

Tesis

Guillermo Stefano Molina Martínez

Marzo 2017

Índice

1. Introducción	2
2. Estadística Bayesiana	3
2.1. Una intuición inicial	3
2.2. Inferencia Bayesiana	4
2.3. Distribuciones <i>a priori</i>	4
2.4. Estimación	5
2.4.1. Estimación puntual	6
2.4.2. Estimación por intervalos	7
2.5. Distribuciones predictivas	8
2.6. Variables Latentes	10
2.7. Distribución Wishart	12
3. Análisis de Factores	13
3.1. Modelo factorial	13
3.2. Estimación por máxima verosimilitud	16
3.2.1. Estimación de la matriz de cargas	16
3.2.2. Estimación por Componentes Principales	18
3.3. Estimación de factores	19
3.4. Determinación del número de factores	20
4. Modelos Dinámicos Lineales	22
4.1. Modelos de Espacio Estado	22
4.2. El Modelo Dinámico Lineal	23
4.3. Análisis Prospectivo	25
4.3.1. Análisis prospectivo V_t constante y conocida	26
4.3.2. Análisis prospectivo V_t constante y desconocida	27
4.4. Análisis retrospectivo	30
4.4.1. Análisis retrospectivo V_t conocida y constante	31
4.4.2. Análisis retrospectivo V_t desconocida y constante	32

4.5. Factores de Descuento	33
4.5.1. Factores de descuento en la varianza de evolución W_t	33
4.5.2. Factores de descuento en la varianza de observación V_t de análisis pro- spectivo	34
4.5.3. Factores de descuento en la varianza de observación V_t de análisis re- trospectivo	37
5. Apéndice	38
5.1. Modelos de Espacio Estado	38

1. Introducción

2. Estadística Bayesiana

2.1. Una intuición inicial

La estadística bayesiana hereda su nombre del teorema más famoso del estadístico Thomas Bayes, el cual fue planteado para describir la probabilidad de ocurrencia de un evento sujeta a condiciones relacionadas a él. El paradigma Bayesiano se basa en el Teorema de Bayes para llevar a cabo un aprendizaje adicional sobre los conocimientos que se tienen sobre un evento, la intuición es la que sigue.

Definición. Dados A y B dos eventos tales que $p(B) > 0$, la probabilidad condicional A dado B se define como

$$p(A|B) = \frac{p(A \cap B)}{p(B)}$$

Suponer que se tiene **H**, una hipótesis que quiere ser probada y se posee evidencia para poner a prueba la veracidad de ella, la cual se denota como **E**, por la definición de probabilidad condicional, se tiene que

$$p(E \cap H) = p(E|H)p(H) = p(H|E)p(E)$$

Si la evidencia es verídica, es decir $p(E) > 0$, se obtiene la fórmula del Teorema de Bayes:

$$p(H|E) = \frac{p(E|H)p(H)}{p(E)}$$

2.2. Inferencia Bayesiana

En la práctica se utilizan las variables θ , una característica de la población en cuestión que sustituye a la hipótesis \mathbf{H} , la cual es considerada variable aleatoria por ser desconocida y $\underline{x} = \{x_1, x_2, \dots, x_n\}$, un conjunto de observaciones de la población, que sustituyen a la evidencia \mathbf{E} , sobre la cual se busca hacer la inferencia.

En el sentido Bayesiano, los términos de esta expresión son conocidos como sigue:

$$f(\theta|\underline{x}) = \frac{f(\underline{x}|\theta)f(\theta)}{f(\underline{x})}$$

- $f(\theta)$ es llamada distribución *a priori*, el conocimiento muchas veces empírico que se tiene del evento antes de conocer la validez de la hipótesis.
- $f(\underline{x}|\theta)$ es la verosimilitud $f(\underline{x}|\theta) = L(\theta; \underline{x})$, asumiendo que \underline{x} sea fijo dado θ .
- $f(\underline{x})$ es la constante de proporcionalidad.
- $f(\theta|\underline{x})$ se llama distribución posterior o *a posteriori*, es la probabilidad de la hipótesis dada la evidencia \underline{x} . También es conocido como *kernel* de la distribución, el cual se dice que es proporcional a la verosimilitud por la *a priori*:

$$f(\theta|\underline{x}) \propto L(\theta; \underline{x})f(\theta)$$

2.3. Distribuciones *a priori*

Una distribución *a priori* para una variable θ posee todas las características de la distribución de una variable aleatoria con la excepción de que θ no es observable. Debido a lo anterior, la distribución de la variable θ está construida con los conocimientos o suposiciones

que se tienen acerca de su verdadero valor. Las distribuciones *a priori* de una misma variable pueden variar según la cantidad de información que se posea sobre ella o las fuentes que se consulten para obtenerla, por lo que cada una de éstas influirá en la forma y valores de la distribución posterior obtenida después del proceso de inferencia. Existen dos métodos populares para determinar las distribuciones *a priori*, uno es el de distribuciones conjugadas y el otro es el de distribuciones iniciales no informativas.

Usando distribuciones conjugadas se puede añadir información adicional a un modelo, el método consiste en basar la distribución *a priori* en la forma de la verosimilitud. Se dice que se tiene un par conjugado cuando la distribución *a priori* y la posterior pertenecen a la misma familia de distribuciones. Se llama familia de distribuciones al conjunto de distribuciones de probabilidad que poseen, entre otras cosas, la misma forma funcional y el mismo soporte.

Existen ocasiones en las que es poco conveniente usar las creencias o información *a priori* debido a un posible sesgo en éstas que podrían aportar a la distribución final o la forma distribucional no es conjugable, por lo que en situaciones como éstas se adoptan distribuciones *a priori* que aportan información mínima o nula y reducen el sesgo subjetivo, se conocen como distribuciones no informativas.

2.4. Estimación

El problema de estimación estadística surge de la necesidad de aproximar el valor de uno o varios parámetros que definen alguna característica de una población, esta característica se puede representar como una variable aleatoria X con densidad $f(x; \theta)$ conocida y un

parámetro θ desconocido (Mood, 1963.). Para llevar a cabo la estimación, se cuenta con una muestra aleatoria de valores $\hat{x} = \{x_1, x_2, \dots, x_n\}$, los cuales dependen del valor θ . Existen dos tipos de estimación: puntual y por intervalos de confianza, llamados intervalos de credibilidad cuando se calculan por el método Bayesiano.

2.4.1. Estimación puntual

La estimación puntual se lleva a cabo cuando se asigna el valor de una estadística en función de la muestra para representar al parámetro desconocido θ . El valor asignado se denomina *estimador puntual*. En estadística matemática, el método que produce el mejor estimador insesgado es el de máxima verosimilitud, el cual basado en la muestra y su distribución, busca obtener el valor de θ que maximiza la verosimilitud $L(\theta; \underline{x})$.

En estadística Bayesiana, la estimación puntual se lleva a cabo penalizando la diferencia que existe entre el estimador $\hat{\theta}$ y el valor real θ que se supone aleatorio. La forma de llevar a cabo dicha penalización es a través de *funciones de pérdida* que se denotan $L(\hat{\theta}, \theta)$. El estimador puntual, de manera similar al método de máxima verosimilitud, se encuentra minimizando la pérdida esperada, en este caso, encontrando $\hat{\theta}$ que minimice bajo la distribución posterior.

$$\min_{\hat{\theta}} E_{\theta|\underline{x}}(L(\hat{\theta}, \theta))$$

Las funciones de pérdida más comunes involucran como resultado una característica muchas veces fácil de describir acerca de la distribución posterior: la función de pérdida cuadrática tiene por estimador $\hat{\theta} = E(\theta|\underline{x})$, para la función de valor absoluto, $\hat{\theta}$ es la mediana de la distribución y para la función de pérdida 0-1, el estimador es la moda de la distribución.

2.4.2. Estimación por intervalos

En estadística matemática, un intervalo de confianza al $(1 - \alpha)$ es aquel en el que en el rango de valores que lo definen, podrá ser observado el valor real del parámetro en un $(1 - \alpha) * 100\%$ de los casos. El análisis de intervalo de confianza depende del cálculo de variables pivotaes que permiten estimar el rango de los intervalos dado cierto α de confianza.

En estadística Bayesiana, un intervalo de credibilidad se define de la misma manera que en estadística matemática, pero su cálculo depende únicamente de la distribución posterior obtenida por el método Bayesiano. Un intervalo de credibilidad se define como sigue:

$$p(a < \theta < b|\underline{x}) = 1 - \alpha$$

sustituyendo la igualdad por un \geq en el caso discreto.

Los intervalos de credibilidad más comunes son los intervalos simétricos y los de máxima credibilidad.

Intervalo simétrico es aquel que deja la misma masa de probabilidad en cada cola de la distribución.

$$p(a < \theta < b|\underline{x}) = 1 - \alpha$$
$$p(\theta \leq a) = p(\theta \geq b) = \frac{\alpha}{2}$$

Intervalo de máxima credibilidad es aquel que minimiza la longitud del intervalo sujeto al valor de α .

$$\begin{aligned}
& \underset{a,b}{\text{mín}} && b - a \\
& \text{s.a.} && p(a < \theta < b) = 1 - \alpha \\
& && a, b \in \mathbb{R}
\end{aligned}$$

2.5. Distribuciones predictivas

A pesar de que la idea fundamental de la estadística Bayesiana radica en encontrar el parámetro θ para la distribución posterior $f(\theta|\underline{x})$, muchas veces el verdadero problema proviene de encontrar cómo sería una observación futura, $x_{n+1}|\underline{x}$: su distribución, esperanza y varianza. Para esto se utilizan las distribuciones predictivas.

Las distribuciones predictivas se basan en el mismo fundamento de la estadística Bayesiana: utilizan la distribución de el parámetro θ para determinar la distribución de una $x|\theta$. Las distribuciones predictivas pueden ser a priori o posteriores.

Las distribuciones **predictivas a priori** ocupan el conocimiento que se tiene acerca del parámetro θ así como la distribución conocida de las variables $\underline{x}|\theta$, la cual se asume para la observación x_{n+1} para determinar la distribución de la observación siguiente sin tomar en cuenta los valores observados para \underline{x} . De esta manera se está eliminando la influencia del parámetro θ el cual se mantiene desconocido.

$$\begin{aligned}
f(\underline{x}) &= \int_{\theta} f(\underline{x}, \theta) d\theta \\
&= \int_{\theta} f(\underline{x}|\theta) f(\theta) d\theta
\end{aligned}$$

Por otro lado, las distribuciones **predictivas posteriores** determinan la distribución para la observación siguiente basadas en el método Bayesiano previamente explicado: con base en la distribución posterior $f(\theta|\underline{x})$ y la distribución conocida para $x_i|\theta$, se busca la distribución de la observación siguiente dadas las observaciones pasadas. En éstas distribuciones, la influencia de la variable θ se elimina después de conocer su distribución posterior $f(\theta|\underline{x})$.

$$\begin{aligned}
f(x_{n+1}|\underline{x}) &= \int_{\theta} f(x_{n+1}, \theta|\underline{x}) d\theta \\
&= \int_{\theta} f(x_{n+1}|\underline{x}) f(\theta|\underline{x}) d\theta
\end{aligned}$$

En éstas distribuciones es fácil probar que existe una forma analítica para determinar el valor esperando de la siguiente observación en caso de no ser necesario el conocimiento de la distribución, esto es posible dado el conocimiento de la Ley de Esperanzas Iteradas y el Teorema de Green.

$$\begin{aligned}
E[x_{n+1}] &= \int_{x_{n+1}} x_{n+1} f(x_{n+1}) dx_{n+1} \\
&= \int_{x_{n+1}} x_{n+1} \left[\int_{\theta} f(x_{n+1}|\theta) f(\theta) d\theta \right] dx_{n+1} \\
&= \int_{x_{n+1}} \int_{\theta} x_{n+1} f(x_{n+1}|\theta) f(\theta) d\theta dx_{n+1} \\
&= \int_{\theta} \left[\int_{x_{n+1}} x_{n+1} f(x_{n+1}|\theta) dx_{n+1} \right] f(\theta) d\theta \\
&= \int_{\theta} E[x_{n+1}|\theta] f(\theta) d\theta
\end{aligned}$$

De la misma manera, se puede probar para el valor esperado de la predictiva posterior, la cual se define

$$\begin{aligned}
E[x_{n+1}|\underline{x}] &= \int_{x_{n+1}} x_{n+1} f(x_{n+1}|\underline{x}) dx_{n+1} \\
&= \int_{\theta} E[x_{n+1}|\theta] f(\theta|\underline{x}) d\theta
\end{aligned}$$

2.6. Variables Latentes

Existen ocasiones en las que se tienen variables en el modelo que no pueden ser medidas ni observadas directamente y deben ser medidas a través de otros indicadores que sí son observables. Las *variables latentes* conectan esos dos conceptos, en el sentido de estadística Bayesiana esto funciona como sigue:

$$p(\underline{x}|\theta) = \int_{\Gamma} p(\underline{x}|\gamma)p(\gamma|\theta)d\gamma$$

donde $p(\gamma|\theta)$ no es una a priori, es una parte del modelo y γ es la variable latente.

Algunas implicaciones

Con el conocimiento de variables latentes, la verosimilitud de una muestra independiente con un parámetro desconocido se puede describir como

$$\begin{aligned}\mathcal{L}(\theta|\underline{x}) &= \prod_{i=1}^n p(x_i|\theta) \\ &= \prod_{i=1}^n \int_{\Lambda} p(x_i|\lambda_i)p(\lambda_i|\theta)d\lambda\end{aligned}$$

Si $p(x_i|\lambda)p(\lambda|\theta)$ no existe analíticamente o no se quiere estudiarla directamente, se puede añadir a θ , el parámetro del modelo, el conjunto de variables latentes $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$:

$$\mathcal{L}(\theta, \{\lambda_i\}_{i=1}^n|\underline{x}) = \prod_{i=1}^n p(x_i|\lambda_i)p(\lambda_i|\theta)$$

a esto se le conoce como verosimilitud extendida.

Se puede calcular la distribución final de $\theta, \{\lambda_i\}_{i=1}^n$ usando el teorema de Bayes

$$p(\theta, \{\lambda_i\}_{i=1}^n|\underline{x}) \propto \prod_{i=1}^n p(x_i|\lambda_i)p(\lambda_i|\theta)\Pi(\theta)$$

donde $x_i|\lambda_i$ es la variable observable, $\lambda_i|\theta$ es la variable latente y θ es el parámetro de la distribución a priori.

2.7. Distribución Wishart

La distribución Wishart es de gran importancia en el análisis multivariado, se relaciona directamente con la distribución normal multivariada ya que permite estimar la matriz de varianzas para ella, también se interpreta como una generalización multivariada de la distribución χ^2 o una distribución Gamma en caso de que sus grados de libertad no sean enteros. En el caso del análisis factorial como el que se aborda en este trabajo, permite estimar la matriz de cargas de los factores a través de la estimación por máxima verosimilitud. La distribución fue propuesta por John Wishart en 1928 con la intención de estimar la matriz de covarianzas para el modelo normal multivariado.

Sea $V \in \mathbb{R}^{p \times p}$ una matriz simétrica positiva definida. V sigue una distribución Wishart p -dimensional con matriz de escala Σ y n grados de libertad ($p \leq n$) si la distribución conjunta de los parámetros de V es continua con función de densidad

$$p(V) = \frac{c|V|^{\frac{n-p-1}{2}}}{|\Sigma|^{\frac{n}{2}}} \exp\left(-\frac{1}{2}\text{tr}(\Sigma^{-1}V)\right) \quad V > 0, \quad \Sigma > 0$$

donde $p(V) = 0$ en otro caso. La constante c se define como

$$c = \left[2^{\frac{np}{2}} \pi^{\frac{p(p-1)}{4}} \prod_{j=1}^p \Gamma\left(\frac{n+1-j}{2}\right) \right]^{-1}$$

En caso de que $n < p$, la distribución no existe. La relación entre V y los parámetros se va a denotar como $V \sim W(\Sigma, p, n)$.

3. Análisis de Factores

Existen ocasiones en las que se lleva a cabo un estudio estadístico en el que las variables deseadas son creadas por el sujeto que está llevando a cabo el estudio y no son cuantificables con variables observables o medibles de manera fácil. Éstas variables se llaman *latentes*, las cuales se identifican por no ser medibles pero se pueden relacionar a otras que sí pueden ser medidas, llamadas *variables manifiestas*. El análisis de factores es un método para desarrollar las relaciones que existen entre estos tipos de variables, las cuales se definen como combinaciones lineales de varias *variables manifiestas* para formar una *variable latente*, mejor conocida como *factor*. El modelo de factores se define de la siguiente manera:

$$x = \Lambda f + \theta + e$$

donde $x^T = (x_1, x_2, \dots, x_N)$ es el vector de variables *manifiestas* que se relacionan con las *variables latentes* no observadas, o *factores*, denotados por $f = (f_1, f_2, \dots, f_k)$ de dimension $k < N$ y tan pequeño como sea posible, mientras que la matriz Λ contiene la información que relaciona a éstas dos variables, conocida como la matriz de *loadings*, los cuales representan el porcentaje de la varianza explicada por ese factor para la variable x_i . El vector $e \in \mathbb{R}^N$ representa al error generado por el modelo y generalmente posee una varianza pequeña.

3.1. Modelo factorial

Sea el modelo de análisis factorial

$$x = \Lambda f + \theta + e$$

A las suposiciones antes mencionadas se agregan otras con respecto a f :

- $f \perp e$, i.e. f y e son independientes
- $f \sim N(0, I_k)$
- $e \sim N(0, D_\psi)$
- $\Lambda^T D_\psi^{-1} \Lambda \equiv D_J$, $D_J \in \mathbb{R}^{k \times k}$, $D_J = \text{diag}(J_1, J_2, \dots, J_k)$ con $J_1 > J_2 > \dots > J_k$

La necesidad de tener $k < N$ recae en que lo que se busca con el modelo es simplificar la estructura de los datos observados, de manera que su representación debe ser sencilla y consistente.

Press(1982), sugiere asumir las distribuciones de f y e como normales para poder llevar a cabo la estimación por máxima verosimilitud de los parámetros θ , Λ y D_ψ , también sugiere, en el caso de no existir normalidad, llevar a cabo una transformación que induzca normalidad a la variables. La matriz de varianza y covarianza para f , definida como I_k permite que los factores sean invariantes a cambios de escala, mientras que la no correlación entre factores puede cambiar acorde al tipo de modelo factorial que se trabaje.

Los errores e tienen la propiedad de ser mutuamente independientes y media cero para buscar la menor variación posible con respecto a la variable real, pero también se tiene la propiedad de heterocedasticidad que permite que las varianzas de dos errores no sean necesariamente las mismas.

El supuesto para D_J permite eliminar la infinidad de soluciones que admite el modelo inicial, donde cada transformación ortogonal posible genera una matriz Λ diferente. La matriz obtenida no siempre es la que permite llevar a cabo un mejor estudio, por lo que a partir de la solución impuesta se puede trabajar con rotaciones que se ajusten de mejor manera al modelo.

Con los supuestos antes mencionados y dada la metodología del presente trabajo, es

conveniente plantear el problema del análisis de factores en el paradigma Bayesiano. Para llevar a cabo ésto, se define la variable $y = x - \theta$, reesultando el modelo de la forma normal multivariada

$$\begin{aligned}(y|f) &\sim N_p(\Lambda f, D_\psi) \\ f &\sim N_k(0, I_k)\end{aligned}$$

El arreglo del modelo en forma Bayesiana permite dar un nuevo enfoque al análisis en el que se puede estimar las variables por medio del Algoritmo EM o el Muestro de Gibbs.

Dadas las suposiciones nombradas, se tienen las siguientes propiedades respecto a la varianza en el análisis factorial:

La varianza de la variable x_i , σ_i^2 , se define como

$$\sigma_i^2 = \sum_{j=1}^k \lambda_{ij}^2 + \psi_i^2$$

donde λ_{ij} determina el *loading* para el factor f_j con respecto a la variable x_i y ψ_i es la varianza del error correspondiente, e_i . La primera parte de la suma, $h_i^2 = \sum_{j=1}^k \lambda_{ij}^2$, representa la varianza que x_i comparte con las otras variables por medio de los factores, mientras que ψ_i es la varianza intrínseca de la variable y no la comparte con las demás.

La covarianza entre dos variables x_i y x_j con $i \neq j$ se determina como

$$\sigma_{ij} = \sum_{l=1}^k \lambda_{il} \lambda_{lj}$$

Dadas éstas propiedades de la varianza, se obtiene que la matriz de varianzas y covarianzas de x , Σ , se puede representar como la suma de un producto de matrices y una matriz

$$\Sigma = \Lambda\Lambda^T + \Psi, \quad \Psi = \text{diag}(\psi_i)$$

Las condiciones planteadas definen el modelo de factores como una suma de vectores y productos de vectores con matrices que deben ser estimados acorde a los información que se posee. Existen dos métodos principales de estimación para los parámetros Λ y D_Ψ del modelo de factores, uno está basado en el método de análisis de componentes principales y el otro en el método de máxima verosimilitud, siendo éste último el que más se utiliza por la teoría detrás de su cálculo, donde se determina el mejor estimador insesgado para los parámetros.

3.2. Estimación por máxima verosimilitud

Como el método en estimación de parámetros en Inferencia Estadística, el método de máxima verosimilitud para estimar los parámetros en el análisis de factores es un método robusto y ampliamente aceptado. Por lo general el método comienza fijando un valor para m , el número de factores que tendrá el modelo para posteriormente estimar Λ y D_ψ . Posteriormente existe una variedad de pruebas para determinar el número final de factores que tendrá el modelo.

3.2.1. Estimación de la matriz de cargas

Dada la suposición de independencia para las observaciones x_1, x_2, \dots, x_N de X , los estimadores de máxima verosimilitud de la media y la matriz de varianzas y covarianzas son

$$\begin{aligned}\hat{\theta} &= \bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \\ \hat{\Sigma} &= \frac{A}{N} = S = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(x_i - \bar{x})'\end{aligned}$$

Calculados estos estimadores y con las características del modelo, se puede observar que $Var(X) = \Sigma = \Lambda\Lambda' + D_\psi$, por lo que suponiendo que $f \sim N(0, D_\sigma^2)$ con $D_\sigma = diag(\sigma_1, \sigma_2, \dots, \sigma_k)$ y $\sigma_j = var(f_j), j = 1, 2, \dots, k$ entonces se puede llevar a cabo la descomposición

$$\Sigma = var(X) = (\Lambda D_\sigma)(\Lambda D_\sigma)' + D_\psi$$

Como $X \sim N(\theta, \Sigma/N)$, se tiene que A se distribuye Wishart con parámetros (Σ, p, n) con $p < n$, se puede calcular la función de verosimilitud L multiplicando éstas densidades y obteniendo su logaritmo.

$$L = c + \left(\frac{n-p-1}{2} \right) \log|A| - \frac{N}{2} \log|\Lambda\Lambda' + D_\Psi| - \frac{1}{2} tr(\Lambda\Lambda' + D_\Psi)^{-1} A - \frac{N}{2} (\bar{x} - \theta)' (\Lambda\Lambda' + D_\Psi)^{-1} (\bar{x} - \theta)$$

Haciendo uso de la verosimilitud calculada, el supuesto $\theta = \bar{x}$ y diferenciación con respecto a Λ y cada ψ_i , se pueden calcular por máxima verosimilitud los estimadores $\hat{\Lambda}$ y $D_{\hat{\psi}}$ por medio del sistema

$$\begin{aligned}diag(D_{\hat{\psi}} + \hat{\Lambda}\hat{\Lambda}') &= diag(S) \\ SD_{\hat{\psi}}^{-1}\hat{\Lambda} &= \hat{\Lambda}(I + \hat{\Lambda}'D_{\hat{\psi}}^{-1}\hat{\Lambda})\end{aligned}$$

Éste sistema de ecuaciones es complejo algebraicamente, además de que existe una infinidad de soluciones para él. Una solución es presentar la restricción $\Lambda^T D_\psi^{-1} \Lambda \equiv D_J$, $D_J \in \mathbb{R}^{k \times k}$, $D_J = \text{diag}(J_1, J_2, \dots, J_k)$ con $J_1 > J_2 > \dots > J_k$ de manera que se garantiza una solución única. Existen rutinas programadas para solucionar de manera eficiente éste sistema como lo son RMLFA(*Restricted Maximum Likelihood Factor Analysis*) y UMLFA(*Unrestricted Maximum Likelihood Factor Analysis*).

3.2.2. Estimación por Componentes Principales

El Análisis por Componentes Principales se basa en transformaciones ortogonales para convertir conjuntos de datos correlacionados en un conjunto linealmente independiente cuya varianza ponderada se maximice, es decir, cada componente principal adicional, que se calcula de manera secuencial, tendrá menor varianza que el calculado anteriormente, siendo el primero el que concentre la mayor varianza posible de los datos originales. Cabe señalar que el número de componentes principales debe ser menor al número de datos observados originalmente.

Una propiedad del Análisis de Componentes Principales consiste en la representación de los Componentes bajo una base ortogonal de la matriz de varianzas y covarianzas. Esto se obtiene tomando bajo consideración que la matriz de varianzas y covarianzas es una matriz simétrica positiva definida, por lo que se puede descomponer en un producto de matrices ortogonales con una diagonal:

$$\hat{\Sigma} = \hat{\Gamma} \hat{D}_\lambda \hat{\Gamma}'$$

En el modelo de factores, si se lleva a cabo $Y = X - \theta$, se obtiene $Y = \Lambda f + e$ y se sabe que $\Sigma = \Lambda \Lambda' + \Psi$, por lo que suponiendo que la varianza de los errores es muy pequeña, es decir $\Psi \approx 0$, se obtiene $\Sigma = \Lambda \Lambda'$. Una posible solución se presenta por el análisis de

componentes principales antes mencionado, donde a través de un caso de la factorización de Cholesky, la factorización LDLT, se puede llevar a cabo como sigue

$$\Sigma = \Gamma D_\lambda \Gamma' = (\Gamma D_\lambda^{1/2})(\Gamma D_\lambda^{1/2})'$$

con Γ ortogonal y $D_\lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$ con λ_i las raíces latentes de Σ . Una característica importante de ésta factorización es que si $\lambda_1 > \lambda_2 > \dots > \lambda_p$, entonces la solución es única. La matriz de cargas de los factores se define preliminarmente como:

$$\Lambda = \Gamma D_\lambda^{1/2}$$

donde $\Lambda \in \mathbb{R}^{p \times p}$, para ajustar al modelo factorial, suponiendo que se ha llevado a cabo la elección del número óptimo de factores, se necesitan recortar las $p - m$ columnas que corresponden a las raíces latentes de Σ con menor magnitud, por lo que si se utiliza el estimador $\hat{\Sigma}$, se tiene que

$$\hat{\Lambda} = \hat{\Gamma} \hat{D}_\lambda^{1/2}$$

con $\hat{\Lambda} \in \mathbb{R}^{p \times m}$, $\hat{\Gamma} \in \mathbb{R}^{p \times m}$ y $D_\lambda \in \mathbb{R}^{m \times m}$.

3.3. Estimación de factores

La estimación de los factores se lleva a cabo bajo la suposición de la independencia entre cada x_i para x_1, \dots, x_N de manera que cada una se puede representar como

$$x_j = \Lambda f_j + \theta + e_j, \quad j = 1, \dots, N$$

donde $f_j \sim N(0, I)$, $e_j \sim N(0, D_\psi)$ con f_j y e_j independientes para toda j .

La estimación de los factores F se lleva a cabo tratando a cada $E(f_j)$ como una función

lineal de x_j , $(f_j|x_j) = Ax_j + u_j$ $j = 1, \dots, N$, donde $E(u_j) = 0$, $var(u_j) = \sigma^2 I_m$ de manera que la representación matricial del modelo resulta $(F|X) = AX + U$. La estadística A se estima bajo mínimos cuadrados suponiendo a F conocida, con lo que se obtiene $\hat{A} = (FX)'(XX')^{-1}$. Utilizando la forma matricial $FX' = F(F'\Lambda' + \theta + U')$, se puede observar que $E(FX') = N\Lambda'$ y posteriormente suponiendo la ley de grandes números, se puede estimar a FX' como $E(FX')$ para muestras grandes. Estimado el valor $\hat{A} \approx N\hat{\Lambda}(XX')^{-1}$ se obtiene de manera inmediata el estimador para el vector F :

$$\hat{F} = \hat{A}X$$

Press(1982), sugiere utilizar la estimación de la matriz de covarianzas representada de forma matricial, $S = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(x_i - \bar{x})' = \frac{1}{N} XX' - \hat{\theta}\hat{\theta}'$, de manera que $\hat{A} \approx \Lambda'(S + \hat{\theta}\hat{\theta}')^{-1}$ obteniendo el estimador

$$F = \hat{A}X \approx \hat{\Lambda}'(S + \hat{\theta}\hat{\theta}')^{-1}X$$

3.4. Determinación del número de factores

El número de factores es uno de los puntos más importantes del modelo ya que éste trata de minimizar el número de variables de manera que se obtenga la mejor representación de éstas sin afectar la información presentada. Un número muy grande de factores resulta redundante con respecto a la información y un número muy pequeño puede resultar insuficiente para explicar el fenómeno en cuestión. Existen múltiples métodos para determinar el número de factores en el modelo, varios sugieren el uso de pruebas sucesivas sobre verosimilitudes o cocientes de verosimilitudes, otros se basan en superar un umbral de mínima varianza que representar.

Akaike (1973, 1977) propone el uso de su criterio para elegir el número de factores sujeto a la minimización del mismo (Criterio de Información de Akaike). El criterio de Akaike se define como

$$AIC := 2(\text{número de parámetros independientes}) - 2\log L^*$$

con L^* siendo la verosimilitud L maximizada para Λ y D_Ψ elegidos anteriormente sujeta al número de factores elegidos.

4. Modelos Dinámicos Lineales

Los modelos de *espacio estado* se originaron a principios de los sesentas, a pesar del interés que existía por pronosticar utilizando series de tiempo desde tiempo anterior. Con la propuesta del **filtro de Kalman**, el cual considera la representación de procesos aleatorios de Bode-Shannon y el *método de transición de estado* para analizar sistemas dinámicos, se tuvo un nuevo enfoque para estudiar los modelos de *espacio estado* y su uso se volvió común en temas de ingeniería. Ésta propuesta funciona tanto para procesos estacionarios como no estacionarios. La propuesta de Kalman ha sido estudiada por la estadística desde la década de los setenta para el estudio de series de tiempo.

Un modelo de *espacio estado* considera una serie de tiempo como resultado de un sistema dinámico con perturbaciones aleatorias. Una ventaja que presentan es que su cómputo se puede efectuar a través de algoritmos recursivos que facilitan su implementación, algunas aplicaciones de éstos son para estimación y pronóstico en las que se calculan distribuciones condicionales de las variables de interés dada la información poseída. Debido a esto, su uso se complementa con el paradigma Bayesiano. Los modelos de espacio estado permiten modelar series de tiempo tanto univariadas como multivariadas incluso bajo no estacionariedad, cambios estructurales o patrones irregulares.

4.1. Modelos de Espacio Estado

Suponer que se tiene interés en una serie de tiempo $\{y_t\}$ determinada por una variable aleatoria no medible, θ_t . Esta relación se representa en términos de probabilidad con la distribución condicional $(y_t|\theta_t)$, la cual se define de manera que la variable $y_t|\theta_t$ sea independiente de la información anterior $D_{t-1} = \{y_1, y_2, \dots, y_{t-1}\}$.

La variable θ_t , por su parte, se define a través de un proceso de Markov en el que depende

de su observación anterior, θ_{t-1} , por medio de la distribución condicional $(\theta_t|\theta_{t-1})$.

La definición de las probabilidades $P(y_t|\theta_t)$ y $P(\theta_t|\theta_{t-1})$ varía atendiendo a las necesidades de cada modelo, en el presente trabajo se empleará la definición de Modelo Dinámico Lineal que se explica en el siguiente capítulo.

4.2. El Modelo Dinámico Lineal

El modelo dinámico lineal normal se define a partir de la cuarteta

$$\{F, G, V, W\}_t = \{F_t, G_t, V_t, W_t\}$$

para cada tiempo t , donde

- F_t es una matriz de dimensión $n \times r$ conocida.
- G_t es una matriz de dimensión $n \times n$ conocida.
- V_t es una matriz de dimensión $r \times r$ conocida.
- W_t es una matriz de dimensión $n \times n$ conocida.

Ésta cuarteta relaciona el modelo a y_t con el parámetro $\theta_t \in \mathbb{R}^{n \times 1}$ en cada tiempo t a través de las distribuciones

$$\begin{aligned} (y_t|\theta_t) &\sim N(F_t'\theta_t, V_t) \\ (\theta_t|\theta_{t-1}) &\sim N(G_t\theta_{t-1}, W_t) \\ (\theta_0|D_0) &\sim N(m_0, C_0) \end{aligned}$$

que se definen de forma secuencial con la información Y_{t-1}, Y_{t-2} y para algunos momentos

iniciales m_0 y C_0 , éstos últimos para completar el esquema Bayesiano que requiere una distribución inicial.

Otra representación de éstas ecuaciones es

$$\begin{aligned} y_t &= F_t' \theta_t + v_t, & v_t &\sim N(0, V_t) \\ \theta_t &= G_t \theta_{t-1} + w_t, & w_t &\sim N(0, W_t) \\ (\theta_0 | D_0) &\sim N(m_0, C_0) \end{aligned}$$

En las ecuaciones anteriores, v_t y w_t son interna y mutuamente independientes.

La ecuación $y_t = F_t' \theta_t + v_t$ se llama **ecuación de observación** y define la distribución de la muestra y_t condicional al parámetro θ_t , de manera que dada θ_t , y_t es independiente de las demás observaciones y parámetros. La ecuación relaciona y_t y θ_t por medio de una regresión lineal que contiene errores normales multivariados con varianza V_t , la cual puede ser conocida pero cambiante a través del tiempo y afecta a la sucesión de errores $\{v_t\}$ que representa las distorsiones en la observación de Y_t .

La ecuación $\theta_t = G_t \theta_{t-1} + w_t$ se llama **ecuación de estado**, la propiedad de independencia condicional permite observar una evolución de Markov tal que dado θ_{t-1} y los valores conocidos G_t y W_t , θ_t es independiente de la información en $t - 1$, D_{t-1} . Esta evolución se completa con la suma de un vector aleatorio w_t que influye en el desarrollo del sistema.

Para éstas ecuaciones, al tiempo t , se tiene

- F_t es la matriz de diseño compuesta de variables independientes;
- θ_t es el vector de estado;
- $\mu_t = F_t' \theta_t$ es la respuesta media o nivel;

- v_t es el error de observación;
- G_t es la matriz de evolución;
- w_t es el error de evolución con varianza W_t .

De acuerdo a cada modelo, F_t y G_t pueden ser elegidos de acuerdo a las necesidades que éste presenta, aunque existen ocasiones en las que éstos son desconocidos y varían en el tiempo. El parámetro V_t es una perturbación aleatoria sobre la observación de y_t , mientras que W_t influye en el desarrollo, o evolución, del sistema. Éstos dos últimos pueden ser desconocidos o variantes en el tiempo, las secciones siguientes abordarán éstos supuestos y cómo se puede llevar a cabo el estudio de los Modelos Dinámicos Lineales acorde a la información poseída acerca de ellos. En el caso en el que $\{F, G\}$ son constantes, se conoce al modelo como un **Modelo Dinámico Lineal de Series de Tiempo**, mientras que si la cuarteta $\{F, G, V, W\}$ es constante para toda t , el modelo se conoce como **Modelo Dinámico Lineal Constante**.

4.3. Análisis Prospectivo

Una de las propiedades de mayor interés de los Modelos Dinámicos Lineales es que éstas permiten un aprendizaje sucesivo dada la información que se tiene al tiempo $t - 1$ y la obtenida en la realización al tiempo t , por lo que suponiendo la existencia de información D_0 en el tiempo $t = 0$, se puede definir la información disponible en cualquier momento de tiempo t como

$$D_t = \{Y_t, D_{t-1}\}$$

con D_{t-1} la información disponible en $t - 1$ y Y_t la obtenida al tiempo t .

El análisis prospectivo se puede llevar a cabo bajo distintos supuestos para la cuarteta $\{F_t, G_t, V_t, W_t\}$, siendo la diferencia el conocimiento que se tiene acerca de V_t , existen tres casos que se construyen cada uno sobre el anterior, éstos son: V_t conocida, V_t desconocida y V_t desconocida y cambiante en el tiempo. Las primeras dos se pueden estudiar por medio del paradigma Bayesiano, mientras que la última requiere el uso de factores de descuento que se presentan más adelante en el trabajo.¹

4.3.1. Análisis prospectivo V_t constante y conocida

Bajo los supuestos presentados en el Modelo Dinámico Lineal, se presenta el modelo de aprendizaje más simple, en el que bajo únicamente la distribución Normal y la distribución a priori $(\theta_0|D_0) \sim N(m_0, C_0)$ definida con anterioridad, las distribuciones predictivas para cada t dado que la cuarteta $\{F_t, G_t, V_t, W_t\}$ es conocida, son definidas por West y Harrison(1997) como

a) Distribución posterior para θ_{t-1}

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

Para algunas m_{t-1} y C_{t-1} .

b) Distribución inicial para θ_t

$$(\theta_t|D_{t-1}) \sim N(a_t, R_t)$$

donde

$$a_t = G_t m_{t-1}$$

$$R_t = G_t C_{t-1} G_t' + W_t$$

¹Escribir qué capítulo es.

c) Pronóstico a un paso para y_t

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

donde

$$f_t = F_t' a_t \qquad Q_t = F_t' R_t F_t + V_t$$

d) Distribución posterior para θ_t

$$(\theta_t|D_t) \sim N(m_t, C_t)$$

con

$$m_t = a_t + A_t e_t \qquad C_t = R_t - A_t Q_t A_t'$$

donde

$$A_t = R_t F_t' Q_t^{-1} \qquad e_t = Y_t - f_t$$

Los resultados se pueden probar por inducción siguiendo la teoría de la distribución Normal.

4.3.2. Análisis prospectivo V_t constante y desconocida

En muchos modelos, normalmente se tiene que dentro de la cuarteta $\{F_t, G_t, V_t, W_t\}$, las matrices F_t de regresión y G_t de evolución son elegidas según las necesidades planteadas al inicio del estudio y la matriz W_t de evolución también se acostumbra elegir de acuerdo

a factores de descuento, que serán introducidos en la Sección ². La matriz restante, V_t es por lo general muy grande comparada con la varianza de evolución, aportando una mayor cantidad de variabilidad al modelo de predicción y es muchas veces desconocida, por lo que es necesario llevar a cabo el uso de la teoría Bayesiana para aproximar estos valores. En este capítulo se lleva a cabo la suposición de una varianza desconocida y constante, $V_t = V \forall t$, así como el uso de la precisión sustituyendo a la varianza para simplificar el análisis, esto es $\tau = V^{-1}$.

El Modelo Dinámico Lineal dada V_t constante y desconocida, se define como

$$\begin{aligned} Y_t &= F_t' \theta_t + v_t \quad v_t \sim N(0, \tau^{-1}) \\ \theta_t &= G_t \theta_{t-1} + w_t \quad w_t \sim N(0, \tau^{-1} W_t) \\ (\theta_0 | D_0, \tau^{-1}) &\sim N(m_0, \tau^{-1} C_0) \\ (\tau^{-1} | D_0) &\sim \Gamma\left(\frac{n_0}{2}, \frac{n_0 S_0}{2}\right) \end{aligned}$$

donde se asumen como conocidas los valores iniciales de m_0 , C_0 , n_0 y S_0 , además de la tripleta $\{F_t, G_t, W_t\}$. Hay que notar que la distribución inicial está compuesta por un par conjugado Normal/Gamma y que se tiene una distribución inicial para la precisión, de manera que se tiene una distribución inicial para θ_0 y τ^{-1} dada la información inicial D_0 .

Bajo éstos nuevos supuestos y definiciones, el aprendizaje sucesivo bajo la condición de V_t desconocida y constante se lleva a cabo de una forma similar para el vector de estados y las nuevas observaciones, de manera que las distribuciones predictivas también definidas por West y Harrison (1997) son

²determinar sección

a) Distribución posterior para θ_{t-1}

$$\begin{aligned}(\theta_{t-1}|D_{t-1}) &\sim T_{n_{t-1}}(m_{t-1}, C_{t-1}) \\(\tau^{-1}|D_{t-1}) &\sim \Gamma\left(\frac{n_{t-1}}{2}, \frac{n_{t-1}S_{t-1}}{2}\right)\end{aligned}$$

Para algunas m_{t-1} y C_{t-1} .

b) Distribución inicial para θ_t

$$\begin{aligned}(\theta_t|D_{t-1}) &\sim T_{n_{t-1}}(a_t, R_t) \\(\tau^{-1}|D_{t-1}) &\sim \Gamma\left(\frac{n_{t-1}}{2}, \frac{n_{t-1}S_{t-1}}{2}\right)\end{aligned}$$

donde

$$a_t = G_t m_{t-1} \qquad R_t = G_t C_{t-1} G_t' + W_t$$

c) Pronóstico a un paso para y_t

$$(Y_t|D_{t-1}) \sim T_{n_{t-1}}(f_t, Q_t)$$

donde

$$f_t = F_t' a_t \qquad Q_t = F_t' R_t F_t + S_{t-1}$$

d) Distribución posterior para θ_t

$$\begin{aligned}(\theta_t|D_t) &\sim T_{n_{t-1}}(m_t, C_t) \\(\tau^{-1}|D_t) &\sim \Gamma\left(\frac{n_t}{2}, \frac{n_t S_t}{2}\right)\end{aligned}$$

con las siguientes ecuaciones siendo aplicables en los casos donde sean requeridas

$$\begin{aligned}m_t &= a_t + A_t e_t & C_t &= \frac{S_t}{S_{t-1}}(R_t - A_t A_t' Q_t) \\n_t &= n_{t-1} + 1 & d_t &= n_t S_t = d_{t-1} + \frac{e_t^2}{Q_t} \\S_t &= \frac{d_t}{n_t}\end{aligned}$$

donde

$$A_t = R_t F_t' Q_t^{-1} \qquad e_t = Y_t - f_t$$

Los resultados de las relaciones se siguen del caso donde la varianza es conocida, de igual manera se utiliza inducción pero esta vez con la teoría de distribución Normal/Gamma. El estimador S_T corresponde al estimador puntual *a priori* de la varianza observacional.

4.4. Análisis retrospectivo

Existen ocasiones en las que el análisis no solo necesita conocer cómo se llevarán a cabo observaciones futuras, sino también se busca conocer cómo es posible que se hayan desarrollado los fenómenos estudiados antes de poder llevar a cabo observaciones de éstos para

poder tener un estudio global acerca del fenómeno en cuestión.

El análisis retrospectivo ayuda a inferir los valores de los vectores de estado, de manera que se busca obtener la distribución marginal retrospectiva de éstos, $(\theta_{t-k}|D_t)$, conocida la información al momento $t - k$ y después de éste. Al uso de datos para hacer inferencia sobre valores anteriores se le llama *filtrado*. La distribución $(\theta_{t-k}|D_t)$ para $k \geq 1$ se llama distribución de filtro de k pasos para el vector de estado al tiempo t .

Al igual que las distribuciones predictivas, las distribuciones retrospectivas pueden ser calculadas de manera sucesiva hacia atrás en el tiempo usando relaciones que se presentarán en éste capítulo. La estructura de las relaciones se deriva de una forma similar a la llevada a cabo en el análisis prospectivo de la sección anterior. De la misma manera, el análisis retrospectivo se puede llevar a cabo suponiendo V_t constante y conocida, constante y desconocida y variante en el tiempo y desconocida, siendo las primeras dos presentadas a continuación y la última en la sección ³ después de presentar factores de descuento.

4.4.1. Análisis retrospectivo V_t conocida y constante

El caso más simple para el análisis retrospectivo se presenta siendo conocida y constante la varianza de observación $V_t = V \forall t$, el procedimiento supone funciones para momentos del tiempo $1 \leq k < T$, por lo que la distribución marginal retrospectiva fue definido de manera sucesiva por West y Harrison (1997) como

$$(\theta_{T-k}|D_T) \sim N_p(a_T(-k), R_T(-k))$$

donde

³cuál sección

$$\begin{aligned}
a_T(-k) &= m_{T-k} + B_{T-k}(a_T(-k+1) - a_{T-k+1}) \\
R_T(-k) &= C_{T-k} + B_{T-k}(R_T(-k+1) - R_{T-k+1})B'_{T-k} \\
B_T &= C_T G'_{T+1} R'_{T+1}
\end{aligned}$$

utilizando los valores iniciales

$$a_T(0) = m_T; \quad R_T(0) = C_T$$

y definiendo

$$a_{T-k}(1) = a_{T-k+1}; \quad R_{T-k}(1) = R_{T-k+1}$$

4.4.2. Análisis retrospectivo V_t desconocida y constante

En el caso donde se desconoce el valor de V_t pero se sabe que es constante para todo tiempo t se lleva a cabo un análisis de la misma manera que en el caso V_t constante y conocida con un desarrollo de análisis conjugado como el del análisis prospectivo de la Sección 4.3.2, West y Harrison (1997) definen la distribución

$$(\theta_{T-k}|D_T) \sim T_{n_t} \left(a_T(-k), \frac{S_T}{S_{T-k}} R_T(-k) \right)$$

donde se definen $a_t(-k)$ y $R_t(-k)$ de la misma manera que en el caso de V_t conocida y constante, así como S_T , el estimador puntual *a priori* de V .

4.5. Factores de Descuento

Como se muestra en las secciones anteriores, una parte importante en la especificación de los Modelos Dinámicos Lineales consiste en la elección adecuada de una matriz de varianza de observación, V_t , y se exploran los casos en donde ésta es constante para todo tiempo t tanto en el análisis prospectivo como en el retrospectivo, pero también existirán ocasiones donde el fenómeno en cuestión tenga una varianza observacional, de evolución, y en algunos casos ambas, que cambie a través del tiempo. Para que el modelado se llevé a cabo de manera exitosa, es necesario especificar de manera adecuada la estructura y magnitud de la matriz de varianza en cuestión. Ésta sección muestra el uso de la técnica de factores de descuento para modelar a W_t y V_t como lo presentan West y Harrison (1997).

4.5.1. Factores de descuento en la varianza de evolución W_t

En un Modelo Dinámico Lineal, el valor de W_t contrala la variación estocástica en la evolución del modelo y determina su estabilidad en el tiempo, por lo que éste puede conducir a aumentar la incertidumbre acerca de θ_t .

Considerando la ecuación de sistema, en la Sección 4.3.1 se define $R_t = G_t C_{t-1} G_t' + W_t$, lo cual es la varianza del vector de estados al tiempo t considerando la información al tiempo $t - 1$, es decir, $R_t = V(\theta_t | D_{t-1})$. Suponiendo que no existe un error de evolución en el tiempo t , conduciendo a $W_t = 0$, se tendría un vector de estado *ideal*, sin alteraciones estocásticas y teniendo completa certidumbre sobre la evolución de $t - 1$ a t con varianza $P_t = G_t C_{t-1} G_t'$. Cabe señalar que este supuesto sólo tiene una validez *local*. West y Harrison (1997) resaltan la importancia de la matriz W_t explicando que ésta refleja qué tan durable es el modelo: un valor $W_t = 0$ representa una ecuación de evolución estable y globalmente confiable, mientras

que $W_t \rightarrow \infty$ muestra una ecuación de sistema, y por consecuencia un Modelo Dinámico Lineal, poco confiable. El objetivo de utilizar factores de descuento consiste en aumentar la incertidumbre de la varianza *ideal* sin alteraciones en su evolución, P_t , a una varianza realista $R_t = P_t + W_t$, por lo que se tiene que

$$R_t \geq P_t$$

El uso de factores de descuento es una solución sencilla al objetivo planteado, suponiendo el factor de descuento δ tal que $0 < \delta \leq 1$ con el caso $\delta = 1$ correspondiendo a modelos estáticos. El factor de descuento es un valor de δ tal que se obtenga la igualdad

$$V(\theta_t|D_{t-1}) = R_t = \frac{1}{\delta}P_t$$

que al sustituir en la definición original $R_t = P_t + W_t$ resulta al despejar

$$W_t = \frac{1-\delta}{\delta}P_t$$

por lo que conocidos C_0 y δ se puede calcular la serie $\{w_t\}$ por medio de la igualdad anterior y $P_t = G_t C_{t-1} G'_t$. Siendo la relación entre δ y W_t inversa, se obtiene que al contrario de W_t , valores grandes de δ , por lo general en el intervalo $[0,9, 0,999)$, resultan en sistemas de evolución estables, mientras que valores pequeños de δ corresponden a sistemas inestables carentes de sentido y utilidad.

4.5.2. Factores de descuento en la varianza de observación V_t de análisis prospectivo

Como se menciona en las Secciones 4.3.2 y 4.4.2, hay ocasiones en las que la varianza de observación es desconocida pero también puede ocurrir que ésta sea variable en el tiempo por

lo que es necesario estimarla de manera que su evolución pueda ser controlada a través del tiempo. El uso de factores de descuento permite llevarlo a cabo de manera que se cumplen ambas condiciones. Al igual que en el caso de V_t constante y desconocida, el modelo involucra el uso de la precisión $\tau = V_t^{-1}$ y se define utilizando la cuarteta $\{F_t, G_t, \tau^{-1}, \tau^{-1}W_t\}$

$$\begin{aligned} Y_t &= F_t' \theta_t + v_t; & v_t &\sim N(0, \tau^{-1}) \\ \theta_t &= G_t \theta_{t-1} + w_t; & w_t &\sim N(0, \tau^{-1} W_t) \\ (\theta_0 | D_0, \tau^{-1}) &\sim N(m_0, \tau^{-1} C_0) \\ (\tau^{-1} | D_0) &\sim \Gamma\left(\frac{n_0}{2}, \frac{n_0 S_0}{2}\right) \end{aligned}$$

siguiendo el esquema llevado a cabo en la Sección 4.3.2 y utilizando un factor de descuento $0 < \delta < 1$, se definen la distribuciones predictivas como

a) Distribución posterior para θ_{t-1}

$$\begin{aligned} (\theta_{t-1} | D_{t-1}) &\sim T_{n_{t-1}}(m_{t-1}, C_{t-1}) \\ (\tau_{t-1}^{-1} | D_{t-1}) &\sim \Gamma\left(\frac{n_{t-1}}{2}, \frac{n_{t-1} S_{t-1}}{2}\right) \end{aligned}$$

Para algunas m_{t-1} y C_{t-1} .

b) Distribución inicial para θ_t

$$\begin{aligned} (\theta_t | D_{t-1}) &\sim T_{n_{t-1}}(a_t, R_t) \\ (\tau^{-1} | D_{t-1}) &\sim \Gamma\left(\delta \frac{n_{t-1}}{2}, \delta \frac{n_{t-1} S_{t-1}}{2}\right) \end{aligned}$$

donde

$$a_t = G_t m_{t-1} \qquad R_t = G_t C_{t-1} G_t' + W_t$$

c) Pronóstico a un paso para y_t

$$(Y_t | D_{t-1}) \sim T_{\delta_{n_{t-1}}}(f_t, Q_t)$$

donde

$$f_t = F_t' a_t \qquad Q_t = F_t' R_t F_t + S_{t-1}$$

d) Distribución posterior para θ_t

$$\begin{aligned} (\theta_t | D_t) &\sim T_{n_{t-1}}(m_t, C_t) \\ (\tau^{-1} | D_t) &\sim \Gamma\left(\frac{n_t}{2}, \frac{n_t S_t}{2}\right) \end{aligned}$$

con las siguientes ecuaciones siendo aplicables en los casos donde sean requeridas

$$\begin{aligned} m_t &= a_t + A_t e_t & C_t &= \frac{S_t}{S_{t-1}} (R_t - A_t A_t' Q_t) \\ n_t &= n_{t-1} + 1 & d_t &= n_t S_t = d_{t-1} + \frac{e_t^2}{Q_t} \\ S_t &= \frac{d_t}{n_t} \end{aligned}$$

donde

$$A_t = R_t F_t' Q_t^{-1} \quad e_t = Y_t - f_t$$

4.5.3. Factores de descuento en la varianza de observación V_t de análisis retrospectivo

La Sección 4.4 ha mostrado que el proceso de *filtrado* auxilia en ocasiones donde es necesario mirar hacia atrás de las observaciones que se poseen para entender mejor el fenómeno en cuestión. Al igual que en el análisis prospectivo, en el proceso de *filtrado* al desconocer la varianza de observación y ser ésta variable en el tiempo es de gran utilidad el uso de un factor de descuento para modelar la precisión $\tau = V_t^{-1}$. Siguiendo la distribución del vector de estados de la Sección 4.4.2, se tiene que

$$(\theta_{T-k}|D_T) \sim T_{n_t} \left(a_T(-k), \frac{S_T}{S_{T-k}} R_T(-k) \right)$$

con

$$\begin{aligned} a_T(-k) &= m_{T-k} + B_{T-k}(a_T(-k+1) - a_{T-k+1}) \\ R_T(-k) &= C_{T-k} + B_{T-k}(R_T(-k+1) - R_{T-k+1})B_{T-k}' \\ B_T &= C_T G_{T+1}' R_{T+1}' \end{aligned}$$

West y Harrison (1997) hacen énfasis en que las distribuciones de filtrado para la precisión no son Gamma ya que la forma cerrada de actualización secuencial no es válida para el proceso de *filtrado*, pero se pueden aproximar por ésta distribución de la forma

$$p(\tau_{T-k}|D_t) \approx \Gamma\left(\frac{n_T(-k)}{2}, \frac{n_T(-k)S_T(-k)}{2}\right)$$

con

$$\begin{aligned} n_T(-k) &= (1 - \delta)n_{T-k} + \delta n_{T-k+1} \\ S_T(-k)^{-1} &= (1 - \delta)S_{T-k}^{-1} + \delta S_T(-k+1)^{-1} \end{aligned}$$

que se inicializan como $n_T(0) = n_T$ y $S_T(0) = S_T$

5. Apéndice

5.1. Modelos de Espacio Estado

Suponer una serie de tiempo $(Y_t)_{t \geq 1}$, se dice que es una *cadena de Markov* si para $t > 1$

$$P(y_t|y_1, y_2, \dots, y_{t-1}) = P(y_t|y_{t-1})$$

es decir, cada observación es dependiente únicamente de la observación anterior a ella.

En un modelo de *espacio estado* se supone una cadena de Markov no observable (θ_t) llamada proceso de estado y que Y_t es una medida imprecisa de θ_t . En la mayoría de aplicaciones estadísticas y económicas, (θ_t) es una variable latente, se puede pensar como una serie de tiempo auxiliar que facilita el cálculo de la distribución de probabilidad de la serie de tiempo observable, (Y_t)

Un modelo de *espacio estado* formalmente consiste de dos series de tiempo: $(\theta_t : t = 0, 1, \dots) \in \mathbb{R}^p$ y $(Y_t : t = 0, 1, \dots) \in \mathbb{R}^m$ que satisfacen

- (θ_t) es una cadena de Markov.
- Condicionadas en (θ_t) , las (Y_t) son independientes y Y_t depende únicamente de θ_t .

La consecuencia de éstas suposiciones es que un modelo de *espacio estado* se puede definir completamente por la distribución inicial $P(\theta_0)$ y las densidades condicionales $P(\theta_t|\theta_{t-1})$ y $P(y_t|\theta_t)$, $t \geq 1$, donde para cada $t > 0$,

$$P(\theta_0, \theta_1, \dots, \theta_t, y_0, y_1, \dots, y_t) = P(\theta_0) \prod_{j=1}^t P(\theta_j|\theta_{j-1})P(y_j|\theta_j)$$

a través de ésta fórmula se puede derivar cualquier otra distribución a través de condicionamiento o marginalización.