{1}{A_t - A} \tag{3.3}$$

para cada t, por sustracción se tiene

$$\begin{aligned} u_{t+1} &= \frac{1}{A_{t+1} - A} \\ &= \frac{1}{\frac{A_t+r}{A_t+r+1} - \frac{A+r}{A+r+1}} \\ &= \frac{1}{\frac{(A_t+r)(A+r+1)-(A+r)(A_t+r+1)}{(A_t+r+1)(A+r+1)}} \\ &= \frac{(A_t+r+1)(A+r+1)}{A_t - A} \\ &= u_t(A_t+r+1)(A+r+1) \ . \end{aligned}$$

Ahora, ya que $\delta = 1 - A$ y $r = \frac{A^2}{(1-A)}$ se tiene que $A+r+1 = \frac{1}{\delta}$ y que $A_t+r+1 = \frac{1}{u_t}+\frac{1}{\delta}$. En efecto;

$$\begin{aligned}
A + r + 1 &= \left(A + \frac{A^2}{(1-A)}\right) + 1 \\
&= \frac{A - A^2 + A^2}{(1-A)} + 1 \\
&= \frac{A}{(1-A)} + 1 \\
&= \frac{A + 1 - A}{(1-A)} \\
&= \frac{1}{\delta}
\end{aligned}$$

y

$$\begin{aligned}
A_t + r + 1 &= \left(\frac{1}{u_t} + A\right) + \frac{A^2}{(1-A)} + 1 \\
&= \frac{1}{u_t} + \left(A + \frac{A^2}{(1-A)}\right) + 1 \\
&= \frac{1}{u_t} + \frac{A - A^2 + A^2}{(1-A)} + 1 \\
&= \frac{1}{u_t} + \left(\frac{A}{(1-A)} + 1\right) \\
&= \frac{1}{u_t} + \frac{A + 1 - A}{1 - A} \\
&= \frac{1}{u_t} + \frac{1}{1 - A} \\
&= \frac{1}{u_t} + \frac{1}{\delta} \ .
\end{aligned}$$

Así que

$$\begin{aligned}
\delta^2 u_{t+1} &= \delta^2 u_t (A_t + r + 1)(A + r + 1) \\
&= \delta^2 u_t \left[(\frac{1}{u_t} + \frac{1}{\delta})(\frac{1}{\delta}) \right] \\
&= \delta^2 u_t \left[\frac{1}{u_t \delta} + \frac{1}{\delta^2} \right] \\
&= \delta + u_t
\end{aligned}$$

luego $u_t = \delta^2 u_{t+1} - \delta$, esto implica que

$$\begin{aligned}
u_1 &= \delta^2 u_2 - \delta \\
&= \delta^2[\delta^2 u_3 - \delta] - \delta \\
&= \delta^4 u_3 - \delta^3 - \delta \\
&= \delta^4[\delta^2 u_4 - \delta] - \delta^3 - \delta \\
&= \delta^6 u_4 - \delta^5 - \delta^3 - \delta \\
&= \delta^6[\delta^2 u_5 - \delta] - \delta^5 - \delta^3 - \delta \\
&= \delta^8 u_5 - \delta^7 - \delta^5 - \delta^3 - \delta \\
&. \\
&. \\
&.
\end{aligned}$$

en general

$$\begin{aligned}
u_1 &= \delta^{2t-2} u_t - \sum_{i=1}^{t-1} \delta^{2i-1}, \quad para\ t = 2... \\
&= \delta^{2t-2} u_t - \frac{1}{\delta} \sum_{i=1}^{t-1} \delta^{2i}, \ \ 0 < \delta < 1 \\
&= \delta^{2t-2} u_t - \frac{1}{\delta} \sum_{i=1}^{t-1} (\delta^2)^i \\
&= \delta^{2t-2} u_t - \frac{1}{\delta} \left(\frac{1 - \delta^{2(t-1)}}{1 - \delta^2} \delta^2 \right) \\
&= \delta^{2t-2} u_t - \frac{\delta(1 - \delta^{2(t-1)})}{1 - \delta^2}
\end{aligned}$$

así

$$\begin{aligned}
\delta^{2t-2} u_t &= u_1 + \frac{\delta(1 - \delta^{2(t-1)})}{1 - \delta^2} \\
&= \frac{u_1(1 - \delta^2) + \delta(1 - \delta^{2(t-1)})}{1 - \delta^2}
\end{aligned}$$

por tanto

$$u_t = \frac{u_1(1 - \delta^2) + \delta(1 - \delta^{2(t-1)})}{(1 - \delta^2)\delta^{2t-2}} \tag{3.4}$$

ahora igualando (3.3) y (3.4) se tiene

$$\begin{aligned}\frac{1}{A_t - A} &= \frac{u_1(1-\delta^2)+\delta(1-\delta^{2(t-1)})}{(1-\delta^2)\delta^{2t-2}}\\ \Longrightarrow A_t - A &= \frac{(1-\delta^2)\delta^{2(t-1)}}{u_1(1-\delta^2)+\delta(1-\delta^{2(t-1)})}\ .\end{aligned}$$

Notemos que $1-\delta^2 = A(1+\delta)$ y sustituyendo $u_1 = \frac{1}{(A_1 - A)}$, se tiene que

$$\begin{aligned}A_t - A &= \frac{A(1+\delta)\delta^{2(t-1)}}{\frac{1}{(A_1-A)}[A(1+\delta)]+\delta(1-\delta^{2(t-1)})}\\ &= \frac{[A(1+\delta)\delta^{2(t-1)}](A_1-A)}{A(1+\delta)+(A_1-A)(\delta(1-\delta^{2t-2}))}\\ &= \frac{A(1+\delta)(A_1-A)\delta^{2(t-1)}}{[A(1+\delta)+A_1(\delta-\delta^{2t-1})-A(\delta-\delta^{2t-1}))]}\\ &= \frac{A(1+\delta)(A_1-A)\delta^{2(t-1)}}{[A+A\delta+A_1(\delta-\delta^{2t-1})-A\delta+A\delta^{2t-1}]}\\ &= \frac{A(1+\delta)(A_1-A)\delta^{2(t-1)}}{[A_1(\delta-\delta^{2t-1})+A(1+\delta^{2t-1})]}\end{aligned}$$

por tanto

$$A_t - A = \frac{A(A_1-A)(1+\delta)\delta^{2(t-1)}}{[(1+\delta^{2t-1})A+(\delta-\delta^{2t-1})A_1]}\ .$$

Esto a su vez implica que,

$$\begin{aligned}A_t &= \frac{A(A_1-A)(1+\delta)\delta^{2(t-1)}}{[(1+\delta^{2t-1})A+(\delta-\delta^{2t-1})A_1]}+A\\ &= \frac{A(A_1-A)(1+\delta)\delta^{2(t-1)}+A[(1+\delta^{2t-1})A+(\delta-\delta^{2t-1})A_1]}{(1+\delta^{2t-1})A+(\delta-\delta^{2t-1})A_1}\\ &= A\left[\frac{A_1\delta^{2(t-1)}-A\delta^{2(t-1)}+A_1\delta^{2t-1}-A\delta^{2t-1}+A+A\delta^{2t-1}+A_1\delta-A_1\delta^{2t-1}}{(1+\delta^{2t-1})A+(\delta-\delta^{2t-1})A_1}\right]\\ &= A\left[\frac{(1-\delta^{2t-2})A+(\delta+\delta^{2t-2})A_1}{(1+\delta^{2t-1})A+(\delta-\delta^{2t-1})A_1}\right]\end{aligned}$$

luego,

$$A_t = A\left[\frac{(1-\delta^{2t-2})A+(\delta+\delta^{2t-2})A_1}{(1+\delta^{2t-1})A+(\delta-\delta^{2t-1})A_1}\right],$$

de lo que podemos deducir lo siguiente:

a) En el caso de una priori inicial muy vaga en la que C_0^{-1} es cercana a cero, $A_1 \approx 1$, y A_t es monótona decreciente con

$$\begin{aligned} A_t &\approx A\left[\frac{(1-\delta^{2t-2})(1-\delta)+(\delta+\delta^{2t-2})}{(1+\delta^{2t-1})(1-\delta)+(\delta-\delta^{2t-1})}\right] \\ &= A\left[\frac{1-\delta-\delta^{2t-2}+\delta^{2t-1}+\delta+\delta^{2t-2}}{1-\delta+\delta^{2t-1}-\delta^{2t}+\delta-\delta^{2t-1}}\right] \\ &= \frac{A(1+\delta^{2t-1})}{(1-\delta^{2t}} \end{aligned}$$

Por tanto

$$A_t \approx \frac{A(1+\delta^{2t-1})}{(1-\delta^{2t})}$$

o, escribiendo $\delta = 1 - A_t$ se tiene

$$\begin{aligned} 1-\delta_t &\approx \frac{(1-\delta)(1+\delta^{2t-1})}{(1-\delta^{2t})} \\ &= \frac{1+\delta^{2t-1}-\delta-\delta^{2t}}{1-\delta^{2t}} \\ \Longrightarrow \delta_t &\approx 1-\frac{(1+\delta^{2t-1}-\delta-\delta^{2t})}{(1-\delta^{2t})} \\ &= \frac{1-\delta^{2t}-1-\delta^{2t-1}+\delta+\delta^{2t}}{(1-\delta^{2t})} \\ &= \frac{\delta(1-\delta^{2t-2})}{(1-\delta^{2t})} \end{aligned}$$

Por tanto

$$\delta_t \approx \frac{\delta(1-\delta^{2t-2})}{(1-\delta^{2t})}$$

b) En el otro extremo cuando la priori inicial es muy precisa con C_0 próximo a cero, entonces $A_1 \approx 0$ y A_t es monótona creciente con

$$A_t \approx \frac{A(1-\delta^{2t-2})}{(1+\delta^{2t-1})}$$

ó

$$\delta_t \approx \frac{\delta(1+\delta^{2t-3})}{(1+\delta^{2t-1})}$$

c) En general, la convergencia es extremadamente rápida, siendo monótonamente creciente para $A_1 < A$, y monótonamente decreciente para $A_1 > A$.

Ejemplo **3.2.1.** Una compañia de mercado farmacéutico de droga llamada KURIT actualmente vende un promedio de 100 unidades por mes. El consejo médico conduce a un cambio en la formulación de la droga que se espera lleve a una mayor demanda del mercado para el producto. Se ha acordado que a partir de enero, $t = 1$, la nueva formulación con nuevos envases sustituirá al actual del producto, pero el precio y la marca sigue siendo KURIT sin cambio. Con el fin de planificar la producción, la existencia de materias primas y suministros, a corto plazo, es necesario pronósticar demandas futuras. El medicamento es utilizado regularmente por los pacientes a fin de que la demanda tiende a ser localmente constante en el tiempo, así, un modelo polinomial de primer orden es asumido para el total de las ventas de KURIT por mes. Las fluctuaciones de las ventas y la variación observacional acerca del nivel se espera varíen mes a mes en el nivel de demanda, así que ω_t es pequeño comparado con ν_t. De acuerdo con esto, el Modelo Lineal Dinámico que funcionó con éxito en la formulación vieja fué el modelo constante, con $V = 100$ y $W = 5$, y estos valores se toman para la nueva formulación. En diciembre, $t = 0$, desde un punto de vista del mercado para el nuevo producto es que la demanda aumente probablemente un 30 % para 130 unidades por mes. Se cree que la demanda es poco probable que se haya reducido más de 10 unidades o haya crecido más de 70. Este rango de 80 unidades es tomado como representación para μ_0 con 4 desviaciones estándar, así, el punto de vista inicial de la compañia a priori se describe por $m_0 = 130$ y $C_0 = 400$. De modo que

$$(\mu_0|D_0) \sim N[130, 400] \ .$$

Por consiguiente el modelo operacional de las ventas Y_t en el mes t es:

$$Y_t = \mu_t + \nu_t \, , \ \ \nu_t \sim N[0, 100] \, ,$$

$$\mu_t = \mu_{t-1} + \omega_t \, , \ \ \omega_t \sim N[0, 5] \, ,$$

$$(\mu_0|D_0) \sim N[130, 400].$$

Aquí $r = W/V = 0{,}05$, una señal baja para el ruido en este tipo de aplicación.

Las observaciones durante los próximos meses y las diversas componentes de la predicción de un paso y la actualización recurrente de las relaciones se dan en la figura (3.1) y la figura (3.2), cifras que proporciona una gráfica de las observaciones y los pronósticos de un paso en el tiempo y la figura (3.3) es una gráfica del coeficiente de adaptación A_t a lo largo del tiempo, hasta septiembre, $t = 10$.

Meses	Distribucion	d Prediccion	C.Adapt	Observacion	Error	Informacion Posterior			
t	Qt=Rt + 100	ft=mt-1	At=Rt/Qt	Yt	et=Yt-ft	mt=mt-1 +Atet	Ct=AtVt	Rt=Ct-1+Wt	constantes
0						130	400		Vt=100
1	505	130,0	0,80	150	20,0	146,0	80	405	Wt=5
2	185	146,0	0,46	136	-10,0	141,4	46	85	
3	151	141,4	0,34	143	1,6	142,0	34	51	
4	139	142,0	0,28	154	12,0	145,3	28	39	
5	133	145,3	0,25	135	-10,3	142,8	25	33	
6	130	142,8	0,23	148	5,2	144,0	23	30	
7	128	144,0	0,22	128	-16,0	140,5	22	28	
8	127	140,5	0,21	149	8,5	142,3	21	27	
9	126	142,3	0,21	146	3,7	143,1	21	26	
10	126	143,1	0,20				20	26	

Figura 3.1: Tabla de Valores

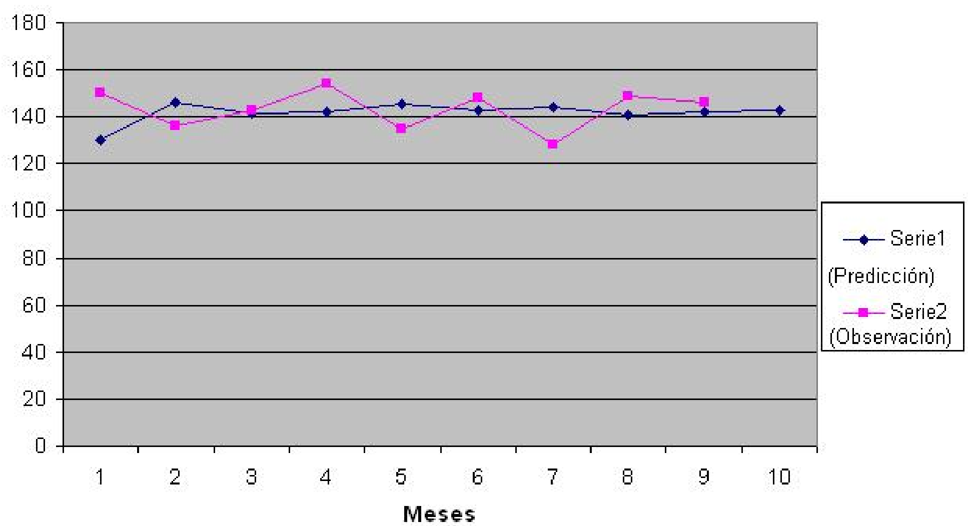

Figura 3.2: Grafica de las Observaciones y la Predicción

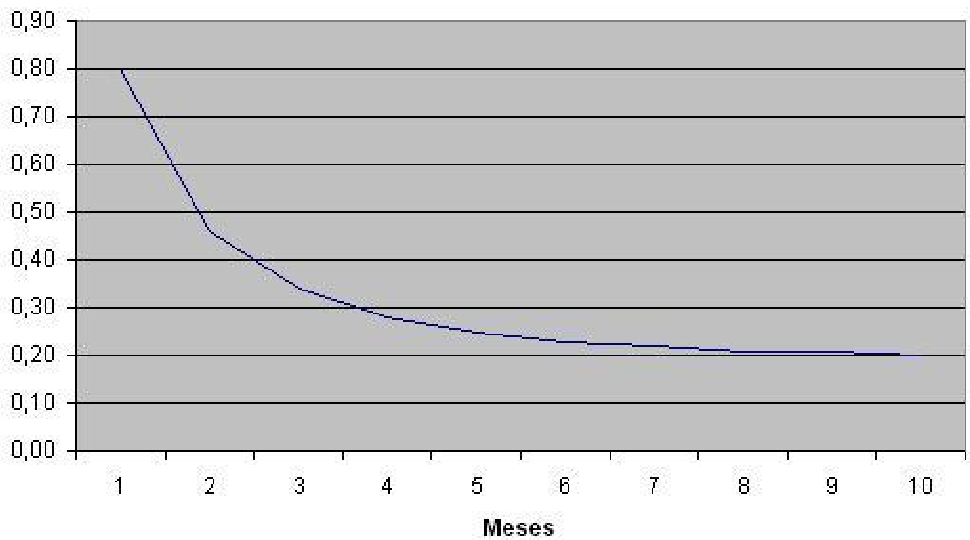

Figura 3.3: Coeficiente de Adaptación

Conclusión

Estos modelo se utilizan eficazmente en numerosas aplicaciones, sobre todo en predicción a corto plazo para la planificación de producción y control de existencias. Por ejemplo, en modelización de la demanda de mercado de un producto, μ_t representa la verdadera demanda del mercado en el tiempo t. Se supone que localmente en el tiempo, con unos pocos períodos hacia delante o hacia atrás, la demanda es aproximadamente constante. Cambios significativos durante períodos más largos de tiempo se esperan, pero la media cero y la independencia natural de la serie ω_t implica que el predictor no desea anticipar la forma de esta variación a largo plazo, simplemente lo describe como un proceso puramente estocástico.

REFERENCIAS BIBLIOGRÁFICAS

[1] Anderson, T.W.(1958). An Introduction to Multivariate Statistical Analysis. New York: John Wiley and Sons.

[2] Berger J.O. (1985). Statistical Decision Theory and Bayesian Analysis. 2nd ed. New York: Springer.

[3] Bernardo, J.M. y Smith A.F.M. (1994). Bayesian Theory. New York: Wiley.

[4] Box, G.E.P y Tiao, G.C. (1973). Bayesian Inference in Statistical Analysis. Addison-Wesley Publishing Co.

[5] Bradley, P.C. y Thomas, A.L. (1996). Bayes and Empirical Bayes Methods for Data Analysis. Chapman & Hall.

[6] Casella, G. y Berger, R.L. (1990). Statistical Inference. Thomson.

[7] Congdon Peter (2001). Statistical Modelling. New York: John Wiley and Sons.

[8] Johnson, R. y Wichern, D. (1998). Applied Multivariate Statistical Analysis. 4ta. Edition. Prentice Hall.

[9] De Groot, M.H. (1971). Optimal Statical Decisions. McGraw-Hill, New York.

[10] Peña Daniel. (2002). Análisis de Datos Multivariantes. McGraw-Hill, Madrid.

[11] Press, James S. (1982). Applied Multivariate Analysis: using Bayesian and Frequentist Methods of Inference. Second Edition. Malabar, Florida.

[12] West, M. and Harrison, J. (1997). Bayesian Forecasting and Dynamic Models, 2nd.ed. Springer, New York.