Universidad de
SanAndrés

Universidad de San Andrés

Departamento de Economía

Licenciatura en Economía

Un modelo mínimo para pronosticar el impacto de la pandemia del COVID-19

Autor: Daiana Lucía Garcia Alvarez

Legajo: 27085

Mentor: Hugo Fort Quijano

Victoria, Provincia de Buenos Aires, Agosto de 2020

Índice general

1. Introducción **5**

2. Modelo cinemático: el modelo Gaussiano como la aproximación lineal **8**

2.1. Curva total de fallecidos 12

2.2. Robustez del modelo cinemático 13

3. Aplicación del modelo Gaussiano a algunos países **14**

3.1. Italia 14

3.2. España 20

3.3. Alemania 21

3.4. Reino Unido 22

3.5. Bélgica 23

3.6. Francia 24

3.7. Holanda 25

3.8. Turquía 26

3.9. EEUU 27

3.10. Comentarios finales 29

4. Comparación con los modelos dinámicos **30**

4.1. Modelo SI o modelo logístico 30

4.2. Modelo SIR 31

4.3. Más allá del modelo SIR 32

5. Aproximación Gaussiana del modelo SIR **34**

5.1. Solución analítica completa del modelo SIR en la aproximación Gaussiana 35

5.2. Comparación con la aproximación de Kermack & McKendrick (1927) 37

6. La tasa de reproducción básica $\rho_0 = \frac{\beta S(0)}{\gamma}$ **40**

Universidad de SanAndrés

7. Solución analítica exacta de la trayectoria $S(R)$ en el modelo SIR **42**

7.1. Cálculo exacto del pico de infectados I_0 43

7.2. Estimación de ρ_0 para un grupo de países 45

7.3. La ecuación trascendente para la población asintótica de susceptibles $S(\infty)$ 47

7.4. Cálculo exacto de la fracción de fallecidos finales $f(\infty)$ 49

8. Resúmen y Conclusiones **52**

9. Apéndice A: Aproximación de segundo orden (el modelo sech^2 de KM) **54**

10. Apéndice B: Resolución de la ecuación logística generalizada **56**

11. Apéndice C: Estimación de los parámetros gaussianos ($I(0)$, t_0 y k) para cada uno de los países **58**

Bibliografía **69**

QUAERERE VERUM
Universidad de
SanAndrés

Resumen

En el presente trabajo desarrollamos, en principio, un modelo cinemático utilizando los datos de la curva de infectados y fallecidos en el marco de la pandemia del COVID-19, basado en dos hipótesis de carácter general que determinan que la curva de infectados activos de la misma se puede aproximar a un pico gaussiano. La curva de fallecidos acumulados $F(\infty)$ es proporcional al área bajo la curva de la gaussiana. Aplicamos el presente modelo mínimo para predecir el principal impacto de la pandemia, las cantidad de fallecidos finales, para Estados Unidos y un grupo de países europeos. Observamos que las curvas ajustan perfectamente a los datos empíricos hasta los días previos al pico de la pandemia. Luego, se observa como las medidas de restricción amortiguan la curva de infectados y fallecidos. Consecuentemente, el resultado asintótico $F(\infty)$ de nuestro modelo resulta ser una cota superior a los datos observados, funcionando el modelo gaussiano como una herramienta de pronóstico.

Para determinar las limitaciones y alcance de nuestro modelo, establecemos la conexión entre el modelo gaussiano y el modelo dinámico compartimental SIR (Kermack y McKendrick, 1927), que plantea una sistema de ecuaciones diferenciales de primer orden para determinar la dinámica de las poblaciones. De ese modo, la aproximación gaussiana resulta ser una robusta aproximación de la secante hiperbólica cuadrado obtenida por Kermack y McKendrick (1927), que, a su vez, es una una aproximación a segundo orden en la tasa de reproducción básica ρ_0 de las ecuaciones exactas del modelo SIR. Dicha conexión también permite establecer un contacto entre los parámetros cinemáticos de nuestro modelo gaussiano y los parámetros dinámicos del modelo SIR. La solución analítica exacta de las trayectoria de las poblaciones permite hallar una ecuación trascendente que determina los valores asintóticos de las mismas. Esto último nos permite acotar el rango de validez de nuestro modelo y la aproximación de segundo orden Kermarck y McKendrick en su predicción del valor asintótico de fallecidos acumulados $F(\infty)$.

Universidad de
SanAndrés

Capítulo 1

Introducción

La presente pandemia provocó que los estados se hayan encontrado ante el dilema de minimizar los riesgos de salud para su población y a su vez, intentar minimizar las consecuencias sobre la economía. Como las medidas que se han tomado para salvaguardar la salud inevitablemente están causando una depresión económica a nivel mundial, un análisis racional del problema nos obliga a hacer un balance de los costos y beneficios de las mismas para intentar minimizar las potenciales pérdidas.

El objetivo principal del presente trabajo es pronosticar el impacto del COVID-19: el número final de fallecidos. Buscaremos predecir el número total de fallecidos que puede esperarse en un país en el marco de una epidemia con ciertos grados de restricción. Nuestro enfoque, en principio, no buscará ahondar en las causas de la dinámica epidemiológica, sino lograr plantear un modelo simple y con la menor cantidad de parámetros posibles. A medida que complejizamos el modelo aumentando el número de parámetros, nuestra capacidad predictiva disminuye, o dicho de otra manera, con un modelo lo suficientemente complejo podemos describir prácticamente cualquier fenómeno. Pero el objetivo de la ciencia no solo trata de describir la realidad, sino de explicarla en términos del menor número de hipótesis posibles.

Tampoco se trata de predecir exactamente la cinemática final de los acontecimientos. Como hemos visto hasta ahora, la mayoría de los modelos pueden ser utilizados prácticamente en el momento en el cual la epidemia finalizó o está en su etapa final. Desde el punto de vista predictivo, estos no aportan dado que la mayoría de las medidas económicas y sociales para ese momento ya fueron adoptadas a ciegas o peor aún, basadas en estimaciones erróneas. Por lo tanto no se trata de tener una estimación exacta del número final de fallecidos en un modelo altamente complejo, sino realizar tempranamente una estimación que, aunque sea aproximada, sea robusta en el orden de magnitud.

Para lograr el objetivo planteado, en la Sección 2 desarrollamos un modelo cinemático que se focaliza en la curva de infectados y fallecidos de una epidemia. El modelo propuesto se basa en dos hipótesis de carácter general que determinan que la curva de infectados activos en una epidemia se puede aproximar a un pico gaussiano $I(t)$. Nuestro modelo tiene, en principio, trés parámetros: la altura del pico de infectados I_0, el tiempo en que ocurre dicho pico t_0 y un parámetro k que mide el ancho de la gaussiana. La curva de fallecidos acumulados $F(t)$ es proporcional al

área bajo la curva de la gaussiana o función error. La constante de proporcionalidad es la tasa de mortalidad μ.

En la Sección 3 utilizamos el modelo planteado para predecir la cantidad de fallecidos finales para un grupo de países europeos y EEUU. El caso de Italia se explica en detalle como ajustamos los parámetros μ, I_0, t_0 y k de la curva de fallecidos $F(t)$. La curva de $F(t)$ es una sigmoidea caracterizada por la función error. En todo el grupo de países bajo análisis, observamos que desde el principio hasta el pico de la pandemia los datos empíricos siguen al modelo. Sin embargo, a partir de ese momento, se observa como las medidas de restricción adoptadas amortiguan la curva de infectados y fallecidos llevando a que siempre el resultado final sea menor que la estimación del modelo del número final de fallecidos $F(\infty)$. De esta forma en los distintos países, dichas estimaciones resultan ser un cota superior de nuestro pronóstico basados en el ajuste de parámetros en el tramo inicial de las curvas. En base a esto, podemos ver que: a) los datos empíricos hasta el pico validan el modelo y b) el modelo gaussiano puede ser utilizado como una robusta herramienta de pronóstico.

En la Sección 4 realizamos un breve repaso de los distintos modelos que describen la dinámicas de poblaciones de una epidemia. Dichos modelos plantean un sistema de ecuaciones diferenciales de primer orden que determinan la dinámica de las poblaciones. Todos ellos tienen origen en el trabajo original de Kermack y McKendrick (1927) (KM) conocido como el modelo SIR (*Susceptible*, *Infected* and *Recovered*) cuyas siglas refiere a las poblaciones $S(t)$, $I(t)$ y $R(t)$ en las que se compartimenta la población total N bajo estudio.

En la Sección 5, establecemos la conexión entre el modelo gaussiano y el pico de la secante hiperbólica cuadrado derivada originalmente por KM, que discutimos en detalle en los Apéndices A y B. En el Apéndice A mostramos la aproximación de segundo orden (en la tasa de reproducción básica inicial $\rho_0{=}\frac{\beta S(0)}{\gamma}$), que conduce a la ecuación diferencial de KM. La misma resulta una ecuación logística generalizada cuya solución obtenemos en el Apéndice B. La solución de KM es una ecuación dinámica en donde todos los parámetros del modelo estan completamente determinados por las condiciones iniciales $S(0) \simeq N$, $I(0) \simeq 0$ y $R(0) = 0$ de las poblaciones y los parámetros β (tasa de transmisión) y γ (tasa de recuperación) del modelo SIR. La conexión con el trabajo de KM muestra que si ajustamos la función error con la sigmoidea de KM de modo tal que la cantidad de fallecidos finales sea la misma para ambas curva, entonces dichos modelos también predicen los mismos resultados en todo rango de tiempo en torno al pico de infectados, donde vimos que es válida la aproximación gaussiana. De esta forma, logramos establecer un vínculo entre los parámetros de nuestro modelo cinemático y los parámetros del modelo SIR.

En la Sección 6 estimamos un importante concepto de toda epidemia, la tasa de reproducción básica instantanea $\rho(t){=}\frac{\beta S(t)}{\gamma}$. Dicho concepto surge de reparametrizar las ecuaciones de movimiento del modelo SIR en función de un tiempo natural γt. Dicha tasa es determinante para lograr un buen monitoreo de la evolución de cualquier epidemia ya que cuando $\rho(t) > 1$, la epidemia se desarrolla, cuando $\rho(t) = 1$ se llega al pico y cuando $\rho(t) < 1$, la misma se "apaga".

Por último, en la Sección 7 discutimos la solución exacta de la trayectoria de las poblaciones en el plano S - R (Murray, 2002) y obtuvimos ρ_0 para el grupo de países bajo análisis. La ecuación de la trayectoria nos permite obtener los valores exactos de las poblaciones en el pico de la pandemia ($t{=}t_0$) a todo orden en el parámetro γ, las cuales a su vez, nos permiten comparar dichos resultados con los que se obtienen de la aproximación de KM. Del

mismo modo, se pueden obtener así los valores asintóticos exactos que nos permiten evaluar el rango de validez de las aproximaciones sigmoideas (a segundo orden en ρ_0) de los fallecidos.

QUAERERE VERUM
Universidad de
SanAndrés

Capítulo 2

Modelo cinemático: el modelo Gaussiano como la aproximación lineal

Nuestro modelo sera un modelo descriptivo que en principio se focalizara en la cinemática de la curva de infectados. El modelo no intenta capturar todas las variables que explican las particularidades de la dinámica de la presente pandemia. Al igual que sucede en física en principio podemos apelar a dos niveles descriptivos de la realidad. El más simple es un nivel cinemático principalmente desarrollado por Galileo en su *Dialogo entre dos mundos* (Galileo, 1632) en el que no se indaga sobre las causas del movimiento sino que tan solo se lo describe mediante ecuaciones horarias parametrizadas en el tiempo. Posteriormente, Newton en sus *Principia* (Newton, 1687) llevo esa descripción a un nivel más profundo, el nivel dinámico en el que se intenta describir las causas de las ecuaciones de movimiento que siguen las trayectorias de las partículas. Aunque como el mismo confiesa en famosa frase *hypotheses non fingo*, solo le fue posible caracterizar las fuerzas que provocan la aceleración de la gravedad descripta por Galileo pero sin llegar a explicar su origen. Como veremos lo mismo ocurre en epidemiologia en donde ciertos modelos de dinámica de poblaciones juegan el rol de la segunda ley de Newton.Esto quedera reservado a los modelos dinámicos mas complejos como el Multi-Risk SIR (Murray, 2002, p. 320), entre otros. En nuestro caso, plantearemos ciertas hipótesis cinemáticas de caracter general que a posterior se veran sustentadas en la observación de las curvas empíricas de infectados y fallecidos de los distintos países bajo análisis. Una explicación causal o de por qué las curvas tienen determinado comportamiento, se observara a partir de modelos dinámicos como el Susceptible-Infected-Recovered (SIR) originalmente desarrollado por Kermack y McKendrick en 1927 (Murray, 2002); (Brauer, 2008) que junto a sus extensiones representan el "caballo de batalla"de los modelos que se han utilizado para las dinamicas epidemiológicas. En principio, comenzaremos haciendo ciertas hipótesis que pueden resultar mas o menos intuitivas y que matematicamente determinan la forma final de las curvas. Estas son hipótesis simples y generales, ya que estan basadas en argumentos comunes que se presentan en todas las extensiones del modelo SIR. La primera de ellas es: **Hipótesis 1: En toda epidemia la velocidad de crecimiento del número de infectados es proporcional al número de infectados que la población tiene en ese instante de tiempo.**

Universidad de SanAndres

$$dI = K(t)I(t)dt. \tag{2.1}$$

En efecto, cuanto más tiempo transcurra y más infectados haya en la población en ese instante, mayor es la probabilidad de infectar al resto de la población sana (no inmune). Es decir, que la probabilidad que tiene un individio sano de infectarse a un dado tiempo es proporcional a la cantidad de agentes trasmisores que interactuan con él en ese lapso de tiempo.

La ecuación diferencial que gobierna el fenómeno es una ecuación diferencial de primer orden del crecimiento porcentual siendo $K(t)$ la tasa con la cual se infecta la población en ese instante

$$\boxed{\frac{1}{I}\frac{dI}{dt} = K(t).} \tag{2.2}$$

En este punto no vamos a hacer ninguna hipótesis adicional sobre la estructura interna de $K(t)$, ni explicar el detalle de su composición. Luego veremos que, en el marco de distintos modelos dinámicos, $K(t)$ adopta distintas formas más o menos complejas de acuerdo a las hipótesis específicas de cada modelo. La hipótesis más simple que uno puede hacer sobre la función $K(t)$ es que la misma es una constante, esto significa que el crecimiento de la epidemia sería una exponencial. Pero esta hipótesis resulta ingenua porque como se observa experimentalmente, eso solo ocurre al principio de toda epidemia. Este resultado es esperable ya que la biología no permite que ello se sostenga más allá de un rango de tiempo. Tampoco en economía los crecimientos exponenciales duran para siempre, todos ellos se ven limitados por las condiciones de contorno del problema ya que los recursos que alimentan todo crecimiento exponencial son limitados.

En el caso de las epidemia hay razones de carácter general que se pueden utilizar para explicar dicho comportamiento, la más simple es la siguiente: a medida que crece el número de recuperados, dicha población actua de contrapeso en la velocidad de crecimiento de la población de infectados. Pero no solo por una razon contable (parte de la población infectada se recupera) sino porque además las chances de generar nuevos infectados va disminuyendo con el transcurso del tiempo. No olvidemos que, como explicamos más arriba, un nuevo infectado es producto de la interación entre un individuo sano pasible de recibir la infección y otro enfermo. Los nuevos contagios $(dI)_c$ en un lapso dt son proporcionales al número de encuentros (fdt) entre individuos infectados $I(t)$ y sanos $S(t)$, y al número de posibles encuentros $I(t) \times S(t)$, y a la probabilidad intrinseca de transmisión i_c en un encuentro individual.

$$(dI)_c = i_c(fdt)I(t)S(t), \tag{2.3}$$

donde f es la frecuencia de dichos encuentros. Lo que podemos reescribir como,

$$\left(\frac{dI}{dt}\right)_c = i_c f I(t)S(t). \tag{2.4}$$

A medida que la epidemia se desarrolla la cantidad de individuos pasibles de ser infectados $S(t)$, es decir susceptibles, se va agotando y por lo tanto, las probabilidades de generar nuevos infectados disminuyen. Si bien para cada uno de los portadores las chances de contagiar p_c son las mismas, la cantidad de potenciales receptores de

la enfermedad disminuye con el tiempo. De esta forma, la tasa de crecimiento de la pandemia de a poco comienza a decrecer como se observa experimentalmente.

La descripción anterior seria la dinámica simple en el marco de una epidemia de conejos que son individuos que no tienen información sobre lo que esta ocurriendo. Pero los humanos como siempre complicamos las cosas agregando un factor distintivo. Nosotros tomamos decisiones en base a la información disponible y esto genera que nuestro comportamiento social sea distintinto a medida que todos recibimos un flujo de información, que puede ser incompleta, sobre el desarrollo de la epidemia. Consecuentemente, se puede esperar que la probabilidad de contagio sea mucho mayor al principio de la misma, que luego de esa fase inicial en la que todos ya estamos advertidos sobre los potenciales peligros. Esta información por momentos sesgada e imperfecta, genera que el individuo promedio modifique sus pautas de comportamiento. La cantidad de interacciones entre humanos por unidad de tiempo f disminuye, las interacciones se vuelven mas distantes, algunos toman más medidas higienicas y de protección; y el resultado efectivo es que la probabilidad intrinseca i_c de que un individuo enfermo contagie a uno sano disminuye estabilizandose el producto $i_c f$ en un nuevo valor. Podriamos sugerir que esto sucede en una forma mas acentuada en nuestra "era de la información", independientemente de que lo que los estados recomienden o regulen para que esto suceda. En definitiva, no reaccionamos de la misma manera una vez informados y esta es la razón por la cual la tasa de crecimiento de toda epidemia es mucho mayor al principio. Una vez que esa pauta en "grado de alerta"se ha estabilizado, la dinámica sigue nuevamente la lógica de la población de conejos con una dinámica de parámetros estables.

Esta dinámica social que acabamos de describir no esta contemplada en ninguno de los modelos matemáticos conocidos, ya que la misma implica un cambio en los parámetros del modelo dinámico (en el caso del modelo SIR veremos que corresponde a un cambio en su coeficiente $\beta = i_c f$). Los modelos conocidos describen dinámicas naturales, pero no hay modelos cinéticos para la dinámica de sus parametros. Es decir, nuestros modelos matemáticos standard no describen como i_c y f pueden verse alterados en el tiempo. Esta es una de las razones principales por las cuales resulta tan dificil predecir en base a los mismos, más aún si nos atamos a la hipótesis de que sus parámetros no cambian con el tiempo. Por eso nuestra estrategia se focalizara en reducir al máximo la cantidad de parámetos para de ese modo poderlos monitorear mejor.

Otro de las limitaciones de los modelos, es la ausencia de la dimensión espacial. En la mayoría de ellos se asume que la población se concentra en un pequeño territorio aislado como puede ser una ciudad o un crucero. Dentro de ese territorio la densidad es más o menos homogénea y por lo tanto la frecuencia f de encuentros entre humanos es aproximadamente la misma. Si tenemos un pais extenso como EEUU, uno podría aplicar esa misma lógica para cada una de sus ciudades y luego sumar para obtener un modelo global. Si cada una de ellas se mantuviese aislada en ese caso lo que se obtendría es un modelo de parámetros promedio pero estables en el tiempo. Pero, que ocurre si se abren las compuertas y permitimos migraciones entre las distintas ciudades? Es claro que nuestro modelo agregado deja de funcionar ya que al migrar infectados de zonas S de determinada densidad a zonas S' de de otra densidad, ese bono demográfico δI de infectados encuentra territorio "virgen"de población sana en donde volver a hacer efectiva la transmisión del virus en S'. Esto se observa claaramente es los rebrotes que se dieron en las últimas

semanas en EEUU.

$$\left(\frac{dI'}{dt}\right)_c = i_c f'(I' + \delta I)S'(t) \tag{2.5}$$

En suma, lo único que sabemos con certeza es que si los parámetros de comportamiento social se mantienen estables y no hay migraciones, la tasa $K(t)$ tiende a decrecer en el tiempo. Esto nos lleva a formular la segunda hipótesis fundamental de nuestro modelo descriptivo:

Hipotesis 2: En toda epidemia, en donde los parámetros de comportamiento social se mantienen estables y el flujo migratorio controlado, la tasa de crecimiento porcentual decrece con el tiempo y a primer orden puede aproximarse por una recta.

En base a dicha hipótesis efectuamos un desarrollo en serie de la función $K(t)$ a primer orden,

$$K(t) = K(0) + \frac{dK}{dt}(0)t + ...$$

renombramos las constantes como

$$\begin{aligned} K(0) &= kt_0 \\ \frac{dK}{dt}(0) &= -k \end{aligned}$$

siendo k por la **Hipótesis 2** una constante positiva, podemos reescribir la aproximación de la función como

$$\boxed{K(t) = -k(t - t_0).} \tag{2.6}$$

La **Hipótesis 1 y 2** combinadas permiten reescribir como (2.2)

$$\boxed{\tfrac{1}{I}\tfrac{dI}{dt} = -k(t - t_0)} \tag{2.7}$$

Otra forma de visualizar el problema es reescribir el primer miembro como una derivada logarítmica

$$\frac{d\ln I}{dt} = -k(t - t_0) \tag{2.8}$$

derivando nuevamente respecto del tiempo obtenemos la aceleración de $\ln I$

$$\frac{d^2 \ln I}{dt^2} = -k. \tag{2.9}$$

Lo que estamos diciendo es que la aceleración es una constante. Siguiendo el método cinemático calcularemos las trayectorias de los infectados. Lo denominamos como cinemático, porque no estamos intentamos desentrañar la náturaleza dinámica del fenómeno, simplemente observamos como se comporta en el tiempo. Para nosotros, por el momento, k sera simplemente una constante del modelo que iremos monitorendo durante el transcurso de la epidemia. Integrando (2.8) obtenemos una parábola

$$\ln I = -\frac{k}{2}(t - t_0)^2 + \ln I_0 \tag{2.10}$$

donde $\ln I_0$ es una nueva constante de integración, de lo que resulta que la población de infectados no es más que una Gaussiana

$$\boxed{I(t) = I_0 \exp -\tfrac{k}{2}(t - t_0)^2} \tag{2.11}$$

Universidad de San Andrés

siendo $I_0 = I(t_0)$ la población máxima y t_0 el tiempo en el cual se alcanza dicho máximo.

Nuestro modelo tiene un solo parámetro k y dos constantes de integración I_0 y t_0. La variable k mide la desaceleración de la epidemia en un gráfico semilogarítmico. Este es nuestro único parámetro de control relevante, ya que si el mismo se mantiene estable, sabemos que "los arboles no creceran exponencialmente hasta el cielo"sino que finalmente la población de infectados hallara su máximo en I_0 al tiempo t_0.

Es claro que k representa el parámetro de riesgo principal de toda epidemia ya que cuanto más grande es k, la Gaussiana se vuelva cada vez más aguda. De hecho existe un vínculo entre la población inicial $I(0) = I(t = 0)$ y el pico máximo de la Gaussiana,

$$I(0) = I_0 \exp -\frac{k}{2}{t_0}^2 \tag{2.12}$$

o lo que es lo mismo invirtiendo la ecuación anterior,

$$I_0 = I(0) \exp \frac{k}{2}{t_0}^2 \tag{2.13}$$

lo que significa que a igual población inicial de infectados, el pico máximo de la epidemia crece exponencialmente con $\frac{k}{2}{t_0}^2$.

2.1. Curva total de fallecidos

Nuestro objetivo principal es evaluar los daños de la pandemia, esto se traduce como el número final de fallecidos. Una vez obtenida la curva de infectados $I(t)$ la idea del cálculo es simple y puede resumirse de la siguiente forma:

En principio siempre podemos escribir que el incremento del número de fallecidos a un dado tiempo es igual a la probabilidad $i(\tau)d\tau$ de fallecer en el intervalo $d\tau$ por el número total de infectados,

$$dF(t) = I(\tau + \varphi)p(\tau)dt$$

donde φ contempla un posible desfasaje que se produce en el procesamiento o registro de la curva de infectados, es decir la demora de tiempo en que un nuevo infectado es contabilizado luego de realizarle un test de resultado positivo.

La cantidad de fallecidos a un dado instante sera entonces la integral

$$F(t) = \int_0^t i(\tau)I(\tau + \varphi)d\tau.$$

La aproximación que se hace habitualmente es suponer que la probabilidad $i(\tau)$ por unidad de tiempo de que un individuo de la población de infectados muera, es una constante independiente del tiempo μ, a pesar de que la muestra va cambiando con el tiempo agregando nuevos individuos infectados (μ es lo que se denomina tasa de mortalidad).

En ese caso podemos sacar factor comun μ fuera de la integral y la expresión finalmente resulta

$$\boxed{F(t) = \mu I_0 \int_0^t \exp -\tfrac{k}{2}(\tau - t_0 + \varphi)^2 d\tau.} \tag{2.14}$$

en otras palabras la curva de fallecidos no es otra cosa que la función error (Glaisher, 1871).

2.2. Robustez del modelo cinemático

Como observaremos en la siguiente sección, el modelo cinemático presenta ventajas a nivel predictivo. Por ser un modelo simple con solo trés parámetros, la cantidad de datos necesaria para ajustar el modelo es mucho menor que en el caso de modelos más complejos con máyor número de parámetros. Esta ventaja nos permite hacer estimaciones, aunque aproximadas, mucho más tempranas sobre cual será el desarrollo final de la epidemia. Los modelos complejos no son útiles a nivel predictivo porque se necesita mucho más tiempo para recolectar datos, y su convergencia con los datos experimentales se logra solo al final cuando la pandemia ya practicamente esta finalizada. En el caso de los modelos complejos, las mediciones de los parámetros involucrados no son sencillas o incluso posibles. Por otro lado, al aumentar el número de parámetros a estimar también aumenta la componente del error relacionada a la propagación de los errores. Incluso algunos parámetros pueden depender fuertemente del lugar. Como podemos observar en modelos complejos, los problemas asociados a sus estimaciones también se complejizan.

QUAERERE VERUM
Universidad de
SanAndrés

Capítulo 3

Aplicación del modelo Gaussiano a algunos países

A pesar de que existe un gran debate sobre las estadísticas que realizan los países sobre el Covid-19, elegimos analizar un grupo amplio de países donde la epidemia, es decir el brote principal o inicial ya se ha desarrollado por completo. Este grupo adopto medidas heterogeneas respecto al grado de restricciones que explicaremos en detalle más adelante. Los países a analizar son Italia, España, Reino Unido, Alemania, Bélgica, Francia, Holanda, Turquía y EEUU. En el caso de EEUU, analizaremos el caso global a nivel país, la ciudad de Nueva York, donde ocurrió el brote inicial principal, y los 3 estados con más poblados que tuvieron brotes de Covid-19 (California, Illnois, Texas). Los datos fueron extraidos de Our World in Data [1] y Worldometers [2].

Universidad de SanAndrés

3.1. Italia

La Figura 3.1 nos muestra una imagen aproximada del "diario del lunes"del desarrollo de la epidemia en Italia. La pregunta que nos haremos es si nuestro modelo permite predecir al menos el orden de magnitud del desarrollo final solamente utilizando los datos que poseíamos en el mes de marzo.

La curva azul representa el número de casos confirmados acumulados ($I_m(t) + R_m(t) + F_m(t)$), eso incluye los los infectados activos ($I_m(t)$ curva amarilla, con la apariencia de una Gaussiana) y los recuperados ($R_m(t)$ curva gris) y los que fallecieron ($F_m(t)$ curva naranja).

Según estos datos, el máximo de la epidemia se produjo el 19 de abril, el día 50 contando a partir del primero de marzo. Sin embargo, no sabemos con certeza si eso es lo que realmente ocurrió, solamente que nuestros datos contienen ruido. Se estima que las cantidades son solo una muestras de las cantidades reales. Existe un porcentaje alto de infectados asintomáticos que se recuperó y nunca fueron registrados. Según la WHO[3], se estima que con

[1] https://ourworldindata.org/covid-deaths

[2] https://www.worldometers.info/coronavirus/

[3] https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/q-a-similarities-and-

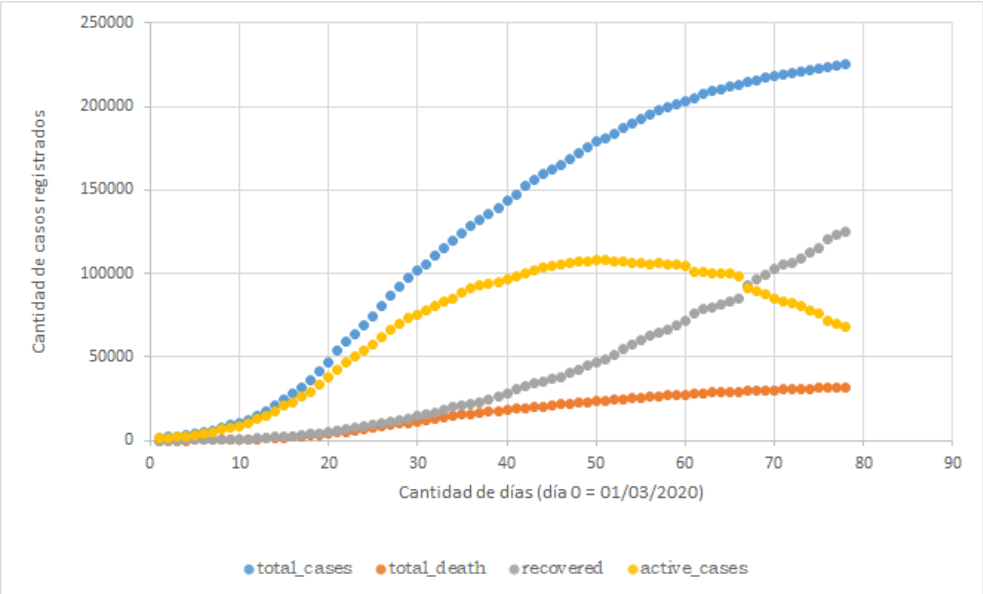


Figura 3.1: Cantidad de casos registrados en Italia entre el 01/03/2020 - 17/05/2020

los datos existentes hasta el momento que el 80 % de los infectados son asintomáticos. Utilizamos el subíndice m para diferenciar las muestras de las cantidades reales. Estas últimas se obtienen multiplicando las muestras por un coeficiente c mayor que la unidad.

$$\begin{aligned} I(t) + R(t) + F(t) &= c\,(I_m(t) + R_m(t) + F_m(t)) \\ R(t) &= c_R R_m(t) \\ F(t) &= c_F F_m(t) \end{aligned}$$

Aquí se nos presenta un primer problema: la muestra de la cantidad de infectados activos $I_m(t)$ se obtiene por diferencia y nada garantiza que los coeficientes sean los mismos. Es probable que c_F sea cercano a la unidad, por lo que representa nuestro único dato más confiable, pero nada asegura que c y c_R sean iguales. Es decir, que no podemos estar seguros de que todas las curvas sean las reales a menos de un mismo factor de escala. Se presume que c_R es mayor que c porque la muestra de recuperados esta sesgada a partir de la muestra de registrados ya que, por falta y finitud de los recursos, solo se registraron en general los pacientes con síntomas y correspondientes a la población de mayor riesgo. Muchos recuperados asintomáticos, fueron omitidos en las muestras registradas. Por ello, la curva real de recuperados probablemente es mayor y la curva real amarilla que se obtiene por diferencia, restando la gris y la naranja a la celeste, haya alcanzado su pico máximo unos días antes.

Teniendo en cuenta todo esto, dado que los únicos datos disponibles son los presentes con los "ruidos"que mencionamos, intentaremos estimar el número final de fallecidos. Es importante remarcar, que se presenta un error sistemático que no podemos corregir y que supera significativamente al error estadístico.

Para obtener los parámetos de la curva Gaussiana que ajusta con los presentes datos, utilizamos un conocido

differences-covid-19-and-influenza

método de análisis de señales (Caruana R. *et al.*, 1986). Dicho método consiste simplemente en efectuar un gráfico semilogarítmico de los datos disponibles para luego ajustarlos con una parábola.

Tomamos los datos del día 13 al 29 de marzo, y vemos que dichos datos ajustan con la parábola

$$\ln I_m(t) = -3.525\,1 \times 10^{-3}x^2 + 0{,}254\,22x + 7.077\,8$$

como muestra la Figura 3.2,

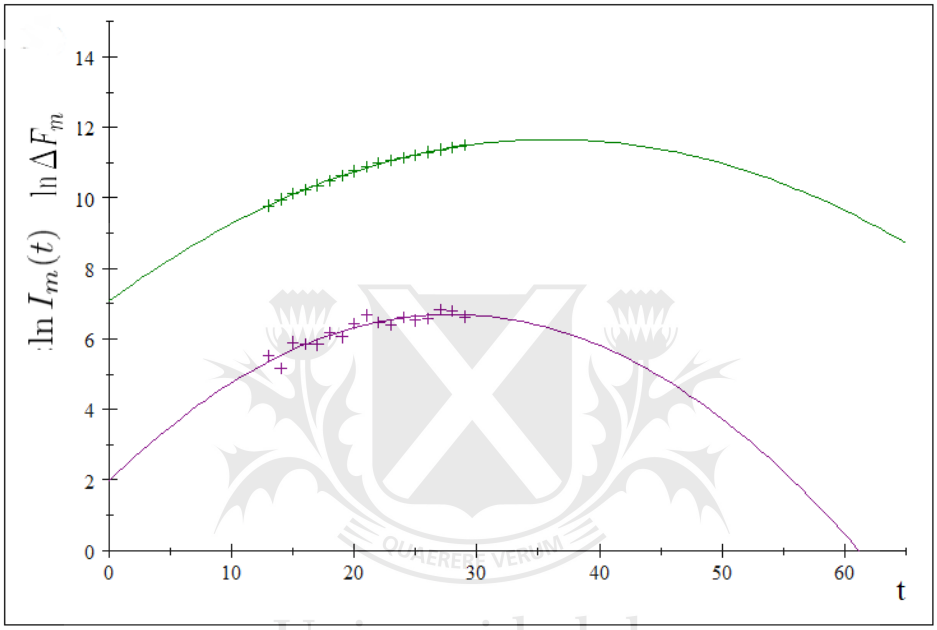


Figura 3.2: Italia:$\ln I_m(t)$ (verde) y $\ln \Delta F_m$ (violeta)

Como podemos observar, los datos en verde ajustan suavemente una parábola. Si hubiésemos utilizado los datos correspondientes al ln del incremento de fallecidos diarios (curva violeta), su curva teórica también debería ajustar a la parábola

$$\ln \frac{dF(t)}{dt} = -\frac{k}{2}(t - t_0 + \varphi)^2 + \ln \mu I_0 \qquad (3.1)$$

pero como vemos, dichos datos tienen mucho más ruido pero probablemente menor error sistemático.

Si lo comparamos con la parábola de la curva teórica,

$$\ln I_m(t) = -\frac{1}{2}kt^2 + kt_0t - \frac{1}{2}kt_0^2 + \ln I_{m0} \qquad (3.2)$$

igualando los coeficientes de ambos polinomios, tenemos

$$\begin{aligned} \frac{1}{2}k &= 3.525\,1 \times 10^{-3} \\ kt_0 &= 0{,}254\,22 \\ \ln I_{m0} &= \frac{1}{2}kt_0^2 + 7.077\,8 \end{aligned}$$

de lo que se sigue que los parámetos de la Gaussiana de la muestra de infectados activos son

$$\begin{aligned} k &= 2 \times 3.525\,1 \times 10^{-3} = 0{,}0070502 \\ t_0 &= \frac{0{,}254\,22}{0{,}0070502} = 36.059 \\ \ln I_{m0} &= \frac{1}{2}\,(0{,}0070502)\,(36.059)^2 + 7.077\,8 = 11.661 \\ I_{m0} &= \exp(11{,}661) = 1.159\,6 \times 10^5 \end{aligned}$$

Por lo que la curva finalmente resulta,

$$I_m(t) = 1.159\,6 \times 10^5 \exp\left[-\frac{0{,}0070}{2}(t-36)^2\right]$$

La curva de infectados reales sería en principio

$$I(t) = c_I \times (1.159\,6 \times 10^5) \exp\left[-\frac{0{,}0070}{2}(t-36)^2\right]$$

donde c_I es un factor desconocido que normaliza la población de la muestra a la real. La curva de infectados activos pasa los 100000 infectados a los 50 días, pero como dijimos más arriba no es una muestra totalmente confiable, por lo tanto tampoco estamos seguros que $t_0 = 36$ ajuste al parámetro real. Solo vamos a asumir que $k = 0{,}0070$ es una estimación cercana al parámetro real. Con ese dato intentaremos hacer un nuevo ajuste de la muestra de la población de fallecidos, que estimamos son los datos más confiables.

Para estimarla vamos a tomar como base la expresión anterior de la Gaussiana,

$$\begin{aligned} F(t) &= \rho \int_0^t (1.159\,6 \times 10^5) \exp\left[-\frac{0{,}0070}{2}(x-36+\varphi)^2\right] dx \\ \rho &= \mu \times c_I \end{aligned}$$

y ajustaremos los coeficientes ρ y φ con los datos de la muestra de fallecidos hasta el día 29. Siguiendo este procedimiento es claro que cualquier error en los parámetros I_0 y t_0 poco importa porque cualquier diferencia en los mismos sera corregida por los nuevos parámetos de ajuste de la curva de fallecidos. Podríamos haber propuesto directamente ajustar tambien k con los mismos datos, pero pasar de dos a tres dimensiones en el espacio de parámetros tiene el costo de poder caer en un mínimo local, que sea un falso mínimo global. Si sabemos que uno de los trés parámetros esta en el orden correcto, es más sencillo moverse en el plano (ρ, φ) para que nuestra curva se acerque lo más posible a los datos.

La Figura 3.3 fue obtenida por iteración gráfica,

Universidad de San Andrés

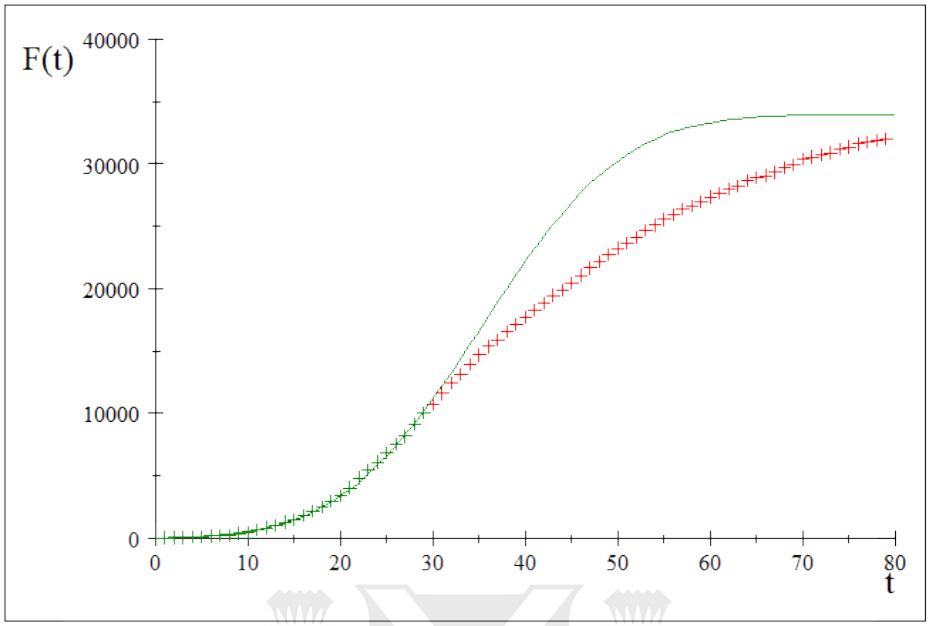


Figura 3.3: Italia - Cantidad de fallecidos desde 01/03/2020

Los puntos en verde fueron los utilizados para obtener la curva pronóstico (verde) y los puntos en rojos son los datos empíricos del desarrollo final de la pandemia. La desviación en la estimación final es del orden del 6 %. Viendo el resultado final (puntos en rojo), una posible hipótesis es que ocurrió un cambio en los parámetros de la curva teórica. Es decir, durante el tiempo en que las condiciones iniciales se mantuvieron estables respetando las normas básicas de cualquier experimento, el modelo siguió a los datos empíricos. Luego, dichas condiciones se fueron alterando.

Italia aplicó una cuarentena desde el 9 de marzo y una cuarentena estricta dejando solamente servicios esenciales desde el 11 de marzo al 13 de abril [4] Como consecuencia de la variación de los parámetros provocada por las distintas medidas de restricción, la curva se aplana desviandose paulatinamente del modelo. Gráficamente, esta variación se observa a partir de $t = 30$, es decir desde el día 30 de marzo. Sin embargo, no notamos un aplanamiento brusco de la curva como ocurre con algunos países que analizaremos a continuación. Esto puede deberse a que las medidas fueran impuestas muy tardiamente o que las medidas no alteraron significativamente los parámetros de la curva; el comportamiento se mantuvo homogéneo.

Si hubiésemos utilizado la curva violeta que corresponde al ln del incremento diario de fallecidos ($\ln \Delta F_m$)

$$\ln \Delta F_m(t) = -5.\,165\,6 \times 10^{-3} x^2 + 0{,}312\,22x + 2.\,024\,6$$

[4] https://www.repubblica.it/politica/2020/03/11/news/coronavirus_conte_italia_governo_misure-250988471

para compararla contra la curva teórica

$$\begin{aligned} \ln \frac{\Delta F_m(t)}{\Delta t} &= -\frac{1}{2}kt^2 + k(t_0-\varphi)t - \frac{1}{2}k(t_0-\varphi)^2 + \ln \mu I_{m0} \\ \Delta t &= 1 \end{aligned}$$

igualando los coeficientes de ambos polinomios, hubiéramos obtenido

$$\begin{aligned} \frac{1}{2}k &= 5.165\,6 \times 10^{-3} \\ k(t_0-\varphi) &= 0{,}312\,22 \end{aligned}$$

de lo que se sigue que los parámetos de la Gaussiana serían

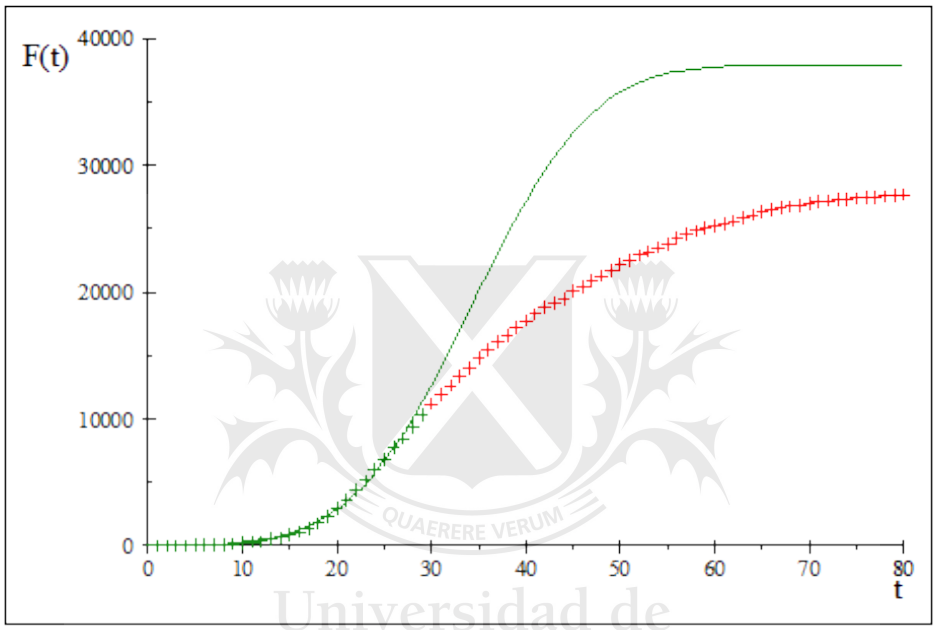
$$\begin{aligned} k &= 2 \times 5.165\,6 \times 10^{-3} = 0{,}010331 \\ t_0 - \varphi &= \frac{0{,}33813}{0{,}010331} = 32.7 \end{aligned}$$

Como anticpamos más arriba, observamos que siguiendo este método el valor de k no es confiable. Y aunque el valor de $t_0 - \varphi = 34{,}7$ tampoco lo es, nos da la pauta de que el t_0 real debe ser menor que 35 (obtenido a partir de la curva de infectados). Por ello, para el resto de los países utilizaremos solamente la curva de infectados activos para estimar el valor de k. A continuación se observan los resultados para los distintos países utilizando la misma metodología. El desarrollo matemático de las estimaciones se encuentra en el Apéndice C.

QUAERERE VERUM
Universidad de
SanAndrés

3.2. España

España, al igual que Italia, al principio del brote de Covid-19 no impuso ningún tipo de restricción. Del 15-27 de marzo se declaró un Estado de Alarma [5] cuando ya contaban con 10.000 casos y 136 muertos. A esto le siguió, una cuarentena estricta hasta el 12 de abril [6] donde suspendieron todas las actividades no esenciales. Como sabemos, los efectos de una cuarentena se observan en una baja del incremento del número de infectados a los 15 días y en el número de fallecidos con ese mismo lag o aún más tarde.



QUAERERE VERUM
Universidad de
SanAndrés

Figura 3.4: España - Cantidad de fallecidos desde 05/03/2020

Como podemos observar en la Figura 3.4, los efectos deberían notarse en principio a partir de $t = 25$. Desde $t = 30$ en adelante, se observa un cambio en la tendencia indicando que hubo una variación en los parámetros de la curva teórica. Sin embargo, durante el tiempo que las condiciones iniciales se mantuvieron estables, el modelo sigue los datos empíricos (hasta $t = 30$). En el caso de España, a diferencia de Italia, las medidas de restricción provocaron una variación en los parámetros inciales que llevaron a que la desviación de la estimación final sea del orden del 36 %, es decir un aplanamiento más pronunciado de la curva teórica.

[5] https://www.theguardian.com/world/2020/mar/14/spain-government-set-to-order-nationwide-coronavirus-lockdown

[6] https://elpais.com/espana/2020-03-28/el-gobierno-amplia-el-confinamiento-los-trabajadores-de-actividades-no-esenciales-deberan-quedarse-en-casa.html

3.3. Alemania

Alemania, a diferencia de España y Italia, no impuso medidas de restricción estrictas para detener el desarrollo de la epidemia. Desde el 22 de marzo al 3 de abril las autoridades alemanas suspendieron reuniones de más de 2 personas, el cierre restaurantes y servicios, e impuso el uso obligatorio de máscaras de protección[7]. Sin embargo, se permitió el resto de las actividades manteniendo el distanciamiento social y los debidos protocolos sanitarios.

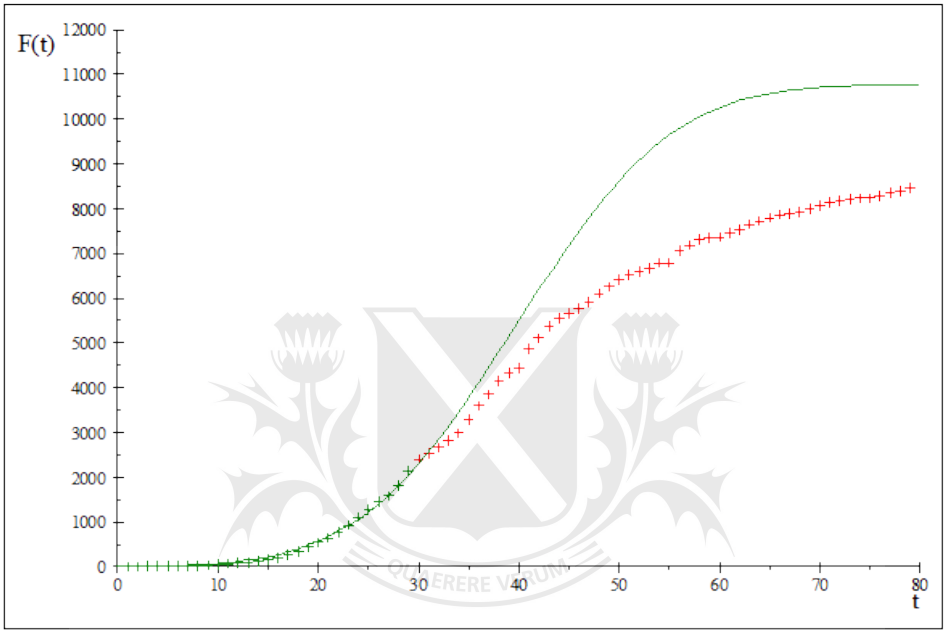


Figura 3.5: Alemania - Cantidad de fallecidos desde 09/03/2020

A pesar de que la cuarentena no fue estricta, los efectos de dichas medidas de distanciamiento social se deberían observar a partir de $t = 28$. Como observamos graficamente, a partir de $t = 30$ se observa un aplanamiento de la curva que lleva a que la desviación de la estimación final sea del orden del 27 %. Sin embargo, durante el tiempo en que las condiciones iniciales se mantuvieron estables, el modelo siguió los datos empíricos hasta $t = 45$. Podemos afirmar que Alemania posiblemente logro un aplanamiento similar al de España sin tener que llegar a una cuarentena estricta otorgandole mayor prioridad a las medidas sanitarias y de distanciamiento.

[7] https://www.welt.de/politik/deutschland/article206725829/Coronavirus-Deutschland-Kontaktverbote-zu-mehr-als-zwei-Personen-Friseure-zu.html

3.4. Reino Unido

Reino Unido, en contraste con España y Italia, no impuso una cuarentena obligatoria para detener el desarrollo de la epidemia. El 16 de marzo el primer ministro Boris Johnson[8] aconsejo no asistir a bares ni restaurantes, salir solamente para realizar comprar esenciales, comenzar a hacer teletrabajo y evitar, en la medida de lo posible, realizar viajes. Ante el desborde en las cantidad de casos, el 17 de marzo se suspendieron todas las operaciones no urgentes en los hospitales para liberar 30.000 camas. Sin embargo, el 20 de marzo[9] se ordenó el cierre de bares, restaurantes y gimnasios, y se le pidió a los ciudadanos no salir de sus casas (para dicha fecha ya habian registrado 158 muertes y unos 3300 infectados).

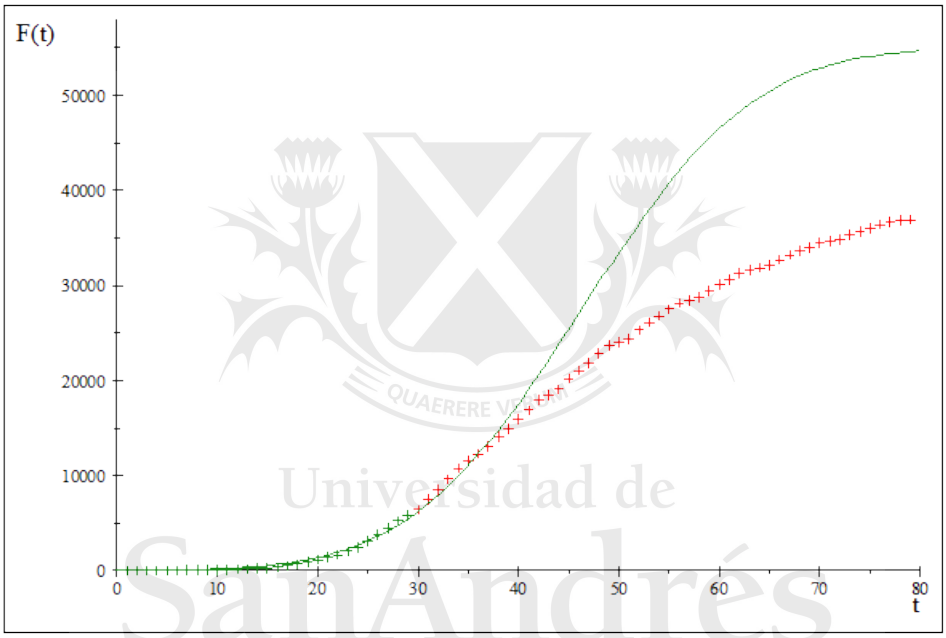


Figura 3.6: Reino Unido - Cantidad de fallecidos desde 09/03/2020

Si esta última medida fue acatada por los ciudadanos, los efectos deberían verse a partir de $t = 26$. A partir de la Figura 3.6, se observa un aplanamiento a partir de $t = 35$ que lleva a una desviación de la estimación final del orden del 47 %. Sin embargo, las condiciones iniciales se mantuvieron estables hasta $t = 45$ donde el modelo siguió aproximadamente a los datos empíricos. A pesar de que se aplicó tardiamente una cuarentena estricta, se observa una variación en los parámetros que provocó un aplanamiento de la curva teórica.

[8] https://www.gov.uk/government/speeches/pm-statement-on-coronavirus-16-march-2020
[9] https://www.gov.uk/government/speeches/pm-statement-on-coronavirus-20-march-2020

Universidad de San Andrés

3.5. Bélgica

Bélgica, al igual que el Reino Unido, tomo las medidas de forma tardía. El 12 de marzo[10] ordenó cerrar escuelas, cafés, restaurantes y bares, pero no se le pidió a la gente que se quede en casa. El 20 de marzo[11] cerró sus fronteras a todos los viajes no-esenciales. Estas medidas fueron sostenidas hasta el 3 de mayo[12]. Si las médidas fueron efectivas, los efectos se deberían notar en el gráfico de la Figura 3.7 a partir de $t = 15$.

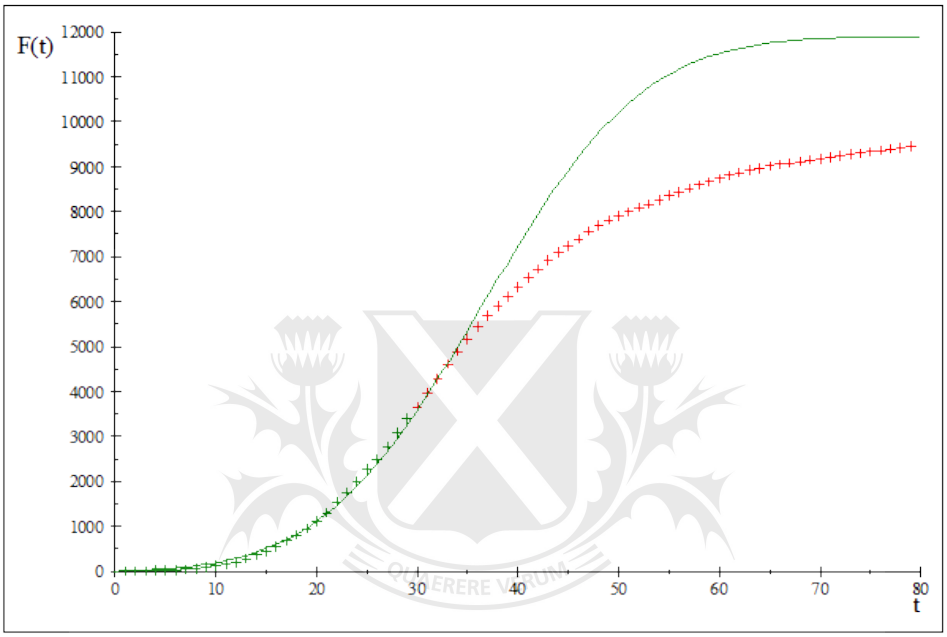


Figura 3.7: Bélgica - Cantidad de fallecidos desde 12/03/2020

Sin embargo, los datos se desvian de la tendencia en $t = 35$, es decir cuando los efectos del cierre de fronteras empezaron a mostrar resultados (pasados los 15 días del cierre). Esto nos podría indicar que el cierre de fronteras era una medida igual de primordial que el distanciamiento social. A partir de este punto, se observa un aplanamiento de la curva que lleva a que la desviacion de la estimación final sea del orden del 25 %. Sin embargo, durante el tiempo en que las condiciones iniciales se mantuvieron estables (hasta $t = 45$), el modelo siguió a los datos empíricos.

[10] https://www.belgium.be/en/news/2020/coronavirus_reinforced_measures
[11] https://www.vrt.be/vrtnws/en/2020/03/20/borders-closed-to-non-essential-travel/

[12] https://crisiscentrum.be/nl/news/crisisbeheer/verlenging-van-de-maatregelen-tem-3-mei

Universidad de San Andrés

3.6. Francia

Francia, tomó medidas cuando tenía 48 muertes y 2300 casos. Impuso un lockdown desde el 17 de marzo al 10 de mayo cerrando colegios, servicios no esenciales y limitando el uso del transporte público[13][14]. Durante esta cuarentena, si se permitió el ejercicio al aire libre a diferencia de Italia y España.

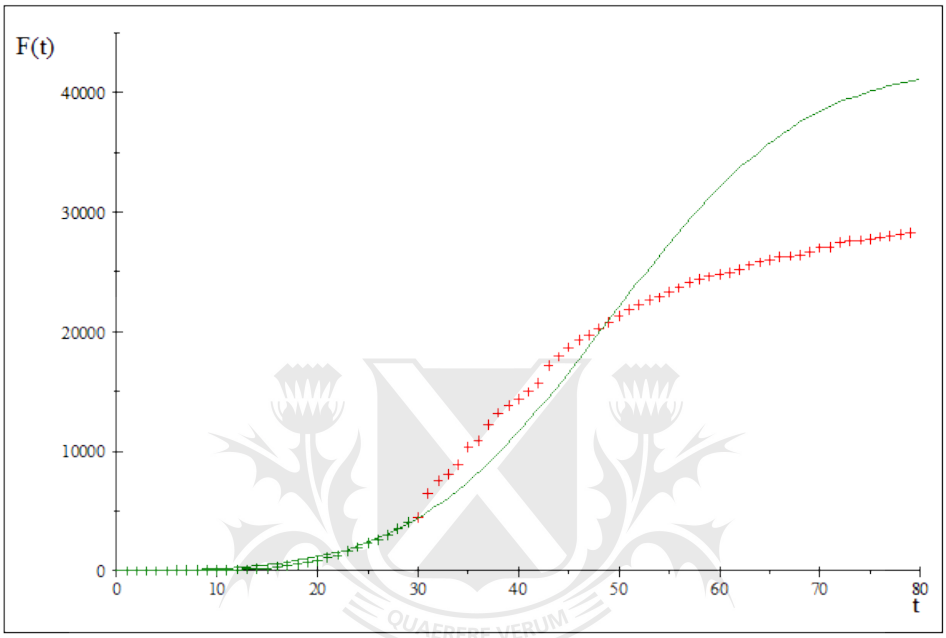


Figura 3.8: Francia - Cantidad de fallecidos desde 05/03/2020

Si las medidas fueron efectivas, los efectos se deberían notarse en el gráfico de la Figura 3.8 a partir de $t = 27$. Como podemos obsevar, los datos se desvian de la tendencia en $t = 31$. A partir de este punto, se observa un salto brusco en la pendiente de la curva que lleva a que la desviacion de la estimación final sea del orden del 45 %. Lo llamativo del presente caso es que hay un cambio de régimen muy notorio. En este caso no lo podemos interpretar unicamente como un aplanamiento de la curva, sino como un salto discontinuo de parámetros en $t = 31$ y luego un aplanamiento paulatino de esa nueva sigmoide como consecuencia de las medidas de restricción.

[13] https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000041746694&dateTexte=&categorieLien=id
[14] https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000041722656&dateTexte=&categorieLien=id

Universidad de San Andrés

3.7. Holanda

Holanda, tuvo una política distinta al resto de los países, al igual que Reino Unido y Alemania buscó en principio evitar el confinamiento obligatorio y estricto. El 12 de marzo suspendió clases presenciales y el 15 marzo[15] se confirmo que las escuelas, bares, restaurants y clubes seguirian cerrados en principio hasta el 3 de abril. El 16 de marzo, Mark Rutte[16] anunció que iban a buscar reforzar las medidas tempranas y buscar la inmunidad de rebaño mientras se controlaba la epidemia.

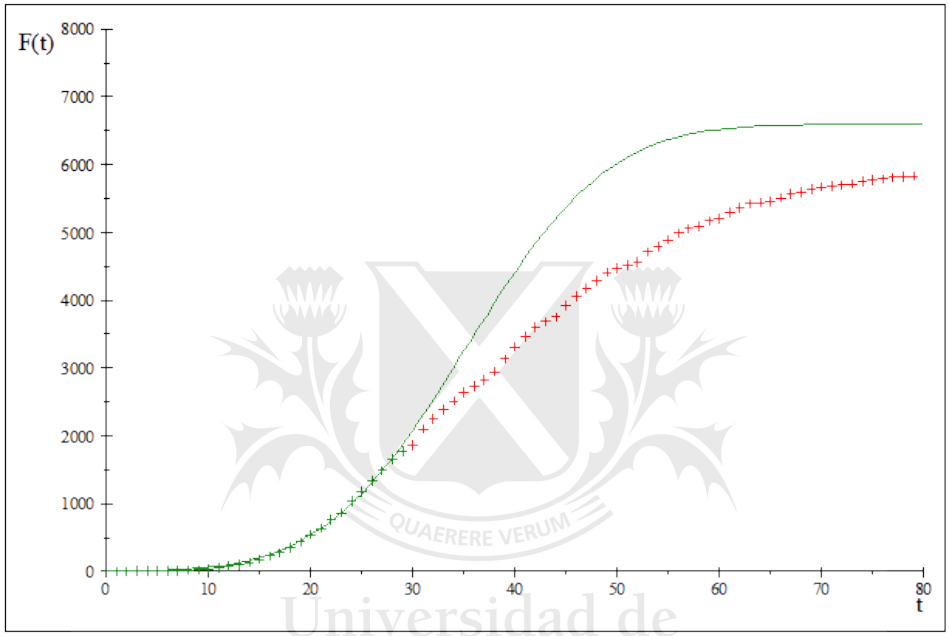


Figura 3.9: Holanda - Cantidad de fallecidos desde 09/03/2020

Si las medidas de distanciamiento social fueron efectivas, los efectos se deberían notar en el gráfico de la Figura 3.9 a partir de $t = 21$. Como podemos obsevar gráficamente, los datos se desvían de la tendencia en $t = 30$. A partir de este punto, se observa un aplanamiento de la curva que lleva a que la desviacion de la estimación final sea del orden del 13 %. Como consecuencia de las medidas de restricción no alteraron significativamente los parámetros iniciales, la curva se aplanó paulatinamente.

[15] https://nos.nl/liveblog/2326799-grand-prix-van-australie-gaat-niet-door-belgie-sluit-horeca-tot-3-april.html
[16] https://www.government.nl/documents/speeches/2020/03/16/television-address-by-prime-minister-mark-rutte-of-the-netherlands

Universidad de SanAndrés

3.8. Turquía

Turquía cerró tempranamente sus fronteras en Febrero. En marzo comenzo una campaña de desinfección del transporte público y los lugares públicos. El 31 de marzo[17] declaro un confinamiento con la excepción de todo aquellas personas que provean bienes y servicios esenciales por 15 días.

Finalmente, esta medida se extendió a todo abril y las medidas de desconfinamiento comenzaron en mayo[18].

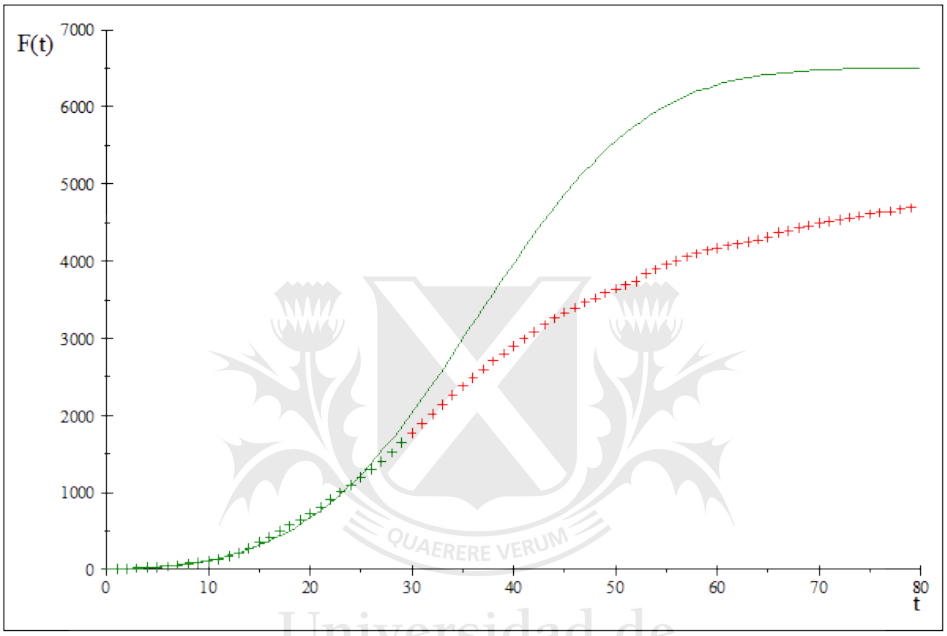


Figura 3.10: Turquía - Cantidad de fallecidos desde 20/03/2020

Si las medidas de distanciamiento social fueron efectivas, los efectos se deberían notar en el gráfico de la Figura 3.10 a partir de $t = 26$. Como consecuencia de la variación de los parámetros provocado por las medidas tomadas, los datos se desvian de la tendencia en $t = 30$. A partir de este punto, se observa un aplanamiento pronunciado de la curva que lleva a que la desviacion de la estimación final sea del orden del 38 %. Al igual que en el caso de Francia, no se observa la forma de sigmoide que esperamos y el pronostico es robusto hasta $t = 40$.

[17] https://www.ttb.org.tr/haber_goster.php?Guid=7d00f1fc-7384-11ea-b12d-d839943d748d

[18] https://www.aa.com.tr/tr/turkiye/cumhurbaskani-erdogan-normal-hayata-donusu-kademe-kademe-baslatacagiz/1828617

Universidad de San Andrés

3.9. EEUU

El caso de EEUU es el que mejor deja en evidencia una de las grandes límitaciones de nuestro modelo que comparte con el modelo SIR: la ausencia de una variable que capture la dimensión espacial. Por ello, analizaremos el caso de la Ciudad de Nueva York, lugar donde se ocurrio el primer y principal brote de coronavirus, que por ser un espacio más localizado se espera que se sostengan las hipótesis de nuestro modelo. A partir de ello, evaluaremos si es posible realizar una estimación global bajo nuestra metodología que al menos nos pueda brindar un pronóstico del orden de magnitud del desarrollo final de la epidemia.

Ciudad de Nueva York

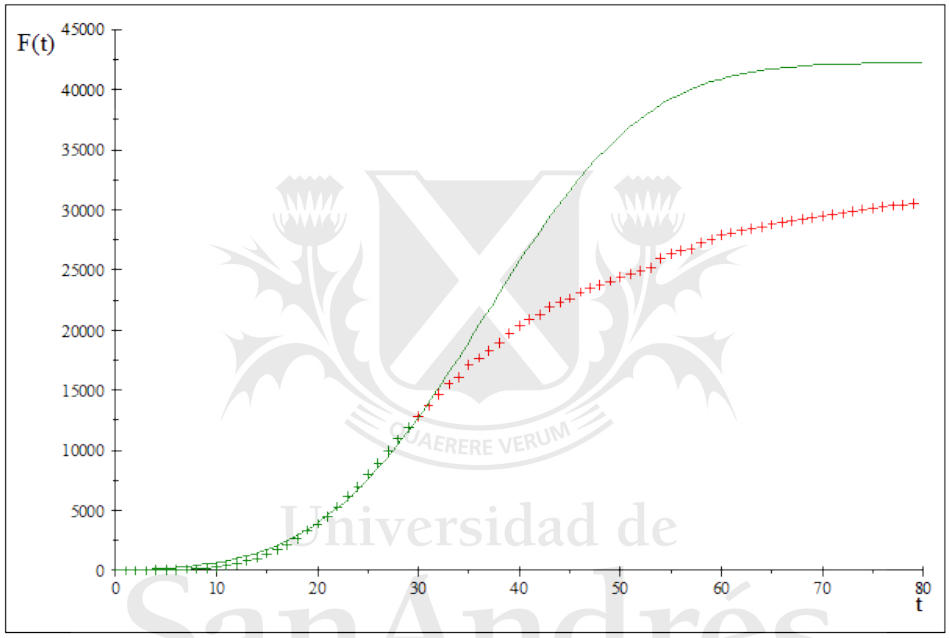


Figura 3.11: Ciudad de Nueva York - Cantidad de fallecidos desde 14/03/2020

Nueva York tuvo un brote de casos descontrolado y el 15 de abril se declaró obligatorio el uso de mascarilla donde no se fuera posible el distanciamiento social. El 1 de mayo cerraron escuelas y universidades[19]. Si las medidas de distanciamiento social fueron efectivas (uso de mascarillas), los efectos se deberían notar en el gráfico de la Figura 3.11 a partir de $t = 30$. Si consideramos la medida de cerrar escuelas y universidades, entonces los efectos deben verse en $t = 40$. Como podemos obsevar gráficamente, los datos se desvian de la tendencia en $t = 30$ y más notoriamente en $t = 40$. A partir de este punto, se observa un aplanamiento brusoc de la curva que lleva a que la desviacion de la estimación final sea del orden del 38 %. Este es otro caso donde la forma de sigmoide final muestra un cambio significativo de sus parámetros (condiciones iniciales) de la curva teórica provocado por las medidas de restricción.

[19]https://www.nytimes.com/2020/03/15/nyregion/nyc-schools-closed.html

EEUU

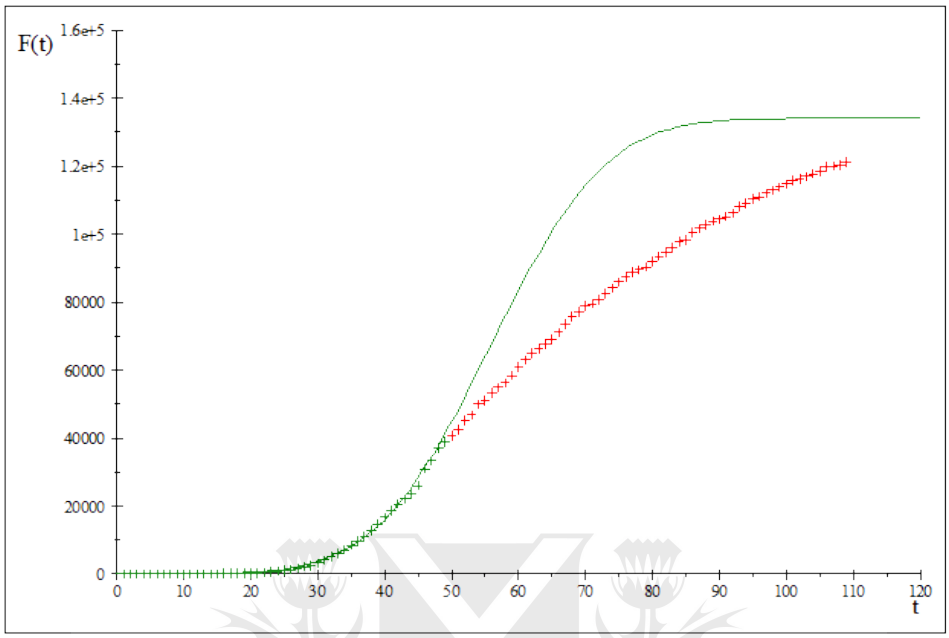


Figura 3.12: EEUU - Cantidad de fallecidos desde 02/03/2020

No podemos mencionar las respectivas medidas en EEUU porque no se tomaron medidas homogéneas a nivel país, es decir cada estado adoptó politicas de distanciamiento social y confinamiento diferentes. El 18 de marzo, se cerraron parcialmente las fronteras y solo se permitieron los viajes esenciales. Por lo tanto, a partir del gráfico de la Figura 3.12 se puede observar un cambio en la tendencia en $t = 50$, correspondiente al día 15/04, fecha estimada en la cual se produjo el pico de muertos en la Ciudad de Nueva York. Si el resto de los estados no estarían comenzando con sus brotes locales, la estimación que observamos sería confiable. A partir de este punto, se observa un cambio en los parámetros de la curva teórica.

Como conclusión, se podría proponer un modelo gaussiano por ciudad o región con sus respectivos desfasajes y luego agregarlas a una estimación global. Sin embargo, con la evolución temprano de los brotes en algunas ciudades todavia no se puede realizar una estimación robusta sobre la evolución de la epidemia. Como podemos observar, en un país como EEUU, cuyo superficie es tan extensa (3era a nivel mundial) la variable geográfica es determinante para lograr un correcto modelo predictivo.

Universidad de San Andrés

3.10. Comentarios finales

A continuación se presentan los resultados de los parámetros de modelo gaussiano por país[20].

País	k	I_0	N (*millones*)	$F(\infty) - F_($ exp) (%)
ESP	0,010	131.660	45	36
IT	0,0070	115.960	60	6
ALE	0,0067	113.600	83	27
BEL	0,0063	30.485	11	25
FRA	0,0043	182.960	67	45
HOL	0,0084	20.618	17	13
TUR	0,0059	100.001	82	38
UK	0,0055	150.690	56	47
NY	0,0063	30.394	8.4	38

Cuadro 3.1: Resultados de los parámetros del modelo gaussiano por país.

Como podemos observar, el resultado del valor de k para el grupo de países bajo análisis se encuentra en el intervalo (0.0043, 0.0100). Para este grupo de países, el k promedio es 0.0067. Esto nos indica que la variable que mide el ancho de la gaussiana es estable en países con distintas poblaciones totales y distintas densidades poblacionales. A pesar de que tenemos países como Bélgica, ocho veces menor en población respecto de Turquía, el k no varía significativamente. Por otro lado, la amplitud del modelo I_0 si tiene una fuerte relación con la población total. Casos como Holanda, Bélgica o Nueva York, nos muestran que la amplitud es menor a menor población. Más adelante, desarrollaremos el vínculo dinámico que existe entre N y I_0.

Respecto al análisis previo, hay dos puntos a remarcar. Primero, durante un intervalo de tiempo, en que las condiciones iniciales se mantuvieron estables respetando las normas básicas de cualquier experimento, el modelo siguió a los datos empiricos en todos los países bajo análisis. Segundo, como consecuencia de las variaciones en los parámetros provocado por las distintas medidas de restricción que adoptaron los países, la curva se aplana desviándose del modelo incial. En algunos casos como Italia, Bélgica, Holanda y Alemania, el desvio es suave y paulatino; en cambio, en casos como España, Francia, Turquía, Reiuno Unido y Nueva York, el desvio es más pronunciado.

[20]Los resultados de la estimaciones se encuentran en el Apéndice C y los datos de la población total fueron extraidos de The World Bank.

Capítulo 4

Comparación con los modelos dinámicos

Como mencionamos anteriormente, el objetivo del modelo Gaussiano que planteamos no busca explicar las causas por las cuales dicha curva ajusta al número de infectados ni cuales son las limitaciones de dicha aproximación. Para ello, necesitamos introducir el modelo que representa el "caballo de batalla.[en] epidemiología, el cual explica la dinámica de poblaciónes de una epidemia. Pero primero comencemos por el modelo más simple.

4.1. Modelo SI o modelo logístico

El modelo dinámico más simple que uno puede plantear para analizar una epidemia es el modelo SI. Dicho modelo divide a la población a la población total N en dos grupos: los infectados I y la población sana susceptible de infectarse S.

$$N = S + I \tag{4.1}$$

La dinámica del modelo plantea el siguiente sistema de ecuaciones

$$\boxed{\begin{array}{l} \frac{dS}{dt} = -\beta IS \\ \\ \frac{dI}{dt} = \beta SI \end{array}} \tag{4.2}$$

La segunda ecuación no es otra cosas que la ecuación (2.4) que discutimos anteriormente en términos probabilísticos, la velocidad de infectados crece en forma proporcional al número de encuentros entre infectados y sanos SI, y la constante β no es otra cosa que la frecuencia media de encuentros f por la probabilidad p_c de transmisión en un encuentro individual

$$\beta = f p_c.$$

La segunda ecuación nos indica que la cantidad que se infecta por unidad de tiempo, es la misma en que la población sana disminuye, de modo tal, que en todo tiempo la suma de ambas poblaciónes permanezca constante

$$\frac{d(S+I)}{dt} = \frac{dN}{dt} = 0$$

Aunque el sistema tiene dos variables, el vínculo entre ambas poblaciónes permite eliminar una, por ejemplo

$$I = (N - S)$$

para reemplazarla en la segunda y obtener una ecuación diferencial dependiente de una sola de ellas

$$\frac{1}{I}\frac{dI}{dt} = \beta(N - I) \tag{4.3}$$

que es la ecuación diferencial originalmente planteada por Verhulst(1845)

$$K(t) = \frac{1}{I}\frac{dI}{dt} = \frac{\beta}{N}(1 - \frac{I}{N}) \tag{4.4}$$

Lo interesante para destacar de dicha ecuación es que en ella se ve claramente que la tasa de crecimiento exponencial $K(t)$, que al principio de la epidemia aproximadamente resulta $\frac{\beta}{N}$, se va agotando con el factor $(1 - \frac{I}{N})$ a medida que la proporción de infectados de la población total $\frac{I}{N}$ crece. Finalmente, cuando toda la población se infecta, el exponente sencillamente se agota. La solución de la ecuación de Verhulst(1845) es la función logística

$$I(t) = \frac{NI(0)e^{\frac{\beta}{N}t}}{N + I(0)(e^{\frac{\beta}{N}t} - 1)} \tag{4.5}$$

como podemos verificar por simple diferenciación, la constante $I(0)$ es la población inicial $I(0) = I(t = 0)$.

De esta solución es claro que al comienzo de la epidemia cuando $\frac{\beta}{N}t \ll 1$ domina el numerador y su evolución es exponencial,

$$I(t) = I(0)e^{\frac{\beta}{N}t}$$

mientras que para tiempos lo suficientemente grandes se termina infectando toda la población.

$$\lim_{t\to\infty} I(t) = N$$

El principal defecto del modelo logístico es que no tiene en cuenta que en toda epidemia parte de la población se recupera. Para cubrir dicha falencia, surgió el modelo epidemiológico SIR ($Susceptible - Infected - Recovered$).

4.2. Modelo SIR

El modelo epimediológico SIR (Murray, 2002) divide al conjunto total de la población N en trés grupos o subpoblaciónes de individuos:

$$\boxed{N = S + I + R} \tag{4.6}$$

Donde S es la población susceptible, individuos sin inmunidad al agente infeccioso, y que por tanto puede ser infectada si es expuesta. I es la población infectada de indiviuos que están infectados en un momento dado y pueden transmitir la infección a individuos de la población susceptible con la que entran en contacto. R es la población recuperada de individuos que son inmunes a la infección, y consecuentemente no afectan a la transmisión cuando entran en contacto con otros individuos.

La dinamica de las poblaciónes del modelo esta descripta por trés ecuaciones diferenciales de primer orden:

$$\boxed{\begin{array}{ll} \frac{dS}{dt} = -\beta IS & (a) \\ \\ \frac{dI}{dt} = \beta SI - \gamma I & (b) \\ \\ \frac{dR}{dt} = \gamma I & (c) \end{array}} \tag{4.7}$$

donde β es la tasa de transmisión y γ la tasa de recuperacion. En base a lo que ya hemos expuesto con nuestro modelo es sencillo entender el significado de las ecuaciones.

La ecuación (a), al igual que en el modelo anterior, nos indica que la población de susceptibles disminuye en el tiempo proporcionalmente a una cierta tasa de trasmición β y a la cantidad de infectados y susceptibles en ese instante de tiempo βIS.

La ecuación (b) es la ecuación del modelo SI agregando una tasa de recuperacion por unidad de tiempo. La misma nos dice que βSI es la cantidad que se infecta por unidad de tiempo, pero para obtener la velocidad total de crecimiento de infectados hay que restarle la velocidad a la que la población se recupera $\frac{dR}{dt}$.

La ecuación (c) nos señala que la velocidad a la cual se recupera la población es proporcional a la tasa γ de recuperación por la población infectada en ese instante. Nuevamente, la probabilidad de incrementar la población recuperada es proporcional a la población infectada en ese instante.

$$dR = p_r I(t) dt$$

el coeficiente γ es p_r.

Notemos que sumando miembro a miembro las tres ecuaciones del modelo tenemos que

$$\frac{d(S+I+R)}{dt} = 0 \tag{4.8}$$

por lo que las ecuaciones de la dinámica de poblaciónes son compatibles con el vínculo $N = S + I + R = cte$. Dichas ecuaciones indican que en todo momento de la epidemia la población total de individuos es constante. Esto refleja que plantea un sistema aislado, no capturan a la población fallecida y tampoco posibles flujos migratorios. Bajo el modelo SIR los fallecidos son contabilizados en la categoría de infectados. Lo que resulta un claro error contable, ya que los muertos no se recuperan ni son infectados activos que puedan contagiar. No obstante ello, si la tasa de mortalidad es baja esta es una aproximación tolerable.

Como son ecuaciones de primer orden, para integrar dicho sistema es necesario conocer los datos de las poblaciónes iniciales $S(0)$, $I(0)$ y $R(0)$, y además los coeficientes β y γ.

4.3. Más allá del modelo SIR

El modelo SIR como todo modelo matemático hace hipótesis simplicadas de la realidad. Como consecuencia de ello tiene sus limitaciones. Una de las limitaciones es que no contabilza la cantidad de muertos que produce la

Universidad de San Andrés

epidemia, simplemente los agrupa en el grupo de infectados. Una extensión del modelo que corrige esa falencia, divide a la población en *Susceptible* (S), $Infected$ (I), *Recovered* (R) y *Deaths* (F).

$$N = S + I + R + F \tag{4.9}$$

Este nuevo modelo denominado modelo $SIRD$, fue utilizado para describir la Epidemia del Ébola (Osemwinyen y Diakhaby, 2015) y tiene el sistema de ecuaciones,

$$\boxed{\begin{array}{ll} \frac{dS}{dt} = -\beta IS & (a) \\ \frac{dI}{dt} = \beta SI - \gamma I - \mu I & (b) \\ \frac{dR}{dt} = \gamma I & (c) \\ \frac{dF}{dt} = \mu I. & (d) \end{array}} \tag{4.10}$$

La nueva ecuación del sistema, la ecuación (d), contabiliza un nuevo grupo: los fallecidos, dicha subpoblación crece a una tasa de mortalidad μ (que es la probabilidad de que un individuo infectado muera por unidad de tiempo) por unidad de tiempo. La ecuación (b) del sistema SIR también se ve modificada restando a la velocidad de crecimiento de infectados también la velocidad de fallecidos $\frac{dD}{dt}$.

Modelos más complejos pueden verse en libros como *Mathematical Biology* (Murray, 2002) y *Mathematical Epidemiology* (Brauer et al, 2008) donde se divide a la población en distintos grupos etarios y se les agrega un riesgo de infección distinto a cada grupo. Recientemente, Acemoglu *et al.* (2020) presentaron un Multi-Risk SIR model teniendo en cuenta distintos grupos etarios con distintos riesgos de infección y a su vez, contemplando variables de encierro y target de las políticas aplicadas. A partir de ello, realizaron simulaciones teóricas de robustez del modelo variando los parámetros. En definitiva, todos estas extensiones del modelo original SIR o modelos epidemiológicos presentan una ecuación (b) modificada para la población de infectados que admite la forma general del modelo Gaussiano.

$$\frac{1}{I}\frac{dI}{dt} = K(t)$$

Por ello, podemos afirmar que el modelo Gaussiano que elegimos en realidad engloba una serie amplia de modelos. El objetivo de la sección anterior era mostrar la ventaja predictiva de nuestro modelo presentando los modelos más simples de la literatura. A medida que se complejizan los modelos y se aumenta el número de parámetros o variables explicativas, es necesario contar cada vez con información más completa y precisa para poder integrar el sistema de ecuaciones. Este es el problema en el cual caemos frecuentemente los economistas, la maldición de la dimensionalidad. Cada vez que agregamos una variable explicativa, necesitamos aumentar exponencialmente el número de observaciones. Es decir, que para realizar predicciones en la mayoría de los casos es mejor recurrir a modelos menos precisos pero más robustos que presenten la menor cantidad de parámetros posibles. Siempre nos enfrentamos al dilema del *trade off* entre poder explicativo vs poder predictivo.

Universidad de San Andrés

Capítulo 5

Aproximación Gaussiana del modelo SIR

Cuando más arriba utilizamos el análisis cinemático, lo que hicimos no fue plantear el sistema completo de ecuaciones diferenciales de primer orden sino centrarnos en la ecuación (b) de la población de infectados que corresponde a una tasa de crecimiento del modelo gaussiano

$$K(t) = \beta S - \gamma$$

La hipótesis que hicimos para llegar a la curva Gaussiana

$$\begin{aligned} I(t) &= P_0 \exp -\frac{k}{2}(t - t_0)^2 \\ I(0) &= P_0 \exp -\frac{k}{2}{t_0}^2 \end{aligned}$$

es que hay un régimen en que la tasa $K(t)$ disminuye linealmente con el tiempo como

$$K(t) = -k(t - t_0)$$

En un modelo SIR simplificado donde se asume que las tasas β y γ son constantes en el tiempo, eso se traduce a que $S(t)$ debe ser una función lineal.

Integrando numéricamente el sistema se obtiene una solución del tipo:

QUAERERE VERUM
Universidad de
SanAndrés

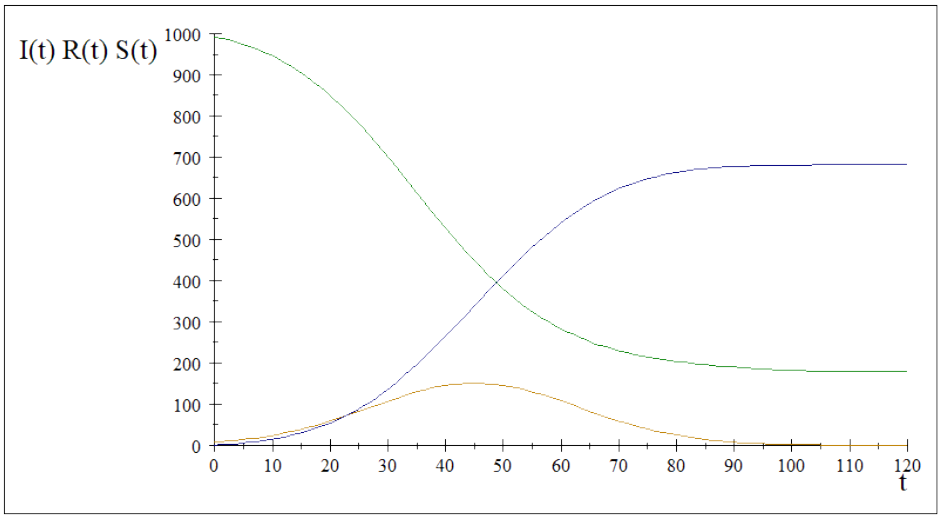


Figura 5.1: $I(r)$ (amarillo), $R(t)$ (azul), $S(t)$ (verde)

donde la curva verde es la población suceptible S, la amarilla la infectada I y la azul la recuperada R. Notemos primero que nuestra hipótesis genérica de que $K(t) = \beta S - \gamma$ es una función decreciente con el tiempo se satisface ya que $S(t)$ decrece en forma monótona con el tiempo. Ademas, vemos que hay un intervalo temporal en que la curva verde se puede aproximar a una recta, la función $K(t)$ es lineal, y por ello, la curva amarilla de infectados se puede aproximar a una Gaussiana. Es importante aclarar que la aproximación Gaussiana es válida en el intervalo central de la curva amarilla y no en las colas.

5.1. Solución analítica completa del modelo SIR en la aproximación Gaussiana

La aproximación Gaussiana ofrece una solución analítica completa del modelo SIR.

Como las soluciones exactas requieren una integración numérica, la ventaja de obtener expresiones sencillas aunque aproximadas, es que permiten obtener conclusiones generales en base a los parámetros del modelo sin recurrir a soluciones más complejas.

Una vez obtenida la solución Gaussiana de la la ecuación (b)

$$I(t) = I_0 \exp -\frac{k}{2}(t - t_0)^2 \tag{5.1}$$

nuestra solución se completa integrando la ecuación (c), que es simplemente el área bajo la curva o la función error al igual que en el caso de los fallecidos

$$R(t) = \gamma \int_0^t I(\tau)d\tau = I(t) = \gamma I_0 \int_0^t \exp -\frac{k}{2}(\tau - t_0)^2 d\tau \tag{5.2}$$

La población de suceptibles se obtiene utilizando la ecuación que vincula a las subpoblaciónes,

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

$$S(t) = N - I_0 \exp -\frac{k}{2}(t - t_0)^2 - \gamma I_0 \int_0^t \exp -\frac{k}{2}(\tau - t_0)^2 d\tau \tag{5.3}$$

Elegimos los parámetros

$$\begin{aligned} k &= 0{,}002 \\ I_0 &= 150 \\ t_0 &= 45 \\ \gamma &= 0{,}1 \\ N &= 1000 \end{aligned}$$

y obtenemos las curvas de poblaciónes, dichos parámetros fueron ajustados de modo tal que el dato inicial N y

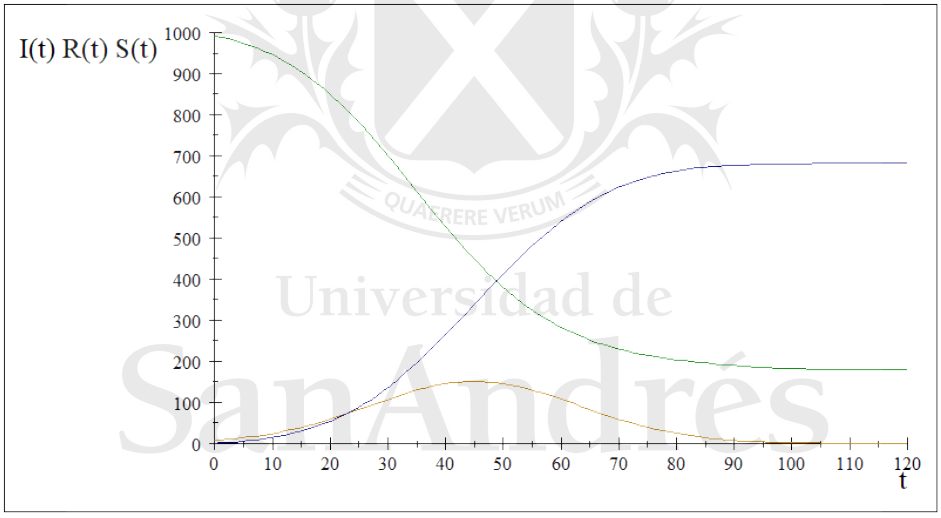


Figura 5.2: $I(r)$ (amarillo), $R(t)$ (azul), $S(t)$ (verde)

el pico (N_0, t_0) coincidiera con el del gráfico de la solución exacta anterior mencionado. Los parámetros γ y k se eligieron de modo tal que las poblaciónes finales de $S(t)$ y $F(t)$ también coincidieran con dicha solución.

QUAERERE VERUM
Universidad de San Andrés

5.2. Comparación con la aproximación de Kermack & McKendrick (1927)

En su trabajo original Kermack y McKendrick obtuvieron una solución aproximada a segundo orden en el parámetro $\frac{\beta N}{\gamma}$, de las ecuaciones de modelo SIR. En lugar de un pico Gaussiano obtuvieron un pico para la población de infectados con la forma de una secante hiperbólica cuadrado (Kermack y McKendrick 1927,p.713)[1]

$$I(t) = \frac{I_0}{\cosh^2 \lambda(t - t_0)}, \tag{5.4}$$

donde el parámetro $\lambda = \frac{\alpha\gamma}{2}$, siendo $\alpha \simeq \frac{\beta N}{\gamma} - 1$, como mostramos en el Apéndice A, es decir

$$\lambda \simeq \frac{\gamma}{2}(\frac{\beta N}{\gamma} - 1) \tag{5.5}$$

Dicha expresión también permite calcular el número de fallecidos acumulados como el área bajo la curva de $I(t)$ multiplicado por la tasa de mortalidad

$$F(t) = \mu I_0 \int_0^t \frac{dt}{\cosh^2 \lambda(t - t_0)} = \frac{\mu I_0}{\lambda} \frac{2}{1 + e^{-2\lambda(t-t_0)}} \tag{5.6}$$

La expresión anterior, que es análoga a la distribución de Fermi-Dirac (Fermi, 1926); (Dirac, 1926), es al igual que la función error una curva sigmoidea que nos permite obtener el valores asintóticos de la cantidad total de fallecidos en la epidemia

$$F(\infty) = \frac{2\mu I_0}{\lambda} \tag{5.7}$$

y el área bajo la curva de infectados $(F(\infty)/\mu)$

$$\int_{-\infty}^{\infty} I_{KM}(t)dt = \frac{2I_0}{\lambda}. \tag{5.8}$$

En el modelo Gaussiano, si el pico de la Gaussiana esta lo suficientemente alejado del origen $t_0 \gg 0$ el área de la cola izquierda es despreciable frente a al área total de la curva, por lo que podemos escribir la siguiente expresión para la cantidad de infectados acumulados en el infinito

$$\int_{-\infty}^{\infty} I_G(t)dt = I_0 \int_{-\infty}^{\infty} \exp -\frac{k}{2}(t - t_0)^2 dt = I_0 \int_{-\infty}^{\infty} \exp -\frac{k}{2}t^2 dt.$$

Efectuando la integral Gaussiana, obtenemos la expresión

$$\int_{-\infty}^{\infty} I_G(t)dt = I_0\sqrt{\frac{2\pi}{k}}. \tag{5.9}$$

Igualando las expresiones (5.8) y (5.9) obtenermos la siguiente relación entre los parámetros λ y k de ambos modelos

$$k = \frac{\pi}{2}\lambda^2, \tag{5.10}$$

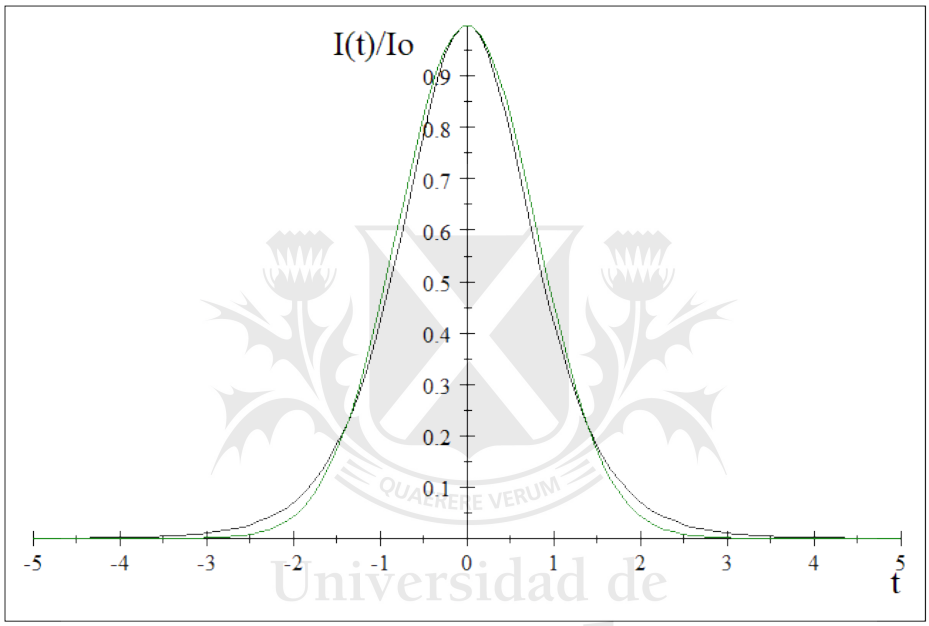
[1] En realidad escribieron la expresión para $\frac{dR}{dt}$ que es γI por la ecuación (c) del modelo SIR.

o reemplazando por (5.5)

$$k = \frac{\pi\gamma^2}{8}(\frac{\beta N}{\gamma} - 1)^2. \tag{5.11}$$

Bajo estas condiciones ambos modelos van a predecir la misma cantidad de fallecidos finales.

Nos preguntamos que forma han de tener ambos picos bajo esta relación de parámetros. Para ello, graficamos una Gaussiana y una secante hiperbólica centrada en $t_0 = 0$ de la misma altura I_0 y con $k = \frac{\pi}{2}, \lambda = 1$. El resultado se muestra en la Figura 5.3



Universidad de SanAndrés

Figura 5.3: Gaussiana (verde), sech^2 (negro)

$$\frac{I(t)}{I_0} = \frac{1}{\cosh^2(t)}$$

$$\frac{I(t)}{I_0} = \exp\left[-\frac{\pi}{4}t^2\right]$$

Vemos que con esta elección de parámetros ambos modelos predicen aproximadamente lo mismos resultados en el intervalo central en torno al pico. Las predicciones para el caso de los fallecidos acumulados de ambos modelos se pueden apreciar en la Figura 5.4 ($\mu = 1$)

$$\frac{F(t)}{I_0} = \frac{2}{1 + e^{-2t}}$$

$$\frac{F(t)}{I_0} = \int_{-\infty}^{t} \exp\left[-\frac{\pi}{4}\tau^2\right] d\tau$$

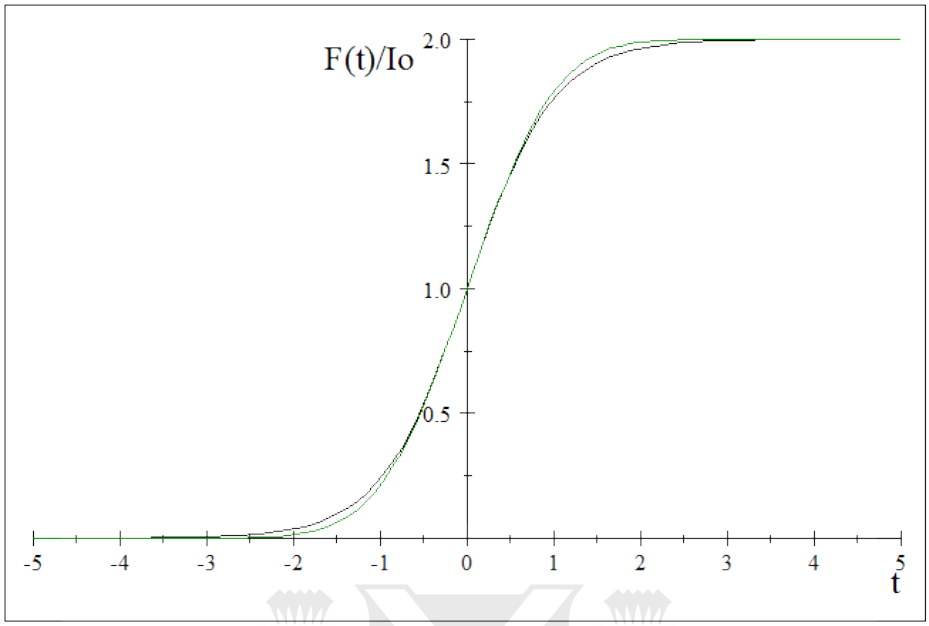


Figura 5.4: Modelo gaussiano (verde), modelo de la sech^2 (negro)

Resumiendo, el modelo Gaussiano es en principio un modelo cinemático con trés parámetros libres. Cuando ajustamos dichos parámetros con los del modelo de Kermack y McKendrick, el modelo Gaussiano a segundo orden en $\frac{\beta N}{\gamma}$ predice esencialmente los mismos resultados que el modelo SIR. Esta última identificacion de parámetros es importante porque nos permitirá hacer contacto con los parámetros epidemiológicos. La elección del modelo Gaussiano se basa en una razón de orden práctico: es más sencillo ajustar los datos con dicho modelo.

Universidad de SanAndrés

Capítulo 6

La tasa de reproducción básica $\rho_0 = \dfrac{\beta S(0)}{\gamma}$

En la sección anterior mencionamos el parámetro $\frac{\beta N}{\gamma}$, dicho parámetro es un número adimensional muy importante ya que define el desarrollo de toda epidemia. Para entender su significado consideremos nuevamente la dos últimas ecuaciones del modelo SIR

$$\begin{aligned} \frac{dI}{dt} &= \beta SI - \gamma I \qquad (b) \\ \frac{dR}{dt} &= \gamma I \qquad (c) \end{aligned}$$

Dichas ecuaciones, dividiendo miembro a miembro por γ, se pueden reescribir como

$$\boxed{\begin{aligned} \tfrac{1}{I}\tfrac{dI}{d(\gamma t)} &= \rho(t) - 1 \qquad (b) \\ \tfrac{dR}{d(\gamma t)} &= I \qquad (c) \end{aligned}} \tag{6.1}$$

donde

$$\rho(t) = \frac{\beta S(t)}{\gamma}$$

es un número adimensional que denominaremos tasa de reproducción instantánea. En particular como al comienzo de la epidemia el número de recuperados es nulo y el número de infectados es muy pequeño

$$S(0) = N - I(0) - R(0) \simeq N \tag{6.2}$$

dicho número es aproximadamente[1]

$$\rho_0 = \rho(0) \simeq \frac{\beta N}{\gamma}$$

Notese que en todo tiempo si $\rho(t) - 1$ es mayor que cero, o lo que es lo mismo $\rho(t) > 1$, la tasa de infectados es positiva y dicha población crece en el parámetro γt y en consecuencia la epidemia se desarrolla. Por el contrario

[1] De aquí en mas anotaremos el valor inicial $\rho(0) = \rho_0$ para respetar la convencion usual de la literatura. Pero no se debe confunfir el subindice 0 como una notación que hace alusión al pico de la epidemia.

si $\rho(t) < 1$, la población de infectados decrece. Ahi radica la importancia de $\rho(t)$ para el control de la misma. Precisamente cuando la epidemia alcanza el pico de infectados $\frac{dI}{d(\gamma t)}(t_0) = 0$ y en consecuencia $\rho(t_0) = 1$.

Para entender en una forma mas sencilla el significado de este número a nivel epidemiológico es instructivo eligir la escala de tiempos de tal modo que $d(\gamma t) = 1$, lo que equivale a tomar en lugar un día como unidad de tiempo, tomar un intervalos de tiempo igual a la inversa de la tasa de recuperación $dt = \frac{1}{\gamma}$.

En dichos intervalos uno puede leer la ecuaciones del modelo en términos de diferencias finitas que ocurren en lapsos $\Delta t = \frac{1}{\gamma}$

$$\begin{aligned} \frac{\Delta I}{I} &= \rho(t) - 1 \qquad (b) \\ \frac{\Delta R}{I} &= 1 \qquad (c) \end{aligned}$$

La ecuación (c) simplemente nos dice que trascurrido dicho lapso se recupera un infectado por período $\frac{1}{\gamma}$, por cada infectado que hay en la población. $\rho(t)$ se interpreta ahora como la tasa discreta con que crecen los infectados en todo instante. Supongamos por ejemplo que $\rho(t) = 3$, eso significa que en el lapso $\frac{1}{\gamma}$ en que se recupera 1, cada infectado contagia a 3 en promedio, por lo que generan $3 - 1 = 2$ nuevos infectados netos por cada infectado en dicho lapso.

QUAERERE VERUM
Universidad de
SanAndrés

Capítulo 7

Solución analítica exacta de la trayectoria $S(R)$ en el modelo SIR

Hasta ahora hemos considerado soluciones aproximadas a segundo orden del modelo. Las trayectorias de las poblaciónes $S(t)$, $I(t)$ y $R(t)$, pero el modelo admite una integral exacta de la trayectoria en el plano de fases $S-R$ (Murray, 2002).

En efecto podemos dividiendo miembro a miembro la primera por la tercera ecuación del modelo

$$\boxed{\begin{array}{ll} \frac{dS}{dt} = -\beta IS & (a) \\ \\ \frac{dR}{dt} = \gamma I & (c) \end{array}} \tag{7.1}$$

podemos eliminar el tiempo para obtener la ecuación

$$\begin{aligned} \frac{dS}{dR} &= -\rho_0 \frac{S}{S(0)} \\ \rho_0 &= \frac{\beta S(0)}{\gamma} \end{aligned}$$

Que podemos reescribir como

$$\frac{dS}{S} = d\ln S = -\rho_0 \frac{dR}{S(0)}$$

e integrado obtenemos

$$\ln \frac{S}{S(0)} = -\rho_0 \frac{R}{S(0)}$$

ya que $R(0) = 0$, es decir,

$$S(R) = S(0) e^{-\rho_0 \frac{R}{S(0)}} \tag{7.2}$$

la anterior es la ecuación exacta de la trayectoria de la poblaciónes en el plano $S - R$.

Volviendo a la ecuación (c) teniendo en cuenta que

$$I = N - R - S$$

podemos finalmente escribir

$$\frac{dR}{d(\gamma t)} = N - R - S(0)e^{-\rho_0 \frac{R}{S(0)}} \tag{7.3}$$

que es una ecuación exacta para la variable $R(t)$.

Kermack y McKendrick resolvieron esta última ecuación expandiendo la exponencial $e^{-\rho_0 \frac{R}{S(0)}}$ a segundo orden ρ_0 (Murray, 2002). Los detalles de dicho cálculo que conducen a la solución (5.4) se muestran en el Apéndice A.

En la sección siguiente vamos a continuar analizando la ecuación de la trayectoria $S = S(R)$ ya que dicha ecuación permite inferir los porcentajes de población que terminaran infectados, recuperados y todavía suceptibles al final de la misma.

7.1. Cálculo exacto del pico de infectados I_0

En la aproximación $S(0) \simeq N$, la ecuación de la trayectoria (7.2) se puede reecribir como

$$S = Ne^{-\rho_0 \frac{R}{N}} \tag{7.4}$$

Utilizando letras minúsculas para anotar las fracciones población

$$\begin{aligned} s &= S/N \\ r &= R/N \\ i &= I/N \end{aligned}$$

$$s + r + i = 1 \tag{7.5}$$

en términos de las fracciones de población s y r la ecuación de la trayectoria $s(r)$ es simplemente una exponencial

$$s = e^{-\rho_0 r} \tag{7.6}$$

cuyo único parámetro es la tasa de reproducción basica inicial ρ_0. O lo que es lo mismo invirtiendo la ecuación anterior para obtener r en función de s

$$r = -\frac{1}{\rho_0} \ln s \tag{7.7}$$

QUAERERE VERUM
Universidad de
SanAndrés

De la anterior se sigue entonces que la fracción de recuperados en el pico es

$$r_0 = -\frac{1}{\rho_0}\ln \rho_0^{-1} = \frac{1}{\rho_0}\ln \rho_0 \tag{7.8}$$

Por otro lado de la ecuación (6.1 (b)) del modelo SIR recordando que en el pico de la epidemia

$$\rho(t_0) = \frac{\beta S_0}{\gamma} = 1$$

se sigue que la población de susceptibes en el pico resulta

$$S_0 = \frac{\gamma}{\beta},$$

y por lo tant,o la fracción de susceptibles en el pico es simplemente la inversa de la tasa de reproducción básica inicial

$$s_0 = \frac{S_0}{N} = \frac{\gamma}{\beta N} = \frac{1}{\rho_0} \tag{7.9}$$

Finalmente, de la ecuación de vínculo entre las fracciones de población, la fracción de infectados en el pico es

$$i_0 = 1 - s_0 - r_0 = 1 - \frac{1}{\rho_0} - \frac{1}{\rho_0}\ln \rho_0 \tag{7.10}$$

La anterior es una expresión exacta. Es interesante comparar la misma con la expresión que se deriva de la solución de KM a segundo orden en ρ_0 que discutimos en el Apéndice A[1]

$$I(t_0) = \frac{N\left(1-\rho_0\right)^2}{2\rho_0^2} \tag{7.11}$$

que en términos de fracciones se reduce a

$$i_0 = \frac{1}{2}(1 - \frac{1}{\rho_0})^2 \tag{7.12}$$

Si graficamos en la Figura 7.1 ambas soluciones $i_0(\rho_0)$,

[1] La expresión obtenida por KM es $I(t_0) = \frac{S(0)\alpha^2}{2\rho_0^2}$ y en la aproximación $N \simeq S(0)$ el parámetro $\alpha \simeq \rho_0 - 1$.

Universidad de San Andrés

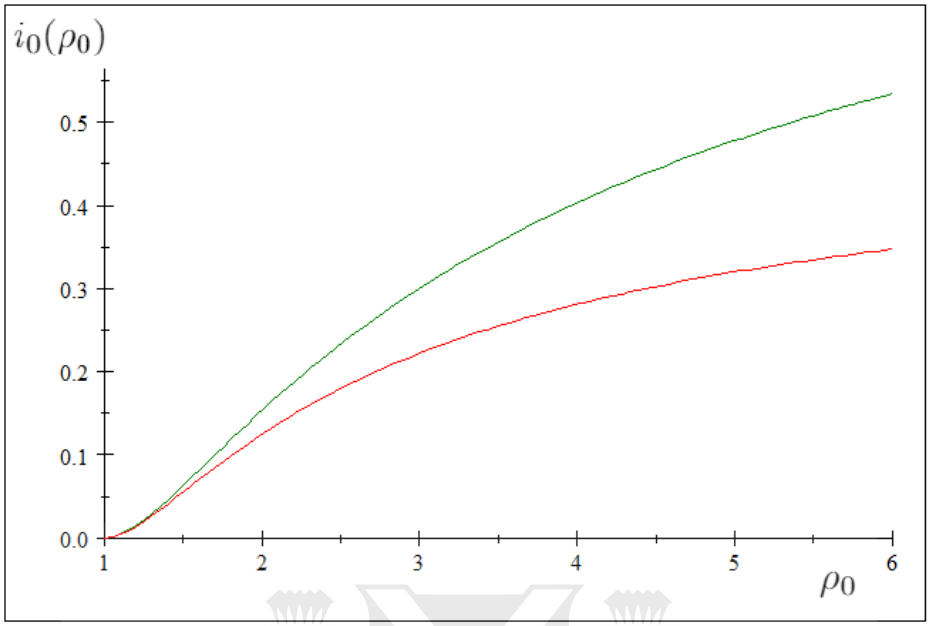


Figura 7.1: Solución exacta de $i_0(\rho_0)$ (verde) y a 2do orden (rojo)

vemos que la aproximación de segundo orden en rojo se aparta sensiblemente de la solución exacta en verde en epidemias en las que ρ_0 es elevado. Para epidemias en la que ρ_0 se mantiene por debajo de 1,5 la aproximación de KM es una buena aproximación. El la siguiente sección vamos a hacer una estimación del orden de magnitud del ρ_0 para el caso de la epidemia del Covid-19 en algunos países para testear cuan buena resulta la aproximación de segundo orden.

Universidad de SanAndrés

7.2. Estimación de ρ_0 para un grupo de países

La tasa de reproducción básica ρ_0, es de gran interés para la evaluación y planificación durante el brote inicial de toda epidemia. La razón de ello es que dicho parámetro determina completamente el impacto de la misma a nivel de la curva de infectados ya que como hemos visto el pico de infectados es solo una función de ρ_0 y la población inical N

$$I_0 = I_0(\rho_0, N) \tag{7.13}$$

y es sabido cuan importante es conocer y/o controlar dicho pico con anterioridad para que el mismo no sobrepase la capacidad del sistema de salud. Como hemos observado en los resultados de nuestro análisis previo, $I_0(\rho_0, N)$ es una función creciente de ρ_0 de modo que I_0 se puede amortiguar o reducir imponiendo medidas de distanciamiento social que disminuyan ρ_0.

Esta relación funcional nos permite realizar una estimación del ρ_0 para el grupo de países bajo análisis para los que conocemos los datos del pico de la muestra de infectados I_{0m}. Sin embargo, necesitamos normalizar la variable

I_{0m}. Como hemos mencionado al principio del análisis, los datos que tenemos es una muestra sesgada del nivel real de I_0, existe un porcentaje alto de infectados no capturados en los datos empíricos. Por lo tanto, tenemos que ahondar en dicha discusión para poder realizar una estimación robusta en el nivel de magnitud de ρ_0.

Al principio de la presente pandemia, se realizaron comparaciones, en cierto punto apresuradas, con la Gripe Española de 1918-1920 cuya tasa de mortalidad fue del 2 %. Como menciona (Barro y Weng, 2020) si la presente pandemia fuera similar, implicaría la muerte de 150 millones de personas en el mundo. Han pasado 6 meses desde que arrancó la epidemia y han muerto hasta el día de la fecha medio millón de personas. Por lo tanto, el tiempo transcurrido nos dió dos lecciones: las muestras de los infectados y los fallecidos ciertamente están sesgadas hacia los grupos de riesgo, y esta epidemia tiene una tasa de mortalidad menor de lo que se estima en base a los datos empíricos.

Atkeson (2020) realizo dos simulaciones utilizando el modelo SIR en dos escenarios, uno con un ρ_0 constante y un ρ_0 decreciente. Bajo esos escenarios, evaluó dos tasas de mortalidad $\mu = 0,01$ y $\mu = 0,001$. Su principal conclusión es que en las etapas tempranas de la epidemia donde hay baja estadística de mortalidad, es imposible determinar cuál es la tasa de mortalidad. Al igual que Manski y Molinari (2020), propone que hay que esperar a que se realicen testeos más masivos o algún estudio de testeo aleatorio masivo como el que propone Alemania.

Sin embargo, existen trabajos que abren el debate y buscan dar una estimación del porcentaje de asintomáticos que no es capturado en las estadísticas (Hortaçsu, Liu y Schwierg, 2020); (Stock *et al.*, 2020); (Bendavid *et al.*,2020); (Russell *et al.*, 2020). El estudio de Bendavid *et al.* (2020), testea a una muestra representativo de Santa Clara para el 1ero de abril y obtiene un ratio de reporte del 50-85 %. El trabajo de Hortaçsu, Liu y Schwierg (2020) utiliza los datos de EEUU y del testeo aleátorio llevado a cabo en Islandia para encontrar una estimación del porcentaje de infectados no capturados y de la tasa de mortalidad durante la primera mitad de marzo. Estiman un rango de tasa de reporte en los registros del orden del 4 % - 14 % cuando hay un lag de 8 días y una tasa de reporte del orden del 1,5 - 10 % cuando hay un lag de 5 días. Respecto a la tasa de mortalidad, sus resultados estiman una tasa de mortalidad real menor a 1,2 %. Sin embargo, sus resultados muestran una alta varianza en distintos países debido posiblemente a la heterogeneidad de la composición demográfica, y la calidad y capacidad del sistema sanitario.

El estudio de Stock *et al.*(2020) toma los datos públicos de dos regimenes de testeo islandeces durante el primer mes del brote en dicho país para estimar el porcentaje de casos no detectados en las estadísticas oficiales. Su principal conclusión es que el nivel de infectados no capturado es del orden del 89 %- 93 % en principio, y cuando se amplía el criterio de elegibilidad para ser testeado este orden baja a un 80 %- 90 %.

El trabajo de Russell *et al.* (2020) que analiza los datos de crucero Diamond Princess estiman un case fataly ratio de 2,6 % y un infected fataly ratio de 1,3 %. Sin embargo, como mencionan los autores los pasajeros de un crucero tienen un cierto status económico, la media de edad era 58 años y solamente fallecieron personas mayores a 70 años. Por lo tanto, estas conclusiones tienen limitaciones en su posible aplicación a otros países.

En resumen, los trabajos mencionados arriban a resultados que estan en linea con el nivel de positividad de los testeos que recomienda la WHO: entre un 3 % - 12 % para poder tener un buen seguimiento de la epidemia en los países. En nuestro caso, en base a lo expuesto, asumiremos que el promedio el 90 % de los infectados no fueron

Universidad de San Andrés

registrados para estimar el ρ_0 en función de I_0 para el grupo de países bajo análisis.

En otras palabras, vamos a multiplicar por un factor 10 los datos de la muestra de infectados en el pico que obtuvimos en nuestro análisis de datos en la Seccion 3 para obtener una estimación del porcentaje de población infectada en el pico $i_0 = I_0/N$ y con ello el orden de magnitud de ρ_0 invirtiendo la relación (7.12) que en principio matematicamente admite dos soluciones

$$\rho_0 = \frac{1}{1 \pm \sqrt{2i_0}}. \tag{7.14}$$

Nos quedamos con la de signo negativo que corresponde a la solución real de desarrollo de la epidemia ($\rho_0 > 1$). La solución con el signo positivo de la raiz corresponde a un $\rho_0 < 1$.

País	I_{0m} *(millones)*	I_0	N *(millones)*	$i_0 = I_0/N$	$\rho_0 = \frac{1}{1-\sqrt{2i_0}}$
ESP	$0,131$	1,31	45	0,029	1,32
IT	$0,116$	1,16	60	0,019	1,24
ALE	$0,114$	1,14	83	0,014	1. 20
BEL	$0,0305$	0,30	11	0,027	1. 30
FRA	$0,183$	1,83	67	0,027	1. 30
HOL	$0,0206$	0,20	17	0,012	1,18
TUR	$0,100$	1,00	82	0,012	1,18
UK	$0,151$	1,51	56	0,027	1. 30
NY	$0,0304$	0,30	8,4	0,036	1,36

Cuadro 7.1: Resultados de la estimación de ρ_0 para el grupo de países

Universidad de SanAndrés

A partir de los resultados obtenidos, observamos un intervalo de ρ_0 (1,18;1,36) para el grupo de países. Es decir, que los ρ_0 caen en el intervalo en el cual es válida la aproximación de segundo orden de KM.

7.3. La ecuación trascendente para la población asintótica de susceptibles $S(\infty)$

De la ecuación (7.2) vemos que al final de la epidemia ($t = \infty$) la población de susceptibles resulta

$$S(\infty) = S(0)e^{-\rho_0 \frac{R(\infty)}{S(0)}} \tag{7.15}$$

pero como al final de la misma la cantidad de infectados $I(\infty) = 0$ y la cantidad de recuperados $R(\infty) = N - S(\infty)$ tenemos

$$S(\infty) = S(0)e^{-\rho_0 \frac{N-S(\infty)}{S(0)}} \tag{7.16}$$

que es una ecuación trascendente para $S(\infty)$. Podemos hacer nuevamente la aproximación de que al comienzo de la epidemia $S(0) \simeq N$ de lo que se sigue que

$$\frac{S(\infty)}{N} = e^{-\rho_0(1-\frac{S(\infty)}{N})} \tag{7.17}$$

En términos de la fracción de susceptbles $s = S/N$ la anterior se reescribe como

$$s(\infty) = e^{-\rho_0(1-s(\infty))} \tag{7.18}$$

y una ecuación similar vale para la fracción de recuperados

$$r(\infty) = 1 - s(\infty) = 1 - e^{-\rho_0 r(\infty)} \tag{7.19}$$

ya que en el infinito la fracción de infectados es nula

$$i(\infty) = 0. \tag{7.20}$$

Las ecuaciones trascendentes anteriores, son soluciones exactas del modelo, por lo tanto importantes para evaluar los límites de la aproximación de segundo orden en ρ_0 del modelo Gaussiano (o su equivalente secante hiperbólica cuadrado) que hemos utilizado en las secciones previas.

La expansión a segundo orden de la ecuación trascendente para los recuperados nos da una ecuación

$$\begin{aligned} r(\infty) &= 1-(1-\rho_0 r(\infty)+\frac{\rho_0^2 r^2(\infty)}{2}) \\ r(\infty) &= \rho_0 r(\infty) - \frac{\rho_0^2 r^2(\infty)}{2} \\ 1 &= \rho_0 - \frac{\rho_0^2 r(\infty)}{2} \end{aligned}$$

de lo que finalmente resulta

$$\boxed{r(\infty) = \frac{2(\rho_0-1)}{\rho_0^2}} \tag{7.21}$$

Para evaluar cuan buena es la aproximación de segundo orden graficamos en la Figura 7.2 la misma contra la solución exacta de la ecuación trascentente.

Universidad de San Andrés

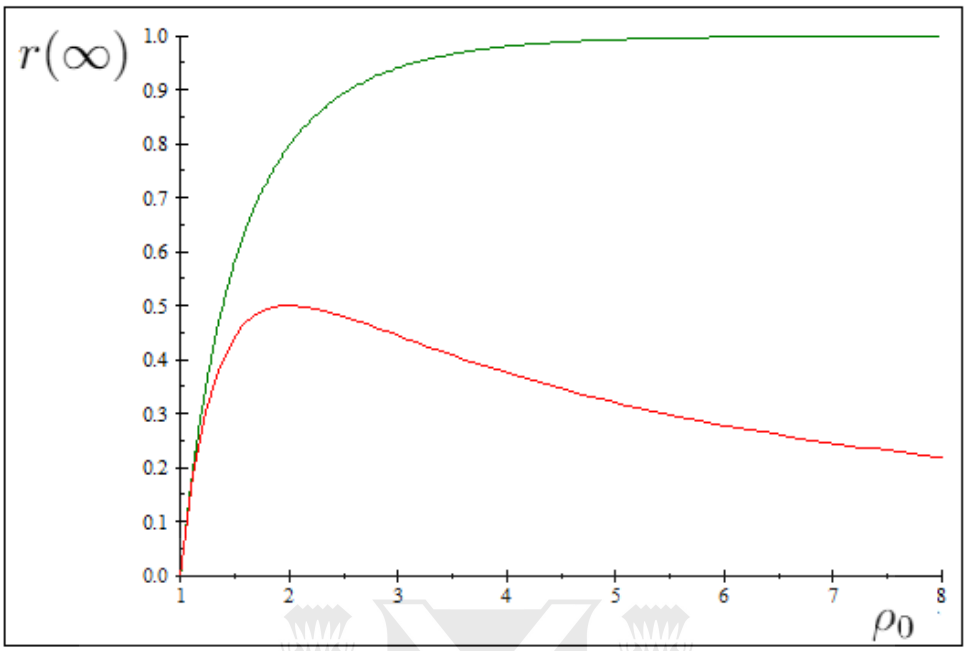


Figura 7.2: Solución exacta de $r(\infty)$ (verde) y a 2do orden (rojo)

Vemos nuevamente que la aproximación utilizada se aparta de la solución exacta en el rango (1,5;2).

7.4. Cálculo exacto de la fracción de fallecidos finales $f(\infty)$

Omitiendo el posible desfasaje φ en la curva de infectados, la cantidad de fallecidos en todo instante es proporcional a la cantidad de infectados acumulados,

$$F(t) = \mu \int_0^t I(\tau) d\tau. \tag{7.22}$$

Analogamente si integramos la tercera ecuación del modelo SIR

$$R(t) = \gamma \int_0^t I(\tau) d\tau. \tag{7.23}$$

De la ecuaciones anteriores se sigue que en todo instante la cantidadde fallecidos y recuperados deben ser proporcionales

$$F(t) = \frac{\mu}{\gamma} R(t) \tag{7.24}$$

y lo mismo ocurre para sus respectivas fracciones de población

$$f(t) = \frac{\mu}{\gamma} r(t). \tag{7.25}$$

Dicha proporcionalidad es importante porque la ecuación trascendente para la fracción de recuperados (7.19) que discutimos en la sección anterior nos permite escribir una ecuación trascendente para la fracción de fallecidos finales $f(\infty)$

$$\frac{\gamma}{\mu} f(\infty) = 1 - e^{-\frac{\gamma}{\mu}\rho_0 f(\infty)} \tag{7.26}$$

Y utilizando la ecuación (7.21) podemos obtener la expresión solución a segundo orden en ρ_0

$$\boxed{f(\infty) = \frac{\mu}{\gamma}\frac{2(\rho_0 - 1)}{\rho_0^2}} \tag{7.27}$$

Más abajo en la Figura 7.3 graficamos la aproximación contra la solución exacta de la ecuación trascendente para un cociente $\frac{\gamma}{\mu} = 10$, que esencialmente es un reescaleo en el eje horizontal del gráfico de la sección anterior.

Notese que γ no es un parámetro homogéneo con la edad. En realidad, siendo estrictos, deberíamos definir un $\gamma(\tau)$ y $\mu(\tau)$ para cada edad biológica τ. El γ del modelo SIRD son en realidad un promedio pesado por la distribución de frecuencias $f(\tau)$ de la pirámide poblacional $\gamma = \sum_\tau f(\tau)\gamma(\tau)$ (de la misma forma para μ). Esta la razón por la que no podemos confiar en la muestra de recuperados $R_m(t)$, la cual esta sesgada a las frecuencias de τ más elevado.

Este último resultado teórico resume conceptualmente el presente trabajo. La ecuación (7.27) nos dice que la fracción final de fallecidos es proporcional al cociente $\frac{\mu}{\gamma}$ de las tasas de mortalidad por un una función de la tasa de reproducción básica inicial ρ_0.

Es decir, el número final de fallecidos en una epidemia depende del cociente de dos parámetros biológicos que no podemos controlar y de la tasa de reproducción básica inicial $\rho_0 \simeq \frac{\beta N}{\gamma}$, que esencialemente depende de la tasa de transmisión β.

La razón por la cual no utilizamos esta expresión y tuvimos que recurrir al modelo cinemático es que al comienzo de la epidemia no es posible determinar ninguno de estos trés parámetros. De hecho, uno podría intentar una determinación cinemática de los mismos como

$$\mu = \frac{F(t)}{\int_0^t I(\tau)d\tau} \tag{7.28}$$

$$\gamma = \frac{R(t)}{\int_0^t I(\tau)d\tau} \tag{7.29}$$

pero como dijimos más arriba las poblaciónes de fallecidos y recuperados estan sesgadas por un muestreo sobre la población mas vulnerable y tampoco conocemos el número real de infectados de la población.

Si al menos conocieramos γ por medio de una estadística confiable, podriamos reescalear el tiempo ordinario como γt, y también determinar cinemáticamente ρ_0, ya que de (6.1) se sigue que al comienzo de la epidema la misma crece con una tasa

$$\rho_0 - 1 = \frac{1}{I}\frac{dI}{d(\gamma t)} \tag{7.30}$$

pero en esta pandemia esos datos no estuvieron a nuestro alcance.

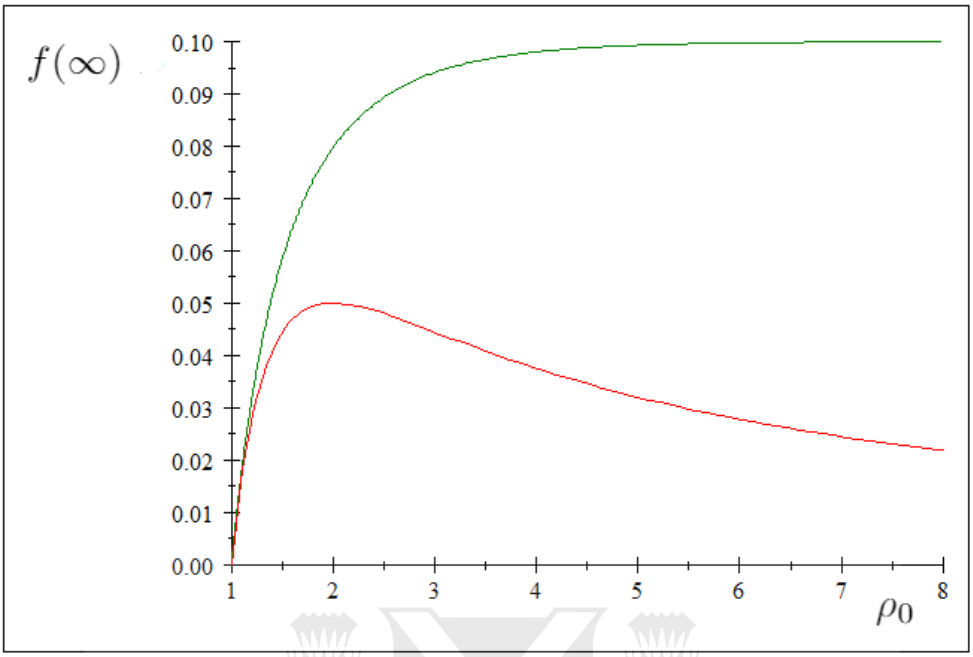


Figura 7.3: Solución exacta de $f(\infty)$ (verde) y a 2do orden (rojo)

Por último queremos destacar que tanto una expresión analítica cerrada como aproximada para $f(\infty)$, nos permitiría estimar la eficacia de las medida de restricción. En efecto las medidas de restricción disminuyen el ρ_0 natural y con ello la cantidad de fallecidos finales $f(\infty)$. En otras palabras, nos permitirían estimar cuales serian las muertes que hubiésemos tenido si no se hubiésen aplicado las mismas.

Universidad de SanAndrés

Capítulo 8

Resúmen y Conclusiones

El objetivo del presente trabajo fue estimar el número final de fallecidos de una epidemia en una población de N habitantes. Para ello, planteamos un modelo cinemático de trés parámetos I_0, t_0 y k que nos condujo a una función error

$$\boxed{F(t) = \mu I_0 \int_0^t \exp -\frac{k}{2}(\tau - t_0 + \varphi)^2 d\tau} \tag{8.1}$$

donde t_0 y I_0, son el tiempo en que la población de infectados alcanza el pico de la epidemia y la magnitud de dicha población en el pico, respectivamente. A nivel del modelo cinemático el parametro k es simplemente una medida del ancho de la Gaussiana que obtuvimos a partir de un ajuste de la población de infectados.

Sin necesidad de conocer a priori la tasa de mortalidad μ, el ajuste global de los parámetros (μI_0) y $(t_0 + \varphi)$ en el modelo cinemático nos permitió hacer un pronóstico del número final de fallecidos para varios países.

Cuando ajustamos los parámetos cinemáticos con los parámetros del modelo SIR de KM, de modo tal que el número final de fallecidos $F(\infty)$ resulte ser el mismo, vemos que la sigmoide que predice el modelo dinámico KM (en la aproximación de segundo orden en la tasa de reproducción básica inicial $\rho_0 = \frac{\beta N}{\gamma}$))

$$F(t) = \frac{\mu I_0}{\lambda} \frac{2}{1 + e^{-2\lambda(t - t_0)}} \tag{8.2}$$

predice los mismos resultados en todo el rango central del pico de la epidemia que la función error del modelo gaussiano.

El parámetro k del modelo Gaussiano queda entonces vinculado con los parámetros dinámicos del modelo SIR como

$$k = \frac{\pi \gamma^2}{8} \left[(\rho_0 - 1)\right]^2 \tag{8.3}$$

Finalmente la solución analítica exacta del modelo KM, utilizando la ecuación trascendente, nos permitió evaluar el rango de validez de ambas aproximaciones sigmoideas a segundo orden que predicen el número final de fallecidos.

$$\boxed{f(\infty) = \frac{\mu}{\gamma}\frac{2(\rho_0 - 1)}{\rho_0^2}.} \tag{8.4}$$

De lo anterior concluimos que:

• El modelo cinemático Gaussiano es facil de ajustar en comparación con el modelo de KM, esencialmente porque ajustar una Gaussiana no es otra cosa que ajustar la parábola del gráfico semilogarítmico de la curva de infectados.

• El modelo es robusto a nivel predictivo porque es un modelo que minimiza el número de parámetros. Esta ventaja, nos permite minimizar los errores que surgirían de una sobreparámetrización, ya que los datos empiricos con los que contamos son altamente "ruidosos".

• En consecuencia, el modelo permitió hacer predicciones robustas en el orden de magnitud para el número final de fallecidos del grupo de países bajo análisis.

• A pesar de que lo ajustamos con el modelo SIR de KM, las hipótesis principales del modelo Gausssiano en realidad contemplan extensiones más complejas del modelo SIR.

• Si conociéramos los parámetros biológicos μ, γ en forma independiente, la aproximación a segundo orden para $f(\infty)$ (o mejor aún la ecuación trascendente que derivamos para ella para todo orden) nos permitiría estimar la eficacia de las medidas de restricción así como las muertes adicionales esperadas en los posibles rebrotes post pandemia. En otras palabras, podriamos estimar cuales serían las muertes que hubiésemos tenido si no se hubiésen aplicado dichas medidas, así como las muertes esperadas post pandemia cuando estas se relajan y la epidemia evoluciona con a su tasa de reprodcción básica natural.

Este trabajo surgió a partir de una motivación económica: plantear un modelo en el cual se minimicen los riesgos de salud para la población conjuntamente con las consecuencias sobre la economía. En el presente trabajo, nos enfocamos en estimar el número final de fallecidos en función del periodo de tiempo de restricción (T) de la actividad económica. A medida que aumenta T, el número final de fallecidos decrece pero la caída del producto aumenta. En un trabajo posterior, buscaremos establecer una equivalencia entre la caída del producto (PBI) y los muertos asociados a este (aumentó de la pobreza estructural). Utilizando el modelo de crecimiento de Solow, encontraremos un tiempo óptimo para las medidas de restricción.

La otra linea de investigación que surge a partir de este trabajo es continuar con el análisis de las estimaciones de los parámetros biológicos (μ y γ), los cuales permitirían predecir con mayor robustez la dinámica final de la epidemia, y así lograr un modelo económico aún más sólido.

Universidad de San Andrés

Capítulo 9

Apéndice A: Aproximación de segundo orden (el modelo sech^2 de KM)

Los infectados y recuparados no crecen exponencialmente en todo rango de tiempo. Si en la expansión de (7.3) retenemos hasta el segundo orden en la exponencial

$$e^{-\rho_0 \frac{R}{S(0)}} = 1 - \frac{\rho_0}{S(0)} R + \frac{\rho_0^2}{2S^2(0)} R^2$$

tenemos

$$\frac{dR}{d(\gamma t)} = N - R - S(0)(1 - \frac{\rho_0}{S(0)} R + \frac{\rho_0^2}{2S^2(0)} R^2)$$

y el término cuadrático en R funciona como un limitante al régimen exponencial, en forma análoga a lo que sucede en la ecuación logística de Verhulst. Dicha ecuación tiene como condición inicial $R(0) = 0$ ya que la población inicial de recuperados es nula al comienzo de la epidemia. La misma puede reescribirse como

$$\frac{dR}{d(\gamma t)} = (N - S(0)) + (\rho_0 - 1)R - \frac{\rho_0^2}{2S(0)} R^2 \tag{9.1}$$

Que haciendo las sustituciones

$$\begin{aligned} g &\rightarrow R \\ t &\rightarrow \gamma t \end{aligned}$$

es la ecuación logística generalizada que discutimos en el Apéndice B

$\frac{dR}{d(\gamma t)} = a + bR - cR^2$		$LG - a$
$R(\gamma t) = \frac{1}{2c}\left[b + \alpha \tanh(\frac{\alpha\gamma(t-t_0)}{2})\right]$		$LG - b$
$\alpha = \sqrt{b^2 + 4ac}$		$LG - c$
$\frac{\alpha\gamma t_0}{2} = \tanh^{-1}(\frac{b}{\alpha})$,	$R(0) = 0$	$LG - d$

con

$$\begin{aligned} a &= N - S(0) \\ b &= \rho_0 - 1 \\ c &= \frac{\rho_0^2}{2S(0)} \end{aligned}$$

Reemplazando a, b y c finalmente resulta

$$\alpha = \sqrt{(\rho_0 - 1)^2 + \frac{2(N - S(0))\rho_0^2}{S(0)}} \tag{9.2}$$

$$\frac{\alpha\gamma t_0}{2} = \tanh^{-1}(\frac{\rho_0 - 1}{\alpha}) \tag{9.3}$$

$$R(\gamma t) = \frac{S(0)}{\rho_0^2}\left[N - S(0) + \alpha \tanh(\frac{\alpha\gamma(t - t_0)}{2})\right] \tag{9.4}$$

Teniendo en cuenta que

$$\frac{d \tanh x}{dx} = 1 - \tanh^2 x = \frac{1}{\cosh^2 x}$$

La ecuación de infectados resulta

$$I(t) = \frac{dR(t)}{\gamma dt} = \frac{I_0}{\cosh^2(\frac{\alpha\gamma}{2}(t - t_0))} \tag{9.5}$$

donde

$$I_0 = I(t_0) = \frac{S(0)\alpha^2}{2r_0^2} \tag{9.6}$$

es la cantidad de infectados en el pico.

Notese que como el número inicial de infectados $I(0) = N - S(0)$ es despreciable frente a la población total, $N \simeq S(0)$ el parámetro α se puede aproximar como

$$\alpha \simeq \rho_0 - 1.$$

Universidad de SanAndrés

Capítulo 10

Apéndice B: Resolución de la ecuación logística generalizada

La estrategia para integrar la ecuación logística generalizada del tipo

$$\boxed{\frac{df}{dt} = a + bf - cf^2 \qquad LG - a}$$

sera factorizar el polinomio $P(f)$ del segundo miembro dicha ecuación

$$P(f) = a + bf - cf^2$$

de modo de llevarlo a la ecuación de Verhulst. Factorizando

$$P(f) = -c(f - r_1)(f - r_2)$$

donde las raices r_1 y $r_2$de dicho polinomio son

$$\begin{aligned} r_1 &= \frac{b - \alpha}{2c} \\ r_2 &= \frac{b + \alpha}{2c} \end{aligned}$$

$$\boxed{\alpha = \sqrt{b^2 + 4ac} \quad LG - c}$$

$$r_2 - r_1 = \alpha/c$$

Vemos que en ese caso la ecuación logística generalizada (LG-a) puede reescribirse como

$$\frac{df}{dt} = -c(f - r_1)(f - r_2)$$

Si ahora definó la variable

$$g = f - r_1$$

$$g - \alpha/c = g + r_1 - r_2 = f - r_2$$

La variable g satisface la ecuación Verhulst

$$\frac{dg}{dt} = -cg(g - \frac{\alpha}{c}) \ = \alpha g - cg^2 = \alpha g(1 - \frac{g}{\alpha/c})$$

Cuya solución es de la forma

$$g(t) = \frac{\alpha/c}{1 + e^{-\alpha(t-t_0)}}$$

donde t_0 es una constante de integración que determina la condición inicial $g(0)$.

Utilizando la identidad

$$\frac{1}{1+e^{-x}} = \frac{1}{2} + \frac{1}{2}\tanh(\frac{x}{2})$$

podemos reescribir dicha solución como

$$g(t) = \frac{1}{2c}\left[\alpha + \alpha\tanh(\frac{\alpha(t-t_0)}{2})\right]$$

y volviendo a la variable f

$$f(t) = g(t) + r_1 = g(t) + \frac{b-\alpha}{2c}$$

$$f(t) = \frac{b-\alpha}{2c} + \frac{\alpha}{2c} + \frac{\alpha}{2c}\tanh(\frac{\alpha(t-t_0)}{2})$$

Por lo que finalmente podemos escribir la solución de la ecuación $(LG-a)$ como

$$\boxed{f(t) = \tfrac{1}{2c}\left[b + \alpha\tanh(\tfrac{\alpha(t-t_0)}{2})\right] \qquad LG-b}$$

Para $t = 0$

$$f(0) = \frac{1}{2c}\left[b - \alpha\tanh\frac{\alpha t_0}{2}\right]$$

es decir

$$\tanh\frac{\alpha t_0}{2} = \frac{b - 2cf(0)}{\alpha}$$

por lo que

$$\frac{\alpha t_0}{2} = \tanh^{-1}(\frac{b-2cf(0)}{\alpha}),$$

y para el caso particular de la condición inicial $f(0) = 0$

$$\boxed{\tfrac{\alpha t_0}{2} = \tanh^{-1}(\tfrac{b}{\alpha}) \qquad LG-c}$$

Universidad de San Andrés

Capítulo 11

Apéndice C: Estimación de los parámetros gaussianos ($I(0)$, t_0 y k) para cada uno de los países

España

Tomamos los datos del día 17 de marzo al 01 de abril, y vemos que dichos datos ajustan con la parábola

$$\ln I_m(t) = -5.322 \times 10^{-3} x^2 + 0{,}356\,16x + 5.829\,4$$

como muestra el siguiente gráfico, Igualando los coeficientes de ambos polinomios, tenemos

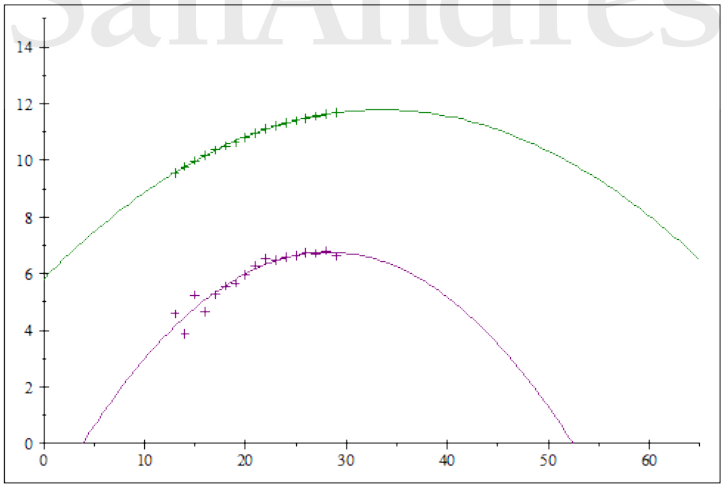


Figura 11.1: España - $\ln I_m(t)$ (verde) y $\ln \Delta F_m$ (violeta)

$$\begin{aligned} \frac{1}{2}k &= -5.322 \times 10^{-3} \\ kt_0 &= 0{,}356\,16 \\ \ln I_{m0} &= \frac{1}{2}kt_0^2 + 5.829\,4 \end{aligned}$$

de lo que se sigue que los parámetos de la Gaussiana de la muestra de infectados activos son

$$\begin{aligned} k &= 2 \times 5.322 \times 10^{-3} = 0{,}010644 \\ t_0 &= \frac{0{,}356\,16}{0{,}010644} = 33{,}46 \\ \ln I_{m0} &= \frac{1}{2}\,(0{,}010644)\,(33{,}46)^2 + 5{,}8294 = 11.788 \\ I_{m0} &= \exp(11{,}788) = 1.3166 \times 10^5 \end{aligned}$$

Por lo que la curva finalmente resulta,

$$I_m(t) = 1.3166 \times 10^5 \exp\left[-\frac{0{,}010}{2}(t - 33{,}5)^2\right]$$

La curva de infectados reales sería en principio

$$I(t) = c_I \times (1.3166 \times 10^5) \exp\left[-\frac{0{,}010}{2}(t - 33{,}5)^2\right]$$

Para estimarla vamos a tomar como base la expresión anterior de la Gaussiana,

$$F(t) = \rho \int_0^t (1.3166 \times 10^5) \exp\left[-\frac{0{,}010}{2}(x - 33{,}5 + \varphi)^2\right] dx$$

Alemania

Tomamos los datos del día 22 de marzo al 7 de abril, y vemos que dichos datos ajustan con la parábola

$$\ln I_m(t) = -3.368 \times 10^{-3}x^2 + 0{,}237\,41x + 7.457\,3$$

como muestra el siguiente gráfico, Igualando los coeficientes de ambos polinomios, tenemos

QUAERERE VERUM Universidad de San Andrés

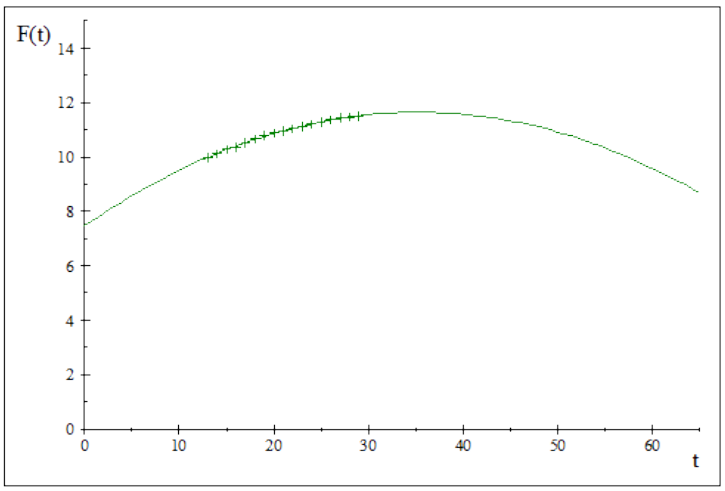


Figura 11.2: Alemania - $\ln I_m(t)$ (verde)

$$\begin{aligned} \frac{1}{2}k &= 3.368 \times 10^{-3} \\ kt_0 &= 0{,}237\,41 \\ \ln I_{m0} &= \frac{1}{2}kt_0^2 + 7.457\,3 \end{aligned}$$

de lo que se sigue que los parámetos de la Gaussiana de la muestra de infectados activos son

$$\begin{aligned} k &= 2 \times 3.368 \times 10^{-3} = 0{,}006736 \\ t_0 &= \frac{0{,}237\,41}{0{,}006736} = 35{,}245 \\ \ln I_{m0} &= \frac{1}{2}\,(0{,}006736)\,(35{,}245)^2 + 7.457\,3 = 11.641 \\ I_{m0} &= \exp(11{,}641) = 1.136\,6 \times 10^5 \end{aligned}$$

Por lo que la curva finalmente resulta,

$$I_m(t) = 1.136\ \times 10^5 \exp\left[-\frac{0{,}0067}{2}(t-35)^2\right]$$

La curva de infectados reales sería en principio

$$I(t) = c_I \times (1.136\ \times 10^5) \exp\left[-\frac{0{,}0067}{2}(t-35)^2\right]$$

Para estimarla vamos a tomar como base la expresión anterior de la Gaussiana,

$$F(t) = \rho \int_0^t (1.136 \times 10^5) \exp\left[-\frac{0{,}0067}{2}(x-35+\varphi)^2\right] dx$$

Reino Unido

Tomamos los datos del día 21 de marzo al 6 de abril, y vemos que dichos datos ajustan con la parábola

$$\ln I_m(t) = -2.765 \times 10^{-3} x^2 + 0{,}272\,87x + 5.191\,6$$

como muestra el siguiente gráfico, Igualando los coeficientes de ambos polinomios, tenemos

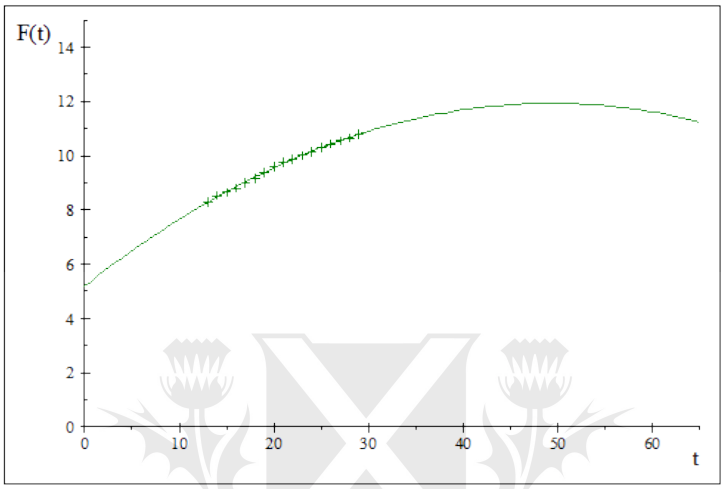


Figura 11.3: Reino Unido - $\ln I_m(t)$ (verde)

Universidad de San Andrés

$$\begin{aligned} \frac{1}{2}k &= 2.765 \times 10^{-3} \\ kt_0 &= 0{,}272\,87 \\ \ln I_{m0} &= \frac{1}{2}kt_0^2 + 7.457\,3 \end{aligned}$$

de lo que se sigue que los parámetos de la Gaussiana de la muestra de infectados activos son

$$\begin{aligned} k &= 2 \times 2.765 \times 10^{-3} = 0{,}005\,53 \\ t_0 &= \frac{0{,}272\,87}{0{,}005\,53} = 49{,}34 \\ \ln I_{m0} &= \frac{1}{2}\,(0{,}005\,53)\,(49{,}34)^2 + 5.191\,6 = 11.923 \\ I_{m0} &= \exp(11{,}923) = 1.506\,9 \times 10^5 \end{aligned}$$

Por lo que la curva finalmente resulta,

$$I_m(t) = 1.506\,9 \times 10^5 \exp\left[-\frac{0{,}0055}{2}(t-49)^2\right]$$

La curva de infectados reales sería en principio

$$I(t) = c_I \times (1.506\,9 \times 10^5) \exp\left[-\frac{0{,}0055}{2}(t-49)^2\right]$$

Para estimarla vamos a tomar como base la expresión anterior de la Gaussiana,

$$F(t) = \rho \int_0^t (1.506\,9 \times 10^5) \exp\left[-\frac{0{,}0055}{2}(x-49+\varphi)^2\right] dx$$

Bélgica

Tomamos los datos del día 24 de marzo al 9 de abril, y vemos que dichos datos ajustan con la parábola

$$\ln I_m(t) = -3.141\,3 \times 10^{-3} x^2 + 0{,}217\,61 x + 6.555$$

como muestra el siguiente gráfico,

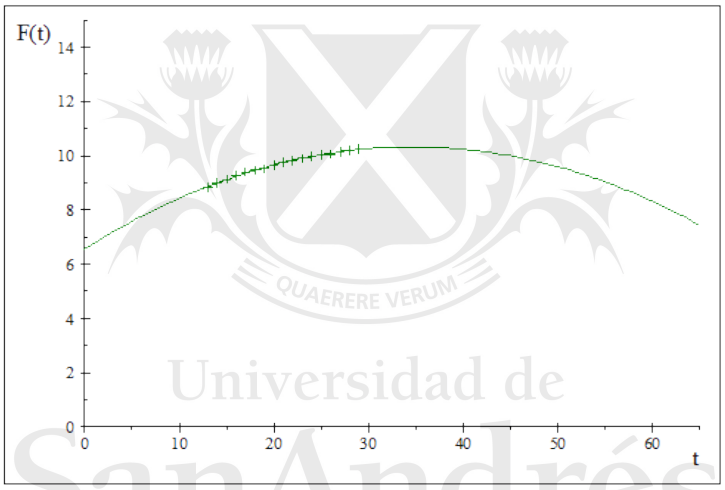


Figura 11.4: Bélgica - $\ln I_m(t)$ (verde)

de lo que se sigue que los parámetos de la Gaussiana de la muestra de infectados activos son

$$\begin{aligned} k &= 2 \times 3.141\,3 \times 10^{-3} = 0{,}00628 \\ t_0 &= \frac{0{,}217\,61}{0{,}00628} = 34{,}65 \\ \ln I_{m0} &= \frac{1}{2}(0{,}00628)(34{,}65)^2 + 6{,}555 = 10{,}325 \\ I_{m0} &= \exp(10{,}325) = 30{,}485 \end{aligned}$$

Por lo que la curva finalmente resulta,

$$I_m(t) = 30{,}485 \exp\left[-\frac{0{,}0063}{2}(t-35)^2\right]$$

La curva de infectados reales sería en principio

$$I(t) = c_I \times 30{,}485 \exp\left[-\frac{0{,}0063}{2}(t-35)^2\right]$$

Para estimarla vamos a tomar como base la expresión anterior de la Gaussiana,

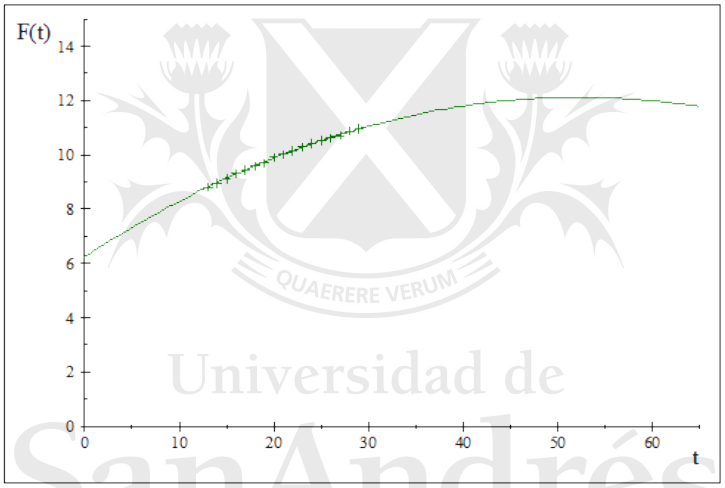
$$F(t) = \rho \int_0^t (30{,}485) \exp\left[-\frac{0{,}0063}{2}(x-35+\varphi)^2\right] dx$$

Francia

Tomamos los datos del día 17 de marzo al 2 de abril, y vemos que dichos datos ajustan con la parábola

$$\ln I_m(t) = -2.\,147\,8 \times 10^{-3} x^2 + 0{,}224\,60x + 6.\,238\,7$$

como muestra el siguiente gráfico,



QUAERERE VERUM
Universidad de
SanAndrés

Figura 11.5: Francia - $\ln I_m(t)$ (verde)

de lo que se sigue que los parámetos de la Gaussiana de la muestra de infectados activos son

$$\begin{aligned} k &= 2 \times 2.\,147\,8 \times 10^{-3} = 0{,}00429 \\ t_0 &= \frac{0{,}22460}{0{,}00429} = 52{,}354 \\ \ln I_{m0} &= \frac{1}{2}(0{,}00429)(52{,}354)^2 + 6{,}238 = 12{,}117 \\ I_{m0} &= \exp(12{,}117) = 182960 \end{aligned}$$

Por lo que la curva finalmente resulta,

$$I_m(t) = 182960 \exp\left[-\frac{0{,}0043}{2}(t-52)^2\right]$$

La curva de infectados reales sería en principio

$$I(t) = c_I \times 182960 \exp\left[-\frac{0{,}0043}{2}(t-52)^2\right]$$

Para estimarla vamos a tomar como base la expresión anterior de la Gaussiana,

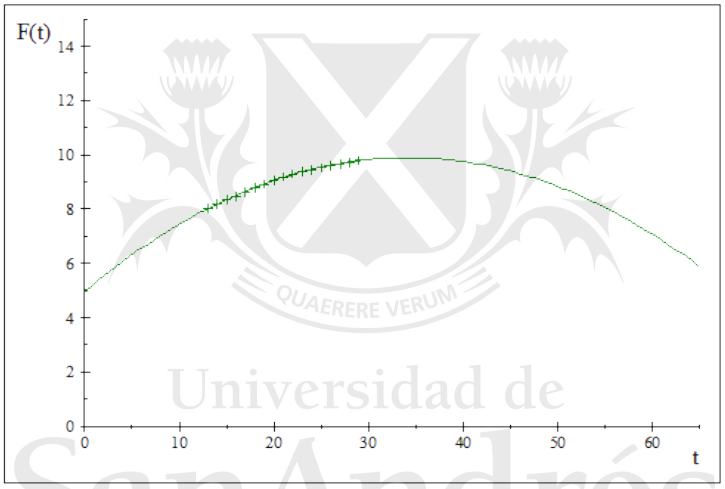
$$F(t) = \rho \int_0^t (182{,}960) \exp\left[-\frac{0{,}0043}{2}(x-52+\varphi)^2\right] dx$$

Holanda

Tomamos los datos del día 21 de marzo al 6 de abril, y vemos que dichos datos ajustan con la parábola

$$\ln I_m(t) = -4.\,221\,9 \times 10^{-3} x^2 + 0{,}288\,24x + 4.\,966\,2$$

como muestra el siguiente gráfico,



QUAERERE VERUM
Universidad de
SanAndrés

Figura 11.6: Holanda - $\ln I_m(t)$ (verde)

de lo que se sigue que los parámetos de la Gaussiana de la muestra de infectados activos son

$$\begin{aligned} k &= 2 \times 4.\,221\,9 \times 10^{-3} = 0{,}0084438 \\ t_0 &= \frac{0{,}288\,24}{0{,}0084} = 34{,}31 \\ \ln I_{m0} &= \frac{1}{2}(0{,}00844)(34{,}31)^2 + 4.\,9662 = 9{,}9339 \\ I_{m0} &= \exp(9{,}9339) = 20618 \end{aligned}$$

Por lo que la curva finalmente resulta,

$$I_m(t) = (20618) \exp\left[-\frac{0{,}0084}{2}(t-34)^2\right]$$

La curva de infectados reales sería en principio

$$I(t) = c_I \times (20618) \exp\left[-\frac{0{,}0084}{2}(t-34)^2\right]$$

Para estimarla vamos a tomar como base la expresión anterior de la Gaussiana,

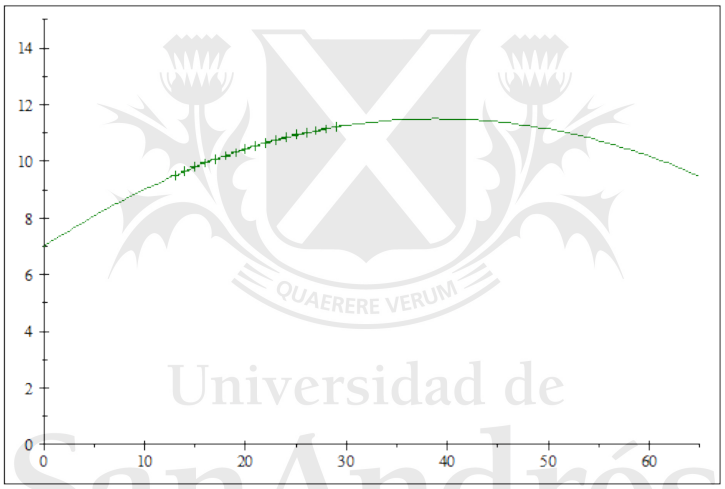
$$F(t) = \rho \int_0^t (20618) \exp\left[-\frac{0{,}0084}{2}(x-34+\varphi)^2\right] dx$$

Turquía

Tomamos los datos del día 1 al 15 de abril, y vemos que dichos datos ajustan con la parábola

$$\ln I_m(t) = -2.\,986\,8 \times 10^{-3} x^2 + 0{,}232\,27x + 6.\,994$$

como muestra el siguiente gráfico, de lo que se sigue que los parámetos de la Gaussiana de la muestra de infectados



QUAERERE VERUM
Universidad de
SanAndrés

Figura 11.7: Turquía - $\ln I_m(t)$ (verde)

activos son

$$\begin{aligned} k &= 2 \times 2.\,986\,8 \times 10^{-3} = 0{,}0059736 \\ t_0 &= \frac{0{,}23227}{0{,}00597} = 38{,}91 \\ \ln I_{m0} &= \frac{1}{2}\,(0{,}00597)\,(38{,}91)^2 + 6{,}994 = 11{,}513 \\ I_{m0} &= \exp(11{,}513) = 100001 \end{aligned}$$

Por lo que la curva finalmente resulta,

$$I_m(t) = (100001) \exp\left[-\frac{0{,}0059}{2}(t-39)^2\right]$$

La curva de infectados reales sería en principio

$$I(t) = c_I \times (100001) \exp\left[-\frac{0{,}0059}{2}(t-39)^2\right]$$

Para estimarla vamos a tomar como base la expresión anterior de la Gaussiana,

$$F(t) = \rho \int_0^t (100001) \exp\left[-\frac{0{,}0059}{2}(x-39+\varphi)^2\right] dx$$

Caso EEUU

Ciudad de Nueva York

Tomamos los datos del día 26 de marzo al 11 de abril, y vemos que dichos datos ajustan con la parábola

$$\ln I_m(t) = -3.\,141\,4 \times 10^{-3} x^2 + 0{,}226\,52x + 8.\,188\,7$$

como muestra el siguiente gráfico,

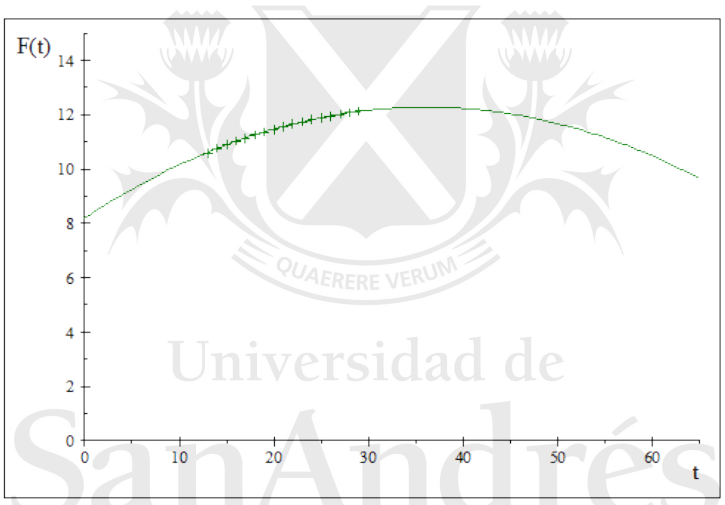


Figura 11.8: Nueva York - $\ln I_m(t)$ (verde)

de lo que se sigue que los parámetos de la Gaussiana de la muestra de infectados activos son

$$\begin{aligned} k &= 2 \times 3.\,141\,4 \times 10^{-3} = 0{,}00628 \\ t_0 &= \frac{0{,}226\,52}{0{,}00628} = 36{,}07 \\ \ln I_{m0} &= \frac{1}{2}(0{,}00628)(26{,}07)^2 + 8.\,188 = 10{,}322 \\ I_{m0} &= \exp(10{,}322) = 30394 \end{aligned}$$

Por lo que la curva finalmente resulta,

$$I_m(t) = (30394)\exp\left[-\frac{0{,}0063}{2}(t-36)^2\right]$$

Para estimarla vamos a tomar como base la expresión anterior de la Gaussiana,

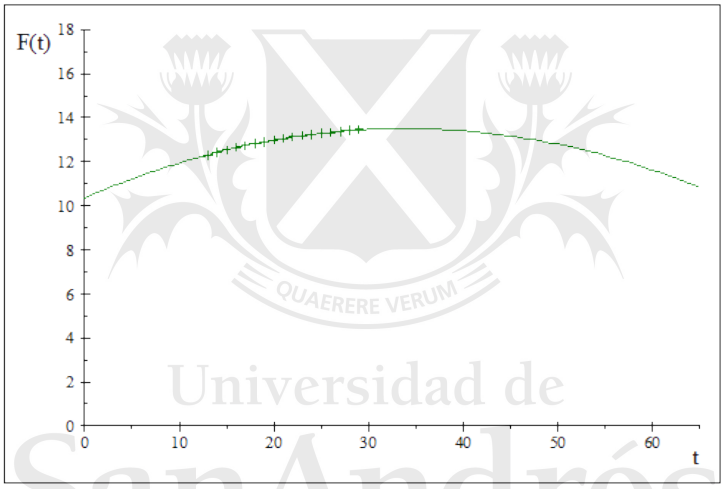
$$F(t) = \rho\int_0^t (30394)\exp\left[-\frac{0{,}0063}{2}(x-36+\varphi)^2\right]dx$$

EEUU

Tomamos los datos del día 2 al 18 de abril, y vemos que dichos datos ajustan con la parábola

$$\ln I_m(t) = -2.764\,2\times 10^{-3}x^2 + 0{,}187\,01x + 10.343$$

como muestra el siguiente gráfico,



QUAERERE VERUM
Universidad de
SanAndrés

Figura 11.9: EEUU - $\ln I_m(t)$ (verde)

de lo que se sigue que los parámetos de la Gaussiana de la muestra de infectados activos son

$$\begin{aligned} k &= 2\times 2.764\,2\times 10^{-3} = 0{,}005528 \\ t_0 &= \frac{0{,}1870}{0{,}00597} = 31{,}32 \\ \ln I_{m0} &= \frac{1}{2}(0{,}00552)(31{,}21)^2 + 10{,}34 = 13{,}03 \\ I_{m0} &= \exp(13{,}03) = 455890 \end{aligned}$$

Por lo que la curva finalmente resulta,

$$I_m(t) = (455890)\exp\left[-\frac{0{,}0055}{2}(t-31)^2\right]$$

Para estimarla vamos a tomar como base la expresión anterior de la Gaussiana,

$$F(t) = \rho \int_0^t (455890) \exp\left[-\frac{0{,}0055}{2}(x - 31 + \varphi)^2\right] dx$$

QUAERERE VERUM
Universidad de
SanAndrés

Bibliografía

ACEMOGLU, Daron, Chernozhukov, Victor,Werning, Iván y Whinston, Michael D.. (2020). A Multi-Risk SIR Model with Optimal Targeted Lockdown. *National Bureau of Economic Research.* Working Paper 27102.

ATKESON, Andrew. (2020). How Deadly Is COVID-19? Understanding The Difficulties With Estimation Of Its Fatality. *National Bureau of Economics and Research.* Working Paper 26965.

BARRO, Robert J., Ursúa, José F., y Weng, Joanna. (2020). The Coronavirus and the Great Influenza Pandemic: Lessons from the "Spanish Flu" for the Coronavirus's Potential Effects on Mortality and Economic Activity. *National Bureau of Economics and Research.* Working Paper 26866.

BENDAVID, E., B. Mulaney, N. Sood, S. Shah, E. Ling, R. Bromley-Dulfano, C. Lai, Z. Weissberg, R. Saavedra, J. Tedrow, et al. (2020). Covid-19 Antibody Seroprevalence in Santa Clara County, California. *medRxiv.* https://doi.org/10.1101/2020.04.14.20062463.

BRAUER, Fred, Pauline van den Driessche, and J. Wu. (2008). Mathematical Epidemiology, Springer-Verlag, USA: New York.

CARUANA, R., Searle R., Heller T., y Shupack S. (1986). Fast algorithm for the resolution of spectra. *Anal. Chem.* 58 (6), 1162–1167.

DIRAC, Paul A. M. (1926). On the Theory of Quantum Mechanics. *Proceeding of The Royal Society A.* 112 (776), 661-677.

FERMI, Enrico. (1926). Sulla quantizzazione del gas perfetto monoatomico. *Rendiconti Lincei (in Italian).* 3, 145–9., translated as Zannoni, Alberto (1999-12-14). On the Quantization of the Monoatomic Ideal Gas. arXiv:cond-mat/9912229.

GALILEO, Galilei. (2011). Diálogo sobre los dos máximos sistemas del mundo ptolemaico y copernicano. (Antonio Beltrán, trad.) Madrid: Alianza Editorial. (Obra original públicada en 1632).

GLAISHER, James Whitbread Lee. . (1871).On a class of definite integrals.*London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science.* 4 (Taylor and Francis). 42 (277), 294-302.

Universidad de San Andrés

HORTAÇSU, Ali, Liu, Jiarui y Schwieg,Timothy. (2020). Estimating the Fraction of Unreported Infections in Epidemics with a Known Epicenter: an Application to COVID-19. *National Bureau of Economics and Research.* Working Paper No. 27028

KERMACK, William Ogilvy, A. G. McKendrick, y Gilbert Thomas Walker. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character.*115 (772), 700–721.

MANSKI, Charles F. y Molinari, Francesca. (2020). Estimating the COVID-19 Infection Rate: Anatomy of an Inference Problem. *National Bureau of Economics and Research.* Working Paper No. 27023.

MURRAY, J. D. (2002). Mathematical Biology: I. An Introduction. Springer-Verlag, USA:New York.

NEWTON, Issac. (1993). Principios matemáticos de la Filosofía natural [Philosophiae Naturalis Principia Mathematica]. (Antonio Escohotado, trad.) Barcelona: Ediciones Altaya, S.A. (Obra original públicada en 1687).

OSEMWINYEN, Amenaghawon, Diakhaby, Aboubakary. (2015). Mathematical Modelling of the Transmission Dynamics of Ebola Virus. *Applied and Computational Mathematics.* 4, 313–320.

RUSSELL, Timothy W. , Hellewell, Joel , Jarvis, Christopher, van Zandvoort, Kevin , Abbott, Sam , Ratnayake, Ruwan , CMMID COVID-19 working group , Flasche, Stefan , Eggo, Rosalind M. , Edmunds, W. John y Kucharski, Adam J . (2020). Estimating the infection and case fatality ratio for coronavirus disease (COVID-19) using age-adjusted data from the outbreak on the Diamond Princess cruise ship, February 2020. *Euro Surveill.* 25(12). https://doi.org/10.2807/1560-7917.

STOCK, James H., Aspelund, Karl M., Droste, Michael y Walker, Christopher D. (2020). Identification and Estimation of Undetected COVID-19 Cases Using Testing Data from Iceland. *medRxiv.* https://doi.org/10.1101/2020.04.06.20055582.

VERHULST, Pierre-François. (1845). Recherches mathématiques sur la loi d'accroissement de la population. Mathematical Researches into the Law of Population Growth Increase. *Nouveaux Mémoires de l'Académie Royale des Sciences et Belles-Lettres de Bruxelles.* 18, 8.

Universidad de San Andrés