



# Ajuste y calibración, con la llegada de nuevos datos, de un modelo epidemiológico simple

|                          |  |
|--------------------------|--|
| <b>Apellidos, nombre</b> | Benítez, Julio <sup>1</sup> (jbenitez@upv.es)<br>Carpitella, Silvia <sup>2</sup> (silvia.carpitella@unipa.it)<br>Izquierdo Sebastián, Joaquín <sup>1</sup> (jizquier@upv.es) |
| <b>Departamento</b>      | <sup>1</sup> Departamento de Matemática Aplicada<br><sup>2</sup> Department of Decision-Making Theory  |
| <b>Centro</b>            | <sup>1</sup> Universitat Politècnica de València<br><sup>2</sup> Institute of Information Theory and Automation, Praga   |



## 1 Resumen

Los denominados sistemas dinámicos están presentes en cualquier ámbito evolutivo. Modelos matemáticos adecuados (físicos, sociales, económicos, biológicos, etc.) constituyen un laboratorio 'in vitro' de tales sistemas, que permiten su estudio, simulación y conocimiento. Tales modelos dependen de determinados parámetros que caracterizan el modelo. Partiendo de condiciones iniciales dadas, permiten estimar la evolución del sistema y facilitan así la toma de decisiones. Para que la precisión de un modelo sea aceptable, los resultados que produce son comparados con datos provenientes de una monitorización adecuada, tan en tiempo real como sea posible, del sistema. Con los datos monitorizados, el modelo puede ser ajustado. Sin embargo, la propia dinámica del modelo es, a su vez, cambiante, por lo que los parámetros deben ser reajustados en determinados momentos y el modelo recalibrado. Es el caso de un modelo epidémico cuando se utiliza para el seguimiento de una epidemia real, en el que distintos parámetros pueden cambiar, por ejemplo, cuando un gobierno impone el confinamiento, lo que disminuye la ratio de contactos. En este artículo presentamos un modelo epidemiológico simple y procedemos al ajuste y reajuste del mismo utilizando datos obtenidos en diferentes momentos de tiempo. En la asignatura Matemáticas II de la doble titulación TELECO+ADE, los alumnos utilizan herramientas suficientes para entender la idea subyacente al ajuste y reajuste de modelos. El modelo epidemiológico que presentamos en este artículo es suficientemente simple como para ser resoluble 'a mano'. Esto permite introducir al alumno en el importante campo de los sistemas dinámicos y su calibración, algo esencial para que los resultados de predicción producidos sean útiles para la toma de decisiones.

## 2 Introducción

Los sistemas dinámicos tienen sus orígenes en la mecánica newtoniana. Al igual que en otras disciplinas de la ingeniería y otras ciencias sociales y naturales, un sistema dinámico viene descrito por una relación implícita que describe el estado del sistema a lo largo del tiempo, habitualmente en el corto o muy corto plazo. La relación es, con frecuencia, una ecuación diferencial (escalar, si involucra a solo una magnitud, o vectorial si involucra a varias magnitudes interdependientes). El estado del sistema se determina integrando, a partir de una condición inicial, el problema diferencial ya sea de manera analítica (las menos de las veces) o numérica. Cuando esto es posible, se pueden determinar los estados futuros, mediante una colección de puntos (valores de estado) sucesivos, que constituyen lo que se denomina una trayectoria u órbita.

En el campo de las Ciencias de la Salud, una línea de trabajo fundamental es el de la evolución de las epidemias. Es perfectamente claro, especialmente en estos momentos en que la humanidad está afectada por la pandemia de la COVID-19, que la evolución de una epidemia es un fenómeno evolutivo perfectamente clasificable como sistema dinámico.

Existen diversos modelos epidemiológicos, entre los que destaca el modelo denominado SIR, por las letras utilizadas para describir a los grupos más representativos en que se distribuye una población durante una epidemia [Anderson, 1991]. Utilizando la nomenclatura anglosajona, la  $S = S(t)$  representa a los "susceptibles", la  $I = I(t)$  a los "infectivos" y la  $R = R(t)$  a los "removals" (curados e inmunes, muertos o aislados). El sistema de ecuaciones diferenciales para este



modelo es:  $S' = -rSI$ ;  $I' = rSI - gI$ ;  $R' = gI$ , donde la prima indica derivada respecto del tiempo,  $r$  es la tasa de infección y  $g$  es la tasa con que los infectados son 'sacados' del sistema. El problema se completa con una condición inicial  $S_0, I_0, R_0$ , siendo estos valores los números iniciales de individuos susceptibles, infectados y excluidos del sistema, y  $S_0 = C - I_0 - R_0$ , siendo  $C$  la población total bajo el efecto de la epidemia.

En un modelo SIR elemental, como el nombrado, se puede probar que la tasa de infección es función de la capacidad infecciosa del virus y de la tasa de encuentros entre infectados y susceptibles. De entrada, la capacidad infecciosa del virus no es modificable, pero sí lo es la tasa de encuentros si se adoptan medidas adecuadas de aislamiento o confinamiento. Cuando se declara una epidemia en una población, son varios los parámetros que deberían ser conocidos para poder extraer información a futuro del modelo. En concreto, para el modelo SIR, se debe conocer los parámetros  $r, g, C, I_0$  (nótese que lo más normal es suponer que  $R_0 = 0$ , ya que todavía no ha habido tiempo de sacar a nadie del sistema sea por vía de curación o por vía de muerte). Con los datos obtenidos durante los primeros días de la epidemia solo es posible modelar de manera elemental justo esas primeras etapas del brote epidemiológico. No cabe esperar que ese modelo sirva para la duración completa de la epidemia. Consecuentemente, se siguen tomando datos de los números de individuos para el modelo presentado: infectados y excluidos (curados y muertos del sistema). Los datos obtenidos sirven para poner a prueba el modelo y, en cuanto se aprecian diferencias sustantivas, se procede a la recalibración. Con frecuencia, tal recalibración puede ser muy radical, implicando no solo el recálculo de algunos parámetros, sino hasta el cambio del propio modelo a uno más complejo que atienda a factores no considerados en un modelo inicial simple. Esto permite ajustar la capacidad del modelo resultante, con el que se tratará de seguir prediciendo la evolución de la epidemia (ver, por ejemplo, [Fanelli y Piazza, 2000] que, entre otros muchos autores, publican *updates* de tales modelos tratando de ajustar los datos en diversas regiones del mundo).

Como el modelo SIR es no lineal y este artículo tiene como grupo objetivo directo o indirecto a estudiantes de primeros cursos de matemáticas (en particular, de ecuaciones diferenciales), utilizaremos un modelo muy simple, a nivel de un primer curso de ecuaciones diferenciales, útil, para ejemplificar la dinámica del ajuste y reajuste de parámetros de un modelo, para alumnos.

Habitualmente estos ajustes se realizan utilizando técnicas de aproximación mínimo-cuadráticas (ver, por ejemplo, [Izquierdo y Torregrosa, 1997], ya que los datos no suelen seguir una evolución describible mediante funciones cerradas.

En este artículo, utilizaremos datos tomados en distintos momentos de la evolución de la epidemia, y consideraremos dos modelos sencillos (el malthusiano y el logístico) que permiten comprender en cierta medida la dinámica epidémica, pero, sobre todo, muestran con claridad el proceso (a veces, artesanal) de ajuste y reajuste (calibración) de tales modelos.

### 3 Objetivos

Tras concluir con la lectura de este documento, serás capaz de:

- Ajustar un modelo epidemiológico sencillo a los datos observados inicialmente.

- Reajustar un modelo epidemiológico más complejo a los nuevos datos reales que se van observando en etapas más avanzadas de la epidemia.
- Extender los procedimientos anteriores a modelos más complejos y de otras ramas de la ciencia.

## 4 Desarrollo

Para leer este artículo deberás tener nociones de:

| Requisitos  |
|---|
| 1. Conocimientos elementales sobre ecuaciones diferenciales                     |
| 2. Conocimientos elementales de métodos de linealización de modelos no lineales |
| 3. Ajuste mínimo cuadrático   |

Tabla 1. Requisitos básicos

### DESCRIPCIÓN DEL PROCESO

En este apartado describimos con detalle el modelo epidemiológico que vamos a utilizar, comentamos las hipótesis en que se basa, dejando claras sus limitaciones, y obtenemos la solución analítica del modelo matemático.

#### Descripción del modelo

En este modelo se supone que un reducido número de infectados,  $P_0$ , aparece en una población cerrada constituida por  $M$  individuos, siendo todos susceptible de contraer la enfermedad. Se define como  $P(t)$  al número de individuos infectados de la población en el tiempo  $t$ . Si se supone que la infección puede extenderse por igual a todos los individuos susceptibles, entonces  $P(t)$  crece desde su valor inicial,  $P_0$ .

Una ecuación posible para  $P$  es la de Verhulst o logística

$$P' = kP \left(1 - \frac{P}{M}\right),$$

siendo  $k$  una constante positiva que mide la ratio de infección, que expresa que la variación de infectados es proporcional al número actual de infectados, con el límite de la saturación dado por el paréntesis.

La solución analítica para este problema viene dada por

$$P = \frac{MP_0}{P_0 + (M - P_0)e^{-kt}}$$

Se observa con claridad que este modelo es no lineal, por lo que determinadas preguntas acerca del comportamiento de la magnitud de interés, pueden resultar clave a la hora de la toma de decisiones. Por ejemplo, ¿cuándo se alcanzará la tasa máxima de infección?, ¿cuál es el comportamiento asintótico de la solución (si lo tiene)?

## Llegada y uso de los primeros datos

Como, obviamente, el proceso es desconocido, no se conocen valores precisos de los parámetros del problema, es decir, de

- $k$ , capacidad de infección del virus ya que, en principio, el virus puede ser completamente desconocido y no hay ningún control sobre el contacto entre individuos;
- $P_0$ , número inicial de individuos infectados (quizá un único paciente cero);
- $M$ , capacidad máxima de la población (que puede no ser conocida con precisión; por ejemplo, porque alguna parte de la población esté completamente aislada y no es, por tanto, susceptible).

En estas condiciones solo queda esperar a tener datos para poder tener estimaciones de los parámetros.

En este artículo vamos a suponer que los primeros datos se obtienen a los pocos días, según recoge la tabla 1.

**Tabla 1.** Datos de infectados de los primeros días

| $t$ (días)       | 6  | 7  | 8  | 9  |
|------------------|----|----|----|----|
| $P$ (infectados) | 16 | 20 | 24 | 29 |

No parece sensato tratar de calibrar el modelo con esa cantidad tan escasa y temprana de datos. No obstante, dado que se conoce que estos procesos tienen inicialmente un carácter prácticamente exponencial, tales datos pueden ser utilizados para obtener una buena estimación del parámetro  $k$ , utilizando un modelo malthusiano simple:

$$P' = kP,$$

ya que, inicialmente  $P$  es muy pequeño comparado con  $M$ , por lo que el paréntesis del modelo logístico es prácticamente igual a 1.

Es perfectamente sabido que la solución de este modelo es:

$$P = P_0 e^{kt}.$$

Aunque este modelo tampoco es lineal, es fácilmente linealizable:

$$\log P = \log P_0 + kt.$$

Utilizando mínimos cuadrados se puede obtener una estimación de los parámetros  $k$  y  $P_0$ , obteniendo,  $P_0 = 4.971$ ;  $k = 0.197$ , es decir,

$$P(t) = 4.971e^{0.197t}.$$

Obviamente este modelo no puede ser utilizado para toda la epidemia, ya que pronto alcanzará valores no compatibles con el fenómeno modelado, como veremos en el siguiente apartado.

## Recepción de datos más avanzados

Supondremos ahora que se dispone de nuevos datos dados por la tabla 2.

**Tabla 2.** Datos de infectados de días más avanzados

|                  |     |     |     |
|------------------|-----|-----|-----|
| $t$ (días)       | 20  | 25  | 30  |
| $P$ (infectados) | 163 | 281 | 403 |

En primer lugar, observamos cómo el modelo malthusiano proporciona, para los días 20, 25 y 30, valores de infectados 254, 678 y 1813, respectivamente, que claramente sobreestiman los valores medidos, como típicamente ocurre con todos los modelos malthusianos para tiempos avanzados.

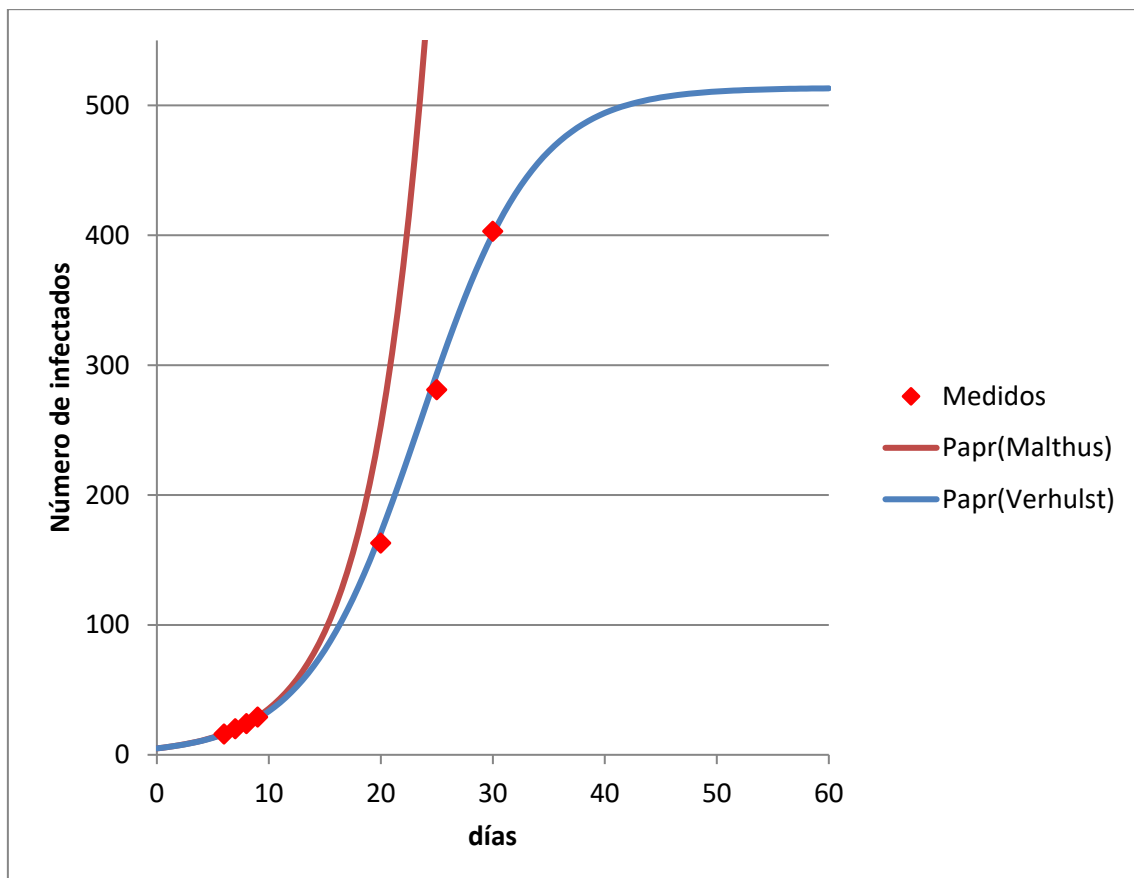
En este artículo utilizamos estos datos para ajustar, finalmente, el modelo completo, aunque utilizando los valores de los parámetros  $k$  y  $P_0$  obtenidos anteriormente. Así, el modelo resulta

$$P = \frac{4.97M}{4.97 + (M - 4.97)e^{-0.197t}},$$

donde resta obtener el valor de  $M$ .

De nuevo, este modelo es no lineal, aunque también puede ser linealizado:

$$(P_0 e^{kt} - 1)M = PP_0(e^{kt} - 1).$$



**Figura 1.** Diferentes elementos del problema considerado



Al particularizar esta ecuación para los tres nuevos datos considerados, se obtiene un sistema algebraico incompatible cuya solución mínimo-cuadrática proporciona el valor  $M = 513.5$ , como población total a la que se aplica el modelo, que, finalmente, permite estimar la población de infectados como:

$$P = \frac{2552.51}{4.97 + 508.53e^{-0.197t}}$$

La representación gráfica de los elementos de este problema puede observarse con detalle en la figura 1, que presenta los puntos medidos, la aproximación malthusiana y la aproximación logística o de Verhulst, finalmente obtenida.

## Discusión y trabajo a desarrollar

Obviamente, este es un ejemplo muy simple de carácter académico no utilizable en el caso de una epidemia real, como la COVID-19, que padece nuestro mundo desde finales de 2019. El artículo solo pretende mostrar al alumno un prototipo muy elemental de la metodología a seguir en procesos, descritos por modelos matemáticos conocidos y perfectamente establecidos a nivel teórico, ante datos desconocidos y/o inciertos. La finalidad es que el alumno sepa que cuando se trabaja con modelos más completos (con menos simplificaciones) y, por tanto, más complejos, las ideas sobre la calibración de modelos que se utilizan son semejantes y, con frecuencia, se pueden llevar a cabo mediante herramientas y técnicas elementales de aproximación, como las que el alumno ve en la asignatura Matemáticas II del doble grado TELECO+ADE.

Para profundizar más en la modelación matemática y sus ajustes, y en la extracción de información útil en la toma de algunas decisiones, el lector puede utilizar los módulos de aprendizaje [ep1](#) y [ep2](#) del repositorio *RiuNet* de la Universitat Politècnica de València, para realizar los siguientes experimentos e introspecciones.

- 1) Reproducir la curva de Verhulst de la Figura 1 mediante [ep1](#), utilizando la opción disponible de "representación de la serie temporal".
- 2) Comprobar cómo afecta la tasa de infección a la evolución de infectados, y observar el carácter asintótico de la solución (hay que utilizar la opción que representa la serie temporal, y quizá para algunos valores del coeficiente sea preciso alargar el periodo de simulación). Con [ep2](#) se puede hacer una comparativa para dos valores de la tasa de infección.
- 3) Observar cuándo se alcanza la tasa máxima de infección; para esto será interesante utilizar la opción que permite representar el plano de fases del número de infectados respecto de su derivada.
- 4) Observar cómo el pico (el máximo) de la derivada es más bajo para valores más pequeños de la tasa de infección. Relacionar tanto valores absolutos del número de infectados, como valores de su variación, con la capacidad del sistema sanitario, y deducir por qué es necesario rebajar al máximo el pico. Por ejemplo, suponer que la capacidad máxima de admisión en el hospital de la zona es de cierto número de infectados por día y obtener cuál debería ser la tasa de infección para que dicho valor no se supere. (Obsérvese que para modificar la tasa de infección las autoridades deben decretar medidas adecuadas, tales como el confinamiento, ya que, en principio la capacidad de infección del virus no es controlable.)
- 5) Otras ideas que te inspiren el 'juego' con dichos módulos de aprendizaje.



## 5 Cierre

En este artículo te hemos presentado un mecanismo para ajustar (y reajustar), es decir, calibrar progresivamente, mediante técnicas mínimo-cuadráticas, los parámetros de un sencillo modelo epidemiológico matemático descrito mediante ecuaciones diferenciales, utilizando datos reales (simulados en este artículo) de la evolución de una epidemia que se recogen a diario, y que deben encajar en el modelo, si este debe tener buenas capacidades de predicción. El mecanismo descrito está basado en técnicas vistas en la asignatura Matemáticas II del doble grado TELECO+ADE de la Universitat Politècnica de València. Tras describir someramente el modelo diferencial, se proporciona la solución, que resulta ser el conocido modelo logístico, que es no lineal, y se realiza una calibración de dicho modelo en dos etapas. En la primera, utilizando datos tempranos, se ajustan el coeficiente de infección y el número inicial de infectados; y mediante otro proceso de linealización, se ajusta la población máxima afectada por la infección para construir el modelo completo. Finalmente, utilizando dos objetos de aprendizaje se propone la realización de un conjunto de experimentos que permiten realizar una introspección en la obtención y la interpretación de modelos matemáticos.

## 6 Bibliografía

- R.M. Anderson. Discussion: the Kermack-McKendrick epidemic thresh-old theorem. *Bulletin of mathematical biology*, 53(1):1–32, 1991.
- D. Fanelli, F. Piazza, Analysis and forecast of COVID-19 spreading. Preprint en <https://www.researchgate.net/publication/339943580>, 2020.
- J. Izquierdo, J.R. Torregrosa. *Álgebra y ecuaciones diferenciales*. Tomos I y II, Universidad Politècnica de Valencia, SPUPV-97.669, 1997.
- J. Izquierdo. Laboratorio virtual <http://hdl.handle.net/10251/146211>, 2020.
- J. Izquierdo. Laboratorio virtual <http://hdl.handle.net/10251/145901>, 2020.