

TESIS CARRERA DE LICENCIATURA EN FÍSICA

INSERTE TITULO DE LA TESIS ACA

Denise S. Cammarota
Licencianda

Dra. Karina F. Laneri
Director

Miembros del Jurado
Inserte jurado aquí

26 de Enero de 2021

Física Estadística e Interdisciplinaria – Centro Atómico Bariloche

Instituto Balseiro
Universidad Nacional de Cuyo
Comisión Nacional de Energía Atómica
Argentina

A mi familia.

Índice de símbolos

COVID-19: Coronavirus Disease 2019

SARS-CoV-2: Severe Acute Respiratory Syndrome Coronavirus 2

OMS: Organización Mundial de la Salud

UCI: Unidad de Cuidados Intensivos

CABA: Ciudad Autónoma de Buenos Aires

SIR: Susceptible-Infectado-Recuperado (Susceptible-Infected-Recovered)

Índice de contenidos

| | |
|--|-----------|
| Índice de símbolos | v |
| Índice de contenidos | vii |
| Índice de figuras | ix |
| Índice de tablas | xi |
| 1. Introducción | 1 |
| 1.1. Breve línea temporal de la propagación. | 2 |
| 1.2. Datos oficiales de la República Argentina. | 5 |
| 2. Modelo semiempírico | 7 |
| 2.1. Número reproductivo empírico | 7 |
| 2.1.1. Efectos de borde | 8 |
| 2.2. Número de casos activos | 8 |
| 2.3. Predicción de casos | 9 |
| 2.4. Diagramas de Riesgo | 9 |
| 2.5. Conclusiones | 10 |
| 3. Análisis de las predicciones | 11 |
| 3.1. Análisis de los diagramas de riesgo y del número reproductivo \mathcal{R}_t^7 . | 11 |
| 3.2. Análisis del error en las predicciones. | 14 |
| 3.2.1. Fenómenos y exponentes críticos. | 14 |
| 3.2.2. Estudio de los histograma convencionales de error predictivo. | 15 |
| 3.2.3. Diseño y estudio de nuevos histogramas de error predictivo. | 18 |
| 3.3. Análisis del delay temporal óptimo. | 21 |
| 3.4. Conclusiones. | 22 |
| 4. Optimización de las predicciones | 23 |
| 4.1. Número reproductivo empírico. | 23 |
| 4.1.1. Optimización con 3 pesos. | 23 |

| | |
|--|-----------|
| 4.1.2. Optimización con 5 pesos. | 23 |
| 4.2. Optimización utilizando la librería SciPy. | 23 |
| 4.2.1. La función <code>curve_fit</code> de la librería SciPy. | 23 |
| 4.3. Optimización con redes neuronales utilizando la librería Keras. | 23 |
| 4.3.1. Breve introducción sobre redes neuronales artificiales. | 23 |
| 4.3.2. Estructura de las redes neuronales utilizadas en la optimización. | 23 |
| 4.3.3. Resultados. | 23 |
| 4.4. Conclusiones. | 23 |
| 5. Modelo SIR y SIR modificado | 25 |
| 5.1. Modelo SIR sin demografía clásico | 25 |
| 5.2. Modelo SIR sin demografía modificado | 27 |
| Bibliografía | 29 |
| Agradecimientos | 31 |

Índice de figuras

| | |
|---|----|
| 1.1. Diferentes síntomas de la COVID-19, en función del tiempo desde la exposición a la enfermedad y la edad de los pacientes. Adaptado de . . . | 2 |
| 1.2. Línea de tiempo mundial del origen y de la propagación de la COVID-19, adaptada de | 3 |
| 1.4. Casos confirmados acumulados cada 100 mil habitantes hasta el día 11 de enero de 2021, para cada provincia de Argentina. | 4 |
| 1.5. Nuevos casos reportados por día en CABA, algunas de sus comunas, la Provincia de Buenos Aires y algunos de sus municipios. En rojo, los últimos 10 días de datos. Se observa una caída aparente de casos en los últimos datos. | 5 |
| 2.1. Evolución temporal de casos acumulados A_t^{14} para la Argentina y para CABA. En línea punteada, la predicción para esta cantidad realizada una semana anterior. Se observa un buen acuerdo entre ambas cantidades. | 9 |
| 2.2. Diagramas de riesgo para la Argentina y para CABA. Las flechas indican las fechas de inicio y fin del registro de casos para la realización de estos diagramas. | 10 |
| 3.1. Diagramas de riesgo para Argentina y las provincias de Buenos Aires, Neuquén y Santa Fe. Las flechas indican el inicio y el último registro de casos en la región correspondiente. | 12 |
| 3.2. \mathcal{R}_t^7 para Argentina y las provincias de Buenos Aires, Neuquén y Santa Fe en función del tiempo. | 13 |
| 3.3. (a) Diagrama de riesgo y (b) Evolución de R_t^7 para la ciudad de San Carlos de Bariloche. | 14 |
| 3.4. Histogramas convencionales de error absoluto promedio en función de R_t^7 para Argentina y las provincias de Buenos Aires, Neuquén y de Santa Fe. | 16 |
| 3.5. Error absoluto promedio en función del valor R_t^7 central de cada bin de la Figura 3.4, en escala log-log para Argentina y las provincias de Buenos Aires, Neuquén y de Santa Fe. | 17 |

| | |
|--|----|
| 3.6. Histogramas modificados de error absoluto promedio en función de R_t^7 para Argentina y las provincias de Buenos Aires, Neuquén y de Santa Fe. Cada bin corresponde a un número constante de 20 observaciones. . | 19 |
| 3.7. Error absoluto promedio en función del valor R_t^7 central de cada bin de la Figura 3.6, en escala log-log para Argentina y las provincias de Buenos Aires, Neuquén y de Santa Fe. Cada punto de los gráficos corresponde a 20 observaciones. | 20 |
| 3.8. Estudio del error absoluto en función de r para diferentes localidades y provincias. | 21 |
| 3.9. Estudio del error absoluto en función de r , teniendo en cuenta todas las provincias y localidades a la vez. | 22 |
| 5.1. Diagrama de flujo representativo de un modelo SIR clásico. | 26 |

Índice de tablas

Capítulo 1

Introducción

La COVID-19 (Coronavirus Disease 2019) es una enfermedad causada por el virus SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2). Este pertenece a la familia de los *coronavirus*, que son conocidos por provocar infecciones respiratorias en animales y en humanos. Además del SARS-CoV-2, se pueden citar como ejemplos de coronavirus que infectan humanos el SARS-CoV (Severe Acute Respiratory Syndrome Coronavirus) y el MERS-CoV (Middle East Respiratory Syndrome Coronavirus), los cuales provocaron epidemias en los años 2002 y 2012 respectivamente. Estos tres virus son de origen zoonótico, y son capaces de provocar enfermedades severas e incluso fatales en los seres humanos.

Los síntomas más comunes de la COVID-19 son fiebre, tos seca y fatiga. Otros síntomas posibles incluyen pérdida de los sentidos del olfato y del gusto, congestión nasal, conjuntivitis y dolor de garganta. En casos severos, la enfermedad puede manifestarse falta de aire, pérdida del apetito y dolor en el tórax.

Según la OMS (Organización Mundial de la Salud), el virus se propaga de persona a persona a través de pequeñas gotas de la nariz o de la boca, que son emitidas por un individuo infectado (por ejemplo, al hablar, estornudar, toser o exhalar) y otra persona puede inhalar. Adicionalmente, estas partículas pueden caer en superficies u objetos cercanos. Entonces, otras personas contraen la enfermedad tocando estas superficies y luego llevándose las manos a los ojos, nariz o boca.

Un sujeto expuesto al virus comienza a manifestar síntomas en promedio de 5-6 días luego de la exposición, aunque puede variar en un rango de 1-14 días. De acuerdo con las evidencias recientes, todos los grupos etarios son susceptibles a contraer la enfermedad. Sin embargo, las manifestaciones clínicas difieren con la edad y el estado de salud de las personas. La mayoría de las personas menores de 50 años se recuperan tras esta etapa inicial. En las personas mayores de 60 años o con comorbilidades, se suelen observar manifestaciones severas como neumonía o falta de aire alrededor de 8 días luego de la exposición al virus. Durante esta etapa, puede ser necesaria la

hospitalización de la persona en la UCI (Unidad de Cuidados Intensivos) y ventilación mecánica. Un pequeño porcentaje de pacientes, usualmente mayores de 68 años o con comorbilidades, desarrollan un estado crítico que puede llevarlos a la muerte en un rango de 12-20 días después del primer contacto con la enfermedad. Este estado se caracteriza por síntomas como fallo multiorgánico o SDRA (Síndrome de Dificultad Respiratoria Aguda). Los diferentes síntomas de la COVID-19 y su relación con la edad de los pacientes se ilustra en la Figura 1.1 adaptada de

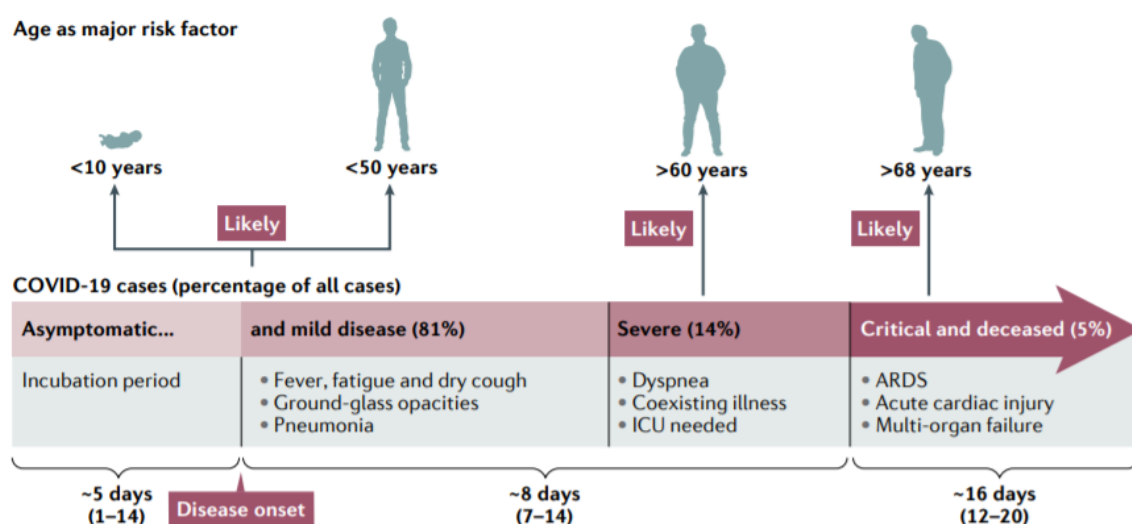


Figura 1.1: Diferentes síntomas de la COVID-19, en función del tiempo desde la exposición a la enfermedad y la edad de los pacientes. Adaptado de

1.1. Breve línea temporal de la propagación.

Los primeros casos de COVID-19 fueron reportados hacia finales de diciembre de 2019 por varios centros de atención médica de la ciudad de Wuhan, provincia de Hubei, China. Estos pacientes presentaban síntomas compatibles con una neumonía cuya causa era entonces desconocida, e inicialmente muchos de ellos estaban conectados al mercado *Huanan Seafood Market* en Wuhan. De acuerdo con estudios retrospectivos, los primeros casos podrían remontarse alrededor del 8-12 de diciembre del 2019.

El día 31 de diciembre de 2019, las autoridades sanitarias de Wuhan reportaron este brote a la OMS. A comienzos del año 2020, el *Huanan Seafood Market* fue cerrado para su sanitización, dada la continuada relación entre nuevos casos de neumonía y este establecimiento. Pronto, nuevos casos que no se conectaban con el lugar aportaron evidencia de una transmisión persona a persona.

El 8 de enero de 2020, se anunció que un nuevo tipo de *coronavirus* era la causa de la nueva enfermedad. Inicialmente, fue llamado 2019-nCov (2019 novel Coronavirus), pero posteriormente OMS cambió su nombre a SARS-CoV-2. Hacia fines de enero de 2020, se detectaron casos de COVID-19 en las 34 provincias de la República Popular

China, tras el aislamiento de la ciudad de Wuhan. Adicionalmente, se reportaron casos importados en países como Tailandia y los Estados Unidos de América. El día 30 de enero la OMS declaró el estado de emergencia sanitaria en todo el mundo. Los casos continuaron en ascenso en China hasta mediados de Febrero, y provocaron estrictas medidas sanitarias de las autoridades chinas, como el aislamiento. Luego, se observó una clara disminución de los mismos en ese país, indicando que el brote había sido controlado.

Sin embargo, la propagación internacional se aceleró, resultando en un ascenso rápido de los casos en Europa y en los Estados Unidos, a pesar de las medidas tomadas por las autoridades sanitarias de todo el mundo. El día 11 de Marzo la OMS declaró a la COVID-19 una pandemia mundial y el 13 del mismo mes declaró a Europa como el nuevo epicentro de la pandemia. Hacia el mes de julio, la OMS volvió a redefinir a América, y específicamente a Latinoamérica, como nuevo epicentro de los contagios. Una línea temporal de los acontecimientos se presenta en la Figura 1.2

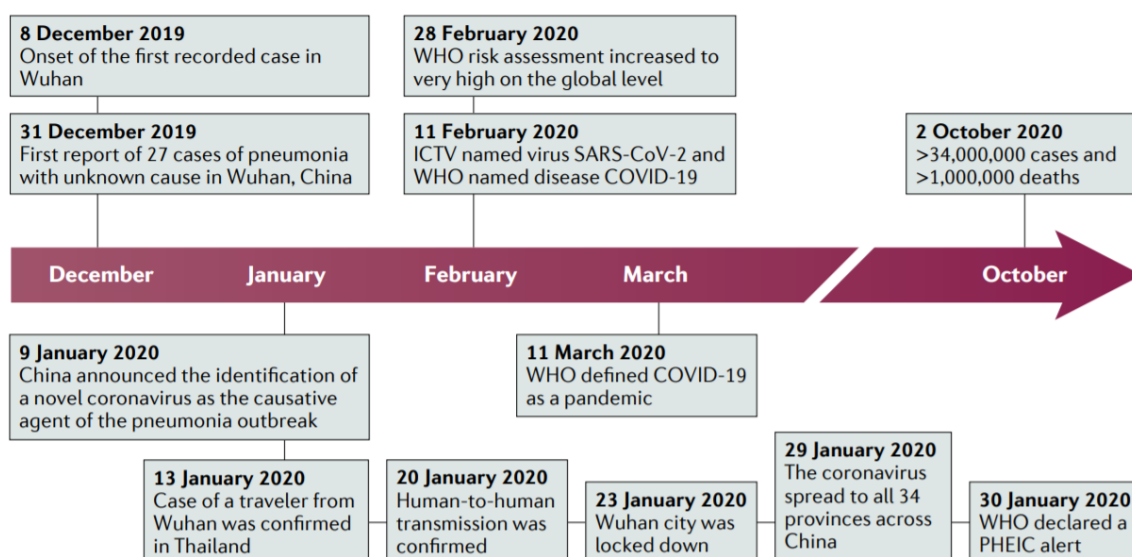
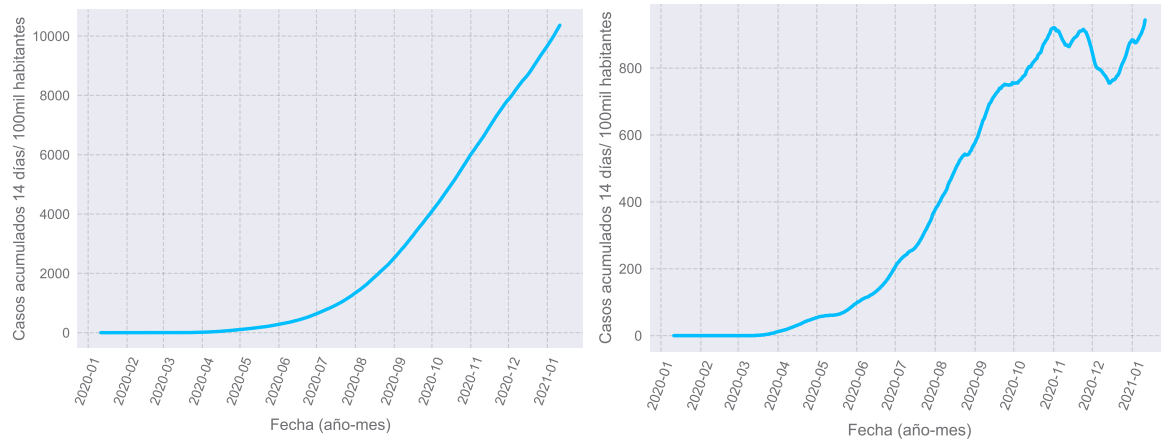


Figura 1.2: Línea de tiempo mundial del origen y de la propagación de la COVID-19, adaptada de

El primer caso registrado en Argentina fue reportado el día 3 de marzo de 2020 como un caso importado de Italia. Le siguieron más casos similares de viajeros que retornaban al país desde diferentes regiones de Europa, y de contactos estrechos con estas personas. La primera mención oficial de la existencia de transmisión comunitaria en nuestro país se hizo el 1 de abril de 2020. Como medida preventiva, se estableció el ASPO (Aislamiento Social Preventivo y Obligatorio) que rigió en la mayor parte del territorio nacional desde el 20 de marzo del 2020 hasta el 8 de noviembre del mismo año.

En Argentina, hacia el día 11 de enero de 2021, se han registrado cerca de 1.7 millones de casos y alrededor de 45 mil fallecimientos, en todas las provincias del país.

En la Figura 1.3a se presentan la cantidad de casos nacionalmente acumulados cada 100 mil habitantes en total, desde el comienzo de la pandemia y en una ventana de 14 días.



(a) Evolución temporal de los casos acumulados cada 100 mil habitantes. (b) Evolución temporal de los casos acumulados cada 100 mil habitantes en una ventana de 14 días.

Adicionalmente, en la Figura 1.4 se tiene un mapa de Argentina indicando la cantidad de casos confirmados cada 100 mil habitantes por provincias. Se observa que todas las provincias han tenido personas infectadas, siendo algunas más afectadas que otras.

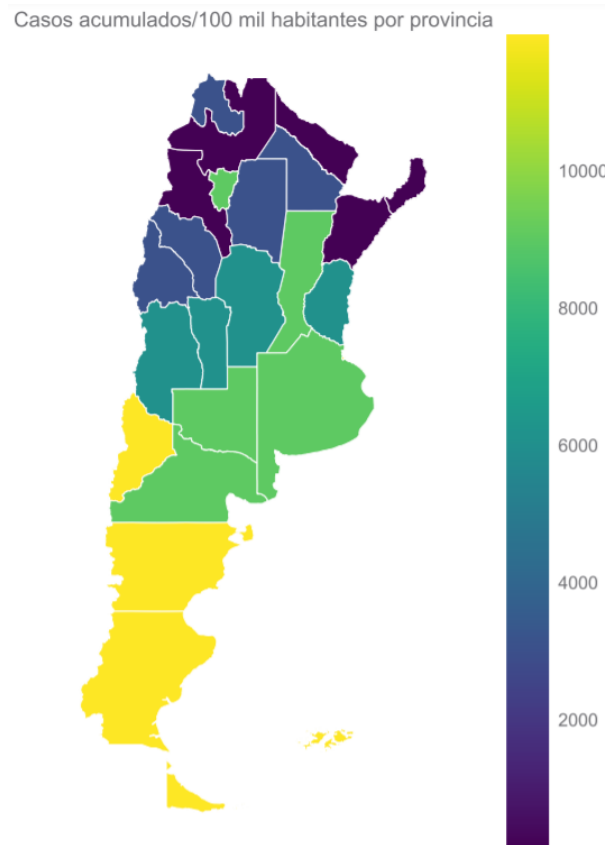


Figura 1.4: Casos confirmados acumulados cada 100 mil habitantes hasta el día 11 de enero de 2021, para cada provincia de Argentina.

A nivel mundial, al mes de enero de 2021, la enfermedad se ha propagado a más de 200 países y territorios de los 6 continentes. De acuerdo con la OMS, se han registrado alrededor 88 millones de contagios y 1.9 millones de muertes.

1.2. Datos oficiales de la República Argentina.

Los datos de test positivos en función del tiempo que se utilizan para la realización de esta tesis son calculados a partir de los datos públicos del Ministerio de Salud de la República Argentina. Estos son actualizados todos los días alrededor de las 20:00 horas, hora local en el sitio oficial. Estos incluyen todas las determinaciones realizadas en el país, junto con una fecha oficial de apertura del caso y la fecha de inicio de síntomas. En caso de que esta última se encuentre disponible, se la utiliza como el día a partir de la cual una persona está infectada y puede contagiar a otros. En caso contrario, se elige la fecha aleatoriamente en un rango de 8 días centrado en la fecha de apertura. Esta decisión se basó en la estrategia usada en ante la falta de un documento oficial que explicita el criterio para asignar una fecha de apertura.

Otra particularidad de los datos de Argentina es el retraso de origen administrativo en la carga de datos en varios municipios, que puede llegar a ser de alrededor de 10 días. Por ejemplo, en varias localidades de la provincia de Buenos Aires ó en la Ciudad Autónoma de Buenos Aires (CABA). Esto genera que se observe una falsa disminución de casos teniendo en cuenta los días más recientes. Esto se ilustra en la Figura 1.5, donde se presentan los casos confirmados diariamente en función del tiempo para diferentes municipios de CABA y de la Provincia de Buenos Aires respectivamente.

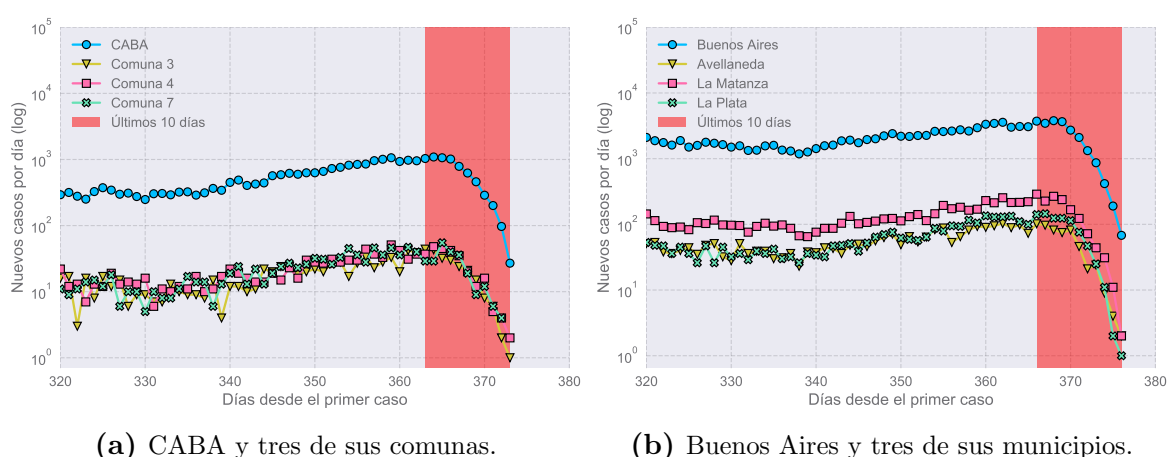


Figura 1.5: Nuevos casos reportados por día en CABA, algunas de sus comunas, la Provincia de Buenos Aires y algunos de sus municipios. En rojo, los últimos 10 días de datos. Se observa una caída aparente de casos en los últimos datos.

Capítulo 2

Modelo semiempírico

2.1. Número reproductivo empírico

En el modelado matemático de enfermedades infecciosas, el número reproductivo empírico a un tiempo determinado \mathcal{R}_t mide la cantidad de casos secundarios de enfermedad por cada caso primario. Es decir, es la cantidad de personas contagiadas por una persona infecciosa durante su período activo de contagio. De esta manera, da una idea de la velocidad de propagación de una epidemia en función del tiempo. Con frecuencia se hace referencia al número reproductivo básico \mathcal{R}_0 , que es el número reproductivo a $t = 0$, en una población susceptible en equilibrio.

Para la epidemia de la COVID-19, es útil definir un número reproductivo empírico según la ecuación:

$$\mathcal{R}_t = \frac{N_{t-1} + N_t + N_{t+1}}{N_{t-6} + N_{t-5} + N_{t-4}}, \quad (2.1)$$

donde N_t es el número de nuevos casos reportados correspondientes al día t luego del comienzo de síntomas del primer positivo detectado. El promediado de varios días antes y después del número t permite disminuir las posibles fluctuaciones en N_t .

Heurísticamente, la anterior definición puede entenderse de la siguiente manera: el desplazamiento atrás en el tiempo alrededor de una ventana de 5 día tiene en cuenta que las personas infectadas comienzan a contagiar a otros luego de un determinado período de tiempo tras su primera exposición al virus. Adicionalmente, se supone que estos casos tienen impacto sobre los nuevos casos reportados en los días $t - 1, t, t + 1$. Es decir, estos últimos son casos secundarios de los casos a tiempos $t - 6, t - 5, t - 4$.

Cabe aclarar que esta definición de \mathcal{R}_t está hecha en base a los casos positivos reportados, que es diferente que el número de nuevos casos totales. Así, en principio, permitiría estimar únicamente la cantidad de nuevos tests positivos a un determinado día en el futuro. Sin embargo, si se asume que el número total de personas infectadas es proporcional a N_t , entonces \mathcal{R}_t sirve para el conjunto de todos los casos. De esta

manera, se puede esperar que si $\mathcal{R}_t > 1$ la epidemia continuará, mientras que si $\mathcal{R}_t < 1$ la epidemia se extinguirá.

En la práctica, el número \mathcal{R}_t presenta fuertes fluctuaciones. Por ello, se define \mathcal{R}_t^7 como el promedio en una ventana de 7 días de esta cantidad, es decir que:

$$\mathcal{R}_t^7 = \sum_{i=-3}^3 \frac{R_{t+i}}{7} = \frac{R_{t-3} + R_{t-2} + R_{t-1} + R_t + R_{t+1} + R_{t+2} + R_{t+3}}{7}. \quad (2.2)$$

2.1.1. Efectos de borde

Se observa que las definiciones de \mathcal{R}_t y de \mathcal{R}_t^7 presentan problemas cuando se consideran días cercanos al presente. Esto se debe a que para calcular las cantidades mencionadas se necesitarían los valores de N_t para días $t \geq \text{hoy}$, donde *hoy* indica el día del ultimo reporte de casos. Con el objetivo de solucionar este problema, se calculan promedios de N_t a 7 días y se completa la serie. Este proceso se realiza hasta obtener estimaciones de N_t para $t \geq \text{hoy} - 4$. Así, esta estrategia permite solucionar el asunto de los efectos de borde. Sin embargo, es esperable que los valores de N_t y sus cantidades derivadas sufran reajustes a medida que se actualizan los datos, por lo menos hasta que $t < \text{hoy} - 4$.

2.2. Número de casos activos

El número \mathcal{R}_t^7 permite estimar el número de casos secundarios por cada infectado reportado. Así, si se conoce el número total de infectados en un día determinado, se pueden estimar los contagiados a un tiempo después multiplicándolo por \mathcal{R}_t^7 . Sin embargo, es imposible conocer con exactitud el número total de personas infecciosas a un tiempo determinado.

Una estimación posible para el numero de nuevos casos reportados cada 100 mil habitantes es la incidencia acumulada cada 100 mil habitantes en los últimos 14 días antes de t , denotada por A_t^{14} . Esta se define en la ecuación 2.3, y para calcularla se debe sumar el número de casos reportados en los 14 días anteriores al día que se considera. Se consideran 14 días como un número adecuado, dado que este es el período medio de recuperación de un infectado.

$$A_t^{14} = \frac{100000}{N_{pop}} \sum_{i=t-13}^t N_i = N_{t-13} + N_{t-12} + \dots + N_{t-1} + N_t. \quad (2.3)$$

donde N_{pop} es el número de habitantes de la población en estudio.

2.3. Predicción de casos

A partir de los cálculos de \mathcal{R}_t^7 y A_t^{14} para un día t pueden estimarse la cantidad de casos reportados en el futuro cercano, simplemente multiplicando ambas cantidades. En trabajos previos, se ha correspondido este producto a los nuevos casos del día $t + 7$. Es decir, que si se considera que:

$$\frac{100000}{N_{pop}} N_t = A_t^{14} \simeq \mathcal{R}_{t-7}^7 A_{t-7}^{14}, \quad (2.4)$$

el acuerdo entre predicción y los reportes de casos reales parece ser correcto empíricamente. La cantidad a la derecha de la igualdad $\mathcal{R}_t^7 A_t^{14}$ recibe en la bibliografía el nombre de índice de crecimiento potencial, y se denota con la sigla P_t .

Finalmente, dos ejemplos que ilustran la predicción de casos se presentan en la Figura Allí, se grafican los test positivos cada 100 mil habitantes hasta el 11 de enero de 2021 y la predicción efectuada una semana antes para Argentina en total y para CABA. En ambos casos, se señalan los últimos 10 días dado que los retrasos en la carga de datos pueden limitar la capacidad predictiva de las herramientas presentadas.

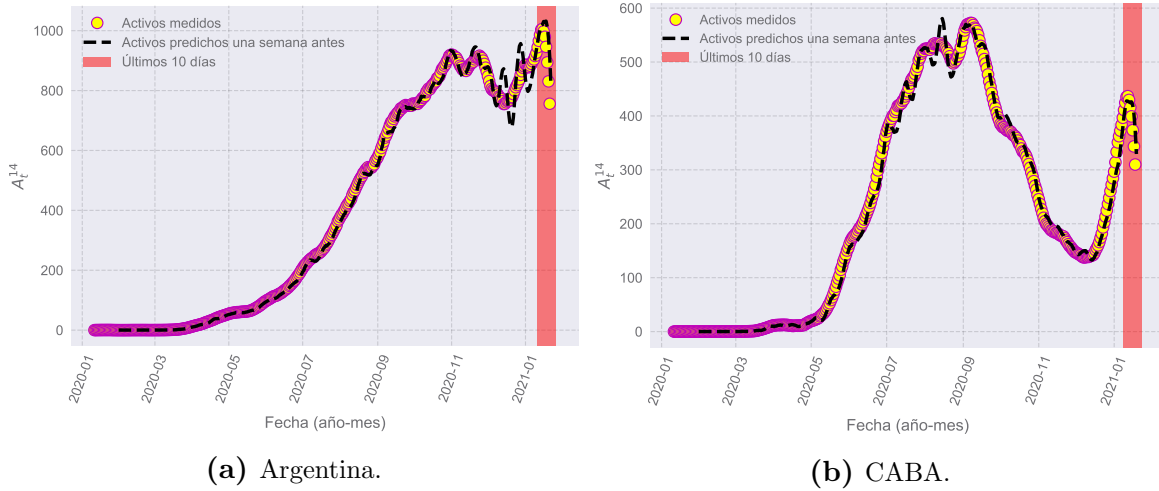


Figura 2.1: Evolución temporal de casos acumulados A_t^{14} para la Argentina y para CABA. En línea punteada, la predicción para esta cantidad realizada una semana anterior. Se observa un buen acuerdo entre ambas cantidades.

2.4. Diagramas de Riesgo

El riesgo sanitario en los próximos días puede caracterizarse a partir del índice P_t , que da una estimación de la cantidad esperable de test con resultado positivo cada 100 mil habitantes.

Por un lado, si DTL es la cantidad de test realizados cada 100 mil habitantes, entonces la situación será de riesgo si $P > DTL$. Lo mismo sucederá si la cantidad de personas que requieran atención médica en el futuro sobrepasa la capacidad del sistema

sanitario, que es limitada. Para estimar esta cantidad, se supone que una fracción f de los casos positivos que desarrollan síntomas graves y requieren atención en la UCI, y se considera que C es el numero de cuidados disponibles. De esta manera, se puede decir que hay riesgo cuando $P\dot{f} > C$. En el caso de Argentina, se puede estimar que $f \approx 0,035$, sin tener en cuenta rangos etarios.

Una representación grafica posible del riesgo de la situación sanitaria esta dada por los diagramas de riesgos. En estos, se grafica en el eje horizontal la incidencia A_t^{14} y en el eje vertical el numero reproductivo \mathcal{R}_t^7 . Adicionalmente, para representar P , se utiliza una escala de colores. Ejemplos de diagramas de riesgo se pueden visualizar en la Figura 2.2a , donde se muestran el diagrama de riesgo nacional y para CABA. Se observa que la carga incompleta de los datos en los últimos días crea una aparente evolución reciente hacia situaciones de menor riesgo.

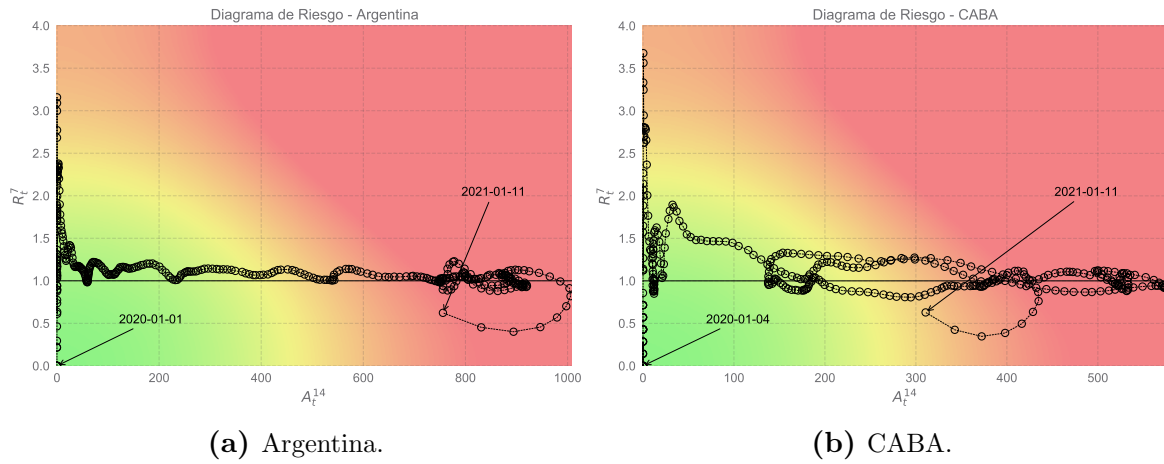


Figura 2.2: Diagramas de riesgo para la Argentina y para CABA. Las flechas indican las fechas de inicio y fin del registro de casos para la realización de estos diagramas.

2.5. Conclusiones

En este capítulo, se han introducido herramientas existentes que permiten generar predicciones aceptables para los nuevos casos reportados de la COVID-19 en el futuro cercano. Adicionalmente, se presentó el concepto de los diagramas de riesgo, que dan una idea cualitativa y cuantitativa del riesgo del colapso del sistema sanitario de una población. En próximos capítulos, nos dedicaremos a estudiar el error de las estimaciones resultantes, y posteriormente a intentar reducirlo. Finalmente, se han introducido y discutido algunas particularidades de los reportes de casos positivos de la Argentina que son importantes a la hora de evaluar la calidad de la predicción realizada.

Capítulo 3

Análisis de las predicciones

En el capítulo 2 presentamos las herramientas básicas desarrolladas en para la predicción de casos confirmados de la COVID-19. En este capítulo, dedicaremos la sección 3.1 a hacer algunas observaciones particulares sobre los diagramas de riesgo. Por otra parte, en la sección 3.2 utilizaremos diferentes estrategias para el análisis del error predictivo. Finalmente, en la sección 3.3 estudiaremos la existencia de una cantidad de días de delay óptima como posibilidad para mejorar las predicciones, sin modificar otros aspectos de los métodos del capítulo anterior.

3.1. Análisis de los diagramas de riesgo y del número reproductivo \mathcal{R}_t^7 .

En la sección 2.4 se introdujo el concepto de diagrama de riesgo, útil para evaluar la evolución temporal del riesgo de colapso del sistema sanitario de una población. En la Figura 3.1 se presentan diagramas de riesgo para la Argentina y algunas de sus provincias. En particular, la Provincia de Buenos Aires, la Provincia de Neuquén y la Provincia de Santa Fe. A diferencia de los ejemplos de la Figura 2.2, en esta ocasión se sustraen los 10 últimos días de datos para no tener en cuenta la carga incompleta de los mismos. En lo que continúa de esta tesis, a menos que se indique lo contrario, se consideran entonces datos hasta el 1 de enero de 2021 inclusive.

En estos diagramas de riesgo se evoluciona temporalmente de una situación de menor a una de mayor riesgo, como es de esperarse. Lógicamente, cuando inicia la epidemia, tanto R_t^7 como A_t^{14} son cercanos a cero. Por otra parte, hasta el último día de datos utilizados se distingue que todas las localidades permanecen en una región de riesgo medio o alto. En cambio, si la epidemia hubiese entrado en remisión, se distinguiría un acercamiento hacia el origen.

Una observación importante se puede hacer sobre cómo es la evolución temporal de estos diagramas. En los cuatro casos, el número R_t^7 es inicialmente muy elevado,

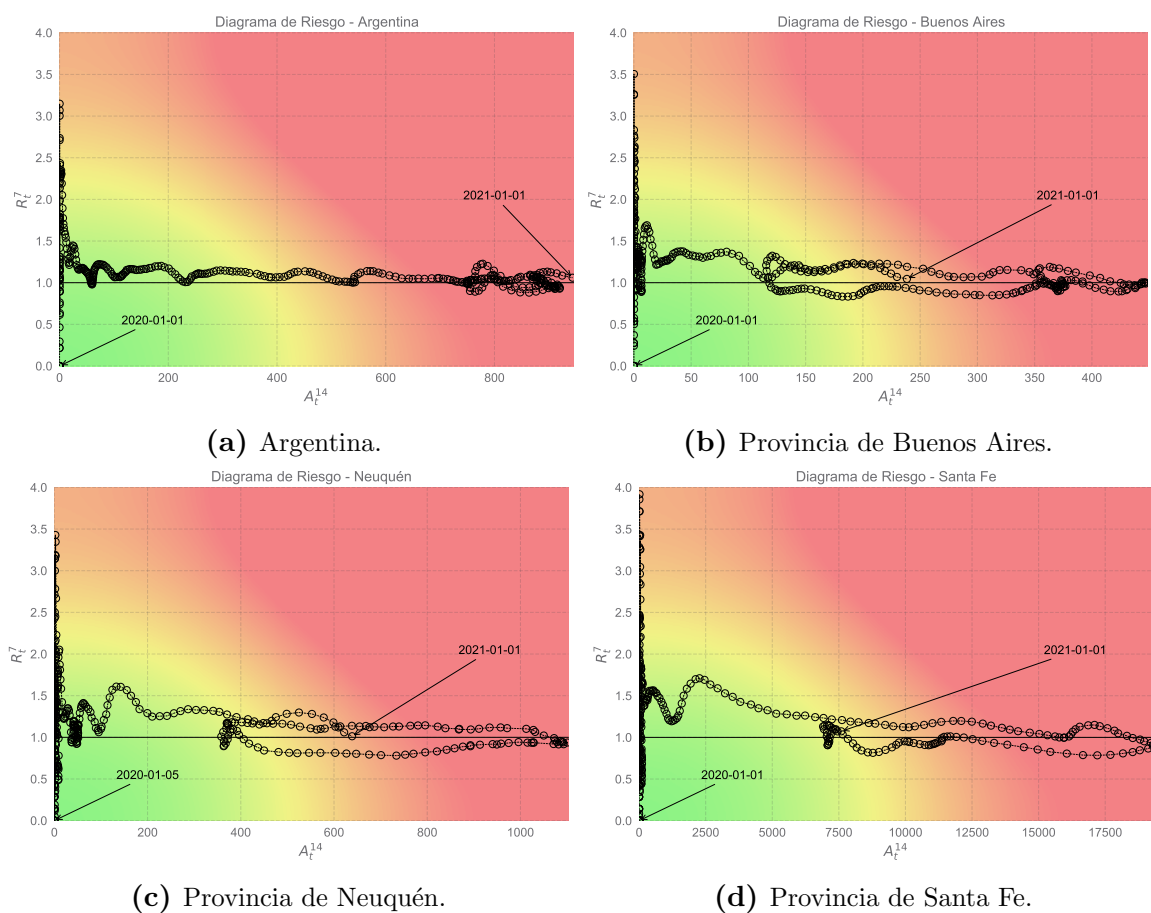


Figura 3.1: Diagramas de riesgo para Argentina y las provincias de Buenos Aires, Neuquén y Santa Fe. Las flechas indican el inicio y el último registro de casos en la región correspondiente.

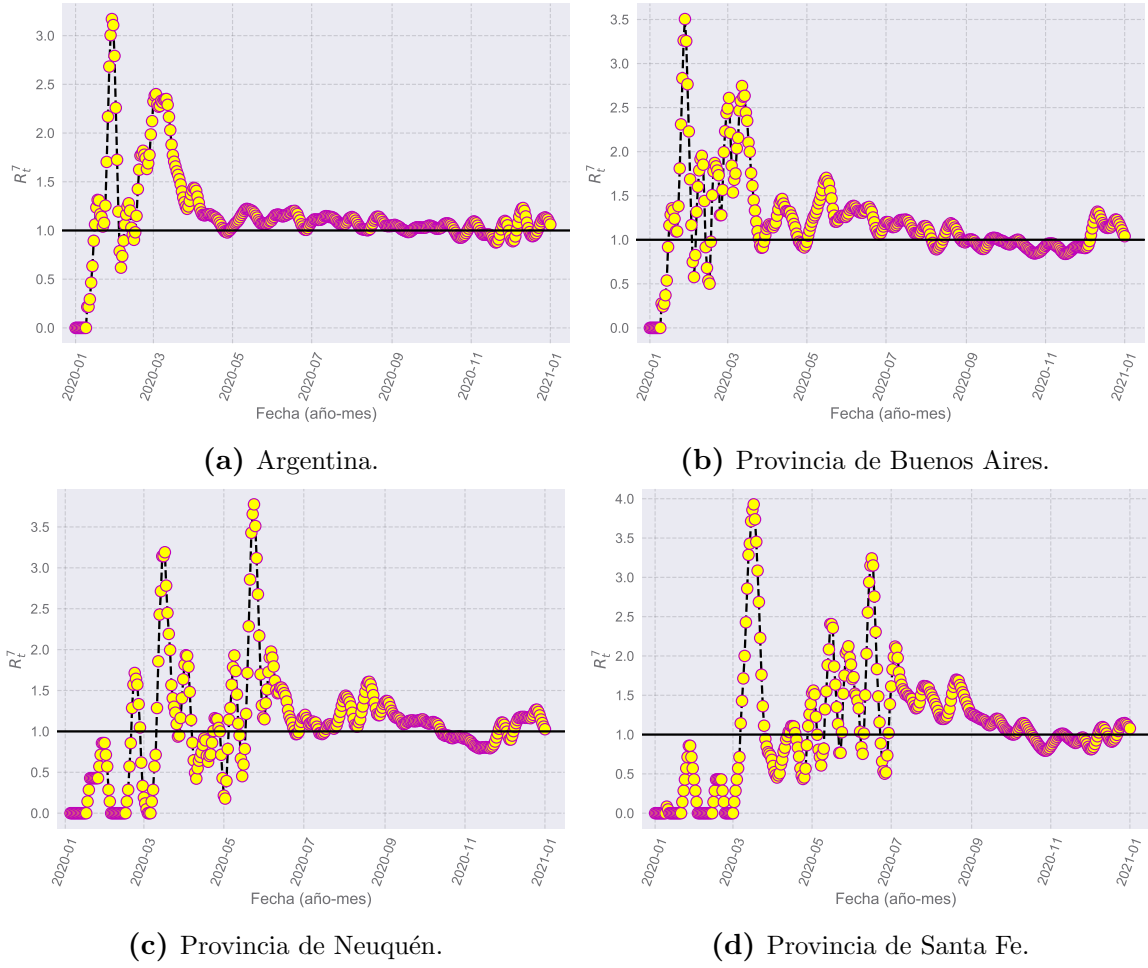


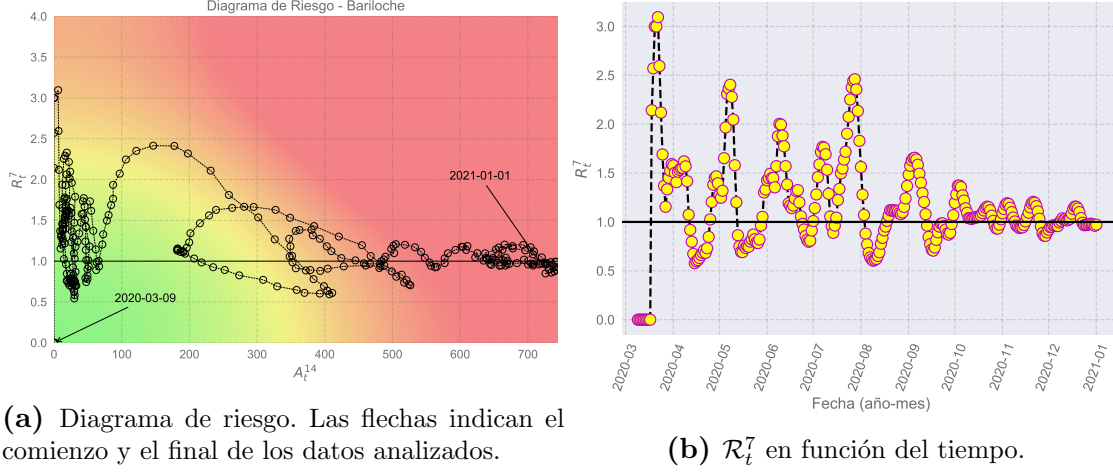
Figura 3.2: \mathcal{R}_t^7 para Argentina y las provincias de Buenos Aires, Neuquén y Santa Fe en función del tiempo.

llegando a valores cercanos a 3 – 3,5, pero luego se torna cercano a la unidad en mayor ó menor medida. Esto se repite en muchas provincias y localidades argentinas, no únicamente las aquí presentadas.

Esta observación nace de la simple inspección de los diagramas de riesgo, y no parece depender de los valores que toma la variable A_t^{14} . Es útil graficar la evolución temporal de este numero reproductivo. Esto se hace en la Figura 3.2 para las poblaciones mencionadas anteriormente. Es claro de aquí la evolución inicial no es equivalente en todos los casos. Por ejemplo, para Argentina y la Provincia de Buenos Aires, el período inicial de valores de \mathcal{R}_t^7 es más corto, y luego rápidamente $\mathcal{R}_t^7 \rightarrow 1$. Lo mismo no sucede para las provincias de Neuquén y de Santa Fe. Sin embargo, en todos los casos, eventualmente el número reproductivo se acerca oscilando en torno a la unidad.

No obstante, no todos los diagramas de riesgo son análogos a los presentados en 3.1. Particularmente, se distingue entre ellos el caso de la ciudad de San Carlos de Bariloche, en la Provincia de Río Negro. En la Figura 3.3a se puede observar su diagrama de riesgo, que presenta una dinámica con ciclos en el plano $\mathcal{R}_t^7 - A_t^{14}$. Esto se refleja en las

oscilaciones de R_t^7 en la Figura 3.3b. Allí, se tiene que R_t^7 toma el mismo valor por lo menos dos veces. Esto es necesario, pero claramente no suficiente para constituir una de las oscilaciones descriptas.



(a) Diagrama de riesgo. Las flechas indican el comienzo y el final de los datos analizados.

(b) R_t^7 en función del tiempo.

Figura 3.3: (a) Diagrama de riesgo y (b) Evolución de R_t^7 para la ciudad de San Carlos de Bariloche.

3.2. Análisis del error en las predicciones.

En esta sección, procederemos al análisis del error predictivo. Primero, hacemos una breve introducción sobre la teoría de los fenómenos y exponentes críticos. Luego, procederemos a analizar el error en búsqueda de algo similar a un exponente crítico. Para ello, se desarrollan dos propuestas en las secciones 3.2.2 y 3.2.3.

3.2.1. Fenómenos y exponentes críticos.

El estudio de los fenómenos críticos se ha centrado recientemente en la definición y el estudio de una serie de índices, llamados exponentes críticos. En primer lugar, para entender los exponentes críticos, es necesario definir la variable adimensional ϵ como:

$$\epsilon = \frac{T - T_c}{T_c}, \quad (3.1)$$

donde T es un parámetro y T_c es el valor que este toma en el punto crítico. De esta manera, ϵ permite medir la diferencia entre T y T_c .

Considérese una función $f(\epsilon)$ cuyo comportamiento se quiere describir cerca de $T \sim T_c$, es decir, de $\epsilon \sim 0$. Asumimos que f es positiva, continua para valores positivos suficientemente chicos de ϵ y que el límite:

$$\lambda = \lim_{\epsilon \rightarrow 0} \frac{\ln f(\epsilon)}{\ln \epsilon} \quad (3.2)$$

existe. Entonces λ se denomina el exponente crítico asociado a la función $f(\epsilon)$. De esta manera, lo suficientemente cerca del punto crítico se cumple aproximadamente que:

$$f(\epsilon) = A\epsilon^\lambda, \quad (3.3)$$

donde A es una constante.

En consecuencia, los gráficos $\log - \log$ de $f(\epsilon)$ en función de ϵ resultan ser líneas rectas, por lo menos lo suficientemente cerca de $\epsilon = 0$. Entonces, el exponente λ es fácilmente determinable.

3.2.2. Estudio de los histograma convencionales de error predictivo.

Con el objetivo estudiar el error, se realizaron histogramas del error absoluto en función de los valores de R_t^7 correspondientes, para diferentes localidades. Para ello, dada una población, se calculó para cada día t la predicción según la ecuación 2.4 y luego el error absoluto. Dado que correspondía con un día t , se lo asoció inmediatamente con el valor de R_t^7 en ese día. Luego, se dividieron los valores de R_t^7 en bins de 0,2 y se le asignó a cada uno de ellos el promedio del error absoluto para observaciones pertenecientes a cada bin. Los resultados de este procedimiento para Argentina y las provincias de Buenos Aires, Neuquén y de Santa Fe se presentan en la Figura 3.4.

Allí, se observa que el error es máximo para valores de R_t^7 cercanos a la unidad. Esto se repite en gran parte de las localidades analizadas. Sin embargo, se observa que no todos los histogramas son completamente iguales. Por ejemplo, existen localidades como la Provincia de Neuquén que presentan otros máximos locales del error absoluto promedio, para valores mayores de R_t^7 , como se observa en la Figura 3.4c.

Por otra parte, para contemplar la posibilidad de un fenómeno crítico, se graficó en escala log-log el error absoluto promedio en función de la distancia del valor central de cada bin a $R_t^7 = 1$, el cual podría funcionar como un parámetro crítico. Como ya se observó, cerca de este valor el error absoluto promedio es máximo. Adicionalmente, como se discutió en la sección 3.1, es el valor al que se acerca el número reproductivo a largo plazo.

En la Figura 3.5 se presentan los gráficos mencionados para las poblaciones de Argentina y las provincias de Buenos Aires, Neuquén y Santa Fe. En ellos, se indica el valor central del bin tal que el error es máximo en cada caso, al que se denomina $R_{t\max}^7$. Adicionalmente, se separan gráficamente los datos según si el valor de R_t^7 es mayor o menor a $R_{t\max}^7$.

De la inspección de la figura anterior, no se desprende un comportamiento similar al esperado de los fenómenos críticos descritos en 3.2.1. Claramente, la relación entre

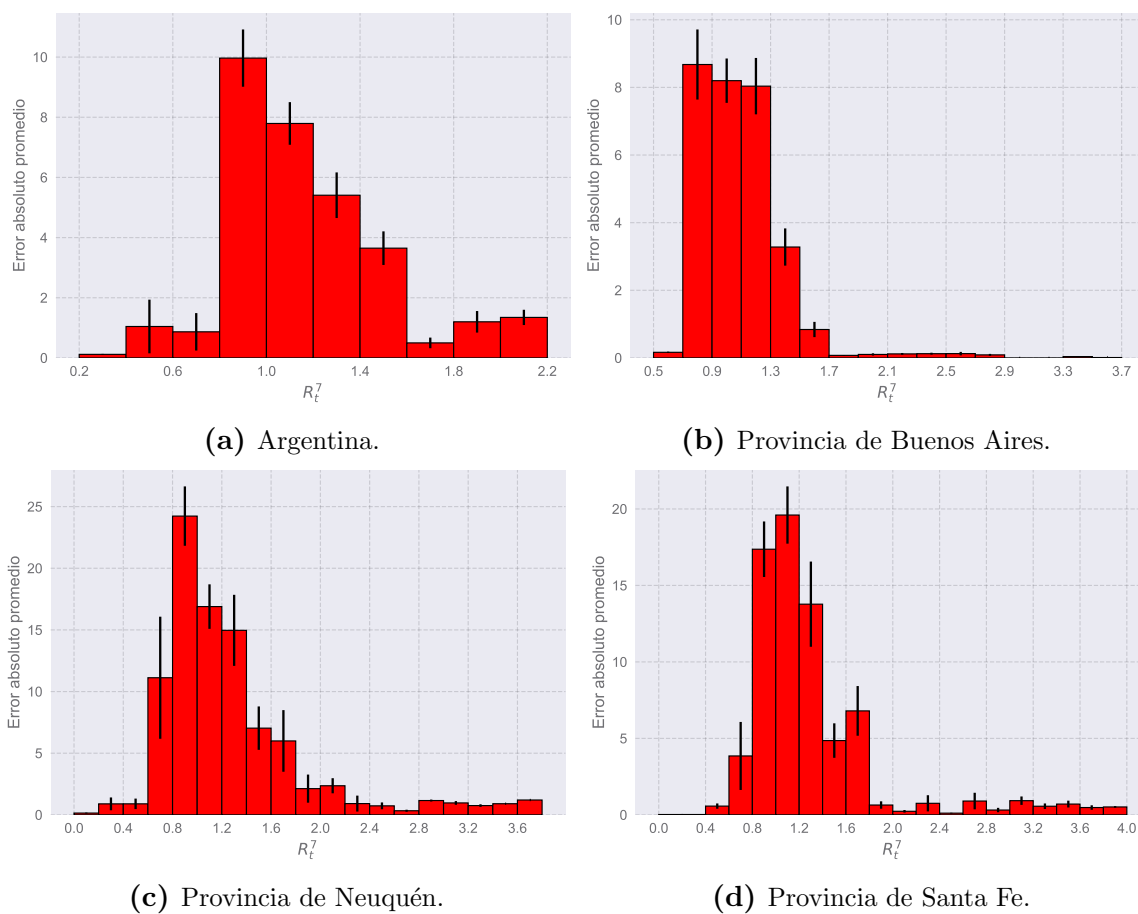


Figura 3.4: Histogramas convencionales de error absoluto promedio en función de R_t^7 para Argentina y las provincias de Buenos Aires, Neuquén y de Santa Fe.

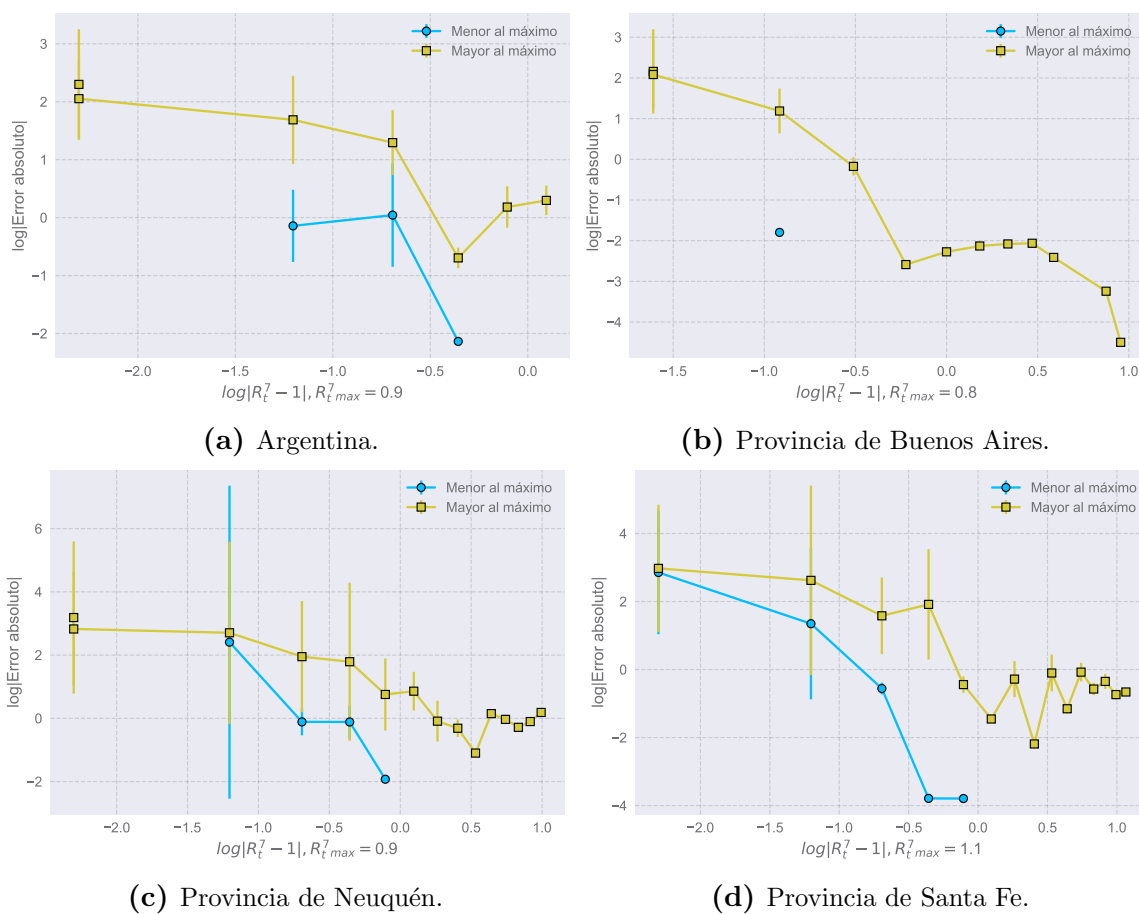


Figura 3.5: Error absoluto promedio en función del valor R_t^7 central de cada bin de la Figura 3.4, en escala log-log para Argentina y las provincias de Buenos Aires, Neuquén y de Santa Fe.

el logaritmo del error absoluto promedio y $\log(|R_t^7 - 1|)$ no se corresponde con una dependencia lineal.

3.2.3. Diseño y estudio de nuevos histogramas de error predictivo.

En el caso de los histogramas anteriores, se organizan las observaciones según su valor de R_t^7 en bins. Así, resulta que hay bins con muchas observaciones, como sucede en torno a la unidad, que tienen del orden de las decenas o centenas. En cambio, para los valores R_t^7 cercanos a los valores máximos o mínimos alcanzados en cada localidad, puede no haber ni siquiera media decena de observaciones. Esto puede provocar que la estadística en estas regiones no sea buena. Para solucionar este problema, se buscó crear histogramas que tuviesen un número fijo de observaciones en cada bin, en contraposición con el caso de los histogramas convencionales.

Para ello, se ordenaron los datos de cada localidad de menor a mayor teniendo en cuenta su R_t^7 y se los dividió en grupos de 20 registros, habiendo asignado previamente la predicción y el error correspondientes. Así, se calculó para cada grupo de 20 datos el debido error absoluto promedio y la desviación estándar del promedio. De esta manera, se obtienen bins con igual cantidad de datos y con valores R_t^7 similares entre sí. Los resultados de este procedimiento para Argentina y las provincias de Buenos Aires, Neuquén y de Santa Fe se presentan en la Figura 3.6.

De estos histogramas modificados, se observa claramente que los valores cercanos a la unidad son más frecuentes que los otros, simplemente notando que los bins en esta región de los gráficos son mucho más angostos. En cambio, hacia los extremos del mismo, se tienen que tener en cuenta rangos más grandes de R_t^7 para contar 20 datos y, por lo tanto, los bins se ven más anchos. Por otra parte, estos histogramas no presentan la misma estructura que los de la Figura 3.4 un máximo absoluto y a lo sumo algún máximo local que toma un valor mucho menor. En varios de los casos, se observan dos máximos de valores muy similares. Finalmente, la desviación del promedio resulta ser mucho mayor en general.

Análogamente al caso presentado en la sección anterior, se busca algún tipo de comportamiento asociado a los fenómenos críticos en torno a $R_t^7 = 1$. Para ello, se repite el procedimiento realizado para obtener de los histogramas convencionales los gráficos de la Figura 3.5, pero a partir de los nuevos histogramas modificados. Los resultados se presentan en la Figura 3.7. Para cada una de las poblaciones en estudio, se indica el valor de $R_{t_{max}}^7$ y se distinguen los datos con valor de R_t^7 superior e inferior. Al igual que en el caso anterior, se tiene que los valores de $R_{t_{max}}^7$ se encuentran cerca de la unidad. Finalmente, no se observa una relación entre el error absoluto promedio y $|R_t^7 - 1|$ que sea como la de la ecuación 3.3.

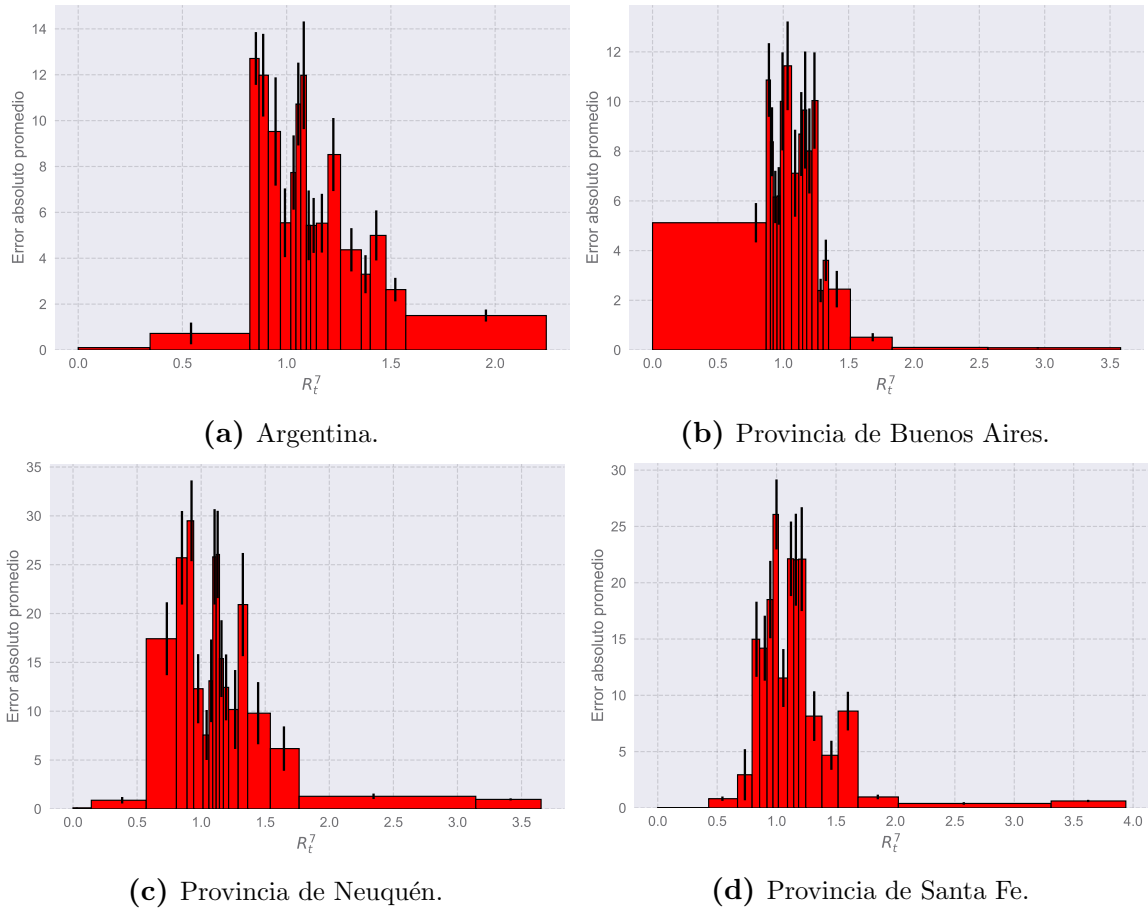


Figura 3.6: Histogramas modificados de error absoluto promedio en función de R_t^7 para Argentina y las provincias de Buenos Aires, Neuquén y de Santa Fe. Cada bin corresponde a un número constante de 20 observaciones.

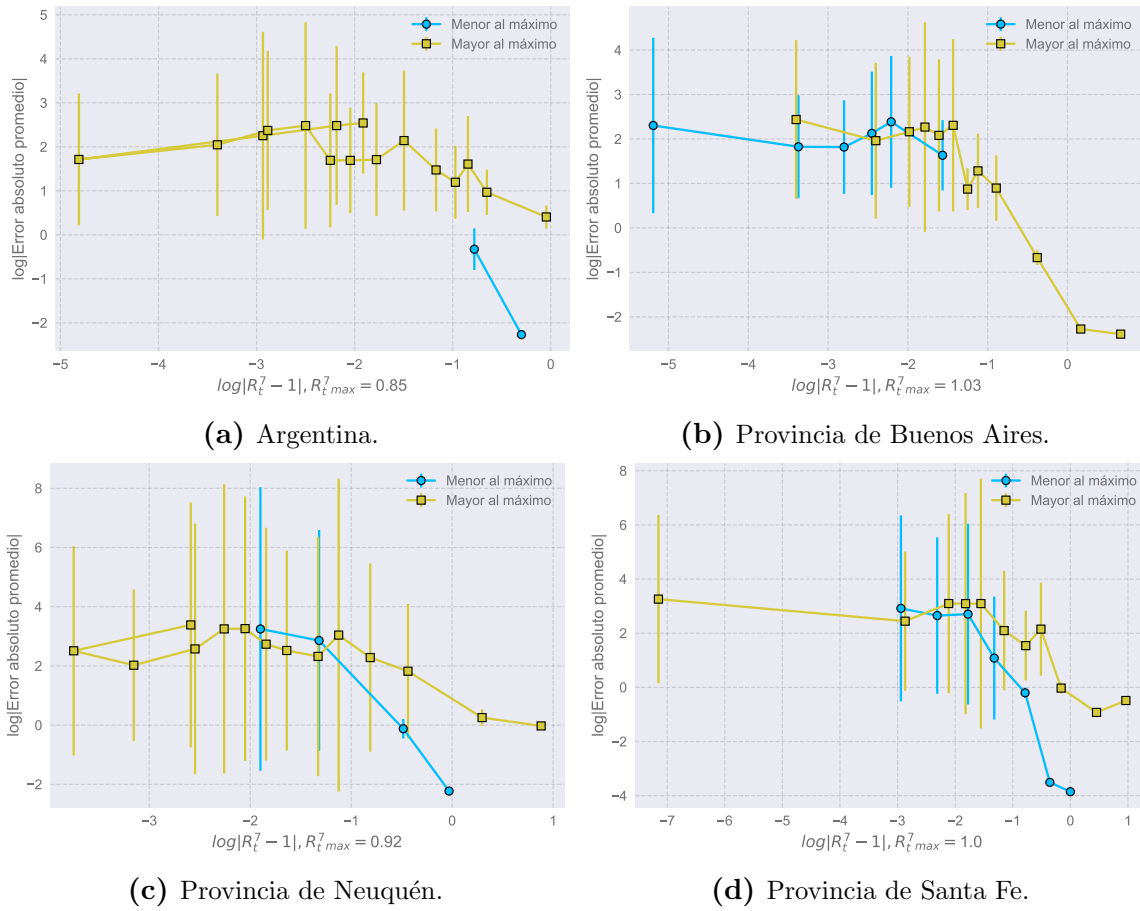


Figura 3.7: Error absoluto promedio en función del valor R_t^7 central de cada bin de la Figura 3.6, en escala log-log para Argentina y las provincias de Buenos Aires, Neuquén y de Santa Fe. Cada punto de los gráficos corresponde a 20 observaciones.

3.3. Análisis del delay temporal óptimo.

Si se recuerdan los conceptos tratados previamente, especialmente en el capítulo 2, se tiene que para predecir los casos a un día t se utilizan los datos a un día $t - r$, con r un retraso de 7 días. De esta manera, se pueden ajustar los casos y predecirlos a 7 días en el futuro razonablemente. Sin embargo, no necesariamente $r = 7$ días es un retraso óptimo con estas herramientas.

Para dedicarse a hallar un retraso r óptimo, se seleccionaron localidades y provincias de la Argentina, que se detallan en el Apéndice 1 junto con sus abreviaturas correspondientes. Entonces, para cada una de estas poblaciones, se calculó el error absoluto promedio de la predicción para diferentes días r en un rango 5 – 8 días. Estos resultados se muestran en la Figura 3.8 con su error correspondiente. Se observa que existen diferencias al utilizar diferentes valores de r , y que las tendencias son similares para la mayoría de las localidades.

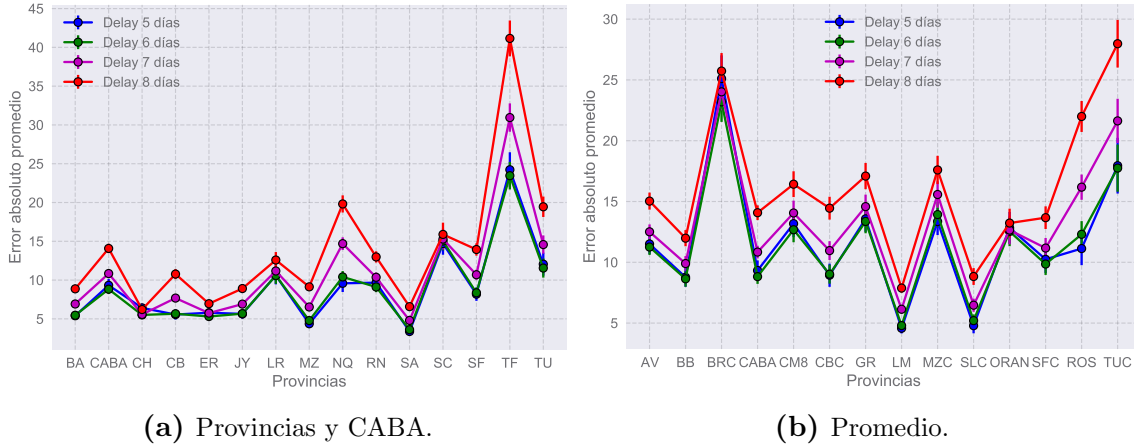


Figura 3.8: Estudio del error absoluto en función de r para diferentes localidades y provincias.

Finalmente, resulta útil promediar el error absoluto promedio para todas las provincias y localidades, lo cual se hace en la Figura ?? . Allí, observamos que para el total de las localidades y de las provincias los resultados son similares. En primer lugar, se tiene que no hay una relación monótona entre r y el error promedio. Por otra parte, se observa que $r = 7$ no el valor óptimo de r . Finalmente, se tiene que el error absoluto promedio es mínimo para $r = 6$, y que para $r = 5$ toma un valor similar al mínimo.

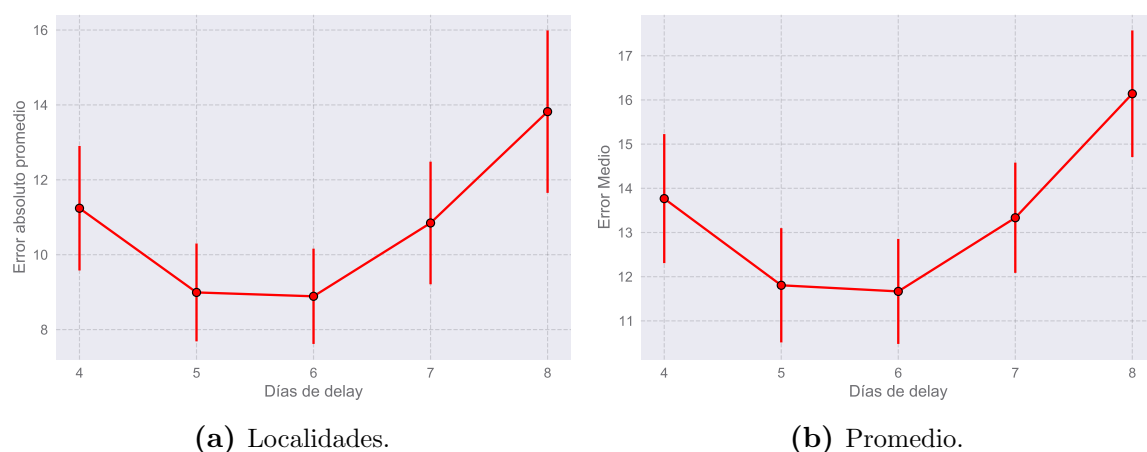


Figura 3.9: Estudio del error absoluto en función de r , teniendo en cuenta todas las provincias y localidades a la vez.

3.4. Conclusiones.

En este capítulo nos hemos dedicado a observar y estudiar algunas propiedades de los diagramas de riesgo y del número reproductivo R_t^7 . En particular, se destaca la tendencia de que $R_t^7 \rightarrow 1$ al largo plazo en varios de las poblaciones en estudio. Luego, se introdujeron conceptos básicos de la teoría de fenómenos críticos, y estudiamos posteriormente la existencia de un fenómeno de tal naturaleza en la relación entre el error absoluto promedio y $R_t^7 - 1$. Sin embargo, no se halló evidencia de que suceda tal cosa. Finalmente, se buscó un número de días de retraso r óptimo para realizar predicciones a partir de las herramientas desarrollados en el capítulo 2.

Capítulo 4

Optimización de las predicciones

4.1. Número reproductivo empírico.

4.1.1. Optimización con 3 pesos.

4.1.2. Optimización con 5 pesos.

4.2. Optimización utilizando la librería SciPy.

4.2.1. La función `curve_fit` de la librería SciPy.

4.3. Optimización con redes neuronales utilizando la librería Keras.

4.3.1. Breve introducción sobre redes neuronales artificiales.

4.3.2. Estructura de las redes neuronales utilizadas en la optimización.

4.3.3. Resultados.

4.4. Conclusiones.

Capítulo 5

Modelo SIR y SIR modificado

El objetivo de este capítulo es introducir y estudiar modelos matemáticos para el modelado de enfermedades infecciosas. En la sección 5.1 se presenta el conocido modelo SIR, uno de los modelos más simples para modelar la propagación de enfermedades infecciosas. En la sección 5.2 se presenta una modificación del mismo. En el caso de ambos modelos, se utilizan las herramientas del capítulo 2 para analizar sus resultados y se los contrasta con propiedades de poblaciones reales.

5.1. Modelo SIR sin demografía clásico

Los modelos más simples para modelar enfermedades agudas (es decir, que causan infección por un periodo de tiempo corto sucedido por inmunidad) son los modelos del tipo SIR. Este formalismo consiste en dividir a los individuos de una población en estudio en tres grupos: susceptibles (S), infectados (I) y recuperados (R). En el caso más simple, si se ignoran todos los aspectos demográficos y se considera inmunidad de por vida, se pueden dar dos transiciones definidas entre grupos: $S \rightarrow I$ y $I \rightarrow R$. La primera corresponde a un individuo susceptible que se transforma en infectado por algún mecanismo de transmisión de la enfermedad. En cambio, la segunda transición corresponde a una persona infectada que, tras un período infeccioso determinado se recupera y se torna inmune. Este último podría ser variable, pero usualmente se realiza la simplificación de considerarlo constante.

En la Figura 5.1 se presenta un diagrama de flujo que representa conceptualmente un modelo de este tipo. Allí, las flechas sólidas negras representan los movimientos posibles entre clases, mientras que la flecha gris rayada representa que la tasa a la que se infectan individuos susceptibles puede depender de la enfermedad infecciosa.

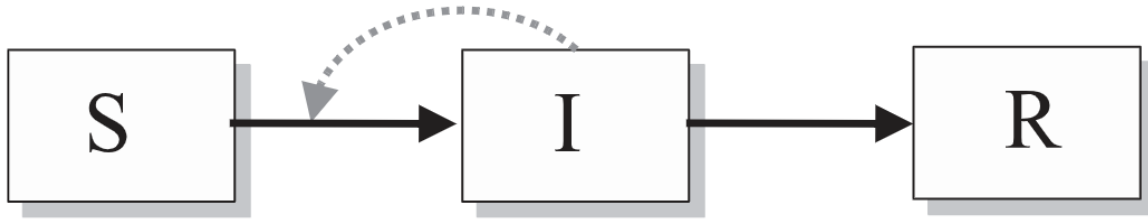


Figura 5.1: Diagrama de flujo representativo de un modelo SIR clásico.

Considérese entonces una población de N individuos, de los cuales S son susceptibles, I están infectados y R se han recuperado, tal que $S + I + R = N$. Alternativamente, se pueden definir las variables: $s = S/N$, $i = I/N$ y $r = R/N$, de manera que $s + i + r = 1$. Utilizando estas cantidades, el modelo previamente introducido se describe matemáticamente con el sistema dinámico:

$$\frac{ds}{dt} = -\beta s \quad (5.1a)$$

$$\frac{di}{dt} = \beta si - \gamma i \quad (5.1b)$$

$$\frac{dr}{dt} = \gamma i, \quad (5.1c)$$

donde γ es la inversa del período infeccioso, es decir, es la tasa de recuperación. Por otra parte, β es la transmisión o tasa de contagio que multiplica a las proporciones de susceptibles y de infectados, ya que las infecciones se provocan por contacto entre estas dos poblaciones. Estas ecuaciones tienen condiciones iniciales $s(0) \geq 0$, $i(0) \geq 0$ y $r = 1 - s(0) - i(0)$. A pesar de ser uno de los modelos más simples para tratar enfermedades infecciosas, no es resoluble analíticamente.

Este modelo ha servido para ilustrar varios principios epidemiológicos cualitativamente. Por ejemplo, el fenómeno de umbral o de *threshold*, que consiste en que es necesaria una mínima fracción de susceptibles inicial $s(0)$ para que una enfermedad infecciosa se propague. Esto se puede ver fácilmente reescribiendo la ecuación 5.1b como:

$$\frac{di}{dt} = i(\beta s - \gamma), \quad (5.2)$$

de donde se observa que para que $\frac{di}{dt}(0) > 0$ debe suceder que $s(0) > \frac{\gamma}{\beta}$. De lo contrario, $\frac{di}{dt}(0) < 0$ y la epidemia finaliza. La cantidad $\frac{\gamma}{\beta}$ suele llamarse la tasa relativa de disipación y el fenómeno de umbral puede interpretarse como la necesidad de que $\frac{\gamma}{\beta}$ sea lo suficientemente pequeña para permitir la propagación de la enfermedad. A partir esta tasa, se define el número reproductivo básico R_0 como su inversa. Este número fue inicialmente introducido y definido en el capítulo 2. El fenómeno del umbral se

puede interpretar en base a esta cantidad como: si inicialmente tenemos una población totalmente susceptible, la enfermedad se propagará si y solo si $R_0 > 1$.

5.2. Modelo SIR sin demografía modificado

Bibliografía

- [1] Harrington, L. C., Scott, T. W., Lerdthusnee, K., Coleman, R. C., Costero, A., Clark, G. G., *et al.* Dispersal of the dengue vector *Aedes aegypti* within and between rural communities. *The American Journal of Tropical Medicine and Hygiene*, **72** (2), 3, 2005. URL <http://www.ajtmh.org/content/journals/10.4269/ajtmh.2005.72.209>.

Agradecimientos

Inserte agradecimientos aquí

