

UN MODELO SENCILLO PARA REPRODUCIR LA PANDEMIA Covid-19

Enrique Castillo^{1,2}

¹ *Real Academia de Ingeniería, España*

¹ *Real Academia de Ciencias, España*

Resumen

Se presenta una colección de modelos sencillos para estudiar la propagación del Covid-19. Tras normalizar y discutir algunas propiedades que deben tener los modelos para ser válidos, se demuestra que las curvas de afectados (fallecidos o infectados) acumulados tienen que ser funciones de distribución. Se recomienda utilizar modelos que den el crecimiento de los fallecimientos en función del número inicial de éstos y del tiempo transcurrido, demostrando que estas funciones no pueden ser arbitrarias, sino que deben satisfacer la ecuación funcional de la traslación, que se resuelve y se interpreta adecuadamente. La solución de esta ecuación muestra que cualquier intervalo de la curva de afectados contiene información sobre toda ella, por lo que un intervalo suficientemente amplio de la misma permite predecir el futuro, o recuperar el pasado, de la misma. La aplicación al caso particular de la distribución normal da ajustes increíblemente buenos para los fallecimientos. Finalmente se dan las conclusiones más importantes del trabajo y algunas consideraciones sobre las cautelas a considerar para finalizar el confinamiento.

Palabras clave: Covid-19, modelos estadísticos para propagación de la pandemia, métodos bayesianos, ecuaciones funcionales, modelos de regresión robusta, recomendaciones y cautelas para dejar la cuarentena.

1. Introduction

En esta breve nota se presenta un modelo muy sencillo para reproducir la propagación del coronavirus según los datos oficiales aportados¹. Más precisamente, se busca una curva suficientemente regular $M = M(t)$, que de el número de personas afectadas (fallecidas o infectadas)² hasta el instante t . Como resultado relevante, se identifican las curvas a utilizar, que tras ser normalizadas con el número final de fallecidos, resultan ser funciones de distribución.

Para resolver un problema, es muy importante conocerlo bien y entender las hipótesis y fundamentos de los modelos a utilizar. En estos casos se han utilizado modelos como,

¹Los datos aportados exhiben una muy mala calidad y alteran de forma manifiesta los posibles modelos, no sólo por no saber exactamente qué miden estos datos, sino porque se sabe que las diferentes comunidades autónomas miden los fallecimientos de forma diferente y hay algunos retrasos que son sistemáticos, como los producidos los fines de semana.

²A partir de ahora hablaremos de fallecidos, pero entendiendo que todo se aplica también al caso de los infectados.

por ejemplo, los SIR o SEIR, en sus variadas versiones que incluyen modelos con y sin retrasos, rangos de edades, variaciones temporales de parámetros, etc., y otros muchos, que tratan de reproducir cómo se produce el contagio y la transmisión de la enfermedad. Esto es un soporte físico y médico para los modelos muy valioso. Por ello, el tipo de curvas que deben utilizarse está bastante bien estudiado y discutido, lo que nos sirve de base para nuestra propuesta.

En particular, para reproducir la misma dinámica de los modelos anteriores, nos interesan las funciones crecientes sigmoideas, es decir con un punto de inflexión en $t = t_m$, que corresponde al máximo de la función derivada $M'(t)$. Esto quiere decir que inicialmente $M'(t)$ crece y posteriormente decrece, a partir de t_m , que conduce a disminuir los contagios hasta que éstos se anulan (ya sea repentina o asintóticamente).

Para comenzar, es importante señalar que siempre debe usarse un modelo adimensional. Por ello, en nuestro caso, suponemos el modelo

$$M^* = \frac{M}{M_\infty} = M^*(t^*), \quad (1)$$

donde M_∞ es el número final de personas fallecidas durante todo el proceso, y $t^* = t/t_m$. Nótese que utilizamos para adimensionalizar los dos parámetros más relevantes, el número de fallecidos totales al final del proceso, M_∞ , y el instante, t_m , en el que se alcanza el máximo de la curva derivada (la diaria).

Nótese que utilizamos variables adimensionales, tal como recomienda el teorema II de Buckingham, y asteriscos para referirnos a ellas.

Como dicha curva debe ser creciente y el rango ya normalizado de $M^*(t^*)$ es el intervalo $[0, 1]$, la función $M^*(t^*)$, del número total de fallecidos en función del tiempo, es una función de distribución y su derivada, $M^{*'}(t^*)$, que da el número de fallecidos diarios, es la correspondiente función de densidad.

Por tanto, tenemos tantos modelos como funciones de distribución sigmoideas para poder elegir.

En particular, la función $M^*(t^*)$ la elegiremos de una familia paramétrica de distribuciones, que dependerá de un número p de parámetros, entre los que es conveniente elegir t_m como uno de ellos, a los que habrá que elegir el parámetro M_∞ , anteriormente indicado. Una vez elegida esta familia buscaremos una curva de la familia, estimando los parámetros, o, si queremos hacerlo mejor, buscaremos mezclas de ellas, tal como hacen los modelos bayesianos, para nuestra propuesta final³.

Casos particulares interesantes son las familias normal, gamma, logística, t de Student, Weibull, Gumbel y Frechet, estas tres últimas en sus versiones de máximos y mínimos, etc. Ello nos conduce a un máximo de 4 parámetros, en estos casos particulares. Hay que señalar que en el caso normal, la media μ es el parámetro asociado a t_m .

³Nótese la gran diferencia que hay entre elegir una curva de la familia y considerar cualquiera de las mezclas posibles entre ellas. Esto puede interpretarse como que cada persona o grupo responde a una distribución de la familia y que la distribución a posteriori nos indica qué pesos se asignan a cada una de ellas.

2. Algunas consideraciones previas

Para resolver nuestro problema podemos elegir la función $M^*(t^*)$ que da el número de fallecidos en función del tiempo t^* , o, mejor todavía, podemos buscar una familia de funciones, $M^*(t^*; M_0^*)$, que nos permitan proceder por etapas, es decir, que nos dé el número de fallecidos cuando se parte de un número inicial de fallecidos M_0^* y se incrementa el tiempo en t^* unidades.

Hay que indicar que la función $M(t^*; M_0^*)$ no puede ser arbitraria, sino que debe cumplir una condición importante de consistencia, que establece lo siguiente:

El número de fallecidos que se obtienen a partir de cualquier instante en el que hay ya M_0^ fallecidos tras $t_1^* + t_2^*$, debe dar lo mismo si se calculan éstos directamente, mediante $M^*(t_1^* + t_2^*; M_0^*)$, que si se calculan los fallecidos en t_1^* , y usándolos como valor inicial, se calculan los que resultan tras t_2^* , es decir, mediante $M^*(t_2^*; M^*(t_1^*; M_0^*))$.*

En otras palabras, la función $M^*(t^*; M_0^*)$ debe satisfacer la ecuación funcional:

$$M^*(t_1^* + t_2^*; M_0^*) = M^*(t_2^*; M^*(t_1^*; M_0^*)), \quad (2)$$

que es la conocida ecuación funcional de la traslación, cuya solución es⁴:

$$M^*(t^*; M_0^*) = \phi \left(t^* + \phi^{-1}(M_0^*) \right), \quad (3)$$

donde $\phi(t^*)$ es una función invertible (creciente) arbitraria y $\phi^{-1}(M_0^*)$ define una traslación para cada valor de M_0^* .

Por tanto, la familia buscada $M^*(t^*; M_0^*)$ puede generarse a partir de la función $\phi(t^*)$ incluyendo todas sus posibles traslaciones, es decir, esta familia debe ser estable frente a traslaciones.

Además, en la ecuación (3) comprobamos, como debe ser, que

$$M^*(0; M_0^*) = \phi \left(\phi^{-1}(M_0^*) \right) = M_0^*, \quad (4)$$

y

$$M^*(0; 0) = \phi \left(\phi^{-1}(0) \right) = 0. \quad (5)$$

Es interesante ver que los grados de libertad que se tenían inicialmente para elegir $M^*(t^*; M_0^*)$, es decir, una función de dos argumentos, quedan reducidos severamente a elegir una función de un único argumento e invertible, si se quiere que se cumpla la condición de compatibilidad, dejando claro que, de no elegir esta función así, los resultados finales dependerán de los instantes elegidos para hacer la propagación.

Puesto que toda función de la forma (3) cumple la condición de compatibilidad (2), cabe preguntarse si dada una función de éstas, existirá más de una función ϕ que la reproduzca, es decir, podemos plantearnos el problema de la unicidad.

⁴Esto es cierto bajo condiciones muy débiles de continuidad puntual de la función.

Theorem 1 (Teorema de unicidad) Si existen dos funciones ϕ_1 y ϕ_2 que generan la misma función, es decir, satisfacen la condición:

$$\phi_1(t^* + \phi_1^{-1}(M_0^*)) = \phi_2(t^* + \phi_2^{-1}(M_0^*)), \quad (6)$$

entonces, debe ser

$$\phi_1(t^*) = \phi_2(t^* - \mu^*), \quad (7)$$

donde μ^* es una constante arbitraria.

La ecuación (6) es una ecuación funcional con dos funciones incógnitas, ϕ_1 y ϕ_2 , cuya solución es (7), lo que quiere decir que la solución es única, salvo una traslación.

Una ilustración gráfica de este resultado se presenta en la figura 1, donde se ve que partiendo de M_0^* , a la izquierda, obtenemos $\phi_1^{-1}(M_0^*)$, utilizando la curva roja, le sumamos t^* y luego, usando la curva roja de nuevo, se obtiene $\phi_1(t^* + \phi_1^{-1}(M_0^*))$. Se puede ver, que haciendo lo mismo con la curva azul, se obtiene el mismo resultado, por ser la curva azul una traslación de la curva roja anterior.

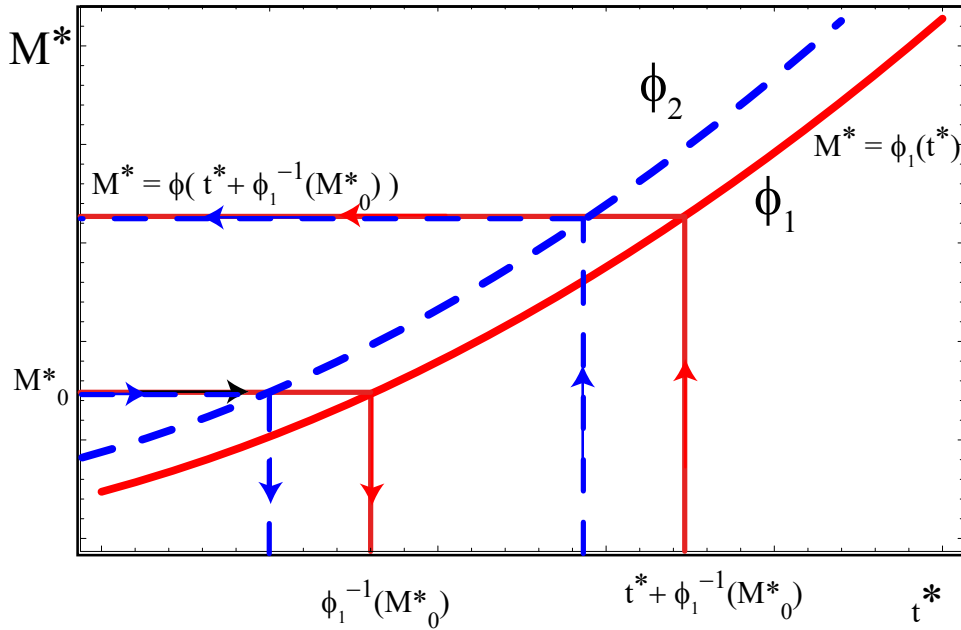


Figura 1: Ilustración gráfica del teorema de unicidad.

Otra pregunta interesante que podemos hacernos es: ¿qué es la función $\phi()$ y cómo se elige en un caso práctico? La respuesta a esta pregunta es la siguiente.

De la ecuación (3) resulta

$$M_0^*(t^*) = M^*(t^*; 0) = \phi(t^* + \phi^{-1}(0)), \quad (8)$$

donde se ha denotado $M_0^*(t^*)$ a $M^*(t^*; 0)$, lo que prueba que $M_0^*(t^*) = \phi(t^*)$ si $\phi^{-1}(0) = 0$, es decir que la función $\phi(t^*)$ es una de las funciones de la familia $M^*(t^*; M_0^*)$.

Por tanto, la familia buscada $M^*(t^*; M_0^*)$ puede generarse a partir de un miembro cualquiera de la familia, por ejemplo $\phi(t^*)$ incluyendo todas sus posibles traslaciones.

3. El ADN del contagio está contenido en cualquier intervalo de la curva

La relevancia de la ecuación de la traslación (2) va mucho más allá de lo que parece inicialmente. La expresión (3) nos da $M^*(t^*; M_0^*)$, mostrando que la dependencia de M_0^* queda definida por la función $\phi^{-1}(M_0^*)$, y que la dependencia de t^* es la dada en (3), es decir, perfectamente conocida en todos los casos de compatibilidad.

De hecho, el ADN de la variación temporal de las curvas de crecimiento de los fallecimientos está en cualquier intervalo temporal de las mismas. Quiere esto decir que conocida $M^*(t^*; M_0^*)$ en cualquier intervalo de t^* , puede conocerse ya la variación temporal completa. Sin embargo, dado que la información local no es suficientemente precisa, hay que utilizar un intervalo temporal finito suficientemente amplio de la misma, pero la idea sigue siendo válida, es decir, se puede predecir el comportamiento futuro de la misma a partir del pasado, y viceversa.

Además, la ventaja de usar $M^*(t^*; M_0^*)$ en vez de $M^*(t^*)$ es que podemos elegir cualquier valor inicial M_0^* , ya que el modelo es independiente de M_0^* . De esta forma, podemos evitar los primeros datos de la propagación, ya que en el origen de la epidemia, los datos presentan una gran variabilidad. Una vez iniciado el proceso de contagio, con valores más altos de M_0^* , al haber más personas que contagian, se estabiliza el proceso y la estimación es más robusta.

Si elegimos una familia de distribuciones, es decir, si se usa $M^*(t^*; M_0^*, \theta^*)$, donde θ^* es el vector de sus parámetros, puede utilizarse un intervalo temporal, que implica un intervalo de observaciones de $M^*(t^*; M_0^*)$, para estimar éstos y así predecir el futuro y, si se desea, recuperar el pasado. Obviamente, las estimaciones serán más precisas cuanto más cercano esté este intervalo del instante que se desea predecir, y cuanto más amplio sea éste.

Como se ha indicado, la necesidad de un intervalo temporal amplio se debe a la imposibilidad de obtener datos suficientemente precisos.

La explicación a todo esto está en que el modelo elegido, es decir, el hecho de que sean funciones de distribución tras la normalización, junto con las familias concretas de distribuciones elegidas incluyen ya la dinámica del sistema de la pandemia. Esto es lo mismo que ocurre con el médico, que, al examinar al enfermo y diagnosticar la enfermedad, es capaz de predecir con anticipación las etapas que seguirá el paciente, de forma tanto más precisa cuanto más cercana esté el instante a predecir.

Todo esto pone de manifiesto la profundidad y belleza de las ecuaciones funcionales y los modelos matemáticos, que son capaces de reproducir muchos aspectos de la realidad

que parecen inverosímiles⁵.

4. Modelo propuesto

Una vez aclaradas las ideas anteriores, proponemos, como simple ejemplo, la familia normal de distribuciones para nuestro modelo, que resulta ser:

$$M^*(t^*) = \Phi\left(\frac{t^* - \mu^*}{\sigma^*}\right), \quad (9)$$

donde $\Phi()$ es la función de distribución de la normal $N(0, 1)$, es decir, nuestros parámetros son $M_\infty^*, \mu^* = \mu/t_m$ y $\sigma^* = \sigma/t_m$, donde las letras griegas μ y σ representan la medias y las desviaciones típicas, respectivamente

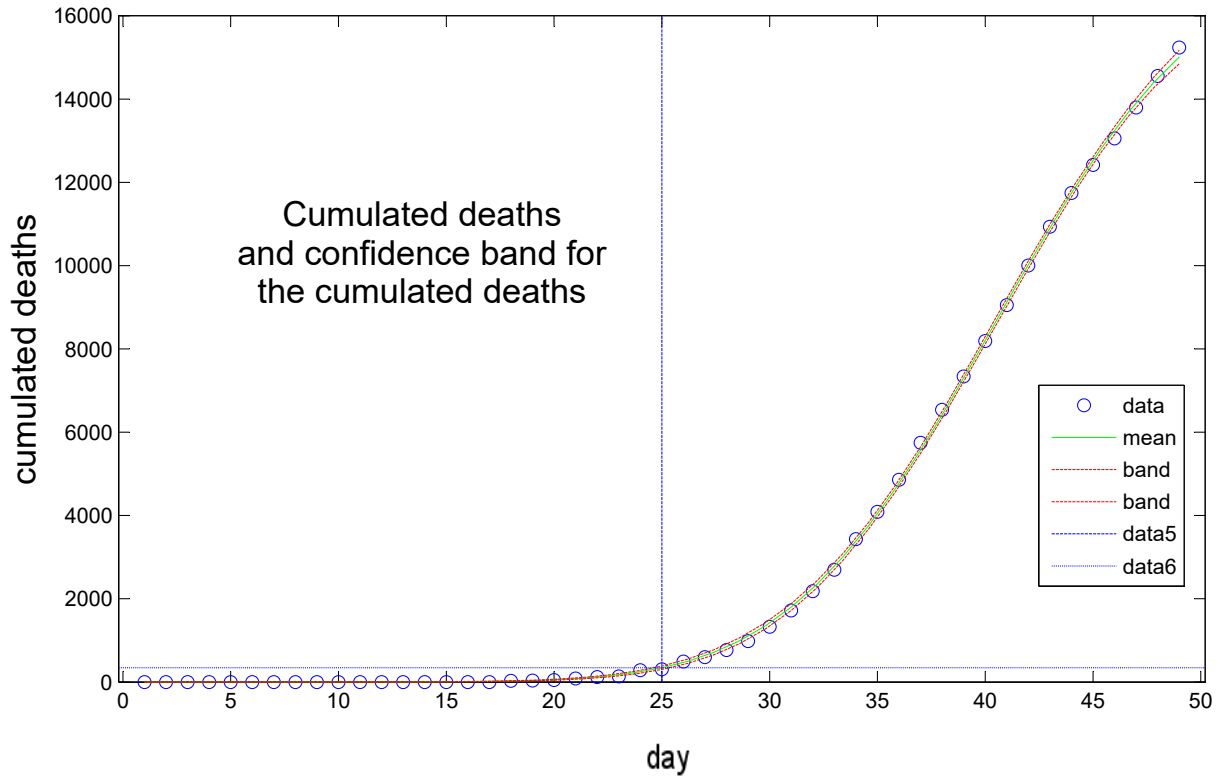


Figura 2: Muertes acumuladas y banda de confianza 0.95 para la media de la curva.

⁵Este comportamiento ya lo habíamos observado en otros modelos, especialmente en los modelos de fatiga, en los que llevamos muchos años trabajando. El ADN del comportamiento a fatiga de un material, es decir, el progreso de las grietas lo llevan éstas marcado antes de que las ensayemos, y, de hecho, no hace falta llegar a su rotura para predecir cuando ocurrirá ésta. Y no sólo esto, sino que podemos adivinar cómo fue el proceso previo de formación de las grietas, incluso cuando no era observable.

Si estimamos los parámetros mediante este modelo de regresión, utilizando matlab y su función “nlinfit” para regresión robusta, y la herramienta “nlintool”, se obtienen los modelos de las figuras 2 y 3 y el de la figura 4, si ajustamos la función de distribución y la de densidad, respectivamente.

El ajuste es asombroso para la curva acumulada y bastante bueno para la función de densidad⁶

La diferente anchura de las bandas es debida a que una de ellas se refiere a la predicción de la media y la otra, a la de datos aislados de cada día, siendo la predicción de la primera mucho más fácil (más precisa) que la de la segunda⁷.

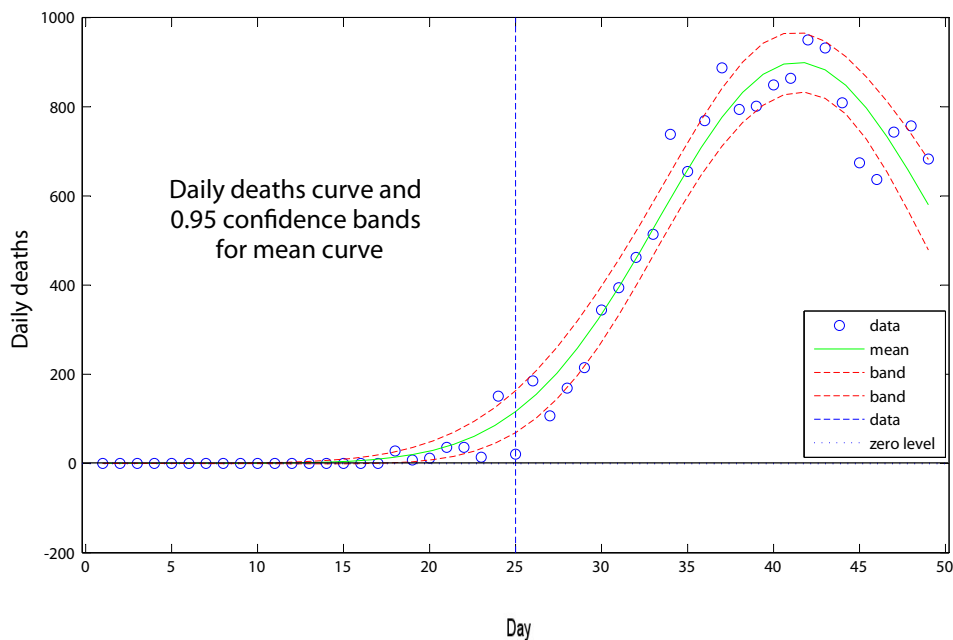


Figura 3: Muertes diarias y banda de confianza 0.95 para la media de la curva.

Finalmente habría que señalar que este modelo normal implica una simetría en las curvas de propagación a los dos lados del máximo de la curva de fallecimientos diarios, lo que puede que corresponda a la realidad. Por tanto, pueden utilizarse los otros modelos indicados o incluso otros diferentes, en su caso.

⁶Me comunican Roberto Mínguez y Javier Girón que la distribución logística también da muy buenos resultados, y Javier Samper que este modelo normal se ajusta muy bien tanto a fallecidos como a contagiados.

⁷Los datos diarios están claramente alterados, especialmente los fines de semana, y podrían corregirse en una etapa de preproceso.

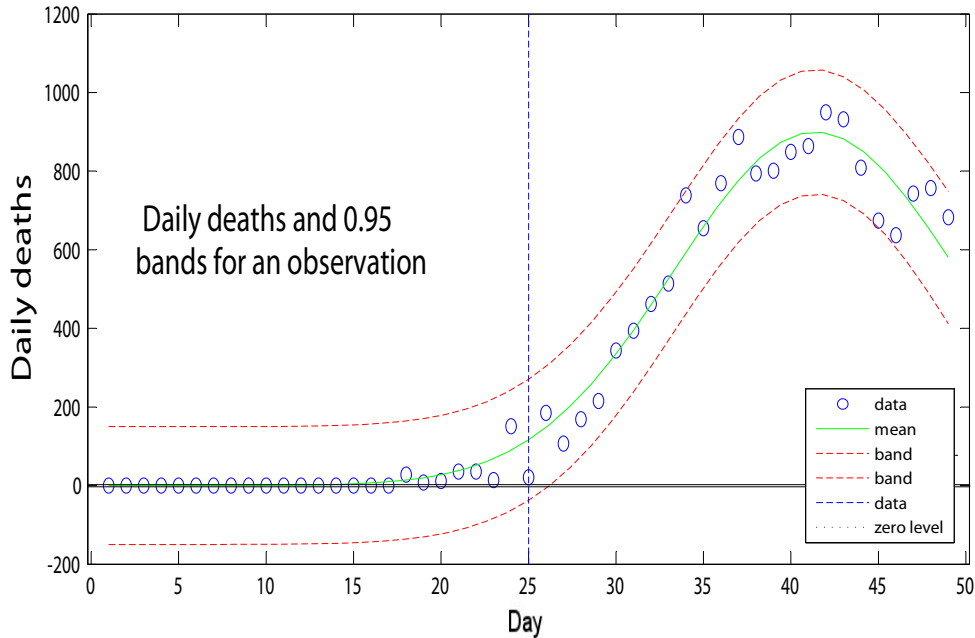


Figura 4: Muertes diarias y banda de confianza 0.95 para una observación.

Para terminar, sería muy interesante utilizar estos modelos considerados como bayesianos con muestreo MCMC, como los utilizados por WinBUGS u OpenBUGS. Desgraciadamente, no he podido ajustar el modelo que esperaba, debido posiblemente a deficiencias en la implementación de estos métodos en el software libre o a la mala calidad de los datos, que afectan al resultado.

Finalmente indicar que los métodos bayesianos, al considerar los parámetros aleatorios incluyen como curvas candidatas a todas las combinaciones lineales convexas de las de la familia elegida. Además, si la muestra es grande, los resultados son prácticamente independientes de las distribuciones “a priori” elegidas.

5. Conclusiones

Las conclusiones más importantes que pueden derivarse de lo anterior son las siguientes:

1. Es importante utilizar siempre modelos adimensionales para resolver los problemas, lo que sigue siendo válido en este caso de propagación de la epidemia.
2. Dado el carácter aleatorio del problema, aunque hay muchos métodos deterministas interesantes para estudiarlo y derivar su dinámica, éstos son claramente insuficientes y es necesario utilizar métodos estocásticos o aleatorios.

3. Para el caso que nos ocupa, deben utilizarse modelos crecientes sigmoidales con un punto de inflexión, que marca el esperado y deseado máximo de la curva.
4. Hay que tener cuidado de elegir modelos que permitan actualizar el número acumulado de fallecimientos a partir de cualquier instante, para lo que no puede elegirse una función arbitraria. La ecuación funcional de la traslación nos aclara cuáles son válidas y cuáles no lo son.
5. Las funciones ϕ , que generan estas funciones válidas y se explica lo que son y representan, deben ser invertibles.
6. A cualquier función creciente que cubra el rango total, basta añadirle un parámetro de localización (traslación) para obtener una familia válida para esta actualización indicada, si bien pueden y deben añadirse más parámetros.
7. Con la adimensionalización indicada, las curvas resultantes son funciones de distribución, por lo que, dependiendo de la tendencia, podremos conocer cuáles pueden ser las óptimas en cada caso.
8. La variación temporal de la curva está contenida en cualquier intervalo de la misma y puede ser calculada a partir de un intervalo finito suficientemente amplio de la misma. Ello permite predecir el futuro y recuperar el pasado, aunque éste sea no observable.
9. Los métodos bayesianos consideran no sólo las curvas de la familia, sino combinaciones lineales convexas de ellas (mixturas), lo que implica modelos totalmente generales.
10. Dada la falta de calidad y las deficiencias de los datos disponibles, todo lo anterior es sólo una ayuda en la búsqueda de soluciones para salir del estado de confinamiento, no teniendo sentido mejorar los métodos utilizados con datos tan deficientes y criterios tan cambiantes.
11. Indicar que el máximo de la curva derivada corresponde a la primera fase del proceso, existiendo una segunda fase, que puede ser más larga que la anterior. Por ello, considero muy arriesgado empezar tan pronto a abandonar el confinamiento. Un error en este sentido tendría consecuencias catastróficas, pues al daño de tener que repetir el proceso anterior, se añadirían las consecuencias psicológicas de la misma.
12. Finalmente, dado que la letalidad del virus es casi nula en personas jóvenes sanas, éstas deberían ser las primeras en salir, pero sólo si no entran en contacto con personas de alta edad y/o riesgo. Los gobiernos no deberían olvidar que para decidir quiénes abandonan el confinamiento, no basta con analizar los riesgos del transporte y del centro de trabajo de los trabajadores que salen, sino que hay que considerar, principalmente, el riesgo y la edad de las personas con las que conviven. Especial

consideración deben tener las residencias de ancianos, a las que, en casi todos los casos, se les ha llevado el virus desde fuera de ellas.