Curso de Aprendizaje Automático
para el INE

Introducción a Modelos Dinámicos Lineales

David Rios Insua
AXA-ICMAT Chair y Real Academia de Ciencias
16 de Mayo, INE

ICMAT INSTITUTO DE CIENCIAS MATEMÁTICAS

Agenda

- **Recordatorio modelo normal-normal**
- Series temporales: Aproximaciones
- Ejemplo introductorio de modelo dinámico lineal
- Modelo dinámico lineal (y librería dlm)
- Modelos dinámicos no lineales
- Referencias

Modelo normal-normal

Observaciones ciid dados θ, σ^2 son $\mathrm{N}(\theta, \sigma^2)$

A priori impropia $f(\theta) \propto 1$

$$f(\theta \mid \mathbf{x}) \propto \exp\left(-\frac{n}{2}\left(\frac{\theta^2}{\sigma^2} - 2\frac{\theta\bar{x}}{\sigma^2}\right)\right) \qquad \theta \mid \mathbf{x} \sim \mathrm{N}\left(\bar{x}, \tfrac{\sigma^2}{n}\right)$$

A priori propia $\theta \sim \mathrm{N}(\mu_0, \sigma_0^2)$

$$f(\theta \mid \mathbf{x}) \propto \exp\left(-\frac{1}{2}\theta^2\left(\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}\right) - 2\theta\left(\frac{\sum x_i}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}\right)\right)$$

$$\theta \mid \mathbf{x} \sim \mathrm{N}\left(\tfrac{n\bar{x}/\sigma^2 + \mu_0/\sigma_0^2}{n/\sigma^2 + 1/\sigma_0^2}, \left(\tfrac{n}{\sigma^2} + \tfrac{1}{\sigma_0^2}\right)^{-1}\right)$$

Modelo normal-normal

Con a priori no informativa, la predictiva

$$X_{n+1} \mid x \sim \mathrm{N}\left(\bar{x}, \frac{n+1}{n}\sigma^2\right)$$

Intervalo predictivo

$$\left[\bar{x} - z_{\alpha/2}\sigma\sqrt{(n+1)/n}, \bar{x} + z_{\alpha/2}\sigma\sqrt{(n+1)/n}\right]$$

Agenda

- Recordatorio modelo normal-normal
- **Series temporales: Aproximaciones**
- Ejemplo introductorio de modelo dinámico lineal
- Modelo dinámico lineal (y librería dlm)
- Modelos dinámicos no lineales
- Referencias

Modelos paramétricos en series temporales

- Intercambiabilidad: Independencia condicionada a los parámetros.
- Dependencia de las observaciones respecto de su propio parámetro (siendo estos independientes o condicionalmente independientes)
- Cuando las observaciones son una serie temporal, i.e. son observaciones en el tiempo, típicamente tendremos que modelizar la influencia del tiempo sobre la serie, tal vez a través de la evolución de los parámetros.

Series temporales: Tres Aproximaciones Principales

- Modelos en el dominio de tiempos. ARIMA. Box Jenkins
- Modelos en dominio de frecuencias. Análisis espectral.
- Modelos de espacio de estados. La serie es la salida de un sistema dinámico sujeto a perturbaciones aleatorias

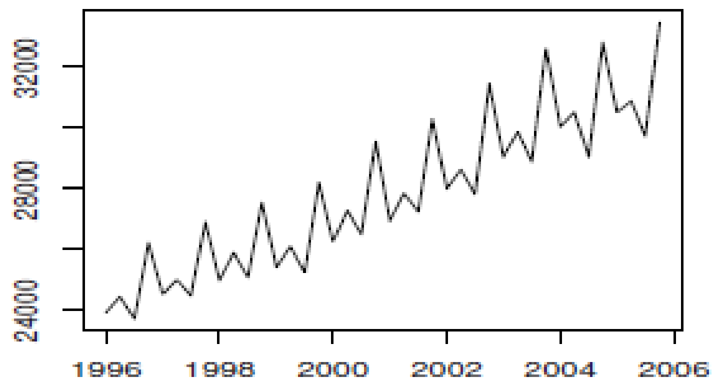
Modelos de espacio de estados: Ventajas

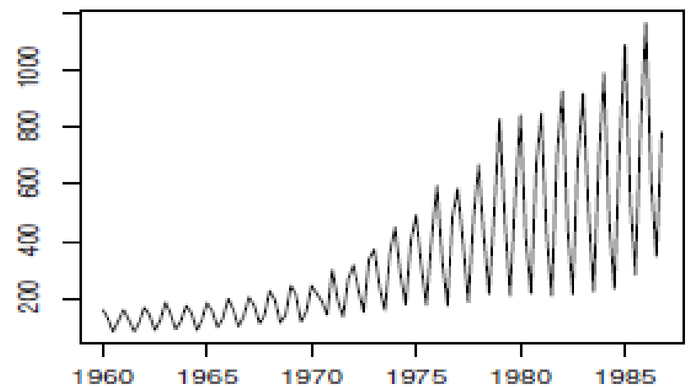
- Interpretación natural como combinación de componentes: tendencia, estacional, regresión, autorregresión.
- Estructura probabilística potente que ofrece un marco de modelización flexible para muchas aplicaciones.
- Computación mediante algoritmos recursivos.
- Tratamiento muy natural desde la perspectiva bayesiana
  - Tratamiento de datos ausentes
- Uni y multivariante
- No estacionariedad, cambios estructurales, patrones irregulares tratables de forma natural
- Lineal (filtro de Kalman) y no lineal (gracias a métodos MCMC).
- Muy útil en muchos campos de aplicación
- Ejemplos de estadística oficial en documento Bayoff

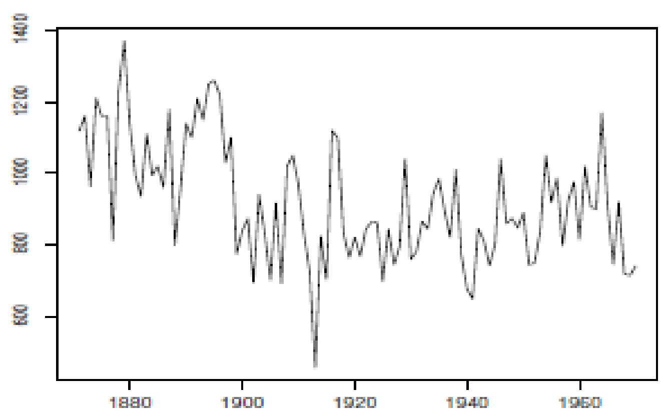
Modelos de espacio de estados: Historia

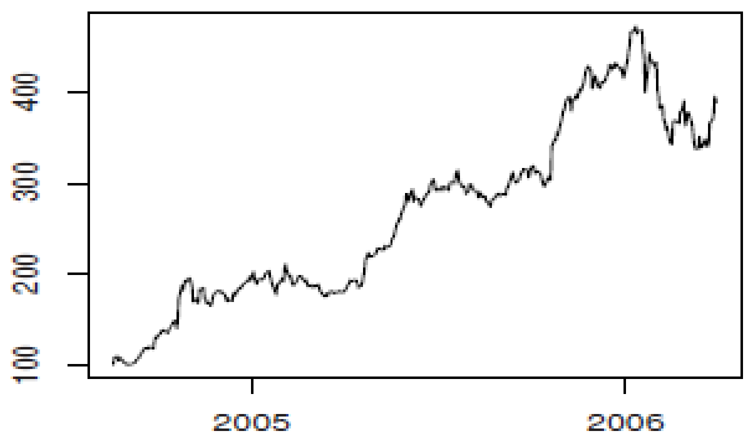
- Kolmogorov (1941)
- Wold (1938) y Wiener (1949)
- Kalman (1960)
- Orbitas del Voyager, Reconocimiento del habla, guiado de aviones,…
- Harrison y Stevens (1976), Akaike (1974)
- West, Harrison (1997)
- dlm (Petris) (2011)
- bsts (Scott) (2014)

Modelos de espacio de estados: Rasgos en series temporales









Agenda

- Recordatorio modelo normal-normal
- Series temporales: Aproximaciones
- **Ejemplo introductorio de modelo dinámico lineal**
- Modelo dinámico lineal (y librería dlm)
- Modelos dinámicos no lineales
- Referencias

Modelos de espacio de estados
Un ejemplo sencillo

- Queremos obtener la posición de un objeto a partir de medidas sujetas a errores aleatorios

$$\theta \qquad\qquad (Y_t : t = 1, 2, \ldots)$$

$$Y_t = \theta + \epsilon_t, \qquad \epsilon_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$$

$$Y_1, Y_2, \ldots | \theta \overset{\text{iid}}{\sim} \mathcal{N}(\theta, \sigma^2). \qquad\qquad \theta \sim \mathcal{N}(m_0, C_0)$$

$$\theta | y_{1:n} \sim \mathcal{N}(m_n, C_n),$$

$$m_n = \mathrm{E}(\theta | y_{1:n}) = \frac{C_0}{C_0 + \sigma^2/n}\bar{y} + \frac{\sigma^2/n}{C_0 + \sigma^2/n} m_0$$

$$C_n = \mathrm{Var}(\theta | y_{1:n}) = \left(\frac{n}{\sigma^2} + \frac{1}{C_0}\right)^{-1} = \frac{\sigma^2 C_0}{\sigma^2 + nC_0}.$$

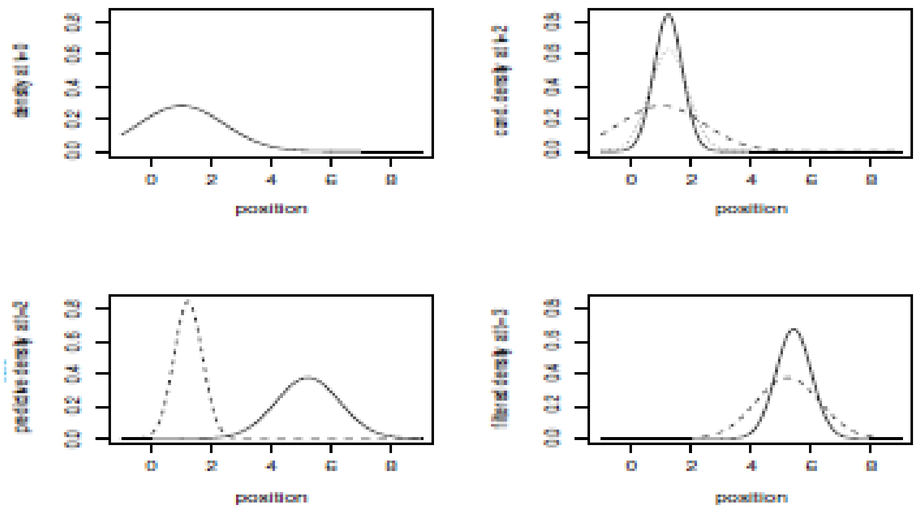
Recursivamente

$$\begin{aligned} m_n &= \frac{C_{n-1}}{C_{n-1} + \sigma^2} y_n + \left(1 - \frac{C_{n-1}}{C_{n-1} + \sigma^2}\right) m_{n-1} \\ &= m_{n-1} + \frac{C_{n-1}}{C_{n-1} + \sigma^2}(y_n - m_{n-1}) \end{aligned}$$

$$C_n = \left(\frac{1}{\sigma^2} + \frac{1}{C_{n-1}}\right)^{-1} = \frac{\sigma^2 C_{n-1}}{\sigma^2 + C_{n-1}}.$$

Modelos de espacio de estados
Un ejemplo sencillo

Tiempo	Observación	Media	Varianza
0	-	1	2
1	1.3	1.24	0.4
2	1.2	1.222	0.222
3	……..	………….	………….



Modelos de espacio de estados
Un ejemplo sencillo

- Introducimos una componente dinámica en el problema. A partir de t=2, el objeto comienza a moverse con cierta velocidad

$$\theta_t = \theta_{t-1} + \nu + w_t, \qquad w_t \sim \mathcal{N}(0, \sigma_w^2) \qquad\qquad Y_t = \theta_t + \epsilon_t, \qquad \epsilon_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$$

con ambos procesos de ruidos independientes.

- Paso inicial: $\theta_2|y_{1:2} \sim \mathcal{N}(m_2 = 1.222, C_2 = 0.222)$
- Predicción: $\theta_3|y_{1:2} \sim \mathcal{N}(a_3, R_3)$ $\qquad Y_3|y_{1:2} \sim \mathcal{N}(f_3, Q_3)$

$$a_3 = \mathrm{E}(\theta_2 + \nu + w_3|y_{1:2}) = m_2 + \nu = 5.722 \qquad f_3 = \mathrm{E}(\theta_3 + \epsilon_3|y_{1:2}) = a_3 = 5.722$$

$$R_3 = \mathrm{Var}(\theta_2 + \nu + w_3|y_{1:2}) = C_2 + \sigma_w^2 = 1.122 \qquad Q_3 = \mathrm{Var}(\theta_3 + \epsilon_3|y_{1:2}) = R_3 + \sigma^2 = 1.622$$

Modelos de espacio de estados
Un ejemplo sencillo

- Paso inicial: $\theta_2|y_{1:2} \sim \mathcal{N}(m_2 = 1.222, C_2 = 0.222)$
- Predicción: $\theta_3|y_{1:2} \sim \mathcal{N}(a_3, R_3)$ $\quad Y_3|y_{1:2} \sim \mathcal{N}(f_3, Q_3)$

$$a_3 = \mathrm{E}(\theta_2 + \nu + w_3|y_{1:2}) = m_2 + \nu = 5.722 \qquad f_3 = \mathrm{E}(\theta_3 + \epsilon_3|y_{1:2}) = a_3 = 5.722$$

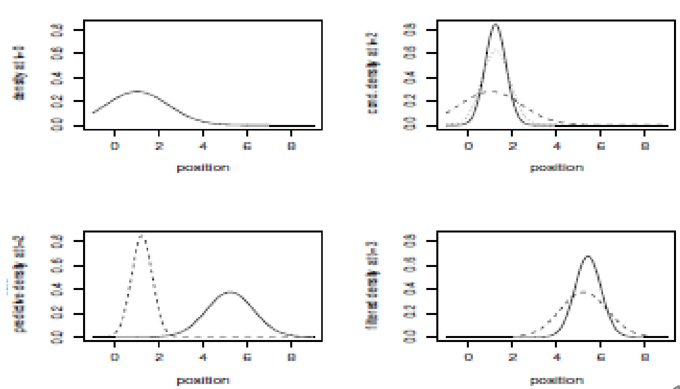
$$R_3 = \mathrm{Var}(\theta_2 + \nu + w_3|y_{1:2}) = C_2 + \sigma_w^2 = 1.122 \qquad Q_3 = \mathrm{Var}(\theta_3 + \epsilon_3|y_{1:2}) = R_3 + \sigma^2 = 1.622$$

- Filtrado: $Y_3 = 5$ $\qquad e_t = y_t - f_t = -0.722$

$$\theta_3|y_1, y_2, y_3 \sim \mathcal{N}(m_3, C_3)$$

$$m_3 = a_3 + \frac{R_3}{R_3 + \sigma^2}(y_3 - f_3) = 5.568$$

$$C_3 = \frac{\sigma^2 R_3}{\sigma^2 + R_3} = R_3 - \frac{R_3}{R_3 + \sigma^2} R_3 = 0.346$$



Modelos de espacio de estados Lecciones del ejemplo introductorio

- Un proceso observable determinado por un proceso latente, salvo error normal
- El proceso latente sólo depende del estado anterior de forma lineal, salvo error normal
- Podemos hacer predicción y estimación de forma secuencial, a media que vamos recibiendo datos

- La linealidad y la normalidad es propia de los modelos dinámicos lineales
- La estructura de dependencia temporal, de los modelos de espacio de estados

Agenda

- Recordatorio modelo normal-normal
- Series temporales: Aproximaciones
- Ejemplo introductorio de modelo dinámico lineal
- **Modelo dinámico lineal (y librería dlm)**
- Modelos dinámicos no lineales
- Referencias

Modelos de espacio de estados

- Una serie temporal (θ_t) p-variante y otra serie temporal Y_t m-variante tales que
  - (θ_t) es una cadena de Markov
  - Dado (θ_t), Y_t son independientes e Y_t sólo depende de θ_t

$$
\begin{array}{ccccccccccccccc}
\theta_0 & \longrightarrow & \theta_1 & \longrightarrow & \theta_2 & \longrightarrow & \cdots & \longrightarrow & \theta_{t-1} & \longrightarrow & \theta_t & \longrightarrow & \theta_{t+1} & \longrightarrow & \cdots \\
 & & \downarrow & & \downarrow & & & & \downarrow & & \downarrow & & \downarrow & & \\
 & & Y_1 & & Y_2 & & & & Y_{t-1} & & Y_t & & Y_{t+1} & &
\end{array}
$$

$$\pi(\theta_{0:t}, y_{1:t}) = \pi(\theta_0) \cdot \prod_{j=1}^{t} \pi(\theta_j|\theta_{j-1})\pi(y_j|\theta_j)$$

Modelo dinámico lineal

- Es un modelo de espacio de estados con
  - Ecuación de observación

$$Y_t = F_t\theta_t + v_t, \qquad v_t \sim \mathcal{N}_m(0, V_t)$$

  - Ecuación de estado (o de sistema)

$$\theta_t = G_t\theta_{t-1} + w_t, \qquad w_t \sim \mathcal{N}_p(0, W_t)$$

  - Distribución a priori

$$\theta_0 \sim \mathcal{N}_p(m_0, C_0)$$

Modelo de espacio de estados

- Es un modelo con
  - Ecuación de observación

$$Y_t = h_t(\theta_t, v_t)$$

  - Ecuación de estado (o de sistema)

$$\theta_t = g_t(\theta_{t-1}, w_t)$$

  - Distribución a priori

$$\theta_0 \sim \mathcal{N}_p(m_0, C_0)$$

MDL: Camino aleatorio con ruido Modelo de nivel local

- Es un modelo de espacio de estados con

$$
\begin{aligned}
Y_t &= \mu_t + v_t, & v_t &\sim \mathcal{N}(0, V) \\
\mu_t &= \mu_{t-1} + w_t, & w_t &\sim \mathcal{N}(0, W)
\end{aligned}
$$

MDL: Modelo de crecimiento lineal

- Es un modelo de espacio de estados con

$$
\begin{aligned}
Y_t &= \mu_t + v_t, & v_t &\sim \mathcal{N}(0, V), \\
\mu_t &= \mu_{t-1} + \beta_{t-1} + w_{t,1}, & w_{t,1} &\sim \mathcal{N}(0, \sigma_\mu^2) \\
\beta_t &= \beta_{t-1} + w_{t,2}, & w_{t,2} &\sim \mathcal{N}(0, \sigma_\beta^2)
\end{aligned}
$$

Ejercicio

- Simulamos en R observaciones del modelo de nivel local. Vemos el efecto de las varianzas V y W.
- Simulamos en R observaciones del modelo de crecimiento lineal
- Diseñamos un DLM con crecimiento cuadrático y simulamos del mismo

Ejercicio

- Describimos cómo tratar con DLMs en el paquete dlm
- DLMs constantes almacenados como objetos dlm con parámetros FF, V, GG, W, $C0$ $m0$
- Especificado con la función dlm

```
rw <- dlm(m0 = 0, C0 = 10, FF = 1, V = 1.4, GG = 1, W = 0.2)
```

Problemas de interés frente a modelos de espacio de estados (DLMs)

- Filtrado
- Predicción
- Predicción a k pasos
- Suavizado conjunto
- Suavizado marginal
- Suavizado con retardo fijo
- Suavizado de intervalo fijo

Estimación y Predicción

Objetivo $\pi(\theta_s|y_{1:t})$

Filtrado $s = t$

Predicción $s > t$

Suavizado $s < t$

$$
\begin{array}{ccccccccccccccc}
\theta_0 & \longrightarrow & \theta_1 & \longrightarrow & \theta_2 & \longrightarrow & \cdots & \longrightarrow & \theta_{t-1} & \longrightarrow & \theta_t & \longrightarrow & \theta_{t+1} & \longrightarrow & \cdots \\
 & & \downarrow & & \downarrow & & & & \downarrow & & \downarrow & & \downarrow & & \\
 & & Y_1 & & Y_2 & & & & Y_{t-1} & & Y_t & & Y_{t+1} & &
\end{array}
$$

Estimación y Predicción

- Densidad predictiva del estado a un paso

$$\pi(\theta_t|y_{1:t-1}) = \int \pi(\theta_t|\theta_{t-1})\pi(\theta_{t-1}|y_{1:t-1})\,\mathrm{d}\theta_{t-1}$$

- Densidad predictiva a un paso

$$\pi(y_t|y_{1:t-1}) = \int \pi(y_t|\theta_t)\pi(\theta_t|y_{1:t-1})\,\mathrm{d}\theta_t$$

- Densidad de filtrado

$$\pi(\theta_t|y_{1:t}) = \frac{\pi(y_t|\theta_t)\pi(\theta_t|y_{1:t-1})}{\pi(y_t|y_{1:t-1})}$$

Estimación y Predicción en el DLM (Filtro de Kalman)

- En un momento dado, $\theta_{t-1}|y_{1:t-1} \sim \mathcal{N}(m_{t-1}, C_{t-1})$
- Densidad predictiva del estado a un paso, normal

$$a_t = \mathrm{E}(\theta_t|y_{1:t-1}) = G_t m_{t-1},$$
$$R_t = \mathrm{Var}(\theta_t|y_{1:t-1}) = G_t C_{t-1} G_t' + W_t$$

- Densidad predictiva a un paso, normal

$$f_t = \mathrm{E}(Y_t|y_{1:t-1}) = F_t a_t,$$
$$Q_t = \mathrm{Var}(Y_t|y_{1:t-1}) = F_t R_t F_t' + V_t$$

- Densidad de filtrado

$$m_t = \mathrm{E}(\theta_t|y_{1:t}) = a_t + R_t F_t' Q_t^{-1} e_t, \qquad e_t = Y_t - f_t$$
$$C_t = \mathrm{Var}(\theta_t|y_{1:t}) = R_t - R_t F_t' Q_t^{-1} F_t R_t$$

Ejercicio

- Desarrolla las fórmulas de predicción para el modelo de nivel local

Filtrado con dlm

- Usamos la función dlmFilter, cuyos argumentos son los datos y un objeto dlm
- La salida es una lista con atributo dlmFiltered que incluye los datos, las medias, y las varianzas

Ejercicio

- Hacemos filtrado con DLM de los datos del Nilo

Predicción a k pasos

- Queremos mirar algo más hacia delante en el futuro

$$\begin{array}{ccccccc} \theta_t & \longrightarrow & \theta_{t+1} & \longrightarrow & \cdots & \longrightarrow & \theta_{t+k} \\ | & & & & & & | \\ Y_{1:t} & & & & & & Y_{t+k} \end{array}$$

- La predicción a k pasos del estado es

$$\pi(\theta_{t+k}|y_{1:t}) = \int \pi(\theta_{t+k}|\theta_{t+k-1})\pi(\theta_{t+k-1}|y_{1:t})\,\mathrm{d}\theta_{t+k-1}$$

- La predicción a k pasos de la observación es

$$\pi(y_{t+k}|y_{1:t}) = \int \pi(y_{t+k}|\theta_{t+k})\pi(\theta_{t+k}|y_{1:t})\,\mathrm{d}\theta_{t+k}$$

Predicción a k pasos en el DLM

- La predicción a k pasos del estado es normal con media y varianza

$$\begin{aligned} a_t(k) &= G_{t+k} a_{t,k-1}, \\ R_t(k) &= G_{t+k} R_{t,k-1} G'_{t+k} + W_{t+k} \end{aligned}$$

- La predicción a k pasos de la observación es normal con media y varianza

$$\begin{aligned} f_t(k) &= F_{t+k} a_t(k), \\ Q_t(k) &= F_{t+k} R_t(k) F'_{t+k} + V_t \end{aligned}$$

Ejercicio

- Hacemos predicción con DLM de los datos de Nilo

Ejercicio

- Aplicamos suavizado a la serie de datos del Nilo

Validación de modelos

- Los errores de predicción se pueden escribir

$$e_t = Y_t - \mathrm{E}(Y_t | y_{1:t-1}) = Y_t - f_t$$

$$\begin{aligned} e_t &= Y_t - F_t a_t = F_t \theta_t + v_t - F_t a_t \\ &= F_t(\theta_t - a_t) + v_t. \end{aligned}$$

- Tienen propiedades que facilitan la validación del modelo. Entre otras,
  – Valor esperado es cero
  – Observaciones de un proceso gaussiano
  – $\tilde{e}_t = e_t / \sqrt{Q_t}$ proceso gaussiano de ruido blanco

Ejercicio

- Validamos el DLM con los datos del Nilo

Especificación de Modelos
El principio de superposición

- La suma de DLMs independientes es un DLM

$$Y_t = Y_{1,t} + \cdots + Y_{h,t}$$

$$\begin{aligned} Y_{i,t} &= F_{i,t}\theta_{i,t} + v_{i,t}, \qquad & v_{i,t} &\sim \mathcal{N}(0, V_{i,t}), \\ \theta_{i,t} &= G_{i,t}\theta_{i,t-1} + w_{i,t}, \qquad & w_{i,t} &\sim \mathcal{N}(0, W_{i,t}) \end{aligned}$$

resulta

$$\begin{aligned} Y_t &= F_t\theta_t + v_t, \qquad & v_t &\sim \mathcal{N}(0, V_t), \\ \theta_t &= G_t\theta_{t-1} + w_t, \qquad & w_t &\sim \mathcal{N}(0, W_t), \end{aligned}$$

$$\theta_t = \begin{bmatrix} \theta_{1,t} \\ \vdots \\ \theta_{h,t} \end{bmatrix} \qquad F_t = [F_{1,t} | \cdots | F_{h,t}] \qquad G_t = \begin{bmatrix} G_{1,t} & & \\ & \ddots & \\ & & G_{h,t} \end{bmatrix} \quad W_t = \begin{bmatrix} W_{1,t} & & \\ & \ddots & \\ & & W_{h,t} \end{bmatrix} \qquad V_t = \sum_{i=1}^{j} V_{i,t}$$

Estrategia de construcción de modelos basada en bloques

Bloques de tendencia (o polinómicos)

- Descripción general para orden n

$$\begin{cases} Y_t = \theta_{t,1} + v_t \\ \theta_{t,j} = \theta_{t-1,j} + \theta_{t-1,j+1} + w_{t,j} \quad j = 1, \ldots, n-1 \\ \theta_{t,n} = \theta_{t-1,n} + w_{t,n}. \end{cases}$$

$$F = (1, 0, \ldots, 0)$$

$$G = \begin{bmatrix} 1 & 1 & 0 & & \ldots & 0 \\ 0 & 1 & 1 & 0 & \ldots & 0 \\ \vdots & & & & \ddots & \vdots \\ 0 & & \ldots & 0 & 1 & 1 \\ 0 & & \ldots & & 0 & 1 \end{bmatrix}$$

$$W = \operatorname{diag}(W_1, \ldots, W_n).$$

- La función de predicción es un polinomio de orden n-1

$$f_t(k) = E(Y_{t+k} | y_{1:t}) = a_{t,0} + a_{t,1} k + \cdots + a_{t,n-1} k^{n-1}$$

- Considera los casos particulares en que n=1,2,3

Ejercicio

- Construimos un modelo para los datos de inversiones en España de 1960 a 2000

Bloques estacionales

En fenómenos estacionales tenemos, al menos, dos estrategias de aproximación

- Modelos de factores estacionales
- Modelos estacionales en forma de Fourier (un ejemplo luego)

Bloques estacionales

Consideramos una serie puramente estacional. Podemos verla como desviaciones estacionales alrededor de su media

Estrategia inicial: Un parámetro por estación, en cada estación actúa el parámetro correspondiente, los parámetros van rotando, se añade una restricción por razones de identificabilidad (habitualmente suma 0)

Bloques estacionales

Una representación más parsimoniosa para un fenómeno de ciclo s es:

- Vector de estados de dimensión s-1

$$F = (1, 0, \ldots, 0)$$

$$G = \begin{bmatrix} -1 & -1 & \ldots & -1 & -1 \\ 1 & 0 & & 0 & 0 \\ 0 & 1 & & 0 & 0 \\ & & \ddots & & \\ 0 & 0 & & 1 & 0 \end{bmatrix}.$$

Ejercicio

- Especificamos un modelo estacional para datos trimestrales

- Un ejemplo luego con series de Fourier!!!

Bloques de regresión (dinámica)

- Pasamos del modelo de regresión estándar

$$Y_t = \beta_1 + \beta_2 x_t + \epsilon_t, \quad \epsilon_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$$

- Al modelo de regresión dinámica

$$Y_t = \beta_{1,t} + \beta_{2,t} x_t + \epsilon_t, \quad \epsilon_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$$

$$\begin{aligned} Y_t &= x_t' \theta_t + v_t, & v_t &\sim \mathcal{N}(0, \sigma_t^2) \\ \theta_t &= G_t \theta_{t-1} + w_t, & w_t &\sim \mathcal{N}_p(0, W_t) \end{aligned}$$

Bloques ARIMA

- Para cualquier proceso ARIMA es posible encontrar un DLM cuyo proceso Y observable tiene la misma distribución

ARMA(p,q)

$$Y_t = \mu + \sum_{j=1}^{p} \phi_j (Y_{t-j} - \mu) + \sum_{j=1}^{q} \psi_j \epsilon_{t-j} + \epsilon_t$$

Bloques ARIMA

- En la estrategia por bloques que describimos, a menudo nos basta con un modelo AR de orden bajo, por ejemplo AR(2)

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \epsilon_t, \qquad \epsilon_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$$

- Para ello hacemos

$$F = \begin{bmatrix} 1 & 0 \end{bmatrix}, \qquad V = 0,$$

$$G = \begin{bmatrix} \phi_1 & 1 \\ \phi_2 & 0 \end{bmatrix}, \quad W = \begin{bmatrix} \sigma^2 & 0 \\ 0 & 0 \end{bmatrix}$$

Estrategia de modelización

- Tendencia + Término estacional + Término regresión + Término AR de orden bajo (1)

Modelos con parámetros desconocidos

- Hasta el momento suponemos las matrices F, G, V, W completamente conocidas. No siempre es así. Introducimos parámetros para aprender

$$(\theta_0, \theta_1, \ldots, \theta_n, Y_1, \ldots, Y_n, \psi) \sim \pi(\theta_0|\psi)\pi(\psi) \prod_{t=1}^{n} \pi(y_t|\theta_t, \psi)\pi(\theta_t|\theta_{t-1}, \psi)$$

- A posteriori conjunta y marginal

$$\pi(\theta_s, \psi|y_{1:t}) = \pi(\theta_s|\psi, y_{1:t})\pi(\psi|y_{1:t}) \qquad \pi(\theta_t|y_{1:t}) = \int \pi(\theta_t|\psi, y_{1:t})\pi(\psi|y_{1:t})d\psi$$

- Suavizado

$$\pi(\theta_{0:t}, \psi|y_{1:t}) = \pi(\theta_{0:t}|\psi, y_{1:t})\pi(\psi|y_{1:t})$$

- En principio, podemos usar MCMC (Gibbs) pero acaba resultando ser más eficiente el uso de Montecarlo secuencial, que describimos en la parte no lineal

Modelos con parámetros desconocidos: el caso conjugado

- F, G conocidas. V, W desconocidas
- V, W y C_0 conocidas salvo un factor de escala

$$V_t = \sigma^2 \tilde{V}_t, \qquad W_t = \sigma^2 \tilde{W}_t, \qquad C_0 = \sigma^2 \tilde{C}_0$$

$$Y_t|\theta_t, \phi \sim \mathcal{N}_m(F_t\theta_t, \phi^{-1}\tilde{V}_t)$$

$$\theta_t|\theta_{t-1}, \phi \sim \mathcal{N}_p(G_t\theta_{t-1}, \phi^{-1}\tilde{W}_t)$$

$$\phi \sim \mathcal{G}(\alpha_0, \beta_0), \quad \theta_0|\phi \sim \mathcal{N}(m_0, \phi^{-1}\tilde{C}_0)$$

$$(\theta_0, \phi) \sim \mathcal{NG}(m_0, \tilde{C}_0, \alpha_0, \beta_0)$$

$$(\theta_{t-1}, \phi)|y_{1:t-1} \sim \mathcal{NG}(m_{t-1}, \tilde{C}_{t-1}, \alpha_{t-1}, \beta_{t-1})$$

- La predicción del estado a un paso es normal gamma
- La predicción de la observación a un paso es una t de Student

Modelos con parámetros desconocidos: el caso conjugado

- F, G conocidas. V, W desconocidas
- Especificación de W mediante descuento

$$W_t = \frac{1-\delta}{\delta} P_t \qquad \mathrm{Var}(G_t \theta_{t-1} | y_{1:t-1})$$

- Puede usarse también para V

- Estimar V, W por EMV

Ejemplo: Control de sueltas en un embalse

Problema: Decidir sueltas óptimas en un embalse

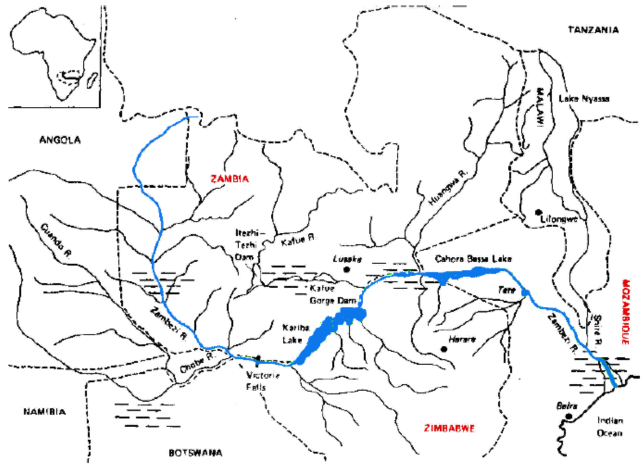
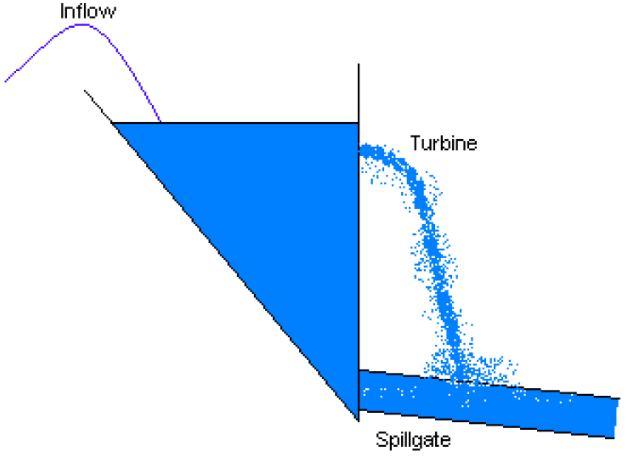


Figure 1: ZAMBIA, ZIMBABWE & MOZAMBIQUE



Ejemplo: Control de sueltas en un embalse

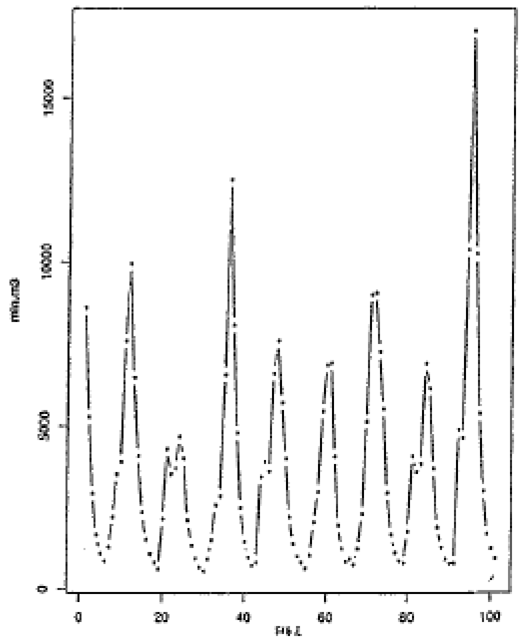
Tomar decisiones en relación con sueltas de agua. Maximizar

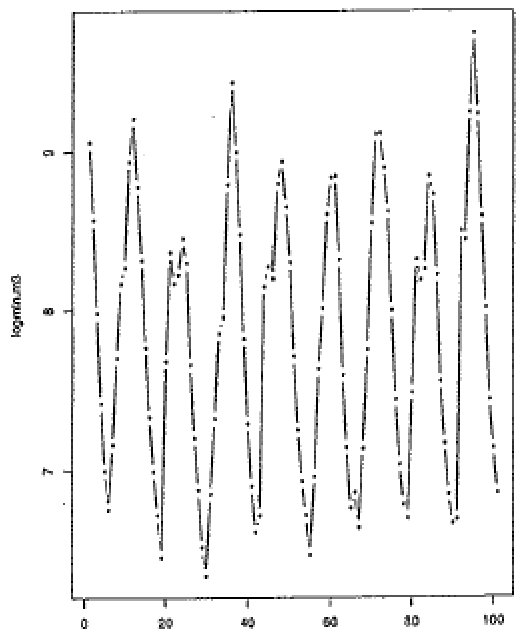
$$\int f(u_{2,T+1}, k_{T+1}, \ldots, u_{2,T+r}, k_{T+r}, s_{T+r+1}) dH(i_{T+1}, \ldots, i_{T+r} | D_T)$$

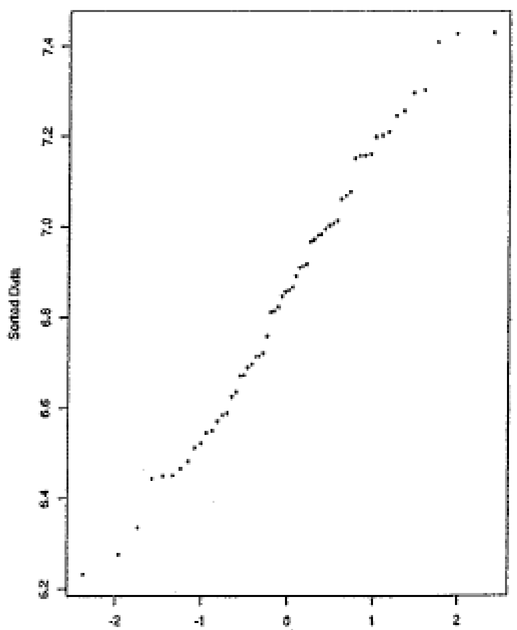
$$\int f(u_{2,T+1}, k_{T+1}, \delta(s_{T+1}, x^*_{T+1})) dH(i_{T+1} | D_T)$$

Ejemplo: Predicción de afluencias a embalse

Algunas vistas del análisis exploratorio empleado







Ejemplo: Predicción de afluencias a embalse

Modelo inicial ideado

$$y_t = (1,1,0,1,0,1,0,1,0,1,0,1)'\mathbf{z}_t + v_t, \quad v_t \sim N(0,v)$$

$$\mathbf{z_t} = \text{block} \;\; \text{diag}(1, G_1, ..., G_5, -1)\mathbf{z}_{t-1} + \mathbf{w_t}, \;\; \mathbf{w_t} \sim N(0, vW_t^*) \quad G_i = \begin{pmatrix} cos(\pi i/6) & sin(\pi i/6) \\ -sin(\pi i/6) & cos(\pi i/6) \end{pmatrix}$$

$$\mathbf{z}_0 | v^{-1} \sim N(\mathbf{m}_0, vC^*)$$

$$v^{-1} \sim Gamma(n_0/2, d_0/2)$$

Información a priori

	OC	NV	DE	JA	FE	MR	AP	MY	JU	JL	AU	SE
EI	6.55	6.9	7.6	8.29	8.7	9.1	8.7	8.3	8	7.6	7.3	6.9
MAD	.82	.67	1.33	.97	1.45	1.27	1.27	1.3	1.2	1	.7	.66

LEV	SOC	SNV	SDE	SJA	SFE	SMR	SAP	SMY	SJU	SJL	SAU	SSE
7.8	-1.28	-.92	-.23	.47	.87	1.27	.87	.47	.17	-.23	-.53	-.93

Ejemplo: Predicción de afluencias a embalse

Más información a priori $[7.3, 8.3]$

LEV	SOC	SNV	SDE	SJA	SFE	SMR	SAP	SMY	SJU	SJL	SAU	SSE
.06	.02	.00	.16	.06	.25	.16	.16	.16	.12	.06	.01	.00

	1	2	3	4	5	6
a_i	-1.027	-0.141	-.058	-.042	.010	-.022
b_i	.329	-0.172	.018	-.056	.040	

Evaluación de la importancia de las componentes estacionales

Harmonic	1	2	3	4	5	6
F statistic	161.2	1.65	.044	.09	.07	646.9

Ejemplo: Predicción de afluencias a embalse

- *Observation equation*

$$y_t = (1,1,0,1)'\mathbf{z}_t + v_t, \quad v_t \sim N(0, v)$$

- *System equation*

$$\mathbf{z}_t = G\mathbf{z}_{t-1} + \mathbf{w}_t, \ \ \mathbf{w}_t \sim N(\mathbf{0}, \begin{pmatrix} vW_t^* & 0 \\ 0 & \sigma^2 \end{pmatrix})$$

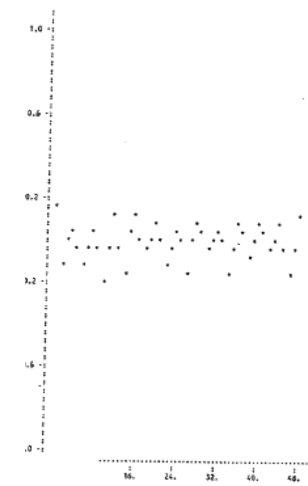
$$G = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\pi/6) & \sin(\pi/6) & 0 \\ 0 & -\sin(\pi/6) & \cos(\pi/6) & 0 \\ 0 & 0 & 0 & r^* \end{pmatrix}$$

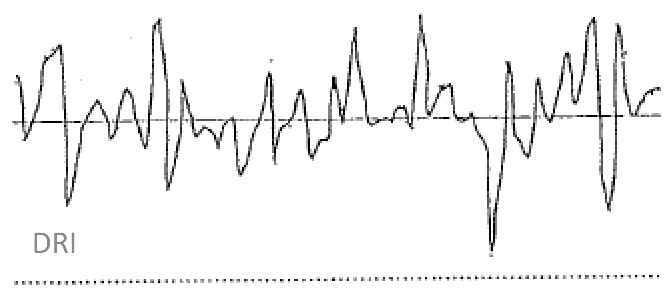
$$\mathbf{z}_0 | v^{-1} \sim N(\mathbf{m}_0, vC^*)$$

$$v^{-1} \sim Gamma(n_0/2, d_0/2)$$

(7.8,-1.02,.33,0)

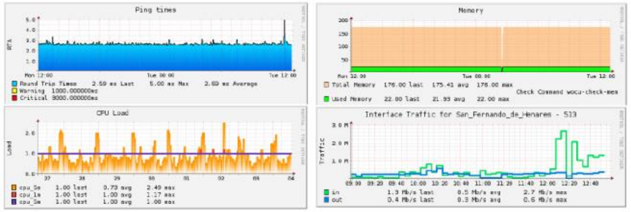
$$C^* = \begin{pmatrix} .02 & 0 & 0 & 0 \\ 0 & .002 & .0007 & 0 \\ 0 & .0007 & .003 & 0 \\ 0 & 0 & 0 & \sigma^2 \end{pmatrix}$$



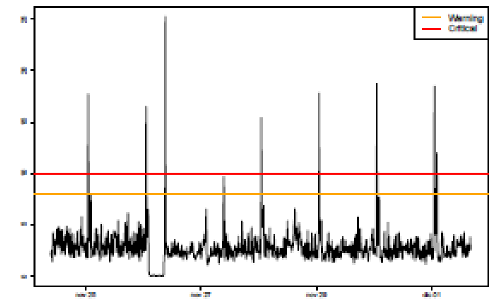


Monitorización de redes para seguridad

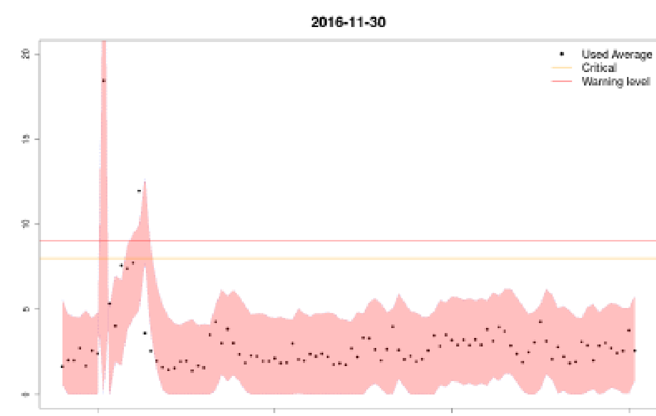
- Sistema recoge medidas de cientos de miles de dispositivos conectados a Internet!!!!
- Por cada dispositivo varias medidas cada muy poco tiempo (1 min, 5 min, 10 min)
- Big Data!!!
- Sistema de monitorización descriptivo: si se alcanzan valores críticos, alarma
- Capacidad predictiva???



Monitorización de redes para seguridad



- Tres requisitos
  - Automático
  - Flexible y Versátil
  - Escalable (en tiempo y memoria). Tiempo real
- Diseñar una clase genérica de modelos de predicción Tendencia+estacional(s)+estallidos
- Procedimiento automatizado de identificación
- Alarmas (corto y largo plazo) si
  - Predicen niveles críticos
  - Cambios repentinos detectados



Modelos para series multivariantes

- Estudiar cada serie independientemente
- Suponer que todas las series siguen la misma dinámica

$$
\begin{aligned}
Y_{i,t} &= F\theta_t^{(i)} + v_{i,t}, \qquad v_{i,t} \sim \mathcal{N}(0, V_i) \\
\theta_t^{(i)} &= G\theta_{t-1}^{(i)} + w_t^{(i)}, \quad w_{i,t} \sim \mathcal{N}_p(0, W_i)
\end{aligned}
$$

- DLMs jerárquicos

$$
\begin{aligned}
Y_t &= F_{y,t}\theta_t + v_t, \qquad v_t \sim \mathcal{N}_m(0, V_{y,t}), \\
\theta_t &= F_{\theta,t}\lambda_t + \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}_P(0, V_{\theta,t}), \\
\lambda_t &= G_t\lambda_{t-1} + w_t, \quad w_t \sim \mathcal{N}_k(0, W_t),
\end{aligned}
$$

Modelos para series multivariantes

- Modelos de factores comunes

$$Y_t = A\mu_t + v_t, \qquad v_t \sim \mathcal{N}(0, V),$$
$$\mu_t = \mu_{t-1} + w_t, \quad w_t \sim \mathcal{N}(0, W),$$

Agenda

- Recordatorio modelo normal-normal
- Series temporales: Aproximaciones
- Ejemplo introductorio de modelo dinámico lineal
- Modelo dinámico lineal (y librería dlm)
- **Modelos dinámicos no lineales**
- Referencias

Modelos de espacio de estados Definición formal

- Una serie temporal (θ_t) p-variante y otra serie temporal Y_t m-variante tales que
  - (θ_t) es una cadena de Markov
  - Dado (θ_t), Y_t son independientes e Y_t sólo depende de θ_t

$$
\begin{array}{ccccccccccccccc}
\theta_0 & \longrightarrow & \theta_1 & \longrightarrow & \theta_2 & \longrightarrow & \cdots & \longrightarrow & \theta_{t-1} & \longrightarrow & \theta_t & \longrightarrow & \theta_{t+1} & \longrightarrow & \cdots \\
 & & \downarrow & & \downarrow & & & & \downarrow & & \downarrow & & \downarrow & & \\
 & & Y_1 & & Y_2 & & & & Y_{t-1} & & Y_t & & Y_{t+1} & &
\end{array}
$$

$$\pi(\theta_{0:t}, y_{1:t}) = \pi(\theta_0) \cdot \prod_{j=1}^{t} \pi(\theta_j|\theta_{j-1})\pi(y_j|\theta_j)$$

Modelo de espacio de estados

- Es un modelo con
  - Ecuación de observación

$$Y_t = h_t(\theta_t, v_t)$$

  - Ecuación de estado (o de sistema)

$$\theta_t = g_t(\theta_{t-1}, w_t)$$

  - Distribución a priori sobre el estado inicial

Algunos modelos no lineales, no gaussianos

- Modelos de espacio de estados de la familia exponencial.
- Modelos de Markov ocultos
- Modelos de redes neuronales
- Modelos de volatilidad estocástica

$$
\begin{aligned}
Y_t &= \exp\left\{\frac{1}{2}\theta_t\right\} w_t, & w_t &\sim \mathcal{N}(0,1), \\
\theta_t &= \eta + \phi\theta_{t-1} + v_t, & v_t &\sim \mathcal{N}(0,\sigma^2)
\end{aligned}
$$

Estimación y Predicción

Objetivo $\pi(\theta_s | y_{1:t})$

Filtrado $s = t$

Predicción de estados $s > t$

Suavizado $s < t$

$$
\begin{array}{ccccccccccccccc}
\theta_0 & \longrightarrow & \theta_1 & \longrightarrow & \theta_2 & \longrightarrow & \cdots & \longrightarrow & \theta_{t-1} & \longrightarrow & \theta_t & \longrightarrow & \theta_{t+1} & \longrightarrow & \cdots \\
 & & \downarrow & & \downarrow & & & & \downarrow & & \downarrow & & \downarrow & & \\
 & & Y_1 & & Y_2 & & & & Y_{t-1} & & Y_t & & Y_{t+1} & &
\end{array}
$$

Estimación y Predicción

- Densidad predictiva del estado a un paso

$$\pi(\theta_t|y_{1:t-1}) = \int \pi(\theta_t|\theta_{t-1})\pi(\theta_{t-1}|y_{1:t-1})\,\mathrm{d}\theta_{t-1}$$

- Densidad predictiva a un paso

$$\pi(y_t|y_{1:t-1}) = \int \pi(y_t|\theta_t)\pi(\theta_t|y_{1:t-1})\,\mathrm{d}\theta_t$$

- Densidad de filtrado

$$\pi(\theta_t|y_{1:t}) = \frac{\pi(y_t|\theta_t)\pi(\theta_t|y_{1:t-1})}{\pi(y_t|y_{1:t-1})}$$

Filtrado y predicción

- Off-line. Pueden usarse métodos MCMC en lotes. FFBS (forward filtering backward sampling dentro de Gibbs o muestreador híbrido).
- On-line. No vale. MCMC no es secuencial.
  - Linealización.
  - SMC.

Muestreo por importancia

- Deseamos calcular $\mathrm{E}_\pi(f(X)) = \int f(x)\pi(x)\,\mathrm{d}x$
- Si g es una densidad por importancia

$$\mathrm{E}_\pi(f(X)) = \int f(x)\frac{\pi(x)}{g(x)}g(x)\,\mathrm{d}x = \mathrm{E}_g(f(X)w^\star(X)) \qquad w^\star(x) = \pi(x)/g(x)$$

- Aproximación MC por importancia. Muestra de g

$$\frac{1}{N}\sum_{i=1}^{N} f(x^{(i)})w^\star(x^{(i)}) \approx \mathrm{E}_\pi(f(X))$$

- Hacemos, finalmente

$$\mathrm{E}_\pi(f(X)) \approx \frac{1}{N}\sum_{i=1}^{N} f(x^{(i)})w^\star(x^{(i)})$$
$$= \frac{\frac{1}{N}\sum_{i=1}^{N} f(x^{(i)})\tilde{w}^{(i)}}{C} \approx \frac{\sum_{i=1}^{N} f(x^{(i)})\tilde{w}^{(i)}}{\sum_{i=1}^{N}\tilde{w}^{(i)}}$$
$$= \sum_{i=1}^{N} f(x^{(i)})w^{(i)},$$

- Consideramos

con lo que $\hat{\pi} = \sum_{i=1}^{N} w^{(i)}\delta_{x^{(i)}}$

$$\pi \approx \hat{\pi}$$

Filtrado de partículas básico

- Cómo se actualiza $\hat{\pi} = \sum_{i=1}^{N} w^{(i)} \delta_{x^{(i)}}$ cuando pasamos de $\pi(\theta_{0:t-1}|y_{1:t-1})$ a $\pi(\theta_{0:t}|y_{1:t})$
- Suponemos que la distribución por importancia tiene la forma

$$g_t(\theta_{0:t}|y_{1:t}) = g_{t|t-1}(\theta_t|\theta_{0:t-1}, y_{1:t}) \cdot g_{t-1}(\theta_{0:t-1}|y_{1:t-1})$$

- Para los pesos se tiene

$$\begin{aligned} w_t &\propto \frac{\pi(\theta_{0:t}|y_{1:t})}{g_t(\theta_{0:t}|y_{1:t})} \propto \frac{\pi(\theta_{0:t}, y_t|y_{1:t-1})}{g_t(\theta_{0:t}|y_{1:t})} \\ &\propto \frac{\pi(\theta_t, y_t|\theta_{0:t-1}, y_{1:t-1}) \cdot \pi(\theta_{0:t-1}|y_{1:t-1})}{g_{t|t-1}(\theta_t|\theta_{0:t-1}, y_{1:t}) \cdot g_{t-1}(\theta_{0:t-1}|y_{1:t-1})} \\ &\propto \frac{\pi(y_t|\theta_t) \cdot \pi(\theta_t|\theta_{t-1})}{g_{t|t-1}(\theta_t|\theta_{0:t-1}, y_{1:t})} \cdot w_{t-1}. \end{aligned}$$

y se normalizan

- Tamaño muestral efectivo

entre N y 1

$$N_{eff} = \left(\sum_{i=1}^{N} (w_t^{(i)})^2 \right)^{-1}$$

0. Initialize: draw $\theta_0^{(1)}, \dots, \theta_0^{(N)}$ independently from $\pi(\theta_0)$ and set

$$w_0^{(i)} = N^{-1}, \qquad i = 1, \dots, N.$$

1. For $t = 1, \dots, T$:
 1.1) For $i = 1, \dots, N$:
 - Draw $\theta_t^{(i)}$ from $g_{t|t-1}(\theta_t|\theta_{0:t-1}^{(i)}, y_{1:t})$ and set

$$\theta_{0:t}^{(i)} = (\theta_{0:t-1}^{(i)}, \theta_t^{(i)})$$

.

 - Set

$$\tilde{w}_t^{(i)} = w_{t-1}^{(i)} \cdot \frac{\pi(\theta_t^{(i)}, y_t|\theta_{t-1}^{(i)})}{g_{t|t-1}(\theta_t^{(i)}|\theta_{0:t-1}^{(i)}, y_{1:t})}.$$

 1.2) Normalize the weights:

$$w_t^{(i)} = \frac{\tilde{w}_t^{(i)}}{\sum_{j=1}^N \tilde{w}_t^{(j)}}.$$

 1.3) Compute

$$N_{\mathit{eff}} = \left(\sum_{i=1}^N (w_t^{(i)})^2 \right)^{-1}.$$

 1.4) If $N_{\mathit{eff}} < N_0$, resample:
 - Draw a sample of size N from the discrete distribution

$$\mathrm{P}(\theta_{0:t} = \theta_{0:t}^{(i)}) = w_t^{(i)}, \qquad i = 1, \dots, N,$$

 and relabel this sample

$$\theta_{0:t}^{(1)}, \dots, \theta_{0:t}^{(N)}.$$

 - Reset the weights: $w_t^{(i)} = N^{-1}$, $i = 1, \dots, N$.

 1.5) Set $\hat{\pi}_t = \sum_{i=1}^N w_t^{(i)} \delta_{\theta_{0:t}^{(i)}}$.

Filtrado de partículas básico

- En un momento dado tenemos $\hat{\pi}_t = \sum_{i=1}^{N} w^{(i)} \delta_{\theta_{0:t}^{(i)}}$ y usamos la aproximación

$$\pi(\theta_t|y_{1:t}) \approx \sum_{i=1}^{N} w^{(i)} \delta_{\theta_t^{(i)}}$$

- Dos propuestas para la densidad de transición por importancia

$$g_{t|t-1}(\theta_t|\theta_{0:t-1}, y_{1:t}) = \pi(\theta_t|\theta_{t-1}).$$

$$\theta_t|\theta_{t-1}, y_t$$

Filtrado de partículas auxiliar

- El filtro anterior depende de la especificación de las densidades de transición por importancia: difícil (salvo en DLMs)
- Suponemos en (t-1) la aproximación

$$\hat{\pi}_{t-1} = \sum_{i=1}^{N} w_{t-1}^{(i)} \delta_{\theta_{0:t-1}^{(i)}}$$

a la distribución de suavizado

$$\pi\left(\theta_{0:t-1} \middle| y_{1:t-1}\right)$$

Filtrado de partículas auxiliar

$$\begin{aligned}
\pi(\theta_{0:t}|y_{1:t}) &\propto \pi(\theta_{0:t}, y_t|y_{1:t-1}) \\
&= \pi(y_t|\theta_{0:t}, y_{1:t-1}) \cdot \pi(\theta_t|\theta_{0:t-1}, y_{1:t-1}) \cdot \pi(\theta_{0:t-1}|y_{1:t-1}) \\
&= \pi(y_t|\theta_t) \cdot \pi(\theta_t|\theta_{t-1}) \cdot \pi(\theta_{0:t-1}|y_{1:t-1}) \\
&\approx \pi(y_t|\theta_t) \cdot \pi(\theta_t|\theta_{t-1}) \cdot \hat{\pi}_{t-1}(\theta_{0:t-1}) \\
&= \sum_{i=1}^{N} w_{t-1}^{(i)} \pi(y_t|\theta_t) \pi(\theta_t|\theta_{t-1}^{(i)}) \delta_{\theta_{0:t-1}^{(i)}}.
\end{aligned}$$

- Para quitarnos la suma usamos variables latentes

$$\mathrm{P}(I = i) = w_{t-1}^{(i)},$$

$$\theta_{0:t}|I = i \sim C\pi(y_t|\theta_t)\pi(\theta_t|\theta_{t-1}^{(i)})\delta_{\theta_{0:t-1}^{(i)}}$$

- Con lo que

$$\pi^{\text{aux}}(\theta_{0:t}, i|y_{1:t}) \propto w_{t-1}^{(i)} \pi(y_t|\theta_t)\pi(\theta_t|\theta_{t-1}^{(i)})\delta_{\theta_{0:t-1}^{(i)}}$$

Filtrado de partículas auxiliar

$$\pi^{\text{aux}}(\theta_{0:t}, i|y_{1:t}) \propto w_{t-1}^{(i)}\pi(y_t|\theta_t)\pi(\theta_t|\theta_{t-1}^{(i)})\delta_{\theta_{0:t-1}^{(i)}}$$

- Se sugiere

$$g_t(\theta_{0:t}, i|y_{1:t}) \propto w_{t-1}^{(i)}\pi(y_t|\hat{\theta}_t^{(i)})\pi(\theta_t|\theta_{t-1}^{(i)})\delta_{\theta_{0:t-1}^{(i)}}$$

- De la que se muestrea en dos pasos

1. Muestrea la variable de clasificación con

$$\mathrm{P}(I_k = i) \propto w_{t-1}^{(i)}\pi(y_t|\hat{\theta}_t^{(i)}), \qquad i = 1, \ldots, N$$

2. Dada la etiqueta, $\theta_t^{(k)} \sim \pi(\theta_t|\theta_{t-1}^{(i)})$ $\qquad \theta_{0:t}^{(k)} = (\theta_{0:t-1}^{(i)}, \theta_t^{(k)}).$

El peso por importancia es proporcional a

$$\tilde{w}_t^{(k)} = \frac{w_{t-1}^{(I_k)}\pi(y_t|\theta_t^{(k)})\pi(\theta_t^{(k)}|\theta_{t-1}^{(k)})}{w_{t-1}^{(I_k)}\pi(y_t|\hat{\theta}_t^{(k)})\pi(\theta_t^{(k)}|\theta_{t-1}^{(k)})} = \frac{\pi(y_t|\theta_t^{(k)})}{\pi(y_t|\hat{\theta}_t^{(k)})}$$

0. Initialize: draw $\theta_0^{(1)}, \ldots, \theta_0^{(N)}$ independently from $\pi(\theta_0)$ and set

$$w_0^{(i)} = N^{-1}, \qquad i = 1, \ldots, N.$$

1. For $t = 1, \ldots, T$:
   1.1) For $k = 1, \ldots, N$:
   - Draw I_k, with $\mathrm{P}(I_k = i) \propto w_{t-1}^{(i)} \pi(y_t | \hat{\theta}_t^{(i)})$.
   - Draw $\theta_t^{(k)}$ from $\pi(\theta_t | \theta_{t-1} = \theta_{t-1}^{(I_k)})$ and set

$$\theta_{0:t}^{(k)} = (\theta_{0:t-1}^{(I_k)}, \theta_t^{(k)}).$$

   - Set

$$\tilde{w}_t^{(k)} = \frac{\pi(y_t | \theta_t^{(k)})}{\pi(y_t | \hat{\theta}_t^{(k)})}.$$

   1.2) Normalize the weights:

$$w_t^{(i)} = \frac{\tilde{w}_t^{(i)}}{\sum_{j=1}^N \tilde{w}_t^{(j)}}.$$

   1.3) Compute

$$N_{\mathit{eff}} = \left(\sum_{i=1}^N (w_t^{(i)})^2 \right)^{-1}.$$

   1.4) If $N_{\mathit{eff}} < N_0$, resample:
   - Draw a sample of size N from the discrete distribution

$$\mathrm{P}(\theta_{0:t} = \theta_{0:t}^{(i)}) = w_t^{(i)}, \qquad i = 1, \ldots, N,$$

   and relabel this sample

$$\theta_{0:t}^{(1)}, \ldots, \theta_{0:t}^{(N)}.$$

   - Reset the weights: $w_t^{(i)} = N^{-1}$, $i = 1, \ldots, N$.

   1.5) Set $\hat{\pi}_t = \sum_{i=1}^N w_t^{(i)} \delta_{\theta_{0:t}^{(i)}}$.

Filtro con parámetros desconocidos

- En muchos casos hay parámetros desconocidos. Usamos una aproximación

$$\hat{\pi}_{t-1}(\theta_{0:t-1}, \psi) = \sum_{i=1}^{N} w_{t-1}^{(i)} \delta_{(\theta_{0:t-1}^{(i)}, \psi^{(i)})} \approx \pi(\theta_{0:t-1}, \psi | y_{0:t-1}).$$

Y marginalmente

$$\hat{\pi}_{t-1}(\psi) = \sum_{i=1}^{N} w_{t-1}^{(i)} \delta_{\psi^{(i)}} \approx \pi(\psi | y_{0:t-1})$$

Se sustituye por

$$\tilde{\pi}_{t-1}(\psi) = \sum_{i=1}^{N} w_{t-1}^{(i)} \mathcal{N}(\psi; m^{(i)}, h^2 \Sigma).$$

$$\tilde{\pi}_{t-1}(\theta_{0:t-1}, \psi) = \sum_{i=1}^{N} w_{t-1}^{(i)} \mathcal{N}(\psi; m^{(i)}, h^2 \Sigma) \delta_{\theta_{0:t-1}^{(i)}}$$

Filtro con parámetros desconocidos

0. Initialize: draw $(\theta_0^{(1)}, \psi^{(1)}), \ldots, (\theta_0^{(N)}, \psi^{(N)})$ independently from $\pi(\theta_0)\pi(\psi)$. Set $w_0^{(i)} = N^{-1}$, $i = 1, \ldots, N$, and

$$\hat{\pi}_0 = \sum_{i=1}^{N} w_0^{(i)} \delta_{(\theta_0^{(i)}, \psi^{(i)})}.$$

1. For $t = 1, \ldots, T$:
 1.1) Compute $\bar{\psi} = \mathrm{E}_{\hat{\pi}_{t-1}}(\psi)$ and $\Sigma = \mathrm{Var}_{\hat{\pi}_{t-1}}(\psi)$. For $i = 1, \ldots, N$, set

$$m^{(i)} = a\psi^{(i)} + (1-a)\bar{\psi},$$
$$\hat{\theta}_t^{(i)} = \mathrm{E}(\theta_t | \theta_{t-1} = \theta_{t-1}^{(i)}, \psi = m^{(i)}).$$

 1.2) For $k = 1, \ldots, N$:
 - Draw I_k, with $\mathrm{P}(I_k = i) \propto w_{t-1}^{(i)} \pi(y_t | \theta_t = \hat{\theta}_t^{(i)}, \psi = m^{(i)})$.
 - Draw $\psi^{(k)}$ from $\mathcal{N}(m^{(I_k)}, h^2\Sigma)$.
 - Draw $\theta_t^{(k)}$ from $\pi(\theta_t | \theta_{t-1} = \theta_{t-1}^{(I_k)}, \psi = \psi^{(k)})$ and set

$$\theta_{0:t}^{(k)} = (\theta_{0:t-1}^{(I_k)}, \theta_t^{(k)}).$$

 - Set

$$\tilde{w}_t^{(k)} = \frac{\pi(y_t | \theta_t = \theta_t^{(k)}, \psi = \psi^{(k)})}{\pi(y_t | \theta_t = \hat{\theta}_t^{(I_k)}, \psi = m^{(I_k)})}.$$

 1.3) Normalize the weights:

$$w_t^{(i)} = \frac{\tilde{w}_t^{(i)}}{\sum_{j=1}^{N} \tilde{w}_t^{(j)}}.$$

 1.4) Compute

$$N_{eff} = \left(\sum_{i=1}^{N} \left(w_t^{(i)}\right)^2 \right)^{-1}.$$

 1.5) If $N_{eff} < N_0$, resample:
 - Draw a sample of size N from the discrete distribution

$$\mathrm{P}\left((\theta_{0:t}, \psi) = (\theta_{0:t}^{(i)}, \psi^{(i)})\right) = w_t^{(i)}, \qquad i = 1, \ldots, N,$$

 and relabel this sample

$$(\theta_{0:t}^{(1)}, \psi^{(1)}), \ldots, (\theta_{0:t}^{(N)}, \psi^{(N)}).$$

 - Reset the weights: $w_t^{(i)} = N^{-1}$, $i = 1, \ldots, N$.

 1.6) Set $\hat{\pi}_t = \sum_{i=1}^{N} w_t^{(i)} \delta_{(\theta_{0:t}^{(i)}, \psi^{(i)})}$.

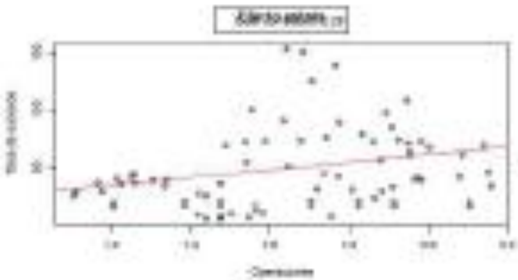
Predicción de incidentes de seguridad aérea

- Proceso de Poisson no homogénea

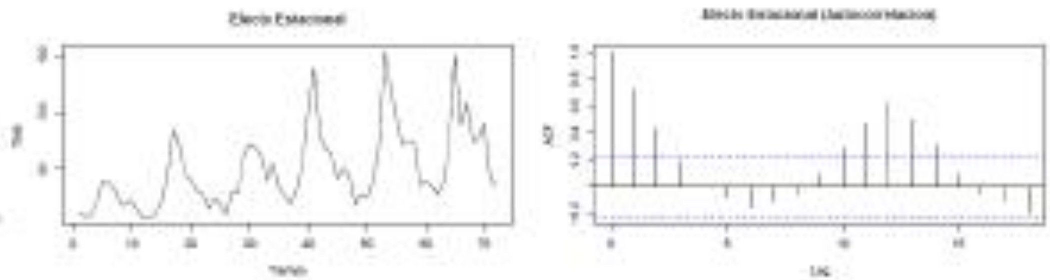
- Número dinámico de operaciones
- Tasa dinámica

- Asignación juicios expertos

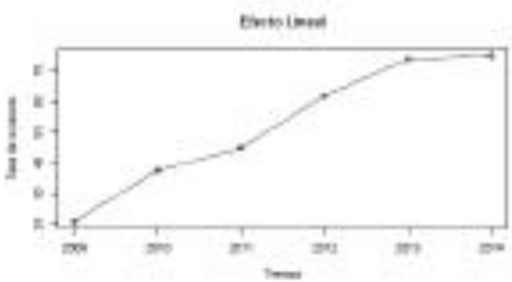
- Predicción incidentes
  - Annual para evaluación y gestión de riesgos
  - Mensual y semanal
    - Monitorización
    - Establecimiento de alarmas



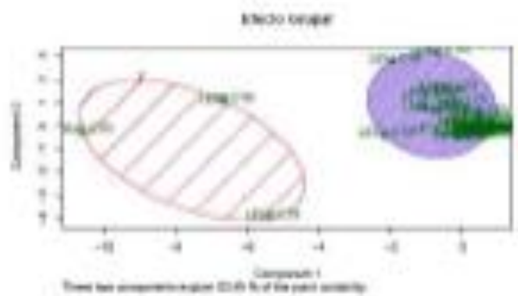
(a)Stress



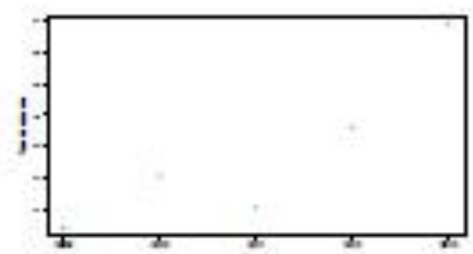
(b)Seasonal



(c)Linear



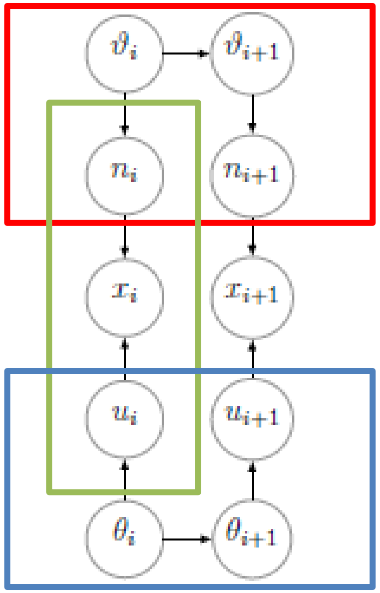
(d)Group



(e)Underreporting

Modelos de predicción

- **ID**



- **Model**

$$
\begin{cases}
\begin{cases}
\begin{cases}
n_i = H_i\vartheta_i + z_i, z_i \sim N(0, \Sigma_i) \\
\vartheta_i = J_i\vartheta_{i-1} + \xi_i, \xi_i \sim N(0, S_i)
\end{cases} \\
\vartheta_0 \sim N(\eta_0, S_0)
\end{cases} \\
x_i|\lambda_i, n_i \sim Po(\lambda_i n_i) \\
\lambda_i = \exp(u_i) \\
\begin{cases}
\begin{cases}
u_i = F_i\theta_i + v_i, v_i \sim N(0, V_i) \\
\theta_i = G_i\theta_{i-1} + w_i, w_i \sim N(0, W_i)
\end{cases} \\
\theta_0 \sim N(\mu_0, W_0),
\end{cases}
\end{cases}
$$

88 modelos

Referencias

- Petris, Petrone, Campagnoli (2009) Dynamic Linear Models with R
- Prado, West (2010) Time Series: Modeling, Computation and Inference
- Rios Insua, Ruggeri, Wiper (2012) Bayesian Analysis of Stochastic Process Models
- West, Harrison (2006) Bayesian Forecasting and Dynamic Models
- https://arxiv.org/abs/1802.06678
- https://arxiv.org/abs/1805.05232
- http://www.stats.ox.ac.uk/~doucet/smc_resources.html
  - https://www.stats.ox.ac.uk/~doucet/doucet_johansen_tutorialPF2011.pdf

Gracias!!

david.rios@icmat.es

SPOR DataLab https://www.icmat.es/spor/