

# TEMA 8.- MODELOS EPIDEMIOLÓGICOS

## CONTENIDOS

1	INTRODUCCIÓN . . . . .	1
2	MODELOS EPIDEMIOLÓGICOS . . . . .	1
3	MODELO DE BERNOULLI (1760) . . . . .	1
4	MODELO SIR . . . . .	3
4.1	Estudio cualitativo del modelo SIR . . . . .	4
4.2	Calibración del modelo para epidemias no severas . . . . .	7
4.3	Calibración del modelo para epidemias severas . . . . .	8
5	OTROS MODELOS TIPO SIR . . . . .	9
5.1	Modelo SIRS . . . . .	9
5.2	Modelo SEIR . . . . .	11
6	BIBLIOGRAFÍA . . . . .	12

## 1. INTRODUCCIÓN

El estudio de los modelos epidemiológicos se suele dejar para cursos de posgrado o de máster, por lo que en la programación inicial de este curso no estaba incluido. Sin embargo, al comienzo de la primavera del año 2020, mientras que estaba impartiendo este curso en el Grado de Matemáticas en mi universidad, todo el mundo se paralizó por la pandemia de COVID-19. En esas circunstancias, no podíamos seguir con el temario oficial del curso como si no hubiera ocurrido nada tan importante. Así que redujimos el temario para incluir dos semanas en las que vimos los modelos epidemiológicos más sencillos. Por ello, en este tema no vamos a ver un estudio cualitativo muy profundo de estos modelos, salvo en los casos más sencillos, usando la computación para visualizar el comportamiento de las soluciones de los modelos más complejos. Para un estudio más detallado, recomendamos el libro de Murray [4]. Un libro muy ameno para saber más de historia de estos modelos es el de Bacaër et al. [1].

## 2. MODELOS EPIDEMIOLÓGICOS

El primer intento de simular matemáticamente cómo se extiende una enfermedad contagiosa en una población se debe a Daniel Bernoulli, que en 1760 publicó un trabajo en el que intentaba modelizar la evolución de la viruela y cómo afectaría una posible inoculación del virus en la población, lo que en dicha época podía suponer que el virus se expandiera aún más. Sin embargo, no fue hasta que Kormack y McKendrick en 1927 propusieron el modelo SIR, que estos modelos empezaron a tener relevancia y aplicabilidad.

Las técnicas que se usan para construir estos modelos son bastante similares a las que se usan en los modelos de dinámica de poblaciones, aunque desde el punto de vista biológico la expansión de agentes patógenos (virus, bacterias y parásitos, principalmente) en una población no se ajusta bien al aumento de una población de una especie depredadora (el patógeno) frente a una presa (la población infectada).

## 3. MODELO DE BERNOULLI (1760)

Con este modelo Bernoulli modelizó la expansión de la viruela aunque se puede aplicar a cualquier otra enfermedad contagiosa que, una vez superada por un individuo, le deja inmune para el resto de su vida. Aunque es un modelo bastante poco aplicable en la realidad, aparte del interés que tiene desde el punto de vista histórico por ser el primero que aplicó las ecuaciones diferenciales para predecir la expansión

de una epidemia, muestra algunas de las herramientas básicas que posteriormente se han usado para construir modelos más sofisticados.

Considere una cohorte<sup>1</sup> de individuos que han nacido en un año determinado (que fijamos como el año  $t = 0$ ) y sea  $y(t)$  el número de esos individuos que siguen vivos  $t$  años después. Por otro lado, llamemos  $x(t)$  al número de individuos de ese grupo que en el año  $t$  todavía no han contraído la enfermedad, por lo que serán *susceptibles* a contagiarse.

Los parámetros que van a determinar el modelo van a ser:

- $\beta$  = ratio de contagio de la viruela, es decir, la tasa que mide la velocidad a la que los susceptibles se contagian.
- $\nu$  = ratio de muerte por viruela, o sea, mide la proporción de contagiados que mueren debido a la enfermedad.
- $\mu(t)$  = ratio de mortalidad por causas diferentes a la viruela.

El modelo describe la velocidad a la que los susceptibles se contagian (y dejan de ser susceptibles),  $dx/dt$ , y la evolución de la mortalidad de toda la cohorte,  $dy/dt$ :

$$\frac{dx}{dt} = -(\beta + \mu(t))x, \quad \frac{dy}{dt} = -\nu\beta x - \mu(t)y, \quad x(0) = x_0, \quad y(0) = x_0, \quad t \geq 0, \quad x_0 > 0 \quad (1)$$

Obsérvese que este modelo es no lineal y no es autónomo. Para resolverlo, Bernoulli propuso considerar la variable  $z = x/y$ , que viene a representar la proporción de personas no contagiadas en  $t$  respecto a los que siguen vivos en dicho año  $t$ . Derivando y usando (1), se tiene que

$$z' = -\beta z(1 - \nu z), \quad z(0) = 1, \quad t \geq 0,$$

lo que se parece a la ecuación logística pero con coeficiente de crecimiento negativo, y que tiene la ventaja importante de que no interviene la ratio  $\mu(t)$ .

En principio, esta ecuación se resuelve relativamente fácil por variables separadas, pero hay que tener en cuenta el signo de  $(1 - \nu z)$  para aplicar correctamente los logaritmos. Esto se ve mejor si antes de resolverla estudiamos cualitativamente esta ecuación, que es de la forma  $z' = f(z)$ . Claramente se ve que tiene dos puntos de equilibrio  $z = 0$  y  $z = 1/\nu$ . Por tanto, salvo las soluciones con  $z_0 = 0$  ó  $z_0 = 1/\nu$ , ninguna solución puede cortar estas soluciones constantes  $z(t) = 0$  ó  $z(t) = 1/\nu$  para  $t \geq 0$ . Así que si  $z_0 \in (0, 1/\nu)$ , se tiene que  $z(t) \in (0, 1/\nu)$ ,  $\forall t \geq 0$ , mientras que en el caso  $z_0 > 1/\nu$ ,  $z(t) > 1/\nu$ ,  $\forall t \geq 0$ . Por otra parte, es lógico que  $\nu < 1$ , pues en caso contrario tendríamos que se mueren más individuos por viruela que los contagiados por viruela, es decir, se tendría  $\nu\beta > \beta$ , lo que no tendría mucho sentido. Así que  $1/\nu > 1 = z(0)$  y, necesariamente,  $0 \leq z(t) < 1/\nu$ ,  $\forall t \geq 0$ . En consecuencia,  $z'(t) < 0$ ,  $\forall t \geq 0$ , y  $z$  es decreciente.

Sabiendo esto, podemos separar las variables

$$\frac{dz}{z(1 - \nu z)} = -\beta dt \Rightarrow \int_1^z \frac{ds}{s(1 - \nu s)} = -\beta \int_0^t du \Rightarrow \int_1^z \frac{ds}{s} + \int_1^z \frac{\nu ds}{1 - \nu s} = -\beta t,$$

donde  $0 < z < 1/\nu$ . Integrando, resulta

$$\ln \left( \frac{(1 - \nu)z}{1 - \nu z} \right) = -\beta t \Rightarrow \frac{(1 - \nu)z}{1 - \nu z} = e^{-\beta t}$$

de donde se obtiene la proporción de individuos que siguen sanos respecto a los que han sobrevivido en  $t$ :

$$z(t) = \frac{e^{-\beta t}}{1 - \nu + \nu e^{-\beta t}}, \quad t \geq 0$$

Bernoulli estimó en su trabajo que  $\nu = \beta = 1/8$  por lo que, por ejemplo, afirmó que la proporción de personas no contagiadas entre los que en ese momento tenían 20 años ( $t = 20$ ) era  $z(20) = 0.0927$ , es decir, prácticamente un 9% de los que tenían 20 años. En cambio, la proporción entre los de 5 años era  $z(5) = 0.568$ , o sea, casi un 57% de los niños de 5 años tenía la viruela.

Una característica que está implícita en este modelo y que aparece en algunas epidemias es la dependencia de la edad de los individuos a la hora de contagiarse y sobrevivir a la infección.

<sup>1</sup>Este nombre se usa mucho en epidemiología para denominar al conjunto de individuos en el que se expande la enfermedad.

#### 4. MODELO SIR

El modelo presentado por Kormack y McKendrick en 1927 sigue siendo actualmente relevante para simular la expansión de ciertas enfermedades contagiosas que producen inmunidad a los individuos que las superan.<sup>2</sup> Una de las características más interesantes de este modelo es que consigue reproducir el hecho de que una epidemia no se acaba solo cuando toda la población está infectada, sino que existen otros factores que determinan que se pare la expansión de una enfermedad.

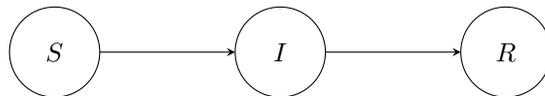
En este modelo, a diferencia del de Bernoulli, se supone que el número total de individuos  $N$  de la población considerada permanece constante a lo largo del tiempo, pues se supone que el ciclo vital del patógeno es mucho menor que el de los individuos que se pueden contagiar.

El proceso de expansión arranca cuando un pequeño grupo de infectados se introduce en una población mucho mayor. Además, como dijimos arriba, se supone que cuando una persona se infecta o bien termina superando la enfermedad y queda inmunizada para el resto de su vida, o bien muere.

Así, se separa la población total en tres clases disjuntas o *compartimentos*<sup>3</sup> y se considera la evolución temporal del número de individuos en estas clases. Si denotamos el tiempo por  $t \geq 0$ , donde  $t$  representa horas, días o semanas, según la infección:

- $S = S(t)$  = número de individuos *susceptibles* de contraer la enfermedad en  $t$ .
- $I = I(t)$  = número de individuos *infectados* en  $t$ .
- $R = R(t)$  = número de individuos *recuperados* en  $t$ , es decir, los que han superado la enfermedad (y son inmunes), los infectados que se aíslan del resto de la población y los que han fallecido, ya que se considera que estos individuos dejan de influir en el desarrollo de la epidemia.

Se entiende que un individuo podrá seguir una evolución esquematizada por



Para construir el modelo se hacen las siguientes suposiciones sobre la transmisión y el período de incubación del patógeno:

- El aumento de infectados es proporcional al número de infectados y de susceptibles. En consecuencia, los susceptibles disminuyen en la misma cantidad.
- La velocidad a la que los individuos pasan de estar infectados a estar recuperados es proporcional al número de infectados.
- El período de incubación es suficientemente pequeño como para ser despreciable, es decir, que un individuo susceptible pasa a ser infectado en el mismo instante en el que se contagia con el patógeno.

Modelizar la primera de estas suposiciones es quizás lo más complicada. Se puede afirmar que la velocidad a la que aumentan los infectados es  $rSI$  para cierta constante  $r > 0$  llamada **tasa de contagio** de la enfermedad. Para precisar numéricamente este parámetro tenemos que tener en cuenta los siguientes valores que se tienen que estimar según los datos que se recaben de la epidemia.

En primer lugar, hay que calcular el número medio  $\lambda$  de contactos por persona por unidad de tiempo. Como los encuentros podrían ser entre dos susceptibles o dos infectados, lo que no influiría en la propagación de la enfermedad, habrá que tener en cuenta la probabilidad de que en un encuentro cualquiera haya un infectado. Esta probabilidad se calcula dividiendo el número de casos favorables a que esto ocurra,  $I(N - 1)$  (pues cada infectado se puede encontrar con  $N - 1$  personas), entre el número total de encuentros posibles  $N(N - 1)$ , por tanto, es

$$\frac{I(N - 1)}{N(N - 1)} = \frac{I}{N}.$$

<sup>2</sup>En general, las infecciones provocadas por bacterias no suelen dejar inmunes a los supervivientes como en el caso de la peste bubónica. En cambio, muchas enfermedades provocadas por virus sí lo hacen, como en el caso de la viruela, el sarampión, etc.

<sup>3</sup>Por esta separación estricta entre clases, los modelos se suelen llamar *compartimentales*.

Por otra parte, dependiendo de qué enfermedad se esté estudiando, habrá una probabilidad  $p$  de que haya contagio de la enfermedad en un encuentro entre un susceptible y un infectado.

Supongamos que un susceptible al azar camina por la calle. En una unidad de tiempo tendrá  $\lambda$  encuentros, de ellos  $I/N$  serán con infectados, y existe una probabilidad  $p$  de que dichos contactos acaben en contagio. Por tanto, la probabilidad de que un susceptible se contagie por unidad de tiempo es  $\lambda \cdot (I/N) \cdot p$ . Como hay  $S$  susceptibles, el número medio de susceptibles que se contagian será  $\lambda \cdot (I/N) \cdot p \cdot S$  por unidad de tiempo y, por tanto, la tasa de contagio es

$$r = \frac{\lambda p}{N} \tag{2}$$

Muchos autores prefieren expresarla en función del producto

$$\beta := \lambda p, \quad r = \frac{\beta}{N} \tag{3}$$

donde  $\beta$  es el *promedio de contactos con contagio por unidad de tiempo*, ya que es el producto del promedio de contactos por persona por unidad de tiempo por la probabilidad de que haya contagio en un contacto entre un susceptible y un infectado. Además, un valor interesante es  $1/\beta$  que es el tiempo medio entre contactos en los que hay contagio.

En lo que respecta a la segunda suposición, esto es, el cálculo de la tasa de cambio del segundo compartimento  $I$  al tercero  $R$ , es claro que será  $aI$  para cierta constante  $a > 0$  que se llama **tasa de recuperación**. En la práctica lo que se suele medir es  $1/a$ , que es el *tiempo medio de recuperación de la enfermedad o tasa de recuperación por unidad de tiempo*. Por ejemplo, si estamos midiendo el tiempo  $t$  en días y se sabe experimentalmente que el tiempo medio que tarda en recuperarse un infectado es 5 días, se considera que en un día se ha recuperado  $1/5$  de los infectados, es decir,  $a = 1/5$ .

Agrupando todo esto, el modelo SIR de Kormack y McKendrick (1927) es

$$\begin{aligned} \frac{dS}{dt} &= -rSI & S(0) &= S_0 > 0 \\ \frac{dI}{dt} &= rSI - aI & I(0) &= I_0 > 0 \\ \frac{dR}{dt} &= aI & R(0) &= 0 \end{aligned} \tag{4}$$

Hay que observar que se tiene

$$\frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0$$

lo que implica que  $S(t) + I(t) + R(t)$  es constante para todo  $t$  y, necesariamente se respeta la suposición inicial de que

$$S(t) + I(t) + R(t) = N, \quad \forall t \geq 0 \tag{5}$$

En principio, (4) es un sistema diferencial de orden 3 pero debido a este invariante (5),  $R = N - S - I$ , por lo que basta resolver el sistema formado por las dos primeras ecuaciones, es decir, realmente tenemos que resolver un sistema de dimensión 2. Si intentamos aplicar la linealización para extraer conclusiones sobre el comportamiento de sus soluciones, tenemos el inconveniente de que tiene infinitos puntos de equilibrio  $(S, 0)$ ,  $\forall S \in \mathbb{R}$ , por lo que no son aislados y la linealización no nos ayuda.

### 4.1 Estudio cualitativo del modelo SIR

Vamos a intentar averiguar cómo se comportan las soluciones de (4)-(5) a partir del propio sistema diferencial, entendiendo que solo nos van a interesar las soluciones positivas  $S \geq 0, I > 0$ .

En primer lugar, de la primera ecuación de (4) se tiene que  $S$  va a ser siempre decreciente, es decir,  $S(t) \leq S_0, \forall t \geq 0$ . Además, de la segunda ecuación evaluada en  $t = 0$  tenemos

$$I'(0) = rS(0)I(0) - aI(0) = I_0(rS_0 - a).$$

En consecuencia, si  $S_0 < a/r$ ,  $S(t) \leq S_0 < a/r$  y, por tanto,

$$I'(t) = I(t)(rS(t) - a) \leq 0, \quad \forall t \geq 0,$$

lo que implica que  $I(t)$  es decreciente para cualquier  $t \geq 0$  (recuérdese que  $I_0 = N - S_0 > 0$ ). Es más,  $I(t) \rightarrow 0$  cuando  $t \rightarrow \infty$  y la enfermedad no se expande. En otras palabras, *existe un valor umbral*

$$\rho = a/r \tag{6}$$

de tal modo que si la población inicialmente susceptible  $S_0$  está por debajo de ese umbral, la enfermedad no se expande y no hay epidemia.

Este umbral en la práctica se puede expresar considerando que  $S_0 = N$ , lo que es una buena aproximación cuando  $N$  es grande, y definiendo el valor

$$R_0 = \frac{rN}{a} \tag{7}$$

por lo que si  $R_0 < 1$ , la enfermedad no se expande. Este parámetro  $R_0$  se suele llamar **tasa básica de reproducción de la epidemia** y se puede interpretar también como *el número de nuevos infectados producidos por un sólo infectado si toda la población es susceptible*. Esto se puede entender si recordamos que  $rN = \beta$  y se escribe

$$R_0 = \frac{\beta}{a} = \frac{\frac{1}{\frac{1}{\beta}}}{\frac{1}{a}} = \frac{\text{tiempo medio de recuperación de la enfermedad}}{\text{tiempo medio entre contactos con contagio}}$$

Por ejemplo, si un infectado tarda 4 días en recuperarse de la enfermedad, se introduce en una población en la que solo hay susceptibles, y el tiempo medio que tarda en contagiar a otro es 2 días, se tiene  $R_0 = 4/2 = 2$ . O, lo que es lo mismo,  $R_0$  es el promedio del número de susceptibles que un infectado contagia antes de recuperarse (curarse o morirse).

Evidentemente, conocer esta tasa de reproducción es fundamental para controlar una epidemia. La manera más eficaz de disminuir  $R_0$  es reducir el número de susceptibles mediante la vacunación. Para que sea efectiva una vacuna, tendrá que administrarse a una proporción de individuos suficiente  $pN$ ,  $0 < p < 1$ , de tal modo que el número de susceptibles  $S_0 = N - pN$  esté por debajo del umbral  $(1 - p)N < \rho$ . Como  $\rho = N/R_0$ , la proporción  $p$  de vacunados necesaria tendrá que ser entonces

$$p > 1 - \frac{1}{R_0}.$$

Obsérvese que esto implica que no hace falta vacunar al 100% de la población para que la vacuna sea efectiva. La vacunación no sólo protege a un individuo sino que da protección a toda la población a lo largo del tiempo pues mantiene la tasa de reproducción por debajo de 1, evitando que la epidemia pueda volver a rebrotar. Esto es lo que se llama *inmunidad de grupo*<sup>4</sup> y es la razón de por qué muchas personas no vacunan a sus hijos y éstos no se enferman, lo que suele dar la impresión errónea de que las vacunas son inútiles.

El parámetro  $R_0$  (7) también explica el hecho de que las epidemias se suelen parar antes de que toda la población se contagie. Por otro lado, esta tasa también explica que desde el momento que  $R_0 \gtrsim 1$  la epidemia puede brotar y extenderse muy rápidamente.

Volviendo al estudio cualitativo del sistema (4), se tiene que las trayectorias en el plano  $(S, I)$  verifican

$$\frac{dI}{dS} = \frac{rSI - aI}{-rSI} = -1 + \frac{\rho}{S}, \quad (S > 0, I > 0),$$

Separando las variables e integrando

$$dI = \left(-1 + \frac{\rho}{S}\right) dS \Rightarrow \int_{I_0}^I du = \int_{S_0}^S \left(-1 + \frac{\rho}{s}\right) ds \Rightarrow I - I_0 = -S + S_0 + \rho \ln S - \rho \ln S_0,$$

por lo que

$$I + S - \rho \ln S = I_0 + S_0 - \rho \ln S_0 = \text{constante},$$

o, lo que es lo mismo, las trayectorias en el plano  $(S, I)$  son las curvas de nivel de la función  $H(S, I) = I + S - \rho \ln S$ .

<sup>4</sup>Si traducimos directamente del inglés "*herd immunity*" deberíamos decir "inmunidad de rebaño".

Como  $I_0 + S_0 = N$ , todas las trayectorias comienzan sobre la recta  $I + S = N$  del plano  $(S, I)$  y permanecen dentro del triángulo formado por esta recta y los ejes  $S = 0$  e  $I = 0$ , pues  $0 \leq S + I \leq N$  para todo  $t > 0$ . Estudiando como son estas funciones de la forma  $I = N - S + \rho \ln(S/S_0)$  se ve que tienen que ser como se ven en la Fig. 1, en la que se ve que todas estas curvas son cóncavas y alcanzan su máximo en el punto  $S = \rho$ . Con esta información podemos predecir el mayor número de contagiados que va a provocar la epidemia, que es

$$I_{max} = I_0 + S_0 - \rho \ln S_0 - \rho + \rho \ln \rho = N - \rho + \ln \frac{\rho}{S_0}$$

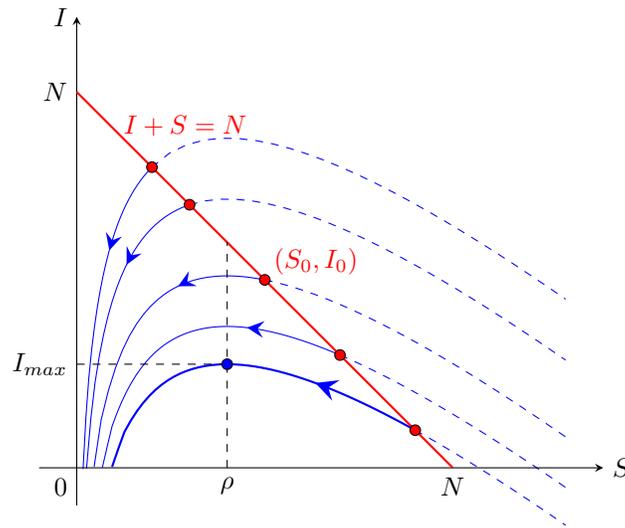


Figura 1: Trayectorias en el plano  $(S, I)$  del modelo SIR (4)

Además, la Fig. 1 también nos muestra claramente que si comenzamos con un valor inicial de infectados  $I_0$  de tal forma que los susceptibles inicialmente sean  $S_0 = N - I_0 > \rho$ , la epidemia evolucionará desde  $I_0$  hasta  $I_{max}$  y a partir de ahí comenzará a remitir. De hecho, si  $I_0$  está próximo a  $I_{max}$ , la enfermedad se expandirá muy poco. Por otro lado, si  $S_0 < \rho$ , la epidemia no prospera.

El eje  $I = 0$  requiere una atención especial ya que cualquier punto de esta recta es un punto crítico del sistema (punto singular, no aislado). Además, en cualquier caso  $(S_0, I_0)$  con  $I_0 \neq 0$ , la solución  $I(t) \rightarrow 0$  cuando  $t \rightarrow \infty$ . Hay que entender que aunque en la Fig. 1 parece que cualquier trayectoria acaba tocando el eje  $I = 0$ , realmente no es así, pues esas curvas no están definidas en  $I = 0$ . Así que podríamos pensar erróneamente que la gráfica demuestra que todas las soluciones verifican que  $S(t) \rightarrow S(\infty) < \rho$  cuando  $t \rightarrow \infty$ . Esto es cierto *si*  $S(\infty)$  existe, lo que se demuestra teniendo en cuenta la tercera ecuación del modelo (4), que al dividirla entre la segunda resulta, cuando  $S \neq 0, I \neq 0$ ,

$$\frac{dS}{dR} = -\frac{1}{\rho}S \Rightarrow \frac{dS}{S} = -\frac{1}{\rho}dR \Rightarrow S = S_0 e^{-R/\rho} \geq S_0 e^{-N/\rho} > 0.$$

Por tanto,  $S$  es una función decreciente acotada inferiormente por 0, por lo que tiene límite cuando  $t \rightarrow \infty$ ,  $S(\infty)$ , que necesariamente, tiene que ser

$$0 < S(\infty) < \rho.$$

Como  $I(\infty) = 0$ , forzosamente  $R(\infty) = N - S(\infty)$ , y de la relación entre  $S$  y  $R$  anterior,

$$S(\infty) = S_0 \exp\left(-\frac{N - S(\infty)}{\rho}\right)$$

por lo que  $z = S(\infty)$  es la raíz positiva  $0 < z < \rho$  de la ecuación

$$z = S_0 \exp\left(-\frac{N - z}{\rho}\right),$$

que es fácilmente resoluble por cualquier método numérico sencillo con mucha precisión. Así, una vez que hemos obtenido este valor  $S(\infty)$ , conoceremos cual es el *número total de infectados por la epidemia*

$$I_{total} = S_0 + I_0 - S(\infty)$$

Una idea que se extrae de que  $I(\infty) = 0$  y  $S(\infty) > 0$  es que *la epidemia se para por la falta de infectados, no por la falta de susceptibles*.

## 4.2 Calibración del modelo para epidemias no severas

En general, el proceso de calibración de este modelo para un caso concreto es complicado. Habitualmente, los datos recogidos por las autoridades sanitarias sobre el número de nuevos infectados cada día no es fiable. Aparte de las dificultades que dicha recogida de datos conlleva en la práctica, lo normal es que un enfermo sea registrado cuando ya lleva varios días o semanas infectado. A menudo, el dato que suele ser más fiable es el del número de recuperados cada día (fallecidos o dados de alta), es decir, en términos del modelo SIR, el valor  $dR/dt$ . De los cálculos anteriores tenemos que

$$\frac{dR}{dt} = aI = a(N - R - S) = a(N - R - S_0 e^{-R/\rho}), \quad R(0) = 0.$$

Esta ecuación se puede resolver numéricamente, si conocemos previamente los parámetros necesarios  $N, a, r, S_0$ . Cuando Kormack y McKendrick diseñaron su modelo SIR en 1927 los métodos numéricos no estaban tan desarrollados como hoy en día, por lo que propusieron una simplificación del modelo bajo ciertas condiciones. Con esta simplificación, además, el problema numérico resultante también es más simple, con lo que se mejora la precisión del modelo.

Así, en el caso en que  $R/\rho$  es pequeño (al menos  $R/\rho < 1$ ), lo que es esperable en las llamadas epidemias *no severas*, es decir, epidemias que afectan a poca población en comparación con la población total, Kormack y McKendrick propusieron aproximar la exponencial del segundo miembro por los primeros desarrollos de su desarrollo de Taylor,

$$e^{-R/\rho} \approx 1 - \frac{R}{\rho} + \frac{R^2}{2\rho^2} \tag{8}$$

lo que es una buena aproximación. Llevándolo a la ecuación diferencial se obtiene la ecuación aproximada

$$\frac{dR}{dt} = a \left( N - S_0 + \left( \frac{S_0}{\rho} - 1 \right) R - \frac{S_0 R^2}{2\rho^2} \right)$$

que se puede resolver por separación de variables. Después de algunos cálculos elementales se obtiene la solución<sup>5</sup>

$$R(t) = \frac{\rho^2}{S_0} \left( \left( \frac{S_0}{\rho} - 1 \right) + \alpha \tanh \left( \frac{\alpha a t}{2} - \phi \right) \right) \tag{9}$$

donde

$$\alpha = \left( \left( \frac{S_0}{\rho} - 1 \right)^2 + \frac{2S_0(N - S_0)}{\rho^2} \right)^{1/2}, \quad \phi = \tanh^{-1} \left( \frac{\frac{S_0}{\rho} - 1}{\alpha} \right).$$

Si se tiene buena información sobre el número de recuperados de la enfermedad se puede ajustar el modelo usando esta ecuación. Además, como  $\tanh x \rightarrow 1$  cuando  $x \rightarrow \infty$ , se tiene

$$R(\infty) = \frac{\rho^2}{S_0} \left( \left( \frac{S_0}{\rho} - 1 \right) + \alpha \right),$$

y es natural suponer que  $N - S_0 \approx 0$ ,

$$\alpha \approx \left( \frac{S_0}{\rho} - 1 \right) \Rightarrow R(\infty) \approx \frac{\rho^2}{S_0} 2 \left( \frac{S_0}{\rho} - 1 \right) = \frac{2\rho}{S_0} (S_0 - \rho),$$

<sup>5</sup>Recuérdese que  $\tanh x = \frac{e^x - e^{-x}}{e^x + e^{-x}}$  y  $\tanh^{-1} x = \frac{1}{2} \ln \left( \frac{1+x}{1-x} \right)$ .

Table 1: Número de infectados por gripe diarios en el internado

$t$ (días)	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$I(t)$	3	8	28	75	221	291	255	235	190	125	70	28	12	5

lo que hace patente el factor umbral que ya vimos anteriormente: basta que el número de susceptibles sea  $N \approx S_0 = \rho + \varepsilon$ ,  $\varepsilon > 0$ , para que la epidemia arranque y  $R(\infty) = 2\varepsilon$ .

Por otro lado, para calibrar el modelo en algunos casos, también es útil derivar (9) y usar esta expresión alternativa de  $dR/dt$ ,

$$\frac{dR}{dt} = \frac{a\alpha^2\rho^2}{2S_0} \operatorname{sech}^2\left(\frac{\alpha at}{2} - \phi\right) \tag{10}$$

Así lo aplicaron, por ejemplo, en el caso de una epidemia de peste en Bombay que duró casi un año en 1905, en la que entre un 80 y un 90% de infectados morían, por lo que el número de recuperados por semana  $dR/dt$  se puede aproximar por el número de muertos por semana. Además, se estimó que la epidemia no era *severa*, es decir, que no afectaba a un alto porcentaje de la población ( $R/\rho < 1$ ), por lo que se buscó la mejor función del tipo  $f(t) = A \operatorname{sech}^2(Bt + C)$  que ajustaba por mínimos cuadrados los datos del número de muertos por semana, obteniendo

$$\frac{dR}{dt} = 890 \operatorname{sech}^2(0.2t - 3.4),$$

lo que simuló con bastante precisión el desarrollo real de la epidemia. Obsérvese que con esto tenemos tres valores determinados

$$\frac{a\alpha^2\rho^2}{2S_0} = 890, \quad \frac{\alpha a}{2} = 0.2, \quad \phi = 3.4.$$

Como  $\alpha = \alpha(\rho, S_0, N)$ ,  $\phi = \phi(\rho, S_0)$ , tenemos tres ecuaciones no lineales con cuatro incógnitas  $\{a, \rho, S_0, N\}$ . Desde que se obtenga una estimación de alguna de ellas a partir de los datos, se podrán calibrar los otros tres parámetros y, en consecuencia, el modelo (4).

### 4.3 Calibración del modelo para epidemias severas

En el caso de una epidemia severa ( $R/\rho > 1$ ), no se puede usar la aproximación de la exponencial (8), por lo que el ajuste de los parámetros es más complicado y se necesitará conocer mejor las características de la enfermedad y recoger datos reales del proceso de expansión.

En un boletín del *British Medical Journal* en 1978 informaron de la rápida expansión de una epidemia de gripe entre los alumnos de un internado en el norte de Inglaterra entre el 22 de enero y el 4 de febrero de 1978. Según detallan en el informe, uno de los internos que había pasado las vacaciones de Navidad con su familia en Hong Kong volvió al internado ya con síntomas febriles y en un período de 14 días se contagiaron 512 niños de los 763 que residían en el colegio. En aquella ocasión, los síntomas se aliviaban recluyendo a los niños a la cama y solo unos pocos tuvieron síntomas graves.

La ventaja en este caso fue que en esos años ya se tenía bastante experiencia médica y conocimiento científico de las epidemias, por lo se hizo una recopilación exacta del número de infectados cada día, que está recogida en la Tabla 1.

Los chicos pasaban en la cama entre dos y cuatro días, por lo que  $1/a \in [2, 4]$  o, lo que es lo mismo,  $a \in [0.25, 0.5]$ . Por otra parte, ya que en el primer día se contagiaron 2 directamente del "paciente 0", como  $N = 763, I_0 = 1, S_0 = 762$ , se puede hacer una primera aproximación

$$\frac{dS}{dt} = -rSI \approx -2 \Rightarrow r \approx \frac{2}{762 \cdot 1} = 2.62 \cdot 10^{-3},$$

por lo que podemos afirmar que  $r \in [10^{-3}, 5 \cdot 10^{-3}]$ .

**Ejercicio.-** Construir un programa en Matlab/Octave que estime los valores de  $(a, r) \in [0.25, 0.5] \times [10^{-3}, 5 \cdot 10^{-3}]$  que aproximan lo mejor posible los datos obtenidos de la Tabla 1.

Indicaciones a seguir:

- Tomar  $Na = 100$  valores equiespaciados de  $a$  en  $[0.25, 0.5]$  y  $Nr = 100$  valores equiespaciados de  $r$  en  $[10^{-3}, 5 \cdot 10^{-3}]$ .
- Para cada par de valores  $(a, r)$  de los anteriores, resolver el sistema formado por las dos primeras ecuaciones del modelo SIR (4) usando la rutina **ode45**, obteniendo una solución aproximada  $I_{ap}(t)$ .
- Calcular la norma uniforme del vector

$$error(a, r) = (I(t) - I_{ap}(t))_{t=1}^{14}$$

donde  $I(t)$  se toma de la Tabla 1.

- Tomar el par  $(a, r)$  que da el menor valor de  $\|error(a, r)\|_{\infty}$  de todos. Este par nos da la tasa de recuperación y la tasa de contagio del modelo calibrado.
- Hacer un gráfico en el que se compare el número de infectados con el número de susceptibles y el de recuperados  $R(t) = N - S(t) - I(t)$  que nos da el modelo ya calibrado.
- Calcular el parámetro  $\rho$  correspondiente, la tasa básica de reproducción y el tiempo medio entre contactos.

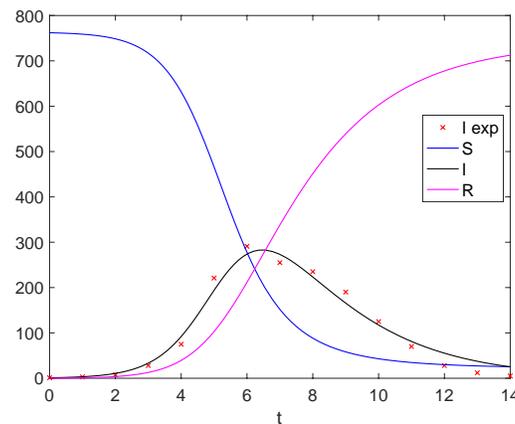
Con este programa podemos ver que los valores óptimos son

$$a = 0.4525, \quad r = 2.16 \cdot 10^{-3}$$

con los que se obtiene el menor error respecto los datos experimentales y con los que se obtienen los parámetros

$$\rho = 209.491, \quad R_0 = 3.63739, \quad \beta = 1.64808,$$

siendo el período medio de recuperación  $1/a = 2.2$  días y el tiempo medio entre contactos igual a 0.6 días. Con estos valores óptimos obtenemos el gráfico siguiente que predice la evolución de cada uno de los compartimentos, en el que se han dibujado con cruces rojas los valores experimentales dados en la Tabla 1. Obsérvese cómo coincide el pico de infectados con el momento que el número de susceptibles baja del valor umbral  $\rho = 209.491$ .



En cualquier caso, está claro que la fiabilidad de este modelo (y de cualquier otro que intente modelizar este fenómeno) depende críticamente de la fiabilidad de los datos de la epidemia y del conocimiento que se tenga de la enfermedad.

## 5. OTROS MODELOS TIPO SIR

### 5.1 Modelo SIRS

Un modelo un poco más general que el SIR (4) considera el caso de una enfermedad en la que cierta proporción de los recuperados no se inmunizan y vuelven a ser susceptibles:

$$\begin{aligned} \frac{dS}{dt} &= -rSI + bR & S(0) &= S_0 > 0 \\ \frac{dI}{dt} &= rSI - aI & I(0) &= I_0 > 0 \\ \frac{dR}{dt} &= aI - bR & R(0) &= 0 \end{aligned} \tag{11}$$

En este caso, también se da que

$$\frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0$$

y, por tanto también se tiene el invariante (5).

Un individuo de la población según este modelo seguiría un esquema del tipo



por lo que suele recibir el nombre de **modelo SIRS**.

El estudio cualitativo de este modelo es quizás más directo que el del modelo SIR anterior, ya que si sustituimos  $R = N - I - S$  en la primera ecuación, lo que tenemos es un sistema de dimensión 2:

$$\begin{aligned} \frac{dS}{dt} &= -rSI + b(N - S - I) & S(0) &= S_0 > 0 \\ \frac{dI}{dt} &= rSI - aI & I(0) &= I_0 > 0 \end{aligned} \tag{12}$$

con  $S_0 + I_0 = N$ , y se tienen sólo dos puntos de equilibrio aislados

$$\begin{aligned} S_1 &= N, & I_1 &= 0 & (R_1 &= 0) \\ S_2 &= \rho, & I_2 &= b \frac{N - \rho}{a + b} & (R_2 &= N - S_2 - I_2) \end{aligned}$$

donde  $\rho = a/r$  es el mismo valor umbral (6) del modelo SIR anterior. El primer punto corresponde al caso en que no hay infectados y la epidemia se ha erradicado, mientras que para que el segundo sea un caso a considerar, tiene que ocurrir que  $I_2 > 0$ , es decir, que  $N > \rho$  o, lo que es lo mismo,

$$R_0 = \frac{\beta}{a} = \frac{rN}{a} > 1,$$

con lo que aparece la misma tasa básica de reproducción de la epidemia que en el caso SIR y se tiene el mismo efecto umbral.

Al ser puntos de equilibrio aislados la linealización nos ayuda a determinar qué ocurre con las trayectorias de las soluciones cercanas a estas soluciones estacionarias. Si denotamos

$$F(S, I) = -rSI + b(N - S - I), \quad G(S, I) = rSI - aI,$$

tenemos la matriz jacobiana

$$J(S, I) = \frac{\partial(F, G)}{\partial(S, I)} = \begin{pmatrix} -(rI + b) & -(rS + b) \\ rI & rS - a \end{pmatrix}$$

Si evaluamos en el primer punto  $(S_1, I_1) = (N, 0)$  tenemos

$$J(N, 0) = \begin{pmatrix} -b & -(rN + b) \\ 0 & rN - a \end{pmatrix}$$

por lo que tiene dos autovalores reales  $-b < 0$  y  $rN - a$ . Si  $R_0 < 1$ , entonces  $rN - a < 0$  y  $J$  tiene dos autovalores reales negativos, con lo que la solución estacionaria  $(N, 0)$  es un nodo atractor. Es decir, en este caso, la epidemia decaería rápidamente. En cambio, si  $R_0 > 1$ , tenemos un autovalor positivo por lo que el punto  $(N, 0)$  es un punto de silla y, en consecuencia, no podemos garantizar la erradicación de la epidemia.

En lo que respecta al segundo punto  $(S_2, I_2)$  tenemos

$$J(S_2, I_2) = \begin{pmatrix} -(rI_2 + b) & -(a + b) \\ rI_2 & 0 \end{pmatrix},$$

así que la traza de  $J$  es  $T = -(rI_2 + b) < 0$  y su determinante  $D = rI_2(a + b) > 0$ , por lo que esta solución estacionaria, siempre que  $R_0 > 1$  (es decir, si  $N > \rho$ , pues en caso contrario este punto no estaría en el dominio  $S \geq 0, I > 0$  de interés), será atractora. Ahora bien, dependiendo del signo de  $T^2 - 4D$  podrá ser un nodo (propio o impropio) o un foco atractor.

En consecuencia, si no hay ningún otro factor que pudiera intervenir en este proceso (vacunación o medicación, por ejemplo), este modelo predice que si no hay inmunidad, la enfermedad se mantendrá dentro de la población a lo largo del tiempo. Esto es lo que ha ocurrido con muchas enfermedades a lo largo de la historia, como la malaria.

## 5.2 Modelo SEIR

En algunas epidemias, los modelos anteriores no se ajustan bien ya que en su expansión influyen mucho aquellos individuos que ya están contagiados pero que durante un cierto período (*período de incubación*) no son capaces de infectar a otros. En el modelo SEIR además de los tres compartimentos  $S$ ,  $I$  y  $R$ , se añade el compartimento  $E(t)$  de los **individuos expuestos**, es decir, el grupo de individuos que portan la enfermedad sin capacidad de contagiarla en el momento  $t$ :

$$\begin{aligned} \frac{dS}{dt} &= -rSI & S(0) &= S_0 > 0 \\ \frac{dE}{dt} &= rSI - cE & E(0) &= 0 \\ \frac{dI}{dt} &= cE - aI & I(0) &= I_0 > 0 \\ \frac{dR}{dt} &= aI & R(0) &= 0 \end{aligned} \tag{13}$$

Los parámetros  $r > 0$  y  $a > 0$  significan lo mismo que en los modelos anteriores, mientras que ahora aparece la *ratio de incubación*  $c > 0$ , que se mide a partir de valor  $1/c$  que es el *período medio de incubación de la enfermedad*, con lo que en este modelo la evolución de un individuo seguirá el esquema



Además, vuelve a aparecer el invariante

$$\frac{dS}{dt} + \frac{dE}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0 \Rightarrow S + E + I + R = N,$$

con el que podemos reducir el sistema (13) a uno de dimensión 3.

Si consideramos  $\tilde{I} = E + I$ , la suma total de los infectados en la población (sean contagiosos o no), sumando la segunda y tercera ecuaciones en (13) tenemos

$$\frac{d\tilde{I}}{dt} = I(rS - a),$$

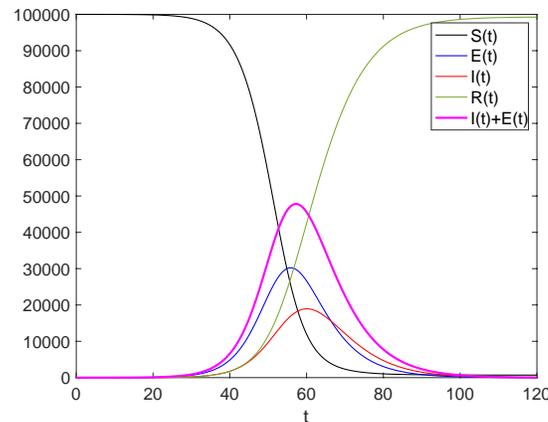
por lo que para que el número de infectados crezca necesariamente volveremos a tener un valor umbral  $\rho = a/r$  de tal modo que si  $S_0 < \rho$  la expansión de la epidemia decae, y se vuelve a tener una *tasa básica de reproducción*  $R_0 = \beta/a$ .

El efecto del nuevo compartimento  $E$  se observa en que, si el período de incubación es relativamente alto, la expansión de la epidemia se ralentiza en comparación con el modelo SIR.

Las herramientas teóricas necesarias para un estudio cualitativo más profundo de este problema se escapan al nivel de un curso de Grado, pero sí que podemos trabajar dicho modelo numéricamente.

**Ejercicio:** varios autores afirman que la pandemia provocada por el coronavirus SARS-CoV-2 se ajusta al modelo SEIR (13) de expansión. Vamos a suponer que el período medio de recuperación de la enfermedad es de 5 días ( $a = 1/5$ ) aproximadamente, con un período medio de incubación previo de 7 días ( $c = 1/7$ ), en los que se cree que un infectado no es capaz de contagiar. Las estimaciones de estos valores han ido variando con las sucesivas cepas y variantes pero esta elección de  $a$  y  $c$  es razonable. La estimación de la ratio de contagio  $r$  sigue siendo algo en discusión. Hay estudios que calculan que el promedio de contactos con contagio por unidad de tiempo  $\beta$  puede estar entre 0.59 y 1.68. Usando Matlab/Octave, estudiar la evolución de la enfermedad durante 120 días, tomando como un valor de referencia  $\beta = 1$  en una población de  $N = 100\,000$  personas a la que llega un único infectado, según el modelo SEIR (13).

Construyendo este programa obtendremos la gráfica siguiente donde podemos ver la evolución de los cuatro compartimentos a lo largo de los 120 días y, además, el número total de contagiados  $I(t) + E(t)$ .



Obsérvese que en el pico de la curva de  $I(t) + E(t)$  llega a haber casi 50000 infectados, es decir, un 50% de la población total. Durante los peores momentos de la epidemia (hasta el día 70 aproximadamente) la curva de expuestos es bastante mayor que la de infectados. Como en el modelo SIR, a partir de que el número de susceptibles baja de un cierto valor (alrededor de 20000 individuos) la epidemia empieza a decrecer.

## 6. BIBLIOGRAFÍA

- [1] N. Bacaër, R. Bravo de la Parra, J. Ripoll, *Breve historia de los modelos matemáticos en dinámica de poblaciones*, Cassini, París, 2008.
- [2] L. Edelstein-Keshet, *Mathematical models in biology*, SIAM, 2005.
- [3] W. O. Kermack, A. G. McKendrick, *Contributions to the mathematical theory of epidemics*, Proc. R. Soc. Lond. A, 1927, 1932 y 1933.
- [4] J. D. Murray, *Mathematical biology I. An introduction*, 3rd edition, Springer, 2002.