

DEDUCCION DEL FILTRO DE KALMAN EN EL CASO DE MODELOS DE ESTADOS GAUSSIANOS SINGULARES

por

Fabio H. Nieto*

Resumen

Nieto, F.H.: Dedución del Filtro de Kalman en el caso de modelos de estados Gaussianos singulares. Rev. Acad. Colomb. Cienc. 18 (71): 539-544, 1993. ISSN 0370-3908.

En ciertos modelos de estados para series temporales, los términos de error tienen distribución multinormal singular. En este artículo se presenta la deducción matemática del Filtro de Kalman para estos modelos.

Palabras claves: Distribución multiformal singular, modelo de estados, Filtro de Kalman.

Abstract

In some state space models for time series, the error terms have singular multinormal distribution. In this paper the mathematical deduction of the Kalman Filter for these models is presented.

Key words: Singular multinormal distribution, state space models, Kalman Filter.

1. Introducción

Un modelo de estados¹ para una serie cronológica multivariada $\{Y_t\}$ de dimensión n consiste del siguiente par de ecuaciones

$$\left. \begin{aligned} Y_t &= Z_t \alpha_t + \varepsilon_t \\ \alpha_t &= T_t \alpha_{t-1} + \eta_t \end{aligned} \right\} \quad t = 1, 2, \dots \quad (1.1)$$

donde para cada $t \geq 1$, α_t es un vector (no observable) de dimensión m , Z_t y T_t son matrices conocidas de dimensiones $n \times m$ y $m \times m$, respectivamente, y $\{\varepsilon_t\}$ y $\{\eta_t\}$ son sucesiones de vectores aleatorios de dimensiones n y m , respectivamente.

Las hipótesis estadísticas usuales para el modelo (1.1) son las siguientes:

- i) Las sucesiones $\{\varepsilon_t\}$ y $\{\eta_t\}$ son cada una mutuamente independientes y además $\varepsilon_t \sim N(0, H_t)$, $\eta_t \sim N(0, Q_t)$ para cada $t \geq 1$.
- ii) $E(\varepsilon_t \eta_s^T) = 0$ para cada $t, s \geq 1$. Aquí y en lo sucesivo la " ' " denotará transposición de matrices.
- iii) $\alpha_0 \sim N(a_0, P_0)$, y para cada $t \geq 1$, α_0 es independiente estocásticamente de ε_t y η_t .

Se supone además que H_t y Q_t para $t \geq 1$ y a_0, P_0 son conocidos.

En la mayoría de artículos y textos sobre el modelo de estados y el Filtro de Kalman se asume que las matrices H_t y Q_t son definidas positivas, puede consultarse por ejemplo Meinhold y Singpurwalla (1983), West y Harrison (1989) o Harvey (1989) que al parecer son los más sobresalientes en el ámbito estadístico. En estos trabajos el Filtro de Kalman se deduce usando la teoría sobre la distribución multinormal convencional. En algunos (pocos además) se permite que esas matrices sean definidas no negativas, es el

* Profesor Asociado, Departamento de Matemáticas y Estadística, Universidad Nacional de Colombia, Santafé de Bogotá, D.C.

1. En este artículo la frase inglesa state space model será traducida como modelo de estados.

caso de Anderson y Moore (1979), Kohn y Ansley (1983) y Catlin (1989) y la deducción del Filtro de Kalman se hace utilizando la teoría de espacios con producto interno. Harvey (1989) afirma que para cubrir el caso de matrices H_t y Q_t definidas no negativas se pueden reemplazar las matrices inversas que aparecen en las ecuaciones clásicas (covarianzas definidas positivas) por inversas generalizadas. Sin embargo no se indica allí una justificación sobre la afirmación, ni se propone una referencia donde se pueda consultar la demostración de esa proposición. A juicio del autor del presente trabajo, esa es una deficiencia de tipo didáctico de su texto. Una afirmación análoga a la de Harvey (1989) aparece en el libro de Aoki (1990), Secc. 5.2, con la misma deficiencia destacada antes.

El propósito del presente trabajo es presentar una deducción del Filtro de Kalman en el caso en que los vectores aleatorios ε_t y η_t tengan distribución multinormal singular, lo que implica que H_t y Q_t sean definidas no negativas. De esta forma se evita el uso de la teoría de espacios con producto interno y se suple la deficiencia didáctica del trabajo de Harvey.

El trabajo está organizado de la siguiente manera: En la sección 2 se presentan algunos resultados básicos sobre la distribución multinormal singular con sus demostraciones. Algunos de estos resultados sólo se citan sin demostración o se omiten, en algunos textos clásicos sobre Análisis Estadístico Multivariante, por ejemplo Mardia y otros (1979) o Anderson (1984). Esto se puede considerar como otro hecho destacable del presente trabajo. En la sección 3 se presenta la deducción del filtro de Kalman.

2. Algunos resultados sobre la distribución multinormal singular.

En esta sección se sigue, aunque no necesariamente se repite, el enfoque de Anderson (1984) sobre la distribución multinormal singular (o degenerada).

Definición 1. Un vector aleatorio X de dimensión $p > 1$ con $E(X) = \mu$ y $\text{Var}(X) = \Sigma$ tiene distribución multinormal *singular* si existe una matriz A de dimensión $p \times r$, un vector aleatorio Y de dimensión r con distribución multinormal y un vector λ de dimensión p , con $r(< p)$ el rango de Σ , tales que $X = AY + \lambda$.

Por notación se escribe $X \sim N_S(\mu, \Sigma)$ para indicar que X tiene distribución multinormal singular con media μ y matriz de covarianzas Σ .

En lo sucesivo y a menos que se diga lo contrario, la dimensión de X y el rango de Σ se seguirán tomando iguales a p y r , respectivamente.

Algunas observaciones sobre la definición anterior son pertinentes:

a) En general se podría definir la distribución multinormal en esa forma pues en el caso Σ de rango completo se puede tomar $A = I_p$, $Y = X$ y $\lambda = 0$. De hecho Anderson (1984) lo hace así.

b) Y tiene distribución multinormal significa que $T = \text{Var}(Y)$ es definida positiva y su densidad está dada por

$$f(y) = k \exp\{-(1/2)(y - v)'T^{-1}(y - v)\}, y \in \mathbb{R}^n,$$

donde k es una constante y $E(Y) = v$.

c) La relación entre μ y v , y entre Σ y T está dada por $\mu = Av + \lambda$ y $\Sigma = ATA'$.

Definición 2. Al vector Y de la definición 1 se le denomina la *parte no singular* de X .

Definición 3. Sean X_1 y X_2 vectores aleatorios con distribución multinormal *singular* cada uno. X_1 y X_2 se denominan independientes (estocásticamente) si sus partes no singulares son independientes estocásticamente.

A continuación se presentan los resultados que serán utilizados en la sección 3.

Teorema 2.1 Si $X \sim N_S(\mu, \Sigma)$ entonces $Z = DX \sim N_S(D\mu, D\Sigma D')$ donde D es una matriz de dimensión $q \times p$ con $q \leq p$.

Demostración. Ver Anderson (1984), pág. 33.

Teorema 2.2 Si $X \sim N_S(\mu, \Sigma)$, entonces la función característica de la distribución de X está dada por

$$\phi(t) = \exp(it'\mu - (1/2)t'\Sigma t), t \in \mathbb{R}^p$$

donde i es el número complejo tal que $i^2 = -1$.

Demostración. Como $X \sim N_S(\mu, \Sigma)$ entonces existen una matriz A con el número de columnas igual al rango de Σ , un vector aleatorio $Y \sim N(v, T)$ y un vector λ tal que $X = AY + \lambda$.

Entonces,

$$\begin{aligned} \phi(t) &= E[\exp(it'X)] \\ &= E\{\exp[it'(AY + \lambda)]\} \\ &= E\{\exp(it'AY) \exp(it'\lambda)\} \\ &= \exp(it'\lambda) \exp(it'Av - (1/2)t'ATA't) \\ &= \exp(it'\mu - (1/2)t'\Sigma t) \end{aligned}$$

para cada $t \in \mathbb{R}^p$.

Se recuerda en este punto que toda distribución de probabilidad está *completamente identificada* por su función característica.

Lema 2.3 Si $X_1 \sim N_S(\mu_1, \Sigma_1)$, $X_2 \sim N_S(\mu_2, \Sigma_2)$ y X_1 y X_2 son independientes entonces $Z = X_1 + X_2 \sim N_S(\mu_1 + \mu_2, \Sigma_1 + \Sigma_2)$.

Demostración. De las hipótesis se obtiene que

$$X_1 = A_1 Y_1 + \lambda_1 \text{ y } X_2 = A_2 Y_2 + \lambda_2$$

para algunas matrices A_1, A_2 , algunos vectores λ_1, λ_2 y $Y_i \sim N(\mu_i, \Sigma_i)$, $i = 1, 2$.

Entonces,

$$\begin{aligned} \phi_Z(t) &= E\{\exp[it'(X_1 + X_2)]\} \\ &= E\{\exp[it'X_1 + it'X_2]\} \\ &= \exp[it'(\lambda_1 + \lambda_2)] E\{\exp[it'A_1 Y_1] \exp[it'A_2 Y_2]\} \\ &= \exp[it'(\lambda_1 + \lambda_2)] \exp[it'A_1 v_1 - (1/2)t'A_1 T_1 A_1' t] \\ &\quad \exp[it'A_2 v_2 - (1/2)t'A_2 T_2 A_2' t] \\ &= \exp[it'(\mu_1 + \mu_2) - (1/2)t'(\Sigma_1 + \Sigma_2)t] \end{aligned}$$

Por lo tanto, $Z = X_1 + X_2 \sim N_S(\mu_1 + \mu_2, \Sigma_1 + \Sigma_2)$.

Teorema 2.4 Si X y Y son vectores aleatorios con distribución multinormal singular cada uno y X y Y son independientes entonces $Z = DX + EY$ tiene distribución multinormal singular donde D y E son de dimensión $m \times p$ y $n \times q$, respectivamente, con $m \leq p$, $n \leq q$ y q es la dimensión de Y .

Demostración. Por el teorema 2.1, DX y EY tienen distribución multinormal singular. Sean $X_1 = DX$ y $X_2 = EY$. Se puede demostrar que las partes no singulares de X_1 y X_2 son independientes estocásticamente, luego X_1 y X_2 son independientes. Aplicando el Lema 2.3 se obtiene que $Z = X_1 + X_2$ tiene distribución multinormal singular.

Corolario. Si X_1, \dots, X_k tienen distribución multinormal singular y sus partes no singulares conforman un conjunto mutuamente independiente entonces $Z = D_1 X_1 + \dots + D_k X_k$ tiene distribución multinormal singular para todo $k \geq 1$ donde A_i es una matriz de dimensión $m_i \times p_i$, $m_i \leq p_i$, $i = 1, \dots, k$.

Demostración. La demostración se obtiene procediendo por inducción sobre k .

Definición 4. Una sucesión $\{X_i\}$ de vectores aleatorios se denomina *mutuamente independiente* si para cualquier subconjunto finito el correspondiente conjunto de partes no singulares es mutuamente independiente.

Teorema 2.5 El vector aleatorio X tiene distribución multinormal singular si y solo si para cada $t \in \mathbb{R}^p$, $t'X$ tiene distribución normal o $t'X$ tiene distribución degenerada.

Demostración. Supóngase que $X \sim N_S(\mu, \Sigma)$. Sea $t \in \mathbb{R}^p$. Existen dos posibilidades acerca de t : (a) $t'\Sigma t > 0$,

o, (b) $t'\Sigma t = 0$. En el caso (a), $t'X \sim N(t'\mu, t'\Sigma t)$, puesto que su función característica está dada por

$$\begin{aligned} \phi_{t'X}(s) &= E\{\exp[is(t'X)]\} \\ &= E\{\exp[i(st')X]\} \\ &= \phi_X(st) \\ &= \exp[ist'\mu - (1/2)(st')\Sigma(st)] \\ &= \exp[is(t'\mu) - (1/2)s^2(t'\Sigma t)] \end{aligned}$$

para cada $s \in \mathbb{R}$.

En el caso (b), $t'X = t'\mu$ con probabilidad 1, por tanto $t'X$ tiene distribución degenerada.

Recíprocamente supóngase que para cada $t \in \mathbb{R}^p$, $t'X$ está distribuido normalmente ó $t'X$ tiene distribución degenerada. Sean $\mu = E(X)$ y $\Sigma = \text{Var}(X)$. En el primer caso $\text{Var}(t'X) = t'\Sigma t > 0$ y para este t se tiene que

$$\begin{aligned} \phi_X(t) &= E[\exp(it'X)] \\ &= E[\exp(i1(t'X))] \\ &= \phi_{t'X}(1) \\ &= \exp(it'\mu - (1/2)t'\Sigma t) \end{aligned}$$

Si $t'X$ tiene distribución degenerada entonces $t'\Sigma = \text{Var}(t'X) = 0$ y su función característica está dada por

$$\phi_{t'X}(s) = \exp[is(t'\mu)], \quad s \in \mathbb{R}.$$

Entonces para este t

$$\begin{aligned} \phi_X(t) &= \phi_{t'X}(1) \\ &= \exp(it'\mu) \\ &= \exp(it'\mu - (1/2)t'\Sigma t). \end{aligned}$$

Teniendo en cuenta que

$$\mathbb{R}^p = \{t : t'\Sigma t > 0\} \cup \{t : t'\Sigma t = 0\},$$

donde la unión es disyunta, se concluye que $X \sim N_S(\mu, \Sigma)$.

El siguiente es un resultado sobre distribuciones univariadas.

Teorema 2.6. Si X y Y son variables aleatorias tales que $X \sim N(\mu, \sigma^2)$ y Y tiene distribución degenerada entonces $X + Y \sim N(\mu + c, \sigma^2)$ donde $P_T(Y = c) = 1$.

Demostración. Sea Ω el espacio muestral en el que X y Y están definidas. La función de distribución acumulada de $X + Y$ está dada por

$$\begin{aligned} F_{X+Y}(z) &= P_T(X + Y \leq z) \\ &= P_T(\{\omega \in \Omega : X(\omega) + Y(\omega) \leq z\}) \end{aligned}$$

$$= \Pr(\{\omega \in \Omega : X(\omega) \leq z - c\})^2$$

ya que $P_r(\{\omega \in \Omega : Y(\omega) \neq c\}) = 0$. Ahora,

$$P_r(X \leq z - c) = \int_{-\infty}^{z-c} \frac{1}{\sqrt{2\pi} \sigma} \exp\left\{-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2\right\} dx.$$

Haciendo $y = x + c$, la anterior integral se convierte en

$$\int_{-\infty}^z \frac{1}{\sqrt{2\pi} \sigma} \exp\left\{-\frac{1}{2}\left(\frac{y - \mu - c}{\sigma}\right)^2\right\} dy.$$

En consecuencia $X + Y \sim N(\mu + c, \sigma^2)$.

3. Deducción del Filtro de Kalman.

Considérese el modelo de estados (1.1) pero ahora con las siguientes hipótesis estadísticas:

- i) $\epsilon_t \sim N_s(0, H_t)$, $\eta_t \sim N_s(0, H_t)$ para cada $t \geq 1$.
- ii) Las sucesiones $\{\epsilon_t\}$ y $\{\eta_t\}$ son cada una mutuamente independientes y ϵ_t y $\{\eta_s\}$ mutuamente independientes (estocásticamente) para cada $t \geq 1$.
- iii) $\alpha_0 \sim N_s(a_0, P_0)$ y α_0 , $\{\epsilon_t\}$ y $\{\eta_t\}$ mutuamente independientes.

En lo sucesivo el modelo (1.1) con las anteriores hipótesis se llamará el *modelo de estados gaussiano singular*.

Teorema. Si una serie cronológica multivariada $\{Y_t\}$ obedece el modelo de estados gaussiano singular entonces la distribución de α_t dadas las observaciones Y_1, \dots, Y_t es multinormal singular con media

$$a_t = a_{t|t-1} + P_{t|t-1} Z_t' F_t^- (Y_t - Z_t a_{t|t-1}) \quad (3.1)$$

y matriz de covarianzas

$$P_t = P_{t|t-1} - P_{t|t-1} Z_t' F_t^- Z_t P_{t|t-1}, \quad (3.2)$$

para cada $t \geq 1$, donde

$$a_{t|t-1} = T_t a_{t-1}, \quad (3.3)$$

$$P_{t|t-1} = T_t P_{t-1} T_t' + Q_t, \quad (3.4)$$

$$F_t = Z_t P_{t|t-1} Z_t' + H_t \quad (3.5)$$

ya F_t^- indica la pseudoinversa de F_t .

Demostración. Se procede por inducción sobre t . El vector de estado en $t = 1$ está dado por

$$\alpha_1 = T_1 \alpha_0 + \eta_1.$$

Luego por el teorema 2.4, α_1 tiene distribución multinormal singular con media $a_{1|0} = T_1 a_0$ y matriz de covarianzas $P_{1|0} = T_1 P_0 T_1' + Q_1$. La notación $1|0$ indica que la distribución de α_1 depende (en un sentido funcional) de la información en $t = 0$ (representada únicamente por el conocimiento de la distribución de α_0).

Con el fin de obtener la distribución de α_1 condicional sobre el valor que tome Y_1 se debe tener en cuenta que:

$$\begin{aligned} Y_1 &= Z_1 \alpha_1 + \epsilon_1 \\ &= Z_1 (T_1 \alpha_0 + \eta_1) + \epsilon_1 \\ &= (Z_1 T_1) \alpha_0 + Z_1 \eta_1 + \epsilon_1 \end{aligned}$$

Sea $V_1 = (\alpha_1' Y_1)'$ y sea X un vector de dimensión $m + n$ tal que $X = (X_1' X_2)'$ donde X_1 y X_2 son de dimensión m y n respectivamente. Entonces:

$$\begin{aligned} X' V_1 &= (X_1' X_2') \begin{pmatrix} \alpha_1 \\ Y_1 \end{pmatrix} \\ &= X_1' \alpha_1 + X_2' Y_1 \\ &= (X_1' T_1 + X_2' Z_1 T_1) \alpha_0 + (X_1' + X_2' Z_1) \eta_1 + X_2' \epsilon_1 \end{aligned}$$

Del teorema 2.5 y posiblemente el 2.6 se sigue que V_1 tiene distribución multinormal singular. La media esta dada por

$$\begin{bmatrix} a_{1|0} & (Z_1 a_{1|0})' \end{bmatrix}$$

y la matriz de covarianzas por

$$\begin{pmatrix} P_{1|0} & P_{1|0} Z_1' \\ Z_1 P_{1|0} & Z_1 P_{1|0} Z_1' + H_1 \end{pmatrix}.$$

Aplicando los resultados de Marsaglia (1964) se obtiene que la distribución de α_1 condicional sobre un valor particular de Y_1 es multinormal singular con media

$$a_1 = a_{1|0} + P_{1|0} Z_1' F_1^- (Y_1 - Z_1 a_{1|0})$$

y matriz de covarianzas

$$P_1 = P_{1|0} - P_{1|0} Z_1' F_1^- Z_1 P_{1|0}$$

donde

$$F_1 = Z_1 P_{1|0} H_1' + H_1.$$

Ahora se supone que el teorema es válido para $t - 1$, es decir que $\alpha_{t-1} \sim N_s(a_{t-1}, P_{t-1})$, y se demuestra para t . Se puede demostrar que bajo las

² $\{\omega \in \Omega : X(\omega) + Y(\omega) \leq z\} = \{\omega \in \Omega : Y(\omega) = c, X(\omega) + Y(\omega) \leq z\} \cup \{\omega \in \Omega : Y(\omega) \neq c, X(\omega) + Y(\omega) \leq z\}$
 Por tanto
 $P_r(\{\omega \in \Omega : X(\omega) + Y(\omega) \leq z\}) = P_r(\{\omega \in \Omega : Y(\omega) = c, X(\omega) + Y(\omega) \leq z\}) + P_r(\{\omega \in \Omega : Y(\omega) \neq c, X(\omega) + Y(\omega) \leq z\})$.

hipótesis del modelo tanto α_{t-1} y η_t como α_t y ε_t son independientes estocásticamente*. Así que procediendo como en el caso $t = 1$ se obtiene la demostración del teorema.

Algunas observaciones son importantes en este punto:

- (a) A las ecuaciones (3.3) y (3.4) se les llama las ecuaciones de predicción y a las ecuaciones (3.1) y (3.2) las de actualización y estas constituyen estrictamente el Filtro de Kalman.
- (b) La notación $t-1$ indica que la distribución de α_t antes de observar Y_t depende determinísticamente de toda la información hasta $t-1$, representada por α_0, P_0 y los valores observados de Y_1, \dots, Y_{t-1} .
- (c) Bajo las hipótesis del modelo, antes de observar Y_1, \dots, Y_{t-1} , este conjunto de vectores aleatorios y η_t son mutuamente independientes.
- (d) Se puede demostrar que a_t es el estimador óptimo, en el sentido de error cuadrático medio, de α_t dados Y_1, \dots, Y_t y que P_t es la matriz de covarianzas del error de estimación.

4. Conclusión

La única modificación que se debe realizar al Filtro de Kalman convencional para obtener su versión análoga en el caso de modelos de estados gaussianos singulares es la de reemplazar las inversas convencionales por las pseudoinversas correspondientes. La comprobación de esta afirma

ción se alcanzó utilizando teoría sobre distribuciones multinormales singulares. Esta vía evita el uso de la teoría de espacios con producto interno para deducir el Filtro cuando sólo se asume que $H_t = \text{Var}(\varepsilon_t)$ y $Q_t = \text{Var}(\eta_t)$ son definidas no negativas.

Referencias

- Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, John Wiley and Sons, Inc, New York.
- Anderson, D. O. and Moore, J. B. (1979). *Optimal Filtering*, Prentice-Hall, Inc., New Jersey (U. S. A.).
- Aoki, M. (1990). *State Space Modeling of Time Series*, Springer-Verlag, Berlín.
- Catlin, D. (1989). *Estimation, Control and the discrete Kalman Filter*, Springer-Verlag, Berlín.
- Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge.
- Kohn, R. and Ansley, C. F. (1983). Fixed interval estimation in state space models when some of the data are missing or aggregated, *Biometrika*, Vol. 70, 3, pp 683 - 8.
- Mardia, K. V., Kent, J. J. and Bibby, J. M. (1979). *Multivariate Analysis*, Academic Press, Inc., London.
- Marsaglia, G. (1964). Conditional Means and Covariances of Normal Variables with Singular Covariance Matrix, *Journal of the American Statistical Association*, 59, pp. 1203 - 1204.
- Meinhold, R. J. and Singpurwalla, N. D. (1983). Understanding the Kalman Filter, *The American Statistician*, Vol. 37, N°2.
- West, M. and Harrison, P. J. (1989). *Bayesian Forecasting and Dynamic Models*, Springer-Verlag, Berlín.

* α_i es una combinación lineal de $\alpha_0, \eta_1, \dots, \eta_i$ para cada $i \geq 1$.