

COVID-19 Y LA TASA DE CONTACTOS INFECCIOSOS: UNA HISTORIA NARRADA A COLOR

HÉCTOR PASTÉN, JORGE CASTILLO SEPÚLVEDA

ÍNDICE

1. Introducción	1
2. La tasa de contactos infecciosos diaria	2
3. Zonas verde, gris y roja	2
4. ¿Dónde estamos?	5
5. Lo que pudo ser... pero no fue	7
6. Detalles técnicos	10

1. INTRODUCCIÓN

¿Vamos por el camino correcto en la lucha contra el COVID-19? *Íbamos, hasta el 12 de Abril.*

Una variable crucial en la evolución de una epidemia como el COVID-19 es *la tasa de contactos infecciosos*. Su papel central se debe al menos a tres aspectos:

- *Es nuestra amiga*. En ausencia de una cura o tratamiento, esta es la *única* variable que podemos controlar para intentar frenar el avance de la epidemia.
- *Es informativa*. Determina el ritmo de avance de una epidemia. Por lo tanto, ayuda a hacer predicciones a futuro y a evaluar nuestra situación actual.
- *Es calculable*. Se puede estimar su valor real con buena precisión a partir de los datos medidos, incluso si hubiera imprecisión en las mediciones.

Todo lo anterior es bien sabido por especialistas. Un primer propósito de este artículo es que el público en general también lo sepa.

Además, entregamos una “receta” para facilitar la lectura e interpretación de la tasa de contactos infecciosos. Para eso, definimos una “zona verde”, “zona gris” y “zona roja” para este indicador de acuerdo a los posibles impactos de la epidemia. Dicha separación se determinó gracias un detallado análisis de modelos predictivos contrastado a los datos y limitaciones reales en Chile.

Los datos oficiales en conjunto con nuestro análisis nos llevan a las siguientes conclusiones:

- Hasta antes del 12 de Abril: íbamos en ruta a frenar la epidemia hasta niveles manejables dentro del primer semestre de 2020. Esto se determinó usando los datos oficiales hasta el 12 de Abril y proyectando con modelos epidemiológicos en conjunto con métodos de reconocimiento de patrones basadas en técnicas de inteligencia artificial, específicamente *machine learning*.
- Sin embargo, luego del 12 de Abril la dinámica cambió: tomamos una ruta hacia una evolución mucho más compleja, aunque no necesariamente catastrófica.
- Desde este punto de vista, ahora no parece ser el mejor momento para que las medidas de mitigación y aislamiento sean menos restrictivas.

Date: 23 de abril de 2020.

El artículo se terminó de escribir el 22 de Abril de 2020. El proceso de edición tomó hasta el 23 de Abril.

En nuestro análisis utilizaremos un modelo matemático tipo SIR con varias modificaciones que debimos incorporar para hacer frente a las particularidades de esta epidemia. Dichas modificaciones tienen por propósito obtener un modelo matemático robusto (dada la incertidumbre de los datos) y adecuado a posibles cambios bruscos de parámetros que en otros estudios suelen considerarse constantes. Los principales aspectos técnicos de este modelo están explicados en la sección 6.

2. LA TASA DE CONTACTOS INFECCIOSOS DIARIA

Convención. Los días son numerados por n , siendo $n = 0$ el 3 de Marzo de 2020 (“día cero” de la primera infección). El 12 de Abril, día crítico en los datos que analizaremos, corresponde a $n = 40$. El 22 de Abril, día en que este estudio fue terminado, corresponde a $n = 50$.

La *tasa de contagios* de hoy se define como la cantidad promedio de contagiados nuevos de hoy que fueron debidos a cada contagiado activo de ayer. La tasa de contagios del día n se escribe $c(n)$.

Ejemplo real. El 22 de Abril ($n = 50$) hubo 464 casos nuevos confirmados de COVID-19 en Chile, y el 21 de Abril hubo un total oficial *estimado*¹ de 5716 casos activos. Entonces la tasa de contagios el 22 de Abril fue $c(50) = 464/5716 \approx 0,081 = 8,1\%$.

La tasa de contagios no es el indicador más adecuado, porque a medida que la epidemia avanza los infectados activos tienen más dificultad en encontrar gente sana no-inmune susceptible de ser infectada. Por eso es mejor considerar otra cantidad:

La *tasa de contactos infecciosos* de hoy es la cantidad promedio de contactos potencialmente contagiosos por cada uno de los infectados activos de ayer. Y decimos “potencialmente” porque no todo contacto termina en contagio: si un infectado interactúa con otro infectado o con un inmune, entonces no produce un nuevo contagio. Es símbolos, la tasa de contactos infecciosos del día n la vamos a llamar $k(n)$.

La fórmula precisa para el cálculo de $k(n)$ es dada en la sección 6. Es más complicada que para $c(n)$, pero si la lectora desea saber cuánto vale aproximadamente basta saber que $k(n) \approx c(n)$ mientras la gran mayoría de una población no haya sido infectada. Así, por ejemplo, la aproximación para el 22 de Abril $k(50) \approx 8,1\%$ es adecuada. Sin embargo, nosotros utilizamos la fórmula más complicada de la sección 6 para que nuestro análisis sea válido en periodos de tiempo más extensos.

Observación. Muy cercano a las tasas de contagio y de contacto infeccioso, hay otro indicador epidemiológico más: el *número reproductivo efectivo* R_E . Se define R_E como la cantidad promedio de contagios secundarios producidos por un infectado. Tenemos las aproximaciones $R_E \approx c(n) \cdot d$ y en las primeras etapas de una epidemia $R_E \approx k(n) \cdot d$ donde d es la duración promedio de la enfermedad (aproximadamente $d \approx 14$ según MINSAL). Sin embargo, los modelos predictivos tipo SIR usan directamente la tasa de contactos infecciosos $k(n)$, no R_E , por lo cual usaremos $k(n)$.

Además, R_E no es informativa en lo que respecta a aspectos temporales. Por ejemplo, $R_E > 1$ indica que la epidemia crece, pero no nos da información sobre el ritmo de crecimiento. Similarmente, $R_E < 1$ indica que eventualmente la epidemia estará controlada pero no nos indica en qué plazo.

3. ZONAS VERDE, GRIS Y ROJA

Notación. El número de contagiados activos el día n se escribirá $I(n)$.

La evolución de $I(n)$ está gobernada por la tasa de contactos infecciosos $k(n)$ no solamente como concepto, sino que de una forma bastante precisa dada por ecuaciones de modelos matemáticos. Así, se observa una evolución de $I(n)$ notoriamente distinta en tres situaciones:

¹El número de casos activos requiere saber el número de recuperados, del cual solo se manejan estimaciones.

- *Zona verde.* $k(n)$ se mantiene por debajo de un 6%. En este caso la epidemia se controla en un plazo razonable. El número de infectados activos $I(n)$ prontamente comenzaría a decrecer.
- *Zona gris.* $k(n)$ se mantiene entre un 6% y un 8%. En este caso la epidemia tiene una evolución compleja, aunque no necesariamente catastrófica. El número de infectados activos podría crecer, mantenerse estable, o decrecer, pero todo de forma demasiado lenta como para ser considerado una victoria, demandando un gran esfuerzo en medidas de mitigación y recursos. Además, una evolución en esta zona es muy sensible a pequeños cambios².
- *Zona roja.* $k(n)$ se mantiene por sobre un 8%. En este caso la epidemia avanza a un ritmo acelerado con consecuencias catastróficas.

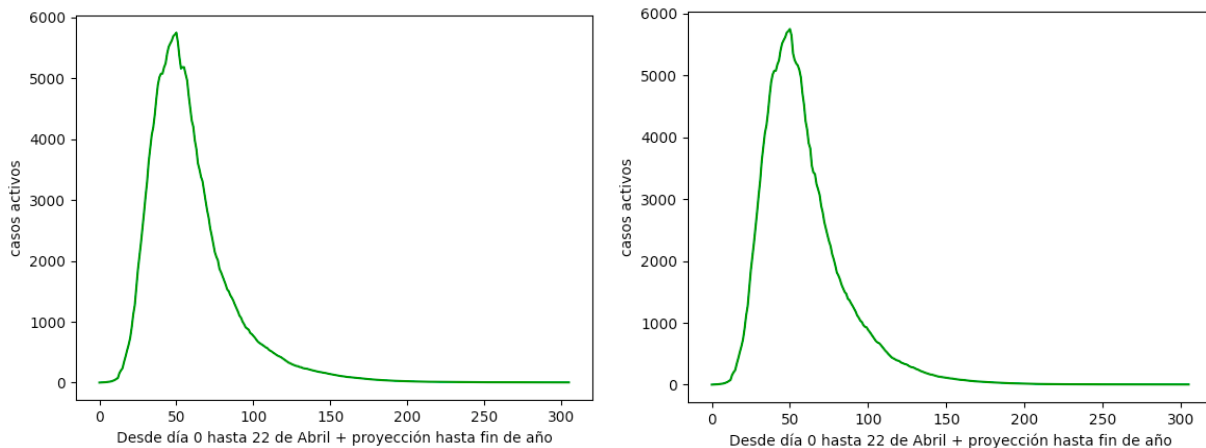
Aclaración. Estos regímenes de la tasa de contactos infecciosos deben entenderse en el sentido de tendencias. Por ejemplo, si la epidemia evoluciona principalmente con $k(n) < 6\%$ entendemos que estamos en la zona verde aunque ocasionalmente pueda haber algún día con $k(n) > 6\%$.

La separación en esas tres zonas no es arbitraria. Se elaboró en base a comparar la capacidad estimada de camas UTI y UCI del sistema de salud chileno disponibles para pacientes afectados por el COVID-19 (en el rango de 1000–3000 camas³) contra las cifras pronosticadas por nuestro modelo epidemiológico en cada escenario. Además, en los casos donde el número de infectados activos decrece lentamente (parte inferior de la zona gris) o se mantiene relativamente estable (parte media de la zona gris), se toma en consideración el tiempo que es necesario mantener medidas severas (e.g. cuarentenas, cierre de colegios, etc.), lo cual es un indicador del impacto socio-económico.

Para esto, hicimos un barrido de experimentos aleatorios con $k(n)$ variando en distintos intervalos comprendidos entre el 4% y 10%. La aleatoriedad es necesaria en un estudio realista de la tasa de contactos infecciosos porque $k(n)$ *no se mantiene constante*, sino que sigue tendencias variando de forma errática —esto será evidente al analizar datos reales en la próxima sección.

A continuación presentamos algunos ejemplos típicos que encontramos en este estudio. Damos algunos gráficos de evolución del número de infectados activos $I(n)$ usando datos oficiales a la fecha y suponiendo que la tasa de contactos infecciosos $k(n)$ se mantiene en determinados intervalos de la zona verde, gris o roja desde hoy hasta fin de año.

Ejemplo en zona verde. Evolución con $k(n)$ tomando valores aleatorios entre 4% y 6% (zona verde) desde hoy hasta fin de año, de forma independiente en ambos experimentos.

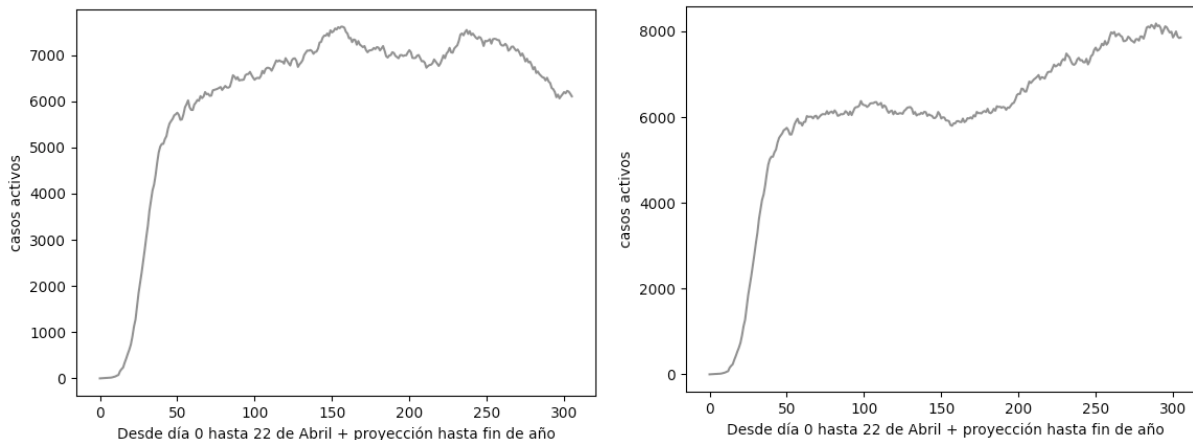


²Tiene aspectos de un sistema caótico.

³Ciertamente, eso es más de lo que existe disponible hoy en Chile. Sin embargo, se ha mencionado la intención de llegar a esa capacidad. A pesar de que el día de hoy no vemos evidencia de que esa meta se logre o no, no vamos a discutirlo aquí.

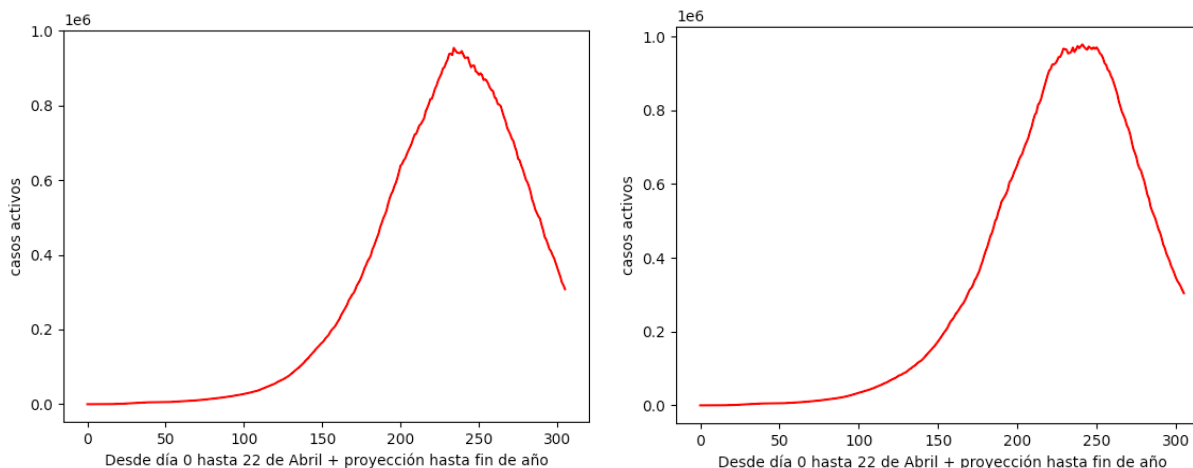
En ambos casos, la epidemia se frenaría hasta niveles manejables entre Mayo y Junio, con un peak inferior a los 6 000 casos activos.

Ejemplo en zona gris. Evolución con $k(n)$ tomando valores aleatorios entre 6% y 8% (zona gris) desde hoy hasta fin de año, de forma independiente en ambos experimentos.



Aunque se ven al borde de lo manejable por el sistema de salud, en realidad la estrategia de quedarse en la zona gris no es viable desde un punto de vista socio-económico: el número de infectados activos se mantiene relativamente estable pero *suponiendo que se mantienen todas las medidas de aislamiento actuales durante el resto del año 2020* (relajar dichas medidas movería $k(n)$ a la zona roja).

Ejemplo en zona roja. Evolución con $k(n)$ tomando valores aleatorios entre 8% y 10% (zona roja) desde hoy hasta fin de año, de forma independiente en ambos experimentos.



En ambos casos se alcanzaría un peak de alrededor un millón de infectados activos en el segundo semestre. Esto se traduciría en unos 50 000 pacientes que necesitarían hospitalización, de los cuales unos 20 000 necesitarían cuidados intensivos durante la fecha peak.

Observación 1. El comportamiento en zonas verde (buena) y roja (mala) es bastante robusto en el sentido que el resultado es similar a pesar que se trata de experimentos aleatorios independientes. Sin embargo, la evolución en zona gris cambia bastante entre experimentos aleatorios, mostrando gran sensibilidad a pequeños cambios. Una interpretación es que, si la epidemia evoluciona en zona

gris, se llevaría una lucha constante por mantenerla controlada, pues cambios en periodos cortos de tiempo tendrían gran impacto.

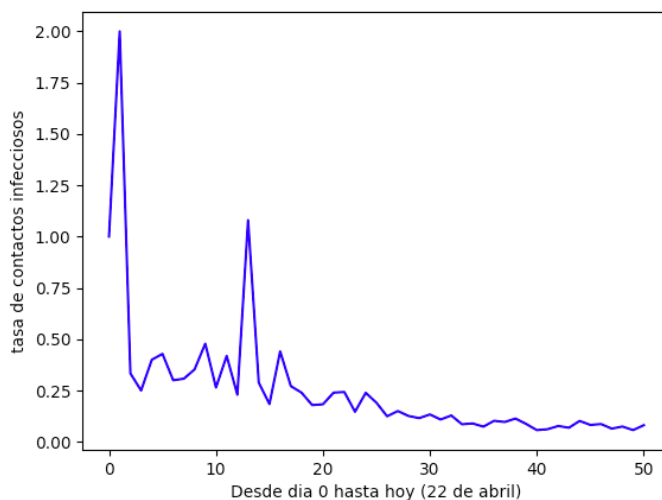
Observación 2. La situación de la zona verde es muy similar a lo que se conoce como “el martillo y el baile” (*the hammer and the dance*, expresión acuñada por Tomas Pueyo⁴). Consiste en una estrategia donde la cantidad de infectados y la tasa de contagios se baja drásticamente en un periodo corto (“el martillo”) seguido de un estilo de vida cuidadoso para mantener lo ganado (“el baile”). Esta estrategia no solo cuida la salud pública sino también minimiza daños económicos y sociales. Ha sido implementada con éxito en países orientales.

Observación 3. Un indicador de que una epidemia no está controlada es cuando el número reproductivo efectivo cumple $R_E > 1$. Por otro lado, un indicador teórico de que una epidemia eventualmente estará bajo control es $R_E < 1$. Sin embargo, estas distinciones solo son aproximadas debido a las ambigüedades relativas a R_E explicadas en la sección 2. De todas formas, con $d = 14$ y la fórmula $R_E \approx k(n) \cdot d$, se obtiene que el caso límite $R_E = 1$ corresponde aproximadamente a $k(n) = 1/14 \approx 7,14\%$. Esto cae en la zona gris, lo cual es consistente con nuestro análisis. Sin embargo, la dicotomía tradicional $R_E > 1$ v/s $R_E < 1$ es insuficiente porque ignora los tiempos de la epidemia. Justamente, la zona gris demarca un rango de posibilidades donde el número de casos activos no crece suficientemente rápido como para ser declarado catastrófico, pero no decae suficientemente rápido como para ser considerado una victoria.

4. ¿DÓNDE ESTAMOS?

La respuesta corta: *estamos en la zona gris*. Eso sería una buena noticia si se tratara de una transición de la zona roja a la verde. Sin embargo, los datos indican otra cosa.

A continuación vemos el gráfico de la tasa de contactos infecciosos $k(n)$ según los datos oficiales en lo que va de la pandemia en Chile⁵, ver la sección 6 para los detalles del cálculo de $k(n)$ a partir de datos oficiales.



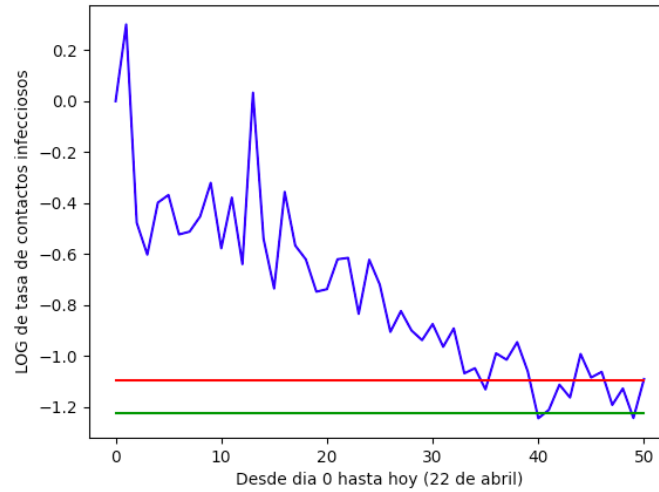
Como la lectora puede apreciar, del gráfico se observa que la tasa de contactos infecciosos ha presentado una cierta tendencia a la baja. Esa tendencia se explica por las cuarentenas, cierre de instituciones educativas, toques de queda, y por supuesto, la cooperación de la ciudadanía.

⁴Ver: <https://medium.com/@tomaspueyo/coronavirus-the-hammer-and-the-dance-be9337092b56>

⁵El valor $k(0)$ no está definido, así que declaramos “ $k(0) = 1$ ” para poder visualizar el gráfico en la escala temporal correcta.

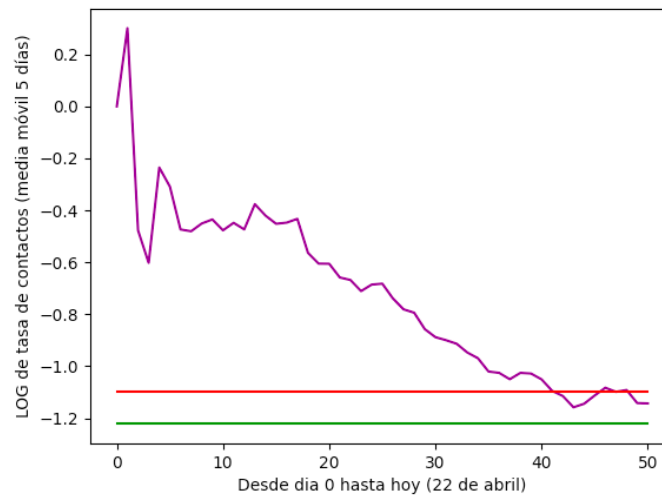
Sin embargo, dada la importancia de entender bien el avance de la epidemia, **no es suficiente con observar que la tasa de contactos infecciosos parece tener una tendencia a la baja. Se hace necesario un análisis mucho más fino.**

Al tratarse de números tan pequeños, el análisis se facilita pasando a escala logarítmica. El gráfico de la tasa de contactos infecciosos diaria $k(n)$ en escala \log_{10} es el siguiente:



Se aprecia una tendencia de decrecimiento hasta aproximadamente $n = 40$ (12 de Abril). Lo que está sobre la línea roja es la zona roja, lo que está entre la roja y la verde es la zona gris, y lo que está por debajo de la línea verde es la zona verde —todo esto en escala logarítmica. Se hace evidente que desde el inicio de la epidemia hasta la semana terminada el domingo 12 de Abril ($n = 40$) íbamos en ruta de transición desde la zona roja a la zona verde, pero aproximadamente desde $n = 40$ en adelante la tendencia cambió y ahora nos encontramos atrapados en la zona gris.

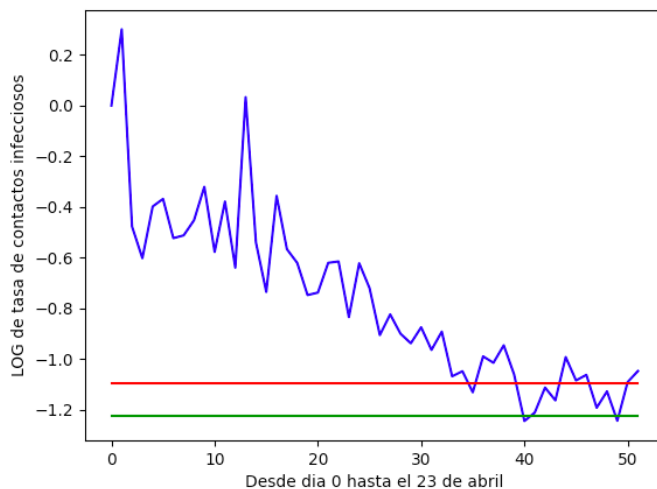
Se podría pensar que este análisis puede verse afectado por oscilaciones propias del sistema de diagnóstico. Para suavizar este efecto, podemos considerar el “promedio móvil cada 5 días seguidos” (*5-days rolling average*) de los datos del gráfico anterior. Se obtiene lo siguiente:



La tendencia es exactamente la misma: salvo irregularidades de los primeros días, hay un decrecimiento visiblemente lineal desde la zona roja en dirección a la zona verde, pero al llegar a la zona gris (aproximadamente $n = 40$, domingo 12 de Abril) se estanca.

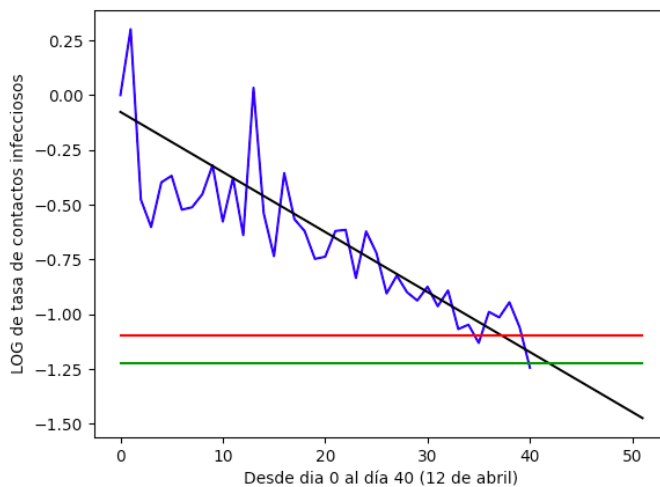
Observación. La información y gráficos presentados en esta sección son hasta el día de hoy (22 de Abril), calculados con cifras oficiales. No es una proyección, y no depende de ningún modelo matemático. En particular, hacemos hincapié en el hecho que el notorio quiebre del 12 de Abril ($n = 40$) en el comportamiento de la tasa de contactos infecciosos que se hace evidente en los gráficos no es una predicción matemática, sino que es una observación empírica sobre datos oficiales.

Nota agregada en edición. El estudio finalizó el 22 de Abril, pero el proceso de edición del artículo tomó hasta el 23 de Abril. En ese día aparecieron datos de $n = 51$ (23 de Abril). No incluimos los datos nuevos en el estudio porque de lo contrario el desfase se repetiría el día siguiente. Sin embargo, agregamos el gráfico actualizado de $k(n)$ con $n = 0, 1, \dots, 51$ en escala logarítmica para la información de la lectora. El nuevo dato del 23 de Abril es el segmento final de la línea quebrada azul, y vemos que se adentra en la zona roja.



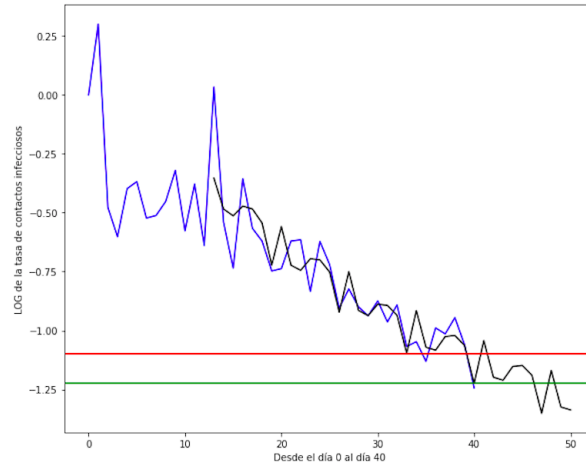
5. LO QUE PUDO SER... PERO NO FUE

De la sección anterior vemos que los logaritmos de la tasa de contactos infecciosos $k(n)$ desde el 3 de Marzo (día 0) hasta el 12 de Abril ($n = 40$) parecían tener un patrón bien definido de tendencia a la baja. Más precisamente:

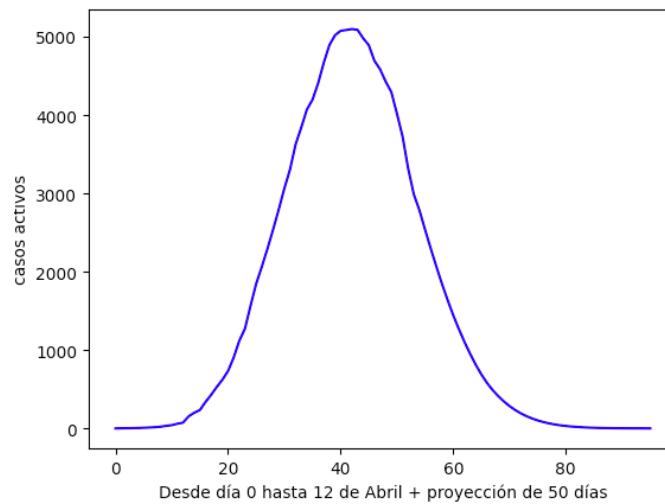


El gráfico tiene los logaritmos de la tasa de contactos infecciosos hasta el 12 de Abril ($n = 40$) sobrepuestos con la recta $y = -0,0274 \cdot x - 0,0766$ que se obtuvo por regresión lineal de mínimos cuadrados en el rango $n = 13, \dots, 40$. Partimos en $n = 13$ porque es el primer día en que hubo más de 100 casos, y es usual en el estudio de epidemias comenzar los análisis cuantitativos en ese punto.

Sin embargo, dada la importancia de la tasa de contactos infecciosos en el avance de la epidemia, no podemos desconocer que la recta no es un ajuste perfecto y existe un cierto comportamiento errático. Es necesario un análisis más sofisticado que una simple regresión lineal. Implementamos algoritmos de *machine learning*⁶ para reconocimiento de patrones y así modelar mejor este comportamiento. El resultado es la curva negra del siguiente gráfico sobrepuesta a los datos:

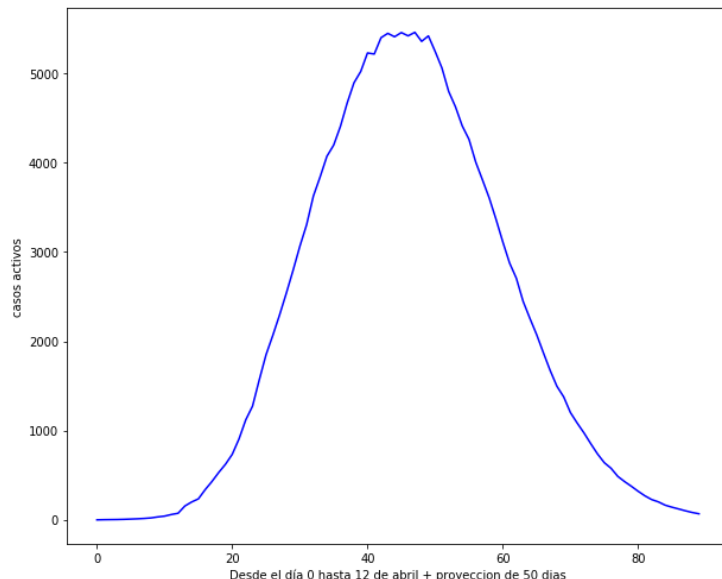


¿Qué hubiese ocurrido si después del 12 de Abril se hubiera mantenido esta tendencia a la baja? Implementamos nuestro modelo epidemiológico en dos casos para responder a esta pregunta, obteniendo los siguientes gráficos de proyección para el número de casos activos $I(n)$ con datos reales hasta $n = 40$ (12 de abril) y proyectando esta tendencia por 50 días. Primero, proyectando el avance de $k(n)$ con la regresión lineal para $\log_{10} k(n)$ mencionada antes:



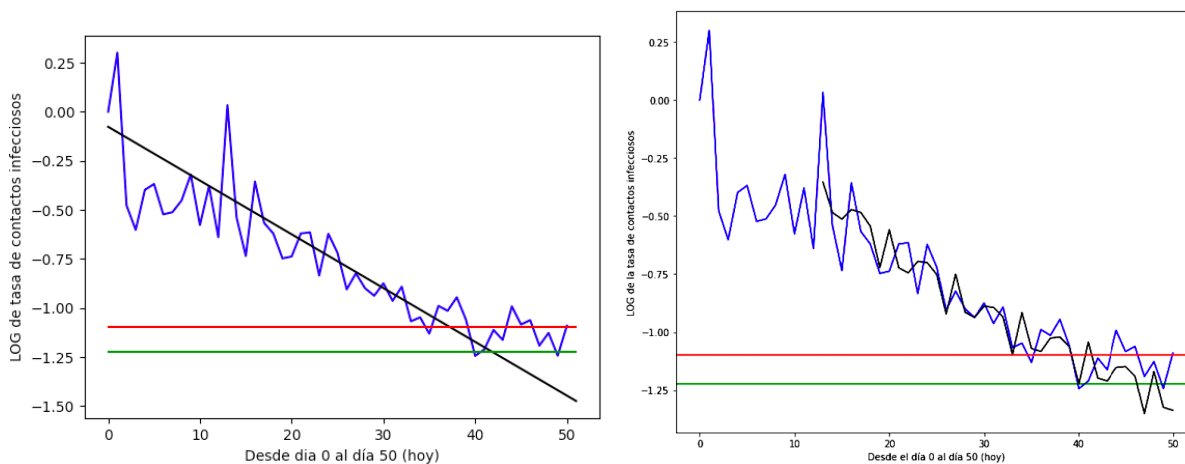
⁶Una rama de la inteligencia artificial particularmente útil en el reconocimiento de patrones. Se trata de métodos ampliamente utilizados hoy en día para entender la evolución de tendencias con regularidades complejas.

Y el siguiente gráfico para el número de infectados activos $I(n)$ es obtenido proyectando la evolución de la tasa de contactos infecciosos $k(n)$ con algoritmos de *machine learning*:



Como se puede observar, a pesar de la diferencia de ambos métodos, se llega a la misma conclusión: *si se hubiera seguido la tendencia de la tasa de contactos vista hasta el 12 de Abril, el avance de la epidemia se pudo haber frenado hasta niveles controlables en un plazo no mayor a fin de Mayo - inicio de Junio ($n = 80$), con un peak del orden de 5000 - 6000 casos activos.*

Sin embargo, este auspicioso escenario se nos escapó y **hoy ya no estamos en esa ruta**. El avance real de la tasa de contactos infecciosos $k(n)$ luego del 12 de Abril se desajusta bruscamente de las tendencias previas y parece estar estancado en la zona gris:



No sabemos por qué ocurrió este quiebre el 12 de Abril. En esos días (y los posteriores) hubo varios eventos que pudieron tener influencia: el éxodo desde Santiago por Semana Santa, el levantamiento de cuarentenas en algunas comunas de Santiago, y un cierto ambiente de triunfalismo. Sería muy interesante analizar si estos factores realmente son los responsables del cambio de tendencia de $k(n)$, pero un análisis de ese tipo requiere información de la cual no disponemos, y por ende no nos vamos a aventurar a realizar dicho análisis.

6. DETALLES TÉCNICOS

A continuación explicamos las principales modificaciones que hicimos al modelo SIR estándar para adecuar nuestras proyecciones a la realidad de los datos relativos al COVID-19 en Chile.

- La *tasa de remoción* γ utilizada en modelos tipo SIR es difícil de estimar en el escenario actual. Por eso, en su lugar utilizamos un parámetro de *delay* d que es el número de días promedio que dura la enfermedad. Si bien d no se conoce con exactitud, los análisis son bastante robustos en variación de este parámetro (ver comentarios sobre d más abajo).
- En otros estudios con modelos tipo SIR es común ajustar la tasa de contactos infecciosos β como una *constante numérica* sujeta a la condición que las curvas proyectadas por el modelo se ajusten a los datos observados. Sin embargo, consideramos que en la situación actual este paradigma es inadecuado: la tasa de contactos infecciosos *real* cambia violentamente en periodos breves. Por eso, hemos considerado la tasa de contactos infecciosos $k(n)$ como una función de la fecha, no como una constante numérica ajustable.

Ahora explicamos el cálculo de $k(n)$ usando datos reales. Sea H el número de habitantes de la comunidad estudiada, sea $S(n)$ el número de personas susceptibles el día n , sea $I(n)$ el número de infectados activos el día n , y sea $I^*(n)$ el número de personas que comienzan su infección el día n . Entonces, la definición precisa de la tasa de contactos infecciosos $k(n)$ es

$$k(n) = \frac{I^*(n) \cdot H}{I(n-1) \cdot S(n-1)}.$$

Los datos oficiales entregan diariamente el número $T(n)$ total acumulativo de casos al día n . También se entrega $I^*(n)$, pero es información redundante porque $I^*(n) = T(n) - T(n-1)$. Por otro lado, se entrega la cantidad de fallecidos y cantidad *estimada* de recuperados, de la cual se puede deducir una estimación de la cantidad de casos activos $I(n)$. Esto da todos los datos necesarios.

Naturalmente, por temas logísticos el cálculo está sujeto a algunas imprecisiones, tales como:

- Los datos de diagnósticos vienen desfasados. Pero esto no afecta las tendencias estudiadas, como se observa del promedio móvil en la sección 4.
- El número de recuperados entregado por el MINSAL es una estimación basada en aproximar $d = 14$. Esto no afecta aspectos cualitativos del análisis ni las fechas, lo cual verificamos re-calculando todo con varios valores de d , desde $d = 14$ hasta $d = 21$. Decidimos reportar el caso $d = 14$ para facilitar la comparación con cifras oficiales.
- El número de casos diagnosticados es menor al de casos reales por un factor del orden de 5 (esto queda en evidencia comparando tasas de mortalidad en Chile con las internacionales). Las cantidades $I^*(n)$ e $I(n-1)$ son homogéneas en este factor. Así, si $S(n)$ porcentualmente es similar a H , el error en la fórmula que define a $k(n)$ mayormente se cancela.

Concluimos que calcular $k(n)$ con datos oficiales da una buena estimación del $k(n)$ real.

El *forecasting* no-lineal de la sección 5 fue realizado por detección de patrones con técnicas de *machine learning*. Se implementó una combinación de dos métodos: de base se utilizó el procedimiento *prophet*, y sobre esto se utilizó *grid search* para optimizar sus hiperparámetros.

Finalmente, los experimentos aleatorios consistieron en simular $k(n)$ como una función estocástica a valores en distintos intervalos test. El propósito era capturar la irregularidad propia de la tasa de contactos infecciosos real observada en los datos oficiales, y evaluar cuales aspectos de las proyecciones eran robustos en estas irregularidades.

(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`