

Series Temporales

Master en Modelización Matemática, Estadística y
Computación
Segunda Parte

F.Tusell

Dpto. Métodos Cuantitativos

Curso 2021–2022



Universidad Euskal Herriko
del País Vasco Unibertsitatea

Indice II

Usos de la transformada de Fourier

Análisis espectral univariante

Procesos estocásticos en el dominio de la frecuencia
Estimación de $f(\lambda)$
Tapering
Filtrado lineal
Métodos paramétricos

Análisis espectral multivariante

Densidad espectral cruzada
Coherencia y fase

Indice I

Modelo en el espacio de los estados

Representación en el espacio de los estados
Casos particulares de interés

El filtro de Kalman

Lemas y resultados previos
Obtención del filtro (caso gaussiano)

Predicción lineal

Mejor aproximación lineal, ortogonalidad, proyección.
Filtro de Kalman como mejor aproximación lineal

Modelos estructurales

Nivel local, tendencia local
Tratamiento de la estacionalidad

Estimación máximo verosímil

Análisis armónico

Requisitos de trigonometría
Aproximación mediante funciones periódicas
Transformadas de Fourier

Introducción

Vamos a:

- ▶ Introducir un nuevo modelo muy general y flexible.
- ▶ Ver como explotarlo para aproximar el estado de un fenómeno que evoluciona.
- ▶ Ver como estimar sus parámetros desconocidos, si los hubiera (“identificación” en la jerga de los ingenieros, “estimación” en la de los estadísticos).

Ecuación de estado - I

- ▶ Consideramos un fenómeno descrito por un vector de magnitudes (el “estado”, α_t), en principio inobservable directamente, que evoluciona en el tiempo de acuerdo a:

$$\alpha_{t+1} = \mathbf{T}_t \alpha_t + \mathbf{R}_t \eta_t \quad (1)$$

- ▶ Las matrices \mathbf{T}_t y \mathbf{R}_t se suponen conocidas y $\eta_t \sim N(\mathbf{0}, \mathbf{Q}_t)$.
- ▶ Las perturbaciones η_k se suponen independientes en el tiempo:

$$E[\eta_k \eta_j'] = \mathbf{Q}_k \delta_{kj}.$$

Ecuación de observación

- ▶ Podemos observar un vector \mathbf{y}_t que se genera como

$$\mathbf{y}_t = \mathbf{Z}_t \alpha_t + \epsilon_t \quad (2)$$

- ▶ Se supone $\epsilon_t \sim N(\mathbf{0}, \mathbf{H}_t)$ y conocidas las matrices \mathbf{Z}_t , \mathbf{H}_t .
- ▶ Las perturbaciones ϵ_k se suponen independientes en el tiempo:

$$E[\epsilon_k \epsilon_j'] = \mathbf{H}_k \delta_{kj}.$$

Adicionalmente, supondremos para todo j, k (incluso cuando $j = k$) que:

$$E[\epsilon_k \eta_j'] = 0$$

Ecuación de estado - II

- ▶ La propiedad markoviana (dependencia último retardo) no es una limitación.
- ▶ Si α_t fuera un escalar, el modelo tendría “memoria de pez”...
- ▶ ...pero es un vector que puede a su vez incorporar retardos.
- ▶ Por ejemplo, α_t puede ser un vector formado por las cantidades de agua (desconocidas, en principio) evaporadas sobre una región en los últimos siete días.

Inicialización

- ▶ En el momento $t = 1$ supondremos que

$$\alpha_1 \sim N(\mathbf{a}_1, \mathbf{P}_1); \quad (3)$$

- ▶ supondremos además que α_1 es independiente de ϵ_t y de η_t .
- ▶ Alternativamente, podemos suponer el estado inicial α_1 fijo y conocido.
- ▶ Aún otra alternativa es suponer incertidumbre total acerca del estado inicial α_1 , tomando $\mathbf{P}_1 \propto I \cdot \infty$.

$$\begin{aligned} \alpha_{t+1} &= \mathbf{T}_t \alpha_t + \mathbf{R}_t \eta_t && \text{(ecuación de estado)} \\ \mathbf{y}_t &= \mathbf{Z}_t \alpha_t + \epsilon_t && \text{(ecuación de observación)} \\ \eta_t &\stackrel{\text{i.i.d.}}{\sim} N(\mathbf{0}, \mathbf{Q}_t) \\ \epsilon_t &\stackrel{\text{i.i.d.}}{\sim} N(\mathbf{0}, \mathbf{H}_t) \\ \alpha_1 &\sim N(\mathbf{a}_1, \mathbf{P}_1) \text{ independientemente de } \eta_t, \epsilon_t \quad \forall t. \end{aligned}$$

- ▶ No conocemos el vector α_t de tres coordenadas que proporcionan la posición de un móvil en un sistema de referencia tridimensional.
- ▶ Podemos contar sin embargo con observaciones de un receptor GPS que nos proporciona dichas tres coordenadas con error.
- ▶ Si tenemos más información sobre el móvil, podemos incrementar la dimensión del vector de estado con velocidades, aceleraciones, etc.
- ▶ Idea esencial: conjugar **observaciones** y **estructura del problema** para estimar el estado.

Ejemplo (II): Observamos posición y velocidad

$$\begin{aligned} \begin{bmatrix} \alpha_{1,t+1} \\ \alpha_{2,t+1} \\ \alpha_{3,t+1} \\ \alpha_{4,t+1} \\ \alpha_{5,t+1} \\ \alpha_{6,t+1} \end{bmatrix} &= \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_{1,t} \\ \alpha_{2,t} \\ \alpha_{3,t} \\ \alpha_{4,t} \\ \alpha_{5,t} \\ \alpha_{6,t} \end{bmatrix} + \begin{bmatrix} \eta_{1,t} \\ \eta_{2,t} \\ \eta_{3,t} \\ \eta_{4,t} \\ \eta_{5,t} \\ \eta_{6,t} \end{bmatrix} \\ &\text{(ecuación de estado)} \\ \mathbf{y}_t = \begin{bmatrix} e_{1,t} \\ e_{2,t} \\ e_{3,t} \\ v_{1,t} \\ v_{2,t} \\ v_{3,t} \end{bmatrix} &= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_{1,t} \\ \alpha_{2,t} \\ \alpha_{3,t} \\ \alpha_{4,t} \\ \alpha_{5,t} \\ \alpha_{6,t} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \\ \epsilon_{4,t} \\ \epsilon_{5,t} \\ \epsilon_{6,t} \end{bmatrix} \\ &\text{(ecuación de observación)} \end{aligned}$$

Ejemplo (III): Observamos sólo posición

$$\begin{aligned} \begin{bmatrix} \alpha_{1,t+1} \\ \alpha_{2,t+1} \\ \alpha_{3,t+1} \\ \alpha_{4,t+1} \\ \alpha_{5,t+1} \\ \alpha_{6,t+1} \\ \alpha_{7,t+1} \\ \alpha_{8,t+1} \\ \alpha_{9,t+1} \end{bmatrix} &= \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \alpha_{1,t} \\ \alpha_{2,t} \\ \alpha_{3,t} \\ \alpha_{4,t} \\ \alpha_{5,t} \\ \alpha_{6,t} \\ \alpha_{7,t} \\ \alpha_{8,t} \\ \alpha_{9,t} \end{bmatrix} + \begin{bmatrix} \eta_{1,t} \\ \eta_{2,t} \\ \eta_{3,t} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \\ &\text{(ecuación de estado)} \\ \mathbf{y}_t = \begin{bmatrix} e_{1,t} \\ e_{2,t} \\ e_{3,t} \end{bmatrix} &= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \alpha_t + \epsilon_t \\ &\text{(ecuación de observación)} \end{aligned}$$

Modelo de regresión lineal con parámetros fijos

- ▶ Si hacemos

$$\begin{aligned}\beta_{t+1} &= I\beta_t + \mathbf{0}\eta_t \\ \mathbf{y}_t &= \mathbf{x}_t'\beta_t + \epsilon_t\end{aligned}$$

tenemos un modelo en espacio de estado con $\alpha_t = \beta_t$, $\mathbf{T}_t = I$ y $\mathbf{R}_t = \mathbf{0}$. El estado no pueda cambiar.

- ▶ Podemos inicializar mediante $\alpha_1 = \beta_1 \sim N(\mathbf{0}, I \cdot \infty)$, y obtener MCO procesando una observación cada vez.
- ▶ Aunque los parámetros sean fijos, β_t cambia con t (¿por qué?).

Modelo estructural de nivel aleatorio

- ▶ Supongamos

$$\begin{aligned}\mu_{t+1} &= \mu_t + \eta_t \\ \mathbf{y}_t &= \mu_t + \epsilon_t\end{aligned}$$

Supongamos μ_1 inicializado convenientemente, $\eta_t \sim N(\mathbf{0}, \mathbf{Q}_t)$ y $\epsilon_t \sim N(\mathbf{0}, \mathbf{H}_t)$.

- ▶ μ_t se interpreta como nivel aleatorio en torno al cual fluctúa \mathbf{y}_t .
- ▶ En el espacio de los estados,

$$\begin{aligned}\mu_{t+1} &= I\mu_t + \eta_t \\ \mathbf{y}_t &= I\mu_t + \epsilon_t,\end{aligned}$$

es decir, $\mathbf{T}_t = I$ y $\mathbf{Z}_t = I$.

Modelo de regresión lineal con parámetros variables

- ▶ El mismo modelo anterior

$$\begin{aligned}\beta_{t+1} &= I\beta_t + \eta_t \\ \mathbf{y}_t &= \mathbf{x}_t'\beta_t + \epsilon_t\end{aligned}$$

con $\alpha_1 = \beta_1 \sim N(\mathbf{0}, I \cdot \infty)$ y la matriz de covarianzas $\text{Var}(\eta_t) = \mathbf{Q} \neq \mathbf{0}$ seleccionada convenientemente, da un modelo de parámetros variables.

- ▶ En lugar de tomar una distribución difusa para β_1 podemos emplear $\beta_1 \sim N(\boldsymbol{\mu}, \boldsymbol{\Lambda})$; ello permite dar entrada a información *a priori* sobre β que podamos tener.
- ▶ Otras posibilidades: \mathbf{Q}_t puede no ser fija, para permitir diferentes grados de variabilidad de β en diferentes momentos.

Modelo MA(2): $y_t = \nu_t + \theta_1\nu_{t-1} + \theta_2\nu_{t-2}$

Una posible representación en espacio de estado sería:

$$\begin{bmatrix} y_{t+1} \\ \theta_1\nu_{t+1} + \theta_2\nu_t \\ \theta_2\nu_{t+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y_t \\ \theta_1\nu_t + \theta_2\nu_{t-1} \\ \theta_2\nu_t \end{bmatrix} + \begin{bmatrix} 1 \\ \theta_1 \\ \theta_2 \end{bmatrix} \nu_{t+1}$$

(ecuación de estado)

$$y_t = [1 \quad 0 \dots 0] \begin{bmatrix} y_t \\ \theta_1\nu_t + \theta_2\nu_{t-1} \\ \theta_2\nu_t \end{bmatrix}$$

(ecuación de observación)

Modelo ARMA(2, 1): $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \nu_t + \theta_1 \nu_{t-1}$

Una posible representación en espacio de estado sería:

$$\begin{bmatrix} y_{t+1} \\ \phi_2 y_t + \theta_1 \nu_{t+1} \end{bmatrix} = \begin{bmatrix} \phi_1 & 1 \\ \phi_2 & 0 \end{bmatrix} \begin{bmatrix} y_t \\ \phi_2 y_{t-1} + \theta_1 \nu_t \end{bmatrix} + \begin{bmatrix} 1 \\ \theta_1 \end{bmatrix} \nu_{t+1}$$

(ecuación de estado)

$$y_t = [1 \quad 0] \begin{bmatrix} y_t \\ \phi_2 y_{t-1} + \theta_1 \nu_t \end{bmatrix}$$

(ecuación de observación)

Modelo de regresión con perturbaciones ARMA.

Sea $\mathbf{Y} = \mathbf{X}\boldsymbol{\delta} + \mathbf{u}$, siendo $u_t = \alpha_{t,1}$ una sucesión ARMA (por simplicidad suponemos AR en lo que sigue). Basta “apilar” la

dinámica de los errores y la de los parámetros de la regresión $\boldsymbol{\delta} = \boldsymbol{\delta}_t$:

$$\begin{bmatrix} \boldsymbol{\alpha}_t \\ \boldsymbol{\delta}_t \end{bmatrix} = \begin{bmatrix} T_t & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_{t-1} \\ \boldsymbol{\delta}_{t-1} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\eta}_t \\ 0 \end{bmatrix}$$

(ecuación de estado)

$$y_t = [\mathbf{x}_t' \quad 1 \quad 0 \dots 0] \begin{bmatrix} \boldsymbol{\delta}_t \\ \boldsymbol{\alpha}_t \end{bmatrix}$$

(ecuación de observación)

Ventajas de la modelización en espacios de estado

- ▶ Análisis *estructural*. Componentes de interés como tendencia, ciclo, estacionalidad se modelizan, no se cancelan.
- ▶ Ventajas computacionales, gran generalidad.
- ▶ Separación del modelo de observación y del modelo del fenómeno.
- ▶ Fácil tratamiento de observaciones faltantes.
- ▶ Fácil extensión a series multivariantes.
- ▶ Fácil extensión a series irregularmente espaciadas.
- ▶ Fácil incorporación de variables explicativas.

Inconvenientes de la modelización en espacios de estado

- ▶ Prácticamente ninguno: casi en cada aspecto, más útil y potente que la metodología Box-Jenkins.
- ▶ Requiere pensar sobre la forma de la ecuación de estado: no es un “calcetín para todos los pies.”
- ▶ Veremos no obstante calcetines genéricos que convienen a muchos pies.
- ▶ No hay un método mecánico para identificar modelos (pero tampoco lo hay en Box-Jenkins cuando las series son multivariantes).

Convenios notacionales

- ▶ Consideraremos series temporales, en general vectoriales, de las que poseemos una muestra finita $[\mathbf{y}_1, \dots, \mathbf{y}_T]$.
- ▶ $\mathcal{Y}_t \stackrel{\text{def}}{=} [\mathbf{y}_1, \dots, \mathbf{y}_t]$ es la sección de la serie temporal desde el comienzo hasta \mathbf{y}_t inclusive.

Distribuciones condicionadas en la normal (Ia)

- ▶ Consideramos un vector \mathbf{X} particionado en $p + q$ componentes cuya distribución es:

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \sim N \left[\begin{pmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right]$$

- ▶ Entonces, $\mathbf{X}_1 | \mathbf{X}_2 = \mathbf{x}_2$ sigue una distribución normal multivariante, con vector de medias y matriz de covarianzas dadas por:

$$\begin{aligned} E[\mathbf{X}_1 | \mathbf{X}_2 = \mathbf{x}_2] &= \mathbf{m}_1 + \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} (\mathbf{x}_2 - \mathbf{m}_2) \\ \text{Var}(\mathbf{X}_1 | \mathbf{X}_2 = \mathbf{x}_2) &= \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} \end{aligned}$$

Distribuciones condicionadas en la normal (Ib)

- ▶ ¿De dónde sale el resultado anterior?
- ▶ Es preciso hacer el cociente entre la densidad conjunta $f(\mathbf{X}_1, \mathbf{X}_2)$ y la marginal

$$f(\mathbf{X}_2; \mathbf{m}_2, \boldsymbol{\Sigma}_{22}) = \frac{1}{(2\pi)^{q/2} |\boldsymbol{\Sigma}_{22}|^{1/2}} e^{-\frac{1}{2} (\mathbf{X}_2 - \mathbf{m}_2)' \boldsymbol{\Sigma}_{22}^{-1} (\mathbf{X}_2 - \mathbf{m}_2)}$$

- ▶ Al simplificar términos, hay que tener presente la identidad

$$|\boldsymbol{\Sigma}| = |\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21}| |\boldsymbol{\Sigma}_{22}|$$

que resulta de tomar determinantes en la identidad:

$$\begin{bmatrix} I & -\boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{11}^{-1} \\ 0 & I \end{bmatrix} \boldsymbol{\Sigma} \begin{bmatrix} I & 0 \\ -\boldsymbol{\Sigma}_{11} \boldsymbol{\Sigma}_{21}^{-1} & I \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} & 0 \\ 0 & \boldsymbol{\Sigma}_{22} \end{bmatrix}$$

Distribuciones condicionadas en la normal (Ic)

- ▶ Intuición: ¿Que relación hay entre $\boldsymbol{\Sigma}_{11}$ y

$$\text{Var}(\mathbf{X}_1 | \mathbf{X}_2 = \mathbf{x}_2) = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21}?$$

¿Cuándo serán iguales?

- ▶ Claramente, cuando $\boldsymbol{\Sigma}_{12}$ sea un bloque de ceros. ¿Qué implica esto en una distribución normal multivariante?
- ▶ ¿Podría acontecer que

$$\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} \gg \boldsymbol{\Sigma}_{11}$$

en que \gg significa “mayores elementos en la diagonal principal”?

Distribuciones condicionadas en la normal (II)

- ▶ Si \mathbf{X} está particionado en $p + q + r$ componentes y las últimas q y r son mutuamente independientes, tenemos:

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \mathbf{X}_3 \end{pmatrix} \sim N \left[\begin{pmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \\ \mathbf{m}_3 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \boldsymbol{\Sigma}_{13} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} & \mathbf{0} \\ \boldsymbol{\Sigma}_{31} & \mathbf{0} & \boldsymbol{\Sigma}_{33} \end{pmatrix} \right]$$

- ▶ Entonces, $\mathbf{X}_1 | (\mathbf{X}_2, \mathbf{X}_3) = (\mathbf{x}_2, \mathbf{x}_3)$ tiene el vector de medias:

$$E[\mathbf{X}_1 | \mathbf{X}_2, \mathbf{X}_3] = \mathbf{m}_1 + \underbrace{\begin{pmatrix} \boldsymbol{\Sigma}_{12} & \boldsymbol{\Sigma}_{13} \end{pmatrix}}_{\boldsymbol{\Sigma}_{12}} \underbrace{\begin{bmatrix} \boldsymbol{\Sigma}_{22}^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{33}^{-1} \end{bmatrix}}_{\boldsymbol{\Sigma}_{22}^{-1}} \underbrace{\begin{bmatrix} \mathbf{x}_2 - \mathbf{m}_2 \\ \mathbf{x}_3 - \mathbf{m}_3 \end{bmatrix}}_{\boldsymbol{\Sigma}_{21}}$$

El filtro de Kalman

- ▶ Datos

$$\mathbf{a}_t : E(\alpha_t | \mathcal{Y}_{t-1})$$

$$\mathbf{P}_t : \text{Var}(\alpha_t | \mathcal{Y}_{t-1}) = E((\alpha_t - \mathbf{a}_t)(\alpha_t - \mathbf{a}_t)' | \mathcal{Y}_{t-1})$$

$$\mathbf{y}_t : \text{Observacion en } t$$

proporciona $\mathbf{a}_{t+1} = E(\alpha_{t+1} | \mathcal{Y}_t)$ y $\mathbf{P}_{t+1} = \text{Var}(\alpha_{t+1} | \mathcal{Y}_t)$.

- ▶ Procede de manera recursiva.
- ▶ Como α_{t+1} es normal, $\mathbf{a}_{t+1} = E(\alpha_{t+1} | \mathcal{Y}_t)$ y \mathbf{P}_{t+1} completamente especifican lo que podemos saber acerca de su distribución con la información disponible en t .

Distribuciones condicionadas en la normal (III)

- ▶ La matriz de covarianzas condicionada es por su parte

$$\begin{aligned} \text{Var}(\mathbf{X}_1 | \mathbf{X}_2, \mathbf{X}_3) &= \\ &= \boldsymbol{\Sigma}_{11} - [\boldsymbol{\Sigma}_{12} \quad \boldsymbol{\Sigma}_{13}] \begin{bmatrix} \boldsymbol{\Sigma}_{22}^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{33}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_{12}' \\ \boldsymbol{\Sigma}_{13}' \end{bmatrix} \\ &= \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{12}' - \boldsymbol{\Sigma}_{13} \boldsymbol{\Sigma}_{33}^{-1} \boldsymbol{\Sigma}_{13}' \end{aligned}$$

- ▶ El resultado se generaliza de manera obvia.

El filtro de Kalman (II)

- ▶ En virtud de los supuestos,

$$\begin{bmatrix} \alpha_t | \mathcal{Y}_{t-1} \\ \mathbf{y}_t | \mathcal{Y}_{t-1} \end{bmatrix} \sim N \left[\begin{bmatrix} \mathbf{a}_t \\ \mathbf{Z}_t \mathbf{a}_t \end{bmatrix}, \begin{bmatrix} \mathbf{P}_t & \mathbf{P}_t \mathbf{Z}_t' \\ \mathbf{Z}_t \mathbf{P}_t & \mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{H}_t \end{bmatrix} \right]$$

- ▶ Por tanto,

$$E(\alpha_t | \mathcal{Y}_t) = \mathbf{a}_t + \mathbf{P}_t \mathbf{Z}_t' (\mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{H}_t)^{-1} (\mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_t)$$

$$\text{Var}(\alpha_t | \mathcal{Y}_t) = \mathbf{P}_t - \mathbf{P}_t \mathbf{Z}_t' (\mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{H}_t)^{-1} \mathbf{Z}_t \mathbf{P}_t'$$

- ▶ Las dos ecuaciones anteriores recogen la **actualización por medida** (*measurement update*).

El filtro de Kalman (III)

- ▶ Como

$$\begin{aligned}\alpha_{t+1} &= \mathbf{T}_t \alpha_t + \mathbf{R}_t \eta_t \\ \mathbf{y}_t &= \mathbf{Z}_t \alpha_t + \epsilon_t\end{aligned}$$

con $\eta_t \sim N(\mathbf{0}, \mathbf{Q}_t)$ y $\epsilon_t \sim N(\mathbf{0}, \mathbf{H}_t)$ tenemos que:

$$\mathbf{a}_{t+1} = E(\alpha_{t+1} | \mathcal{Y}_t) = \mathbf{T}_t E(\alpha_t | \mathcal{Y}_t)$$

$$\begin{aligned}\mathbf{P}_{t+1} = \text{Var}(\alpha_{t+1} | \mathcal{Y}_t) &= \text{Var}(\mathbf{T}_t \alpha_t + \mathbf{R}_t \eta_t | \mathcal{Y}_t) \\ &= \mathbf{T}_t \text{Var}(\alpha_t | \mathcal{Y}_t) \mathbf{T}_t' + \mathbf{R}_t \mathbf{Q}_t \mathbf{R}_t'\end{aligned}$$

- ▶ Las dos ecuaciones anteriores recogen la **actualización temporal** (*time update*); no hacen uso de nueva información.

El filtro de Kalman (V)

- ▶ Reordenando la primera de las expresiones anteriores,

$$\begin{aligned}\mathbf{a}_{t+1} &= \mathbf{T}_t \mathbf{a}_t + \mathbf{T}_t \mathbf{P}_t \mathbf{Z}_t' (\mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{H}_t)^{-1} (\mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_t) \\ &= \mathbf{T}_t \mathbf{a}_t + \mathbf{K}_t (\mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_t)\end{aligned}$$

en que $\mathbf{K}_t = \mathbf{T}_t \mathbf{P}_t \mathbf{Z}_t' (\mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{H}_t)^{-1}$ se denomina **matriz de ganancia**.

- ▶ La segunda por su parte proporciona:

$$\begin{aligned}\mathbf{P}_{t+1} &= \mathbf{T}_t \mathbf{P}_t \mathbf{T}_t' - \mathbf{T}_t \mathbf{P}_t \mathbf{Z}_t' (\mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{H}_t)^{-1} \mathbf{Z}_t \mathbf{P}_t' \mathbf{T}_t' + \mathbf{R}_t \mathbf{Q}_t \mathbf{R}_t' \\ &= \mathbf{T}_t \mathbf{P}_t \mathbf{L}_t' + \mathbf{R}_t \mathbf{Q}_t \mathbf{R}_t'\end{aligned}$$

en que $\mathbf{L}_t = \mathbf{T}_t - \mathbf{K}_t \mathbf{Z}_t$.

El filtro de Kalman (IV)

- ▶ Tenemos por tanto:

$$\begin{aligned}\mathbf{a}_{t+1} &= E(\alpha_{t+1} | \mathcal{Y}_t) = \mathbf{T}_t E(\alpha_t | \mathcal{Y}_t) \\ \mathbf{P}_{t+1} &= \text{Var}(\alpha_{t+1} | \mathcal{Y}_t) = \mathbf{T}_t \text{Var}(\alpha_t | \mathcal{Y}_t) \mathbf{T}_t' + \mathbf{R}_t \mathbf{Q}_t \mathbf{R}_t'\end{aligned}$$

y:

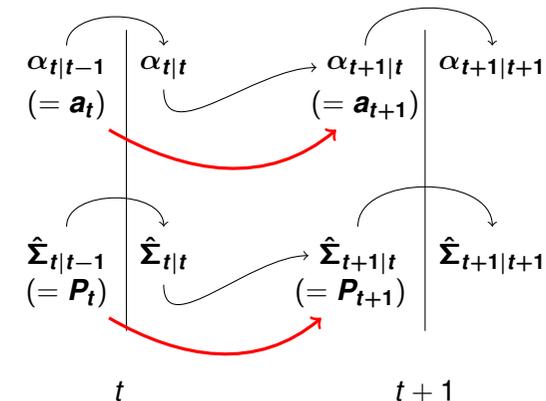
$$\begin{aligned}E(\alpha_t | \mathcal{Y}_t) &= \mathbf{a}_t + \mathbf{P}_t \mathbf{Z}_t' (\mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{H}_t)^{-1} (\mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_t) \\ \text{Var}(\alpha_t | \mathcal{Y}_t) &= \mathbf{P}_t - \mathbf{P}_t \mathbf{Z}_t' (\mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{H}_t)^{-1} \mathbf{Z}_t \mathbf{P}_t'\end{aligned}$$

- ▶ Sustituyendo estos valores en las igualdades previas:

$$\begin{aligned}\mathbf{a}_{t+1} &= \mathbf{T}_t [\mathbf{a}_t + \mathbf{P}_t \mathbf{Z}_t' (\mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{H}_t)^{-1} (\mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_t)] \\ \mathbf{P}_{t+1} &= \mathbf{T}_t [\mathbf{P}_t - \mathbf{P}_t \mathbf{Z}_t' (\mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{H}_t)^{-1} \mathbf{Z}_t \mathbf{P}_t'] \mathbf{T}_t' + \mathbf{R}_t \mathbf{Q}_t \mathbf{R}_t'\end{aligned}$$

Actualización temporal y por medida

$$\begin{aligned}\mathbf{a}_{t+1} &= \mathbf{T}_t [\mathbf{a}_t + \mathbf{P}_t \mathbf{Z}_t' (\mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{H}_t)^{-1} (\mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_t)] \\ \mathbf{P}_{t+1} &= \mathbf{T}_t [\mathbf{P}_t - \mathbf{P}_t \mathbf{Z}_t' (\mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{H}_t)^{-1} \mathbf{Z}_t \mathbf{P}_t'] \mathbf{T}_t' + \mathbf{R}_t \mathbf{Q}_t \mathbf{R}_t'\end{aligned}$$



Ejemplo: Modelo estructural de nivel aleatorio - I

- ▶ Supongamos

$$\begin{aligned}\mu_{t+1} &= \mathbf{T}_t \mu_t + \mathbf{R}_t \eta_t & \eta_t &\stackrel{\text{i.i.d.}}{\sim} N(\mathbf{0}, \mathbf{Q}_t) \\ \mathbf{y}_t &= \mathbf{Z}_t \mu_t + \epsilon_t & \epsilon_t &\stackrel{\text{i.i.d.}}{\sim} N(\mathbf{0}, \mathbf{H}_t)\end{aligned}$$

- ▶ Modelo en espacio de estado con $\mathbf{T}_t = \mathbf{R}_t = \mathbf{Z}_t = \mathbf{I}$.
- ▶ El filtro de Kalman actualiza \mathbf{a}_t de acuerdo con la ecuación:

$$\begin{aligned}\mathbf{a}_{t+1} &= \mathbf{T}_t \mathbf{a}_t + \mathbf{T}_t \mathbf{P}_t \mathbf{Z}_t' (\mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{H}_t)^{-1} (\mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_t) \\ \mathbf{a}_{t+1} &= \mathbf{a}_t + \mathbf{P}_t (\mathbf{P}_t + \mathbf{H}_t)^{-1} (\mathbf{y}_t - \mathbf{a}_t)\end{aligned}$$

- ▶ Intuitivamente atrayente: ¿A qué se reduce en el caso de un paseo aleatorio univariante?

Características del filtro de Kalman - I

Ecuaciones del filtro:

$$\mathbf{a}_{t+1} = \mathbf{T}_t \mathbf{a}_t + \mathbf{K}_t (\mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_t) \quad (4)$$

$$\mathbf{P}_{t+1} = \mathbf{T}_t \mathbf{P}_t \mathbf{L}_t' + \mathbf{R}_t \mathbf{Q}_t \mathbf{R}_t' \quad (5)$$

Las expresiones (4) y (5) permiten actualizar $\mathbf{a}_{t+1} = E(\alpha_{t+1} | \mathbf{y}_t)$ y su matriz de covarianzas $\mathbf{P}_{t+1} = \text{Var}(\alpha_{t+1} | \mathbf{y}_t)$ cuando se dispone de la nueva observación \mathbf{y}_t . Se trata de un filtro lineal.

Ejemplo: Modelo estructural de nivel aleatorio - II

$$\begin{aligned}\mu_{t+1} &= \mathbf{T}_t \mu_t + \mathbf{R}_t \eta_t & \eta_t &\stackrel{\text{i.i.d.}}{\sim} N(\mathbf{0}, \mathbf{Q}_t) \\ \mathbf{y}_t &= \mathbf{Z}_t \mu_t + \epsilon_t & \epsilon_t &\stackrel{\text{i.i.d.}}{\sim} N(\mathbf{0}, \mathbf{H}_t)\end{aligned}$$

- ▶ La actualización de \mathbf{P}_t viene dada por:

$$\mathbf{P}_{t+1} = \mathbf{T}_t \mathbf{P}_t \mathbf{L}_t' + \mathbf{R}_t \mathbf{Q}_t \mathbf{R}_t' = \mathbf{T}_t \mathbf{P}_t (\mathbf{T}_t - \mathbf{K}_t \mathbf{Z}_t)' + \mathbf{R}_t \mathbf{Q}_t \mathbf{R}_t'$$

- ▶ En nuestro caso,

$$\begin{aligned}\mathbf{P}_{t+1} &= \mathbf{T}_t \mathbf{P}_t (\mathbf{T}_t - \mathbf{K}_t \mathbf{Z}_t)' + \mathbf{R}_t \mathbf{Q}_t \mathbf{R}_t' \\ &= \mathbf{P}_t (\mathbf{I} - \mathbf{K}_t)' + \mathbf{Q}_t\end{aligned}$$

- ▶ También intuitivamente atrayente. ¿A qué se reduce en el caso de un paseo aleatorio univariante?

Características del filtro de Kalman - II

Ecuaciones del filtro:

$$\mathbf{a}_{t+1} = \mathbf{T}_t \mathbf{a}_t + \mathbf{K}_t (\mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_t) \quad (6)$$

$$\mathbf{P}_{t+1} = \mathbf{T}_t \mathbf{P}_t \mathbf{L}_t' + \mathbf{R}_t \mathbf{Q}_t \mathbf{R}_t' \quad (7)$$

La ecuación (6) da \mathbf{a}_{t+1} como un promedio ponderado de α_t y la **innovación**:

$$(\mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_t)$$

En Economía suele hablarse de *modelos de corrección de error*.

Características del filtro de Kalman - III

Ecuaciones del filtro:

$$\begin{aligned}\mathbf{a}_{t+1} &= \mathbf{T}_t \mathbf{a}_t + \mathbf{K}_t (\mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_t) \\ \mathbf{P}_{t+1} &= \mathbf{T}_t \mathbf{P}_t \mathbf{L}_t' + \mathbf{R}_t \mathbf{Q}_t \mathbf{R}_t'\end{aligned}$$

Sólo el último valor de \mathbf{a}_t hace falta ("propiedad markoviana") e interviene de modo lineal en la generación de \mathbf{a}_{t+1} . Se trata de un filtro **lineal** y **finito-dimensional**.

Características del filtro de Kalman - IV

Ecuaciones del filtro:

$$\begin{aligned}\mathbf{a}_{t+1} &= \mathbf{T}_t \mathbf{a}_t + \mathbf{K}_t (\mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_t) \\ \mathbf{P}_{t+1} &= \mathbf{T}_t \mathbf{P}_t \mathbf{L}_t' + \mathbf{R}_t \mathbf{Q}_t \mathbf{R}_t'\end{aligned}$$

La matriz

$$\mathbf{K}_t = \mathbf{T}_t \mathbf{P}_t \mathbf{Z}_t' (\mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{H}_t)^{-1}$$

cambia a lo largo del tiempo, incluso si $\mathbf{T}_t = \mathbf{T}$ constante (porque \mathbf{P}_t va variando). El filtro es **temporalmente variable** incluso aunque la dinámica del modelo sea invariante en el tiempo.

Características del filtro de Kalman - V

Ecuaciones del filtro:

$$\begin{aligned}\mathbf{a}_{t+1} &= \mathbf{T}_t \mathbf{a}_t + \mathbf{K}_t (\mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_t) \\ \mathbf{P}_{t+1} &= \mathbf{T}_t \mathbf{P}_t \mathbf{L}_t' + \mathbf{R}_t \mathbf{Q}_t \mathbf{R}_t'\end{aligned}$$

Para calcular \mathbf{P}_{t+1} no se requiere \mathbf{y}_t ; la sucesión completa de matrices \mathbf{P}_t puede calcularse desde el principio.

Características del filtro de Kalman - VI

Ecuaciones del filtro:

$$\begin{aligned}\mathbf{a}_{t+1} &= \mathbf{T}_t \mathbf{a}_t + \mathbf{K}_t (\mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_t) & (8) \\ \mathbf{P}_{t+1} &= \mathbf{T}_t \mathbf{P}_t \mathbf{L}_t' + \mathbf{R}_t \mathbf{Q}_t \mathbf{R}_t' & (9)\end{aligned}$$

La ecuación (9) es una *ecuación de Riccati*.

Características del filtro de Kalman - VII

Ecuaciones del filtro:

$$\begin{aligned}\mathbf{a}_{t+1} &= \mathbf{T}_t \mathbf{a}_t + \mathbf{K}_t (\mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_t) \\ \mathbf{P}_{t+1} &= \mathbf{T}_t \mathbf{P}_t \mathbf{L}_t' + \mathbf{R}_t \mathbf{Q}_t \mathbf{R}_t'\end{aligned}$$

Cuando las distribuciones son gaussianas, los momentos $\mathbf{a}_{t+1} = E(\alpha_{t+1} | \mathbf{y}_t)$ y $\mathbf{P}_{t+1} = \text{Var}(\alpha_{t+1} | \mathbf{y}_t)$ las especifican completamente. El filtro de Kalman es una regla que actualiza la distribución del estado a lo largo del tiempo.

Cuando no hay normalidad, trasladar los momentos no es suficiente y se ha de recurrir a alternativas (EKF, UKF, filtros de partículas. . .).

Tratamiento de datos faltantes (I)

- ▶ ¡No hay que hacer nada especial!
- ▶ Si en

$$\begin{aligned}\alpha_{t+1} &= \mathbf{T}_t \alpha_t + \mathbf{R}_t \eta_t && \text{(ecuación de estado)} \\ \mathbf{y}_t &= \mathbf{Z}_t \alpha_t + \epsilon_t && \text{(ecuación de observación)}\end{aligned}$$

falla la observación de parte de \mathbf{y}_t , simplemente ocurre que la matriz \mathbf{Z}_t “encoge”.

- ▶ Todas las fórmulas del filtro se aplican igual, con la \mathbf{Z}_t “encogida”.

Tratamiento de datos faltantes (II)

- ▶ Si el vector \mathbf{y}_t es inobservable en su totalidad, hacemos en el filtro una actualización temporal (*time update*) sin medida (*measurement update*).
- ▶ Tenemos entonces:

$$\begin{aligned}\mathbf{a}_{t+1} &= \mathbf{T}_t \mathbf{a}_t \\ \mathbf{P}_{t+1} &= \mathbf{T}_t \mathbf{P}_t \mathbf{T}_t' + \mathbf{R}_t \mathbf{Q}_t \mathbf{R}_t'\end{aligned}$$

- ▶ Es equivalente a hacer la innovación igual al vector de ceros, y la ganancia $\mathbf{K}_t = \mathbf{0}$ ($\Rightarrow \mathbf{L}_t = \mathbf{T}_t - \mathbf{K}_t \mathbf{Z}_t = \mathbf{T}_t$).

Insertar sesión práctica N1 aquí.

¿Y si no hay normalidad?

- ▶ Hemos supuesto normalidad en los ruidos, y el filtro ha resultado ser lineal.
- ▶ La linealidad ha sido una **consecuencia** de la normalidad.
- ▶ Podemos prescindir de la hipótesis de normalidad e **imponer** linealidad en el filtro.
- ▶ El resultado, como comprobaremos, es idéntico.
- ▶ La obtención alternativa contempla un modelo algo más general.

Nuevos supuestos:

- ▶ Introducimos inputs deterministas \mathbf{u}_t en la ecuación de estado:

$$\alpha_{t+1} = \mathbf{T}_t \alpha_t + \mathbf{B}_t \mathbf{u}_t + \mathbf{R}_t \eta_t \quad (\text{ecuación de estado})$$

$$\mathbf{y}_t = \mathbf{Z}_t \alpha_t + \epsilon_t \quad (\text{ecuación de observación})$$

$$\eta_t \stackrel{\text{i.i.d.}}{\sim} (\mathbf{0}, \mathbf{Q}_t); \quad \epsilon_t \stackrel{\text{i.i.d.}}{\sim} (\mathbf{0}, \mathbf{H}_t); \quad E[\eta_t \epsilon_t'] = \mathbf{S}_t$$

$$\alpha_1 \sim (\mathbf{a}_1, \mathbf{P}_1) \text{ independiente de } \eta_t, \epsilon_t \quad \forall t.$$

- ▶ La restricción de normalidad ha desaparecido; además de introducir \mathbf{u}_t inputs admitimos covarianza \mathbf{S}_t entre ruidos.

Predicción lineal (I)

- ▶ Desarrollaremos un nuevo método para obtener el filtro de Kalman.
- ▶ La teoría de aproximación lineal puede ser aplicada a nuestro problema.
- ▶ Presentaremos la teoría en un caso simple y conocido, extendiéndola después a nuestro problema.
- ▶ Ventajas: mejorar la intuición, pensar las cosas una sola vez.

Predicción lineal (II)

- ▶ Problema: aproximar un vector $\mathbf{v} \in V$ por un vector en $H \subset V$. Provisionalmente, V de dimensión finita.
- ▶ En el espacio V (y por tanto en H) hay definido un producto interno $\langle \cdot, \cdot \rangle$ y consiguientemente una *norma*,

$$\|\mathbf{v}\|^2 = \langle \mathbf{v}, \mathbf{v} \rangle$$

- ▶ $\mathbf{u} \in H$ será la mejor aproximación de \mathbf{v} si

$$\|\mathbf{v} - \mathbf{u}\|^2 = \min_{\mathbf{a} \in H} \|\mathbf{v} - \mathbf{a}\|^2$$

- ▶ Claramente, solución dependiente de $\langle \cdot, \cdot \rangle$.

Predicción lineal (IIa) - Digresión

- ▶ Un producto interno de $H \times H$ en R debe verificar:

$$\mathbf{u}, \mathbf{v} \in H, \langle \mathbf{u}, \mathbf{v} \rangle \in R$$

$$\mathbf{u}, \mathbf{v} \in H, \langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$$

$$\mathbf{u} \in H, \langle \mathbf{u}, \mathbf{u} \rangle \geq 0; \langle \mathbf{u}, \mathbf{u} \rangle = 0 \Rightarrow \mathbf{u} = \mathbf{0}$$

$$\mathbf{u}, \mathbf{v}, \mathbf{w} \in H, \langle \alpha \mathbf{u} + \beta \mathbf{v}, \mathbf{w} \rangle = \langle \alpha \mathbf{u}, \mathbf{w} \rangle + \langle \beta \mathbf{v}, \mathbf{w} \rangle$$

Predicción lineal (III) - Proyección

- ▶ Consideremos el producto interno ordinario

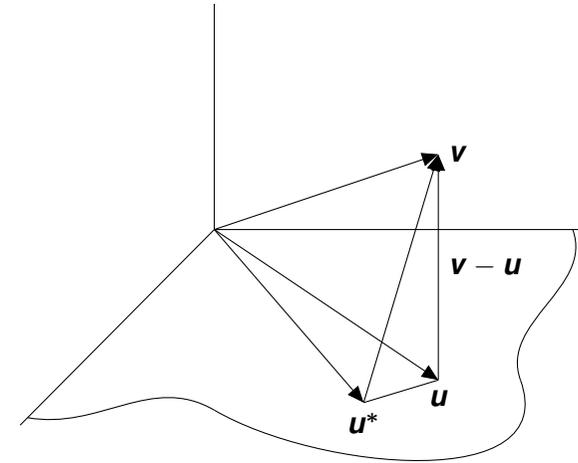
$$\langle \mathbf{w}, \mathbf{z} \rangle = \mathbf{w}'\mathbf{z} = w_1 z_1 + \dots + w_p z_p$$

- ▶ Asociado al mismo está la norma euclídea ordinaria:

$$\|\mathbf{z}\| = \sqrt{\langle \mathbf{z}, \mathbf{z} \rangle} = \sqrt{z_1^2 + \dots + z_p^2},$$

distancia al origen de \mathbf{z} visto como un punto en \mathbb{R}^p .

Predicción lineal (IV) - Geometría problema



Predicción lineal (V) - Condiciones de ortogonalidad

- ▶ La mejor aproximación $\mathbf{u} \in H \subset V$ de $\mathbf{v} \in V$ en la norma inducida por \langle, \rangle verifica:

$$\langle \mathbf{v} - \mathbf{u}, \mathbf{w} \rangle = 0$$

para cualquier $\mathbf{w} \in H$.

- ▶ Se expresa lo anterior así: $(\mathbf{v} - \mathbf{u}) \perp H$
- ▶ Si H generado por los vectores columna de X , la mejor aproximación $\mathbf{u} \in H$ puede expresarse $\mathbf{u} = X\hat{\beta}$ y ha de verificar:

$$(\mathbf{v} - X\hat{\beta}) \perp H \Leftrightarrow (\mathbf{v} - X\hat{\beta})'X = \mathbf{0} \Rightarrow \hat{\beta} = (X'X)^{-1}X'\mathbf{v}$$

(ecuaciones normales).

Predicción lineal (Va) - Condiciones de existencia

- ▶ Estamos suponiendo que la mejor aproximación $\mathbf{u} \in H \subset V$ para cualquier $\mathbf{v} \in V$ existe.
- ▶ No necesariamente ha de ser el caso.
- ▶ Ilustración: ¿Existe el número racional que mejor aproxima π ?
- ▶ Hacer el argumento anterior estanco requiere condiciones sobre el espacio de trabajo.
- ▶ Mencionaremos brevemente lo que se requiere. En la práctica son cosas que no necesitan preocuparnos.

Predicción lineal (VI) - Procesos estocásticos

- ▶ Las condiciones de ortogonalidad se han ilustrado en un espacio con un producto interno (y norma) conocidos.
- ▶ *Nada* en el razonamiento previo está asociado a un particular espacio o producto interno.
- ▶ Podemos hacer uso de lo precedente en un espacio y con un producto interno adecuados al problema de predicción lineal en procesos estocásticos.

Predicción lineal (VII) - Procesos estocásticos

- ▶ Sea V el espacio generado por las $\{y_s\}_{s=-\infty}^t$ y H el generado por $\{y_s\}_{s=-\infty}^{t-1}$ (hemos designado H anteriormente como \mathcal{Y}_{t-1}).
- ▶ Por simplicidad supondremos procesos univariantes centrados. Definamos en V (y en H) el producto interno:

$$\langle y_m, y_n \rangle = E[y_m y_n]$$

- ▶ **Ejercicio:** Comprobar que se trata de un producto interno bien definido.
- ▶ La norma asociada es,

$$\|y_s\|^2 = E[y_s^2] = \text{Var}(y_s)$$

Predicción lineal (VIII) - Procesos estocásticos

- ▶ El problema de predicción mínimo-cuadrática de y_t haciendo uso de información hasta $t - 1$ puede plantearse como:

$$\min_{u \in \mathcal{Y}_{t-1}} \|y_t - u\|^2$$

- ▶ La discusión anterior muestra que la solución u debe verificar:

$$(y_t - u) \perp \mathcal{Y}_{t-1} \Rightarrow E[(y_t - u)z_s] = 0 \quad \forall z_s \in \mathcal{Y}_{t-1}$$

Predicción lineal (XI) - Procesos estocásticos

- ▶ Si tuviéramos una base de \mathcal{Y}_{t-1} , podríamos escribir

$$u = \sum_{s=1}^{\infty} c_s z_{t-s}$$

- ▶ Particularmente cómodas las bases ortonormales. Si $E[z_m z_n] = 0$ para $m \neq n$ y $E[z_n^2] = 1$ para todo n ,

$$E[(y_t - u)z_s] = 0 \Rightarrow E[y_t z_s] - \sum_{k=1}^{\infty} c_k E[z_{t-k} z_s] = 0$$

- ▶ De la última igualdad,

$$c_k = E[y_t z_{t-k}]$$

Cojamos aliento. . .

- ▶ El problema de la óptima predicción lineal (en sentido m.c.) puede ser elegantemente formulado como la búsqueda de una proyección.
- ▶ Facilita las cosas disponer en el espacio sobre el que se proyecta de una base ortonormal, o al menos ortogonal.
- ▶ Podemos obtener la mejor predicción lineal m.c. de α_{t+1} proyectando sobre \mathcal{Y}_t .

Las innovaciones están centradas

- ▶ Recordemos que,

$$E_X[X] = E_Y \{ E_{X|Y}[X|Y = y] \},$$

en que el subíndice denota la distribución respecto a la que se integra.

- ▶ Consecuentemente,

$$\begin{aligned} E[\tilde{y}_t] &= E_{\mathcal{Y}_{t-1}} \left\{ E_{Y_t|\mathcal{Y}_{t-1}}[\tilde{y}_t|\mathcal{Y}_{t-1}] \right\} \\ &= E_{\mathcal{Y}_{t-1}} \left\{ E_{Y_t|\mathcal{Y}_{t-1}}[y_t - E[y_t|\mathcal{Y}_{t-1}]|\mathcal{Y}_{t-1}] \right\} \\ &= 0. \end{aligned}$$

Innovaciones

- ▶ Sea $\{y_t\}_{t=0}^{\infty}$ una sucesión aleatoria gaussiana.
- ▶ Se obtiene $\{\tilde{y}_t\}$ incorrelada por ortogonalización à la Gram-Schmidt:

$$\begin{aligned} \tilde{y}_0 &= y_0 - E[y_0] \\ \tilde{y}_1 &= y_1 - E[y_1|\mathcal{Y}_0] \\ &\vdots \\ \tilde{y}_t &= y_t - E[y_t|\mathcal{Y}_{t-1}] \\ &\vdots \end{aligned}$$

- ▶ $\{\tilde{y}_t\}_{t=0}^{\infty}$ es la sucesión de *innovaciones* asociada a $\{y_t\}_{t=0}^{\infty}$.

Las innovaciones están incorreladas

- ▶ Supongamos, sin pérdida de generalidad, $t > s$. Entonces,

$$E[\tilde{y}_t \tilde{y}_s] = E[(y_t - E[y_t|\mathcal{Y}_{t-1}])\tilde{y}_s]$$

- ▶ $(y_t - E[y_t|\mathcal{Y}_{t-1}])$ es ortogonal a \mathcal{Y}_{t-1} :
- ▶ $\tilde{y}_s \in \mathcal{Y}_{t-1}$.
- ▶ $E[y_t|\mathcal{Y}_{t-1}]$ es la proyección de y_t sobre \mathcal{Y}_{t-1} .

Las innovaciones forman una base ortogonal del espacio generado por $\{y_t\}_{t=0}^{\infty}$

- ▶ Descomposición de Wold procesos estacionarios regulares:

$$y_t = \sum_{k=0}^{\infty} a_k \xi_{t-k}$$

- ▶ El espacio que generan las ξ_t es el mismo que generan las observaciones y las innovaciones:

$$\mathcal{Y}_t = \tilde{\mathcal{Y}}_t = \bar{\Xi}_t$$

- ▶ Consecuentemente, será equivalente proyectar sobre uno o sobre otro.

El desarrollo anterior no es formal...

- ▶ ...pero puede hacerse que lo sea.
- ▶ Con procesos "sin comienzo", $\{y_t\}_{t=-\infty}^{\infty}$, el espacio sobre el que se proyecta es infinito dimensional.
- ▶ Necesitamos que sea completo (contenga todos los límites de sucesiones de Cauchy).
- ▶ Marco adecuado de análisis: espacio de Hilbert (espacio vectorial con producto interno, completo respecto a la norma inducida por dicho producto).
- ▶ La descomposición de Wold es válida para procesos aleatorios estacionarios y regulares.
- ▶ Anderson, T.W. *The Statistical Analysis of Time Series*, Wiley, § 7.6.

Varianza del error de predicción

- ▶ Descomposiciones alternativas, quizá finitas:

$$y_t = \sum_{k=0}^{\infty} a_k \xi_{t-k} \quad y_t = \sum_{k=1}^{\infty} b_k y_{t-k} + \xi_t$$

- ▶ Descomposición mixta finita frecuentemente buena aproximación (fundamento modelización ARMA):

$$y_t = \sum_{k=1}^p b_k y_{t-k} + \sum_{k=1}^q a_k \xi_{t-k} + \xi_t$$

- ▶ La descomposición MA permite calcular fácilmente la varianza del error de predicción:

$$\text{Var}(\hat{y}_{t|m}) = \sum_{k=0}^{m-1} a_k^2 \sigma_{\xi}^2 = \sum_{k=0}^{m-1} \hat{a}_k^2$$

Obtención alternativa del filtro de Kalman (I)

- ▶ Recordemos,

$$\begin{aligned} \alpha_{t+1} &= \mathbf{T}_t \alpha_t + \mathbf{B}_t \mathbf{u}_t + \mathbf{R}_t \eta_t && \text{(ecuación de estado)} \\ \mathbf{y}_t &= \mathbf{Z}_t \alpha_t + \epsilon_t && \text{(ecuación de observación)} \end{aligned}$$

$$\begin{aligned} \eta_t &\stackrel{\text{i.i.d.}}{\sim} (\mathbf{0}, \mathbf{Q}_t); & \epsilon_t &\stackrel{\text{i.i.d.}}{\sim} (\mathbf{0}, \mathbf{H}_t); & E[\eta_t \epsilon_t'] &= \mathbf{S}_t \\ \alpha_1 &\sim (\mathbf{a}_1, \mathbf{P}_1) & \text{independiente de} & \eta_t, \epsilon_t & \forall t. \end{aligned}$$

Obtención alternativa del filtro de Kalman (II)

- ▶ Queremos obtener $\hat{\alpha}_{t|t-1} = \mathbf{a}_t = E(\alpha_t | \mathcal{Y}_{t-1})$. Consideramos procesos centrados.
- ▶ $\mathcal{Y}_{t-1} = \tilde{\mathcal{Y}}_{t-1}$ y podemos proyectar sobre este último subespacio

$$E(\alpha_t | \mathcal{Y}_{t-1}) = E(\alpha_t | \tilde{\mathcal{Y}}_{t-1});$$

- ▶ En virtud de la ortogonalidad,

$$E(\alpha_{t+1} | \tilde{\mathcal{Y}}_t) = E(\alpha_{t+1} | \tilde{\mathbf{y}}_t) + E(\alpha_{t+1} | \tilde{\mathcal{Y}}_{t-1})$$

- ▶ Cf. Anderson-Moore, *Optimal Filtering*, § 5.4.

Obtención alternativa del filtro de Kalman (IV)

$$E(\alpha_{t+1} | \tilde{\mathcal{Y}}_t) = \underbrace{E(\alpha_{t+1} | \tilde{\mathbf{y}}_t)}_{(1)} + \underbrace{E(\alpha_{t+1} | \tilde{\mathcal{Y}}_{t-1})}_{(2)}$$

$$E(\alpha_{t+1} | \tilde{\mathbf{y}}_t) = (\mathbf{T}_t \boldsymbol{\Sigma}_{t|t-1} \mathbf{Z}_t' + \mathbf{R}_t \mathbf{S}_t) [\text{cov}(\tilde{\mathbf{y}}_t, \tilde{\mathbf{y}}_t)]^{-1} \tilde{\mathbf{y}}_t$$

- ▶ Por otra parte,

$$\text{cov}(\tilde{\mathbf{y}}_t, \tilde{\mathbf{y}}_t) = \text{cov}(\mathbf{Z}_t(\alpha_t - \mathbf{a}_t) + \epsilon_t, \mathbf{Z}_t(\alpha_t - \mathbf{a}_t) + \epsilon_t) = \mathbf{Z}_t \boldsymbol{\Sigma}_{t|t-1} \mathbf{Z}_t' + \mathbf{H}_t$$

- ▶ Reemplazando en la expresión anterior tenemos:

$$E(\alpha_{t+1} | \tilde{\mathbf{y}}_t) = (\mathbf{T}_t \boldsymbol{\Sigma}_{t|t-1} \mathbf{Z}_t' + \mathbf{R}_t \mathbf{S}_t) (\mathbf{Z}_t \boldsymbol{\Sigma}_{t|t-1} \mathbf{Z}_t' + \mathbf{H}_t)^{-1} \tilde{\mathbf{y}}_t$$

- ▶ Ello da cuenta de (1).

Obtención alternativa del filtro de Kalman (III)

$$E(\alpha_{t+1} | \tilde{\mathcal{Y}}_t) = \underbrace{E(\alpha_{t+1} | \tilde{\mathbf{y}}_t)}_{(1)} + \underbrace{E(\alpha_{t+1} | \tilde{\mathcal{Y}}_{t-1})}_{(2)}$$

- ▶ Consideremos (1):

$$E(\alpha_{t+1} | \tilde{\mathbf{y}}_t) = \text{cov}(\alpha_{t+1}, \tilde{\mathbf{y}}_t) [\text{cov}(\tilde{\mathbf{y}}_t, \tilde{\mathbf{y}}_t)]^{-1} \tilde{\mathbf{y}}_t$$

- ▶ Sea $\boldsymbol{\Sigma}_{t|t-1} = E[(\alpha_t - \mathbf{a}_t)(\alpha_t - \mathbf{a}_t)']$. Como

$$\begin{aligned} \text{cov}(\alpha_{t+1}, \tilde{\mathbf{y}}_t) &= \text{cov}(\mathbf{T}_t \alpha_t + \mathbf{B}_t \mathbf{u}_t + \mathbf{R}_t \eta_t, \mathbf{Z}_t(\alpha_t - \mathbf{a}_t) + \epsilon_t) \\ &= E[(\mathbf{T}_t \alpha_t + \mathbf{R}_t \eta_t - \mathbf{T}_t E[\alpha_t]) (\mathbf{Z}_t(\alpha_t - \mathbf{a}_t) + \epsilon_t)'] \\ &= E[\mathbf{T}_t \alpha_t (\alpha_t - \mathbf{a}_t)' \mathbf{Z}_t' + \mathbf{R}_t \eta_t \epsilon_t'] \\ &= E[\mathbf{T}_t (\alpha_t - \mathbf{a}_t) (\alpha_t - \mathbf{a}_t)' \mathbf{Z}_t' + \mathbf{R}_t \eta_t \epsilon_t'] \\ &= \mathbf{T}_t \boldsymbol{\Sigma}_{t|t-1} \mathbf{Z}_t' + \mathbf{R}_t \mathbf{S}_t \end{aligned}$$

$$E(\alpha_{t+1} | \tilde{\mathbf{y}}_t) = (\mathbf{T}_t \boldsymbol{\Sigma}_{t|t-1} \mathbf{Z}_t' + \mathbf{R}_t \mathbf{S}_t) [\text{cov}(\tilde{\mathbf{y}}_t, \tilde{\mathbf{y}}_t)]^{-1} \tilde{\mathbf{y}}_t$$

Obtención alternativa del filtro de Kalman (V)

$$E(\alpha_{t+1} | \tilde{\mathcal{Y}}_t) = \underbrace{E(\alpha_{t+1} | \tilde{\mathbf{y}}_t)}_{(1)} + \underbrace{E(\alpha_{t+1} | \tilde{\mathcal{Y}}_{t-1})}_{(2)}$$

- ▶ Examinemos ahora (2):

$$E(\alpha_{t+1} | \tilde{\mathcal{Y}}_{t-1}) = \hat{\alpha}_{t+1|t-1} = \mathbf{T}_t \hat{\alpha}_{t|t-1} + \mathbf{B}_t \mathbf{u}_t$$

- ▶ Reemplazando las expresiones de (1) y (2):

$$E(\alpha_{t+1} | \tilde{\mathcal{Y}}_t) = (\mathbf{T}_t \boldsymbol{\Sigma}_{t|t-1} \mathbf{Z}_t' + \mathbf{R}_t \mathbf{S}_t) (\mathbf{Z}_t \boldsymbol{\Sigma}_{t|t-1} \mathbf{Z}_t' + \mathbf{H}_t)^{-1} \tilde{\mathbf{y}}_t + \mathbf{T}_t \hat{\alpha}_{t|t-1} + \mathbf{B}_t \mathbf{u}_t$$

Obtención alternativa del filtro de Kalman (VI)

- ▶ La última expresión puede simplificarse::

$$\begin{aligned} E \left(\underbrace{\alpha_{t+1}}_{\mathbf{a}_{t+1}} | \tilde{\mathcal{Y}}_t \right) &= (\mathbf{T}_t \boldsymbol{\Sigma}_{t|t-1} \mathbf{Z}_t' + \mathbf{R}_t \mathbf{S}_t) (\mathbf{Z}_t \boldsymbol{\Sigma}_{t|t-1} \mathbf{Z}_t' + \mathbf{H}_t)^{-1} \tilde{\mathbf{y}}_t + \\ &\quad \mathbf{T}_t \hat{\alpha}_{t|t-1} + \mathbf{B}_t \mathbf{u}_t \\ &= (\mathbf{T}_t \boldsymbol{\Sigma}_{t|t-1} \mathbf{Z}_t' + \mathbf{R}_t \mathbf{S}_t) (\mathbf{Z}_t \boldsymbol{\Sigma}_{t|t-1} \mathbf{Z}_t' + \mathbf{H}_t)^{-1} (\mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_t) \\ &\quad + \mathbf{T}_t \hat{\alpha}_{t|t-1} + \mathbf{B}_t \mathbf{u}_t \\ &= \mathbf{T}_t \mathbf{a}_t + \mathbf{B}_t \mathbf{u}_t + \mathbf{K}_t (\mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_t) \end{aligned}$$

con $\mathbf{K}_t = (\mathbf{T}_t \boldsymbol{\Sigma}_{t|t-1} \mathbf{Z}_t' + \mathbf{R}_t \mathbf{S}_t) (\mathbf{Z}_t \boldsymbol{\Sigma}_{t|t-1} \mathbf{Z}_t' + \mathbf{H}_t)^{-1}$.

- ▶ Cuando $\mathbf{S}_t = \mathbf{0}$, \mathbf{K}_t es la matriz de ganancia aparecida anteriormente.
- ▶ Si $\mathbf{B}_t = \mathbf{0}$, la actualización de \mathbf{a}_t es también la que teníamos.

Obtención alternativa del filtro de Kalman (VIII)

- ▶ Consecuentemente

$$\begin{aligned} \mathbf{P}_{t+1} &= E[(\alpha_{t+1} - \mathbf{a}_{t+1})(\alpha_{t+1} - \mathbf{a}_{t+1})'] \\ &= (\mathbf{T}_t - \mathbf{K}_t \mathbf{Z}_t) \mathbf{P}_t (\mathbf{T}_t - \mathbf{K}_t \mathbf{Z}_t)' + \\ &\quad (\mathbf{R}_t - \mathbf{K}_t) \begin{pmatrix} \mathbf{Q}_t & \mathbf{S}_t \\ \mathbf{S}_t & \mathbf{H}_t \end{pmatrix} \begin{pmatrix} \mathbf{R}_t' \\ -\mathbf{K}_t' \end{pmatrix} \end{aligned}$$

- ▶ En el caso en que $\mathbf{S}_t = \mathbf{0}$, este resultado es igual al obtenido anteriormente. Basta sustituir en la expresión anterior

$$\mathbf{K}_t = \mathbf{T}_t \mathbf{P}_t \mathbf{Z}_t' (\mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{H}_t)^{-1}$$

- ▶ **Ejercicio:** Verificar la afirmación anterior.

Obtención alternativa del filtro de Kalman (VII)

¿Cómo evoluciona la matriz de covarianzas del estado?
Restando las dos igualdades siguientes,

$$\begin{aligned} \alpha_{t+1} &= \mathbf{T}_t \alpha_t + \mathbf{B}_t \mathbf{u}_t + \mathbf{R}_t \eta_t \\ \mathbf{a}_{t+1} &= \mathbf{T}_t \mathbf{a}_t + \mathbf{B}_t \mathbf{u}_t + \mathbf{K}_t (\mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_t) \end{aligned}$$

tenemos:

$$\begin{aligned} \alpha_{t+1} - \mathbf{a}_{t+1} &= \mathbf{T}_t (\alpha_t - \mathbf{a}_t) + \mathbf{R}_t \eta_t - \mathbf{K}_t (\mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_t) \\ &= \mathbf{T}_t (\alpha_t - \mathbf{a}_t) + \mathbf{R}_t \eta_t - \mathbf{K}_t (\mathbf{Z}_t \alpha_t + \epsilon_t - \mathbf{Z}_t \mathbf{a}_t) \\ &= (\mathbf{T}_t - \mathbf{K}_t \mathbf{Z}_t) (\alpha_t - \mathbf{a}_t) + \mathbf{R}_t \eta_t - \mathbf{K}_t \epsilon_t \end{aligned}$$

Fuentes de series temporales

Sesión práctica N2 aquí.

Modelos estructurales básicos (BSM)

- ▶ Una familia de modelos muy simples, fáciles de interpretar y útiles en la práctica.
- ▶ Nivel local (“local level”), tendencia lineal local (“local linear trend”) y modelo con estacionalidad.
- ▶ Trabajo seminal de Harvey (varios libros). Un tratamiento moderno en Durbin-Koopman (2001).

Modelo de nivel local

- ▶ Apareció ya como ejemplo.
- ▶ Tiene por expresión:

$$\begin{aligned}\mu_{t+1} &= \mathbf{T}_t \mu_t + \mathbf{R}_t \eta_t & \eta_t &\stackrel{\text{i.i.d.}}{\sim} N(\mathbf{0}, \mathbf{Q}_t) \\ \mathbf{y}_t &= \mathbf{Z}_t \mu_t + \epsilon_t & \epsilon_t &\stackrel{\text{i.i.d.}}{\sim} N(\mathbf{0}, \mathbf{H}_t)\end{aligned}$$

- ▶ Los términos en verde son la unidad (caso univariante) o matrices unidad.
- ▶ En el caso univariante, sólo dos parámetros: $\mathbf{Q}_t = \sigma_\eta$ y $\mathbf{H}_t = \sigma_\epsilon$.

Modelo de tendencia lineal local

- ▶ Tiene por expresión (caso univariante):

$$\begin{aligned}y_t &= \beta_t + \epsilon_t \\ \beta_{t+1} &= \beta_t + \delta_t + \eta_t \\ \delta_{t+1} &= \delta_t + \nu_t\end{aligned}$$

- ▶ El “estado” está formado por $(\beta_t \ \delta_t)$.
- ▶ La matriz de transición del estado es:

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

- ▶ En el caso univariante, sólo tres parámetros: σ_η , σ_ν y σ_ϵ .

Modelos estructurales como ARIMA's (I)

- ▶ Modelos estructurales pueden escribirse como modelos ARIMA en *forma reducida*. Ver Harvey(1989), p. 54–55.
- ▶ La forma reducida incorpora restricciones implícitas en los parámetros.
- ▶ Las restricciones son una ventaja si “nos creemos” el modelo: limitan el ámbito de la estimación y permiten estimar con mayor precisión.

Modelos estructurales como ARIMA's (II)

- **Ejemplo:** Consideremos el modelo de nivel local,

$$\begin{aligned} y_t &= \mu_t + \epsilon_t \\ \mu_t &= \mu_{t-1} + \xi_t \end{aligned}$$

- Equivalentemente,

$$\nabla y_t = \nabla \mu_t + \nabla \epsilon_t = \xi_t + \nabla \epsilon_t$$

- Es ahora fácil calcular la función de autocovarianza:

$$\gamma(0) = \sigma_\xi^2 + 2\sigma_\epsilon^2 \quad \gamma(1) = -\sigma_\epsilon^2 \quad \gamma(\tau) = 0 \quad \forall \tau \geq 2.$$

- ARIMA(0,1,1). Fácil comprobar que $-1/2 \leq \theta_1 \leq 0$.

Modelos estructurales como ARIMA's (IV)

- Utilizando

$$\nabla^2 y_t = \psi_t + \nabla \xi_t + \nabla^2 \epsilon_t$$

se puede calcular:

$$\begin{aligned} \gamma(0) &= 2\sigma_\xi^2 + 6\sigma_\epsilon^2 + \sigma_\psi^2 \\ \gamma(1) &= -\sigma_\xi^2 - 4\sigma_\epsilon^2 \\ \gamma(2) &= \sigma_\epsilon^2 \\ \gamma(\tau) &= 0 \quad \forall \tau \geq 3 \end{aligned}$$

- Estamos ante un ARIMA(0,2,2) restringido.

Modelos estructurales como ARIMA's (III)

- **Ejemplo:** Consideremos el modelo de tendencia local lineal,

$$\begin{aligned} y_t &= \mu_t + \epsilon_t \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + \xi_t \\ \beta_t &= \beta_{t-1} + \psi_t \end{aligned}$$

- Tenemos

$$\begin{aligned} \nabla \beta_t &= \psi_t \\ \nabla \mu_t &= \beta_{t-1} + \xi_t \end{aligned}$$

- Por tanto,

$$\nabla^2 y_t = \nabla \nabla \mu_t + \nabla^2 \epsilon_t = \psi_t + \nabla \xi_{t-1} + \nabla^2 \epsilon_t$$

“Dummies” estacionales (I)

- Cada periodo tiene asociado un efecto característico. ($\alpha_1 = \text{Trim. 1}, \dots, \alpha_4 = \text{Trim. 4}$).
- Por ejemplo (estacionalidad anual, datos trimestrales):

$$\begin{bmatrix} y_t \\ y_{t+1} \\ y_{t+2} \\ y_{t+3} \\ y_{t+4} \\ \vdots \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ \epsilon_{t+1} \\ \epsilon_{t+2} \\ \epsilon_{t+3} \\ \epsilon_{t+4} \\ \vdots \end{bmatrix} =$$

- El “efecto estación” en t es el mismo que en $(t - 4)$
- Redundancia de parámetros.

“Dummies” estacionales (II)

- ▶ La ecuación de observación,

$$y_t = [1 \ 1 \ 0 \ 0 \ 0] \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix} + \epsilon_t$$

debe encontrar en la posición superior del vector de estado $\alpha_1, \alpha_2, \alpha_3, \alpha_4 =$ de manera cíclica.

- ▶ Podríamos también emplear una matriz de observación variable, pero es más complejo.

“Dummies” estacionales (IV)

- ▶ Podemos hacer:

$$\theta_t = \begin{bmatrix} \alpha_{0,t} \\ \alpha_{1,t} \\ \alpha_{2,t} \\ \alpha_{3,t} \\ \alpha_{4,t} \end{bmatrix}$$

- ▶ Como $\alpha_4 = -\alpha_1 - \alpha_2 - \alpha_3$ la ecuación de estado puede escribirse así:

$$\theta_{t+1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & -1 & -1 \end{bmatrix} \theta_t$$

- ▶ Es fácil ver que nos basta con un vector de estado de dimensión 4 (Tendencia + 3 dummies estacionales).

“Dummies” estacionales (III)

- ▶ Necesitamos poner esto en forma de modelo en EE:

$$y_t = [1 \ 1 \ 0 \ 0 \ 0] \begin{bmatrix} \alpha_{0,t} \\ \alpha_{1,t} \\ \alpha_{2,t} \\ \alpha_{3,t} \\ \alpha_{4,t} \end{bmatrix} + \epsilon_t$$

$$\begin{bmatrix} \alpha_{0,t} \\ \alpha_{1,t} \\ \alpha_{2,t} \\ \alpha_{3,t} \\ \alpha_{4,t} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \alpha_{0,t-1} \\ \alpha_{1,t-1} \\ \alpha_{2,t-1} \\ \alpha_{3,t-1} \\ \alpha_{4,t-1} \end{bmatrix}$$

- ▶ La ecuación de estado “recicla” los coeficientes estacionales. Sensato imponer $\alpha_{4,t} = -\alpha_{1,t} - \alpha_{2,t} - \alpha_{3,t}$

“Dummies” estacionales (V)

- ▶ Lo anterior supone estacionalidad “de piñon fijo”.
- ▶ Podemos hacerla evolutiva e integrarla con otras componentes del modelo para obtener, por ejemplo:

$$y_t = \beta_t + \alpha_{1,t} + \epsilon_t$$

$$\beta_{t+1} = \beta_t + \delta_t + \eta_t$$

$$\delta_{t+1} = \delta_t + \nu_t$$

$$\alpha_{1,t+1} = -\alpha_{1,t} - \alpha_{2,t} - \alpha_{3,t} + \phi_t$$

$$\alpha_{2,t+1} = \alpha_{1,t}$$

$$\alpha_{3,t+1} = \alpha_{2,t}$$

- ▶ El estado está formado por $[\beta_t \ \delta_t \ \alpha_{1,t} \ \alpha_{2,t} \ \alpha_{3,t}]$.

Modelo completo con estacionalidad

- ▶ Las ecuaciones de observación y de estado son:

$$y_t = [1 \ 0 \ 1 \ 0 \ 0] \begin{bmatrix} \beta_t \\ \delta_t \\ \alpha_{t,1} \\ \alpha_{t,2} \\ \alpha_{t,3} \end{bmatrix} + \epsilon_t$$

$$\begin{bmatrix} \beta_{t+1} \\ \delta_{t+1} \\ \alpha_{t+1,1} \\ \alpha_{t+1,2} \\ \alpha_{t+1,3} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & -1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \beta_t \\ \delta_t \\ \alpha_{t,1} \\ \alpha_{t,2} \\ \alpha_{t,3} \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \eta_t \\ \nu_t \\ \phi_t \end{bmatrix}$$

¿Que ocurriría con datos mensuales, semanales, diarios...?

- ▶ Mismo principio: con datos mensuales, necesitaríamos 12 – 1 elementos en el vector de estado.
- ▶ Con datos semanales, 52 – 1.
- ▶ Con datos diarios, 365 – 1 (y complicaciones adicionales con los años bisiestos).
- ▶ Vectores de estado tan grandes son poco manejables.
- ▶ Necesitamos forma alternativa de introducir estacionalidad.

Identidades trigonométricas

- ▶ Para el ángulo suma se tiene:

$$\sin(\lambda_j t + \lambda_j) = \cos(\lambda_j) \sin(\lambda_j t) + \sin(\lambda_j) \cos(\lambda_j t)$$

$$\cos(\lambda_j t + \lambda_j) = \cos(\lambda_j) \cos(\lambda_j t) - \sin(\lambda_j) \sin(\lambda_j t)$$

o equivalentemente:

$$\begin{bmatrix} \sin(\lambda_j(t+1)) \\ \cos(\lambda_j(t+1)) \end{bmatrix} = \begin{bmatrix} \cos(\lambda_j) & \sin(\lambda_j) \\ -\sin(\lambda_j) & \cos(\lambda_j) \end{bmatrix} \begin{bmatrix} \sin(\lambda_j t) \\ \cos(\lambda_j t) \end{bmatrix}$$

- ▶ La matriz en el lado derecho de la ecuación “traslada” sinusoides de frecuencia λ_j del momento t al momento $(t+1)$.

Estacionalidad: alternativa Fourier (I)

- ▶ Las ecuaciones de observación y de estado son:

$$y_t = [1 \ 0 \ 1 \ 0] \begin{bmatrix} \beta_t \\ \delta_t \\ \gamma_{t,1} \\ \gamma_{t,2} \end{bmatrix} + \epsilon_t$$

$$\begin{bmatrix} \beta_{t+1} \\ \delta_{t+1} \\ \gamma_{t+1,1} \\ \gamma_{t+1,2} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos(\lambda_j) & \sin(\lambda_j) \\ 0 & 0 & -\sin(\lambda_j) & \cos(\lambda_j) \end{bmatrix} \begin{bmatrix} \beta_t \\ \delta_t \\ \gamma_{t,1} \\ \gamma_{t,2} \end{bmatrix} + \begin{bmatrix} \eta_t \\ \nu_t \\ \omega_{t,1} \\ \omega_{t,2} \end{bmatrix}$$

- ▶ Se pueden emplear tantos armónicos como sea preciso con $\lambda_j = 2\pi j/s, j = 1, \dots, [s/2]$.

Estacionalidad: alternativa Fourier (II)

- ▶ En la representación anterior, salvo por el ruido,

$$\begin{bmatrix} \gamma_{t+1,1} \\ \gamma_{t+1,2} \end{bmatrix} = \begin{bmatrix} \cos(\lambda_j) & \sin(\lambda_j) \\ -\sin(\lambda_j) & \cos(\lambda_j) \end{bmatrix} \begin{bmatrix} \gamma_{t,1} \\ \gamma_{t,2} \end{bmatrix}$$

- ▶ Son las identidades mencionadas: $\gamma_{t,1}$, $\gamma_{t,2}$ son sinusoides de frecuencia angular λ_j .
- ▶ Sólo $\gamma_{t,1}$ interviene en la ecuación de observación, pero $\gamma_{t,2}$ hace falta para describir la transición.

¿No es más simple la primera representación?

- ▶ ¿Por qué no emplear siempre la representación basada en dummies?
- ▶ Si tomamos todos los armónicos posibles, ambas representaciones son equivalentes.
- ▶ La representación utilizando sinusoides tiene la ventaja de (quizá) un vector de estado más pequeño.
- ▶ Permite además soslayar problemas como el de los años bisiestos.

Insertar sesión práctica N3 aquí.

Estacionalidad: alternativa Fourier (III)

- ▶ No es imperativo que nos limitemos a una sola componente armónica.
- ▶ Cada senoide armónica da lugar a un nuevo bloque 2×2 en la matriz de transición.
- ▶ **Ejercicio:** Escribir en detalle las ecuaciones de estado y medida de un modelo de tendencia local lineal con estacionalidades anual y semanal, y datos diarios.

El criterio máximo-verosímil

- ▶ Dada una m.a.s. x_1, \dots, x_n con densidad $f(x; \theta)$ la verosimilitud es:

$$L(\theta; x_1, \dots, x_n) = f(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta)$$

vista como función de θ dada la muestra.

- ▶ Estrategia intuitivamente plausible para seleccionar estimadores.
- ▶ Buenas propiedades en grandes muestras bajo condiciones muy generales (consistencia, eficiencia, insesgadez y normalidad asintóticas).

Cálculo de la verosimilitud con el filtro de Kalman (I)

- ▶ Con datos de serie temporal,

$$f(\mathbf{y}_1, \dots, \mathbf{y}_T; \theta) \neq \prod_{t=1}^T f(\mathbf{y}_t; \theta)$$

- ▶ Sin embargo, sí es cierto que:

$$f(\mathbf{y}_1, \dots, \mathbf{y}_T; \theta) = \prod_{t=1}^T f(\mathbf{y}_t | \mathcal{Y}_{t-1})$$

- ▶ Observemos que,

$$\mathbf{y}_t = \mathbf{Z}_t \alpha_t + \epsilon_t = \mathbf{Z}_t \mathbf{a}_t + \mathbf{Z}_t (\alpha_t - \mathbf{a}_t) + \epsilon_t$$

- ▶ Por tanto,

$$\begin{aligned} E[\mathbf{y}_t | \mathcal{Y}_{t-1}] &= \mathbf{Z}_t \mathbf{a}_t \\ \text{Var}[\mathbf{y}_t | \mathcal{Y}_{t-1}] &= \mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{H}_t \stackrel{\text{def}}{=} \mathbf{F}_t \end{aligned}$$

Cálculo de la verosimilitud con el filtro de Kalman (III)

- ▶ Una pasada del filtro de Kalman permite así calcular cómodamente la verosimilitud para θ dado.
- ▶ Los ingredientes necesarios en $\log L(\theta)$ se calculan ya (salvo $|\mathbf{F}_t|$) en el filtro de Kalman.
- ▶ Maximización numérica de la verosimilitud posible utilizando una rutina numérica que no haga uso de derivadas (e.g. Nelder-Mead)
- ▶ Con algún esfuerzo adicional, se puede hacer que el filtro de Kalman calcule también valores de las derivadas.

Cálculo de la verosimilitud con el filtro de Kalman (II)

- ▶ Llamando $\mathbf{v}_t = \mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_t$, tenemos por tanto:

$$\prod_{t=1}^T f(\mathbf{y}_t | \mathcal{Y}_{t-1}) = \prod_{t=1}^T \frac{1}{(2\pi)^{\frac{p}{2}} |\mathbf{F}_t|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} \mathbf{v}_t' \mathbf{F}_t^{-1} \mathbf{v}_t \right\}$$

- ▶ La expresión anterior depende implícitamente de los parámetros del modelo.
- ▶ Tomando logaritmos, tenemos el logaritmo de la verosimilitud:

$$\log L(\theta) = -\frac{Tp}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^n \left(\log |\mathbf{F}_t| + \mathbf{v}_t' \mathbf{F}_t^{-1} \mathbf{v}_t \right)$$

Cálculo de la verosimilitud con el filtro de Kalman (IV)

- ▶ Hay que tener en cuenta que algunos parámetros están restringidos en su campo de variación (e.g., varianzas de los ruidos no negativas).
- ▶ Otras restricciones son implícitas: \mathbf{Q}_t y \mathbf{H}_t paramétricas y p.s.d. implican restricciones sobre los parámetros.
- ▶ Frecuentes soluciones en la frontera factible (e.g. con varianzas cero de algún o varios ruidos).
- ▶ Parte sustancial del esfuerzo de cálculo es invertir la matriz $\mathbf{F}_t = \mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{H}_t$; tanto más oneroso cuanto mayor dimensión del vector de observación.

Verosimilitud concentrada (I)

- ▶ La verosimilitud sólo depende de los parámetros a través de \mathbf{P}_t y \mathbf{H}_t en $\mathbf{F}_t = \mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{H}_t$
- ▶ Multiplicar por una constante todos los términos en \mathbf{P}_t y \mathbf{H}_t meramente reescala la verosimilitud.
- ▶ Podemos fijar una varianza en “uno” y operar en esa escala. Luego de maximizada la verosimilitud, se restituyen las escalas.
- ▶ **Ventaja:** La optimización se hace en un espacio de dimensión menor ($k - 1$ en lugar de k). La varianza “omitida” cumple automáticamente la condición de no negatividad.

Verosimilitud concentrada (II)

- ▶ **Ejemplo:** En el modelo de nivel local, los parámetros son σ_ϵ^2 y σ_η^2 .
- ▶ La recurrencia $\mathbf{P}_{t+1} = \mathbf{Z}_t \mathbf{P}_t (\mathbf{I} - \mathbf{K}_t)' + \mathbf{R}_t \mathbf{Q}_t \mathbf{R}_t'$ se reduce a:

$$P_{t+1} = P_t \left(\frac{\sigma_\epsilon^2}{P_t + \sigma_\epsilon^2} \right) + \sigma_\eta^2$$

Dividiendo todo entre σ_ϵ^2 tendríamos

$$P_{t+1}^* = P_t^* \left(\frac{1}{P_t^* + 1} \right) + \sigma_\eta^{*2}$$

en que $P_t^* = P_t / \sigma_\epsilon^2$ y $\sigma_\eta^{*2} = \sigma_\eta^2 / \sigma_\epsilon^2$.

Verosimilitud concentrada (III)

- ▶ Sólo tenemos que maximizar ahora respecto de un parámetro, σ_η^{*2} .
- ▶ Una vez que hemos obtenido σ_η^{*2} , filtramos la serie, estimamos

$$\hat{\sigma}_\epsilon^2 = \frac{1}{n} \sum_{i=1}^n (y_t - a_t)^2$$

y recuperamos

$$\begin{aligned} P_t &= P_t^* \times \hat{\sigma}_\epsilon^2 \\ \hat{\sigma}_\eta^2 &= \hat{\sigma}_\eta^{*2} \times \hat{\sigma}_\epsilon^2 \end{aligned}$$

- ▶ Procedimiento general; fijamos en 1 el valor de una varianza, en una de las matrices \mathbf{Q}_t , \mathbf{H}_t .

Proceso secuencial de observaciones (I)

- ▶ Al calcular la verosimilitud

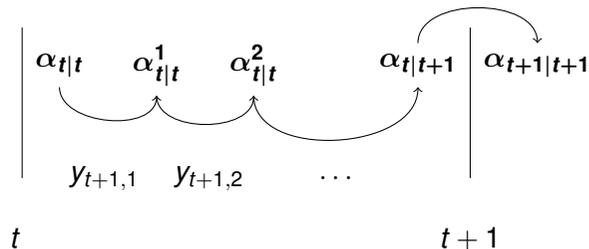
$$\log L(\theta) = -\frac{T\rho}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^n \left(\log |\mathbf{F}_t| + \mathbf{v}_t' \mathbf{F}_t^{-1} \mathbf{v}_t \right)$$

parte sustancial del cálculo es invertir $\mathbf{F}_t = \mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{H}_t$;

- ▶ Tanto más oneroso cuanto mayor dimensión del vector de observación; si la observación fuera escalar, \mathbf{F}_t sería 1×1 y \mathbf{F}_t^{-1} trivial de obtener.
- ▶ Conceptualmente, podemos pretender que vemos \mathbf{y}_t elemento a elemento: $y_{t,1}$, $y_{t,2}$, etc.

Proceso secuencial de observaciones (II)

- ▶ Incorporamos \mathbf{y}_{t+1} por partes, un componente a la vez.
act. temporal



- ▶ En $(t + 1)$ todo \mathbf{y}_{t+1} ha sido incorporado; hacemos una actualización temporal.

Proceso secuencial de observaciones (III)

- ▶ Todo lo que se requiere es utilizar de forma sucesiva las filas de la matriz de observación \mathbf{Z}_t .
- ▶ Procesamos una serie multivariante como si fuera univariante.
- ▶ Para hacer esto, **se requiere que la matriz \mathbf{H}_t sea diagonal.** (¿Por qué?)
- ▶ Si no lo es, pero es invariante en el tiempo ($\mathbf{H}_t = \mathbf{H}$), podemos modificar la ecuación de observación:

$$\mathbf{H}^{-\frac{1}{2}} \mathbf{y}_t = \mathbf{H}^{-\frac{1}{2}} \mathbf{Z}_t \alpha_t + \mathbf{H}^{-\frac{1}{2}} \epsilon_t$$

en que $\epsilon_t^* = \mathbf{H}^{-\frac{1}{2}} \epsilon_t$ cumple la condición requerida.

Proceso secuencial de observaciones (IV)

- ▶ Inversión de la matriz \mathbf{F}_t , sustituida por p cocientes.
- ▶ En lugar de hacer “de una vez”

$$\mathbf{a}_{t|t} = \mathbf{a}_{t|t-1} + \mathbf{P}_t \mathbf{Z}_t' (\mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{H}_t)^{-1} (\mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_{t|t-1})$$

hacemos para $i = 1, \dots, p$:

$$\mathbf{a}_{t|t-1}^{(i)} = \mathbf{a}_{t|t-1}^{(i-1)} + \mathbf{P}_t^{(i-1)} \mathbf{z}_{t,i} \frac{(y_{t,i} - \mathbf{z}_{t,i}' \mathbf{a}_{t|t-1}^{(i-1)})}{\mathbf{z}_{t,i}' \mathbf{P}_t^{(i-1)} \mathbf{z}_{t,i} + \mathbf{H}_{t,ii}}$$

con el convenio $\mathbf{a}_{t|t-1}^{(0)} = \mathbf{a}_{t|t-1} = \mathbf{a}_t$, $\mathbf{P}_t^{(0)} = \mathbf{P}_t$.

- ▶ Análogamente para la matriz de covarianzas \mathbf{P}_t .
- ▶ Tras p medidas, $\mathbf{a}_{t|t-1}^{(p)} = \mathbf{a}_{t|t}$ y $\mathbf{P}_t^{(p)} = \mathbf{P}_{t|t}$.

Proceso secuencial de observaciones (V)

- ▶ Ahorro que puede exceder del 80% del cálculo.
- ▶ Indicado cuando la dimensión del vector de observación es “grande” relativa a la del vector de estado.
- ▶ Resultado idéntico para el vector de estado y matriz de covarianzas que procesando todo a la vez; pero residuos “intermedios” no tienen que ser iguales que cuando procesamos vectores completos.

Suavizado de Kalman (I)

- ▶ Además de predicción o filtrado podemos hacer suavizado:

$$E(\alpha_t | \mathcal{Y}_\tau) \quad t < \tau \leq T.$$

- ▶ Fixed-point smoothing:

$$E(\alpha_t | \mathcal{Y}_\tau) \quad t + 1 \leq \tau \leq T, \quad t \text{ fijo}$$

- ▶ Fixed-lag smoothing:

$$E(\alpha_t | \mathcal{Y}_\tau) \quad \tau = t + K, \quad K \text{ fijo.}$$

- ▶ Fixed interval smoothing: $E(\alpha_t | \mathcal{Y}_T)$.

Suavizado de Kalman (III)

- ▶ Construiremos el suavizador como

$$\tilde{\alpha}_t = \mathbf{K}_f \hat{\alpha}_{f,t} + \mathbf{K}_b \hat{\alpha}_{b,t+1}$$

en que $\hat{\alpha}_{f,t}$ es el resultado del filtro 1 : t y $\hat{\alpha}_{b,t}$ el del filtro $T : (t + 1)$.

- ▶ Si nos limitamos a estimadores insesgados, $\mathbf{K}_b = \mathbf{I} - \mathbf{K}_f$.
- ▶ Sea $\mathbf{e}_f = \alpha_t - \hat{\alpha}_{f,t}$ y $\mathbf{e}_b = \alpha_t - \hat{\alpha}_{b,t+1}$. Entonces,

$$\begin{aligned} \tilde{\alpha}_t - \alpha_t &= \mathbf{K}_f \hat{\alpha}_{f,t} + \mathbf{K}_b \hat{\alpha}_{b,t+1} - \alpha_t \\ \tilde{P} &= E[(\tilde{\alpha}_t - \alpha_t)(\tilde{\alpha}_t - \alpha_t)^T] \\ &= E[(\mathbf{K}_f(\mathbf{e}_f - \mathbf{e}_b) + \mathbf{e}_b)(\mathbf{K}_f(\mathbf{e}_f - \mathbf{e}_b) + \mathbf{e}_b)^T] \\ &= E[\mathbf{K}_f(\mathbf{e}_f \mathbf{e}_f^T + \mathbf{e}_b \mathbf{e}_b^T) \mathbf{K}_f^T + \mathbf{e}_b \mathbf{e}_b^T \\ &\quad - \mathbf{K}_f \mathbf{e}_b \mathbf{e}_b^T - \mathbf{e}_b \mathbf{e}_b^T \mathbf{K}_f^T] \end{aligned}$$

Suavizado de Kalman (II)

- ▶ Diferentes algoritmos. Para *fixed-point* y *fixed-lag*, ver Anderson-Moore, § 7.2 y 7.3.
- ▶ Idea para *fixed-interval*: correr dos filtros de Kalman en sentidos opuestos y combinar los resultados: *forward-backward algorithm* (FB), ver Simon § 9.4.1 ó Gibbs § 9.2.
- ▶ Otros algoritmos para *fixed-interval*: Rauch-Tung-Striebel (RTS) y Bryson-Frazier (BF), mas rápidos que el algoritmo FB: ver Gibbs § 9.2.3 y refs.
- ▶ Mostraremos por su sencillez la obtención del FB.

Suavizado de Kalman (IV)

- ▶ Si derivamos la traza de \tilde{P} respecto de \mathbf{K}_f tenemos:

$$\frac{\partial \text{tr}(\tilde{P})}{\partial \mathbf{K}_f} = 2E[\mathbf{K}_f(\mathbf{e}_f \mathbf{e}_f^T + \mathbf{e}_b \mathbf{e}_b^T) - \mathbf{e}_b \mathbf{e}_b^T]$$

- ▶ De aquí,

$$\mathbf{K}_f(\mathbf{P}_f + \mathbf{P}_b) - \mathbf{P}_b = 0$$

y por tanto (rango completo):

$$\mathbf{K}_f = \mathbf{P}_b(\mathbf{P}_f + \mathbf{P}_b)^{-1}$$

- ▶ Notemos que \mathbf{P}_f es la matriz de covarianzas de error de filtrado y \mathbf{P}_b la matriz de covarianzas del error de predicción una etapa "hacia atrás".

Insertar sesión práctica N4 aquí.

Extensiones: modelos lineales generalizados (I)

- ▶ El modelo de regresión lineal ordinario (MCO) prescribe:

$$y = \mathbf{x}'\beta + \epsilon$$

- ▶ (Junto con una ecuación de estado para β proporciona un modelo en espacio de estado.)
- ▶ Es inadecuado en algunos casos por:
 - ▶ Ser la respuesta de rango limitado.
 - ▶ Ser la respuesta acusadamente no normal.
- ▶ En estos casos, un modelo lineal generalizado (GLM) puede ser la solución.

Extensiones: modelos lineales generalizados (III)

- ▶ Simbólicamente,

$$\begin{aligned} g(\mu) &= \mathbf{x}'\beta \\ y &\sim P(y; \mu) \end{aligned}$$

- ▶ El caso particular $g(\mu) = \mu$ y $P(y; \mu) = N(\mu, \sigma^2)$ nos devuelve al modelo lineal ordinario. . .
- ▶ . . .pero hay otras muchas posibilidades.
- ▶ Casos particulares de interés: regresión de Poisson, logit, probit.

Extensiones: modelos lineales generalizados (II)

- ▶ En el modelo MCO se suponen *dos* cosas diferentes:

$$\begin{aligned} E[y] = \mu(\beta) &= \mathbf{x}'\beta \\ y &\sim N(\mu, \sigma^2) \end{aligned}$$

- ▶ Un modelo lineal generalizado (GLM) hace que
 - ▶ Una *función de la media* $g(\mu)$ dependa linealmente de las \mathbf{x} .
 - ▶ y tenga una distribución no necesariamente normal en torno a esa media (quizá multiplicada por una constante).
- ▶ La función $g(\mu)$ se conoce como función de enlace (*link function*).

Extensiones: modelos lineales generalizados (IV)

- ▶ **Ejemplo:** Regresión de Poisson

$$\begin{aligned} \log(\mu_t) &= \mathbf{x}_t'\beta \\ y_t &\sim \text{Poisson}(\lambda = N_t\mu_t) \end{aligned}$$

- ▶ Aquí, $g() = \log()$ y la distribución es de Poisson. N_t son unidades en riesgo.
- ▶ Notemos que sea cual fuere $\log(\mu_t)$ el parámetro de la Poisson es no negativo.

Extensiones: modelos lineales generalizados (V)

- ▶ **Ejemplo:** Regresión logística

$$\log\left(\frac{p_t}{1-p_t}\right) = \mathbf{x}_t' \boldsymbol{\beta}$$
$$y_t \sim \text{Binomial}(p_t, N_t)$$

- ▶ Aquí, $g() = \log(p/1-p)$ y la distribución es binomial.
- ▶ Notemos que sea cual fuere $\log(p_t/1-p_t)$ el parámetro de la binomial es no negativo.

Modelos de espacio de estado generalizados (II)

- ▶ El filtro de Kalman ordinario ya no sirve.
- ▶ No obstante, es posible (aunque más costosa) la estimación.
- ▶ dlm no la soporta, KFAS sí.
- ▶ Detalles en [3], cap. 10.

Modelos de espacio de estado generalizados (I)

- ▶ Retendremos una dinámica lineal para el vector de estado:

$$\boldsymbol{\alpha}_{t+1} = \mathbf{T}_t \boldsymbol{\alpha}_t + \mathbf{R}_t \boldsymbol{\eta}_t$$

- ▶ En lugar de ecuación de observación, tendremos ahora:

$$y_t \sim p(y_t | \mathbf{Z}_t \boldsymbol{\alpha}_t)$$

- ▶ en que la distribución no necesariamente es normal.
- ▶ El ruido de la ecuación de estado puede también ser o no normal.
- ▶ La distribución $p(y_t | \mathbf{Z}_t \boldsymbol{\alpha}_t)$ típicamente en la familia exponencial.

Flecos

- ▶ Controlabilidad, observabilidad, estabilidad.
- ▶ Divergencia del filtro de Kalman. “Joseph-stabilization”.
- ▶ Algoritmos raíz cuadrada.
- ▶ Inicialización difusa. Filtros de información.
- ▶ Ejemplos de aplicaciones: análisis factorial dinámico, agregación y benchmarking,

Insertar sesión práctica N5 aquí.

Ángulos medidos en radianes

- ▶ *Radian*: Angulo que subtiende un arco de longitud igual al radio de la circunferencia.
- ▶ Como la longitud de la circunferencia es $2\pi r$, 2π radianes equivalen a 360 grados sexagesimales.
- ▶ Consecuentemente, $90 = \pi/2$, $45 = \pi/4$ y $180 = \pi$ radianes.

Funciones trigonometricas (II)

- ▶ Fácilmente se demuestra:

$$\sin(x + y) = \sin(x) \cos(y) + \sin(y) \cos(x)$$

$$\cos(x + y) = \cos(x) \cos(y) - \sin(x) \sin(y)$$

- ▶ En consecuencia

$$\sin(2x) = 2 \sin(x) \cos(x)$$

$$\cos(2x) = \cos^2(x) - \sin^2(x)$$

- ▶ Otra importante relación es:

$$\cos^2(x) + \sin^2(x) = 1$$

para x cualquiera.

Funciones trigonometricas (I)

- ▶ Las funciones $\sin(t)$ y $\cos(t)$ son funciones periódicas. Cuando su argumento t se expresa en radianes, su periodo es de 2π .
- ▶ Consecuentemente,

$$\sin\left(\frac{2\pi t}{N}\right)$$

$$\cos\left(\frac{2\pi t}{N}\right)$$

tienen periodo N (en unidades de t). Su *frecuencia* (en unidades de t^{-1}) es $1/N$.

Funciones trigonometricas (III)

- ▶ Una función como $\cos\left(\frac{2\pi kt}{N} + \phi\right)$ se dice que tiene una *fase* de ϕ radianes (se expresa también en ocasiones en unidades de t).
- ▶ Haciendo uso de

$$\cos(x + y) = \cos(x) \cos(y) - \sin(x) \sin(y)$$

se comprueba que

$$\cos\left(\frac{2\pi kt}{N} + \phi\right) = \cos\left(\frac{2\pi kt}{N}\right) \cos(\phi) - \sin\left(\frac{2\pi kt}{N}\right) \sin(\phi)$$

Funciones trigonometricas (IV)

► En

$$\cos\left(\frac{2\pi kt}{N} + \phi\right) = \cos\left(\frac{2\pi kt}{N}\right)\cos(\phi) - \sin\left(\frac{2\pi kt}{N}\right)\sin(\phi)$$

llamamos:

► *componente en fase a:*

$$\cos(\phi)\cos\left(\frac{2\pi kt}{N}\right)$$

► *componente en cuadratura a:*

$$\sin(\phi)\sin\left(\frac{2\pi kt}{N}\right)$$

Funciones trigonometricas (VI)

► La fórmula de Moivre,

$$e^{iut} = \cos(ut) + i\sin(ut)$$

permite trabajar cómodamente con los componentes en fase y cuadratura simultáneamente.

► Recordemos que el módulo de un número complejo $a + bi$ se define $a^2 + b^2$.

► Se deduce que:

$$|e^{iut}|^2 = \cos^2(ut) + \sin^2(ut) = 1$$

Funciones trigonometricas (V)

► Comparando

$$\cos\left(\frac{2\pi kt}{N} + \phi\right) = \cos\left(\frac{2\pi kt}{N}\right)\cos(\phi) - \sin\left(\frac{2\pi kt}{N}\right)\sin(\phi)$$

con:

$$a\cos\left(\frac{2\pi kt}{N}\right) + b\sin\left(\frac{2\pi kt}{N}\right)$$

vemos que la fase de ésta última expresión es:

$$\text{arc tan}\left(-\frac{b}{a}\right)$$

Frecuencias fundamental y armónicas

► *Frecuencia fundamental:* aquélla cuyo periodo coincide con el intervalo de observación. Si $t = 1, 2, \dots, N$,

$$\sin\left(\frac{2\pi t}{N}\right) \quad \cos\left(\frac{2\pi t}{N}\right)$$

son sinusoides a la frecuencia fundamental.

► Frecuencias *armónicas:* múltiplos enteros de la fundamental:

$$\sin\left(\frac{2\pi kt}{N}\right) \quad \cos\left(\frac{2\pi kt}{N}\right)$$

► *Frecuencia Nyquist* es la del armónico $k = N/2$ (periodo 2 en unidades de t).

Bases de R^N formadas por funciones periódicas

- ▶ Definamos para $k = 0, 1, \dots, N/2$ los vectores en R^N :

$$\mathbf{c}_k = \left\{ \cos\left(\frac{2\pi kt}{N}\right) \right\}_{t=1,2,\dots,N}$$

$$\mathbf{s}_k = \left\{ \sin\left(\frac{2\pi kt}{N}\right) \right\}_{t=1,2,\dots,N};$$

- ▶ $\mathbf{c}_0, \mathbf{c}_1, \dots, \mathbf{c}_{N/2-1}, \mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_{N/2}$ forman una base ortogonal de R^N .

Demostración de la ortogonalidad (II)

- ▶ Para cualesquiera $\mathbf{c}_0, \dots, \mathbf{c}_{N/2-1}, \mathbf{s}_1, \dots, \mathbf{s}_{N/2}$ con $k, l = 0, 1, 2, \dots$ y $k \neq l$,

$$\langle \mathbf{c}_k, \mathbf{c}_l \rangle = \sum_{t=1}^N \frac{e^{i2\pi kt/N} + e^{-i2\pi kt/N}}{2} \frac{e^{i2\pi lt/N} + e^{-i2\pi lt/N}}{2}$$

$$= 0$$

dado que $\sum_{t=1}^N e^{i2\pi(k \pm l)t/N} = 0$ para k, l enteros y distintos.

- ▶ Análogamente $\langle \mathbf{s}_k, \mathbf{s}_l \rangle = 0$ y $\langle \mathbf{s}_k, \mathbf{c}_l \rangle = 0$ (el último, incluso si $k = l$).
- ▶ Por tanto, todos los vectores en

$$\mathbf{c}_0, \mathbf{c}_1, \dots, \mathbf{c}_{N/2-1}, \mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_{N/2}$$

son mutuamente ortogonales.

Demostración de la ortogonalidad (I)

- ▶ Haciendo uso de la fórmula de Moivre,

$$\cos\left(\frac{2\pi kt}{N}\right) = \frac{e^{i2\pi kt/N} + e^{-i2\pi kt/N}}{2}$$

$$\sin\left(\frac{2\pi kt}{N}\right) = \frac{e^{i2\pi kt/N} - e^{-i2\pi kt/N}}{2i}$$

- ▶ Hemos de comprobar que $\langle \mathbf{c}_k, \mathbf{c}_l \rangle = 0$ si $k \neq l$.

Propiedades adicionales de

$\mathbf{c}_0, \dots, \mathbf{c}_{N/2-1}, \mathbf{s}_1, \dots, \mathbf{s}_{N/2}$

- ▶ $|\mathbf{c}_0|^2 = \langle \mathbf{c}_0, \mathbf{c}_0 \rangle = N$.
- ▶ $\langle \mathbf{s}_0, \mathbf{s}_0 \rangle = 0$ y $\langle \mathbf{s}_{N/2}, \mathbf{s}_{N/2} \rangle = N$.
- ▶ Para todos los demás valores de k ,

$$\langle \mathbf{c}_k, \mathbf{c}_k \rangle = \sum_{t=1}^N \cos^2(2\pi kt/N) = \frac{N}{2}$$

$$\langle \mathbf{s}_k, \mathbf{s}_k \rangle = \sum_{t=1}^N \sin^2(2\pi kt/N) = \frac{N}{2}$$

habida cuenta de que $\langle \mathbf{c}_k, \mathbf{c}_k \rangle = \langle \mathbf{s}_k, \mathbf{s}_k \rangle$ y

$$\langle \mathbf{c}_k, \mathbf{c}_k \rangle + \langle \mathbf{s}_k, \mathbf{s}_k \rangle = \sum_{t=1}^N |e^{i2\pi kt/N}|^2 = N$$

Aproximación armónica de una sucesión (I)

- ▶ Consideremos $\mathbf{y} = \{y_t\}_{t=1, \dots, N}$. \mathbf{y} es combinación lineal de $\mathbf{c}_0, \dots, \mathbf{c}_{N/2-1}, \mathbf{s}_1, \dots, \mathbf{s}_{N/2}$:

$$\mathbf{y} = \alpha_0 \mathbf{c}_0 + \sum_{k=1}^{N/2-1} (\alpha_k \mathbf{c}_k + \beta_k \mathbf{s}_k) + \beta_{N/2} \mathbf{s}_{N/2}$$

- ▶ Los α_j, β_j , pueden obtenerse por regresión.
- ▶ Si empleamos todos los “regresores” $\mathbf{c}_0, \dots, \mathbf{c}_{N/2-1}, \mathbf{s}_1, \dots, \mathbf{s}_{N/2}$ se obtiene un ajuste “perfecto”, sin grados de libertad.

Aproximación armónica de una sucesión (III)

- ▶ Análogamente todos los α_k, β_k :

$$\alpha_k = \frac{\langle \mathbf{c}_k, \mathbf{y} \rangle}{\langle \mathbf{c}_k, \mathbf{c}_k \rangle} = \frac{2}{N} \sum_{t=1}^N y_t \cos(2\pi kt/N)$$

$$\beta_k = \frac{\langle \mathbf{s}_k, \mathbf{y} \rangle}{\langle \mathbf{s}_k, \mathbf{s}_k \rangle} = \frac{2}{N} \sum_{t=1}^N y_t \sin(2\pi kt/N)$$

salvo para α_0 y $\beta_{N/2}$ en que las constantes $2/N$ han de ser reemplazadas por $1/N$.

Aproximación armónica de una sucesión (II)

- ▶ Haciendo el producto interno de ambos lados por \mathbf{c}_j obtenemos:

$$\langle \mathbf{c}_j, \mathbf{y} \rangle = \alpha_0 \langle \mathbf{c}_j, \mathbf{c}_0 \rangle + \sum_{k=1}^{N/2-1} (\alpha_k \langle \mathbf{c}_j, \mathbf{c}_k \rangle + \beta_k \langle \mathbf{c}_j, \mathbf{s}_k \rangle) + \beta_{N/2} \langle \mathbf{c}_j, \mathbf{s}_{N/2} \rangle$$

- ▶ Haciendo uso de la ortogonalidad,

$$\alpha_j = \frac{\langle \mathbf{c}_j, \mathbf{y} \rangle}{\langle \mathbf{c}_j, \mathbf{c}_j \rangle} = \frac{2}{N} \sum_{t=1}^N y_t \cos(2\pi jt/N)$$

Aproximación armónica de una sucesión (IV)

- ▶ Una notación más compacta sería:

$$\delta_k = \frac{2}{N} \sum_{t=1}^N y_t e^{i2\pi kt/N},$$

con $k = 1, 2, \dots, N/2 - 1$.

- ▶ Las partes real e imaginaria de δ_k corresponderían a α_k y β_k .
- ▶ Para $k = 0$ ó $k = N/2$, la constante al frente del sumatorio es $1/N$.

Aproximación armónica de una sucesión (V)

- ▶ Recordemos que

$$\mathbf{y} = \alpha_0 \mathbf{c}_0 + \sum_{k=1}^{N/2-1} (\alpha_k \mathbf{c}_k + \beta_k \mathbf{s}_k) + \beta_{N/2} \mathbf{s}_{N/2}$$

- ▶ Una medida de la “importancia” del armónico k -ésimo en la sucesión analizada vendría dada por

$$\alpha_k^2 + \beta_k^2$$

Transformadas de Fourier (I)

- ▶ Denominamos *transformada de Fourier* de la sucesión $\{y_t\}_{t=1,\dots,N}$ a la función

$$\mathcal{Y}(\lambda) = \frac{1}{2\pi N} \sum_{t=1}^N y_t e^{i\lambda t}; \quad (1)$$

la sucesión $\{y_t\}$ puede ser real o compleja, y también infinita, con tal de que sea de cuadrado sumable: $\sum_t |y_t|^2 < \infty$.

- ▶ Denominamos *transformada inversa de Fourier* de una sucesión a una expresión como (1) en que ha cambiado de signo el exponente (y posiblemente la constante frente al sumatorio).

Transformadas de Fourier (Ia)

- ▶ Diferentes textos emplean diferentes normalizaciones:

$$\mathcal{Y}(\lambda) = \frac{1}{2\pi N} \sum_{t=1}^N y_t e^{i\lambda t} \implies \mathcal{Y}(2\pi k/N) = \frac{1}{4\pi} (\alpha_k + i\beta_k)$$

$$\mathcal{Y}(\lambda) = \sum_{t=1}^N y_t e^{i\lambda t} \implies \mathcal{Y}(2\pi k/N) = \frac{N}{2} (\alpha_k + i\beta_k)$$

- ▶ Podemos escoger la que nos convenga.
- ▶ En ocasiones, se intercambian $e^{-i\lambda t}$ y $e^{i\lambda t}$.

Transformadas de Fourier (II)

Propiedad de inversión:

- ▶ Si $\sum_{t=-\infty}^{\infty} |y_t|^2 < \infty$.

$$\mathcal{Y}(\lambda) = \frac{1}{2\pi} \sum_{t=-\infty}^{\infty} y_t e^{i\lambda t} \quad (1)$$

y se verifica:

$$y_t = \int_{-\pi}^{\pi} \mathcal{Y}(\lambda) e^{-i\lambda t} d\lambda \quad (2)$$

- ▶ Se dice que (1)–(2) forman un par de transformadas de Fourier, la segunda siendo la inversa de la primera.

Transformadas de Fourier (IIa)

- ▶ Una “demostración” de (2) se puede lograr sustituyendo $\mathcal{Y}(\lambda)$ por su valor en (1):

$$\begin{aligned}y_t &= \int_{-\pi}^{\pi} \left(\frac{1}{2\pi} \sum_{s=-\infty}^{\infty} y_s e^{i\lambda s} \right) e^{-i\lambda t} d\lambda \\ &= \sum_{s=-\infty}^{\infty} y_s \left(\int_{-\pi}^{\pi} \frac{1}{2\pi} e^{i\lambda(s-t)} d\lambda \right) d\lambda\end{aligned}$$

- ▶ Puede verse que la integral se anula salvo cuando $s = t$, con lo que la igualdad anterior realmente se verifica.

Transformadas de Fourier (IV)

- ▶ La definición de transformada de Fourier (y la propiedad de inversión) son más generales. Para cualquier función de cuadrado integrable, si:

$$\mathcal{G}(\lambda) = \int_{-\infty}^{\infty} g(t) e^{i\lambda t} dt$$

entonces:

$$g(t) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathcal{G}(\lambda) e^{-i\lambda t} d\lambda$$

▶

Transformadas de Fourier (III)

- ▶ La propiedad de inversión muestra que es equivalente conocer $\{y_t\}$ o conocer $\mathcal{Y}(\lambda)$; de una se puede recuperar la otra.
- ▶ La importancia de la propiedad de inversión radica en que podemos trabajar en el dominio en que la manipulación es más simple para luego deshacer el cambio y volver al dominio original.

Transformadas de Fourier (V)

- ▶ Las funciones $\mathcal{G}(\lambda)$ y $g(t)$ en la transparencia anterior forman un *par de transformadas directa en inversa*. Una cancela el efecto de la otra:
- ▶ Pueden emplearse diferentes normalizaciones:

$$\begin{aligned}\mathcal{G}(\lambda) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g(t) e^{i\lambda t} dt \\ g(t) &= \frac{1}{\sqrt{2\pi}} \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathcal{G}(\lambda) e^{-i\lambda t} d\lambda\end{aligned}$$

- ▶ El coeficiente $\frac{1}{\sqrt{2\pi}}$ hace “simétrico” el par de transformadas.

Transformadas de Fourier (VI)

- ▶ De hecho, basta conocer $\mathcal{Y}(\lambda)$ evaluada a las frecuencias armónicas $\lambda_k = 2\pi k/N$ para poder recuperar $\{y_t\}$ (teorema de muestreo).
- ▶ Se tiene así que

$$\mathcal{Y}(\lambda_k) = \frac{1}{2\pi N} \sum_{t=1}^N y_t e^{i\lambda_k t}$$

$$y_t = 2\pi \sum_{k=-N/2}^{N/2-1} \mathcal{Y}(\lambda_k) e^{-i\lambda_k t}$$

La función `fft` de R

- ▶ La función `fft(y)` en R calcula la transformada directa como

$$\sum_{t=0}^{N-1} y_t e^{-i\lambda_k t}$$

- ▶ La función `fft(y, inverse=TRUE)` calcula

$$\sum_{k=0}^{N-1} Y_k e^{i\lambda_k t}$$

- ▶ Para recuperar y es preciso dividir entre N ($=\text{length}(y)$).

Transformadas de Fourier (VII)

- ▶ Diferentes normalizaciones:

$$\mathcal{Y}(\lambda_k) = \frac{1}{2\pi N} \sum_{t=1}^N y_t e^{i\lambda_k t} \implies y_t = 2\pi \sum_{k=-N/2}^{N/2-1} \mathcal{Y}(\lambda_k) e^{-i\lambda_k t}$$

$$\mathcal{Y}(\lambda_k) = \sum_{t=1}^N y_t e^{-i\lambda_k t} \implies y_t = \frac{1}{N} \sum_{k=-N/2}^{N/2-1} \mathcal{Y}(\lambda_k) e^{i\lambda_k t}$$

$$\mathcal{Y}(\lambda_k) = \frac{1}{\sqrt{N}} \sum_{t=1}^N y_t e^{i\lambda_k t} \implies y_t = \frac{1}{\sqrt{N}} \sum_{k=-N/2}^{N/2-1} \mathcal{Y}(\lambda_k) e^{-i\lambda_k t}$$

- ▶ La función `fft` en R calcula la transformada directa de acuerdo con la segunda línea (sin normalizar).

Ejemplo de uso de `fft`

```
> y <- c(1, 4, 8, 7)
> Y <- fft(y)
> Y
[1] 20+0i -7+3i -2+0i -7-3i
> invY <- fft(Y, inverse=TRUE) / length(y)
> invY
[1] 1+0i 4+0i 8+0i 7+0i
> as.real(invY)
[1] 1 4 8 7
#
Comprobación:
> lambda <- 2*pi*(0:3)/4
> sum(y*cos(3*lambda))
[1] -7
> sum(-y*sin(3*lambda))
[1] -3
```

Convolución (I)

Propiedad de convolución

- ▶ Si $\{x_t\}$ y $\{y_t\}$ son sendas sucesiones y definimos una nueva cuyo término general es

$$z_t = \sum_k x_k y_{t-k}$$

entonces:

$$\mathcal{Z}(\lambda) = \mathcal{X}(\lambda) \mathcal{Y}(\lambda)$$

- ▶ La operación de convolución es muy frecuente (ej: al aplicar un filtro lineal a una sucesión) y relativamente laboriosa. El cálculo de $\mathcal{Z}(\lambda)$ en cambio es inmediato, y su transformada inversa proporciona $\{z_t\}$.

Convolución (II)

Propiedad de convolución

- ▶ La propiedad de convolución opera también en sentido inverso: Si

$$\mathcal{Z}(\lambda) = \int_{-\pi}^{\pi} \mathcal{X}(\omega) \mathcal{Y}(\lambda - \omega) d\omega$$

entonces:

$$z_t = x_t y_t \quad \forall t$$

es decir, la convolución en el dominio de la frecuencia es producto en el dominio del tiempo.

Transformada rápida de Fourier (I)

- ▶ La aproximación directa al cálculo de

$$\mathcal{Y}(\lambda_k) = \sum_{t=1}^N y_t e^{-i\lambda_k t}$$

requiere un cálculo $O(N^2)$.

- ▶ Una familia de algoritmos (FFT) hacen el mismo cálculo con un esfuerzo de sólo $O(N \log(N))$.
- ▶ Lo óptimo es que N sea “highly composite” (producto de números primos “pequeños”). Idealmente, $N = 2^n$.

Transformada rápida de Fourier (II)

- ▶ Podemos alargar una sucesión con ceros hasta longitud N' “highly composite”.
- ▶ En R podemos utilizar `nextn` para calcular el siguiente $N' \geq N$ que es altamente compuesto.
- ▶ Salvo para N muy grande o cálculo repetitivo (simulación) no es preciso preocuparse.
- ▶ Reemplazando N por N' las frecuencias armónicas cambian:

$$\frac{2\pi k}{N} \neq \frac{2\pi k}{N'}$$

Ejemplo:

```
> N <- 257
> nextn(N, factors=c(2, 3, 5))
[1] 270
> nextn(N, factors=2)
[1] 512
> 3^3 * 2 * 5
[1] 270
> 2^9
[1] 512
```

Aproximación de funciones

- ▶ Desarrollo en serie de Taylor aproxima una función suficientemente derivable *en un punto*.
- ▶ Desarrollo en serie de Fourier aproxima una función en un intervalo.
- ▶ Buenas propiedades: aproximación “casi” uniforme si la función a aproximar suave (pero fenómeno de Gibbs con discontinuas).
- ▶ ¿Para qué?

La función de densidad espectral

- ▶ Toda sucesión aleatoria estacionaria de segundo orden y media cero puede escribirse así (Cramer, 1941):

$$X_t = \int_{-\pi}^{\pi} e^{i\lambda t} dZ(\lambda)$$

- ▶ $Z(\lambda)$ es un proceso de incrementos ortogonales:

$$E[(Z(\lambda_1) - Z(\lambda_2)) \overline{(Z(\lambda_3) - Z(\lambda_4))}] = 0$$

si $(\lambda_1, \lambda_2) \cap (\lambda_3, \lambda_4) = \emptyset$.

- ▶ Se tiene que $E[Z(d\lambda)\overline{Z(d\lambda)}] = f(\lambda)d\lambda$. Se llama *densidad espectral* a la función $f(\lambda)$.

Interpretación de la densidad espectral (I)

- ▶ La expresión

$$X_t = \int_{-\pi}^{\pi} e^{i\lambda t} dZ(\lambda)$$

muestra X_t como *una suma de oscilaciones incorreladas de diferentes frecuencias*.

- ▶ Las oscilaciones en un pequeño intervalo $d\lambda$ tienen varianza $E[Z(d\lambda)]^2 = f(\lambda)d\lambda$
- ▶ La función de densidad espectral $f(\lambda)$ es una medida de la relevancia de las oscilaciones de frecuencia λ en la evolución de la sucesión aleatoria X_t .

Densidad espectral y función ACF (I)

- ▶ Tenemos (suponemos X_t estacionario de media cero) que:

$$E[X_t \overline{X_{t+\tau}}] = E \left[\int_{-\pi}^{\pi} e^{i\lambda t} dZ(\lambda) \int_{-\pi}^{\pi} e^{-i\omega(t+\tau)} \overline{dZ(\omega)} \right]$$

- ▶ Como $Z(\lambda)$ es de incrementos ortogonales:

$$\begin{aligned} E[X_t \overline{X_{t+\tau}}] &= \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} e^{i\lambda t} e^{-i\omega(t+\tau)} E[Z(d\lambda) \overline{dZ(\omega)}] \\ &= \int_{-\pi}^{\pi} e^{-i\lambda\tau} f(\lambda) d\lambda \end{aligned}$$

- ▶ Por tanto, la ACF es la transformada inversa de Fourier de $f(\lambda)$.

Interpretación de la densidad espectral (II)

- ▶ De

$$\rho(\tau) = E[X_t \overline{X_{t+\tau}}] = \int_{-\pi}^{\pi} e^{-i\lambda\tau} f(\lambda) d\lambda$$

deducimos en particular:

$$\sigma^2 = \rho(0) = \int_{-\pi}^{\pi} f(\lambda) d\lambda$$

- ▶ Por tanto, $f(\lambda)$ proporciona un desglose de la varianza por bandas de frecuencia, corroborando la interpretación anterior.

Densidad espectral y función ACF (II)

- ▶ Tenemos:

$$f(\lambda) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \rho(\tau) e^{i\lambda\tau}$$

- ▶ Para que la anterior expresión converja, basta $\sum_{\tau=-\infty}^{\infty} |\rho(\tau)| < \infty$.
- ▶ Siendo la ACF de una sucesión aleatoria real una función real y par,

$$f(\lambda) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \rho(\tau) \cos(\lambda\tau) = \frac{1}{2\pi} \left\{ \rho(0) + 2 \sum_{\tau=1}^{\infty} \rho(\tau) \cos(\lambda\tau) \right\}$$

Densidad espectral del ruido blanco

- ▶ Consideremos $X_t = \epsilon_t$, “ruido blanco”; $\rho(0) = \sigma_{\epsilon}^2$ y $\rho(\tau) = 0$ para cualquier $\tau \neq 0$.
- ▶ En consecuencia,

$$f(\lambda) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \rho(\tau) \cos(i\lambda\tau) = \frac{\sigma_{\epsilon}^2}{2\pi}$$

- ▶ Todas las frecuencias tienen la misma importancia –como en la luz blanca–.
- ▶ Se verifica

$$\int_{-\pi}^{\pi} f(\lambda) d\lambda = \int_{-\pi}^{\pi} \frac{\sigma_{\epsilon}^2}{2\pi} d\lambda = \sigma_{\epsilon}^2;$$

como sabemos que ocurre de manera general.

Periodograma

- ▶ Llamamos periodograma a:

$$I(\lambda) = \frac{1}{2\pi N} \left| \sum_{t=1}^N y_t e^{i\lambda t} \right|^2 = \frac{1}{2\pi N} \left(\sum_{t=1}^N y_t e^{-i\lambda t} \right) \left(\sum_{s=1}^N y_s e^{i\lambda s} \right)$$

- ▶ $I(\lambda) \propto \alpha_k^2 + \beta_\lambda^2$ en que α_λ y β_k son los coeficientes al “regresar” y_t sobre sinusoides de frecuencia λ

¿Tiene algo que ver $f(\lambda)$ con el periodograma $I(\lambda)$?

- ▶ Puesto que

$$f(\lambda) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \rho(\tau) e^{i\lambda\tau} \quad (1)$$

parece que

$$I(\lambda) = \frac{1}{2\pi} \sum_{\tau=-(N-1)}^{N-1} \hat{\rho}(\tau) e^{i\lambda\tau} \quad (2)$$

sería un estimador natural.

Periodograma $I(\lambda)$ y ACF

- ▶ Se verifica:

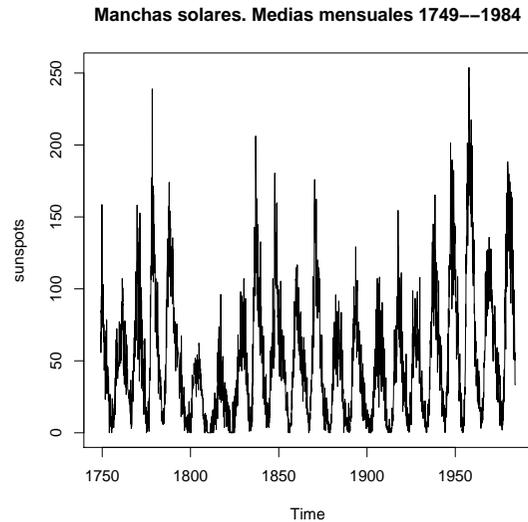
$$\begin{aligned} I(\lambda) &= \frac{1}{2\pi N} \left| \sum_{t=1}^N y_t e^{i\lambda t} \right|^2 = \frac{1}{2\pi N} \left(\sum_{t=1}^N y_t e^{-i\lambda t} \right) \left(\sum_{s=1}^N y_s e^{i\lambda s} \right) \\ &= \frac{1}{2\pi N} \sum_{t=1}^N \sum_{s=1}^N y_t y_s e^{-i\lambda(t-s)} \\ &= \frac{1}{2\pi} \sum_{\tau=-(N-1)}^{N-1} \left[\frac{1}{N} \sum_{t=1}^{N-\tau} y_t y_{t+|\tau|} \right] e^{i\lambda\tau} \\ &= \frac{1}{2\pi} \sum_{\tau=-(N-1)}^{N-1} \hat{\rho}(\tau) e^{i\lambda\tau} \end{aligned}$$

La serie de manchas solares (I)

- ▶ Datos pre-cargados en R. Podemos recuperarlos así:

```
> data(sunspots)
> plot(sunspots)
> str(sunspots)
Time-Series [1:2820] from 1749 to 1984: 58 ...
> plot(sunspots, main="Manchas solares.
      Medias mensuales 1749--1984")
```

La serie de manchas solares (II)

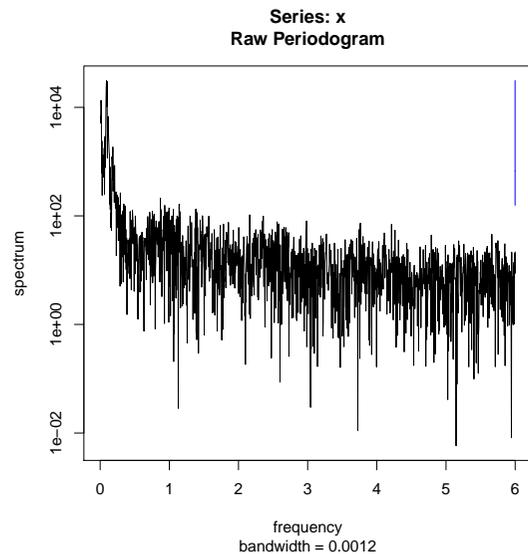


La serie de manchas solares (III)

- ▶ Podemos calcular el periodograma en R así:

```
> spectrum(sunspots)
```
- ▶ Sin argumentos adicionales, `spectrum` calcula el periodograma de la serie.

Periodograma de la serie de manchas solares



Inconsistencia del periodograma

- ▶ Comportamiento muy errático.
- ▶ De una frecuencia a la contigua oscila fuertemente: no es creíble.
- ▶ Un ciclo de ≈ 11 años, claro en los datos, es difícil de ver en el periodograma.
- ▶ Inconsistencia (Schuster, 1898).

Origen de la inconsistencia

- ▶ El número de abscisas crece con la muestra.
- ▶ Supongamos X_t centrada:

$$\begin{aligned} I(\lambda) &= \frac{1}{2\pi N} \left| \sum_{t=1}^N y_t e^{i\lambda t} \right|^2 \\ &= \frac{1}{2\pi N} \left[\left(\sum_{t=1}^N y_t \cos(\lambda t) \right)^2 + \left(\sum_{t=1}^N y_t \sin(\lambda t) \right)^2 \right] \end{aligned}$$

- ▶ Cada sumatorio es aproximadamente una normal de media cero. $I(\lambda) \approx \chi_2^2$ reescalada, sea cual fuere N .

Suavizado del periodograma

- ▶ Suponiendo que $f(\lambda)$ es suave, $f(\lambda_j) \approx f(\lambda_k)$ para frecuencias de Fourier próximas.
- ▶ Por tanto, $E[I(\lambda_j)] \approx E[I(\lambda_k)]$.
- ▶ Si tenemos v.a. de similar valor medio e independientes, el promedio estima el valor medio (casi).
- ▶ La varianza decrecerá:

$$\text{Var} \left(\frac{1}{2k+1} I(\lambda_{j-k}) + \dots + \frac{1}{2k+1} I(\lambda_{j+k}) \right) \approx \frac{1}{2k+1} \text{Var}(I(\lambda_j))$$

Sin embargo, es aproximadamente insesgado

- ▶ Lo podemos comprobar tomando valor medio en

$$I(\lambda) = \frac{1}{2\pi} \sum_{\tau=-(N-1)}^{N-1} \hat{\rho}(\tau) e^{i\lambda\tau}$$

- ▶ Cuando $N \rightarrow \infty$, $E[I(\lambda)] \approx f(\lambda)$.
- ▶ Idea: suavizar para reducir la varianza conservando aproximadamente la insesgadez.

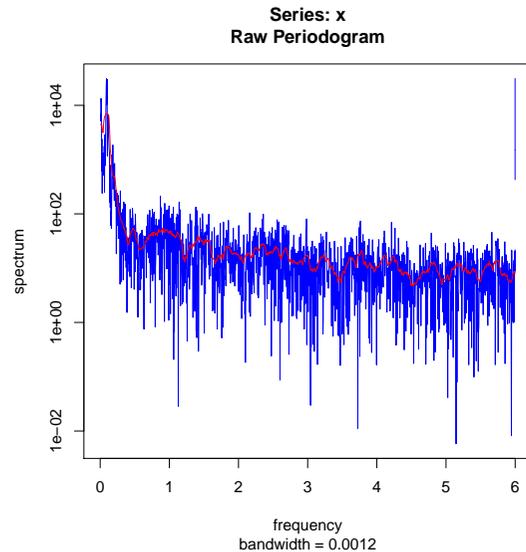
La serie de manchas solares (IV)

- ▶ Podemos suavizar el periodograma en R como sigue. El segundo mandato superpone un periodograma suavizado (línea roja) al bruto (en azul). bruto.

```
> spectrum(sunspots, ci=0.90, col="blue")
```

```
> spectrum(sunspots, spans=21, ci=0.90,
           col="red", add=TRUE)
```

Periodograma bruto y suavizado de sunspots



El periodograma suavizado es una convolución

- ▶ El procedimiento anterior hace la *convolución* de una sucesión (un *kernel*) con otra (el periodograma).

$$\hat{f}(\lambda_k) = \sum_{j=-10}^{10} d_j I(\lambda_{k-j})$$

- ▶ Si no se indica otra cosa, todos los d_j son iguales salvo los extremos, que valen la mitad: $d_j = 0.05$ para $-9, \dots, 9$ y $d_{-10} = d_{10} = 0.025$ (*modified Daniell kernel*).
- ▶ Hay otras posibilidades.

Estimador de covarianza ponderada

- ▶ $I(\lambda)$ es transformada de Fourier de $\hat{\rho}(\tau)$.
- ▶ En lugar de convolver la transformada de Fourier de $\hat{\rho}(\tau)$, podemos transformar $\hat{\rho}(\tau)$ multiplicada por una función de ponderación. Simbólicamente:

$$\mathcal{F}(\hat{\rho}(\tau)) * \mathcal{F}(w(\tau)) = \mathcal{F}(w(\tau)\hat{\rho}(\tau))$$

- ▶ La idea anterior da lugar al estimador llamado de covarianza ponderada:

$$f(\lambda) = \frac{1}{2\pi} \left[\rho(0) + 2 \sum_{\tau=1}^K w(\tau) \rho(\tau) \cos(\lambda\tau) \right]$$

Cuestiones de carpintería

- ▶ A $w(\tau)$ se la llama *ventana de ponderación*.
- ▶ A $W(\lambda) = \mathcal{F}(w(\tau))$ se la llama *ventana espectral*.
- ▶ Se ha estudiado con cuidado la elección de $w(\tau)$ que de lugar a “buenas” ventanas espectrales.
- ▶ El caso más simple de ventana de ponderación es la “ventana truncadora”: $w(\tau) = 1$ para $|\tau| \leq K$.

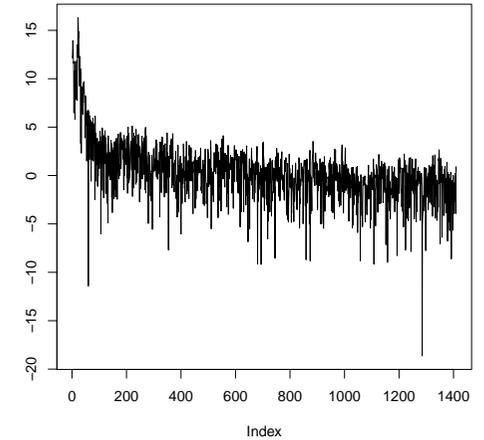
Ejemplo

```
data(sunspots)
N <- length(sunspots)
rho <- acf(sunspots,type="covariance",lag=N)$acf

cov.spec <- function(rho,w,N) {
  r <- rho*w
  r[(N/2+1):N] <- rev(r[2:(N/2+1)])
  Rho <- fft(r)
  return(I <- (1/(2*pi*N)) * Mod(Rho)^2)
}
```

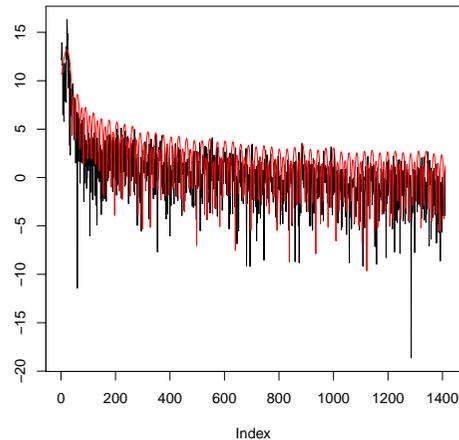
Periodograma obtenido vía covarianza

```
w <- rep(1,N)
I <- cov.spec(rho,w,N)
plot(log(I[1:(N/2)]),
      type="l")
```



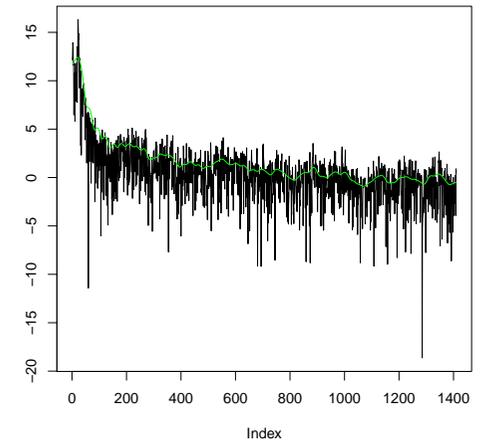
Ventana rectangular con $K = 100$

```
plot(log(I[1:(N/2)]),
      type="l")
w <- rep(0,N)
w[1:100] <- 1
I2 <- cov.spec(rho,w,N)
lines(log(I2[1:(N/2)]),
      col="red")
```

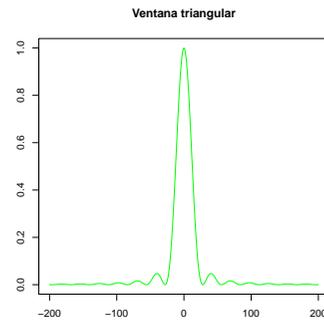
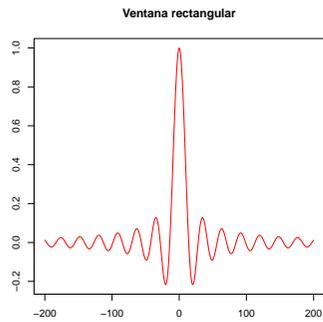


Ventana triangular con $K = 100$

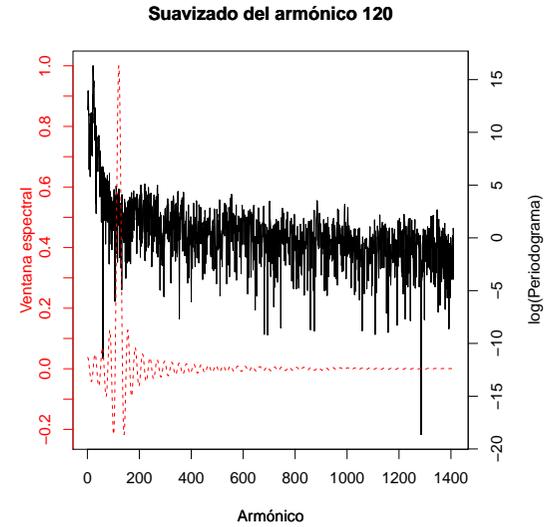
```
w <- rep(0,N)
w[1:100] <- (100:1)/100
I3 <- cov.spec(rho,w,N)
lines(log(I3[1:(N/2)]),
      col="green")
```



Ventanas espectrales asociadas



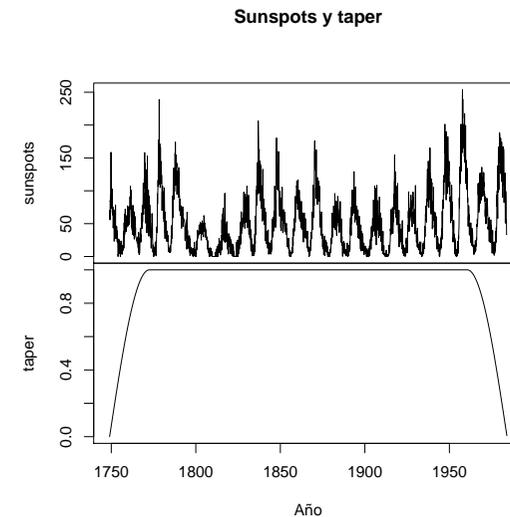
Detalle suavizado con ventana rectangular $K = 100$



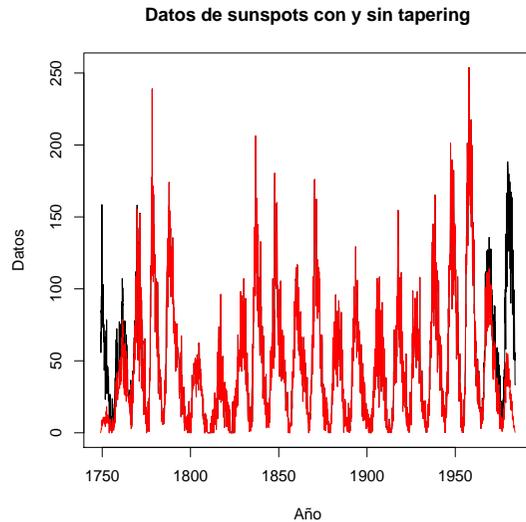
"Tapering"

- ▶ Consiste en amortiguar los datos al comienzo y al final, multiplicándolos por una función que va suavemente a cero.
- ▶ Al **multiplicar** los datos por el *taper*, **convolvemos** su transformada de Fourier con la del *taper*.
- ▶ Parecido a multiplicar la ACF por una ventana, pero hecho directamente sobre los datos.
- ▶ Disminuye el sesgo, pero aumenta la varianza.

"Cosine bell tapering"

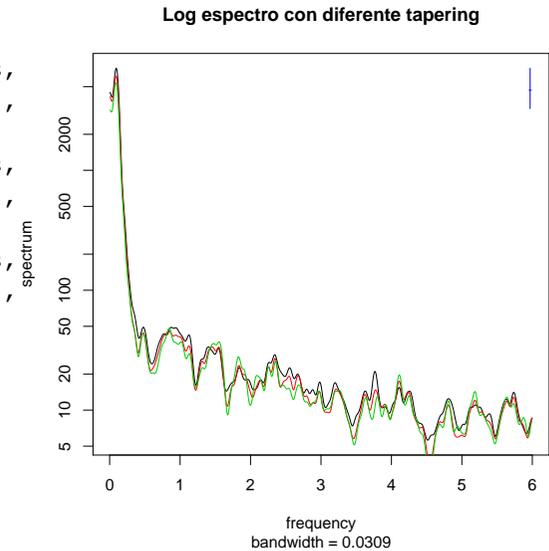


Efecto del “tapering”



Ejemplo: tapering + suavizado

```
s1 <- spectrum(sunspots,
               spans=rep(8,10),
               taper=0.00)
s2 <- spectrum(sunspots,
               spans=rep(8,10),
               taper=0.20)
s3 <- spectrum(sunspots,
               spans=rep(8,10),
               taper=0.40)
```



Multi-tapering

- ▶ En los últimos 25 años se ha popularizado el análisis espectral multi-taper, que parece tener ventajas con los espectros típicos en ciencias físicas.
- ▶ Se emplean *tapers* cuidadosamente escogidos (DPSS, “discrete prolate spheroidal sequences”, de acuerdo con un criterio de optimalidad.
- ▶ Artículo seminal es Thomson(1982), y un tratamiento detallado en Percival & Walden(1992).
- ▶ Implementado en el paquete `sapa` en R.

Filtros lineales invariantes en el tiempo (I)

- ▶ Decimos que un operador \mathcal{L} es un *filtro lineal invariante en el tiempo* si verifica las siguientes propiedades:

- ▶ **Linealidad:**

$$\mathcal{L}(aX_t + bY_t) = a\mathcal{L}(X_t) + b\mathcal{L}(Y_t)$$

- ▶ **Invariancia temporal:**

$$\mathcal{L}(X_t) = \tilde{X}_t \Rightarrow \mathcal{L}(X_{t+\tau}) = \tilde{X}_{t+\tau}$$

- ▶ Medias móviles, retardos, etc. son ejemplos de filtros lineales.

Filtros lineales invariantes en el tiempo (II)

- ▶ Las sinusoides complejas $e^{i\lambda t}$ son “vectores propios” al aplicar un filtro lineal.
- ▶ Tomemos por ejemplo $\mathcal{L}(X_t) = \sum_k a_k X_{t-k}$ (una media móvil de valores de $\{X_t\}$). Si presentamos a este filtro la entrada $e^{i\lambda t}$, obtenemos:

$$e^{i\lambda t} \longrightarrow \boxed{\mathcal{L}} \longrightarrow G(\lambda) e^{i\lambda t}$$

ya que

$$\sum_k a_k e^{i\lambda(t-k)} = \underbrace{\sum_k a_k e^{-i\lambda k}}_{G(\lambda)} e^{i\lambda t}$$

- ▶ $G(\lambda)$ es el “valor propio”.

Filtros lineales invariantes en el tiempo (IV)

- ▶ Como quiera que

$$X_t = \int_{-\pi}^{\pi} e^{i\lambda t} dZ(\lambda)$$

el resultado anterior nos permite deducir que:

$$\begin{aligned} \mathcal{L}(X_t) &= \int_{-\pi}^{\pi} |G(\lambda)| e^{i\phi(\lambda)} e^{i\lambda t} dZ(\lambda) \\ &= \int_{-\pi}^{\pi} e^{i\lambda t} d\tilde{Z}(\lambda) \end{aligned}$$

con

$$d\tilde{Z}(\lambda) = |G(\lambda)| e^{i\phi(\lambda)} dZ(\lambda)$$

Filtros lineales invariantes en el tiempo (III)

- ▶ Podemos examinar con mayor detalle el efecto del filtro si reparamos en que $G(\lambda) = |G(\lambda)| e^{i\phi(\lambda)}$.
- ▶ $|G(\lambda)|$ (módulo) y $\phi(\lambda)$ (argumento) se obtienen de las partes real e imaginaria de $G(\lambda) = A(\lambda) + iB(\lambda)$ así:

$$\begin{aligned} |G(\lambda)| &= \sqrt{A(\lambda)^2 + B(\lambda)^2} \\ \phi(\lambda) &= \arctan \left[\frac{B(\lambda)}{A(\lambda)} \right] \end{aligned}$$

- ▶ Por tanto,

$$e^{i\lambda t} \longrightarrow \boxed{\mathcal{L}} \longrightarrow |G(\lambda)| e^{i\lambda t + \phi(\lambda)};$$

Obtenemos como salida una senoide de la misma frecuencia que la entrada, con su amplitud multiplicada por $|G(\lambda)|$ y un desfase de $\phi(\lambda)$.

Filtros lineales invariantes en el tiempo (V)

- ▶ La densidad espectral de $\mathcal{L}(X_t)$ estará relacionada con la de X_t así:

$$\begin{aligned} f_{\mathcal{L}(X_t)}(\lambda) d\lambda &= E[d\tilde{Z}(\lambda) \overline{d\tilde{Z}(\lambda)}] \\ &= |G(\lambda)|^2 E[dZ(\lambda) \overline{dZ(\lambda)}] \\ &= |G(\lambda)|^2 f_{X_t}(\lambda) d\lambda \end{aligned}$$

- ▶ En definitiva,

$$f_{\mathcal{L}(X_t)}(\lambda) = |G(\lambda)|^2 f_{X_t}(\lambda)$$

- ▶ Se llama *factor de filtro* a la función $|G(\lambda)|^2$. Recoge el efecto del filtro sobre la varianza de la componente de frecuencia λ .

Filtros lineales invariantes en el tiempo (VI)

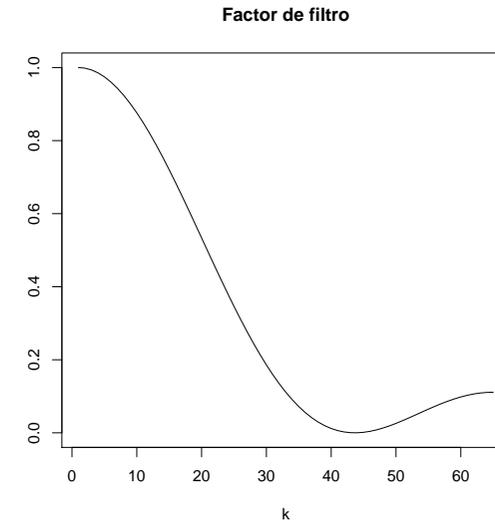
Ejemplo:

- ▶ Consideremos el filtro $\mathcal{L}(X_t) = \frac{1}{3}X_{t-1} + \frac{1}{3}X_t + \frac{1}{3}X_{t+1}$.
- ▶ Entonces,

$$\begin{aligned}G(\lambda) &= \frac{1}{3}e^{-i\lambda} + \frac{1}{3}e^{i \cdot 0} + \frac{1}{3}e^{i\lambda} \\ &= \frac{1}{3}\cos(-\lambda) + \frac{1}{3} + \frac{1}{3}\cos(\lambda) \\ &= \left(\frac{1}{3} + \frac{2}{3}\cos(\lambda)\right)\end{aligned}$$

- ▶ Se trata de una función real: $|G(\lambda)| = G(\lambda)$ y $\phi(\lambda) = 0$ (siempre ocurre con filtros simétricos).

Filtros lineales invariantes en el tiempo (VII)



Filtros lineales invariantes en el tiempo (VIII)

- ▶ En abscisas, k . La frecuencia es $\lambda_k = 2\pi k/N$, con $N = 128$.
- ▶ Examinando la gráfica, vemos que el filtro deja pasar las frecuencias bajas y atenúa fuertemente las correspondientes a $k \approx 44$.
- ▶ $k = 44$ corresponde a $\lambda_k = 2\pi 44/128$ o a un periodo de $128/44 \approx 2.90$ en unidades de t

Preblanqueo (*prewhithening*)

- ▶ Idea original de John Tukey.
- ▶ Trata de evitar los “goteos” (*leakage*).
 - ▶ Hacemos un análisis piloto.
 - ▶ Decidimos que filtro aplicar para que el espectro de la serie filtrada sea aproximadamente ruido blanco.
 - ▶ Estimamos el espectro de la serie transformada.
 - ▶ Restituimos el efecto del filtro.
- ▶ ¡Como el periodograma que se calcula lo es de un “casi” ruido blanco, el goteo “casi” desaparece!

Estimador autorregresivo

- ▶ Puede verse como una variación de la idea anterior.
- ▶ Un modelo autorregresivo puede verse como un filtro blanqueador. Entonces,
 - ▶ Se ajuste un modelo AR (orden, típicamente, mediante AIC):

$$\phi(L)y_t = \epsilon_t$$

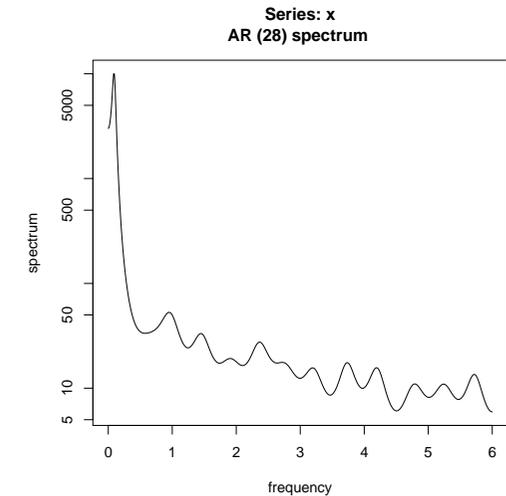
- ▶ Se estima la varianza de la perturbación, σ_ϵ^2 .
- ▶ Se estima el espectro por

$$|\Phi(\lambda)|^2 \hat{f}(\lambda) = \frac{\hat{\sigma}^2}{2\pi} \implies \hat{f}(\lambda) = \frac{\hat{\sigma}^2}{2\pi} \frac{1}{|\Phi(\lambda)|^2}$$

en que $|\Phi(\lambda)|^2$ es el factor de filtro asociado a $\phi(L)$.

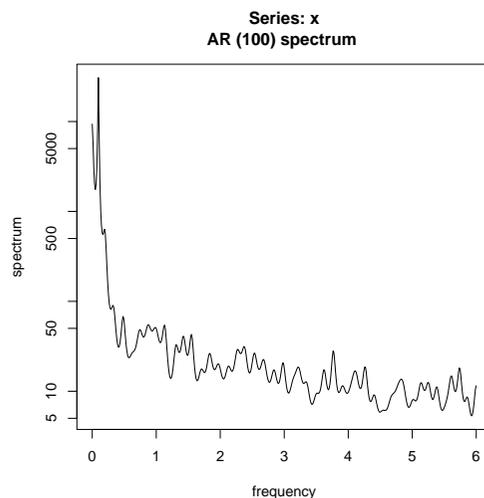
Ejemplo de estimación autorregresiva

```
data (sunspots)
spectrum (sunspots,
method="ar")
```



Relación order AR y suavizado

```
data (sunspots)
spectrum (sunspots,
method="ar",
order=100)
```



Otros estimadores paramétricos

- ▶ También posibles estimadores ARMA

$$f(\lambda) = \frac{\hat{\sigma}^2 \Theta(\lambda)}{2\pi \Phi(\lambda)}$$

- ▶ Más populares los puramente autorregresivos (“all pole” o “todo polos” en la jerga de los ingenieros), por la facilidad para escoger el orden automáticamente.

Variantes estimador autorregresivo

- ▶ Se ha propuesto diversos estimadores (Burg, Yule-Walker, MV (MLE) , MC (LS), diversas variantes de MC), todas equivalentes en grandes muestras.
- ▶ En muestras reducidas puede haber diferencias notables.
- ▶ El estimador paramétrico AR ha sido justificado desde diferentes puntos de vista (Burg, máxima entropía).

Densidad espectral cruzada

- ▶ La función $f_{12}(\lambda)$ tiene una parte real y una parte imaginaria.
- ▶ La parte real mide la covarianza entre oscilaciones de la misma frecuencia “en fase”: $\cos(\lambda t)$ con $\cos(\lambda t)$ y $\sin(\lambda t)$ con $\sin(\lambda t)$.
- ▶ La parte imaginaria mide la covarianza entre oscilaciones “fuera de fase”.
- ▶ $|f_{12}(\lambda)|^2$ es la “covarianza total” al cuadrado.

Procesos bivariantes

- ▶ Supongamos un proceso estacionario bivalente:
 $\mathbf{X}_t = (X_{1t}|X_{2t})^T$. Se tiene,

$$\mathbf{X}_t = \int_{-\pi}^{\pi} e^{i\lambda t} \mathbf{Z}(d\lambda)$$

- ▶ $\mathbf{Z}(d\lambda) = (Z_1(d\lambda)|Z_2(d\lambda))^T$.
- ▶ Como antes, $E[Z_\ell(d\lambda)\overline{Z_\ell(d\lambda)}] = f_\ell(\lambda)d\lambda$, para $\ell = 1, 2$.
- ▶ Ahora, además,

$$E[Z_1(d\lambda)\overline{Z_2(d\lambda)}] = f_{12}(\lambda)d\lambda$$

Coherencia y fase

- ▶ Como $f_1(\lambda)$ y $f_2(\lambda)$ eran respectivamente las varianzas de las oscilaciones de frecuencia λ ,

$$\rho_{12}^2 = \frac{|f_{12}(\lambda)|^2}{f_1(\lambda)f_2(\lambda)}$$

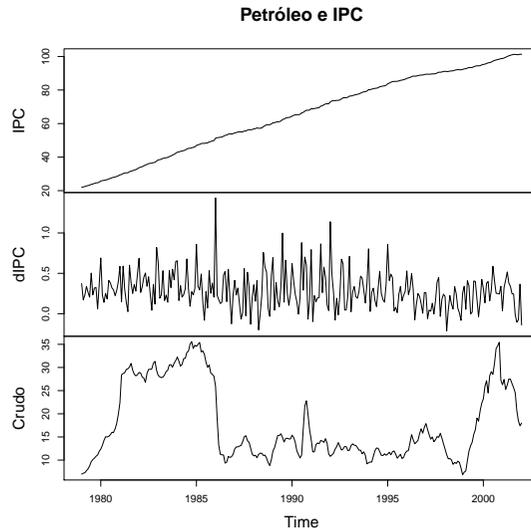
es el “coeficiente de correlación al cuadrado” entre oscilaciones de frecuencia λ componiendo X_{1t} y X_{2t} .

- ▶ Más información en $f_{12}(\lambda)$:

$$\tan^{-1} \left[\frac{\text{Im}(f_{12}(\lambda))}{\text{Re}(f_{12}(\lambda))} \right]$$

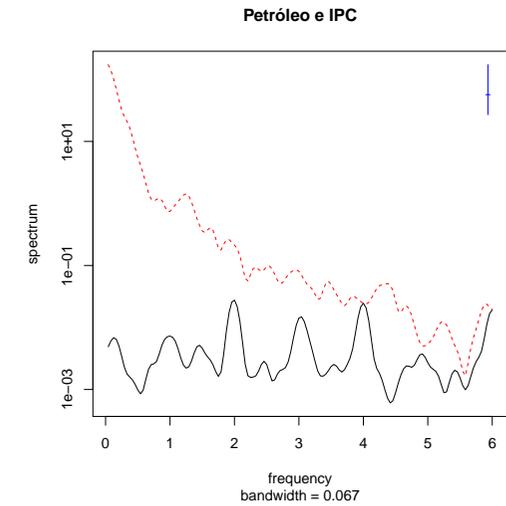
mide el desfase en radianes entre oscilaciones de frecuencia λ componiendo X_{1t} y X_{2t} .

Ejemplo: petróleo e índice de precios



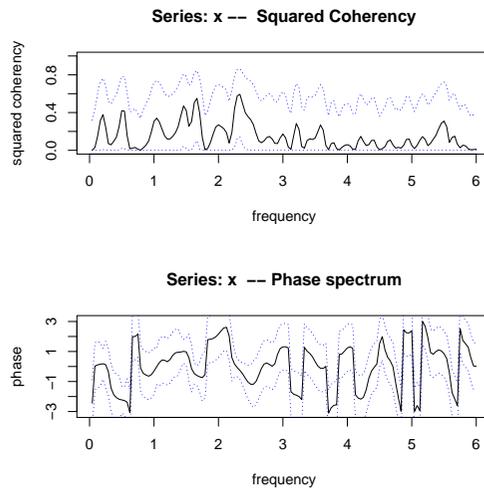
Coherencia y fase de petróleo e IPC (I)

```
datos <- ts.union(dIPC,  
Crudo)  
datos <- window(datos,  
start=c(1979,1),  
end=c(2002,1))  
sp <- spectrum(datos,  
main="Petróleo e IPC",  
spans=rep(3,5))
```



Coherencia y fase de petróleo e IPC (II)

```
par(mfrow=c(2,1))  
plot(sp,plot.type="coh")  
plot(sp,plot.type="phase")
```



Ventajas análisis espectral multivariante

- ▶ Esencialmente cualquier análisis en el dominio del tiempo puede trasladarse al dominio de la frecuencia: componentes principales, análisis factorial, correlación canónica...
- ▶ Resultados muy ricos y descriptivos: ¡un análisis por cada banda de frecuencia!

Inconvenientes análisis espectral multivariante

- ▶ Estimación de la coherencia es MUY delicada. Habitualmente no coincide el suavizado deseable para los espectros y para la densidad espectral cruzada.
- ▶ Intervalos de confianza muy amplios. . .
- ▶ . . .salvo que se tengan series descomunadamente largas.
- ▶ La fase es particularmente difícil de estimar.
- ▶ Por todo ello: instrumento más para las Ciencias de la Naturaleza que para las Ciencias Sociales.

Libros y monografías I

-  B.D.O. Anderson and J.B. Moore.
Optimal Filtering.
Prentice-Hall, 1979.
-  T. W. Anderson.
The Statistical Analysis of Time Series.
Wiley, New York, 1971.
-  J. Durbin and S. J. Koopman.
Time Series Analysis by State Space Methods.
Oxford Univ. Press, New York, 2001.
-  A. Gelb, editor.
Applied Optimal Estimation.
MIT Press, 1974.

Libros y monografías II

-  Bruce P. Gibbs.
*Advanced Kalman Filtering, Least-Squares and Modeling:
A Practical Handbook.*
John Wiley & Sons, 2011.
-  M.S. Grewal and A.P. Andrews.
Kalman Filtering : Theory and Practice Using MATLAB.
Wiley-Interscience, second edition, 2001.
-  D. Simon.
*Optimal State Estimation: Kalman, H Infinity, and Nonlinear
Approaches.*
John Wiley & Sons, United States, 2006.