

Reporte técnico sobre la estimación del R diario con el paquete EpiEstim del software R (4 de mayo de 2020)

Sumario. Se desarrolla la modelación matemática correspondiente a la estimación del R_t , el número de reproducción de la epidemia (es decir, la cantidad de casos secundarios producidos por un infectado primario en cada día t de la epidemia) que brinda el paquete **EpiEstim** [1]. El análisis proviene del artículo de Anne Cori, Neil M. Ferguson, Christophe Fraser y Simon Cauchemez [2] (del Imperial College de Londres) y del respectivo apéndice técnico de soporte al artículo [3].

Desarrollo

La epidemia comenzó en tiempo $t = 1$, observamos la incidencia I_s de la enfermedad, la cantidad de nuevos casos infectados el día s , hasta hoy, que es día T . Un insumo del modelo es entonces el vector de datos I_1, \dots, I_T (que son números enteros no negativos). En este caso se utilizan los datos del SINAE (corregidos por proporcionalidad entre el 16 de marzo y el 11 de abril, dado que se informaba únicamente de test positivos).

La distribución de la infectividad es constante a través del tiempo. Se denota mediante w_1, w_2, \dots, w_T donde $\sum_{i=1}^{\infty} w_i = 1$ y w_s representa la probabilidad de que un infectado en el día 1 infecte a un susceptible en el día $s + 1$. Por eso se toma en el caso paramétrico una infectividad gamma trasladada un día, con parámetros que da el usuario. En particular, en la estimación utilizamos una media de $\mu_0 = 3,95$ y un desvío estándar de $\sigma_0 = 4,75$ tomados de [4].

Se modela la transmisión (el contagio) de un infectado mediante un proceso de Poisson. La tasa mediante la que un infectado en $t - s$ infecta en tiempo t es $R_t w_s$, donde R_t es el número de reproducción de la epidemia (donde vamos “para atrás” en el tiempo), y la cantidad de infectados correspondiente es una distribución de Poisson de parámetro $R_t w_s$. Si agregamos todos los infectados de todos los días anteriores a t , resulta que la infectividad, definiendo

$$\Lambda_t = \sum_{s=1}^{t-1} I_{t-s} w_s = I_{t-1} w_1 + \dots + I_1 w_{t-1}, \quad t \geq 2,$$

tiene distribución de Poisson de parámetro

$$R_t \Lambda_t.$$

Aquí se supone que cada individuo infecta en forma independiente a los demás, utilizándose que la suma de variables de Poisson tiene distribución de Poisson. La versomilitud de la incidencia hoy I_t como función de las observaciones I_1, \dots, I_{t-1} , el perfil de infectividad $w = (w_s)$ y el número de reproducción R_t

es entonces

$$\mathbf{P}(I_t \mid I_{t-1}, \dots, I_1, w, R_t) = e^{-R_t \Lambda_t} \frac{(R_t \Lambda_t)^{I_t}}{I_t!}. \quad (1)$$

Para generar una estimación robusta de R_t , el número de reproducción de la epidemia en tiempo t se supone que la transmisibilidad que se estima hoy fué constante en un intervalo de hace τ días hasta hoy, es decir, fué constante en el intervalo $[t - \tau + 1, t]$. Se designa $R_{t,\tau}$ a ese número de reproducción. En nuestras estimaciones usamos el parámetro $\tau = 7$, que es el que viene por default en el **Epiestim**. Un intervalo mayor produce estimaciones mas suaves, menos sensibles a los cambios diarios. Bajo el supuesto anterior, la verosimilitud de la observación de los casos en el intervalo que es el vector $I_{t-\tau+1} \dots I_t$ dadas las observaciones anteriores $I_1 \dots I_{t-\tau}$, el perfil w y R_t se calcula aplicando la regla de la probabilidad condicional y la fórmula (1). Tenemos

$$\begin{aligned} \mathbf{P}(I_t, \dots, I_{t-\tau+1} \mid I_{t-\tau}, \dots, I_1, w, R_t) &= \prod_{s=t-\tau+1}^t \mathbf{P}(I_s \mid I_{s-1}, \dots, I_1, w, R_t) \\ &= \prod_{s=t-\tau+1}^t e^{-R_s \Lambda_s} \frac{(R_s \Lambda_s)^{I_s}}{I_s!}. \end{aligned} \quad (2)$$

Para obtener la estimación de R_t se utiliza un esquema Bayesiano con distribución a priori Gamma con parámetros (a, b) para $R_{t,\tau}$. En particular, se toman los parámetros $a = 1$ y $b = 5$ para la distribución a priori (que se toma constante para la estimación de cada día), lo que da una media de $\mu = 5$ y $\sigma = 5$. El valor que da el algoritmo como estimación de R_t es la mediana de la distribución a posteriori de R_t , y con la distribución se construye el intervalo de confianza. Recordemos que la densidad Gamma con esos parámetros es

$$p(x|a, b) = \frac{x^{a-1}}{\Gamma(a)b^a} e^{-x/b}. \quad (3)$$

($\Gamma(x)$ es la función Gamma.) Una variable gamma de parámetro $a = 1$ es una exponencial de parámetro $1/b$, y una gamma de parámetro $a = n$ es una suma de n exponenciales de parámetro $1/b$, es decir una Erlang de parámetros n y $1/b$. La esperanza y la varianza de una gamma con parámetros (a, b) viene dada por

$$\mu = ab, \quad \sigma^2 = ab^2,$$

y los parámetros se obtienen mediante

$$a = \frac{\mu^2}{\sigma^2}, \quad b = \frac{\sigma^2}{\mu}.$$

Calculamos ahora la distribución conjunta a posteriori de R_t y los datos obser-

vados, utilizando el cálculo (2) y la densidad en (3):

$$\begin{aligned}
\mathbf{P}(I_t, \dots, I_{t-\tau+1}, R_t \mid I_{t-\tau}, \dots, I_1, w) \\
&= \mathbf{P}(I_t, \dots, I_{t-\tau+1} \mid R_t, I_{t-\tau}, \dots, I_1, w) \mathbf{P}(R_t) \\
&= \prod_{s=t-\tau+1}^t e^{-R_s \Lambda_s} \frac{(R_s \Lambda_s)^{I_s}}{I_s!} \frac{(R_t)^{a-1}}{\Gamma(a) b^a} e^{R_t/b}
\end{aligned}$$

Ahora, bajo nuestro supuesto $R_{t,\tau} = R_{t-\tau+1} = R_{t-\tau+2} = \dots = R_t$, la fórmula anterior es

$$\begin{aligned}
\mathbf{P}(I_t, \dots, I_{t-\tau+1}, R_t \mid I_{t-\tau}, \dots, I_1, w) &= \prod_{s=t-\tau+1}^t e^{-R_s \Lambda_s} \frac{(R_s \Lambda_s)^{I_s}}{I_s!} \frac{(R_t)^{a-1}}{\Gamma(a) b^a} e^{-R_t/b} \\
&= \prod_{s=t-\tau+1}^t e^{-R_{t,\tau} \Lambda_s} \frac{(R_{t,\tau} \Lambda_s)^{I_s}}{I_s!} \frac{(R_{t,\tau})^{a-1}}{\Gamma(a) b^a} e^{-R_{t,\tau}/b} \\
&= C(e^{-(\frac{1}{b} + \sum_{s=t-\tau+1}^t \Lambda_s) R_{t,\tau}} (R_{t,\tau})^{\sum_{s=t-\tau+1}^t I_s + a - 1})
\end{aligned}$$

donde la constante C no depende de $R_{t,\tau}$:

$$C = \prod_{s=t-\tau+1}^t e^{\Lambda_s} \frac{(\Lambda_s)^{I_s}}{I_s!}$$

Obtenemos así que la distribución a posteriori de $R_{t,\tau}$ tiene también distribución Gamma, con parámetros

$$\hat{a} = a + \sum_{s=t-\tau+1}^t I_s, \quad \hat{b} = \frac{1}{\frac{1}{b} + \sum_{s=t-\tau+1}^t \Lambda_s}.$$

En particular, la media y el coeficiente de variación CV (la razón entre el desvío y la media) a posteriori de $R_{t,\tau}$ es

$$\hat{\mu} = \hat{a} \hat{b} = \frac{a + \sum_{s=t-\tau+1}^t I_s}{\frac{1}{b} + \sum_{s=t-\tau+1}^t \Lambda_s}, \quad CV = \frac{\hat{\sigma}}{\hat{\mu}} = \frac{1}{\sqrt{\hat{a}}} = \frac{1}{\sqrt{a + \sum_{s=t-\tau+1}^t I_s}}.$$

Esto permite programar en forma independiente al paquete **EpiEstim** el cálculo de la media y el desvío de los R_t diarios. Para ilustrar este procedimiento incluimos la incidencia al día de hoy, hasta el 3 de mayo y el cálculo del R diario producido por el paquete. Cabe aclarar que el paquete da diversas opciones distintas a la utilizada, consistentes en tomar un intervalo de constancia más largo, otra distribución a priori para R (lo que implica la necesidad de implementar un MCMC para estimar dado que no hay fórmulas explícitas), otra manera de calcular la infectividad con datos de parejas de infectador-infectado y otras.

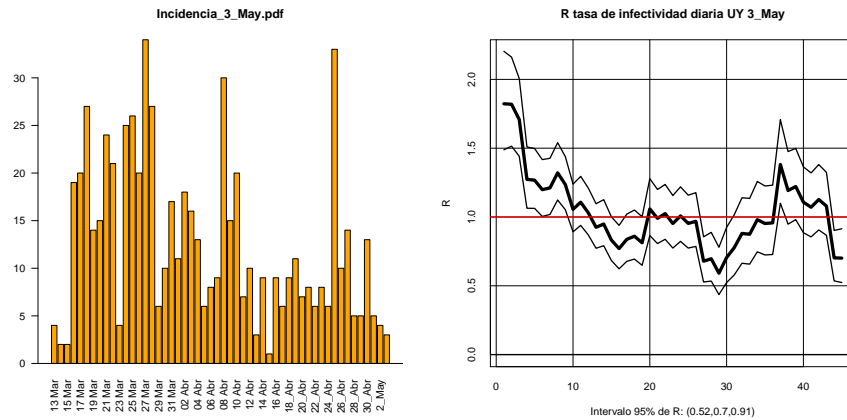


Figura 1: A la izquierda la incidencia reportada por el SINAIE, a la derecha el R diario estimado por el EpiEstim.

Referencias

- [1] <https://cran.r-project.org/package=EpiEstim>
- [2] Anne Cori, Neil M. Ferguson, Christophe Fraser, and Simon Cauchemez. A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics. *American Journal of Epidemiology*. Vol. 178, No. 9 DOI: 10.1093/aje/kwt133 Advance Access publication: September 15, 2013
- [3] A new framework and software to estimate time-varying reproduction numbers during epidemics: web material. (en referencia a [2]). <http://aje.oxfordjournals.org/>
- [4] Du Z, Xu X, Wu Y, Wang L, Cowling BJ, Ancel Meyers L. Serial interval of COVID-19 among publicly reported confirmed cases. *Emerg Infect Dis*. 2020 Jun [to be published]. <https://doi.org/10.3201/eid2606.200357>