

# Proyecto COVID-19. Estimaciones

María Andrea Domínguez      Eduardo Heilbron      Martín Lara  
Daniel López      Alejandro Ortiz      Juan Andrés Rincón  
Daniel Sánchez      Laura Tenjo

Estimar los modelos que aparecen en: <https://arxiv.org/abs/2004.04741> Para esta primera entrega deben presentar los modelos y la estrategia de estimación.

También espero que podamos estimar este para la entrega final: <https://la.mathworks.com/matlabcentral/fileexchange/74658-fitviruscovid19> y usando EM. Usar datos de mortalidad o infectados. Y usar esta metodología para estimar el subreporte: CFR\_Colombia\_underReporting\_25.4.2020 (en la carpeta compartida)

## 1. División de Trabajo

### 1.1. Estimar los modelos de Bertozzi et al. (2020) (4 personas)

#### 1. *Exponential Growth*

– Jorge Alejandro Ortiz

Cuando comenzó el proyecto se propuso con las demás personas que se encuentran trabajando el tema tres de estimaciones utilizar Rstudio cloud de manera que se pudiera colaborar con más facilidad y se pudiera concentrar todos los recursos en un mismo lugar. Con la aceptación de los demás miembros, se creó un espacio dedicado exclusivamente a replicar las estimaciones para este proyecto. A su vez, este espacio se dividió en tres entornos, el primero dedicado al numeral 1.1 de replicar el trabajo de Bertozzi et al., el segundo dedicado al numeral 1.2 de replicar las estimaciones de Mathworks, y el tercero dedicado al numeral 1.3 de replicar los modelos de subreportaje. Así mismo también se compartió un código para importar los datos, de manera que todas las personas utilicen la misma fuente y que siempre se utilice la versión más actualizada de los datos (actualizan diariamente). Estos datos se obtuvieron de Our World in Data quienes recopilan y distribuyen datos del Covid-19 de manera gratuita y cuya base está conformada por una combinación de datos descriptivos por país (como la población total, el número de camas de hospital por 100 mil personas, etc.) y datos particulares de Covid-19 (como el número total de infectados, muertos, recuperados, el número y tipo de pruebas realizadas, etc.).

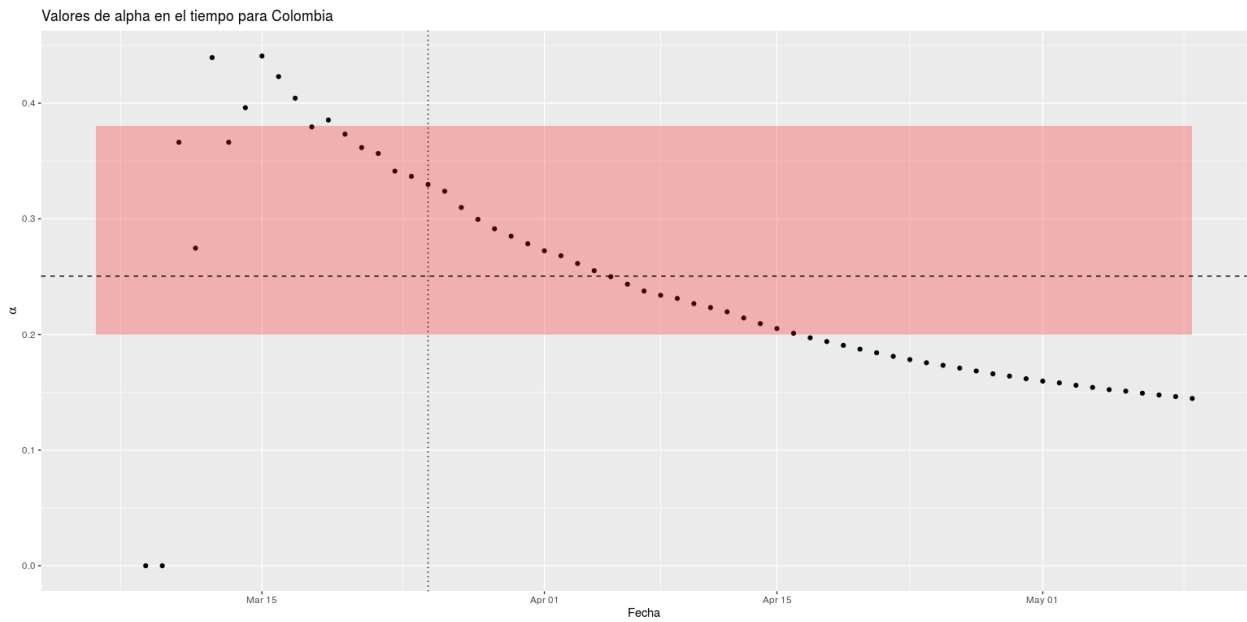
Luego de esto se procedió a trabajar directamente sobre el modelo de crecimiento exponencial. La ecuación fundamental por la que se rige el modelo exponencial es:

$$I_t = I_0 e^{\alpha t}$$

Donde  $I_t$  es el número de infectados en el momento  $t$ ,  $I_0$  es el número inicial de infectados, y  $\alpha$  es el parámetro de crecimiento. De esta ecuación es claro que:

$$\alpha = \frac{1}{t} \ln \left( \frac{I_t}{I_0} \right)$$

En el caso particular de Colombia el primer día que se detectó la presencia del Covid-19 solo se encontró una persona con un resultado positivo de esta enfermedad. Por lo que el valor de  $I_0$  sería uno, en ausencia de otras consideraciones. Así pues, utilizando la ecuación anterior el siguiente gráfico muestra como evoluciona el valor del parámetro de crecimiento en el tiempo con base en los datos de Colombia.



En la gráfica anterior, cada punto representa el valor de  $\alpha$  para ese día. La zona roja muestra el rango de valores estimados por los autores en el artículo para los estados de California Indiana y Nueva York. La línea rayada (horizontal) representa el promedio de los valores del parámetro de crecimiento excluyendo los dos primeros días (este valor es  $\sim 0.25$ ). La línea punteada (vertical) marca el inicio del aislamiento obligatorio a nivel nacional (24 de marzo 2020).

Como es claro del gráfico el valor del parámetro de crecimiento comienza muy elevado y va cayendo en el tiempo. Inicialmente esto se debe simplemente a una sobre estimación debido a que ocurren muchos contagios en pocos días relativos al

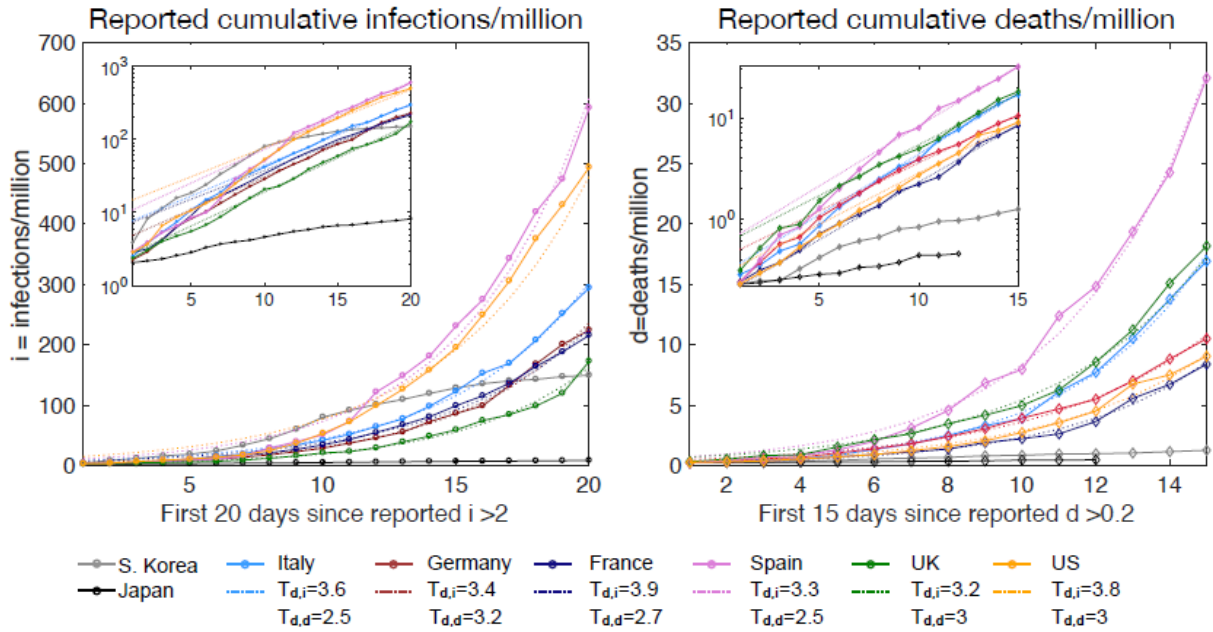
número de contagios anteriores. Si no se hubiera implementado ninguna medida de distanciamiento social se espera que este parámetro se ajuste a su valor poblacional. Sin embargo, debido a que se implementó aislamiento obligatorio en el país desde el 24 de marzo y en Bogotá desde el 20 de marzo es difícil controlar por el efecto de estas medidas sobre el parámetro de interés.

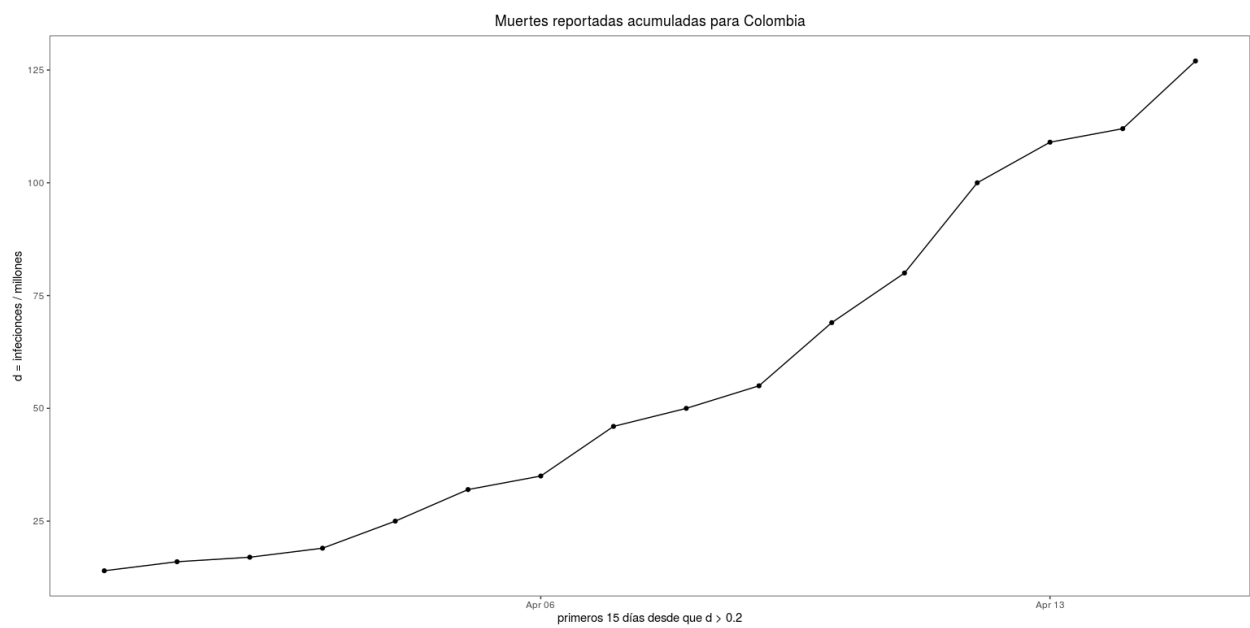
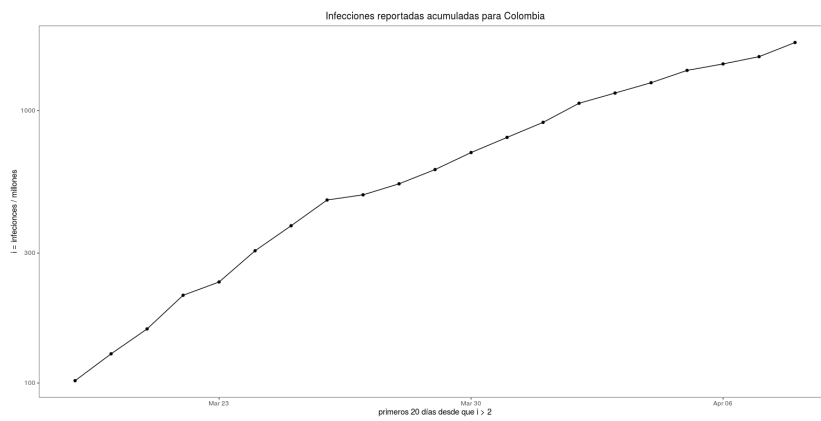
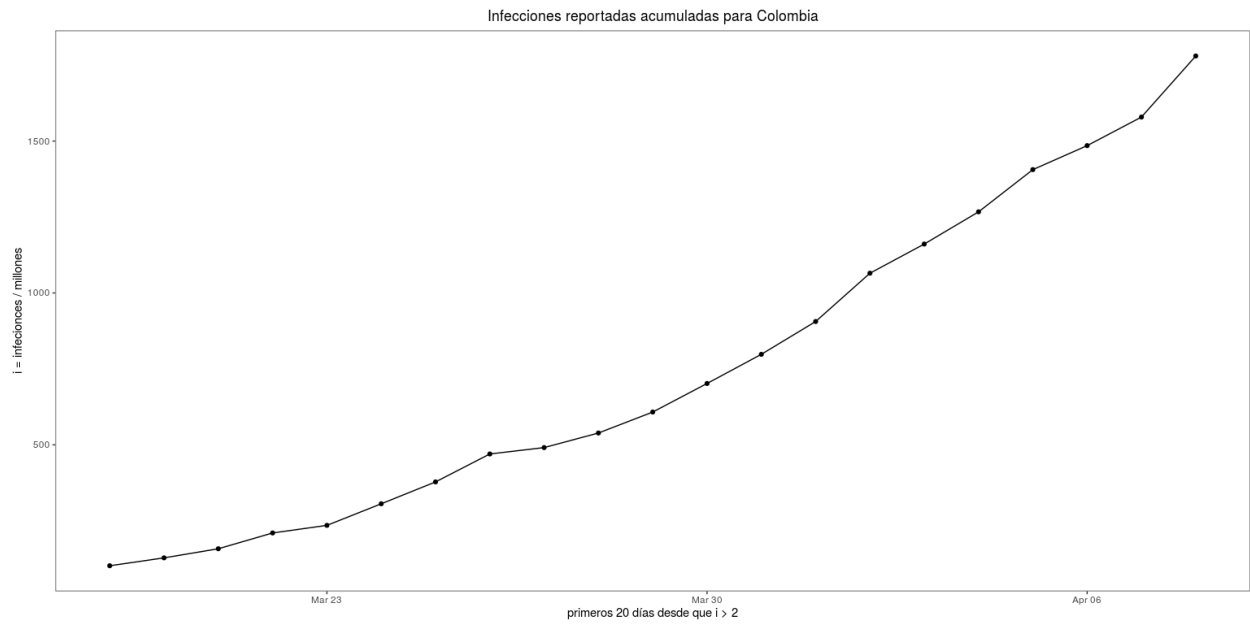
Adicionalmente también se estimó el modelo por MCO tomando provecho de que el modelo se puede escribir de la siguiente forma:

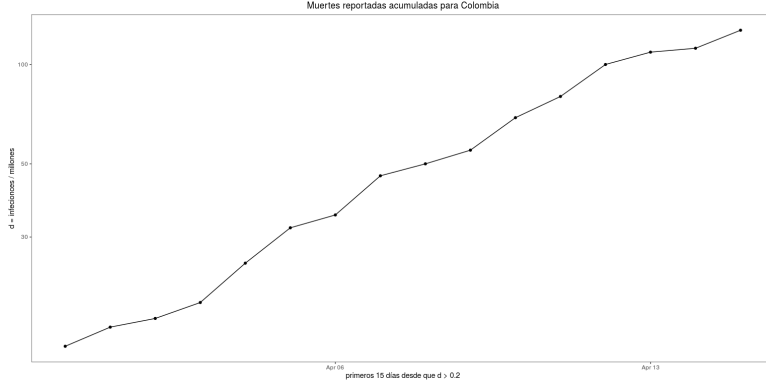
$$\ln \left( \frac{I_t}{I_0} \right) = \alpha t$$

Donde el término estocástico se puede adicionar al final del lado derecho de la ecuación. El resultado de este ejercicio da un valor de  $\alpha = 0.1845$ , altamente significativo (p-valor muy cercano a 0). Es necesario mencionar que se debe evaluar el desempeño de este modelo, así como de los supuestos subyacentes antes de poder implementarlo en cualquier tipo de análisis.

Luego de esto se buscó replicar la figura 1 del artículo y se obtuvieron los siguientes resultados:







La primera imagen corresponde a la figura 1 del artículo y las siguientes cuatro al caso de Colombia. Al igual que en artículo, las imágenes pequeñas se encuentran en escala logarítmica mientras que las grandes se encuentran en escala lineal.

#### Bibliografía:

Organization, W. H. (2019). World health statistics 2019: monitoring health for the SDGs, sustainable development goals. Switzerland: WHO.

Ritchie, H. (2020, 5 10). Coronavirus Source Data. Retrieved from Our World in Data: <https://ourworldindata.org/coronavirus-source-data>

## 2. *Self-exciting branching process*

- Daniel López
- Eduardo Heilbron

Otro modelo que fue estimado en el paper de Bertozzi et al (2020) es el Self-exciting branching process también conocido como el modelo de Hawkes. Es un proceso punto que permite estimaciones precisas de la tasa de infección y el tiempo de transmisión (Bertozzi et al, 2020). El modelo busca estimar la tasa de infecciones expresada como:

$$\lambda(t) = \mu + \sum_{t_i < t} R(t_i)w(t - t_i)$$

donde  $t$  es el tiempo actual y  $t_i$  son las veces con infecciones pasadas. Esto, en esta estimación, será el día actual y los días anteriores con infecciones.  $R(t)$  es la tasa de reproducción sin dimensiones del COVID-19, esto es el número de infecciones secundarias a partir de la infección primaria y está sujeta a cambios en el tiempo. Esta variable logra capturar los efectos de políticas gubernamentales como la Cuarentena. La segunda variable de importancia es  $w(t - t_i)$  que es la distribución inter-temporal entre eventos de infección Weibull con los parámetros  $k$  y  $b$ . Por ultimo,  $\mu$  es el promedio de casos de infección exógenos. Esto se entiende como el numero de casos promedio fuera de Colombia. A partir de la ecuación anterior, se

obtiene:

$$p_{ij} = \frac{R(t_j)w(t_i-t_j)}{\lambda(t_i)}$$

Esta es la versión branching del modelo y nos permite estimar la probabilidad que el caso primario haya causado el caso secundario. La tercera ecuación para estimar el modelo Self-Exciting Point Process fue presentada en Mohler Sledge (2020) y expresa  $R(t)$  como:

$$R(t) = \sum_{k=1}^B r_k I \{t \in I_k\}$$

Donde  $I_k$  son los intervalos discreteando el tiempo, B es el numero de dichos intervalos y  $r_k$  es la tasa estimada de reproducción en el intervalo  $k$ . Conociendo las características del modelo, la estrategia de estimación será la siguiente: Se estiman los parámetros  $k$  y  $b$  de la distribución Weibull ( $W$ ) en la base de datos de contagiados confirmados de Colombia utilizando el paquete de R `riskDistributions`. Luego se estima el parámetro  $r_k$  de la ecuación de  $R(t)$ . Con estos parámetros, se aplica el algoritmo EM (Expectation-Maximization) que calcula reiteradamente actualizaciones de los parámetros en dos pasos:

El Paso E donde se calcula:

$$p_{ij} = \frac{R(t_j)w(t_i-t_j)}{\lambda(t_i)} \text{ y } p_{ii} = \frac{\mu}{\lambda(t_i)}$$

El paso M donde se estima:

$$w(t) \sim MLE_{Weibull}(\{t_i - t_j; p_{ij}\}), \mu = \sum_i \frac{p_{ii}}{T}, r_k = \sum_{i>j} \frac{p_{ij} I\{t \in I_k\}}{N_k}$$

Donde T es el tamaño del periodo de observación, es decir, el numero de días del primer caso hasta hoy.  $N_k$  es el número de eventos en el intervalo  $k$  y la distribución Weibull se estima mediante MLE usando el numero de inter-eventos como observaciones y las probabilidades del proceso branching como ponderaciones. Con la aplicación del algoritmo se obtiene la tasa de infecciones y se puede estimar el día donde se verá el pico de infecciones.

Una extensión interesante de este modelo es la incorporación de un prefactor que incluye disminución de la proporción de la población susceptible al virus, logrando calcular las tendencias a largo plazo del virus. Esta extensión del modelo es caracterizada como:

$$\lambda^h(t) = (\frac{1-N_t}{N})(\mu + \sum_{t>t_i} R(t_i)w(t-t_i))$$

Se espera estimar ambas versiones del modelo y presentarlos gráficamente para observar los atributos más importante de la tendencia epidemiológica del COVID-19.

### 3. *Susceptible-Infected-Resistant (SIR)*

– Laura Daniela Tenjo Galvis

El tercer modelo estimado por Bertozzi et. al (2020) es un modelo de Susceptible-Infectado-Resistente (SIR). Este es uno de los modelos más utilizados para entender el comportamiento y la expansión del COVID-19.

El modelo SIR describe un modelo clásico en el que se divide la población en 3 grandes compartimentos. Al inicio de la pandemia se encuentran las personas Susceptibles a contagiarse (en este caso toda la población), luego una parte de las personas Susceptibles pasan a ser Infectadas para finalmente pasar al grupo de personas Resistentes. Es importante notar que en este modelo inicial las muertes se incluyen el grupo de personas resistentes ya que ya no son susceptibles de contagiarse ni pueden contagiar más personas. En este sentido la lógica del modelo se ve de la siguiente manera:

Susceptible  $\longrightarrow$  Infectada  $\longrightarrow$  Resistente

De manera más formal, el modelo supone una población de tamaño  $N$  en la que  $S$  es el número total de individuos susceptibles a contagiarse,  $I$  es el número de individuos infectados y  $R$  es el número de individuos resistentes. Como se mencionó anteriormente, para simplicidad del modelo, se toman las muertes como una submuestra de  $R$ . Esto es es razonable para un modelo con una tasa de mortalidad relativamente baja (Bertozzi et. al, 2020). Bajo este modelo, también se debe asumir un periodo relativamente corto de tiempo de manera que se asegure que las personas resistentes no se conviertan en susceptibles. lo que significa que la población se puede dividir de la siguiente manera:

$$S(t) + I(t) + R(t) = N$$

Teniendo esto en cuenta, las ecuaciones del modelo SIR son:

$$\begin{aligned}\frac{dS}{dt} &= -\beta \frac{IS}{N} \\ \frac{dI}{dt} &= \beta \frac{IS}{N} - \gamma I \\ \frac{dR}{dt} &= \gamma I\end{aligned}$$

La primera ecuación hace referencia a que el cambio en las personas susceptibles viene del contagio de personas susceptibles (S) por parte de personas infectadas (I)

multiplicado por la probabilidad de contagio o tasa de transmisión  $\beta$ . Este parámetro en la primera ecuación es negativo porque estas personas que se infectan salen del grupo de susceptibles. Siguiendo esta misma lógica estas personas que salen del grupo  $S$  ahora pasan al grupo de infectados que esta representado por  $\beta \frac{IS}{N}$  en la segunda ecuación. La última parte de esta ecuación viene de las personas que dejan de estar infectadas por lo que salen del grupo  $I$  a una tasa de recuperación  $\gamma$ . Finalmente, esto último explica la tercera ecuación que representa a las personas que estaban en el grupo  $I$  y luego pasan a los recuperados ( $R$ ).

De modo que, la solución de este sistema de ecuaciones diferenciales nos permite estimar los parámetros de interés  $\beta$  que es la tasa de transmisión constante,  $\gamma$  que es la tasa de recuperación constante y con estas dos podemos definir un parámetro  $\mathcal{R}_0 = \frac{\beta}{\gamma}$  que es el número de reproducción. Para poder resolver este sistema de ecuaciones se integran las 3 ecuaciones a partir de un valor inicial de  $S$ ,  $I$  y  $R$  en el momento cero ( $S(0) = S_0, I(0) = 0, R(0) = 0$ ).

En cuanto a la estimación del modelo SIR, el modelo se ajusta usando una estimación por máxima verosimilitud con una probabilidad Poisson. Luego se usa el Akaike como criterio para medir el desempeño del modelo. En particular, las estimaciones de Bertozzi et. al (2020) se encuentran usando una grilla de parámetros en la que  $I(0) \in [.005, -1]$ ,  $\mathcal{R}_0 \in [1.5, 5]$ ,  $\gamma \in [.01, .2]$ .

Además, también estiman una variación del modelo SIR que es el modelo SEIR. Este incluye una categoría adicional en la que se añade una parte de la población que se califica como expuesta:

$$\text{Susceptible} \longrightarrow \text{Expuesta} \longrightarrow \text{Infectada} \longrightarrow \text{Resistente}$$

Las ecuaciones que representan este modelo son:

$$\begin{aligned} \frac{dS}{dt} &= -\beta \frac{IS}{N} \\ \frac{dE}{dt} &= \beta \frac{IS}{N} - aE \\ \frac{dI}{dt} &= \beta \frac{IS}{N} - \gamma I \\ \frac{dR}{dt} &= \gamma I \end{aligned}$$

En esta nueva especificación, los parámetros  $\beta$  y  $\gamma$  tienen la misma interpretación del SIR solo que ahora se incluye  $a$  que representa el inverso del tiempo promedio de incubación. El SEIR, así como el SIR, se estima por máxima verosimilitud usando una distribución Poisson.

La ventaja de estos modelos radica en que permiten hacer un análisis del papel de los parámetros como una herramienta para proponer posibles soluciones que tengan en cuenta la dinámica de estos parámetros. Adicionalmente, el modelo SEIR tiene una ventaja sobre



el SIR ya que al incluir las personas expuestas, lo cual es aún más relevante bajo la medida de aislamiento obligatorio en la que una parte de la población está expuesta, permite asegurarnos que los estimados tengan mayor precisión.

Por último, una ventaja de estos modelos es que es posible reescalarlos en términos de una escala ( $\tau$ ) de tiempo para obtener un modelo sin dimensiones. De esta manera el modelo no estará en términos de  $t$  sino de  $\frac{t}{\tau}$  donde  $\tau$  es una constante con unidades de tiempo. Esto es beneficioso en la medida que proporciona una idea de cómo evoluciona el modelo con respecto a un tiempo característico.

## 1.2. Estimar los modelos de Mathworks (4 personas)

- María Andrea Domínguez
- Martin Lara
- Juan Andrés Rincón
- Daniel Sánchez

Pasar los algoritmos de MatLab a R y correrlos con los datos para Colombia. Hacer estimaciones para ciudades, municipios o departamentos sería una extensión buena.

En MathWorks se tiene un método principal para la estimación de la propagación del COVID-19 utilizando los datos para el número total de casos. Este utiliza el modelo SIR (susceptible-infected-removed) para dicha estimación. Adicionalmente, como supuestos toma una población constante, una mezcla uniforme para las personas y una recuperación igualmente probable para las personas infectadas.

En los métodos de MathWorks se tienen 3 funciones principales para la estimación y el análisis de la pandemia. En primer lugar, se tiene una función para la generación de datos que permite utilizar y convertir la información para cada país de número de infectados y fecha inicial que posteriormente serán utilizados en las demás funciones para la estimación de la pandemia. Adicionalmente, se tiene la función `fitVirusCV19` que cumple el objetivo de usar el modelo SIR para la estimación de los datos para cada país. Finalmente, se tiene la función `analyseCV19` encargada de graficar distintos parámetros encontrados para la pandemia como el número de reproducción básico (número de contacto), el número de reproducción, el tamaño, y otros parámetros del modelo como  $\beta$  y  $\gamma$ . Es importante mencionar que estas no son las únicas funciones presentes, también se tienen funciones adicionales como `calcR0` y `plotGrowthRate` que calculan el número básico de reproducción y grafican la tasa de crecimiento de la pandemia respectivamente.

Asimismo, la función `fitVirusCV19` está compuesta de una función principal y diversas funciones secundarias para la estimación de la pandemia. La función principal está diseñada para recibir como entrada los datos del país que se desea evaluar (nombre, vector de número de casos por día y fecha inicial de los datos), y da como salida una estructura de respuesta que incluye variables como el día que inició la pandemia, los parámetros del modelo, el número inicial de casos, el número final de casos, la predicción para el

incremento de infectados por día, la duración del periodo de aceleración y desaceleración, la duración de la pandemia, el último día de la pandemia, el número final de casos, la tasa de infección, entre otras. Adicionalmente, la función está programada para que se puedan indicar como entrada comandos para facilitar la visualización de los datos, aunque estos son corregidos o colocados automáticamente en caso de que no se indiquen o de que al indicarse no sean válidos.

En cuanto a las funciones secundarias, en primer lugar, se tiene una función para realizar la estimación de los parámetros, esta se obtiene a través de la minimización objetivo que corresponde a la suma del error cuadrático para las diferencias de los valores actuales y estimados, y residuales. De igual forma, se tienen funciones secundarias para el cálculo de variables como el número de individuos recuperados, número de casos en el punto de inflexión, el tiempo del pico de la epidemia, el tiempo de finalización de la epidemia, la densidad final, el punto de inflexión, el coeficiente de determinación, la estimación inicial para la regresión logística, y para la preparación de la función SIR, estas toman como entrada los parámetros calculados y su salida es la variable que calculan.

Las ecuaciones utilizadas por el modelo se muestran a continuación

$$\frac{dS}{dt} = -\frac{\beta}{N}IS \quad (1)$$

$$\frac{dI}{dt} = \frac{\beta}{N}IS - I\gamma \quad (2)$$

$$\frac{dR}{dt} = I\gamma \quad (3)$$

En donde  $t$  es el tiempo,  $S(t)$  el número de personas susceptibles,  $I(t)$  el número de personas infectadas,  $R(t)$  el número de personas recuperadas,  $\beta$  es la tasa de contacto y  $1/\gamma$  es el promedio del periodo de infección.

Adicionalmente, se calcula la constante para el total de la población de la siguiente forma

$$N = S + I + R \quad (4)$$

Las condiciones iniciales son  $S(0) = S_0, I(0) = I_0, R(0) = R_0$

Adicionalmente, se tienen las siguientes ecuaciones para el número de personas susceptibles

$$S = S_0 \exp\left[\frac{\beta}{N\gamma}(R - R_0)\right] \quad (5)$$

$$S_\infty = S_0 \exp\left[\frac{\beta}{N\gamma}(R_\infty - R_0)\right] \quad (6)$$

En donde la ecuación (6) hace referencia al número de personas susceptibles en el tiempo límite, y  $R_\infty$  hace referencia al número final de personas recuperadas.

Ahora, dado que el número final de personas infectadas es 0 se tiene:

$$N = S_{\infty} + R_{\infty} \quad (7)$$

Usando las ecuaciones (6) y (7) se obtiene

$$R_{\infty} = N - S_0 \exp\left[\frac{\beta}{N\gamma}(R_{\infty} - R_0)\right] \quad (8)$$

Ahora, es necesario estimar los parámetros  $\beta$ ,  $\gamma$ ,  $S_0$  e  $I_0$  de la información encontrada en la base de datos para cada país. Se debe tener en cuenta que  $R_0 = 0$  y  $I_0 = C_1$ . Es decir, en el tiempo inicial se asume un número de recuperados de 0 y un número de infectados inicial igual al número de casos registrados en el tiempo 0. Lo anterior tiene sentido pues en el tiempo inicial no ha pasado el tiempo suficiente para que el primer individuo infectado se haya recuperado.

Ahora, los datos disponibles se pueden representar como una serie de tiempo correspondiente al total de casos de la siguiente forma

$$C = I + R \quad (9)$$

Finalmente, se realiza la estimación de los parámetros y los valores iniciales a través de la minimización de la función objetivo dada por la diferencia del error cuadrático entre los casos reales y aquellos estimados por el modelo.

$$\text{mín } (C_t - \hat{C}_t(\beta, \gamma, S_0))^2 \quad (10)$$

Teniendo en cuenta lo anterior, en esta sección se busca replicar el modelo presentado anteriormente utilizando la herramienta estadística R, utilizando los datos para Colombia, tanto para todo el país como para departamentos, ciudades y municipios. Una primera aproximación para el problema se encuentra realizando la estimación del modelo con los métodos de MathWorks para los datos de Colombia, al hacerlos se encuentra lo siguiente.

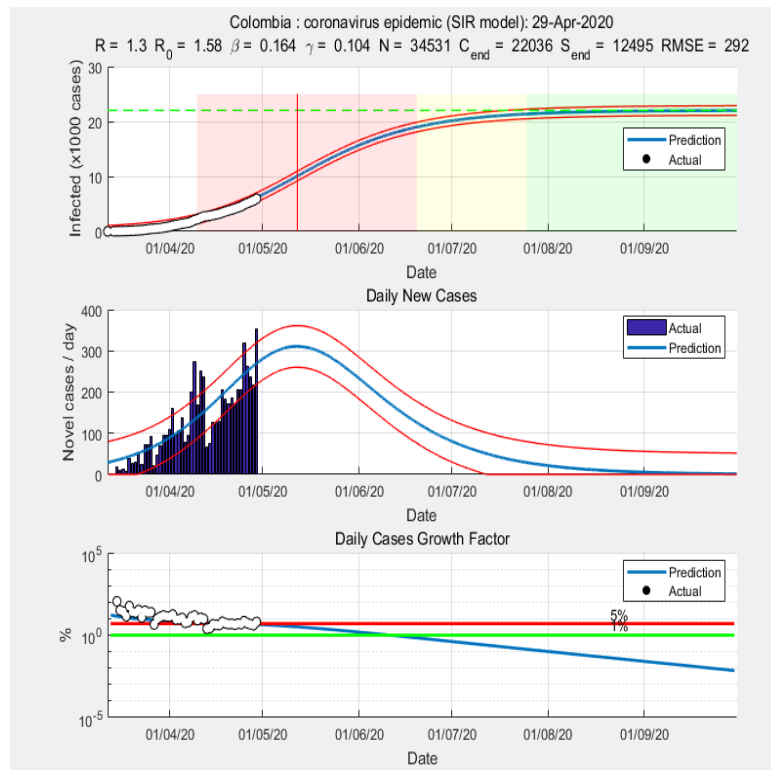


Figura 1: Estimación de COVID-19 en Colombia con el modelo SIR (función fitVirusCV19)

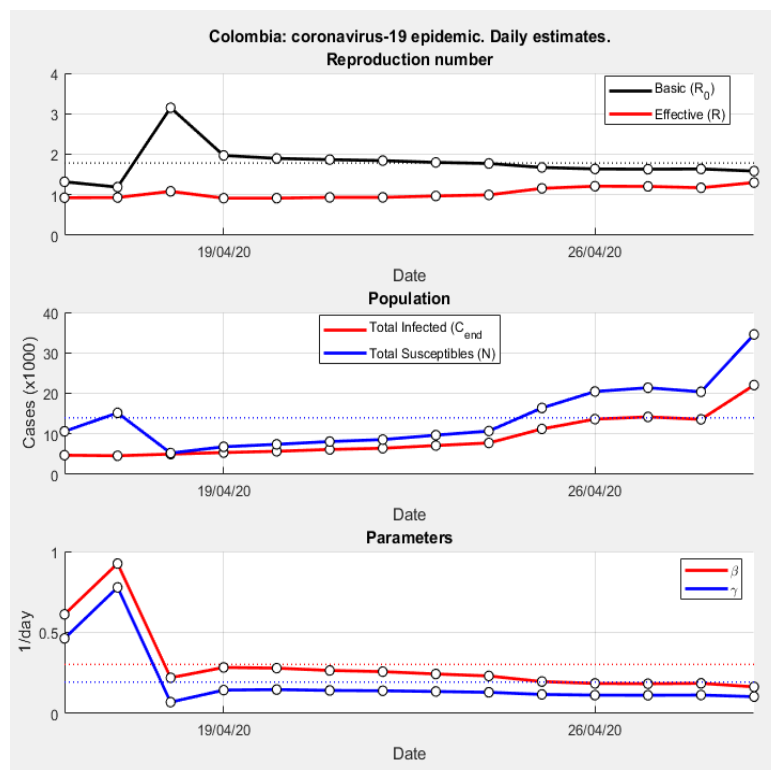


Figura 2: Número de reproducción y tamaño de epidemia de COVID-19 en Colombia con el modelo SIR (función analyseCV19)

Como puede verse, las funciones permiten encontrar el número de infectados en Colombia a través del tiempo. De acuerdo con las predicciones obtenidas en la figura 1, se espera que el número de infectados por cada 1000 casos se estabilice a partir de inicios de Julio. Adicionalmente, puede verse que el número de nuevos casos comienza a disminuir a partir de mediados de Mayo. Adicionalmente, la figura 2 muestra la evaluación de los parámetros  $R_0$ ,  $R_e$ , el total de infectados y susceptibles, y los parámetros del modelo. Los parámetros  $R_0$  y  $R_e$  son de suma importancia pues permiten representar la intensidad de la pandemia y se busca reducirlos al implementar políticas sanitarias. El parámetro  $R$  es el número de casos, en promedio, que van a ser causados por una persona infectada durante el período de contagio.  $R_0$  hace referencia al número básico de reproducción que representa el máximo potencial epidémico de un patógeno, es decir, es un escenario teórico que hace referencia a la situación en que una persona infectada entra en contacto con una comunidad totalmente susceptible. Por otro lado,  $R_e$  depende de la verdadera susceptibilidad de la población. Para los datos obtenidos en Colombia, puede verse que  $R_{efectivo}$  es menor que  $R_0$  pues tiene en cuenta personas inmunes al patógeno. Sin embargo, puede verse que ambos valores se encuentran en alrededor de 1.5 para inicios de Mayo, es decir que cada colombiano infectado a su vez infectará a 1.5 personas más. Se espera obtener resultados similares al implementar esta misma metodología en R.

### 1.3. Estimar los modelos de *underreporting* (2 personas)

Corrigiendo por la tasa de infectados como lo hacen las estimaciones de la U Nacional.

Parra et. al (2020) buscan estimar el porcentaje de casos sintomáticos de COVID-19 reportados en Colombia usando la tasa de mortalidad de casos estimada. Estas estimaciones se obtienen de los datos reportados diariamente por el Instituto Nacional de Salud (INS). El objetivo principal de estas estimaciones es corregir por los retrasos causados por el subreporte y los retrasos entre la confirmación de los casos y los reportes de fallecimientos.

La metodología se basa en la aplicada por Russel et. al (2020) en la que primero se hacen unas estimaciones ingenuos de la tasa de mortalidad. Esta se obtiene dividiendo el número de muertes a la fecha por el número de casos a la fecha. Claramente, esta primera estimación es una ingenua que lleva a resultados sesgados porque solo se incluyen los casos que han sido diagnosticados, lo cual omite la información del subreporte de casos (que puede estar causado tanto por demoras en las pruebas como por personas asintomáticas) y tampoco tiene en cuenta el retraso entre la confirmación de los casos y las muertes. El segundo paso consiste en ajustar el cálculo de la verdadera tasa de mortalidad para que incluya estos retrasos. Esto se logra mediante una estimación del número de casos esperados a partir de la distribución del retraso de confirmación de muertes.

## Referencias

Bertozzi, A. L., Franco, E., Mohler, G., Short, M. B., Sledge, D. (2020). The challenges of modeling and forecasting the spread of COVID-19. arXiv preprint arXiv:2004.04741.

Milan Batista. (2020). fitVirusCOVID19 (<https://www.mathworks.com/matlabcentral/fileexchange/fitviruscovid19>), MATLAB Central File Exchange. Retrieved May 10, 2020.

Parra, N., Vargas, V., Flórez, J. S., Árdila, L. Viviescas, C. (2020). Delay-adjusted casi fatality ration and under-reporting estimates in Colombia. Preliminary Communication.

Russell T W, Hellewell J, Abbott S, Golding N, Gibbs H, Jarvis C I, van Zandvoort K, CMMID nCov working group, Flasche S, Eggo R, Edmunds W J and Kucharski A J Using a delay-adjusted case fatality ratio to estimate under-reporting URL [https://cmmid.github.io/topics/covid19/global\\_cfr\\_estimates.html](https://cmmid.github.io/topics/covid19/global_cfr_estimates.html)