

Diagramas de Riesgo: *Significado, Implementación y Uso*

K. F. Laneri¹ y A. B. Kolton^{1,2}

¹ CNEA, CONICET, Centro Atómico Bariloche, (8400)
Bariloche, Argentina

² Instituto Balseiro (UNCu)

31 de julio de 2020

Resumen

En este documento se describe un índice de riesgo epidemiológico basado en la serie temporal de casos positivos diarios de COVID19 en una población dada. El índice de riesgo estima la cantidad de casos positivos esperables en los próximos días y puede usarse para estimar el número de cuidados intensivos adicionales que se necesitarán en los hospitales o el número de personas que requerirán de un seguimiento.

Número Reproductivo Empírico

En epidemiología matemática clásica se utiliza el Número Reproductivo Efectivo \mathcal{R}_t para medir la velocidad con la que se propaga una epidemia a un tiempo dado. Más específicamente, es una medida del número medio de personas infectadas por una persona infecciosa durante su período activo. Es decir, el número de casos secundarios por cada caso primario. A menudo se habla del número reproductivo básico \mathcal{R}_0 , que es el valor de \mathcal{R}_t a tiempo cero ($t = 0$), es decir, antes de que se inicie la propagación de la epidemia, sin intervención alguna, en una población susceptible en equilibrio. El número de reproducción empírico que definiremos, si bien está emparentado con el efectivo, representa en cambio el número medios de contagios diarios y monitorea la velocidad de contagio en función del tiempo durante la epidemia, es

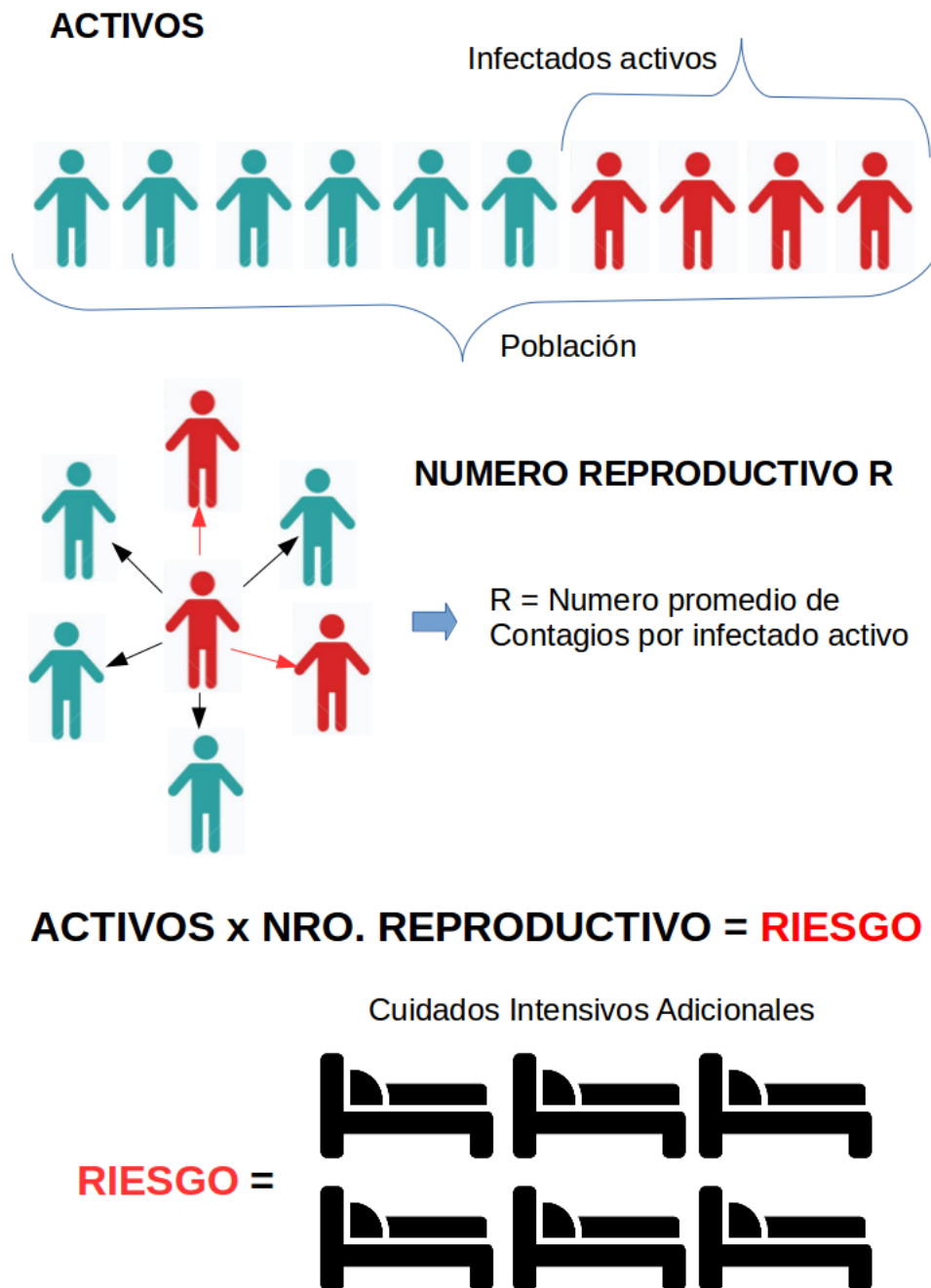


Figura 1: Explicación gráfica del uso del número de reproducción empírico y el número de casos activos para la predicción de un índice de riesgo sanitario.

susceptible a las intervenciones y a los distintos eventos que producen brotes epidémicos.

Para evaluar el Número Reproductivo Empírico, primero definimos R_t , que se estima a partir del número de casos positivos diarios reportados N_t , donde t es el número de día que corresponde al inicio de síntomas del caso positivo. Para eliminar fluctuaciones en N_t , usamos una ventana de casos de 3 días y la dividimos por la misma ventana corrida 5 días atrás:

$$R_t = \frac{N_{t-1} + N_t + N_{t+1}}{N_{t-6} + N_{t-5} + N_{t-4}} \quad (1)$$

El desplazamiento de ~ 5 días contempla, de forma empírica, el hecho de que una persona recién infectada (tenga o no síntomas) no infecta inmediatamente a otras personas, sino luego de un cierto número de días característico de la enfermedad. Es decir, la dinámica de contagio contiene tiempos de espera o demoras. Además estamos suponiendo que los casos secundarios $N_{t-1} + N_t + N_{t+1}$ tienen como casos primarios a $N_{t-6} + N_{t-5} + N_{t-4}$. El número R_t mide entonces la tasa de multiplicación de casos. Como sólo nos basamos en el número de casos positivos reportados, R_t es una medida cuantitativa para estimar cuántas personas en promedio arrojarán un resultado positivo en sus testeos de mañana por cada persona testada positiva hoy. Si suponemos que el número total de personas infectadas (incluyendo asintomáticos y sintomáticos no testados) es proporcional al número de casos positivos reportados, entonces el número estimado R_t vale para todo el conjunto de casos, aún los no testados.

En la práctica, R_t tiene fuertes fluctuaciones, especialmente en poblaciones no muy grandes o en poblaciones con una dinámica de contagio muy heterogénea. Las fluctuaciones pueden tener su origen en un gran número de factores, desde retraso en la carga de los datos o la falta de carga de la fecha de inicio de síntomas, hasta la ocurrencia de eventos sociales de distinta índole. Dichas fluctuaciones complican la evaluación de tendencias, y afectan la calidad de la predicción. Para minimizarlas realizamos a su vez un promedio del valor de R_t en una ventana de 7 días, definiendo el Número R_t^7 :

$$R_t^7 = \sum_{i=-3}^3 \frac{R_{t+i}}{7} = \frac{R_{t-3} + R_{t-2} + R_{t-1} + R_t + R_{t+1} + R_{t+2} + R_{t+3}}{7} \quad (2)$$

Efectos de Borde

La definición de R_t^7 tiene claramente un problema cuando t se acerca al presente, porque puede ocurrir que R_{t+3} , R_{t+2} , o R_{t+1} no puedan ser calculados, ya que necesitaríamos el reporte de casos futuros. Para minimizar este problema recurrimos a la estrategia de completar N_t , solo cuando sea necesario, usando extrapolación hacia adelante. Para ello calculamos el promedio de N_t en una ventana de 7 días y usamos ese valor para completar la serie N_t los días necesarios. Esta extrapolación solamente es necesaria hacerla cuando el tiempo considerado es tal que $t > hoy - 4$ días. Dicha corrección minimiza los efectos de borde pero no evita completamente que los valores R_t^7 para los tiempos más recientes sufran un ligero reajuste en los próximos días hasta estabilizarse en un valor definido cuando $t < hoy - 4$ días.

Número De Casos Activos

Si conocemos R_i^7 tenemos un índice que nos mide el número de personas que se infectan por cada persona infecciosa. Este número, multiplicado por el número de personas infecciosas hoy, nos da la cantidad de personas que se infectarán mañana. Pero no sabemos exactamente el número de personas infecciosas de hoy. Lo que sabemos es que el número de personas infecciosas hoy es aproximadamente proporcional al número de casos positivos de los últimos 14 días, porque 14 días es aproximadamente el tiempo de recuperación de un infectado.

La incidencia acumulada en los últimos 14 días A_t^{14} mide el número de personas cuyo testeo dió positivo en los últimos 14 días por cada 100 mil habitantes (la elección de 100 mil es arbitraria pero fija una escala de valores muy conveniente para la visualización y la comparación de poblaciones). Es una medida del número de casos positivos activos a tiempo t , ya que el período de recuperación es de aproximadamente 14 días. Este número se utiliza como una cota inferior de la población infecciosa, ya que no todos contagiados fueron testeados. La definición de A_t^{14} es entonces

$$A_t^{14} = \frac{100000}{N_{pob}} \sum_{i=t-13}^t N_i = \frac{100000}{N_{pob}} [N_{t-13} + N_{t-12} + \dots + N_{t-1} + N_t] \quad (3)$$

donde N_{pob} es el número de habitantes de la población a la que pertenece la serie temporal de casos N_t .

Indice de Crecimiento Potencial

Si supiéramos hoy el número de infectados que desarrollarán síntomas y darán positivo en unos pocos días, podríamos predecir aproximadamente los posibles nuevos tests positivos mañana multiplicándolo por el Número de Reproducción Empírico R_t^7 . Como no conocemos aquel número, lo estimamos a partir de la Tasa de Ataque A_t^{14} definida anteriormente. Definimos así el Índice de Crecimiento Potencial P :

$$P = A_t^{14} \times R_t^7 \quad (4)$$

Esta cantidad predice el número de nuevos tests positivos esperables en los próximos días a partir de los test positivos hoy, por cada 100 mil habitantes. Con esta predicción podemos hacer una predicción de riesgo epidemiológico para los próximos días.

Indices de Riesgo

El nivel de riesgo asociado a un valor de P está determinado por la capacidad del sistema sanitario. Si llamamos DTL al número de testeos diarios por cada 100000 habitantes, la nueva situación será de riesgo si el número predicho de nuevos test positivos es por ejemplo mayor que la capacidad de testeo $P > DTL$.

Por otro lado si llamamos C a la capacidad hospitalaria de cuidados intensivos por cada 100000 habitantes, la situación será de alto riesgo si P implica una cantidad de casos graves (de cuidados intensivos) que supere a C . Si llamamos f a la fracción de casos positivos que posteriormente desarrollan síntomas graves, esta situación peligrosa ocurre cuando $C < P * f$. El valor estimado a partir de estadísticas para Argentina (sin contemplar rango etario), nos da $f \approx 0,035$, es decir 35 de cada mil casos positivos, requieren un cuidado intensivo. En la figura 1 se resume en forma gráfica y sencilla esta idea.

Diagrama de Riesgo

Podemos graficar P , R_t^7 y A_t^{14} en un solo diagrama, en función del tiempo, con el objeto de visualizar la evolución temporal de los indicadores. Para ello en el eje vertical representamos R_t^7 y en el eje horizontal representamos A_t^{14} , de modo que la epidemia al tiempo t está caracterizada por un punto en el diagrama. Para representar el valor de P usamos una escala de colores de fondo.

Escenarios Posibles

- **Grave** Es claro que si $R_t^7 > 1$ y A_t^{14} son grandes la situación es muy grave (P es alto), ya que la predicción es de un alto número de casos positivos graves que podrían colapsar el sistema sanitario.
- **Moderadamente Grave** Si $R_t^7 > 1$ pero A_t^{14} es pequeño, de forma que P es pequeño, la situación es moderadamente grave, ya que se esperan pocos casos graves. Sin embargo de sostenerse una tasa de contagio alta, con $R_t^7 > 1$ por muchos días, el número de activos aumentará rápidamente, dando lugar a altos valores de riesgo P .
- **Moderadamente Grave** Si $R_t^7 < 1$ es pequeño pero A_t^{14} es grande, de forma que P es pequeño, la situación también es moderadamente grave, ya que se esperan pocos casos graves en los próximos días. Sin embargo si de repente $R_t^7 > 1$ en un determinado momento, el gran número de activos A_t^{14} provocará un brote dando lugar repentinamente a altos valores de riesgo P , lo cual podría colapsar en poco tiempo la capacidad sanitaria.
- **Seguro** La situación mas segura corresponde a $R_t^7 < 1$ y A_t^{14} . Aunque es evidente que las dos variables están conectadas, solo tenemos control directo sobre R_t^7 . Medidas de aislamiento, de seguimiento de contactos, así como la posibilidad de vacunas, apuntan a reducir R_t^7 , y a la larga disminuyen el riesgo P .

En la figura 2 se muestra un ejemplo, para la ciudad de San Carlos de Bariloche y para una de las comunas de la Ciudad Autonoma de Buenos Aires. Cada punto en los diagramas representa la situación un día dado, y el triángulo muestra el indicador calculado para el día más reciente que permiten los datos. El fondo de color representa el grado de riesgo, y para cuantificarlo la barra de color de la derecha hace una correspondencia aproximada entre el color y el número de ciudadanos intensivos adicionales predicho.

Es interesante observar la evolución temporal en el diagrama de riesgo. En la figura 2 observamos que en Bariloche, la epidemia esta controlada por una secuencia de brotes. En un dado brote, el número de activos aumenta por un tiempo pero luego se contiene. El número reproductivo baja por debajo de 1, y los activos empiezan a recuperarse haciendo que el número de activos disminuya y el punto se desplace a la izquierda. Un nuevo brote aparece cuando nuevamente $R_t^7 > 1$. En la comuna 1 de CABA se observan muchos menos de estos brotes y una estructura mas suave, debido al tamaño de la población. En los últimos días disponibles para cada población puede observarse que

Diagrama de Riesgo: San Carlos de Bariloche

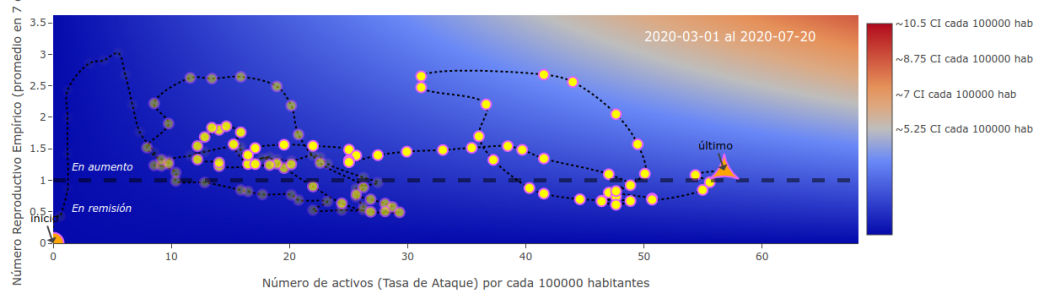


Diagrama de Riesgo: CABA Comuna 1

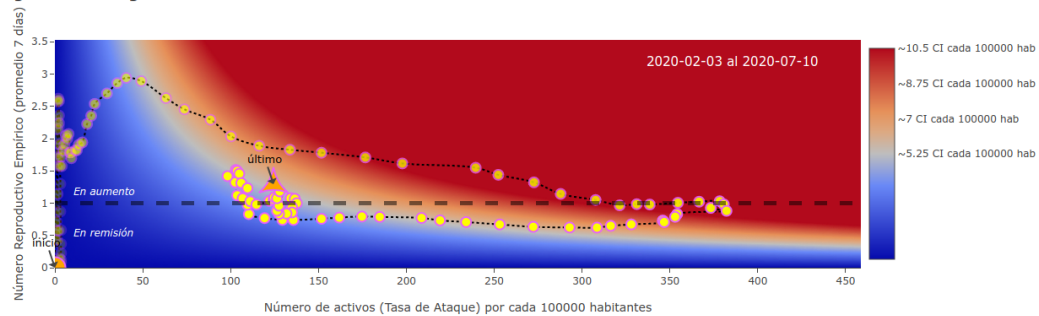


Figura 2: Ejemplos de diagramas de riesgo, para ciudad de San Carlos de Bariloche (arriba) y para la Comuna 1 (Retiro), de la Ciudad Autonoma de Buenos Aires (abajo).

mientras para Bariloche el número de cuidados intensivos adicionales es del orden de 2, para la comuna 1 de CABA es del orden de 5 por cada 100000 habitantes. Por supuesto, la gravedad de cada caso debe evaluarse de acuerdo a la capacidad sanitaria de cada población.

Datos

Todos los datos utilizados para todos los diagramas de riesgo son públicos y provistos por el Ministerio de Salud de la Nación. La calidad de los indicadores depende fuertemente de la calidad de los datos.

Una de las dificultades más importantes para poder hacer predicciones epidemiológicas es la carga lenta de los datos diarios. La figura 3 muestra un ejemplo, extraído de fuentes oficiales del Ministerio de Salud de la Nación (casos diarios). En las poblaciones referidas se puede observar una caída sistemática del número de casos diarios los últimos días. Esta caída es *artificial* y la misma estructura se repite día a día (es decir, solo se desplaza). Se debe a que las bases de datos no se actualizan rápidamente. En las poblaciones de la figura, se aprecia una demora de hasta 10 días.

Las fechas utilizadas corresponden a la *fecha de inicio de síntomas*. Sin embargo otra de las dificultades encontradas con los datos es que la fecha de inicio de síntomas no está siempre reportada. En esos casos se ha utilizado la fecha de apertura, que ocurre varios días más tarde, a veces hasta 10 días mas tarde. Este problema no está resuelto aún pero estaremos implementando próximamente una estimación estadística de la fecha de inicio de síntomas a partir de la de apertura.

Los datos poblacionales utilizados corresponden a la proyección del INDEC para 2020 y pueden encontrarse en

https://sitioanterior.indec.gob.ar/nivel4_default.asp?id_tema_1=2&id_tema_2=24&id_tema_3=119

Calidad de las Predicciones

El número de activos A_t^{14} se puede estimar, razonablemente, acumulando los casos diarios los últimos 14 días, ya que el tiempo de recuperación medio es de 14 días. El número reproductivo, que dice a cuantas personas contagia un infectado en su período infeccioso es sin embargo mucho más complicado. Depende del número medio de personas en contacto con un recién infectado (en el hogar, en el trabajo, en los comercios, etc.), del tipo y efectividad de ese contacto (tipo de actividad, precauciones tomadas, lugar del contacto, medidas de

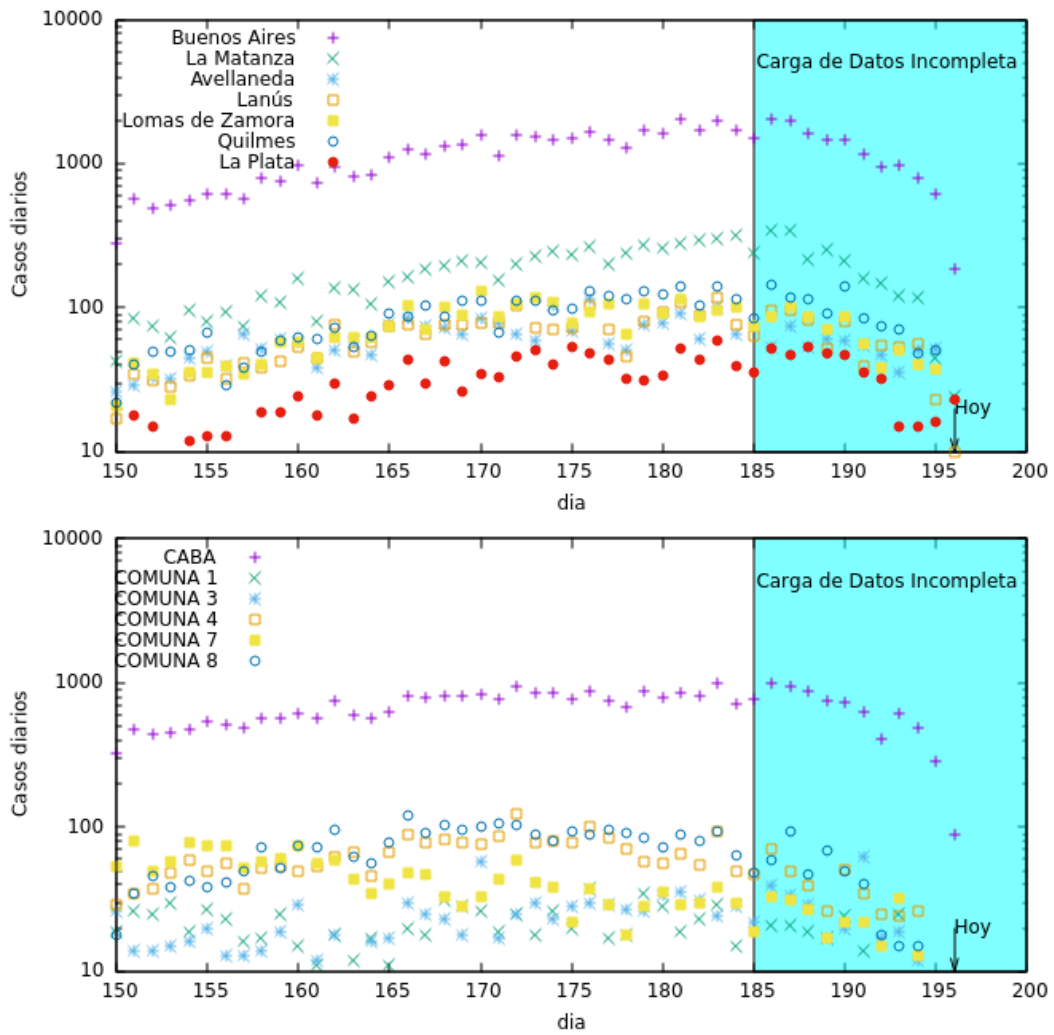


Figura 3: Ejemplos de series temporales públicas obtenidas del Ministerio de Salud de la Nación. Se indica una anomalía, una disminución aparente pero artificial del número de casos positivos los últimos días, que es resultado de la carga lenta de los datos de algunas poblaciones. Estas demoras, que pueden llegar a ser de diez días, son sistemáticas y nos impiden hacer predicciones precisas así como observar la dinámica más reciente.

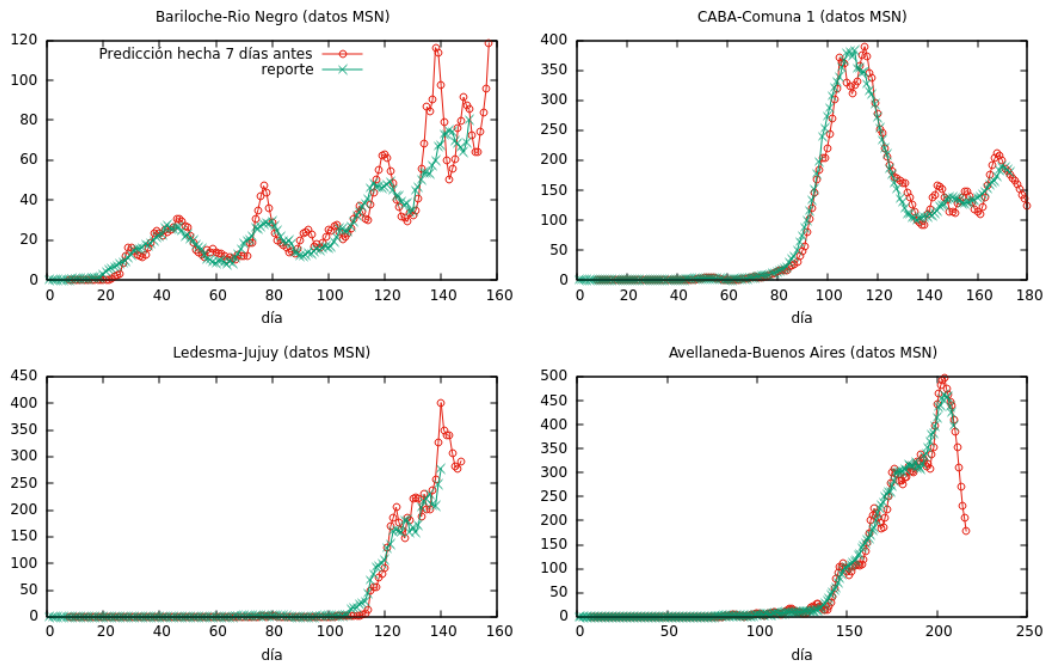


Figura 4: Comparación entre la predicción hecha una semana antes usando el predictor $P = A_{t-7}^{14} R_{t-7}^7$ (en rojo) y el número de activos observado A_t^{14} (en verde).

contención presentes, etc.), y del tiempo medio que el infectado efectivamente contagia. Este tiempo es en general menor que 14 días ya que sería de 14 días solo si el infectado siguiera su rutina normal luego de contraer la enfermedad. Sin embargo, la conducta de un infectado sintomático en general se altera, implicando, según la severidad de los síntomas, algún grado de aislación que cambia el tiempo medio de contagio. Debido a esta complejidad, para calcular el número reproductivo R_t^7 hemos adoptado un enfoque empírico bajo la hipótesis de que R_t^7 es una función de la historia del número de casos diarios. Investigaciones previas, realizadas con datos de diversas ciudades europeas, llegaron a la fórmula de la ecuación 2 como una buena aproximación a la función que optimiza la predicción para el caso específico de COVID19. En base a esta experiencia previa, adoptamos la misma función.

Para testear la calidad de la predicción, en la figura 4 comparamos el número de casos activos A_t^{14} día tras día durante un período de varios días, con la predicción realizada una semana antes $P = A_{t-7}^{14} R_{t-7}^7$. La comparación la realizamos para cuatro localidades de Argentina. Como se puede apreciar, en todos los casos la predicción es relativamente buena en promedio, aunque a veces es susceptible a grandes fluctuaciones implicando un gran error en la predicción.

Advertencia

La dinámica epidémica es muy compleja, y sería naïve pensar que un solo indicador (como el tiempo de duplicación) o dos indicadores como los aquí discutidos nos permitan capturar todos los aspectos relevantes. En particular, la epidemia es compleja debido al alto grado heterogeneidad espacial y temporal dentro de cada población, y entran en juego innumerables variables sociales, económicas, geográficas, sanitarias, etc. Es importante por lo tanto resaltar que es necesario trabajar con una variedad de indicadores seleccionados de manera que brinden, entre todos, una imagen completa del estado de situación.

Consultas

Para consultas relacionadas con los diagramas de riesgo, comunicarse con:

Dra. Karina Laneri
 Grupo de Física Estadística e Interdisciplinaria
 del Centro Atómico Bariloche
karinalaneri@gmail.com

Referencias

Estos cálculos fueron realizados a partir del siguiente tutorial

https://biocomsc.upc.edu/en/shared/avaluacio_risc.pdf

del grupo BIOCOMSC de la Universitat Politècnica de Catalunya. En particular, los diagramas de riesgo aquí descriptos fueron presentados en reportes diarios para otros países (principalmente poblaciones europeas) por el grupo citado anteriormente, y pueden consultarse en el siguiente link:

<https://biocomsc.upc.edu/en/covid-19>