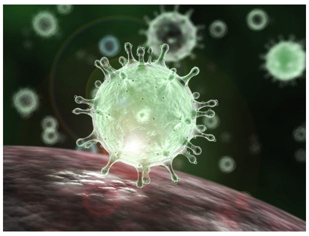
MAT^2
MATerials MATemàtics
Volum 2020, treball no. 1, 12 pp. ISSN: 1887-1097
Publicació electrònica de divulgació del Departament de Matemàtiques
de la Universitat Autònoma de Barcelona
www.mat.uab.cat/matmat

Un modelo simple para el número de infectados por Covid-19

Salomón Rebollo-Perdomo

Explicaré un modelo matemático sencillo que se ajusta y predice el comportamiento de los datos oficiales en España. Estas notas pretenden contribuir a la educación científica de todos para concienciarnos respecto a la situación tan grave del Covid-19 y mostrar, a través de gráficas, la importancia del confinamiento y aislamiento social en la lucha contra esta pandemia.

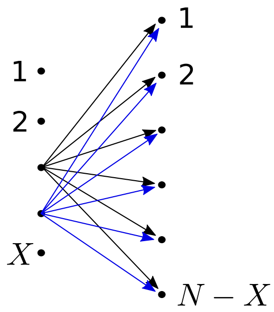


1. Lo elemental

El primer paso es entender cómo se expande el contagio. Pensemos en una fiesta que se desarrolla en un salón, donde cada invitado se desplaza libremente, intercambia saludos y conversa con el resto de invitados. Si en la fiesta hay una cantidad N de invitados y al inicio una cantidad X_0 de ellos están infectados con un virus, por ejemplo el de la gripe, entonces al inicio de la fiesta hay una cantidad $N - X_0$ de no infectados. ¿Cuál es el número X de invitados infectados que habrá al final de la fiesta si esta dura 8 horas?

Una hipótesis (basada en estudios de epidemiología) dice que la razón (tasa o velocidad) de crecimiento, respecto del tiempo, del número de infectados es proporcional al número de interacciones entre infectados y no infectados.

Regresemos a nuestro ejemplo de la fiesta. Podemos analizar las interacciones entre los X infectados y los $N - X$ no infectados de la siguiente forma: ponemos en una columna los infectados y en otra los no infectados. Cada infectado interacciona con el resto de invitados, en particular con los $N - X$ no infectados, esta acción la representamos mediante una flecha como lo muestra la siguiente figura.



Así, el número de interacciones entre los infectados y los no infectados es igual al número de flechas que podemos formar desde la columna izquierda hasta la columna derecha (en la figura hemos colocado solamente las flechas correspondientes a dos infectados). Es sencillo deducir que tal número es el producto de la cantidad X por la cantidad $N - X$.

Si representamos la razón de crecimiento del número de infectados respecto del tiempo por el simbolo dX/dt, entonces la hipótesis que hemos mencionado antes:

> "dX/dt es proporcional al número de interacciones entre los X infectados y $N - X$ no infectados"

se representa matemáticamente como:

$$\frac{dX}{dt} = k\, X\, (N - X),$$

donde k es la constante de proporcionalidad.

¡No se asuste por favor con la expresión! Lo importante que debemos entender de esta es lo siguiente:

> cuanto mayor es el producto $X\,(N - X)$ (interacciones entre infectados y no infectados) más rápido crece el número de infectados.

Sigamos con nuestro ejemplo de la fiesta. Supongamos que cada hora verificamos la cantidad de invitados que tienen "gripe", mediante un test instantáneo a todos los invitados, entonces obtenemos los datos X_0, X_1, X_2, X_3, etc. La razón de crecimiento dX/dt se puede obtener a partir de los datos de la forma:

$$\frac{X_1 - X_0}{1 - 0}, \frac{X_2 - X_1}{2 - 1}, \frac{X_3 - X_2}{3 - 2}, \ldots$$

A continuación explico el significado de estos términos. Pensemos por ejemplo, en el tercer término. El numerador indica la diferencia del número de infectados en la tercera hora menos el número de infectados de la segunda hora, es decir, representa el incremento en el número de infectados desde la segunda hora hasta la tercer hora de la fiesta. El denominador representa la diferencia entre el tiempo en que se tomó el tercer test y el tiempo en que se tomó el segundo test. En otras palabras, el denominador indica la unidad de tiempo que estamos usando para medir el incremento en el número de infectados, en este caso la unidad es una hora.

Bajo el escenario planteado, estamos diciendo que la razón de crecimiento dX/dt en cada hora, desde el comienzo de la fiesta, es:

$$X_1 - X_0, X_2 - X_1, X_3 - X_2, \ldots$$

pues el denominador siempre es 1 (en este caso).

Debido a que conocemos los datos $X_0, X_1, X_2, X_3, \ldots$, entonces podemos estimar el valor de la constante de proporcionalidad k, a través de:

$$\frac{X_1 - X_0}{X_1(N - X_1)}, \frac{X_2 - X_1}{X_2(N - X_2)}, \frac{X_3 - X_2}{X_3(N - X_3)}, \ldots$$

Estas expresiones provienen simplemente de despejar la constante k de nuestra primera ecuación y usar la información del párrafo anterior. Entonces el promedio, k_0, de estos valores calculados es una estimación de k. Regresando a nuestra primera ecuación obtenemos:

$$\frac{dX}{dt} = k_0\, X\, (N - X).$$

Esta expresión nos permite saber cual será el número de infectados después de cuatro horas de iniciada la fiesta. Simplemente, tenemos que usar la expresión:

$$X_4 - X_3 = k_0\, X_4\, (N - X_4)$$

recordando que la razón de crecimiento, dX/dt, es igual a $X_4 - X_3$. Esta ecuación la podemos resolver como aprendimos en la enseñanza media. Reagrupamos respecto de X_4:

$$k_0\, (X_4)^2 - (k_0\, N - 1)\, X_4 - X_3 = 0.$$

¡Esta es una ecuación cuadrática!, su solución positiva es:

$$X_4 = \frac{(k_0\, N - 1) + \sqrt{(k_0\, N - 1)^2 + 4\, k_0\, X_3}}{2\, k_0}.$$

Ahora, ilustremos estas ideas con un ejemplo concreto. La siguiente tabla muestra datos oficiales de España para el Covid-19 hasta el 11 de marzo de 2020.

En la primera columna se representa la fecha. Considerando que a partir del 24 de febrero el número de infectados aumentó diariamente, el 7 en la segunda columna representa el séptimo día de incremento consecutivo de

2-mar-2020	7	125	
3-mar-2020	8	169	0,000000005528
4-mar-2020	9	228	0,000000005494
5-mar-2020	10	282	0,000000004066
6-mar-2020	11	365	0,000000004828
7-mar-2020	12	430	0,000000003209
8-mar-2020	13	674	0,000000007686
9-mar-2020	14	999	0,000000006907
10-mar-2020	15	1622	0,000000008155
11-mar-2020	16	2128	0,000000005049

infectados, el 8 el octavo día y así sucesivamente. En la tercera columna se representa la cantidad de infectados.

Con estos últimos diez datos realizaremos el análisis general expuesto en los párrafos anteriores. Es decir, los datos de la tercera columna serán los valores X_7, X_8, X_9, $X_{10}, \ldots, X_{15}$, X_{16}. La cuarta columna representa la razón de crecimiento diario de infectados, es decir,

$$\frac{X_8 - X_7}{X_8\,(N - X_8)}, \frac{X_9 - X_8}{X_9\,(N - X_9)}, \frac{X_{10} - X_9}{X_{10}\,(N - X_{10})}, \ldots, \frac{X_{16} - X_{15}}{X_{16}\,(N - X_{16})}.$$

El promedio de estas razones de crecimiento es $k_0 = 0{,}000000005658\,$.

Entonces, una predicción para la cantidad de infectados del 12 de marzo sería

$$X_{17} = \frac{(k_0\,N - 1) + \sqrt{(k_0\,N - 1)^2 + 4\,k_0\,X_{16}}}{2\,k_0},$$

donde $N = 47100000$ es la población de España según [6].

Sustituyendo los valores de N, k_0 y X_{16} en la fórmula anterior obtenemos que

$$X_{17} = 2901\,.$$

Usando este valor podemos predecir el número de infectados para el 13 de marzo a través de la fórmula

$$X_{18} = \frac{(k_0\,N - 1) + \sqrt{(k_0\,N - 1)^2 + 4\,k_0\,X_{17}}}{2\,k_0},$$

que arroja el valor $X_{18} = 3954\,$. Aplicando esta idea sucesivamente obtenemos los valores

$$X_{19} = 5390,\ X_{20} = 7347,\ X_{21} = 10015 \quad \text{y} \quad X_{22} = 13652,$$

que corresponden a los días 14, 15, 16 y 17 de marzo de 2020, respectivamente. Los valores informados oficialmente por el Gobierno de España fueron:

		Real	Predicción
12-mar-2020	17	2950	2901
13-mar-2020	18	4209	3954
14-mar-2020	19	5753	5390
15-mar-2020	20	7753	7347
16-mar-2020	21	9191	10015
17-mar-2020	22	11178	13652

Notemos que con esta idea simple se predice con cierta exactitud los valores para los cuatro siguientes días después del día 11 de marzo. Las predicciones para los días 16 y 17 de marzo ya no son buenas. Esto es normal, por ejemplo la predicción meteorológica no va más allá de unos pocos días, los motivos son varios. Dos de los más importantes, que explican por qué el modelo actual falla a mediano y largo plazo son, por un lado, que un infectado en una ciudad no interacciona con todos los no infectados del país y, por otro, que los no infectados al pasar el tiempo tienen miedo de contagiarse y evitan salir, lo que disminuye el número de interacciones entre infectados y no infectados. Esto permite ralentizar el crecimiento del número de infectados. Sin embargo, a pesar de estas cuestiones importantes, resulta que en este caso el modelo simple da una buena aproximación para el corto plazo.

2. La idea simple

Una pregunta surge inmediatamente: ¿podemos modificar el modelo para que se reflejen en él las cuestiones planteadas? La respuesta es sí. La idea simple y sencilla, que da origen al título del artículo, se basa en lo siguiente. En primer lugar, nuestro modelo simple

$$\frac{dX}{dt} = k\, X\, (N - X)$$

ya refleja el hecho de que un infectado solamente interactúe con un porcentaje de la población de no infectados. La constante de proporcionalidad k representa tanto la posibilidad de contagio en una interacción como la cantidad de interacciones de un infectado con los no infectados. Por lo tanto, solamente falta incorporar la información de que cada día habrá menos no infectados que interactúen con los infectados.

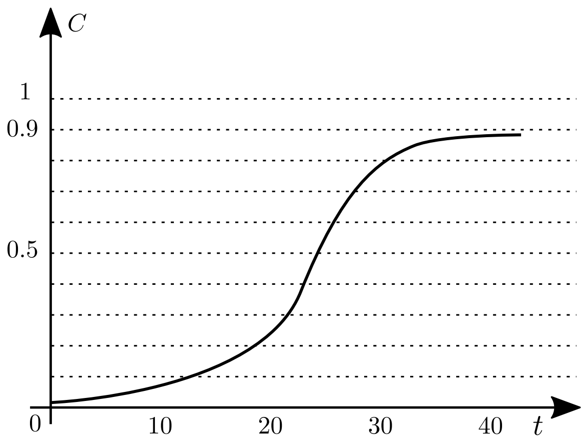
En el modelo básico, la cantidad de no infectados que interactúan con los infectados es $N - X$. Esta cantidad la podemos hacer disminuir al pasar los días realizando la siguiente operación:

$$(1 - C)\, N - X,$$

donde C representa la cantidad de confinados sobre la población total, por lo que C tiene un valor entre 0 y 1. Si el valor de C es cercano a 1, la cantidad $1-C$ es cercana a cero. La consecuencia de esto es que el valor del producto $(1-C)\,N$ es pequeño y por lo tanto la cantidad $(1-C)\,N-X$ de no infectados que interactúan con los X infectados es pequeña. Así, obtenemos una propuesta de modelo:

$$\frac{dX}{dt} = k\,X\,\big((1-C)\,N - X\big).$$

Pero, ¿cómo debería comportarse la cantidad C al pasar el tiempo (los días)? La respuesta es obvia, al inicio de los contagios no hay confinamiento por lo cual C debe ser cercana a 0, pero conforme pasan los días y el número de infectados crece, la cantidad de confinados aumentará, por lo que después de algunos días C debe estar próxima al valor 1. Por ejemplo, si el gobierno decreta cuarentena en el país, entonces de forma gradual la población se confinará y solamente un porcentaje pequeño (policía, médicos, abastecimiento básico, etc.) continuará interactuando. Por lo cual, es natural suponer que el valor de C va dependiendo del tiempo y que tenga un comportamiento como el de la figura siguiente.



La variable t en el eje horizontal de la figura representa el tiempo, en este caso medido en días. Así la curva de la figura indica que durante los primeros 20 días (aproximadamente) menos del 20 % de la población está en casa. El día 25 el 50 % de la población ya está confinada, para el día 35 lo estará más del 80 % y a partir del día 40 (aproximadamente) solamente un 10 % de la población no está en casa.

Como se deduce de la figura, el valor de C depende del tiempo t, lo cual se representa como $C(t)$. Existe gran cantidad de expresiones matemáticas, es decir de funciones $C(t)$, cuyas gráficas tienen la forma descrita en la figura anterior. En general, dichas funciones no son estudiadas en la enseñanza media, pero con un poco de esfuerzo un estudiante puede descubrir alguna de ellas. Una de las funciones más simples con esta propiedad se expresa como:

$$C = C(t) = \frac{C_{\max}}{1 + a\ \exp(b\,t)}, \tag{1}$$

donde $C_{\max}$ es la proporción de confinamiento máxima esperada, a es una constante positiva y el parámetro b en la función exponencial, $\exp(b\,t)$, es una constante real. Notemos que $C(0)$ (el valor de $C(t)$ cuando $t = 0$) representa la proporción de confinamiento al inicio del incremento consecutivo de infectados.

Por ejemplo, supongamos que el mandato gubernamental que obliga al confinamiento entra en vigor el vigésimo quinto día desde iniciado el incremento consecutivo de infectados y que ese mandato contempla:

(I) un 1 % de confinamiento al inicio del incremento consecutivo diario.

(II) un 50 % de confinamiento el día 25,

(III) un 90 % de confinamiento máximo.

Entonces, la función (1) tomaría la forma

$$C(t) = \frac{0{,}9}{1 + a\ \exp(b\,t)},$$

la cual debe cumplir $C(0) = 1/100$ y $C(25) = 50/100$, es decir,

$$\frac{0{,}9}{1 + a} = \frac{1}{100} \qquad \text{y} \qquad \frac{0{,}9}{1 + a\ \exp(25\,b)} = \frac{50}{100}. \tag{2}$$

Estas igualdades forman un sistema de dos ecuaciones en las incógnitas a y b, el cual puede ser resuelto fácilmente con operaciones elementales y propiedades de la función exponencial y su inversa: la función logaritmo natural, de donde obtenemos $a = 89$ y $b = -0{,}281$. Por lo tanto, la fórmula que describe la curva de la figura anterior es:

$$C(t) = \frac{0{,}9}{1 + 89\ \exp(-0{,}281\,t)}.$$

3. El modelo y su solución

Con lo anterior, nuestro modelo inicial básico se transforma en uno de la forma

$$\frac{dX}{dt} = k\,X\,\big((1 - C(t))\,N - X\big), \tag{3}$$

donde, como hemos mencionado, $C(t)$ es una curva que cambia con el tiempo.

Aquí conviene detenerse un momento para explicar algunas diferencias, desde el punto de vista matemático, entre este nuevo modelo y el original.

En el lenguaje de cálculo diferencial nuestro primer modelo:

$$\frac{dX}{dt} = k\,X\,(N - X)$$

es llamado una *ecuación diferencial autónoma* y el modelo (3) es llamado una *ecuación diferencial no autónoma*. La primera diferencia entre ellos es que el lado derecho del primer modelo no depende explícitamente del tiempo t y el lado derecho del modelo (3) sí, pues aparece $C(t)$. Esto da origen a los adjetivos autónoma y no autónoma, respectivamente.

En ambas ecuaciones diferenciales la incógnita X ya no es una variable que toma un valor concreto, como en el caso de las ecuaciones algebraicas, sino que ahora es una función que depende del tiempo: $X = X(t)$. Sin embargo, usualmente en la notación de una ecuación diferencial solo escribimos X y reservamos la notación $X(t)$ para expresar su solución.

Existen diversas técnicas que se pueden aplicar para resolver una ecuación diferencial, de acuerdo al tipo que tengamos. Por ejemplo, la ecuación de nuestro primer modelo, conocida como *ecuación diferencial logística (o de Verhulst)*, pertenece a la familia de *ecuaciones diferenciales de variables separables* y puede resolverse por medio de cálculo integral básico, esto es, podemos encontrar X explícitamente:

$$X(t) = \frac{N\,X_0}{X_0 + (N - X_0)\,\exp(-k\,N\,t)},$$

donde recordemos que X_0 es la cantidad inicial de infectados. Una aplicación de la ecuación diferencial logística al crecimiento de poblaciones se puede hallar en [3].

Una ventaja de tener la solución explícita es que podemos predecir la cantidad de infectados en cualquier instante de tiempo y no solamente cada hora. El lector interesado en profundizar en los diferentes tipos de ecuaciones diferenciales y técnicas para resolverlas puede consultar [1, 4].

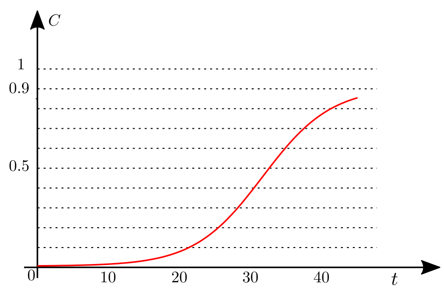
En contraste con el primer modelo, la solución del segundo no es posible encontrarla tan fácilmente, incluso en ocaciones es imposible de obtenerla explícitamente usando las funciones que generalmente se estudian en la educación superior. Sin embargo, ya desde el siglo XVIII el matemático Suizo Leonhard Euler (1707–1783) nos proporcionó una idea sencilla pero brillante para resolver numéricamente una ecuación diferencial. Entonces, con la ayuda de una computadora o una calculadora moderna se puede obtener el comportamiento de la solución de esta ecuación, es decir, podemos conocer (aproximadamente) el número de infectados, X, en cada instante, lo cual se representa como $X(t)$. Así, $X(35)$ representa la cantidad de infectados en el día 35.

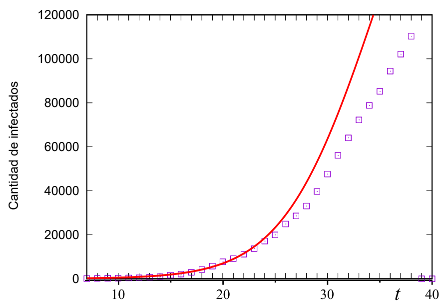
La forma de la curva de C dependerá de las medidas que tome la población o el gobierno. En concreto, depende del día que se decreta la cuarentena y de la cantidad de población que se pretende dejar confinada (que tan amplia es la medida).

En el caso de España, la medida de estado de alarma fue implementada el 14 de marzo de 2020 que corresponde al día 19 desde el incremento consecutivo de infectados. Sin embargo, debido a que los síntomas de infección pueden aparecer entre 10 y 14 días después del contagio, los efectos reales de la cuarentena no serán reflejados hasta 14 días después de implementada la medida, es decir, hasta el día 33. Además, es claro que el confinamiento máximo se alcanza de forma gradual (debido a que hay que hacer la compra, que hay trabajos pendientes de terminar, por apatía, etc.). Suponiendo que al inicio hay un 0,1 % de confinamiento, el dia 19 hay un 50 % de confinamiento y que el confinamiento máximo será del 90 %, entonces, al resolver el sistema (2) correspondiente, se obtiene la siguiente función

$$C(t) = \frac{0{,}9}{1 + 899\,\exp(-0{,}21286\,t)},$$

su gráfica y la comparación entre los valores oficiales hasta el día 38 (02 de abril de 2020) y los obtenidos a través del modelo son los mostrados en las figuras siguientes.



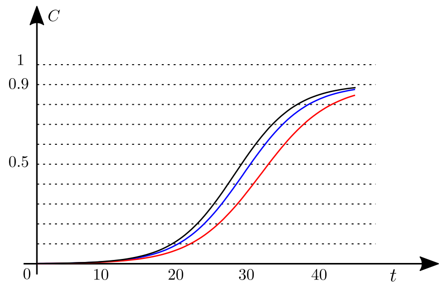


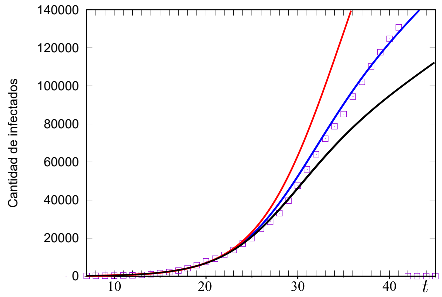
Notar que la curva de confinamiento dice que al momento de la medida (día 33 suponiendo que hasta 14 días después del 19 no hay efectos) el 50 % de la población se queda en casa y que no se logra que más del 85 % se confine hasta pasados 10 días (día 43). El modelo se ajusta bien hasta el día 24 desde que se inició el aumento consecutivo de infectados, el cual corresponde a la fecha 19 de marzo de 2020. Sin embargo, el modelo predice un mayor número de infectados de los que realmente existen (según los datos oficiales) a partir del día 25 (20 de marzo). Una posible explicación de esta discrepancia es que más gente de lo que se espera se esta confinando.

Ahora consideremos dos situaciones: primera, que al día de entrar en vigor la medida el 62 % se queda en casa y que se alcanza más del 80 % en siete días. Segunda, que al día de entrar en vigor la medida el 68 % se queda en casa y que se alcanza más del 80 % en seis días, entonces las funciones a usar son las siguientes:

$$C = C(t) = \frac{0{,}9}{1 + 899\,\exp(-0{,}2302\,t)} \qquad C = C(t) = \frac{0{,}9}{1 + 899\,\exp(-0{,}2403\,t)}$$

Las curvas de estas dos funciones son señaladas en azul y negro, respectivamente, en la figura de abajo en el lado izquierdo (la curva roja, es la curva del caso anterior). La comparación de los datos oficiales hasta el día 41 (5 de abril de 2020) y los dados por el modelo está en la figura del lado derecho.

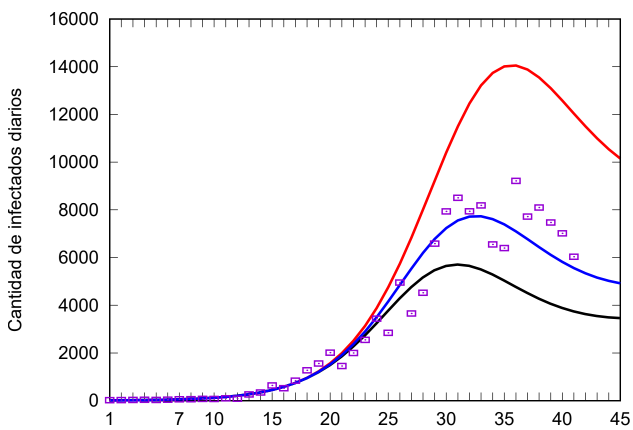




Notemos que las curvas de confinamiento $C(t)$ son muy parecidas, es decir, no hay mucha diferencia numérica entre ellas. Sin embargo, la figura en el lado derecho pone de manifiesto que, parafraseando a Edward Lorenz (1917–2008),

¡pequeños cambios pueden conducir a
consecuencias completamente
divergentes!

La comparación del número de infectados diarios entre los datos oficiales hasta el día 41 (5 de abril de 2020) y los dados por el modelo se presenta en la siguiente figura.



4. Más modelos en epidemiología

Para finalizar estas notas realizaré unas observaciones sobre el modelo planteado en la sección anterior y mencionaré la existencia de modelos más completos en temas epidemiológicos.

Respecto del modelo (3) presentado en la sección anterior cabe mencionar algunos aspectos importantes.

(i) El modelo no considera explícitamente los casos de personas recuperadas que quedan inmunes al virus y tampoco permite la posibilidad de que la tasa de contagio disminuya en el tiempo.

(ii) El modelo (3) con la curva de confinamiento azul es una propuesta (idealizada) que se ajusta a los datos durante un periodo de tiempo, como vimos en la última figura de la sección anterior, pero que no se ajustará (ojalá) para todo el tiempo futuro.

(iii) Una situación más realista respecto de la función $C(t)$ es que tenga cambios más bruscos, producto de nuevas medidas gubernamentales u otros factores.

Entendiendo que el modelo (3) es solamente una aproximación, surge la pregunta: ¿existen modelos que se ajusten más a la situación real? La respuesta es sí, hay una gran variedad de modelos más completos, esto es, que consideran más factores involucrados en el proceso epidemiológico. Uno de los más usados es el llamado modelo SIR. Las siglas de este modelo significan: S susceptible de contagio, I infectado y R removido (recuperado o fallecido). Una descripción y discusión de este modelo es realizada en [5]. Variaciones de este modelo son modelos SEIR y SITR [2] los cuales consideran más aspectos en el proceso de la epidemia. Para tener información sobre más modelos y su fundamentación consultar [1, 2].

Finalmente, es importante mencionar que actualmente existen equipos interdisciplinarios de investigadores alrededor del mundo trabajando en el planteamiento y desarrollo de modelos más sofisticados, que tengan en cuenta más factores y permitan obtener información para tomar medidas y rebajar el número de infectados diarios. Este es uno de los principales objetivos en el combate a la pandemia del Covid-19. Sería imposible nombrar a todos estos equipos. Sin embargo, a continuación dejo algunos sitios web que pueden ser consultados.

- https://www.uab.cat/web/noticies/detall-noticia-1202368197016.html?noticiaid=1345810211799
- https://images.math.cnrs.fr/Modelamiento-de-una-epidemia.html?lang=fr
- https://images.math.cnrs.fr/Modelamiento-de-una-epidemia-segunda-parte.html?lang=fr

- https://www.ucm.es/momat/covid-19
- http://www.cmm.uchile.cl/?p=37663
- https://spiral.imperial.ac.uk:8443/handle/10044/1/77482
- http://www.madrimasd.org/blogs/matematicas/2020/03/28/147534
- https://www.medrxiv.org/content/10.1101/2020.01.31.20019901v2.full.pdf
- https://www.medrxiv.org/content/10.1101/2020.02.16.20023754v2.full.pdf

Agradecimientos. Agradezco al revisor, así como a varios colegas matemáticos y a mi hijo por todas las sugerencias para mejorar estas notas.

Referencias

[1] M. Braun, *Differential Equations and Their Applications: An Introduction to Applied Mathematics*. Springer (1992).

[2] F. Brauer, C. Castillo-Chávez, *Mathematical models in population biology and epidemiology*. Springer; New York (2012).

[3] A. Gasull, J. Torregrosa, *Models de població del món i prediccions*. Materials Matemàtics 2009, treball no. 4, 20 pp.

[4] J. D. Logan, *A first course in differential equations*. Undergraduate Texts in Mathematics. Third edition. Springer (2015).

[5] H. Weiss, *The SIR model and the foundations of public health*. Materials Matemàtics 2013, treball no. 3, 17pp.

[6] https://elpais.com/politica/2020/01/08/actualidad/1578483165_970264.html

Departamento de Matemática,
Universidad del Bío-Bío,
Concepción, Chile.
srebollo@ubiobio.cl

Publicat el 14 d'abril de 2020
Versió corregida de la preliminar del dia 8 d'abril