

EVOLUCIÓN DE LAS EPIDEMIAS: LA MATEMÁTICA DE AISLARSE

HÉCTOR PASTÉN, JORGE CASTILLO SEPÚLVEDA

¡Por favor lea al menos la sección 1 de este artículo! Prometemos que es corta, fácil de leer y tiene muy poca matemática.

Resumen. Este es un artículo sobre los aspectos matemáticos de la epidemia COVID19 en Chile, *orientado a un público general, no solo especialistas*. Explicaremos cómo los científicos logramos predecir el avance de la epidemia usando modelos matemáticos. En favor de la claridad de exposición, nos permitiremos usar la palabra “teorema” de un modo mucho más liberal que en un artículo de matemáticas puras y presentaremos modelos que solo necesitan matemática de liceo, evitando el uso de ecuaciones diferenciales.

- La sección 1 contiene algo de matemática (¡muy sencilla!), pero incluso si usted no se lleva bien con los números, leer la sección 1 es un ejercicio que por responsabilidad debería hacer. Se presenta el **método más sencillo de proyectar el avance de la epidemia** con algunas proyecciones de lo que podría venir para Chile en Abril.
- La sección 2 contiene una versión más refinada del análisis elemental de la sección 1. Aquí se presenta el **modelo exponencial** de crecimiento de epidemias. La evidencia indica que actualmente el ritmo de contagio es cercano a un caso nuevo al día por cada 5 contagiados.
- La sección 3 presenta el **modelo SIR** y una versión simplificada del mismo. Este modelo para predecir el comportamiento de epidemias es más realista y da predicciones más confiables a mayor plazo que el modelo exponencial.
- La sección 4 usa datos reales en el **modelo SIR simplificado** para predecir cómo evolucionará el COVID19 en Chile si continuamos al mismo ritmo que hasta ahora. A este ritmo, la predicción es un primer peak de la enfermedad entre fines de Mayo y comienzos de Junio. El número de casos se saldría de control mucho antes. Además, explicaremos por qué **modelos como el SIR tienen crecimiento exponencial en etapas iniciales**.
- La sección 5 examina las **proyecciones si bajáramos la frecuencia de contagios con una estrategia drástica de aislamiento**. Esto supone una acción inmediata: cada día se empeora drásticamente la situación debido a que en esta etapa temprana de la epidemia el modelo de crecimiento exponencial todavía es válido.

Los modelos matemáticos en este artículo son mucho más simples que los modelos que se necesitarían para hacer predicciones más certeras. Además, la calidad de los datos existentes no es óptima; está sujeta a las limitaciones de la capacidad de diagnóstico en Chile bajo la contingencia. Pero aún así, esperamos que este artículo informe al público general sobre cómo se estudia el avance de una epidemia como el COVID19.

1. ¿CÓMO SE PRONOSTICA EL AVANCE DE UNA EPIDEMIA?

Las enfermedades altamente contagiosas se expanden de una manera relativamente predecible en sus primeras etapas si no se toman fuertes medidas para controlarlas. Por ejemplo, invitamos a la lectora a reflexionar sobre la siguiente observación aproximada: si un enfermo promedio contagia a alguna otra persona cada 4 días, entonces el total de casos ocurridos se *duplicaría* cada 4 días.

En el caso concreto del COVID19 en Chile, la siguiente tabla tiene el número total de casos confirmados por el MINSAL hasta hoy, 29 de Marzo^[1], ver [3]. En la tabla, n es el número de días contando desde el primer caso, y la cantidad $T(n)$ es el número de casos *oficial* total hasta ese día

Date: 30 de marzo de 2020.

¹Este artículo fue terminado el Domingo 29 de Marzo de 2020 en la tarde y se usaron los datos disponibles hasta ese momento. El proceso de edición (revisar errores de escritura, producir gráficos, etc.) nos tomó hasta el Lunes 30

(el número *real* de infectados es obviamente mayor debido a que en muchos casos la infección no presenta síntomas inmediatos, y el afectado no siempre es sometido al test).

fecha	n	$T(n)$ oficial	fecha	n	$T(n)$ oficial	fecha	n	$T(n)$ oficial
3 Marzo	1	1	12 Marzo	10	33	21 Marzo	19	537
4 Marzo	2	3	13 Marzo	11	43	22 Marzo	20	632
5 Marzo	3	4	14 Marzo	12	61	23 Marzo	21	746
6 Marzo	4	5	15 Marzo	13	75	24 Marzo	22	922
7 Marzo	5	7	16 Marzo	14	156	25 Marzo	23	1142
8 Marzo	6	11	17 Marzo	15	201	26 Marzo	24	1306
9 Marzo	7	13	18 Marzo	16	238	27 Marzo	25	1610
10 Marzo	8	17	19 Marzo	17	342	28 Marzo	26	1909
11 Marzo	9	23	20 Marzo	18	434	29 Marzo	27	2139

Pareciera ser que el número de contagiados se duplica cada 3 o 4 días. Eso quiere decir, aproximadamente, que cada enfermo contagia a alguna otra persona cada 3 o 4 días².

Básicamente, si hoy (29 de Marzo) tenemos 2139 contagiados *reportados* (¡los contagiados reales no diagnosticados podrían ser muchos más!), entonces en 4 días más (el 2 de Abril) debería haber unos $2 \times 2139 = 4278$ contagiados. Alguien podría decir que no es un número tan grande para un país con cerca de 19 millones de habitantes, pero ¿qué pasaría si la situación se mantuviera a este ritmo, digamos, por un mes? Los números se verían así:

día	número pronosticado de casos (suponiendo duplicación cada 4 días)
2 Abril	4278
6 Abril	8556
10 Abril	17112
14 Abril	34224
18 Abril	68448
22 Abril	136896
26 Abril	273792
30 Abril	547584

Esto es *más de medio millón de personas* hacia el final de Abril.

¿Qué podemos hacer para evitar un avance tan rápido? Como no disponemos de una cura en este momento, lo mejor que podemos hacer es bajar el ritmo de propagación. Por ejemplo, si cada enfermo contagiara (en promedio) a una nueva persona una vez cada 8 días en lugar de una vez cada 4 días, los números serían así:

día	número esperado de casos diagnosticados (suponiendo duplicación cada 8 días)
6 Abril	4278
14 Abril	8556
22 Abril	17112
30 Abril	34224

Esto es *cerca de 34 mil personas* hacia el final de Abril.

El cerebro humano no está diseñado para comprender números demasiado grandes. Así que para dimensionar mejor la diferencia entre ambos escenarios, piense usted que prefiere: una deuda por medio millón de pesos, o una deuda por 34 mil pesos.

de Marzo. A esa fecha había una nueva actualización de datos, pero no la usamos pues de lo contrario la edición del artículo nos habría tomado hasta el día siguiente, y el desfase se repite.

²Las lectoras más audaces podrían reclamar que esto no toma en cuenta el hecho que hay enfermos recuperándose y que dejan de ser contagiosos. Eso es verdad, y por este motivo más adelante vamos a tomar ese factor en cuenta para presentar un modelo un poco mejor diseñado para aproximar la realidad.

El problema es que nadie sabe con certeza si porta el virus o no, salvo que comience a demostrar síntomas o se haga el examen. Los síntomas no son inmediatos: un portador puede pasar hasta 14 días sin síntomas [4], y por ende, sin saber que está contagiando a otros.

Por eso es un desafío diseñar una estrategia que permita hacer más lenta la velocidad en que los portadores contagian a una nueva persona.

- **Conclusión importante: Hacer que la velocidad de contagio por persona sea solo unos cuantos días más lenta provocaría una diferencia dramática en los números.**
- Reduciendo el contacto entre personas se evita que los infectados contagien a otros, y recíprocamente los sanos no se exponen a ser contagiados.
- Incluso si alguien sano asiste a un evento social y toma todas las medidas de higiene, el asistir al evento fomenta que más gente asista, y esas otras personas podrían dar origen a contagios nuevos.

Hay quienes no pueden aislarse por diversos motivos de fuerza mayor. Pero si usted *puede* quedarse en aislamiento, entonces *tiene la obligación moral* de hacerlo.

2. EL MODELO EXPONENCIAL

Un modelo matemático es una forma de utilizar las matemáticas para tratar de predecir el futuro a partir de datos conocidos.

Por naturaleza, los modelos matemáticos tienen algunas falencias, por ejemplo:

- Simplifican la realidad para que el modelo sea calculable. Esto hace que algunos factores no se tomen en cuenta.
- La vida real es incierta y hay un poco de azar.
- El modelo matemático toma datos iniciales para hacer una predicción. La calidad de la predicción no es mejor que la calidad de los datos de partida. Es decir, si los datos de partida son poco confiables, la predicción será poco confiable.

Sin embargo, los modelos matemáticos bien diseñados y que toman en cuenta sus falencias al analizar una predicción, son una valiosa herramienta en la toma de decisiones.

Lo que hicimos en la sección anterior es un modelo matemático sencillo para tratar de predecir el avance de una epidemia. Este modelo se llama *modelo exponencial*: *En un periodo de tiempo determinado, la cantidad de contagiados se va multiplicando por una constante fija*. En la sección anterior consideramos una versión donde el “periodo de tiempo” eran 4 días, y otra donde eran 8 días. Para simplificar las cosas, podemos estandarizar el periodo de tiempo. Por ejemplo 1 día:

Modelo exponencial, avance diario: Cada día, la cantidad de contagiados se va multiplicando por una constante C fija.

El razonamiento en que se basa este modelo es el siguiente:

Razonamiento del modelo exponencial: Supongamos que cada infectado tiene en promedio k contactos contagiosos al día y que ayer teníamos T_{ayer} infectados. Entonces cada infectado de ayer “aporta” con k infectados nuevos para hoy, en promedio, y eso nos da $k \cdot T_{ayer}$ infectados nuevos para hoy. Así que hoy tendremos el total de infectados de ayer (es decir T_{ayer}) más los nuevos (es decir, $k \cdot T_{ayer}$), dando un total de

$$T_{hoy} = T_{ayer} + k \cdot T_{ayer} = (1 + k) \cdot T_{ayer}.$$

Esto da el modelo exponencial con $C = 1 + k$. En símbolos:

Teorema 2.1 (Fórmula del modelo exponencial). *Sea $T(n)$ el número total de casos al día n de nuestro estudio, según el modelo exponencial. Sea k el número promedio de interacciones contagiosas que un infectado tiene al día. Definimos $C = 1+k$. Entonces, según el modelo exponencial, se cumple que $T(n) = C \cdot T(n - 1)$.*

El $T(n)$ del modelo exponencial es un concepto matemático; no es el número *real* de infectados. Pero el punto es que eligiendo adecuadamente k y C según los datos existentes, la secuencia $T(1), T(2), T(3), \dots$ debería aproximar el número de infectados y esto nos permite hacer predicciones.

Para calcular cuál es el valor de C que más se acerca a la realidad en el caso del COVID19 en Chile, podemos usar los datos disponibles del MINSAL a la fecha. De la tabla de la Sección 1, utilizaremos los datos oficiales de la última semana para tratar de ajustarnos a la tendencia de crecimiento más reciente (los datos de los primeros días no necesariamente reflejan la tendencia predominante ahora):

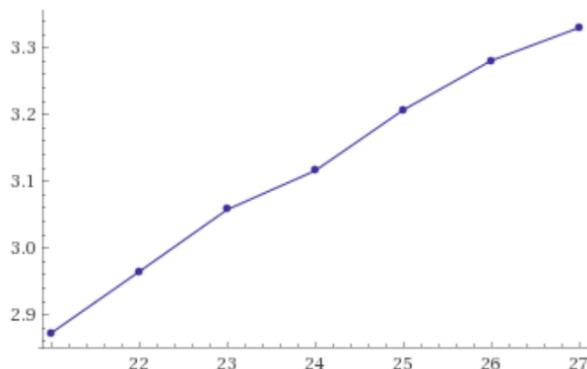
fecha	n	$T(n)$ oficial
23 Marzo	21	746
24 Marzo	22	922
25 Marzo	23	1142
26 Marzo	24	1306
27 Marzo	25	1610
28 Marzo	26	1909
29 Marzo	27	2139

Mirando estos números es difícil adivinar un valor de C . Aquí vienen los logaritmos y la ecuación de la recta a ayudarnos³. Tomando \log_{10} en la fórmula del modelo exponencial se obtiene

$$\log_{10} T(n) = \log_{10} T(n-1) + \log_{10} C.$$

Es decir, cada día la cantidad $\log_{10} T(n)$ aumenta *sumando* la constante $\log_{10} C$. Esto quiere decir que la función $f(x) = \log_{10} T(x)$ según el modelo, es una recta de pendiente $\log_{10} C$.

Hacemos un gráfico de las cantidades $f(n) = \log_{10} T(n)$ según los datos oficiales de la última semana⁴:



¡La tendencia en forma de recta es evidente a la vista! Esto confirma que el modelo exponencial se ajusta a la realidad en esta etapa temprana de la epidemia. Calculando la recta que mejor se ajusta⁵ se obtiene

$$f(x) = 0,0769188x + 1,27237, \quad \text{pendiente} = 0,0769188.$$

El gráfico de esta recta sobre los $\log_{10} T(n)$ de la última semana está en la página siguiente.

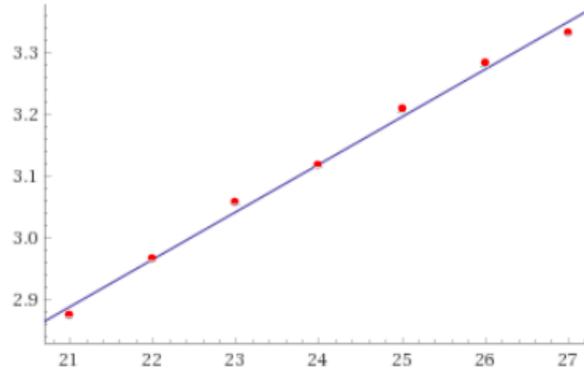
La pendiente es $\log_{10}(C) = 0,0769188$, así que

$$C = 10^{0,0769188} \approx 1,194 \quad \text{y por ende} \quad k = C - 1 = 0,194$$

³Así es; la matemática que se ve en el liceo sirve.

⁴Gráfico hecho con Wolfram|Alpha

⁵Para calcular la ecuación de la recta que mejor se ajusta a los datos, utilicé una herramienta matemática llamada "ajuste por mínimos cuadrados".



Conclusiones del modelo exponencial. El modelo exponencial nos dice que según la tendencia actual:

- Una persona infectada contagia en promedio a $k = 0,194$ otras personas diariamente. En otras palabras, esto significa que aproximadamente por cada 5 personas infectadas, se contagia una nueva persona al día.
- Cada 4 días, la cantidad de infectados se multiplica por $C^4 = 1,194^4 \approx 2,032$. Esto se ajusta bastante bien al análisis más elemental que hicimos en la Sección 1.
- A este ritmo, podemos proyectar que el 30 de Abril (32 días desde hoy) el total de infectados debería ser

$$2139 \cdot C^{32} = 2139 \cdot 1,194^{32} \approx 622800.$$

Esto es bastante más que medio millón de personas.

Estos números son predicciones suponiendo que se sigue la tendencia actual y suponiendo que el modelo exponencial sigue siendo válido. En realidad, los casos *oficialmente confirmados* podrían ser significativamente menos, por ejemplo:

- Si se toman excelentes medidas de aislamiento para bajar el valor de k y así frenar la epidemia (los números oficiales serían más chicos por menos contagios). Esto sería lo ideal.
- Si no se toman suficientes tests para detectar contagios (los números oficiales serían más chicos por menos diagnósticos). Esto sería una falsa ilusión: números chicos que no reflejarían la realidad.
- Si el modelo exponencial deja de ser válido. Esto ocurre cuando pasa mucho tiempo desde el inicio de la epidemia. En la siguiente sección explicaremos cómo dar un mejor modelo matemático que toma en cuenta una serie de otros factores y que da mejores predicciones a mediano plazo.

3. UN MEJOR MODELO: SIR

El modelo exponencial funciona muy bien para analizar etapas iniciales de una epidemia de propagación rápida, pero hay algunas fallas en su diseño:

- No se toma en cuenta que después de un cierto tiempo, los enfermos se recuperan (o lamentablemente, fallecen) y dejan de ser contagiosos (al menos por un tiempo).
- No se toma en cuenta que mientras más infectados hay, menos personas sanas quedan susceptibles de ser contagiadas.

Un modelo matemático mejor diseñado que toma en cuenta estos problemas es el llamado *modelo SIR*, desarrollado en 1927 [1] y que hasta el día de hoy es la base para otros modelos más sofisticados en el estudio de epidemias. La sigla viene de catalogar los miembros de una comunidad (digamos, un país) en tres tipos

- Susceptibles: Personas sanas que no son inmunes. Es decir, son susceptibles a ser infectadas.

- Infectados: Personas portadoras de la enfermedad y que pueden contagiar a otros
- Retirados: Personas que estuvieron enfermas y que se recuperaron (o que lamentablemente fallecieron), y por ende, dejan de ser contagiosas.

El modelo SIR en su forma más sencilla es un sistema de tres ecuaciones diferenciales relacionando la cantidad de personas en estos tres grupos y cómo varían a través del tiempo.

No vamos a escribir las ecuaciones diferenciales aquí⁶ porque no es nuestra intención asustar a nadie; ya tenemos suficiente con el COVID19. En lugar de eso, vamos a explicar un modelo muy parecido al SIR pero que no necesita nada avanzado. ¡Este modelo incluso será entendible para quienes cursen enseñanza media con entusiasmo por las matemáticas!

Vamos a tratar de entender cómo avanza la infección por COVID19 en una comunidad con H habitantes. Por ejemplo, si la comunidad que vamos a estudiar es Chile, entonces H es aproximadamente 19 millones. Vamos a suponer que a lo largo del estudio H es constante: esto es aproximadamente cierto para ciudades grandes (o países) en un periodo de tiempo moderado.

Separamos los miembros de la comunidad en 3 grupos como en el modelo SIR. La cantidad de personas en cada uno de estos grupos el día n de nuestro estudio se va a escribir $S(n)$, $I(n)$ y $R(n)$ respectivamente. El total de habitantes es la suma de estas cantidades

$$H = S(n) + I(n) + R(n)$$

porque cada persona está en exactamente uno de los tres grupos.

Ejemplo hipotético. Por ejemplo, si el tercer día de nuestro estudio tenemos 100 sanos, 30 enfermos contagiosos asintomáticos, 3 enfermos con síntomas, 10 recuperados y 2 fallecidos, entonces $S(3) = 100$, $I(3) = 33$ y $R(3) = 12$. El total de habitantes de esta comunidad (contando los fallecidos) es $H = 100 + 33 + 12 = 145$.

Volviendo al caso general, hay dos cantidades muy importantes relacionadas con la enfermedad:

- d = el número de días en promedio que dura la infección, ya sea que termine en recuperación o fallecimiento.
- k = el número promedio de contactos infecciosos al día que un portador tiene con otras personas (sanas o no). Un “contacto infeccioso” es una interacción en la que una persona sana resultaría contagiada. Por ejemplo, si un infectado se afirma de una puerta del metro dejando el virus, y otra persona toca la puerta con su guante y luego se rasca un ojo sin notarlo.

Observación importante sobre d . Los infectados *nuevos* en un determinado día demoran en promedio d días en pasar al grupo “R”. En símbolos:

$$R(n) - R(n - 1) = I_{\text{nuevos}}(n - d)$$

donde $I_{\text{nuevos}}(n)$ es la cantidad de infectados nuevos el día n .

¿Cómo evoluciona la epidemia? Los infectados nuevos de hoy son la cantidad de infectados de ayer por la cantidad de personas sanas que cada uno de ellos infectó en promedio. Aquí es donde hay una diferencia importante con el modelo exponencial: ¡no toda interacción contagiosa termina en un contagio; solamente las interacciones contagiosas con personas sanas! Deducimos que cada infectado “aporta” con $k \cdot S(n - 1)/H$ infectados nuevos al día, en promedio. Por lo tanto los infectados nuevos son

$$I_{\text{nuevos}}(n) = \frac{kS(n - 1)}{H} \cdot I(n - 1)$$

⁶Quien desee, puede consultar en la versión en inglés de Wikipedia [\[2\]](#), donde está bien explicado el modelo SIR y su historia.

Por otro lado, los que dejan de estar infectados son los que pasan al grupo “R”. Esto es

$$R(n) - R(n - 1).$$

Por lo tanto, el número total de infectados de hoy menos los de ayer son:

$$I(n) - I(n - 1) = \frac{kS(n - 1)}{H} \cdot I(n - 1) - (R(n) - R(n - 1))$$

Finalmente, de las ecuaciones que encerramos en cajas obtenemos la regla que permite saber como avanza la epidemia al pasar de los días:

Teorema 3.1 (Fórmulas del modelo SIR simplificado). *Según el modelo antes descrito, la evolución de las cantidades $S(n)$, $I(n)$ y $R(n)$ cuando el número de días n avanza, es dada por las ecuaciones*

$$\begin{aligned} S(n) &= H - I(n) - R(n) \\ I(n) &= I(n - 1) + \frac{kS(n - 1)}{H} \cdot I(n - 1) - R(n) + R(n - 1) \\ R(n) &= R(n - 1) + I_{\text{nuevos}}(n - d) \\ &= R(n - 1) + \frac{kS(n - d - 1)}{H} \cdot I(n - d - 1) \end{aligned}$$

Para utilizar estas ecuaciones, primero se calcula $R(n)$ (que depende solo de datos de fechas anteriores), después se calcula $I(n)$ (que depende de $R(n)$ y de datos de fechas anteriores), y finalmente se utilizan esos valores en el cálculo de $S(n)$.

4. PREDICCIONES USANDO EL SIR SIMPLIFICADO

Vamos a aplicar nuestro modelo simplificado para estudiar la evolución de la epidemia COVID19 en Chile⁷.

Advertencia. El modelo SIR simplificado se ajusta mejor a la realidad que el modelo exponencial pues toma en cuenta una serie de otros factores importantes. Sin embargo, una predicción más exacta a largo plazo debe incorporar muchos otros aspectos: por ejemplo, segregación de la población en grupos (distintas regiones de Chile), cambio de medidas de prevención (el parámetro k cambiaría en el proceso), el posible colapso del sistema de salud y la posible falta de suficientes tests (el parámetro d se haría mayor), etc.

Básicamente:

- Modelo exponencial: acertado el primer o segundo mes, si la proporción de contagios no cambia (el parámetro k).
- Modelo SIR (o el SIR simplificado): acertado los primeros meses, si la proporción de contagios se mantiene relativamente constante en ese periodo y el sistema de salud no colapsa.
- Para una proyección más a largo plazo (por ejemplo, el primer año), es necesario un modelo matemático más sofisticado.
- En todos los modelos matemáticos, las predicciones pueden ser menos acertadas si no se dispone de datos correctos.

En relación al parámetro k (la cantidad promedio de contactos infecciosos diarios de un infectado con otras personas) hacemos la siguiente observación que permite aproximar k con los datos disponibles en las etapas iniciales de una epidemia:

⁷El modelo debería funcionar mejor si se aplica a un área mas concentrada, como por ejemplo la región metropolitana.

Observación. En las etapas iniciales de una epidemia, la gran mayoría de la población todavía está sana y el número de recuperados o fallecidos todavía es muy bajo. Entonces $S(n)/H$ es muy cercano a 1 cuando la epidemia comienza, y $R(n) - R(n - 1)$ es pequeño. Esto da:

$$I(n) = I(n - 1) + \frac{kS(n - 1)}{H} \cdot I(n - 1) - R(n) + R(n - 1) \approx I(n - 1) + k \cdot I(n - 1) = (1 + k) \cdot I(n - 1)$$

lo que se aproxima a la fórmula del modelo exponencial dada en el Teorema [2.1](#).

Teorema 4.1 (El avance inicial es exponencial). *Al inicio de una epidemia, cuando la cantidad de sanos es muy cercana al total de la población (es decir, cuando la cantidad de personas que se han enfermado es todavía extremadamente chica comparada al total de la población) se cumple que el modelo exponencial y el modelo SIR simplificado se comportan de forma aproximadamente idéntica.*

Por lo tanto, el parámetro k calculado al inicio de una epidemia según el modelo exponencial es una buena aproximación del parámetro k que necesitamos en el modelo SIR simplificado. En el caso del COVID19 en Chile, como ya calculamos en la Sección 2, los datos actuales indican $k = 0,194$.

Los datos de cantidad de personas infectadas que usaremos son los dados por el MINSAL (un registro de fácil acceso se puede encontrar en [3](#)).

Día	n	$T(n)$	$I(n)$	$I(n) - I(n - 1)$	Recuperados	Fallecidos	$R(n)$	$R(n) - R(n - 1)$
3 Marzo	1	1	1					
4 Marzo	2	3	3	2				
5 Marzo	3	4	4	1				
6 Marzo	4	5	5	1				
7 Marzo	5	7	7	2				
8 Marzo	6	11	11	4				
9 Marzo	7	13	13	2				
10 Marzo	8	17	17	4				
11 Marzo	9	23	23	6				
12 Marzo	10	33	33	10				
13 Marzo	11	43	43	10				
14 Marzo	12	61	61	18				
15 Marzo	13	75	75	14				
16 Marzo	14	156	156	81				
17 Marzo	15	201	201	45				
18 Marzo	16	238	238	37				
19 Marzo	17	342	342	104				
20 Marzo	18	434	434	92				
21 Marzo	19	537	528	94	8	1	8	
22 Marzo	20	632	623	95	8	1	9	0
23 Marzo	21	746	733	110	11	2	13	4
24 Marzo	22	922	903	179	17	2	19	6
25 Marzo	23	1142	1117	214	22	3	25	6
26 Marzo	24	1306	1269	152	33	4	37	12
27 Marzo	25	1610	1562	293	43	5	48	11
28 Marzo	26	1909	1842	280	61	6	67	19
29 Marzo	27	2139	2057	215	75	7	82	15

La columna $T(n)$ es el total oficial publicado por el MINSAL. Restando los recuperados oficiales y fallecidos obtenemos la columna $I(n)$ de los infectados vigentes el día n . Idealmente uno quisiera saber el total *real* de infectados. A falta de mejores datos, usaremos estos⁸.

El parámetro d que usaremos es $d = 14$. Esto por dos razones:

- La cuarentena recomendada suele ser de 2 semanas. Esto concuerda que en la mayoría de los casos leves suele haber recuperación dentro de las dos primeras semanas [4].
- Según el modelo, esperamos que aproximadamente se cumpla

$$R(n) - R(n - 1) = I_{nuevos}(n - d)$$

y para los primeros días de la epidemia se cumple $I_{nuevos}(n - d) = I(n - d) - I(n - d - 1)$. En la tabla anterior, los números 0, 4, 6, 6, 12, 11, 19, 15 de la columna $R(n) - R(n - 1)$ parecen coincidir mejor con el bloque 4, 2, 4, 6, 10, 10, 18, 14 de la columna $I(n) - I(n - 1)$, lo que corresponde a tomar $d = 14$ en el modelo.

Con todo esto, podemos aplicar las fórmulas del modelo SIR simplificado (Teorema 3.1) para anticipar la evolución de $S(n)$, $I(n)$ y $R(n)$.

A modo de ejemplo, calcularemos la proyección para el 30 de Marzo, correspondiente a $n = 28$. Primero,

$$S(27) = 19000000 - I(27) - R(27) = 19000000 - 2057 - 82 = 18997861$$

y ahora usamos las fórmulas del Teorema 3.1 para $n = 28$ (con $k = 0,194$ y $d = 14$). Primero $R(28)$ proyectado (recuperados y fallecidos):

$$R(28) = R(27) + I_{nuevos}(14) = 82 + 156 - 75 = 163$$

y ahora el número de infectados proyectado:

$$\begin{aligned} I(28) &= I(27) + \frac{0,194 \cdot S(27)}{19000000} \cdot I(27) - (R(28) - R(27)) \\ &= 2057 + \frac{0,194 \cdot 18997861}{19000000} \cdot 2057 - (163 - 82) \approx 2375 \end{aligned}$$

y obtenemos los valores proyectados $I(28) = 2375$ y $R(28) = 163$.

Nota agregada durante la edición final. *Mientras se editaba la versión final de este artículo, los datos para el Lunes 30 de Marzo se liberaron por el MINSAL:*

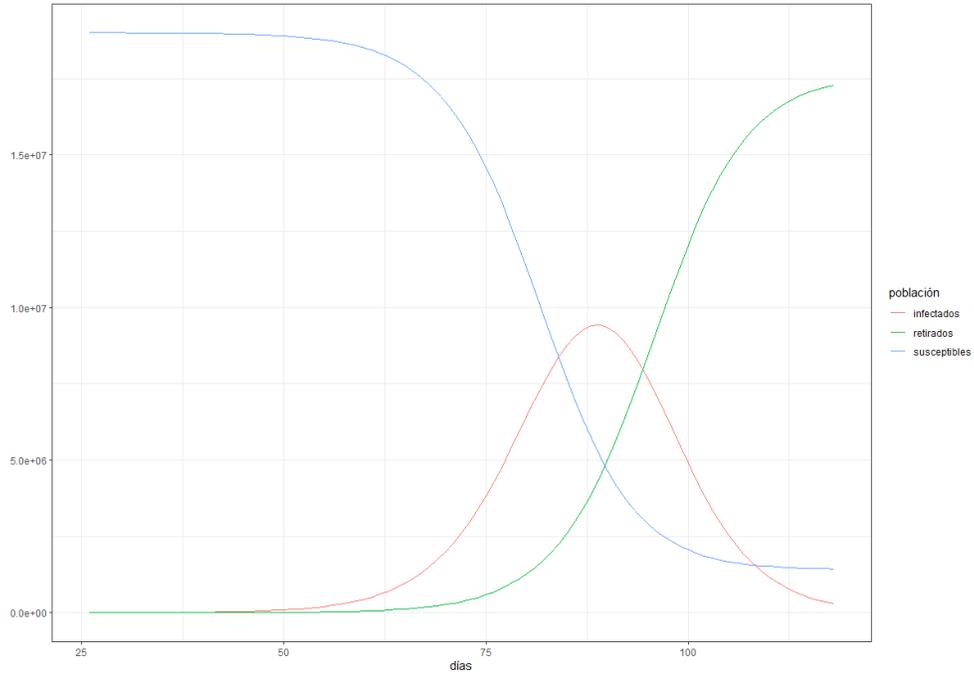
2449 casos totales, 156 recuperados, 8 fallecidos a la fecha. Esto da los valores oficiales $I(28) = 2287$ y $R(28) = 162$. El error de la proyección del modelo en relación a los datos oficiales es un 3,8% para el número de infectados vigentes, y un 0,6% para el número de fallecidos o recuperados. La coincidencia casi exacta para el valor de $R(28)$ (proyección: 163; oficial 162) confirma que la elección de $d = 14$ es adecuada para esta etapa de la epidemia. El error de 3,8% en $I(28)$ es bajo pero en la dirección favorable, y admite al menos dos explicaciones posibles:

- *puede ser que las medidas de aislamiento están empezando a mostrar resultados,*
- *o puede ser que son tantos casos (y posiblemente no toda la gente está dispuesta a solicitar el test en presencia de solo síntomas leves) que se hace difícil aplicar el test de forma suficiente como para tener un conteo fidedigno.*

No sabemos cuál de estas dos explicaciones es la más acertada.

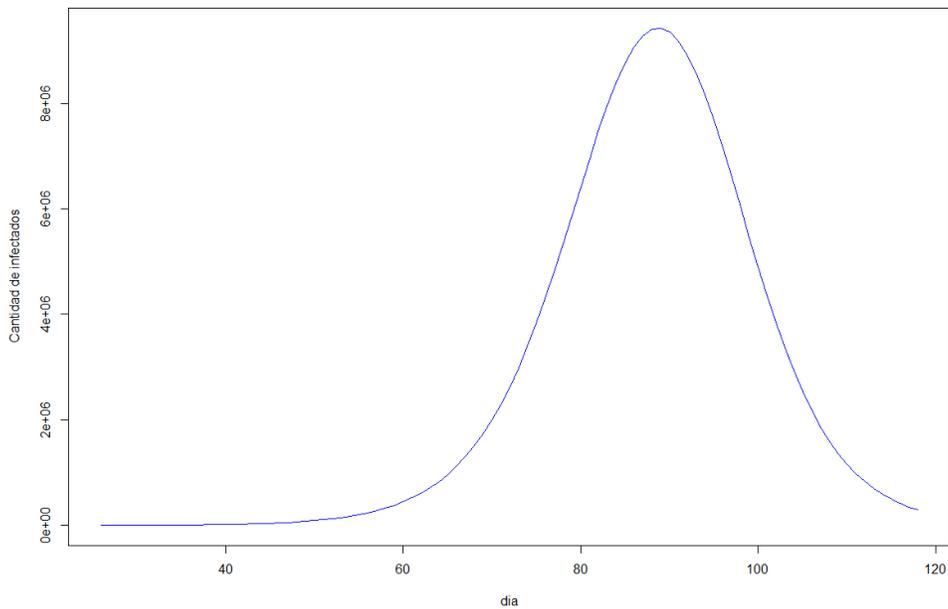
Uno puede hacer un pequeño programa de computador para calcular secuencialmente los números $S(n)$, $I(n)$ y $R(n)$. El gráfico de $S(n)$, $I(n)$ y $R(n)$ proyectados por el modelo para los primeros 120 días es el siguiente:

⁸Se podría tratar de usar otro modelo matemático para estimar el número real de casos en base a la información disponible, contrastando con los datos de países donde se hizo tests de COVID19 a una proporción de la población mucho mayor.



La curva azul es del grupo sano susceptible S , la roja en forma de campana es de los infectados I , y la verde es de los recuperados o fallecidos R .

En más detalle, este es el gráfico solamente de la curva de infectados (ahora en azul):



Visiblemente alcanza un peak alrededor del día 90 contando desde el inicio de los registros (3 de Marzo). Más precisamente, mirando los resultados calculados por el programa de computador que escribimos siguiendo nuestro modelo, se pronostica un peak el día 90 (final del mes de Mayo - inicio de Junio), con un total proyectado de aproximadamente 9 400 000 infectados (en palabras:

más de 9 millones). Este escenario es *suponiendo que la proporción de contagio actual $k = 0,194$ se mantiene*. Por otro lado, este modelo predice que, al ritmo actual, a finales de Abril la cantidad de infectados activos sería del orden de los 300 000; mejor que el modelo exponencial (que proyectaba medio millón), pero aún así es terrible.

Cabe recordar que este modelo es una versión muy simplificada del verdadero modelo SIR porque evitamos usar ecuaciones diferenciales. Así que:

- Hay un margen de error mayor que con modelos matemáticos más sofisticados.
- Con el paso del tiempo, la predicción del modelo es cada vez menos confiable.
- Los datos de los que disponemos son limitados y no necesariamente reflejan la realidad.

Pero a pesar de las falencias de este sencillo modelo matemático y la precaria calidad de los datos utilizados como insumo, lo relevante es que:

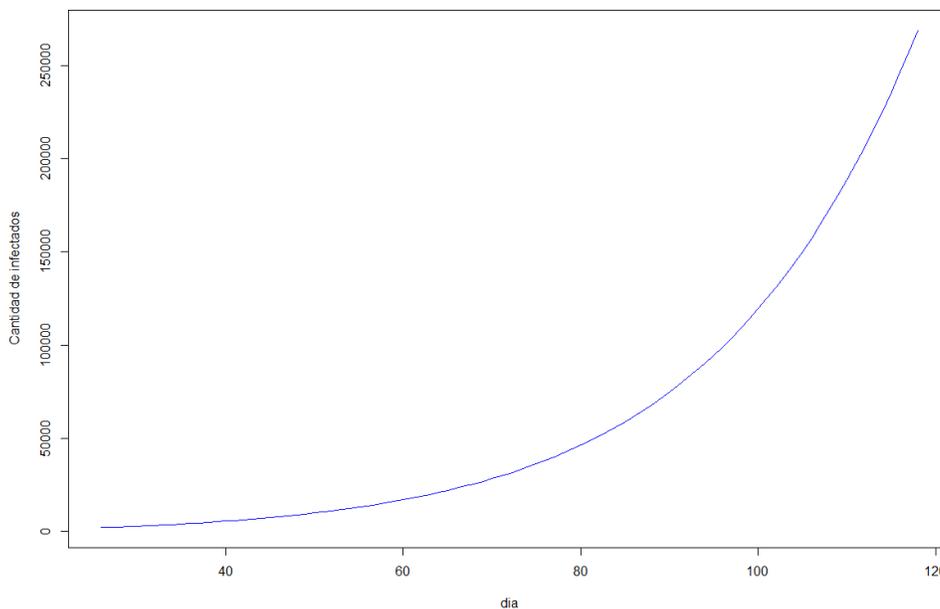
- Si continuamos a este paso, seguramente veremos un peak de la cantidad de contagios muy cerca en el futuro, posiblemente a finales de Mayo o inicios de Junio.
- El número de casos se sale de control mucho antes, posiblemente a fin de Abril.

Así que la peor parte de la epidemia se viene muy pronto, a menos que se tomen acciones drásticas para frenar el avance y bajar el valor del parámetro k .

5. ¿QUÉ PASARÍA SI LOGRAMOS BAJAR EL VALOR DE k ?

Si tomáramos medidas de aislamiento más serias, podríamos bajar el valor de k . ¿Qué efecto tendría esto?

Por ejemplo, imaginemos que desde hoy mismo consiguiéramos $k = 0,1 = 1/10$. Esto correspondería a medidas drásticas de aislamiento en *toda* la población de Chile, tomadas *inmediatamente* para bajar la frecuencia de interacciones contagiosas nuevas a solamente 1 al día por cada 10 infectados (la catástrofe proyectada en la sección anterior corresponde a una por cada 5 infectados, *que es exactamente como van las cosas hasta hoy*). En este escenario hipotético de $k = 0,1$ y continuando con $d = 14$, el gráfico de valores proyectados de $I(n)$ para los primeros 120 días (hasta fin de Junio) sería



Comparemos con el escenario catastrófico $k = 0,194$ analizado antes. Ahora, el día 90 (fin de Mayo - inicio de Junio) la cantidad de contagiados activos sería del orden de los 70 000 (setenta mil). Eso sería un escenario mucho menos devastador, pero de todas formas es *muuy malo*.

Sin embargo, **la realidad hasta el momento es mucho menos auspiciosa** que el escenario hipotético de $k = 0,1$. De hecho, el avance actual del COVID19 en Chile se ajusta más al valor $k = 0,194$ que analizamos antes.

La conclusión es que las medidas para frenar el avance deben tomarse:

- **en toda la población**, pues un valor *promedio* de k más grande sigue siendo suficiente para un crecimiento acelerado;
- **y tan pronto como sea posible**, porque en esta etapa la epidemia todavía crece a ritmo exponencial (Teorema 4.1), y cada día sin tomar medidas efectivas empeora severamente el pronóstico a futuro.



Sobre los autores.

Héctor Pastén Vásquez (PhD) es un matemático de la Pontificia Universidad Católica de Chile. Doctor en Matemáticas de la Universidad de Concepción (2010) y Doctor en Matemáticas de Queen's University (2014). Anteriormente fue investigador en la Universidad de Harvard (2014-2018) y en el Institute for Advanced Study de Princeton (2015-2016). Ganador de la Medalla de Oro al Mérito Académico del Gobernador General de Canadá (2014) y del premio MCA del Mathematical Council of the Americas (2017) entre otros galardones internacionales.

Jorge Castillo Sepulveda (PhD) es ingeniero civil matemático de la Universidad de Concepción (2015) y Doctor en Ciencias de la Complejidad Social de la Universidad del Desarrollo (2019). Actualmente se desempeña como científico de datos en la compañía Z Data Lab, donde lidera las áreas de inteligencia de precios e inteligencia comercial, y es investigador en la Universidad del Desarrollo, donde desempeña labores en proyectos asociados al tratamiento de datos y en docencia de posgrado.

REFERENCIAS

- [1] Kermack, W. O.; McKendrick, A. G. *A Contribution to the Mathematical Theory of Epidemics*. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences. (1927) 115 (772): 700.
- [2] Wikipedia: https://en.wikipedia.org/wiki/Compartmental_models_in_epidemiology
- [3] Wikipedia (con datos del MINSAL): https://es.wikipedia.org/wiki/Pandemia_de_enfermedad_por_coronavirus_de_2020_en_Chile
- [4] Harvard: <https://www.health.harvard.edu/diseases-and-conditions/coronavirus-resource-center>

(H. Pasten) FACULTAD DE MATEMÁTICAS, UC CHILE
Email address: `hector.pasten@mat.uc.cl`

(J.Castillo-Sepúlveda) Z DATA LAB
Email address: `jorgecastillo@zdatalab.com`

ANEXO A: “EVOLUCIÓN DE LAS EPIDEMIAS: LA MATEMÁTICA DE AISLARSE”

HÉCTOR PASTÉN, JORGE CASTILLO SEPÚLVEDA

En el artículo “Evolución de las epidemias: la matemática de aislarse” presentamos una discusión sobre modelos matemáticos para estudiar la evolución de epidemias, orientado a público general. El propósito es crear conciencia sobre la importancia del aislamiento como medida para frenar el crecimiento del COVID19 en Chile, y entregar herramientas educativas sobre cómo los científicos hacemos proyecciones en este tema.

A pesar que el público objetivo del artículo no eran los especialistas sino una audiencia más amplia, con grata sorpresa hemos recibido consultas y comentarios de un carácter más técnico. En este anexo trataremos de elaborar en esos puntos.

No queremos dejar fuera a quienes nos acompañaron leyendo el artículo inicial. Por lo que vamos a tratar de limitar la jerga científica en la medida de lo posible, y exponer las ideas centrales de forma transparente.

Además, vamos a discutir una pregunta tristemente relevante: ¿Qué hacer cuando los datos entregados por la autoridad no son reales?

1. SOBRE EL MODELO SIR SIMPLIFICADO

El modelo SIR original, en su forma más clásica consta de un sistema de ecuaciones diferenciales que modelan la evolución de una población clasificada en 3 grupos: Susceptibles (S), Infectados (I), y Retirados (recuperados o fallecidos, R). Ver [2,3].

El “SIR simplificado” presentado en la segunda mitad del artículo sigue los principios básicos del SIR clásico, pero seguramente los especialistas ya notaron que se diferencia en 2 aspectos:

- Es *discreto*: la evolución de los números $S(n)$, $I(n)$, $R(n)$ es “a saltos” de un día al siguiente.
- La evolución de $R(n)$ no es gobernada por una “tasa de recuperación-mortalidad” como en el SIR clásico. En su lugar, es gobernada por el parámetro d = número de días que tarda la recuperación.

Nos vimos en la necesidad de incorporar estas diferencias con el SIR clásico por dos motivos:

- Queríamos un modelo serio, pero que utilice matemática de un nivel no más avanzado que enseñanza media. De este modo esperamos llevar las ideas fundamentales a una audiencia mucho más amplia, pero siendo rigurosos. Esto nos llevó a preferir un modelo discreto (en lugar de uno de tiempo continuo, que necesitaría ecuaciones diferenciales¹).
- La tasa de recuperación-mortalidad para el COVID19 no es para nada clara. Especialmente en Chile en esta etapa, pues existen muy pocos datos precisos al respecto. Entonces el modelo elegido debía controlar la evolución de R de alguna otra manera con algún parámetro mejor entendido. A pesar de ser un artículo con finalidad educativa, y que la calidad y cantidad de los datos disponibles para el COVID19 en Chile no es óptima, quisimos que el modelo sea tan confiable como sea posible. Por eso trabajamos con el parámetro d .

Date: 31 de marzo de 2020.

¹Dicho sea de paso, las ecuaciones diferenciales del SIR clásico no son complicadas, y para efectos de aplicación basta con resolverlas numéricamente. Sin embargo, a pesar de no ser un tema difícil para los iniciados en las artes ocultas de las matemáticas superiores, decidimos evitar ecuaciones diferenciales porque no se suelen ver en el liceo.

En concreto, el modelo supone que los enfermos nuevos de un día x , aproximadamente son los que se recuperan el día $x + d$. En símbolos, $R(n) - R(n - 1) = I_{\text{nuevos}}(n - d)$.

Obviamente esto se trata de una simplificación de la realidad: cada enfermo tiene sus propios tiempos, y eso hace que no sea tan sencillo determinar una buena elección de d para el modelo.

Aprovechamos de mencionar que el símbolo d no solo es por “días”, sino también por la palabra “delay” que es el término técnico utilizado en matemáticas y otras ciencias para referirnos a parámetros que cumplen roles similares a nuestro d . Se puede decir entonces que el modelo SIR simplificado del artículo es un ejemplo de “modelo SIR discreto con delay”.

2. ELIGIENDO EL PARÁMETRO d

Entonces queda pendiente cómo elegir el parámetro $d =$ número de días que tarda la recuperación (en promedio). La verdad es que en este punto no existe claridad sobre cuánto dura la enfermedad en promedio. Pero esto es lo que se sabe:

- El periodo de incubación usual (infectado sin síntomas) se estima en 5-6 días aproximadamente [1] pero los investigadores advierten que hay gran variabilidad entre casos.
- No se sabe cuántos días antes del primer síntoma el portador comienza a ser contagioso [5].
- Desde el primer síntoma, se estima que en la mayoría de los casos la enfermedad dura de 1 a 2 semanas [5].
- Desde que desaparecen los síntomas, no se sabe cuánto tiempo más la persona sigue siendo significativamente contagiosa. Por precaución, se sugiere extremar cuidados por unos 8 días después del final de los síntomas leves [5].

(Recomendamos consultar [4] por orientación y más información sobre el COVID19.)

Tomando esto en consideración, la elección $d = 14$ (2 semanas) parece razonable para el caso típico: unos 10 – 11 días de síntomas (entre 1 a 2 semanas) donde se puede asumir que es contagioso, unos 3 – 4 días contagioso durante el período típico de incubación, y esperar que una vez desaparecidos los síntomas se aplique extremo cuidado por un tiempo (a pesar de no estar claro si todavía los pacientes que se alivian de los síntomas son significativamente contagiosos en general).

Insistimos: para utilizar el modelo matemático basta saber el caso típico. Pero para tomar medidas en enfermos reales se debe hacer seguimiento y proceder con todas las precauciones ya que cada caso tiene sus propios tiempos. Por ejemplo, en casos mas serios la enfermedad puede tomar hasta 6 semanas o más [5].

Todo esto es nebuloso, pero esperamos que los estudios que van apareciendo logren aclarar estas dudas para llegar a una mejor estimación de d .

Con esto en mente, en el modelo SIR simplificado de nuestro artículo hemos utilizado $d = 14$.

¿Significa esto que todos los enfermos se recuperan luego de 14 días? La respuesta es un rotundo *no*. Como ya mencionamos, hay enfermos que pueden tomar hasta 6 semanas o más. El número 14 es una estimación de la duración típica que basta para el modelo matemático, pero los casos reales estudiados de forma individual presentan mayor variabilidad en su duración.

Por otro lado, como se explicó en el artículo, los datos sobre pacientes recuperados que han sido publicados por el MINSAL son consistentes con esta estimación. Sin embargo, aquí viene un gran “pero”: hoy Martes 31 de Marzo², la subsecretaria de Salud Pública, Paula Daza, declaró que las cifras de pacientes recuperados publicadas por el MINSAL “son estimaciones”, ver por ejemplo [6]. No sabemos qué significa eso exactamente, así que naturalmente esto levanta dudas sobre la calidad de los datos de pacientes recuperados entregados por el MINSAL³.

²Un día después de nuestro artículo, pero ciertamente no relacionado con el mismo.

³Y si los datos de pacientes recuperados son estimaciones, cabe preguntarse hasta qué punto son confiables los datos oficiales sobre casos confirmados.

3. EL MODELO ES ROBUSTO EN d PARA PREDECIR EL PEAK

Pongámonos en la peor situación: ¿Qué podemos hacer los científicos para elaborar proyecciones confiables cuando los datos entregados por la autoridad no son reales?

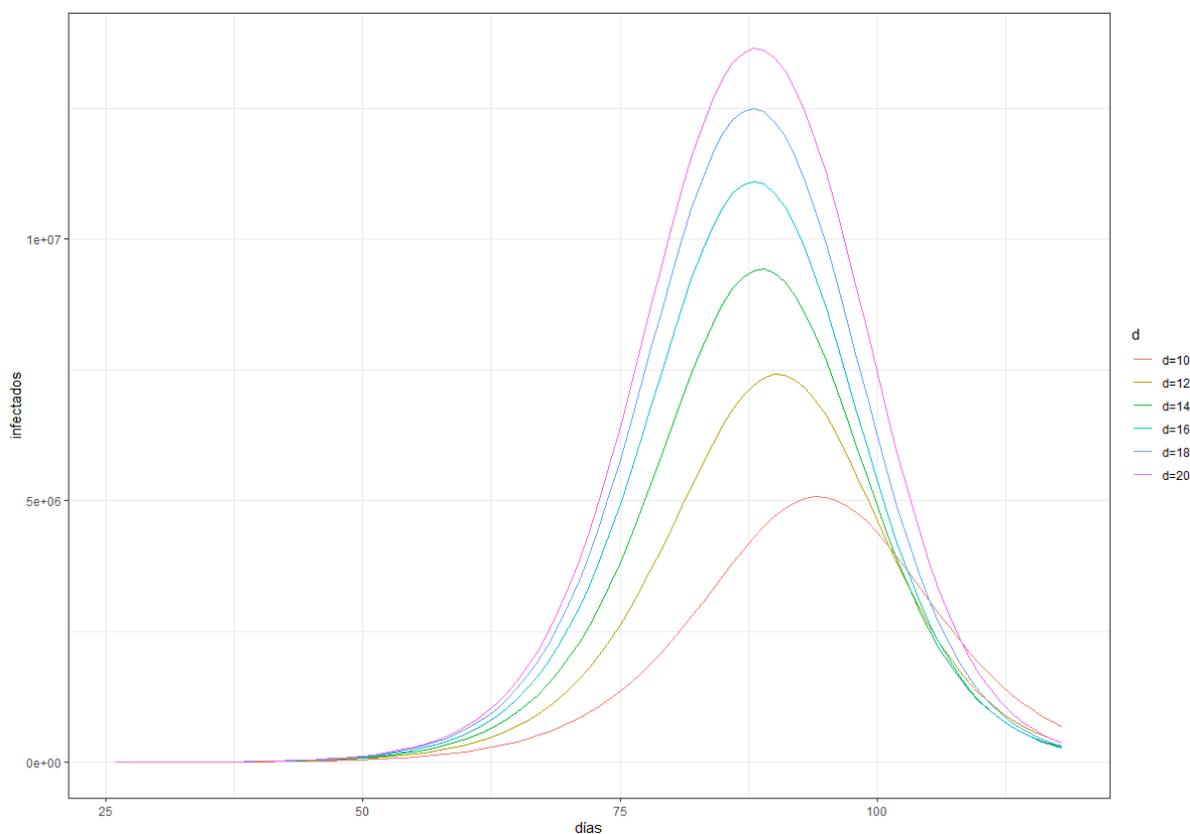
Este es un problema extremadamente serio: *Predecir el avance de la epidemia depende de la calidad de los datos que tenemos.*

Una opción es utilizar los datos *reales* entregados por instituciones internacionales u otros gobiernos. Eso fue lo que hicimos al elegir nuestro $d = 14$. Pero eso tiene una limitación: los datos en otros países son función de muchos factores propios de cada país, como por ejemplo el sistema de salud, capacidad de diagnóstico, clima, densidad poblacional, etc. Por lo tanto solo se puede asumir que el valor $d = 14$ es una estimación de lo que podemos esperar en Chile, pero lamentablemente sin datos confiables de Chile para contrastar.

Aquí viene a ayudarnos un concepto importantísimo en el diseño de modelos matemáticos, y que fue tomado en cuenta cuando diseñamos el modelo SIR simplificado: *robustez*.

Resulta que la predicción de la fecha peak entregada por nuestro modelo SIR simplificado⁴ es *robusta en d* en el siguiente sentido: incluso si el valor real de d para Chile es un poco distinto a $d = 14$, la predicción de la fecha peak sigue siendo aproximadamente la misma.

Sin más preámbulo, estos son los gráficos de evolución de $I(n)$ para $d = 10, 12, 14, 16, 18, 20$, en un lapso de 120 días (hasta fin de Junio) y con el parámetro $k = 0,194$ como fue analizado en la sección 4 del artículo (con datos de la semana pasada utilizados al escribir el artículo):



⁴En el caso cuando el parámetro de contagios actual $k = 0,194$ se mantiene, que es la situación analizada en la sección 4 del artículo. Este caso ocurriría si mantenemos la tendencia de la semana pasada de interacciones contagiosas diarias por infectado, y es una situación que debemos evitar a toda costa.

Por supuesto, las elecciones $d = 10$ y $d = 20$ no son realistas; se incluyen únicamente para testear la robustez de la predicción de la fecha peak dada por el modelo SIR simplificado.

Los datos producidos por nuestro programa de computador que calcula la evolución del modelo⁵ nos dan las siguientes fechas para el peak del número de infectados $I(n)$, utilizando $k = 0,194$ como en la sección 4 del artículo, y según cada valor de d :

d	n peak proyectado	Fecha peak proyectada (en 2020)	Peak proyectado de infectados
10	95	5 de Junio	5,1 millones
12	91	1 de Junio	7,4 millones
14	90	31 de Mayo	9,4 millones
16	89	30 de Mayo	11,1 millones
18	89	30 de Mayo	12,5 millones
20	89	30 de Mayo	13,6 millones

En todos los casos el peak ocurriría con varios millones de infectados, a fin de Mayo - inicio de Junio. Esto ocurre incluso si usamos los valores poco realistas $d = 10$ y $d = 20$ como duración promedio de la enfermedad.

Por lo tanto, el análisis de la sección 4 del artículo utilizando el SIR simplificado es bastante robusto para perturbaciones del parámetro d .

Tomamos esta oportunidad para recordar la conclusión del análisis de la sección 4 del artículo: la peor parte de la epidemia se viene muy pronto, a menos que se tomen acciones drásticas para frenar el avance y bajar el número de interacciones contagiosas diarias por infectado.

REFERENCIAS

- [1] Stephen A. Lauer, Kyra H. Grantz, Qifang Bi, Forrest K. Jones, Qulu Zheng, Hannah R. Meredith, Andrew S. Azman, Nicholas G. Reich, Justin Lessler. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Annals of Internal Medicine*, 2020; DOI: 10.7326/M20-0504
- [2] Kermack, W. O.; McKendrick, A. G. *A Contribution to the Mathematical Theory of Epidemics*. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*. (1927) 115 (772): 700.
- [3] Wikipedia: https://en.wikipedia.org/wiki/Compartmental_models_in_epidemiology
- [4] Harvard: <https://www.health.harvard.edu/diseases-and-conditions/coronavirus-resource-center>
- [5] Harvard: <https://www.health.harvard.edu/diseases-and-conditions/if-youve-been-exposed-to-the-coronavirus>
- [6] Publimetro: <https://www.publimetro.cl/cl/social/2020/03/31/casos-recuperados-estimaciones-la-declaracion-paula-daza-causo-indignacion-redes.html>

(H. Pasten) FACULTAD DE MATEMÁTICAS, UC CHILE
Email address: `hector.pasten@mat.uc.cl`

(J.Castillo-Sepúlveda) Z DATA LAB
Email address: `jorgecastillo@zdatalab.com`

⁵No incluimos las fórmulas de nuevo: son exactamente las fórmulas del Teorema 3.1 de nuestro artículo.