

## Filtro de Kalman

Alan Ledesma Arista.<sup>1</sup>

---

<sup>1</sup>[alan.ledesma@bcrp.gob.pe](mailto:alan.ledesma@bcrp.gob.pe)

- 1 Motivación e introducción
- 2 Representación de espacio de estados
- 3 Filtrado y Estimación
- 4 Introducción: modelos de espacio de los estados no lineales o no gaussianos

# Introducción

- Texto principal: [Durbin y Koopman 2012](#)
- Texto secundario: [Kim y Nelson 1999 \(cap. 3\)](#)
- El principal uso de la metodología de Filtro de Kalman es la estimación de la verosimilitud del modelo
  - Así se puede estimar parámetros y filtrar variables subyacentes del modelo
  - Para ello, el modelo debe estar escrito en su Representación de Espacio de Estados (REE)
- La macroeconomía moderna favorece la rigurosidad microeconómica  $\Rightarrow$  *modelos (semi) estructurales* en los cuales se suelen definir variables teóricas comúnmente no medibles, por ejemplo
  - Producto Natural,
  - Desempleo Natural,
  - Primas de riesgo
  - Expectativas ...
- El Filtro de Kalman (KF por sus siglas en inglés) ofrece una metodología formal, eficiente y óptima para evaluar empíricamente estos modelos.

# Representación de espacio de estados

Todo **modelo lineal de series de tiempo** puede escribirse como una REE, éste consta de **dos sistemas de ecuaciones**:

**1<sup>er</sup> Ecuaciones de espacio** (o de medida): Ecuaciones que describen la relación entre las variables observables (variables de medida) y no observables (variables de estado).

Ejemplo:

$$y_t = \bar{y}_t + \hat{y}_t + \varepsilon_t,$$

donde  $y_t$ ,  $\bar{y}_t$  y  $\hat{y}_t$  son el PBI observado, tendencial y la brecha de producto, respectivamente. La variable  $\varepsilon_t$  es un error de medida.

# Representación de espacio de estados

Todo **modelo lineal de series de tiempo** puede escribirse como una REE, éste consta de **dos sistemas de ecuaciones**:

**2<sup>do</sup> Ecuaciones de estado** (o de transición): Ecuaciones que describen la dinámica de las variables no observables (variables de estado)

Ejemplo:

$$\bar{y}_t = \bar{y}_{t-1} + \bar{\varepsilon}_t,$$

$$\hat{y}_t = \phi_1 \hat{y}_{t-1} + \phi_2 \hat{y}_{t-1} + \hat{\varepsilon}_t,$$

donde  $\bar{\varepsilon}_t$  y  $\hat{\varepsilon}_t$  son innovaciones

# Representación de espacio de estados

Suponga que  $\mathbf{y}_t$  ( $n_y \times 1$ ),  $\boldsymbol{\alpha}_t$  ( $s \times 1$ ) y  $\mathbf{x}_t$  ( $n_x \times 1$ ) son los vectores de variables de medida, de estado y exógenas, respectivamente. Entonces el sistema de espacio de estados es:

$$\mathbf{y}_t = \mathbf{Z}\boldsymbol{\alpha}_t + \mathbf{B}\mathbf{x}_t + \boldsymbol{\varepsilon}_t, \text{ con } \boldsymbol{\varepsilon}_t \stackrel{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{H})$$

$$\boldsymbol{\alpha}_t = \mathbf{c} + \mathbf{T}\boldsymbol{\alpha}_{t-1} + \mathbf{R}\boldsymbol{\eta}_t, \text{ con } \boldsymbol{\eta}_t \stackrel{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{Q})$$

donde  $\mathbf{Z}$ ,  $\mathbf{B}$ ,  $\mathbf{H}$ ,  $\mathbf{c}$ ,  $\mathbf{T}$ ,  $\mathbf{R}$  y  $\mathbf{Q}$  son matrices de coeficientes con dimensiones  $n_y \times s$ ,  $n_y \times n_x$ ,  $n_y \times n_y$ ,  $s \times 1$ ,  $s \times s$ ,  $s \times r$  y  $r \times r$ , respectivamente. Asimismo,  $\boldsymbol{\varepsilon}_t$  ( $n_y \times 1$ ) y  $\boldsymbol{\eta}_t$  ( $r \times 1$ ) son el vector de errores de medida y el de innovaciones de estado, respectivamente.

# Representación de espacio de estados

Del ejemplo mencionado, se tendría que  $\mathbf{y}_t = y_t$ ,  $\boldsymbol{\alpha}_t = [\bar{y}_t \ \hat{y}_t \ \hat{y}_{t-1}]'$  y  $\mathbf{x}_t = \emptyset$ , así la REE es:

$$y_t = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \bar{y}_t \\ \hat{y}_t \\ \hat{y}_{t-1} \end{bmatrix} + \sigma_\varepsilon \varepsilon_{y,t}$$

$$\begin{bmatrix} \bar{y}_t \\ \hat{y}_t \\ \hat{y}_{t-1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \phi_1 & \phi_2 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \bar{y}_{t-1} \\ \hat{y}_{t-1} \\ \hat{y}_{t-2} \end{bmatrix} + \begin{bmatrix} \sigma_{\bar{\varepsilon}} & 0 \\ 0 & \sigma_{\hat{\varepsilon}} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \bar{\varepsilon}_t \\ \hat{\varepsilon}_t \end{bmatrix}$$

# Representación de espacio de estados

Ejemplo: VAR(p):

- Sea el modelo lineal  $\mathbf{y}_t = \mathbf{b}_0 + \sum_{\ell=1}^p \mathbf{b}_\ell \mathbf{y}_{t-\ell} + \mathbf{a}_t$  con  $k$  variables
- Se puede definir  $\mathbf{y}_t = \mathbf{y}_t$ ,  $\boldsymbol{\alpha}_t = [\mathbf{y}_t \dots \mathbf{y}_{t-p+1}]'$ ,  $\boldsymbol{\varepsilon}_t = \mathbf{x}_t = \emptyset$  y  $\boldsymbol{\eta}_t = \mathbf{a}_t$  entonces

$$\mathbf{y}_t = \begin{bmatrix} \mathbf{I}_k & \mathbf{0}_{k,k(p-1)} \end{bmatrix} \boldsymbol{\alpha}_t$$

$$\boldsymbol{\alpha}_{t+1} = \begin{bmatrix} \mathbf{b}_0 \\ \mathbf{0}_{k,1} \\ \mathbf{0}_{k,1} \\ \vdots \\ \mathbf{0}_{k,1} \end{bmatrix} + \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \dots & \mathbf{b}_{p-1} & \mathbf{b}_p \\ \mathbf{I}_k & \mathbf{0}_{k,k} & \dots & \mathbf{0}_{k,k} & \mathbf{0}_{k,k} \\ \mathbf{0}_{k,k} & \mathbf{I}_k & \dots & \mathbf{0}_{k,k} & \mathbf{0}_{k,k} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0}_{k,k} & \mathbf{0}_{k,k} & \dots & \mathbf{I}_k & \mathbf{0}_{k,k} \end{bmatrix} \boldsymbol{\alpha}_{t-1} + \begin{bmatrix} \mathbf{I}_k \\ \mathbf{0}_{k(p-1),k} \end{bmatrix} \boldsymbol{\eta}_t,$$

con  $\boldsymbol{\eta}_t \stackrel{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$



# Representación de espacio de estados

Ejemplo: VARMA(p,q):

- Sea el modelo lineal  $\mathbf{y}_t = \mathbf{b}_0 + \sum_{\ell=1}^p \mathbf{b}_\ell \mathbf{y}_{t-\ell} + \mathbf{a}_t + \sum_{\ell=1}^q \mathbf{c}_\ell \mathbf{a}_{t-\ell}$  con  $k$  variables
- Se puede definir  $\mathbf{y}_t = \mathbf{y}_t$ ,  $\boldsymbol{\alpha}_t = [\mathbf{y}_t \dots \mathbf{y}_{t-p+1} \quad \mathbf{a}_t \dots \mathbf{a}_{t-q+1}]'$ ,  $\varepsilon_t = \mathbf{x}_t = \emptyset$  y  $\boldsymbol{\eta}_t = \mathbf{a}_t$  entonces

$$\mathbf{y}_t = [\mathbf{I}_k \quad \mathbf{0}_{k, k(p+q-1)}] \boldsymbol{\alpha}_t$$

$$\boldsymbol{\alpha}_{t+1} = \mathbf{b}_0 + \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \dots & \mathbf{b}_{p-1} & \mathbf{b}_p & \mathbf{c}_1 & \mathbf{c}_2 & \dots & \mathbf{c}_{q-1} & \mathbf{c}_q \\ \mathbf{I}_k & \mathbf{0}_{k,k} & \dots & \mathbf{0}_{k,k} & \mathbf{0}_{k,k} & \mathbf{0}_{k,k} & \mathbf{0}_{k,k} & \dots & \mathbf{0}_{k,k} & \mathbf{0}_{k,k} \\ \mathbf{0}_{k,k} & \mathbf{I}_k & \dots & \mathbf{0}_{k,k} & \mathbf{0}_{k,k} & \mathbf{0}_{k,k} & \mathbf{0}_{k,k} & \dots & \mathbf{0}_{k,k} & \mathbf{0}_{k,k} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0}_{k,k} & \mathbf{0}_{k,k} & \dots & \mathbf{I}_k & \mathbf{0}_{k,k} & \mathbf{0}_{k,k} & \mathbf{0}_{k,k} & \dots & \mathbf{0}_{k,k} & \mathbf{0}_{k,k} \\ \mathbf{0}_{k,k} & \mathbf{0}_{k,k} & \dots & \mathbf{0}_{k,k} & \mathbf{0}_{k,k} & \mathbf{0}_{k,k} & \mathbf{0}_{k,k} & \dots & \mathbf{0}_{k,k} & \mathbf{0}_{k,k} \\ \mathbf{0}_{k,k} & \mathbf{0}_{k,k} & \dots & \mathbf{0}_{k,k} & \mathbf{0}_{k,k} & \mathbf{I}_k & \mathbf{0}_{k,k} & \dots & \mathbf{0}_{k,k} & \mathbf{0}_{k,k} \\ \mathbf{0}_{k,k} & \mathbf{0}_{k,k} & \dots & \mathbf{0}_{k,k} & \mathbf{0}_{k,k} & \mathbf{0}_{k,k} & \mathbf{I}_k & \dots & \mathbf{0}_{k,k} & \mathbf{0}_{k,k} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0}_{k,k} & \mathbf{0}_{k,k} & \dots & \mathbf{0}_{k,k} & \mathbf{0}_{k,k} & \mathbf{I}_k & \mathbf{0}_{k,k} & \dots & \mathbf{I}_k & \mathbf{0}_{k,k} \end{bmatrix} \boldsymbol{\alpha}_{t-1}$$

$$+ \begin{bmatrix} \mathbf{I}_k \\ \mathbf{0}_{k(p-1),k} \\ \mathbf{I}_k \\ \mathbf{0}_{k(q-1),k} \end{bmatrix} \boldsymbol{\eta}_t, \text{ con } \boldsymbol{\eta}_t \stackrel{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$$

# Representación de espacio de estados

## Factores Dinámicos: Stock y Watson (1991)

Stock y Watson (1991) utilizan el KF para extraer un indicador de coincidencia económica que busca resumir el estado de la economía.

Idea: Los comovimientos de algunas variables macroeconómicas son guiados por un elemento común y subyacente: una variable no observada.

Variables:  $y_{it} = \Delta Y_{it} - \overline{\Delta Y_{it}}$  con  $i \in \{1, 2, 3, 4\}$  que corresponden a los logaritmos de producción industrial, ingreso personal neto, total de manufacturas y empleo no agro-industrial. Todas en diferencias y en desviaciones respecto de su media.

- El modelo lineal es

$$y_{it} = \gamma_i x_t + e_{it}, \quad \forall i \in \{1, \dots, 3\},$$

$$y_{4t} = \gamma_{40} x_t + \gamma_{41} x_{t-1} + \gamma_{42} x_{t-2} + \gamma_{43} x_{t-3} + e_{4t},$$

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + w_t, \quad w_t \sim \mathcal{N}(0, 1)$$

$$e_{it} = \rho_{i1} e_{it-1} + \rho_{i2} e_{it-2} + \epsilon_{it}, \quad \epsilon_{it} \sim \mathcal{N}(0, 1) \forall i \in \{1, \dots, 4\}$$

- ¿Cuál es la REE?

- Variables de medida:  $\mathbf{y}_t = [y_{1t}, y_{2t}, y_{3t}, y_{4t}]'$ , no hay errores de medida ni variables exógenas
- Variables de estado:  $\boldsymbol{\alpha}_t = [x_t, x_{t-1}, x_{t-2}, x_{t-3}, e_{1t}, e_{1t-1}, e_{2t}, e_{2t-1}, e_{3t}, e_{3t-1}, e_{4t}, e_{4t-1}]'$
- Innovaciones de estado:  $\boldsymbol{\eta}_t = [w_t, \epsilon_{1t}, \epsilon_{2t}, \epsilon_{3t}, \epsilon_{4t}]$

# Representación de espacio de estados

## Ejemplo: Modelo de tendencias comunes

- Sea el modelo lineal

$$y_t^* = \bar{y}_t^* + \hat{y}_t^* + \mu_t^* \quad (\text{PBI global})$$

$$y_t = \bar{y}_t + \hat{y}_t + \mu_t \quad (\text{PBI doméstico})$$

$$\hat{y}_t^* = \phi_{1,1,1}\hat{y}_{t-1}^* + \phi_{1,1,2}\hat{y}_{t-2}^* + \varepsilon_t^* \quad (\text{Ciclo global})$$

$$\hat{y}_t = \phi_{2,1,1}\hat{y}_{t-1} + \phi_{2,2,1}\hat{y}_{t-1}^* + \phi_{2,1,2}\hat{y}_{t-2} + \phi_{2,2,2}\hat{y}_{t-2}^* + \varepsilon_t \quad (\text{Ciclo doméstico})$$

$$\bar{y}_t^* = \bar{y}_{t-1}^* + \nu_t^* \quad (\text{Tendencia global})$$

$$\bar{y}_t = (1 - \theta_1)\bar{y}_{t-1} + \theta_1\bar{y}_{t-1}^* \quad (\text{Tendencia doméstica})$$

- ¿Cuál es la REE?

# Filtrado

Si se asume (por ahora) que las matrices  $\mathbf{Z}$ ,  $\mathbf{B}$ ,  $\mathbf{H}$ ,  $\mathbf{c}$ ,  $\mathbf{T}$ ,  $\mathbf{R}$  y  $\mathbf{Q}$  son conocidas, entonces se puede inferir la distribución y realización de los estados.

Si contamos con una muestra  $\{\mathbf{y}_t\}_{t=1}^T$ , el filtrado consta de tres pasos:

- Predicción
- Actualización
- Suavizamiento

# Filtrado

Tomemos la siguiente notación:

$$X_{i|s} = E[X_i | \Omega_s] \text{ con } \Omega_s \equiv \left\{ \{y_j\}_{j=1}^s, \mathbf{Z}, \mathbf{B}, \mathbf{H}, \mathbf{c}, \mathbf{T}, \mathbf{R}, \mathbf{Q} \right\}$$

$X_{i|s}$  es la esperanza condicional de  $X$  al momento  $i$  con información al momento  $s$

- $\mathbf{y}_{t|t-1}$ : predicción de los datos para el momento  $t$  con información hasta  $t - 1$
- $\mathbf{y}_{t|t} = \mathbf{y}_t$ : predicción de los datos para el momento  $t$  con información hasta  $t$
- $\boldsymbol{\alpha}_{t|t-1}$ : predicción de los estados para el momento  $t$  con información hasta  $t - 1$
- $\boldsymbol{\alpha}_{t|t}$ : predicción de los estados para el momento  $t$  con información hasta el momento  $t$   
(¿ $\boldsymbol{\alpha}_{t|t} = \boldsymbol{\alpha}_t$ ?)

# Predicción

Según la REE, la predicción de ' $\alpha$ ' e ' $y$ ' son

$$1.1. \alpha_{t|t-1} = \mathbf{c} + \mathbf{T}\alpha_{t-1|t-1}$$

$$1.2. \mathbf{y}_{t|t-1} = \mathbf{Z}\alpha_{t|t-1} + \mathbf{B}\mathbf{x}_t$$

$$1.3. \text{Error de predicción: } \epsilon_{t|t-1} = \mathbf{y}_t - \mathbf{y}_{t|t-1}$$

y sus respectivas varianzas

$$1.4. \Sigma_{t|t-1}^{\alpha} = \mathbf{T}\Sigma_{t-1|t-1}^{\alpha}\mathbf{T}' + \mathbf{R}\mathbf{Q}\mathbf{R}'$$

$$1.5. \Sigma_{t|t-1}^{\epsilon} = \mathbf{Z}\Sigma_{t|t-1}^{\alpha}\mathbf{Z}' + \mathbf{H}$$

$$\bullet \text{COV}_{t-1}(\alpha_{t|t-1}, \epsilon_{t|t-1}) = \Sigma_{t|t-1}^{\alpha}\mathbf{Z}'$$

entonces

$$\begin{bmatrix} \alpha_t \\ \epsilon_t \end{bmatrix} \Big| \Omega_{t-1} \sim \mathcal{N} \left( \begin{bmatrix} \alpha_{t|t-1} \\ \epsilon_{t|t-1} \end{bmatrix}, \begin{bmatrix} \Sigma_{t|t-1}^{\alpha} & \Sigma_{t|t-1}^{\alpha}\mathbf{Z}' \\ \mathbf{Z}(\Sigma_{t|t-1}^{\alpha})' & \Sigma_{t|t-1}^{\epsilon} \end{bmatrix} \right)$$

# Actualización

## Lema 1: Propiedad de la distribución Normal

Si  $\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_x & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}'_{xy} & \boldsymbol{\Sigma}_y \end{bmatrix} \right)$  entonces

$$\mathbf{x}|\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_y^{-1}(\mathbf{y} - \boldsymbol{\mu}_y), \boldsymbol{\Sigma}_x - \boldsymbol{\Sigma}_{xy}(\boldsymbol{\Sigma}_y)^{-1}\boldsymbol{\Sigma}'_{xy})$$

- En base a esta propiedad y debido a que  $\mathbf{y}_t$  es conocido, se puede actualizar las predicciones en la diapositiva anterior
- Definamos la Ganancia de Kalman ( $\mathbf{K}_t$ ) como la proporción de corrección requerida en los estados (equivalente a  $\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_y^{-1}$  en la expresión de arriba). Entonces la actualización consta de los siguientes pasos

2.1.  $\mathbf{K}_t = \boldsymbol{\Sigma}_{t|t-1}^\alpha \mathbf{Z}' (\boldsymbol{\Sigma}_{t|t-1}^\epsilon)^{-1}$

2.2.  $\boldsymbol{\alpha}_{t|t} = \boldsymbol{\alpha}_{t|t-1} + \mathbf{K}_t \boldsymbol{\epsilon}_{t|t-1}$

2.3.  $\boldsymbol{\Sigma}_{t|t}^\alpha = \boldsymbol{\Sigma}_{t|t-1}^\alpha - \mathbf{K}_t \mathbf{Z} \boldsymbol{\Sigma}_{t|t-1}^\alpha$

# Suavizamiento

- Para mejorar la precisión de la inferencia sobre  $\alpha_t$  se añade un tercer paso que utiliza toda la información de la muestra
- Una extensión del lema 1, nos indica que

$$\alpha_{t|T} = \alpha_{t|t-1} + \sum_{j=t}^T \text{cov}(\alpha_{t|t-1}, \epsilon_{j|j-1})(\Sigma_{j|j-1}^{\epsilon})^{-1} \epsilon_{j|j-1}$$

$$\Sigma_{t|T}^{\alpha} = \Sigma_{t|t-1}^{\alpha} - \sum_{j=t}^T \text{cov}(\alpha_{t|t-1}, \epsilon_{j|j-1})(\Sigma_{j|j-1}^{\epsilon})^{-1} \text{cov}(\alpha_{t|t-1}, \epsilon_{j|j-1})'$$

- Se puede demostrar que esta expresión se reduce a

$$3.1. \mathbf{S}_t \equiv \Sigma_{t|t}^{\alpha} \mathbf{T}' (\Sigma_{t+1|t}^{\alpha})^{-1}$$

$$3.1. \alpha_{t|T} = \alpha_{t|t} + \mathbf{S}_t (\alpha_{t+1|T} - \mathbf{T} \alpha_{t|t} - \mathbf{c})$$

$$3.2. \Sigma_{t|T}^{\alpha} = \Sigma_{t|t}^{\alpha} + \mathbf{S}_t (\Sigma_{\alpha, t+1|T} - \Sigma_{\alpha, t+1|t}) \mathbf{S}_t'$$

- El filtro de Kalman consiste en la recursión desde  $t = 1$  hasta  $t = T$  de los pasos [1.1. - 1.5.] y [2.1.-2.3.] y una recursión desde  $t = T$  hasta  $t = 1$  de los pasos [3.1.-3.2.]



$$\Sigma_{0|0}^{\alpha} - \mathbf{T}\Sigma_{\alpha,0|0}\mathbf{T}' - \mathbf{R}\mathbf{Q}\mathbf{R}' = 0$$

# Verosimilitud

Note que condicional a  $\mathbf{Z}$ ,  $\mathbf{B}$ ,  $\mathbf{H}$ ,  $\mathbf{c}$ ,  $\mathbf{T}$ ,  $\mathbf{R}$  y  $\mathbf{Q}$  y la data, se obtiene una trayectoria de estadoa y por tanto de residuos. Dada la normalidad de los residuos, se puede calcular la probabilidad de que la data en  $\{\mathbf{Y}_t\}_{t=1}^T$  provenga del modelo en estado espacio (verosimilitud).

Sea  $\theta$  el vector de coeficientes en la REE tal que  $\mathbf{Z}(\theta)$ ,  $\mathbf{B}(\theta)$ ,  $\mathbf{H}(\theta)$ ,  $\mathbf{c}(\theta)$ ,  $\mathbf{T}(\theta)$ ,  $\mathbf{R}(\theta)$  y  $\mathbf{Q}(\theta)$  determine el sistema. Entonces la función de verosimilitud es

$$\ln \mathcal{L}(\theta; \{\mathbf{y}_t\}_{t=1}^T) = \text{constante} - \frac{1}{2} \left( \sum_{t=1}^T \log \left| \boldsymbol{\Sigma}_{t|t-1}^{\epsilon} \right| + \boldsymbol{\epsilon}'_{t|t-1} (\boldsymbol{\Sigma}_{t|t-1}^{\epsilon})^{-1} \boldsymbol{\epsilon}_{t|t-1} \right)$$

## Estimación

Dos alternativas:

- **Frecuentista - Máxima verosimilitud:** Implementar algún algoritmo de optimización numérica para maximizar  $\ln \mathcal{L}(\cdot; \{\mathbf{y}_t\}_{t=1}^T)$

$$\hat{\theta} = \arg \max_{\theta} \ln \mathcal{L}(\theta; \{\mathbf{y}_t\}_{t=1}^T)$$

- **Bayesiana:** Note que  $p(\{\mathbf{y}_t\}_{t=1}^T | \theta) = \mathcal{L}(\theta; \{\mathbf{y}_t\}_{t=1}^T)$ , si se formula un prior  $p(\theta)$  entonces la distribución posterior es

$$p(\theta|\{\mathbf{y}_t\}_{t=1}^T) = \frac{p(\{\mathbf{y}_t\}_{t=1}^T|\theta)p(\theta)}{p(\{\mathbf{y}_t\}_{t=1}^T)} \propto p(\{\mathbf{y}_t\}_{t=1}^T|\theta)p(\theta)$$

1. Obtener  $\theta_{mode}$  y  $\Sigma_{mode}^{\theta}$  al maximizar numéricamente  $p(\theta|\{\mathbf{y}_t\}_{t=1}^T)$
2. Obtener 'muchas' realizaciones de  $p(\theta|\{\mathbf{y}_t\}_{t=1}^T)$  con la técnica de muestreo de Metropoli-Hasting
3. Calcular momentos con integración de Monte Carlo

# REE no lineales o no gaussianos

- Principales referencias: [Herbst y Schorfheide 2016 \(cap. 8\)](#) y [Herbst y Schorfheide 2019](#)
- Códigos que replican [Herbst y Schorfheide 2019](#):
  - En Matlab:  
<https://web.sas.upenn.edu/schorf/files/2016/10/MATLABProgramsTPF-14uavml.zip>
  - Para los Fortran geeks: [https://github.com/eph/tempered\\_pf](https://github.com/eph/tempered_pf)
- Una REE no lineal o no gaussiana tiene la forma

$$\mathbf{y}_t = \Psi(\boldsymbol{\alpha}_t, t; \boldsymbol{\theta}) + \mathbf{u}_t \text{ con } \mathbf{u}_t \sim F_u(\cdot; \boldsymbol{\theta}) \text{ y} \quad (1)$$

$$\boldsymbol{\alpha}_t = \Phi(\boldsymbol{\alpha}_{t-1}, \boldsymbol{\epsilon}_t; \boldsymbol{\theta}) \text{ con } \boldsymbol{\epsilon}_t \sim F_\epsilon(\cdot; \boldsymbol{\theta}) \quad (2)$$

la verosimilitud se puede estimar vía Filtro de Partículas (FP)

- Amplia literatura sobre FP, con muchos resúmenes y tutoriales. Por ejemplo, Arulampalam, Maskell, Gordon, and Clapp (2002), Cappe, Godsill, and Moulines (2007), Doucet and Johansen (2011), Creal (2012). Kantas, Doucet, Singh, Maciejowski, and Chopin (2014).
- El **algoritmo de FP bootstrap** básico es notablemente simple, aunque no opera muy bien en la práctica. Por lo tanto, gran parte de la literatura se centra en los refinamientos de este filtro con el propósito de mejorar su eficiencia
- Un FP más eficiente: **Conditionally optimal importance distribution**
- Un FP incluso más eficiente: **The tempered particle filter**

# El algoritmo “bootstrap particle filtering”

- 1 **Inicialización.** Simular las partículas iniciales de la distribución  $\alpha_{0|0} \stackrel{iid}{\sim} p(\alpha_0|\theta)$  y fijar los pesos en uno ( $W_0^j = 1$  para  $j \in \{1, \dots, m\}$  donde  $m$  es el número de partículas)
- 2 **Recurción.** Para  $t \in \{1, \dots, T\}$ :

- 1 **Predecir**  $\alpha_{t|t-1}$ : Propagar las partículas del periodo  $t - 1$  (i.e.,  $\{\alpha_{t-1|t-1}^j, W_{t-1}^j\}$ ) iterando la ecuación de estado hacia adelante

$$\alpha_{t|t-1}^j = \Phi(\alpha_{t-1|t-1}^j, \epsilon_t^j; \theta) \text{ con } \epsilon_t^j \sim F_\epsilon(\cdot; \theta)$$

- 2 **Predecir**  $y_{t|t-1}$ : Definir los pesos incrementales

$$w_t^j = p(y_t - \Psi(\alpha_{t|t-1}^j, t; \theta)) \text{ con } p(\cdot) \text{ a partir de } F_u(\cdot, \theta)$$

La densidad predictiva  $p(y_t|\mathcal{Y}^t; \theta)$  se puede aproximar con

$$\hat{p}(y_t|\mathcal{Y}^t; \theta) = \frac{1}{m} \sum_{j=1}^m w_t^j W_{t-1}^j$$

- 3 **Actualización:** Actualizar los pesos normalizados

$$W_t^j = \frac{w_t^j W_{t-1}^j}{\frac{1}{m} \sum_{j=1}^m w_t^j W_{t-1}^j}$$

# El algoritmo “bootstrap particle filtering”

## ② Continúa paso 2

- ④ **Selección:** Si el número efectivo de partículas  $(\frac{m^2}{\sum_{j=1}^m W_t^j})$  cae a más de la mitad, volver a muestrear las partículas mediante remuestreo multinomial. Así, si  $\{\alpha_{t|t}\}_{j=1}^m$  denota  $m$  simulaciones iid de una distribución multinomial caracterizada por puntos de soporte y pesos  $\{\alpha_{t|t-1}^j, W_t^j\}_{j=1}^m$  y fijamos  $W_t^j = 1$  para  $j \in \{1, \dots, m\}$ .

- ④ **Aproximación de la verosimilitud.** La aproximación de la función log-verosimilitud viene dada por

$$\log \hat{p}(\mathbf{y}^T | \boldsymbol{\theta}) = \sum_{t=1}^T \log \left( \frac{1}{m} \sum_{j=1}^m w_t^j W_{t-1}^j \right) \quad (3)$$

# Un filtro de partículas general

- El FP bootstrap en la iteración de predicción ignora la información contenida en la observación  $\mathbf{y}_t$  lo que puede conducir a una distribución desigual de los pesos de las partículas. En particular si la varianza del error de medición es reducido o si el modelo tiene dificultades para explicar el período  $t$
  - El FP se generaliza al permitir  $\alpha_{t|t-1}^j$  ser simulado desde una densidad genérica de *importance sampling*  $g_t(\cdot | \alpha_{t-1|t-1}^j; \theta)$  en el paso de predicción
  - Solo los pasos (2.1.) y (2.2.) cambian de la siguiente manera
- 2.(a) **Predecir**  $\alpha_{t|t-1}$ : Simular  $\alpha_{t|t-1}^j$  a partir de la densidad  $g_t(\alpha_{t|t-1} | \alpha_{t-1|t-1}^j; \theta)$ , las partículas del periodo  $t-1$   $\{\alpha_{t-1|t-1}^j, W_{t-1}^j\}$  y definir los pesos "importance"

$$\tilde{w}_t^j = \frac{p(\alpha_{t|t-1} | \alpha_{t-1|t-1}^j; \theta)}{g_t(\alpha_{t|t-1} | \alpha_{t-1|t-1}^j; \theta)}$$

- 2.(b) **Predecir**  $\mathbf{y}_{t|t-1}$ : Definir los pesos incrementales

$$w_t^j = p(\mathbf{y}_t - \Psi(\alpha_{t|t-1}^j, t; \theta)) \tilde{w}_t^j \text{ con } p(\cdot) \text{ desde } F_u(\cdot, \theta)$$

La densidad predictiva  $p(\mathbf{y}_t | \mathcal{Y}^t; \theta)$  se puede aproximar por

$$\hat{p}(\mathbf{y}_t | \mathcal{Y}^t; \theta) = \frac{1}{m} \sum_{j=1}^m w_t^j W_{t-1}^j$$

# Un filtro de partículas general

- La distribución condicionalmente óptima se define como la distribución que minimiza la variación de Monte Carlo de los pesos de importancia. En el caso del modelo gaussiano lineal, el “*importance sampling*” genérico óptimo es el siguiente

$$\alpha_{t|t-1}^j | \alpha_{t-1|t-1}^j, \mathbf{y}_t \sim \mathcal{N}(\alpha_{t|t}, \Sigma_{t|t}^\alpha) \quad (4)$$

i.e., un Normal con momentos provenientes del paso de actualización del KF



