



Series Temporales

Parte II. Modelos en espacio de estado y filtro de Kalman Convocatoria de Diciembre de 2020

1. Datos

Se trata de datos cuyo significado y origen se detallan en el Apéndice, recogiendo diversas magnitudes relacionadas con la pandemia.

2. Ejercicios

Haz *una* de estas cosas a tu elección:

1. (Opción A) Especifica un modelo predictivo para el número de personas en UCI.
2. (Opción B) Especifica un modelo que proporcione una estimación suavizada del número de personas infectadas en la CAPV.
3. (Opción C) Especifica un modelo predictivo para el número de muertes.

3. Algunas sugerencias

1. Por lo que hace a los nuevos ingresos en UCI, cabe pensar que estén relacionados con el total de enfermos COVID ingresados en las fechas inmediatamente anteriores: típicamente, tras aparecer síntomas suficientemente serios para requerir hospitalización, transcurren unos días en que el enfermo mejora o evoluciona a situación que aconseja su ingreso en UCI. Puede haber no obstante personas que directamente sean trasladadas a una UCI.

2. Para el número de muertos, cabe pensar que estén relacionados con pacientes ingresados y pacientes en UCI, pero no únicamente con ellos.
3. El número de personas infectadas (con infección “viva”, no que hayan pasado la enfermedad y posean anticuerpos) es estrictamente imposible de estimar con la información aportada. Es razonable suponer que los números de personas ingresadas, en UCI, o muertas son, con diferentes retardos, proporcionales al número de personas en las que está cursando la enfermedad, quizá con retardo. Por ejemplo

$$UCI_t = k \times \alpha_{t-\ell} \quad (1)$$

o quizá

$$UCI_t = k_0\alpha_t + k_1\alpha_{t-1} + \dots + k_\ell\alpha_{t-\ell} \quad (2)$$

en que α_t es un elemento (directamente inobservable) del vector de estado que recoge los enfermos en el momento t .

Sin embargo, k es desconocido. No podemos estimar α_t porque cualquier transformado $\alpha_t^* = s\alpha_t$ de él junto con un valor $k^* = (1/s)k$ sería observacionalmente equivalente: hemos de fijar k o el valor de α_t en algún momento¹.

4. Siempre hay la tentación a desarrollar un modelo muy elaborado que incorpore todo lo que sabemos (o creemos saber) que ocurre (o puede ocurrir). Por ejemplo, preferiríamos *a priori* como más realista (2) sobre (1). El mayor problema al estimar un modelo en espacio de estado es la dimensión elevada: **comienza con modelos muy simples** y sólo en la medida en que parezca justificado y los datos lo soporten añade complejidad.

Referencias

- [1] J. Durbin and S. J. Koopman. *Time Series Analysis by State Space Methods*. Oxford Univ. Press, New York, second edition, 2011.
- [2] A. C. Harvey. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge Univ. Press, Cambridge, 1989.
- [3] Jouni Helske. *KFAS: Kalman Filter and Smoother for Exponential Family State Space Models*, 2016. R package version 1.2.4.
- [4] Giovanni Petris, Sonia Petrone, and Patrizia Campagnoli. *Dynamic Linear Models with R*. Springer Verlag, 2009.

¹Se han desarrollado estudios serológicos como el [ENE-COVID](#); pero lo que permiten estimar es el porcentaje de personas que han desarrollado anticuerpos, es decir de infectados presentes y pasados. Podría imaginarse que los resultados proporcionan observaciones en t de $\sum_{j=1}^t \alpha_j$, siendo α_j el porcentaje de infectados en el periodo j ; pero esta única observación difícilmente permitirá estimar con alguna precisión el nivel de α_j para un j cualquiera.