



Universidad de Valladolid

Facultad de Ciencias

TRABAJO FIN DE GRADO

Grado en Matemáticas

MODELOS ESTOCÁSTICOS EN EPIDEMIOLOGÍA

Autor: Raúl Arnanz Manso

Tutor/es: Eustasio Del Barrio Tellado

*Agradecer a Eustasio del Barrio
el tiempo y la dedicación empleados
para la elaboración de este trabajo.
De la misma manera, agradecer a mi familia
y amigos el apoyo moral
durante toda mi trayectoria universitaria.*

Índice general

1. Resultados previos	3
1.1. Procesos estocásticos	4
1.2. Procesos de puntos y procesos de conteo	7
1.3. Procesos de Poisson	9
1.4. Función de supervivencia y función de riesgo	11
1.5. Intensidades	12
2. Modelos deterministas	17
3. Modelos estocásticos	25
3.1. Proceso de ramificación en tiempo continuo	25
3.2. El modelo Reed-Frost	27
3.3. El modelo estocástico en epidemias general (SIR)	29
3.4. Proceso de Hawkes	33
3.5. Proceso de Hawkes modificado	34
4. Ilustración numérica de los resultados teóricos	37
4.1. Curva epidémica	38
4.2. Distribución del intervalo de serie	39
4.3. Estimación del número básico de reproducción $R(t)$	40
4.4. Infectividad global $\lambda^*(t)$	42
Bibliografía	45

Introducción

A lo largo de la historia de la humanidad, las epidemias han sido uno de los grandes obstáculos, en el ámbito de la supervivencia, que han tenido que superar los seres humanos; enfermedades como la peste negra, el cólera o el dengue han producido un elevado número de muertes en la población. El estudio matemático de estas epidemias han dado lugar a los modelos epidemiológicos. Dentro de este tipo de modelos existen dos enfoques diferenciados, el determinista y el estocástico. El enfoque determinista se basa en ecuaciones diferenciales para realizar la modelización de la situación epidemiológica, en cambio, el enfoque estocástico tiene como principal herramienta los procesos estocásticos.

El objetivo principal de este trabajo es estudiar modelos matemáticos que muestren como se comporta la población infectada a lo largo de una situación con epidemia.

En la primera parte del trabajo, realizaremos un estudio teórico con los principales conceptos que utilizaremos en el desarrollo de modelos estocásticos. Definiremos conceptos como filtración y martingala, también describiremos en que consisten los procesos de puntos y los procesos de conteo, y como se relacionan entre ellos. En el ámbito de la supervivencia, desarrollaremos la función de supervivencia y de riesgo, que nos ayudarán a definir de manera precisa la función de intensidad condicionada.

Además, desarrollaremos un modelo determinista, el modelo SIR. Tras la descripción cualitativa del modelo expondremos algunos inconvenientes que nos llevarán a rechazar dicho modelo. Hemos realizado un estudio cualitativo debido a que el sistema que describe este modelo es no lineal.

Una vez hayamos demostrado que estos modelos no son válidos para modelar la realidad de la situación que queremos estudiar, pasaremos a los modelos estocásticos. Estos habrán resuelto varios de los inconvenientes de los modelos deterministas, y se asemejan más a la situación que nos planteamos. Los modelos de los que hablaremos son los modelos de ramificación y de Reed-Frost que se centrarán en cómo se desarrolla cada generación de individuos infectados. Uno de

los modelos estocásticos más importantes es el modelo SIR estocástico, del que conoceremos su funcionamiento. Por último, expondremos el modelo de Hawkes y una modificación interesante de dicho modelo. Gracias a esta modificación podremos realizar un cálculo del número de reproducción instantáneo, que es el número medio de individuos secundarios infectados por un individuo infeccioso primario que varía a lo largo del tiempo.

Como conclusión, expondremos una serie de gráficas que nos ayudarán a entender el comportamiento de las epidemias en una población susceptible. Todas estas gráficas son resultado de la recolección de datos epidemiológicos en la comunidad autónoma de Castilla y León.

Capítulo 1

Resultados previos

En este capítulo presentamos algunos resultados previos necesarios para describir los modelos estocásticos en epidemiología que se presentan en el Capítulo 3. El objeto fundamental, en el que se basan dichos modelos, son los llamados procesos de conteo o, en una formulación alternativa equivalente, los procesos de puntos. Los procesos de conteo son un tipo de proceso estocástico adecuado para modelar la ocurrencia reiterada a lo largo del tiempo de un suceso de interés (la aparición de nuevos casos de una enfermedad, el ingreso de paciente en hospitales, etc). Por esta razón, comenzaremos el capítulo describiendo la terminología y conceptos fundamentales de los procesos estocásticos. En esta descripción intervendrán los elementos habituales de la Teoría de la Probabilidad, tal como se describe por ejemplo en [1].

Después nos dedicamos a presentar propiamente el concepto de proceso de conteo. Discutiremos distintas formas en las que se pueden especificar un modelo de proceso de conteo. Es especialmente interesante, la descripción en términos del proceso de intensidad asociado. La forma integral de esta intensidad, es decir, la intensidad acumulada presenta un papel importante en la teoría de procesos de conteo. La diferencia entre el proceso de conteo y la intensidad acumulada es una martingala, es decir, ruido estocástico de forma que la información esencial en el modelo de proceso de conteo está contenida en el proceso de intensidad.

Para completar el capítulo presentaremos algunos de los ejemplos más habituales de procesos de conteo, empezando por el más importante de todos: el proceso de Poisson. A partir de este proceso, se pueden construir distintas variaciones (procesos no homogéneos, procesos de renovación, procesos de Cox, etc) que servirán como base a todos los modelos descritos en el capítulo 3.

Únicamente presentaremos algunos resultados omitiendo las demostraciones, pero aportando referencias donde las demostraciones pueden ser consultadas. Las

fuentes principales para este capítulo son [2] y [8].

1.1. Procesos estocásticos

Un proceso estocástico es un modelo matemático para describir un fenómeno aleatorio que evoluciona en el tiempo. Formalmente, un proceso es una colección de variables aleatorias indexadas por el tiempo $X = \{X(t) : t \in \mathbb{T} \subset [0, \infty)\}$ en un espacio probabilístico (Ω, \mathcal{F}, P) .

Cada $X(t)$ es, por lo tanto, una aplicación de $\Omega \rightarrow \mathbb{R}$ con la propiedad de ser $(\mathcal{F}, \mathcal{B})$ -medible, donde \mathcal{B} es la σ -álgebra de Borel.

En realidad, el espacio de llegada de las variables aleatorias $X(t)$ puede ser un subconjunto propio $E \subset \mathbb{R}$. A este conjunto E nos referimos como espacio de estados, ya que dentro de E se encuentran todos los posibles valores que puede tomar $X(t)$.

Cada variable aleatoria del proceso estocástico es una aplicación $X(t) : \Omega \rightarrow \mathbb{R}$ definida por $\omega \mapsto X(t)(\omega)$. Debido a que la notación $X(t)(\omega)$ no es utilizada con frecuencia, usaremos $X(t, \omega)$ o $X_t(\omega)$ en su lugar.

Ahora podemos interpretar un proceso como una función de dos variables $X : \mathbb{T} \times \Omega \rightarrow \mathbb{R}$ definida por $(t, \omega) \mapsto X(t, \omega)$. Luego para cada ω fijo, la aplicación de $\mathbb{T} \rightarrow \mathbb{R}$ que a cada t envía en $X_t(\omega)$ es una trayectoria. Cuando nos referimos a la propiedad de trayectorias del proceso, decimos que tiene trayectorias continuas cuando son continuas para casi todo ω . Se habla de un proceso con trayectorias cadlag si todas las trayectorias son continuas por la derecha con límite por la izquierda, etc.

Por tanto, un proceso estocástico puede interpretarse como una aplicación de Ω en el espacio de funciones de \mathbb{T} en \mathbb{R} . De manera formal, podemos expresar el concepto anterior como $X : \Omega \rightarrow \mathbb{R}^{\mathbb{T}}$, donde

$$\mathbb{R}^{\mathbb{T}} = \{(x(t))_{t \in \mathbb{T}} : x(t) \in \mathbb{R}, \forall t \in \mathbb{T}\}.$$

En $\mathbb{R}^{\mathbb{T}}$ se puede considerar la σ -álgebra $\mathcal{B}^{\mathbb{T}} = \sigma(\mathcal{C})$ (la σ -álgebra producto) que es la menor σ -álgebra que contiene a los conjuntos finito dimensionales (\mathcal{C}), donde

$$\mathcal{C} = \{(X(t))_{t \in \mathbb{T}} : (X(t_1) \in A_1, \dots, X(t_k) \in A_k); t_1, \dots, t_k \in \mathbb{T}, A_1, \dots, A_k \in \mathcal{B}, k \geq 1\}.$$

Se puede probar que X es $\mathcal{F}|\mathcal{B}^{\mathbb{T}}$ -medible, si y sólo si, $X(t)$ es $\mathcal{F}|\mathcal{B}$ -medible. Por tanto, la definición anterior de proceso coincide con la de una aplicación $\mathcal{F}|\mathcal{B}^{\mathbb{T}}$ medible en el sentido anterior. Gracias a esto, podemos hablar de la ley del

proceso. Cuando existe una aplicación de (Ω, \mathcal{F}, P) a $\mathbb{R}^{\mathbb{T}}$ hay una ley inducida y podemos definir $P_X(A) = P(X \in A)$, con $A \in \mathcal{B}^{\mathbb{T}}$, luego podemos hablar de la probabilidad de un conjunto.

Se dice que dos procesos son equivalentes si tienen la misma distribución. Se puede probar (ver [8]) que dos procesos tienen la misma distribución, si y sólo si, estos procesos dan la misma probabilidad a todos los conjuntos de la clase \mathcal{C} , es decir, si y sólo si ocurre

$$P(X(t_1) \in A_1, \dots, X(t_k) \in A_k) = P(Y(t_1) \in A_1, \dots, Y(t_k) \in A_k),$$

$$\forall t_1, \dots, t_k \in \mathbb{T}, A_1, \dots, A_k \in \mathcal{B}^{\mathbb{T}},$$

o equivalentemente

$$(X(t_1), \dots, X(t_k)) \stackrel{d}{=} (Y(t_1), \dots, Y(t_k)).$$

Las distribuciones del proceso evaluadas en una colección finita de instantes, se conocen con el nombre de finito dimensionales. Como consecuencia de lo anterior observamos que dos procesos tienen la misma distribución, si y sólo si, tiene las mismas distribuciones finito dimensionales.

Hablaremos de procesos donde las variables toman valores discretos y el tiempo es continuo. En este caso, nos interesa el concepto de filtración.

Definición 1.1. (Filtración). Sea (Ω, \mathcal{F}, P) un espacio probabilístico. Una *filtración* es una colección de sub σ -álgebras de \mathcal{F} , $\mathbb{F} = \{(\mathcal{F}_t)_{t \in \mathbb{T}}\}$, tales que

- 1.) $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}$ para todo $s < t \in \mathbb{T}$.

La condición 1.) significa que la familia de sub σ -álgebras es creceinte. Generalmente se asume que la filtración satisface

- 2.) $\mathcal{F}_s = \bigcap_{t > s} \mathcal{F}_t$ para todo s .

La condición 2.) indica que, en cierto sentido, la familia es continua por la derecha.

- 3.) $A \subset B \in \mathcal{F}, P(B) = 0 \Rightarrow A \in \mathcal{F}_0$

Nos solemos referir a la condición 3.) diciendo que la familia es completa.

Las condiciones 2.) y 3.) son conocidas como las "condiciones usuales".

La σ -álgebra \mathcal{F}_t se puede interpretar como aquella que contiene a todos los eventos que ocurren o no hasta el instante de tiempo t , luego podemos entender dicho concepto como "la historia hasta el instante t ". Otro concepto interesante

es el de pre- t σ -álgebra \mathcal{F}_{t-} que hace referencia a la σ -álgebra más pequeña que contiene a todos los \mathcal{F}_s , $s < t$ y en este caso podemos interpretar \mathcal{F}_{t-} como todos los eventos que se han producido estrictamente antes del instante t .

Definición 1.2. (Proceso adaptado). Decimos que X es *adaptado* a la filtración \mathbb{F} si $X(t)$ es \mathcal{F}_t -medible para todo $t \geq 0$.

Un ejemplo importante de filtración es la generada por un proceso estocástico X , es decir, el caso en el que $\mathcal{N}_t = \sigma(\{X(s) : s \leq t\})$ y la llamaremos *filtración natural*. Cualquier proceso es adaptado a la filtración natural. En este contexto solemos hacer referencia a una filtración \mathbb{F} más grande, donde cada sub σ -álgebra de \mathcal{F} tiene la forma $\mathcal{F}_t = \mathcal{F}_0 \vee \mathcal{N}_t$, donde \mathcal{N}_t es la filtración natural, siendo cada una de ellas crecientes y continuas por la derecha.

Un resultado importante (ver [9]) afirma que la *filtración* generada por un proceso de salto continuo por la derecha es continua por la derecha, es decir, si un proceso de salto X es un proceso tal que para cada t y ω , $X(s, \omega)$ es constante en $s \in [t, t + \epsilon)$ para algún $\epsilon > 0$. Gracias a este resultado podemos demostrar que las filtraciones generadas cuando los eventos se producen a lo largo de un tiempo continuo son automáticamente continuas por la derecha.

Pasaremos a estudiar la descripción de los procesos *martingala* y *predecible*, que son dos tipos de procesos con una gran importancia en esta teoría.

Definición 1.3. (Martingala) Una *martingala* es un proceso cadlag adaptado M que es integrable, es decir, $E[|M(t)|] < \infty$ para todo $t \in \mathbb{T}$ y cumple la propiedad:

$$E[M(t)|\mathcal{F}_s] = M(s), \text{ para todo } s \leq t. \quad (1.1)$$

El proceso M es una submartingala si reemplazamos la igualdad por la desigualdad $E[M(t)|\mathcal{F}_s] \geq M(s)$ para todo $s \leq t$.

Una clases de procesos complementarios a los proceso *martingala* son los procesos predecibles, que definiremos a continuación.

Si $\mathbb{T} = [0, T]$ o $[0, \infty)$ entonces la σ -álgebra asociada es $\mathcal{B}([0, T])$ o $\mathcal{B}([0, \infty))$, respectivamente. Llamamos a la sub σ -álgebra de $\mathcal{B}([0, T]) \otimes \mathcal{F}$ generada por el conjunto producto de la forma $(s, t] \times U$, donde $U \in \mathcal{F}_s$, $s \leq t$ y $t \in \mathbb{T}$, la σ -álgebra predecible y la denotamos por \mathcal{G} .

Por tanto, sea $X : \mathbb{T} \times \Omega \rightarrow \mathbb{R}$ un proceso estocástico definido por $(t, \omega) \mapsto X(t, \omega)$, X es medible si es $\mathcal{B}([0, T]) \otimes \mathcal{F}$ medible.

Definición 1.4. (Proceso predecible) Un proceso estocástico X es predecible si es $\mathcal{G}|\mathcal{B}$ medible.

De manera menos formal, estos procesos están generados por aquellos procesos adaptados, acotados y continuos por la izquierda.

1.2. Procesos de puntos y procesos de conteo

En esta sección introduciremos los conceptos de proceso de puntos y de conteo en \mathbb{R}_+ y mostraremos que los dos conceptos se pueden interpretar de la misma manera.

Definición 1.5. (Proceso de puntos).

Sea $T = \{T_n : n \in \mathbb{N}\}$ una sucesión de variables aleatorias definidas en el espacio probabilístico (Ω, \mathcal{F}, P) y que toman valores en $\hat{\mathbb{R}}_0 = \mathbb{R}_0 \cup \infty$, siendo $\mathbb{R}_0 = [0, \infty)$. Entonces T es un proceso de puntos si satisface:

1. $T_0 = 0$,
2. $0 < T_1 < T_2 < \dots$,
3. $\lim_{n \rightarrow \infty} T_n = \infty$.

Como podemos observar, un proceso de puntos se puede definir como una secuencia positiva, en algunos casos infinita, y estrictamente creciente de variables aleatorias. La interpretación de la variable T_n será el instante en el que el evento n -ésimo tiene lugar. En el caso de que $T_n = \infty$, lo interpretaremos como que el número total de eventos ocurridos a lo largo de \mathbb{R}_0 es menor a n .

También mencionaremos los procesos de conteo, que se relacionan de forma biyectiva con los procesos de puntos.

Definición 1.6. (Proceso de conteo). Sea T un proceso de puntos en el espacio probabilístico (Ω, \mathcal{F}, P) . El proceso estocástico $N = \{N(t); t \geq 0\}$ en el que se cumple que $N(0) = 0$ y

$$N(t) = \sum_n \mathbf{1}_{\{T_n \leq t\}} \text{ para } t > 0,$$

es el *proceso de conteo* asociado a T .

Observamos que $N(t)$ es el número de sucesos que se producen en el intervalo $(0, t]$ y que es equivalente conocer $N(t), t \geq 0$ o $\{T_n : n \in \mathbb{N}\}$, ya que $N(t) = n$ si y sólo si $T_n \leq t < T_{n+1}$ siendo T_n el instante del n -ésimo salto de $N(t)$. Luego podemos referirnos indistintamente al proceso de puntos o al proceso de conteo.

También podemos observar que de la definición de proceso de puntos, para un tiempo fijo $t < \infty$, $N_t(\omega)$ es una variable aleatoria que toma valores en el conjunto \mathbb{N} .

Por otra parte, podemos considerar que dada una realización $\omega \in \Omega$, entonces $N_t(\omega)$ está representado como una función en t escalonada y continua por la derecha.

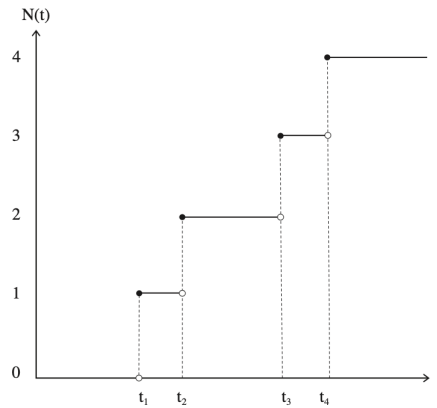


Figura 1.1: Ejemplo de un proceso de conteo $N(t)$

De forma más general, supongamos dada una filtración $\mathbb{F} = \{\mathcal{F}_t : t \in \mathbb{T}\}$ sobre un espacio probabilístico (Ω, \mathcal{F}, P) que satisface las condiciones usuales, entonces definimos un proceso de conteo multivariante $N = (N_1, \dots, N_k)$ como un vector de k procesos de cadlag adaptados, todos ellos con valor nulo cuando el tiempo es igual a cero, presentan trayectorias constantes a trozos y no decrecientes con saltos de tamaño $+1$, sin producirse saltos de manera simultánea entre dos componentes distintas. Debido a que no se producen saltos simultáneos, la suma de sus componentes es también un proceso de conteo que se puede definir como

$$N_{\bullet} = \sum_{h=1}^k N_h.$$

A las diferencias $\tau_i = T_i - T_{i-1}$ las llamaremos tiempos entre sucesos y nos ayudaran a definir un proceso de renovación.

Definición 1.7. Se dice que un proceso de conteo $N = \{N(t) : t \geq 0\}$ es un proceso de renovación si los tiempos entre sucesos $\{\tau_n : n \geq 1\}$ son independientes e idénticamente distribuidos.

1.3. Procesos de Poisson

En esta sección definiremos el proceso de Poisson de tasa λ , que es el ejemplo más importante de proceso de puntos o de proceso de conteo en $[0, \infty)$, y expon-dremos algunos resultados importantes obtenidos a partir de dicho proceso.

Definición 1.8. (Proceso de Poisson) Sea $\tau = \{\tau_n : n \geq 1\}$ un proceso de puntos donde cada variable aleatoria independiente τ_i sigue una distribución exponencial de tasa λ . Sea $T_0 = 0$ y $T_n = \sum_{i=1}^n \tau_i$ para $n \geq 1$. Entonces el proceso de *Poisson* de tasa λ , que es un tipo de proceso de conteo, está definido por

$$N(t) = \text{máx}\{n : T_n \leq t\}, \quad t \geq 0.$$

Usaremos la notación $N(a, b] = N(b) - N(a)$ para $0 \leq a < b$, luego $N(t) = N(0, t]$ y también llamaremos a este proceso, proceso de Poisson homogéneo. Las variables τ_i hacen referencia al tiempo que transcurre entre eventos sucesivos (en nuestro caso, entre una infección y la siguiente), siendo $T_n = \tau_1 + \tau_2 + \dots + \tau_n$ el instante en el que ocurre el n -ésimo evento y $N(t)$ el número de eventos que se han producido hasta el instante t .

Para hacernos una idea de como funciona, nos fijaremos en el siguiente diagrama:

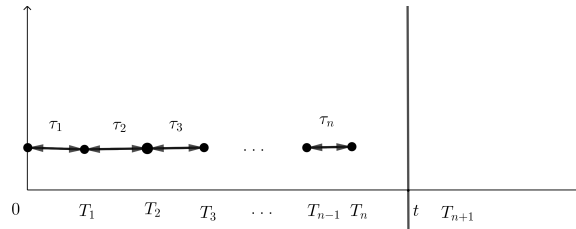


Figura 1.2: Interpretación del proceso de Poisson $N(t) = n$.

Si $\{\tau_n : n \geq 1\}$ son los tiempos entre sucesos en un $PPH(\lambda)$ y s_1, \dots, s_n son los valores que toma cada variable τ_i , entonces τ_1, \dots, τ_n tiene función de densidad conjunta $f_\tau(s_1, \dots, s_n) = \exp(-\lambda(s_1 + \dots + s_n))$ si $s_i > 0, i = 1, \dots, n$. Como ya hemos mencionado anteriormente $T_n = \sum_{i=1}^n \tau_i$ y si realizamos un cambio de variable, se obtiene que T_1, \dots, T_n tienen densidad conjunta $f_T(t_1, \dots, t_n) = \lambda^n \exp(-\lambda t_n)$ si $0 < t_1 < \dots < t_n$. Se observa que $(N(t) = i) \equiv (T_i \leq t < T_{i+1})$. Entonces, si $0 < t_1 < \dots < t_n$,

$$\begin{aligned} P(N(t_1) = i_1, N(t_2) = i_1 + i_2, \dots, N(t_n) = i_1 + \dots + i_n) \\ = P(T_{i_1} \leq t_1 < T_{i_1+1}, T_{i_1+i_2} \leq t_2 < T_{i_1+i_2+1}, \dots, T_{i_1+\dots+i_n} \leq t_n < T_{i_1+\dots+i_n+1}) \end{aligned}$$

Veamos el caso donde $n=1$,

$$\begin{aligned} P(T_{i_1} \leq t_1 < T_{i_1+1}) &= \int_{\{0 < y_1 < \dots < y_{i_1} \leq t_1 < y_{i_1+1}\}} \lambda^{i_1+1} e^{-\lambda y_{i_1+1}} dy_1 \dots dy_{i_1} dy_{i_1+1} \\ &= \lambda^{i_1} \int_{\{0 < y_1 < \dots < y_{i_1} \leq t_1\}} 1 dy_1 \dots dy_{i_1} \int_{t_1}^{\infty} \lambda e^{-\lambda y_{i_1+1}} dy_{i_1+1} = \lambda^{i_1} \frac{t_1^{i_1}}{i_1!} e^{-\lambda t_1} \end{aligned}$$

Si $n=2$

$$\begin{aligned} P(T_{i_1} \leq t_1 < T_{i_1+1}, T_{i_1+i_2} \leq t_2 < T_{i_1+i_2+1}) &= \int_{\{0 < y_1 < \dots < y_{i_1} \leq t_1 < y_{i_1+1} < \dots < y_{i_1+i_2} \leq t_2 < y_{i_1+i_2+1}\}} \lambda^{i_1+1} e^{-\lambda y_{i_1+1}} dy_1 \dots dy_{i_1} dy_{i_1+i_2+1} \\ &= \lambda^{i_1+i_2} \int_{\{0 < y_1 < \dots < y_{i_1} \leq t_1\}} 1 dy_1 \dots dy_{i_1} \int_{\{t_1 < y_{i_1+1} < \dots < y_{i_1+i_2} \leq t_2\}} 1 dy_{i_1+1} \dots dy_{i_1+i_2} \\ &\quad \int_{t_2}^{\infty} \lambda e^{-\lambda y_{i_1+i_2+1}} dy_{i_1+i_2+1} = \lambda^{i_1+i_2} \frac{t_1^{i_1}}{i_1!} \frac{(t_2 - t_1)^{i_2}}{i_2!} e^{-\lambda t_2}. \end{aligned}$$

De manera general, tenemos que

$$\begin{aligned} P(T_{i_1} \leq t_1 < T_{i_1+1}, T_{i_1+i_2} \leq t_2 < T_{i_1+i_2+1}, \dots, T_{i_1+\dots+i_n} \leq t_n < T_{i_1+\dots+i_n+1}) \\ = \lambda^{i_1+\dots+i_n} \frac{t_1^{i_1}}{i_1!} \dots \frac{(t_n - t_{n-1})^{i_n}}{i_n!} e^{-\lambda t_n} \end{aligned}$$

Si utilizamos la equivalencia existente entre un proceso de puntos y un proceso de conteo, tenemos que:

$$\begin{aligned} P(N(t_1) = i_1, N(t_2) - N(t_1) = i_2, \dots, N(t_n) - N(t_{n-1}) = i_n) \\ = \frac{e^{-\lambda t_1} (\lambda t_1)^{i_1}}{i_1!} \frac{e^{-\lambda(t_2-t_1)} (\lambda(t_2-t_1))^{i_2}}{i_2!} \dots \frac{e^{-\lambda(t_n-t_{n-1})} (\lambda(t_n-t_{n-1}))^{i_n}}{i_n!} \end{aligned}$$

Mediante la realización de estos cálculos hemos probado el siguiente resultado.

Proposición 1.9. *Un proceso de conteo es un proceso de Poisson homogéneo de tasa λ , si y sólo si tiene incrementos independientes y estacionarios y, para $s < t$, $N(t) - N(s) = N(s, t]$ sigue una distribución de Poisson de parámetro $\lambda(t - s)$.*

Pasaremos a enunciar el proceso de Poisson no homogéneo que también presenta un interés especial.

Definición 1.10. (Proceso de Poisson no homogéneo). Sea $N = \{N(t); t \geq 0\}$ un proceso de conteo, llamamos proceso de *Poisson* no homogéneo de tasa $\lambda(s)$, no negativa, si verifica las siguientes propiedades:

- i) $N(0) = 0$.
- ii) $N(t)$ tiene incrementos independientes.

- iii) $N(t) - N(s)$ sigue una distribución de Poisson de media $\int_s^t \lambda(r)dr = \Lambda(t) - \Lambda(s)$.

En esta definición hemos utilizado el término de función de intensidad que desarrollamos en la siguiente sección 1.5.

1.4. Función de supervivencia y función de riesgo

En esta sección vamos a discutir varias formas equivalentes de describir la distribución de una variable aleatoria positiva. Estas variables pueden representar la duración de vida de un componente o el tiempo de espera hasta que se observa un suceso. Estas descripciones equivalentes serán útiles en la discusión sobre intensidades de procesos de conteo en una sección posterior.

Sea T una variable aleatoria positiva, que podemos interpretar como el tiempo que ha transcurrido hasta que se produce un suceso, como por ejemplo, el tiempo hasta que un individuo se infecta de una determinada enfermedad. Asumiremos que T tiene función de distribución F y función de densidad f . Por simplicidad asumiremos que la función de densidad es continua. La *función de supervivencia*, $S(t)$, representa la probabilidad de sobrevivir más allá del instante t , es decir,

$$S(t) = P(T > t) = 1 - F(t).$$

Además se tiene que $S(0) = 1$, $\lim_{t \rightarrow \infty} S(t) = 0$ y $S(t_1) \geq S(t_2)$ si $t_1 < t_2$, ya que la función es decreciente.

Ahora definiremos la *función de riesgo*, $h(t)$, que representa la tasa instantánea de probabilidad de no sobrevivir en un instante inmediatamente posterior a t cuando se ha sobrevivido hasta el instante t . De manera formal,

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} P(T \in (t, t + \Delta t] | T > t) = \lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - F(t)}{\Delta t} \frac{1}{S(t)} = \frac{f(t)}{S(t)}$$

En varias ocasiones se utiliza la función de riesgo acumulado, que viene definida por $H(t) = \int_0^t h(s)ds$.

Podemos relacionar la función de supervivencia y la función de riesgo, si conocemos la función de distribución mediante la siguiente relación:

$$h(t) = H'(t) = \frac{f(t)}{S(t)} = \frac{-S'(t)}{S(t)} = (-\log S(t))'$$

y se cumple que $H(0) = -\log S(0) = 0$. Por tanto, obtenemos que $H(t) = -\log S(t)$ o, equivalentemente,

$$S(t) = \exp(-H(t)) = \exp\left(-\int_0^t h(s)ds\right).$$

1.5. Intensidades

En esta sección desarrollaremos el cálculo de la función de intensidad mediante resultados que podemos obtener de [8]. Un proceso de puntos regular es aquel en el que (t_1, \dots, t_n) tiene densidad conjunta para todo n . Definiremos esta intensidad para procesos de puntos regulares.

Supongamos que las observaciones de los proceso ocurren en un intervalo de tiempo $[0, T]$, luego los resultados pueden estar descritos en términos de procesos de puntos en \mathbb{R}_+ . Sea $\{T_n : n \geq 1\}$ un proceso de puntos en el intervalo $(0, T)$, la función de densidad de T_n condicionada por $T_1 = t_1, \dots, T_{n-1} = t_{n-1}$ es

$$p_n(t|t_1, \dots, t_{n-1}) = \frac{f_{T_1, \dots, T_n}(t_1, \dots, t_{n-1}, t)}{f_{T_1, \dots, T_{n-1}}(t_1, \dots, t_{n-1})}, \quad (1.2)$$

es decir, la densidad conjunta de las variables T_1, \dots, T_{n-1}, T_n entre la densidad marginal de T_1, \dots, T_{n-1} y con soporte de T_n en (t_{n-1}, ∞) .

Si $n = 2$, la función de densidad vendrá dada por $f_{T_1, T_2}(t_1, t_2) = f_{T_1}(t_1)p_2(t_2|t_1) = p_1(t_2|t_1)$.

Denotamos la función de supervivencia para la variable T_n condicionada por T_1, \dots, T_{n-1} como

$$\begin{aligned} S_n(t|t_1, \dots, t_{n-1}) &= P(T_n > t|t_1, \dots, t_{n-1}) \\ &= 1 - P(T_n \leq t|t_1, \dots, t_{n-1}) = 1 - \int_{t_{n-1}}^t p_n(s|t_1, \dots, t_{n-1})ds \end{aligned}$$

que representa la probabilidad de supervivencia más allá del instante t .

Ahora tiene sentido que si conocemos la función de supervivencia en el instante t condicionada por t_1, \dots, t_{n-1} podemos definir la función de riesgo de t condicionada por t_1, \dots, t_{n-1} como

$$h_n(t|t_1, \dots, t_{n-1}) = \frac{p_n(t|t_1, \dots, t_{n-1})}{S_n(t|t_1, \dots, t_{n-1})}.$$

De manera habitual se suele utilizar la función de riesgo acumulado que viene dada por

$$H_n(t|t_1, \dots, t_{n-1}) = \int_0^t h_n(s|t_1, \dots, t_n)ds.$$

Definición 1.11. La función de intensidad condicionada de un proceso de puntos regular en \mathbb{R}_+ es la función representada por $\lambda^*(\cdot)$ y definida por

$$\lambda^*(t) = \begin{cases} h_1(t) & t \leq t_1 \\ h_n(t | t_1, \dots, t_{n-1}) & t_{n-1} < t \leq t_n, n \geq 2 \end{cases} \quad (1.3)$$

En verdad, como podemos apreciar en (1.3) la intensidad $\lambda^*(u)$ es $\lambda^*(u|\mathcal{F}_u)$, ya que $\lambda^*(u)$ depende de los sucesos observados hasta el instante u .

Veamos cómo es dicha función en procesos que ya conocemos, como los procesos de Poisson homogéneos y de renovación.

- **PPH(λ).** La función de densidad de t_1 en un proceso de Poisson homogéneo es $p_1(t) = \lambda e^{-\lambda t}$, luego su función de supervivencia será $S_1(t) = e^{-\lambda t}$ y su función de riesgo $h_1(t) = \lambda$. La función de intensidad de t_2 condicionada por t_1 es $p_2(t|t_1) = \lambda e^{-\lambda(t-t_1)}$ siendo $t > t_1$, donde su función de supervivencia es $S_2(t|t_1) = e^{-\lambda(t-t_1)}$ y de riesgo vuelve a ser $h_2(t|t_1) = \lambda$. Como conclusión, este proceso presenta una función de intensidad condicionada constante e igual a λ .
- **Proceso de renovación.** En este proceso las variables $\tau_i = T_i - T_{i-1}, i \geq 1$ son independientes e idénticamente distribuidas, la densidad conjunta de T_1, \dots, T_n es $f_{T_1, \dots, T_n}(t_1, \dots, t_n) = f(t_1) \cdots f(t_n - t_{n-1})$ donde $0 < t_1 < \dots < t_n$. Para $n = 1$, la función de densidad es $p_1(t) = f(t)$ y la de supervivencia $S_1(t) = s(t)$, luego la función de riesgo será $h_1(t) = h(t)$. Para n la función de densidad es

$$p_n(t|t_1, \dots, t_{n-1}) = \frac{f_{T_1, \dots, T_n}(t_1, \dots, t_{n-1}, t)}{f_{T_1, \dots, T_{n-1}}(t_1, \dots, t_{n-1})} = \frac{f(t_1) \cdots f(t_{n-1} - t_{n-2})f(t - t_{n-1})}{f(t_1) \cdots f(t_{n-1} - t_{n-2})}$$

$$= f(t - t_{n-1}),$$

la función de supervivencia es $S_n(t|t_1, \dots, t_{n-1}) = s(t - t_{n-1})$, luego la función de riesgo será $h_n(t|t_1, \dots, t_{n-1}) = h(t - t_{n-1})$. Por tanto, podemos definir la función de intensidad en este proceso como

$$\lambda^*(t) = \begin{cases} h(t) & t \leq t_1 \\ h(t - t_{n-1}) & t_{n-1} < t \leq t_n, n \geq 2 \end{cases}$$

Discutiremos a continuación el proceso de verosimilitud. Sea N un proceso de conteo en $[0, T]$ para T positiva y finita, y supongamos que en el intervalo $[0, T]$ se

han observados $t_1, \dots, t_{N(T)}$. Si $N(T) = n$, teniendo en cuenta que es equivalente a $T_n \leq T < T_{n+1}$, tendremos que

$$\begin{aligned} P((T_1, \dots, T_n) \in A | N(T) = n) &= \frac{P((T_1, \dots, T_n) \in A, T_n \leq T < T_{n+1})}{P(N(T) = n)} \\ &= \frac{1}{P(N(T) = n)} \int_{\{(t_1, \dots, t_n) : (t_1, \dots, t_n) \in A, t_n \leq T < t_{n+1}\}} f_{T_1, \dots, T_{n+1}}(t_1, \dots, t_{n+1}) dt_1 \dots dt_{n+1} \\ &= \frac{1}{P(N(T) = n)} \int_{\{(t_1, \dots, t_n) \in A, t_n \leq T\}} p_1(t_1) p_2(t_2 | t_1) \dots p_n(t_n | t_1, \dots, t_{n-1}) S_{n+1}(T | t_1, \dots, t_n) dt_1 \dots dt_n \\ &= \frac{1}{P(N(T) = n)} \int_{A \cap \{t_n \leq T\}} p_1(t_1) \dots p_n(t_n | t_1, \dots, t_{n-1}) S_{n+1}(T | t_1, \dots, t_n) dt_1 \dots dt_n. \end{aligned}$$

donde $S_{n+1}(T | t_1, \dots, t_n) = \int_T^\infty p_{n+1}(t_{n+1} | t_1, \dots, t_n) dt_{n+1}$.

Por tanto, dado $N(T) = n$, (T_1, \dots, T_n) tienen densidad conjunta

$$L := p_1(t_1) \dots p_n(t_n | t_1, \dots, t_{n-1}) S_{n+1}(T | t_1, \dots, t_n).$$

Si utilizamos (1.2) entonces L es igual a:

$$\begin{aligned} L &= h_1(t_1) \cdot \exp\left(-\int_0^{t_1} h_1(u) du\right) \\ &h_2(t_2 | t_1) \cdot \exp\left(-\int_{t_1}^{t_2} h_2(u | t_1) du\right) \\ &\dots \\ &h_n(t_n | t_1, \dots, t_{n-1}) \cdot \exp\left(-\int_{t_{n-1}}^{t_n} h_n(u | t_1, \dots, t_{n-1}) du\right) \\ &\exp\left(-\int_{t_n}^T h_{n+1}(u | t_1, \dots, t_n) du\right) \end{aligned}$$

siendo la forma compacta $L = \prod_{i=1}^n \lambda^*(t_i) \cdot \exp\left(-\int_0^T \lambda^*(u) du\right)$. Tomando logaritmos tendremos $\log L$:

$$\log L = \sum_{i=1}^n \log \lambda^*(t_i) - \int_0^T \lambda^*(u) du.$$

Dada una función constante a trozos, la integral de Lebesgue-Stieltjes (4.6 de [8]) de una función continua por la izquierda $\lambda^*(t)$ con respecto a una función constante a trozos es la suma de los valores de la función en los puntos de salto multiplicada por la magnitud de los saltos. Por eso, la $\log L$ en un proceso de conteo en el intervalo $[0, T]$ es

$$\log L = \int_0^T \log \lambda^*(u) dN(u) - \int_0^T \lambda^*(u) du.$$

Esta igualdad es válida independiente del valor de $N(T)$ y se utiliza como log-verosimilitud, es decir, como base para la estimación de parámetros en procesos de conteo.

Un resultado importante del que daremos una versión simplificada (**Theorem A3.4.IX** de [8]) es el teorema de Doob-Meyer, de forma que un proceso de conteo admite una descomposición (única) como suma de una martingala más un proceso predecible. Este resultado nos ayudará a describir el término de compensador.

Definición 1.12. Supongamos que $N = \{N(t) : 0 \leq t < \infty\}$ es un proceso de conteo adaptado a la filtración \mathbb{F} y que admite intensidad $\lambda^*(t)$ continua por la derecha. Definimos el compensador $\Lambda^*(t)$ como la integral

$$\Lambda^*(t) = \int_0^t \lambda^*(u) du.$$

Dicha integral verifica que $M(t) = N(t) - \Lambda^*(t)$ es una martingala.

Por últimos, expondremos un proceso que presenta especial interés cuando hablamos de función de intensidad. Un proceso de Cox condicionalmente dado el proceso de intensidad es un proceso de Poisson.

Definición 1.13. (Proceso de Cox) Sea (Ω, \mathcal{F}, P) un espacio con una filtración \mathbb{F} . Sea $N(t)$ un proceso de conteo adaptado a \mathcal{F}_t y $\lambda^*(t)$ un proceso adaptado a \mathcal{F}_t no negativo tal que $\int_0^t \lambda^*(s) ds < \infty$ casi siempre. Si para todo $t \in \mathbb{T}$ y $u \in \mathbb{R}$ se tiene

$$E[\exp(iuN(t)) | \mathcal{F}_t] = \exp((e^{iu} - 1) \int_0^t \lambda^*(s) ds)$$

entonces $N(t)$ es conocido como proceso de Cox con intensidad $\lambda^*(t)$.

Capítulo 2

Modelos deterministas

El modelo determinista que vamos a desarrollar es el modelo SIR. Dicho modelo fue inicialmente propuesto por O. Kermack y A. G. McKendrick en 1927 (ver [4]) con el objetivo de desarrollar matemáticamente la evolución de una epidemia en una población de gran tamaño expuesta al virus.

Una de las características del modelo es que la población sobre la que se realiza el estudio es constante, es decir, el número de nacimientos y defunciones se mantiene estable, al igual que no se tendrán en cuenta las inmigraciones y emigraciones de los individuos que constituyen la población; por tanto, el tamaño de la población sobre la que realizamos el estudio será N , donde $N = S(t) + I(t) + R(t)$ para todo valor de t . Las variables S, I, R hacen referencia al número de susceptibles, infectados y recuperados, respectivamente, en cada instante de tiempo t .

Otro dato a tener en cuenta es que la población es homogénea en el sentido de que dos individuos cualquiera presentan la misma tasa de entrar en contacto β , luego no existe una distinción a la hora de producirse un contacto estrecho entre diferentes individuos. Además, el periodo de latencia que existe desde que el individuo se expone al agente infeccioso hasta que el propio individuo comienza a ser infeccioso es lo suficientemente pequeño como para no tenerlo en cuenta. La última consideración sería que los individuos infectados se recuperarán y pasarán al compartimento de los recuperados con una tasa constante γ . El esquema que representa los estados por los que un individuo puede pasar a lo largo del modelo SIR es la Figura 2.1.

Las ecuaciones diferenciales que modelan la evolución del número de susceptibles, infectados y recuperados en la población de tamaño N son (2.1):



Figura 2.1: Esquema modelo SIR.

$$\begin{cases} \dot{S}(t) = -\beta S(t) \frac{I(t)}{N} \\ \dot{I}(t) = \beta S(t) \frac{I(t)}{N} - \gamma I(t) \\ \dot{R}(t) = \gamma I(t) \end{cases} \quad (2.1)$$

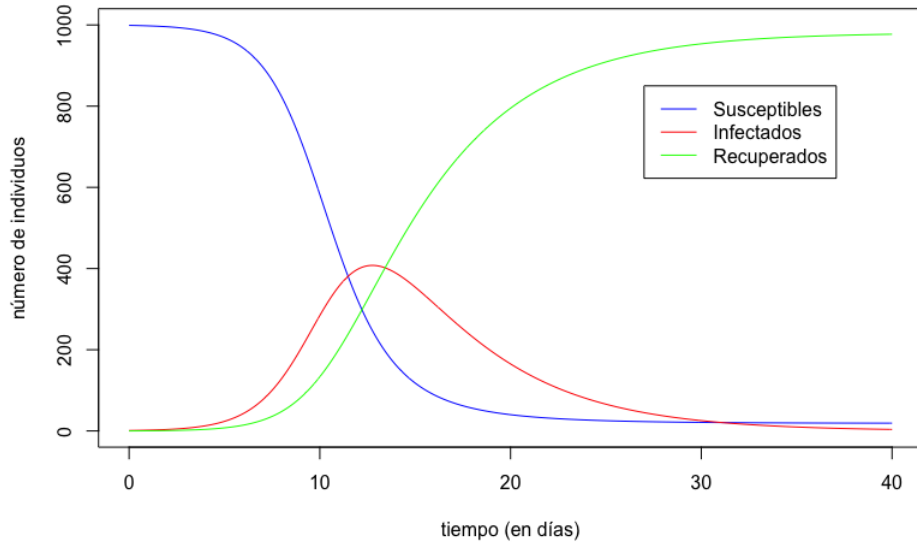
cuyas condiciones iniciales son $S(0) = S_0, I(0) = I_0, R(0) = R_0$. Si suponemos que al comienzo de la epidemia el número de recuperados es 0, $R_0 = 0$, siendo esta suposición la más coherente ya que al inicio del proceso de expansión de la enfermedad solo hay individuos susceptibles e infectados, es decir, $N = S_0 + I_0$.

El signo menos de la primera ecuación de 2.1 hace referencia a la disminución del número de individuos susceptibles que se produce con el paso del tiempo, ya que a medida que se infectan abandonan ese estado S y pasan al estado I . En la segunda ecuación el coeficiente positivo indica que, tras el contacto con individuos infectados, aquellos individuos susceptibles se van a infectar y, por lo tanto, aumentará el número de individuos infectados. Durante el estudio de como evoluciona la población es posible que algunos de los individuos se hayan recuperado, estos dejarán de pertenecer al grupo de infectados I para formar parte del grupo de recuperados R y alcanzar la inmunidad; perteneciendo también al grupo de recuperados R los individuos que han fallecido a causa de la enfermedad infecciosa.

Otro dato a tener en cuenta es que $\dot{S}(t) \leq 0$ y $\dot{R}(t) \geq 0$ según hemos definido en (2.1), luego como $0 \leq S(t) \leq S(0) \leq N$ y $0 \leq R(0) \leq R(t) \leq N$, los límites $S(\infty) = \lim_{t \rightarrow \infty} S(t)$, $R(\infty) = \lim_{t \rightarrow \infty} R(t)$ son finitos y como existe una relación con el número de infectados tenemos que $I(\infty) = \lim_{t \rightarrow \infty} I(t) = N - S(\infty) - R(\infty)$. Por tanto, es fácil deducir que la enfermedad siempre se extingue, es decir, $I(\infty) = 0$ dadas cualquiera de sus condiciones iniciales.

Para que tengamos una visión general de la evolución de cada uno de los compartimentos del modelo SIR, nos fijaremos en la Figura 3.2 y entenderemos la coherencia de la misma. Como podemos observar en el gráfico el número de infectados crece al comienzo de la epidemia hasta que alcanza el número máximo de infectados y desde ahí decrece hasta llegar a cero, que será el momento en el

que la epidemia finalice. Con respecto al número de susceptibles y recuperados, los individuos susceptibles presentan el valor máximo al comienzo de la epidemia, ya que en ese momento es cuando la población presenta un mayor riesgo de exposición a la enfermedad, en cambio el número de recuperados o fallecidos siempre crece hasta que toda la población ha sido infectada y se ha recuperado o ha fallecido algún individuo.



Los parámetros utilizados en el modelo SIR son $N=1000, \beta = 0.85, \gamma = 0.21$.

Figura 2.2: Representación gráfica del modelo SIR

El siguiente paso será realizar un análisis cualitativo del modelo de estudio (ver [5]) y para comenzar, normalizaremos las variables dividiendo cada una por el total de individuos de la población N

$$s(t) = \frac{S(t)}{N}, I(t) = \frac{I(t)}{N}, R(t) = \frac{R(t)}{N}.$$

Si derivamos respecto de t en ambas miembros de cada una de las expresiones anteriores y realizamos las sustituciones necesarias obtenemos el siguiente sistema:

$$\begin{cases} s'(t) = \frac{ds(t)}{dt} = \frac{1}{N} \frac{dS(t)}{dt} = -\beta s(t)i(t) \\ i'(t) = \frac{di(t)}{dt} = \frac{1}{N} \frac{dI(t)}{dt} = \beta s(t)i(t) - \gamma i(t) \\ r'(t) = \frac{dr(t)}{dt} = \frac{1}{N} \frac{dR(t)}{dt} = \gamma i(t) \end{cases} \quad (2.2)$$

Como es posible calcular el número de individuos recuperados a través de la relación $R(t) = N - S(t) - I(t)$, podemos simplificar el sistema (3.2) y obtener un sistema más simple formado únicamente por dos ecuaciones:

$$\begin{cases} s'(t) = -\beta s(t)i(t) \\ i'(t) = \beta s(t)i(t) - \gamma i(t) \end{cases} \quad (2.3)$$

Para conocer si se produce un brote epidémico en la población realizaremos el siguiente análisis:

Caso 1. Si $i'(0) > 0$ tenemos que $i'(0) = i_0(\beta s_0 - \gamma) > 0$ si $s_0 > \frac{\gamma}{\beta}$, luego el número de infectados aumentará y se producirá una epidemia. Entonces para algún $t > 0$, existirá un brote epidémico si $i(t) > i_0$, ya que el valor de $i_0 > 0$.

Caso 2. Si $i'(0) < 0$ tenemos que $i'(0) = i_0(\beta s_0 - \gamma) < 0$ si $s_0 < \frac{\gamma}{\beta}$ y no se producirá una epidemia debido a que si observamos la primera ecuación $s'(t) < 0$ para cualquier instante de tiempo t , luego $s'(t) \leq s_0$ para cualquier $t \geq 0$.

Si nos fijamos en el Caso 1, podemos obtener el siguiente resultado:

Si $i'(0) > 0$, entonces $i'(0) = i_0(\beta s_0 - \gamma) > 0$ y si tenemos en cuenta las condiciones iniciales sabemos que $i_0 > 0$, luego $i'(0) = i_0(\beta s_0 - \gamma) > 0$ si $\beta s_0 > \gamma$, es decir, $\frac{\beta s_0}{\gamma} > 1$, y en este caso se producirá un brote epidémico.

Para el caso 2 tenemos la situación de que el cociente $\frac{\beta s_0}{\gamma} < 1$ y entonces no se producirá un brote epidémico.

En la literatura habitual al cociente que determina si se producirá un brote o no lo denotaremos por $R_0 = \frac{\beta s_0}{\gamma}$ (ver [5]) y lo llamaremos **número básico de reproducción** de la epidemia haciendo referencia al número promedio de infectados secundarios producidos por un individuo infectado anteriormente sobre una población totalmente susceptible.

Para $i(t) \neq 0$, un método para encontrar una solución al sistema (2.3) puede ser el siguiente:

$$\frac{i'(t)}{s'(t)} = \frac{(\beta s(t) - \gamma)i(t)}{-\beta s(t)i(t)} = \frac{\beta s(t) - \gamma}{-\beta s(t)} = -1 + \frac{\gamma}{\beta s(t)}.$$

Utilizamos el método de separación de variables para resolver el sistema y obtenemos el siguiente resultado

$$i(t) = -s(t) + \frac{\gamma}{\beta} \log s(t) + k,$$

donde k representa una constante de integración que se puede obtener a partir de las condiciones iniciales s_0 e i_0 , y gracias a esto, tenemos que las curvas solución del plano fase (s, i) están determinadas por

$$p(s, i) = i(t) + s(t) - \frac{\gamma}{\beta} \log s(t) = k, \quad (2.4)$$

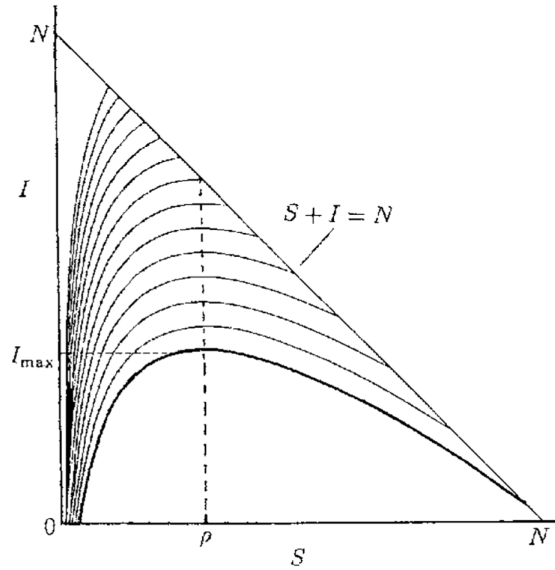


Figura 2.3: Trayectorias del plano fase $p(s,i)$ [4].

como podemos ver en la siguiente figura:

Cada curva de la Figura 2.3 representa la curva solución del modelo (2.3) para distinto valores de las condiciones iniciales y la línea discontinua al valor de $\rho = \frac{\gamma}{\beta}$.

El siguiente paso, será calcular el valor de la constante de integración k , pues $k = p(s_0, i_0) = i_0 + s_0 - \frac{\gamma}{\beta} \log s_0$ y si sustituimos k en 2.4 tenemos que

$$i(t) = -s(t) + \frac{\gamma}{\beta} \log s(t) + i_0 + s_0 - \frac{\gamma}{\beta} \log s_0. \quad (2.5)$$

Como hemos mencionado anteriormente, el tamaño de la población es constante y lo hemos denotado por N , si ahora introducimos una pequeña cantidad de individuos infectados podemos suponer que $s(0) \approx N$ e $i(0) \approx 0$, entonces $R_0 = \frac{\beta N}{\gamma}$. Si nos fijamos en la situación de que el número de infectados con el paso del tiempo tiende a cero $i(\infty) = \lim_{t \rightarrow \infty} i(t) = 0$ y tomando $s(\infty) = \frac{S(\infty)}{N}$, tenemos que $p(s(t), i(t)) = p(s(\infty), 0)$, es decir, $N - \frac{\gamma}{\beta} \log s_0 = s(\infty) - \frac{\gamma}{\beta} \log s(\infty)$, así obtenemos el siguiente resultado

$$\frac{\beta}{\gamma} = \frac{\log \frac{s_0}{s(\infty)}}{N - s(\infty)}. \quad (2.6)$$

Un resultado que podemos observar de la ecuación (2.6) es que $0 < s(\infty) < N$, y esto quiere decir que una parte de la población $N - s(\infty)$ evita ser infectada, luego aún quedan individuos susceptibles que están expuestos a la infección, y como consecuencia de este resultado, podemos entender que la epidemia desaparece por falta de individuos infectados y no por falta de susceptibles.

Gracias a la ecuación 2.6 obtenemos el número *de contactos* necesarios para infectar a un individuo susceptible, dicho número lo denotaremos por ρ , y en base al modelo de Kermack McKendrick (ver [4]) podemos definir $\rho = \frac{\gamma}{\beta}$ y así el número de reemplazos R en el instante $t = 0$ es el producto del número de contactos por el número de individuos susceptibles en el instante inicial, siendo este s_0 , luego $R = \frac{s_0}{\rho}$. En el instante inicial de la epidemia el número básico de reproducción es igual al número de reemplazos, ya que $R = \frac{s_0}{\rho} = \frac{\beta}{\gamma}s_0 = R_0$.

De forma general, podemos conocer el valor del número de reemplazos en cualquier instante de tiempo, luego tenemos que $R = \frac{s(t)}{\rho}$.

Pasaremos ahora a fijarnos en el número máximo de individuos infectados a lo largo de la epidemia al que llamaremos $i_{\text{máx}}$.

El número máximo de individuos infectados en un instante de tiempo es el número de infectados cuando se tienen las condiciones de que $i'(t) = 0$ e $i(t) \neq 0$. De la segunda ecuación del sistema de ecuaciones 2.2 tenemos que $i'(t) = (s(t) - \frac{\gamma}{\beta})i(t)\beta = 0$ si $s(t) = \frac{\gamma}{\beta}$, pues $i(t) > 0$. Si dicho valor de $s(t)$ lo sustituimos en la ecuación 2.5 tenemos que

$$i_{\text{máx}} = s_0 + i_0 - \frac{\gamma}{\beta} \log s_0 - \frac{\gamma}{\beta} + \frac{\gamma}{\beta} \log \frac{\gamma}{\beta},$$

que tras realizar los cálculos pertinentes, se reduce a la siguiente ecuación

$$i_{\text{máx}} = N + \frac{\gamma}{\beta} \log \left(\frac{\frac{\gamma}{\beta}}{s_0} \right) - \frac{\gamma}{\beta},$$

ya que inicialmente hemos considerado que el número de recuperados es cero, luego $N = s_0 + i_0$.

Para tenerminar el análisis cualitativo del modelo, utilizamos las ecuaciones primera y tercera de 2.2. Por tanto, para $s(t) \neq 0$ e $i(t) \neq 0$ tenemos que

$$\frac{s'(t)}{r'(t)} = -\frac{\beta s(t)i(t)}{\gamma i(t)} = -\frac{\beta s(t)}{\gamma},$$

si resolvemos esta ecuación por el método de separación de variables, como hemos hecho anteriormente, se tiene que $\log s(t) = -\frac{\beta}{\gamma}r(t) + cte$, tras despejar $s(t)$ tenemos

$$s(t) = s_0 e^{-\frac{\beta r(t)}{\gamma}}. \quad (2.7)$$

Como sabemos que $i(\infty) = 0$ y como se menciona en repetidas ocasiones, podemos calcular $r(t)$ conociendo los valores de $s(t)$ e $i(t)$, luego tenemos que $r(\infty) = N - s(\infty)$. Entonces $s(\infty)$ será igual a

$$s(\infty) = s_0 e^{-\frac{\beta r(\infty)}{\gamma}} = s_0 e^{-\frac{\beta(N-s(\infty))}{\gamma}}.$$

Por tanto, el número de individuos susceptibles que contraen la enfermedad será

$$i_{max} = i_0 + s_0 - s(\infty),$$

donde $s(\infty)$ es el resultado anterior.

Con el objetivo de conocer el número de personas recuperadas $r(t)$ por unidad de tiempo, es decir, $r'(t)$ podemos obtenerla a partir de la tercera ecuación de 2.2, de la ecuación 2.7 y de la relación $N = i(t) + s(t) + r(t)$, luego tenemos

$$r'(t) = \gamma i(t) = \gamma(N - s(t) - r(t)) = \gamma(N - r(t) - s_0 e^{-\frac{\beta r(t)}{\gamma}})$$

Así si resolvemos está ecuación, podemos obtener el número de personas recuperadas.

Una vez que hemos desarrollado el modelo determinista SIR, nos fijaremos en cómo de realistas es dicho modelo.

Los compartimentos de este modelo representan funciones cuyo conjunto de llegada son los números naturales más el elemento cero, debido a que los compartimentos están contituidos por individuos. Estas funciones pueden no ser continuas, luego no derivables y esta carencia conlleva un problema a la hora de demostrar la fiabilidad del modelo.

Otro punto a tener en cuenta es que la tasa de infección varía respecto del tiempo en la realidad, este parámetro no es constante; ya que es imposible que un individuo se infecte siempre con la misma probabilidad, es decir, el número de contacto con individuos infectados varía con respecto del tiempo.

Por último, hemos comentado antes que el número R_0 se interpreta como el número de individuos que es capaz de infectar un individuo infectado. Sin embargo, para calcular ese número medio debería haber una distribución de probabilidad asociada a ese número de individuos a los que infecta un contagiado. Esa distribución no existe de manera natural en los modelos deterministas descritos aquí.

Capítulo 3

Modelos estocásticos

Como conclusión del capítulo segundo, afirmamos que los modelos deterministas presentan limitaciones importantes para modelar el desarrollo de una epidemia. Para intentar solventar dichas limitaciones del modelo, introduciremos otros tipos de modelos, los modelos estocásticos. En estos modelos intervienen variables aleatorias que siguen una distribución de probabilidad, como por ejemplo, el número de individuos susceptibles, infectados y recuperados.

En los siguientes apartados definiremos modelos estocásticos que mejoran las limitaciones enunciadas, todos ellos los podemos encontrar en [3]

3.1. Proceso de ramificación en tiempo continuo

Un proceso de ramificación ayuda a describir la evolución de los individuos infectados en una epidemia con respecto del tiempo. La idea principal del proceso es considerar que cada individuo infectado actúa de forma independiente y siendo cada uno de los infectados idénticos entre sí, es decir, son variables aleatorias independientes e idénticamente distribuidas. Sean $I_n^{(j)}$ el número de infectados por el infectado j en la generación $n - 1$, que son variables aleatorias independientes e idénticamente distribuidas con valores en \mathbb{N} , siendo $I_n = (I_n^{(1)}, I_n^{(2)}, \dots)$. A partir de las variables $I_n^{(j)}$, podemos calcular las variables que representan a todos los individuos infectados en la generación $n + 1$, sabiendo que el inicio de la epidemia se produce por un único individuo en la generación 0, $X_0 = 1$, $X_{n+1} = \sum_{j=1}^{X_n} I_n^{(j)}$ para todo $n \in \mathbb{N}$. Una herramienta importante en el desarrollo de los procesos de ramificación son las funciones generadoras de probabilidades. La función generadora de probabilidades de la variable aleatoria $I_n^{(j)}$ se define como

$$g(z) = E \left[z^{I_n^{(j)}} \right] = \sum_{k=0}^{\infty} P(I_n^{(j)} = k) z^k,$$

para todo $n \in \mathbb{N}$, y la función generadora para la variable aleatoria X_n será

$$\psi_n(z) = E \left[z^{X_n} \right] = \sum_{k=0}^{\infty} P(X_n = k) z^k$$

Una propiedad importante que relaciona las esperanzas con la función generadora es la siguiente:

$$E \left[z^{X_{n+1}} | X_n = k \right] = E \left[z^{I_{n+1}^{(1)} + \dots + I_{n+1}^{(k)}} \right] = \left[E \left[z^{I_{n+1}^{(1)}} \right] \right]^k = g(z)^k.$$

Luego si no conocemos el valor que toma X_n tenemos $E \left[z^{X_{n+1}} | X_n \right] = g(z)^{X_n}$ y la función generadora será

$$\psi_{n+1}(z) = E \left[g(z)^{X_n} \right] = \psi_n(g(z)).$$

Ahora, iterando la ecuación anterior se obtiene el siguiente resultado:

$$\psi_{n+1}(z) = \psi_0(g^{(n+1)}(z)) \text{ donde } g^{(n+1)}(z) = g \circ \dots \circ g(z).$$

Luego como $X_0 = 1$ entonces se cumple que $\psi_0(z) = P(X_0 = 1)z^1 = z$ y tenemos que $\psi_n(z) = g^{(n)}(z)$.

El siguiente paso, será calcular la probabilidad de extinción de la epidemia. Se define el evento extinción como

$$[E] = \cup_{n=1}^{\infty} [X_n = 0].$$

y sabiendo que se verifica que $\{X_n = 0\} \subset \{X_{n+1} = 0\}$ y la propiedad de continuidad, su probabilidad $P(E)$, será

$$P(E) = P(\cup_{n=1}^{\infty} (X_n = 0)) = \lim_{n \rightarrow \infty} P(X_n = 0) = \lim_{n \rightarrow \infty} \psi_n(0).$$

Si denotamos $z = P(E)$ y como g es continua, entonces

$$g(z) = g(P(E)) = g(\lim_{n \rightarrow \infty} \psi_n(0)) = \lim_{n \rightarrow \infty} g(\psi_n(0)) = \lim_{n \rightarrow \infty} \psi_{n+1}(0) = z$$

luego $P(E)$ es un punto fijo de la función g . Gracias a estos resultados podemos describir las situaciones en las que la epidemia se extingue y en cuales no. Los momentos del proceso se pueden expresar en términos de la derivada de $g(z)$ en $z = 1$. Sea

$$g'(1) = \sum_{k=0}^{\infty} P(I_n^{(j)} = k) k = E[I_n^{(j)}] = m$$

el promedio de reproducción, tal que $m < \infty$. Dependiente del valor que tome el promedio de reproducción m , podemos clasificar el proceso de ramificación de la forma siguiente:

- Si $m \leq 1$ entonces $P(E) = 1$ y la epidemia llegará a la extinción de forma segura.
- Si $m > 1$, entonces $P(E) < 1$ y $P(E)$ es la única solución de la ecuación $z = g(z)$, $z < 1$. Además no podemos asegurar si la epidemia se extinguirá.

Por último, realizaremos una descripción de las principales propiedades que presenta g . Dicha función siempre es creciente y convexa en el intervalo $[0, 1]$. Si $P(I_n^{(j)} = 0) < 1$, g es estrictamente creciente en el intervalo unidad $[0, 1]$, si $P(I_n^{(j)} \leq 0) < 1$ entonces g es estrictamente convexa en $[0, 1]$. Como ya sabemos $g'(1) = E[I_n^{(j)}] = m$ y asumiento que se cumple la situación $P(I_m^{(j)} = 0) < 1$, tenemos que

- Si $m \leq 1$, la única solución de $g(z) = z$ en $[0, 1]$ es $z = 1$.
- Si $m > 1$, existen dos soluciones de la ecuación $g(z) = z$ en $[0, 1]$ y son $z_1 = 1$ y $z_2 \in (0, 1)$.

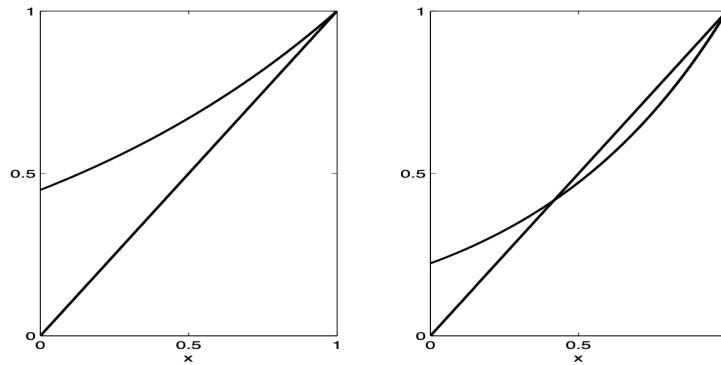


Figura 3.1: En la figura de la izquierda $m \leq 1$ y en la figura de la derecha $m > 1$. Esta representación ha sido extraída de [3].

3.2. El modelo Reed-Frost

El modelo Reed-Frost está definido en una situación en la que no existe el periodo de latencia y donde T es el periodo de infección, que es fijo y constante. Esto implica que los contactos entre individuos son independientes y la probabilidad de tener un contacto estrecho con otro individuo específico es $\pi = 1 - e^{-pcT/N}$, que es equivalente a pcT/N cuando T toma valores cercanos a cero. Sea pc la tasa de contacto entre individuos, que hace referencia al número de contactos

por la probabilidad de que un individuo susceptible sea infectado, y el ritmo de contagio es pc/N . Luego la expresión de $1 - \pi$ es la probabilidad de no producirse ningún contacto durante el periodo de infección T .

La versión con tiempo discreto del modelo de Reed-Frost es basicamente la misma que la versión con tiempo continuo; ya que no nos fijamos en el tiempo real si no en las generaciones de individuos infectados. Un individuo susceptible evita ser infectado en la generación $k + 1$ si evita ser infectado por todos los individuos infectados en la generación anterior k , este hecho sucede con probabilidad $(1 - \pi)^{i_k}$, donde i_k denota el número de individuos infectados en la generación k . La probabilidad de infectarse sería la probabilidad complementaria $1 - (1 - \pi)^{i_k}$. Como consecuencia, si hay i_k individuos que se infectan en la generación k y s_k individuos susceptibles, entonces este proceso sigue una distribución binomial definida como

$$I_{k+1} \sim B(s_k, 1 - (1 - \pi)^{i_k})$$

donde $S_{k+1} = s_k - I_{k+1}$.

Por tanto, la probabilidad de que el número de infectados en la generación $t + 1$ sea igual a i_{t+1} será:

$$P(I_{t+1} = i_{t+1} | S_t = s_t, I_t = i_t) = \binom{s_t}{i_{t+1}} (1 - (1 - \pi)^{i_t})^{i_{t+1}} [(1 - \pi)^{i_t}]^{s_t - i_{t+1}}$$

Podemos utilizar este método de manera iterativa para calcular la probabilidad de un brote en términos de generaciones. Luego si conocemos el número de infectados en cada generación $\{i_0, i_1, \dots, i_T\}$ siendo T el instante en el que finaliza la epidemia, la probabilidad de que el brote se desarrolle con dicho número de infectados es:

$$\begin{aligned} &P(I_1 = i_1 | S_0 = s_0, I_0 = i_0) P(I_2 = i_2 | S_1 = s_1, I_1 = i_1) \cdots P(I_T = i_T | S_{T-1} = s_{T-1}, I_{T-1} = i_{T-1}) \\ &= \prod_{t=0}^{T-1} \binom{s_t}{i_{t+1}} (1 - (1 - \pi)^{i_t})^{i_{t+1}} [(1 - \pi)^{i_t}]^{s_t - i_{t+1}} \end{aligned}$$

Por ejemplo, supongamos que tenemos la situación en la que el número de infectados y susceptibles ha ido variando de la siguiente manera: la generación inicial está constituida por $(S_0 = 3, I_0 = 1)$, la siguiente generación será $(S_1 = 1, I_1 = 2)$ y la última $(S_2 = 0, I_2 = 1)$. Entonces la probabilidad de que se

produzca este suceso será

$$\begin{aligned} & P(I_1 = 2 | S_0 = 3, I_0 = 1) P(I_2 = 1 | S_1 = 1, I_1 = 2) \\ &= \binom{3}{2} (1 - (1 - \pi)^1)^2 [(1 - \pi)^1]^{3-2} \cdot \binom{1}{1} (1 - (1 - \pi)^2)^1 [(1 - \pi)^2]^{1-1} \end{aligned}$$

3.3. El modelo estocástico en epidemias general (SIR)

El tercer modelo que vamos a desarrollar y que presenta un gran interés a lo largo del trabajo, es el modelo SIR estocástico. Este modelo no presenta periodo de latencia y está definido mediante procesos de Markov continuos que vienen dados en una escala de tiempo continua, $t \in [0, \infty]$.

Definición 3.1. Sea $X = \{X(t) : t \geq 0\}$ un proceso estocástico en tiempo continuo que toma valores en el espacio de estados E . Se dice que dicho proceso es una cadena de Markov en tiempo continuo si $\forall s, t \geq 0$ y $\forall i, j, x_u \in E$ con $1 \leq u \leq s$ se cumple que

$$P(X(t+s) = j | X(s) = i, X(u) = x_u) = P(X(t+s) = j | X(s) = i), \quad \forall t > 0.$$

En palabras menos formales, una cadena de Markov en tiempo continuo es un proceso estocástico que verifica que la probabilidad condicionada de un futuro estado en el instante $t+s$, dado el estado actual que se corresponde con el instante s y todos los estados pasados, depende únicamente del estado presente y es independiente del pasado.

Para simplificar la escritura de este tipo de probabilidades, denotaremos a la función de masa de probabilidad asociada a cada variable aleatoria $X(t)$ como:

$$p_i(t) = P(X(t) = i), \quad i \geq 0, \forall t \geq 0.$$

Supongamos que partimos de una población cerrada, es decir, el número total de individuos en nuestra población de estudio siempre es N y los estados por los que pueden experimentar dichos individuos son Susceptibles (S), Infectados (I) y Recuperados (R). En cada instante $t \geq 0$, $S(t), I(t)$ y $R(t)$ representan, respectivamente, la cantidad de individuos susceptibles, infectados y recuperados, que son variables aleatorias discretas tales que:

$$S(t), I(t), R(t) \in \{0, 1, \dots, N\}$$

con $N = S(t) + I(t) + R(t)$, ya que el tamaño de la población es cerrado.

En este modelo, los individuos que se recuperan se vuelven inmunes a la enfermedad, por lo tanto ya nunca vuelven a ser susceptibles ni pueden volver a padecer la enfermedad. En tales circunstancias, las restricciones que se deben verificar son las siguientes:

1. El contagio se produce de un individuo infectado a un individuo susceptible, luego en el momento inicial de la epidemia, por lo menos debe haber un individuo infectado de la población.
2. Un individuo infectado siempre se traslada al bloque de recuperados, ya sea porque se ha recuperado totalmente de la enfermedad o por su defunción.

La descripción de la evolución de la epidemia en la población se realiza mediante el uso de cadenas de Markov bidimensionales en tiempo continuo, ya que el número de recuperados se puede obtener directamente de $R(t) = N - S(t) - I(t)$, y tenemos que

$$X = \{(S(t), I(t)), t \geq 0\}$$

con espacio de estados

$$E = \{(s, i) : 0 \leq s + i \leq N, 0 \leq s \leq N, 1 \leq i \leq N\}$$

y cuyo cardinal es $|E| = \frac{(N+1)(N+2)}{2} - 1$, ya que el par $(N, 0)$ no sería posible debido a que debe haber al menos un paciente cero en la población y el par $(0, N)$ entraría dentro del conjunto E , ya que es posible que toda la población se contagie de la enfermedad, aunque no suele darse dicha situación con mucha frecuencia.

Existen dos posibles transiciones entre estados y son las siguientes:

1. Un individuo susceptibles es contagiado.
2. Un individuo infectado se recupera de la enfermedad.

Podemos expresar estos cambios suponiendo que el estado presente es (s, i) , es decir, el número de susceptibles en el instante t es s y el de infectados i , luego si se infecta un individuo susceptible el estado pasará a ser $(s - 1, i + 1)$ y si se recupera un infectado el estado será $(s, i - 1)$.

El tiempo que se tarda en producir un nuevo contagio es una variable aleatoria exponencial que depende de la composición de la población y de la tasa de contagio β . Los individuos se recuperan de manera independiente y el tiempo que transcurre en recuperarse cada individuo se distribuye mediante una variable

aleatoria exponencial de tasa γ . Por tanto, el tiempo que cada individuo permanece en cualquiera de los estados de la cadena es exponencial y la tasa de permanencia viene dada por:

$$q_{(s,i)} = \beta s \frac{i}{N} + \gamma i$$

si $(s, i) \in E$, donde cada uno de los términos recoge la tasa de cambio de estado a causa de:

- Nueva infección de un susceptible:

$$\beta s \frac{i}{N}$$

- Recuperación de un infectado:

$$\gamma i$$

El generador infinitesimal de la cadena de Markov en tiempo continuo es una matriz cuadrada, de dimensión $|E| \times |E|$ en la que se representan las tasas de cambio y podemos definirla de la siguiente manera: $A = [q_{(s,i),(s^*,i^*)}]$ siendo

$$q_{(s,i),(s^*,i^*)} = \begin{cases} \beta s \frac{i}{N}, & \text{si } (s, i) = (s - 1, i + 1), \\ \gamma i, & \text{si } (s, i) = (s, i - 1) \\ -q_{(s,i)}, & \text{si } (s, i) = (s, i), \\ 0, & \text{en el resto de los casos.} \end{cases}$$

Para describir cuál es la forma del generador infinitesimal A vamos a descomponer el espacio de estados E en niveles $E = \cup_{s=0}^N L(s)$ donde $L(s) = \{(s, i) : 0 \leq i \leq N - s\}$.

Para aclarar un poco mejor el funcionamiento del modelo, vamos a ver cuales son las transiciones entre los estados para $N = 4$. (Ver figura 3.2).

Como vemos en el diagrama, el conjunto de estados absorbentes, aquellos que al entrar en él no se puede salir del mismo, está formado por los estados $E_A = \{(0, i) : 0 \leq i \leq N\}$, que son los estados representados con un círculo negro. El conjunto de estados transitorios esta formado por $E_T = E - E_A$.

Mientras que $L(0)$ contiene todos los estados absorbentes, $\cup_{s=1}^N L(s)$ son los estados transitorios.

Si observamos el diagrama de transición de estados, podemos observar que los niveles $L(v)$ tiene dos estados, menos $L(0)$ que presenta un único estado.

Matricialmente el generador infinitesimal A se expresa como:

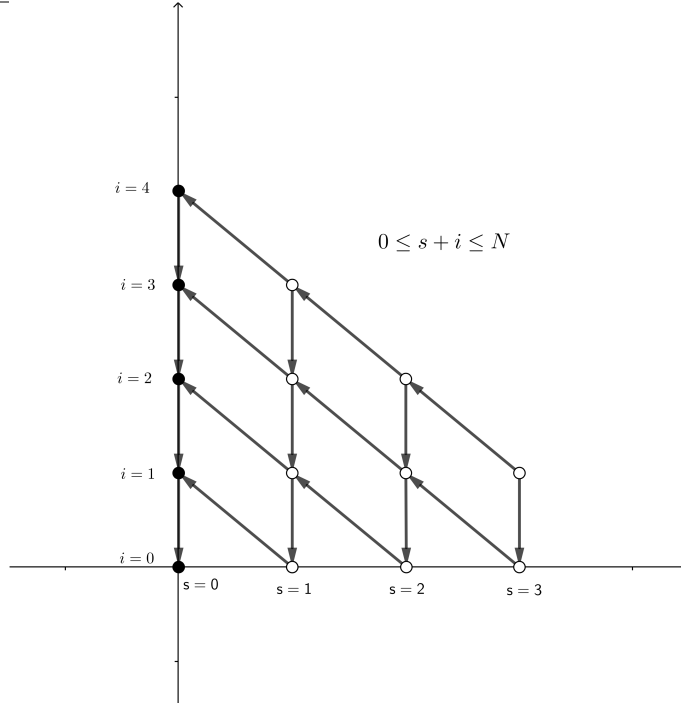


Figura 3.2: Diagrama de transición.

$$A = \begin{pmatrix} A_{0,0} & 0 & 0 & \cdots & 0 \\ 0 & A_{1,1} & 0 & \cdots & 0 \\ 0 & 0 & A_{2,2} & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & A_{N,N} \end{pmatrix}$$

donde, para $0 \leq k \leq N$,

$$A_{k,k} = \begin{pmatrix} -\lambda_{(k,0)} & \lambda_{(k,0)} & 0 & \cdots & 0 \\ \mu_{(1)} & -q_{(k,1)} & \lambda_{(k,1)} & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \mu_{(N)} & -q_{(k,N)} & \lambda_{(k,N)} \\ 0 & \cdots & \cdots & \mu_{(N)} & -q_{(k,N)} \end{pmatrix}$$

$A_{k,k}$ es una matriz de dimensión $(N + 1) \times (N + 1)$ que podemos representar mediante $Q_{k,k} = [q_{(i,j)}^k]$ para $0 \leq i, j \leq N$, donde $q_{(i,j)}^k$ tiene los siguientes valores no nulos:

$$q_{(i,j)}^k = \begin{cases} \mu_{(i)} & \text{si } j = i - 1 \\ -q_{(k,i)} & \text{si } j = i \\ \lambda_{(k,i)} & \text{si } j = i + 1. \end{cases}$$

siendo $\lambda_{(s,i)} = \frac{\beta}{N}si$, $\mu_{(i)} = \gamma i$.

Como el espacio de estados es finito, entonces satistafe sistema de ecuaciones diferenciales de Kolmogorov (ver página 338 de [10]) puede calcularse, sabiendo que $p_{(s,i)}(t) = P(S(t) = s, I(t) = i)$, es

$$\begin{aligned} \frac{dp_{(s,0)}(t)}{dt} &= -p_{(s,0)}(t)\lambda_{(s,0)} + p_{(s,1)}(t)\mu_{(1)}, \\ \frac{dp_{(s,i)}(t)}{dt} &= p_{(s-1,i+1)}(t)\lambda_{(s-1,i+1)} + p_{(s,i-1)}(t)\mu_{(i-1)} - p_{(s,i)}(t)q_{(s,i)}. \end{aligned}$$

para cada $0 \leq s \leq N$.

Es interesante aplicar las ecuaciones diferenciales de Kolmogorov al cálculo de la derivada de la esperanza del número de infectados. Para ello definiremos la esperanza del número de infectados y simplificaremos la derivada aplicando los cálculos de [13].

$$E[I(t)] = \sum_{s=0}^N ip_{(s)}(t) = \sum_{s=0}^N i \sum_{i=0}^{N+1} p_{(s,i)}(t) = \sum_{i=0}^{N+1} \sum_{s=0}^N iP(S(t) = s, I(t) = i)$$

luego

$$\begin{aligned} \frac{dE[I(t)]}{dt} &= \sum_{i=0}^{N+1} \sum_{s=0}^N \left[\frac{\beta i}{N} \left[p_{(s-1,i+1)}(t)(s-1)(i+1) - p_{(s,i)}(t)si \right] + i\gamma \left[p_{(s,i-1)}(t)(i-1) - p_{(s,i)}(t)i \right] \right] \\ &= \beta \sum_{i=0}^{N+1} \sum_{s=0}^N \frac{i}{N} \left[p_{(s-1,i+1)}(t)(s-1)(i+1) - p_{(s,i)}(t)si \right] \\ &\quad + \gamma \sum_{i=0}^{N+1} \sum_{s=0}^N i \left[p_{(s,i-1)}(t)(i-1) - p_{(s,i)}(t)i \right] \\ &= \beta E[S(t)I(t)] - \gamma E[I(t)] \end{aligned}$$

donde $E[S(t)I(t)] = E[S(t)]E[I(t)] + Cov(E[S(t)], E[I(t)])$.

3.4. Proceso de Hawkes

En esta sección introduciremos la definición de los procesos de Hawkes y describiremos algunas de sus principales propiedades.

Definición 3.2. Sea $N = \{N(t) : t \geq 0\}$ un proceso de conteo en el espacio probabilístico (Ω, \mathcal{F}, P) con una filtración $\mathbb{F} = \{(\mathcal{F}_t)_{t \in \mathbb{T}}\}$ que satisface

$$P(N(t+\delta) - N(t) = m | \mathcal{F}_t) = \begin{cases} 1 - \lambda^*(t)\delta + o(\delta), & m = 0 \\ \lambda^*(t)\delta + o(\delta), & m = 1 \\ o(\delta), & m > 1 \end{cases}$$

donde la función de intensidad condicionada $\lambda^*(t)$ es descrita por la ecuación:

$$\lambda^*(t) = \mu + \int_0^t g(t-s)dN(s)$$

siendo $\mu > 0$ una constante, denominada intensidad de fondo, y $g : [0, \infty) \rightarrow [0, \infty)$ una función denominada función de excitación.

La constante μ representa la existencia de una fuente constante de sucesos, por ejemplo, en nuestro caso, representa la inmigración que se está produciendo en la población, luego el número de individuos de la población no es contante.

Por otro lado, la función g satisface la condición de regularidad $\int_0^\infty g(s)ds < 1$. Esta condición me ayuda a garantizar una versión estacionaria del proceso, pero si dicha condición no se da entonces el proceso es explosivo. (**Proposition 6.4.VII** de [8])

Sean t_1, t_2, \dots los tiempos en que ocurren los eventos para cierta realización del proceso de Hawkes, podemos escribir la función intensidad del proceso como función de t y el conjunto de tiempos $\{t_i : t_i \leq t\}$. Luego la función de intensidad tiene la siguiente forma

$$\lambda^*(t) = \mu + \sum_{i:t_i < t} g(t-t_i).$$

La función de excitación más común es la de una exponencial decreciente en todo su dominio, por ejemplo, $g(t) = ae^{-bt}$ donde $a, b > 0$ y se utiliza esta función con el objetivo de simplificar las derivaciones que podamos realizar. La función de intensidad con $g(t)$ concreta tiene la siguiente forma

$$\lambda(t) = \mu + \int_0^t ae^{-b(t-s)}dN(s) = \mu + \sum_{i:t_i < t} ae^{-b(t-t_i)}.$$

3.5. Proceso de Hawkes modificado

En esta última sección del Capítulo 3 discutiremos una variante del proceso de Hawkes que añada flexibilidad para permitir una tasa de transmisión variable en el tiempo.

En esta modificación, que nos ayudará a calcular el número de reproducción instáneo $R(t)$, la constante μ del proceso de Hawkes tiene un valor nulo, luego la población se mantiene constante al eliminar la constante que media la inmigración. Además, en vez de depender g únicamente del tiempo desde que se ha producido un suceso τ , va a depender del tiempo calendario t , es decir, el tiempo desde que se produjo el primero de los sucesos, no únicamente el suceso de

estudio. Por tanto, la función de intensidad del proceso tendrá la forma

$$\lambda^*(t) = \int_0^t g(t, t-s) dN(s).$$

Denotaremos por $l(t)$ a la intensidad promedio del proceso N , es decir, $l(t) = E \left[\int_0^t g(t, t-s) dN(s) \right]$. Utilizando el teorema de descomposición de Doop-Meyer, anteriormente referenciado, podemos deducir que $dN(s) = dM(s) + d\Lambda(s)$, es decir, $dN(s) = dM(s) + \lambda(s)ds$. Sustituyo el valor de $dN(t)$ en $l(t)$ y podemos obtener que

$$\begin{aligned} l(t) &= E[\lambda^*(t)] = E \left[\int_0^t g(t, t-s) dM(s) \right] + E \left[\int_0^t g(t, t-s) \lambda(s) ds \right] \\ &= E \left[\int_0^t g(t, t-s) \lambda^*(s) ds \right] = \int_0^t g(t, t-s) l(s) ds \end{aligned}$$

sabiendo que la esperanza de una martingala es $E[dM(s)] = 0$. Si realizo un cambio de variable $\tau = t - s$ la igualdad anterior será igual a

$$l(t) = \int_0^t g(t, \tau) l(t - \tau) d\tau$$

y como l es nula para valores negativos del argumento, es decir, cuando $\tau > t$ entonces

$$l(t) = \int_0^\infty g(t, \tau) l(t - \tau) d\tau. \quad (3.1)$$

Luego llegamos a la ecuación (3.1), que es la ecuación que se describe en [6], y podemos ver que g depende de los tiempos t y τ , y nos ayudará a calcular el número de reproducción instantáneo. También podemos ver que si M (no es martingala) es un proceso de Poisson no homogéneo la esperanza de $N(b) - N(a)$ será

$$E[N(b) - N(a)] = \Lambda(b) - \Lambda(a) = \int_a^b \lambda^*(r) dr.$$

Tomando límites cuando $a \rightarrow 0$ y $b \rightarrow \infty$, el número medio de sucesos en este proceso será $\int_0^\infty \lambda^*(r) dr$. Con esta observación lo que vemos es que el número medio de individuos contagiados por el individuo infeccioso original es:

$$R(t) = \int_0^\infty g(t, \tau) d\tau.$$

Tenemos que $g(t, \tau)$ es la tasa de transmisibilidad, que podemos separarla de la siguiente forma (ver [6]) $g(t, \tau) = g_1(t)g_2(\tau)$, donde puede elegir que $g_2(\tau)$ sea una función de densidad tal que $\int_0^\infty g_2(\tau) d\tau = 1$. Por tanto, podemos verificar lo siguiente

$$R(t) = \int_0^\infty g(t, \tau) d\tau = \int_0^\infty g_1(t)g_2(\tau) d\tau = g_1(t),$$

es decir, la función $g_1(t)$ será igual al número de reproducción instantáneo $R(t)$. Además, la función $g_2(\tau)$ representa la distribución de cómo se distribuyen los sucesos (infecciones) en función del tiempo τ . Esta distribución es difícilmente estimable y en la práctica se suele aproximar por la distribución del intervalo de serie, que denoto por $\omega(\tau)$, donde el intervalo de serie es el intervalo de tiempo entre el inicio de los síntomas del individuo infeccioso hasta que presenta los síntomas el individuo que ha sido infectado, que a veces es conocido debido a estudios previos. Por lo tanto, la infectividad puede descomponerse como el producto del número de reproducción instantáneo y la distribución del intervalo de serie

$$g(t, \tau) = R(t)\omega(\tau).$$

Si volvemos a la esperanza de la función de intensidad $l(t)$ conociendo estos resultados tenemos que $l(t) = R(t) \int_0^\infty \omega(\tau)l(t - \tau)d\tau$ y si despejamos $R(t)$ tenemos que

$$R(t) = \frac{l(t)}{\int_0^\infty \omega(\tau)l(t - \tau)d\tau}.$$

Ahora observaremos las incidencias de casos en distintos instantes de tiempo que se registran de forma discreta, donde $I(t)$ son los casos registrados entre el instante t y $t + 1$, gracias a esto tenemos la siguiente aproximación

$$R(t) = \frac{l(t)}{\int_0^\infty \omega(\tau)l(t - \tau)d\tau} \approx \frac{I(t)}{\sum_{s=1}^t I(t - s)\omega(s)},$$

que es el estimador de $R(t)$ cuando conocemos el número de incidencia en cada instante t . Luego un individuo contagiado en el día 0, contagiará en el día t , es decir, $I(0)\omega(t)$.

Capítulo 4

Ilustración numérica de los resultados teóricos

El objetivo de este capítulo es observar de manera gráfica los resultados teóricos, expuestos a lo largo del trabajo, que se han obtenido a partir del Software R. Los datos que utilizaremos son aquellos asociados a la pandemia de Covid-19 en la comunidad autónoma de Castilla y León publicados por el Instituto de Salud Carlos III (<https://cneccovid.isciii.es/covid19/>) y a través de la librería EpiEstim hemos obtenidos las gráficas más relevantes en el estudio de una epidemia; gracias a esto podremos entender de manera más simple como se distribuye la enfermedad infecciosa en la población de estudio. Algunos de los procedimientos que he descrito en el capítulo 3, en concreto el proceso de Hawkes modificado, es la base del análisis de la evolución de la pandemia que ofrece cada día el Instituto de Salud Carlos III. En la figura adjunta (Figura 4.1) en el panel derecho se ofrece el número de reproducción instantáneo de la sección 3.5.

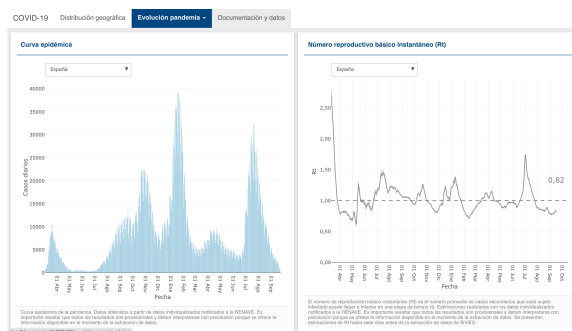


Figura 4.1: Web del Instituto de Salud Carlos III

El virus sobre el que hablamos es un tipo de coronavirus. Los coronavirus son una familia de virus que causan una infección en los seres humanos. El método

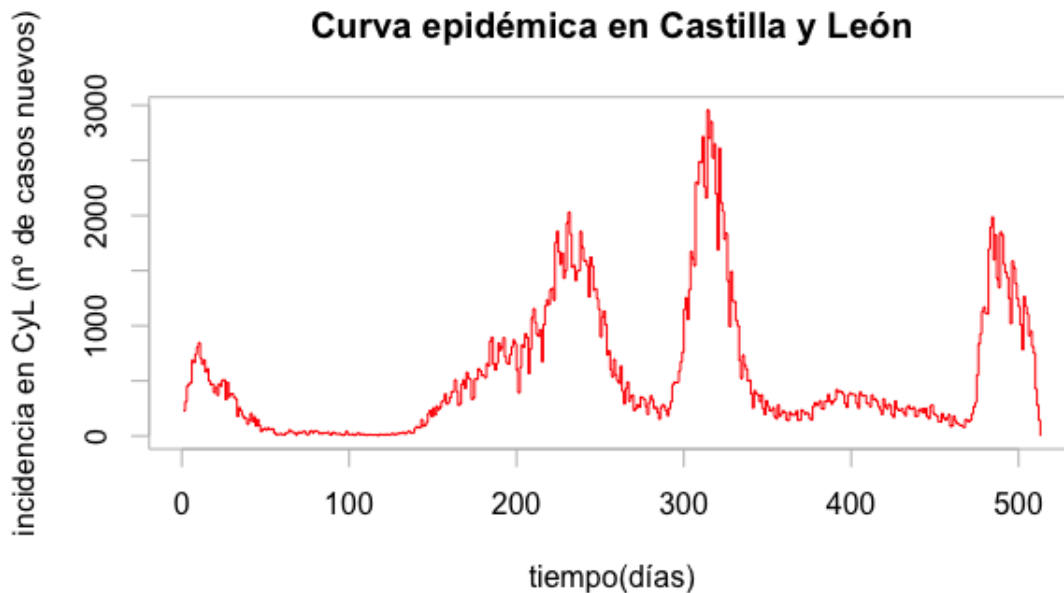
de transmisión se realizó de animales a humanos, por lo que se trata de un tipo de enfermedad zoonótica.

Para conocer el funcionamiento de la librería EpiEstim he utilizado como ayuda la guía [11]

4.1. Curva epidémica

Una curva epidémica es la representación gráfica del número de casos infectados según la fecha de aparición de la enfermedad en cada uno de los individuos de la población. La información que dicha curva nos aporta puede tener una gran utilidad a la hora de tomar medidas restrictivas que frenen la expansión de la enfermedad infecciosa. Confinamiento domiciliario, obligación de mascarillas en todo espacio ajeno al hogar y toque de queda, son algunas de las medidas que las autoridades sanitarias suelen tomar para frenar dicha curva.

Para empezar, mostraremos como es la curva epidémica del número de casos de Covid-19 en la comunidad de Castilla y León (11/03/2020 hasta 08/08/2021) y a partir de ella observaremos las conclusiones que nos aporta.



En la curva epidémica de Castilla y León podemos observar una serie de picos, concretamente cinco picos de diferentes intensidades, que hacen referencia a las cinco olas de Covid-19 que han atravesado los individuos que constituyen la comunidad de Castilla y León . Si tomamos como comienzo de la epidemia el día en el que se declaró el estado de alarma (11-03-2020), aunque la recolección de

datos no fuese tan precisa como en la actualidad, podemos observar el efecto del confinamiento domiciliario; ya que después de 68 días (17 de Mayo) se redujo la incidencia de casos a 7 en toda la comunidad. Después de un periodo de estabilidad de 71 días comenzó el segundo repunte de casos que hace referencia a la segunda ola, situándola temporalmente en el 27 de julio de ese mismo año. Trás un largo periodo de restricciones se consiguió disminuir la incidencia, pero el día 300 se inició la tercera ola que llevó la incidencia acumulada a máximos históricos. El día que se produjo el máximo del que hablamos fue el 18 de enero de 2021, llegando a una incidencia de casos de 2956 solamente en la comunidad de Castilla y León. Con respecto a las otras dos últimas oleadas de Covid aunque presenten una incidencia de alto riesgo, gracias a las vacunas la gravedad de la situación disminuyó.

La magnitud de la enfermedad fue de un nivel desorbitado, de tal forma que no pudo ser controlado de manera inmediata. Desde el día 11 de Marzo de 2020 hasta el 8 de Agosto de 2021 el número de infectados totales en Castilla y León es de 280996.

En ningún momento durante la epidemia se presentaron pacientes con casos aislados de Covid-19, ya que debido a la escasa protección y conocimiento del agente patógeno la dispersión de este fue descontrolada.

4.2. Distribución del intervalo de serie

Recordaremos que el intervalo de serie es el tiempo que transcurre desde el inicio de la enfermedad en el caso primario hasta el inicio de la enfermedad en el caso secundario. Asumiremos que el intervalo de serie está distribuido mediante una distribución Gamma, con media $\mu = 6.5$ y desviación típica $\sigma = 0.62$. Esta elección está justificada por estudios previos (ver [12]).

El resultado principal que podemos deducir de esta gráfica (Figura 4.2) es que los tiempos que pueden trancurren con mayor probabilidad entre la aparición de un individuo con Covid-19 en Castilla y León y el siguiente son 3 días, con la mayor probabilidad que es 0.185418, seguido de dos días y cuatro días con probabilidades respectivas de 0.1780743 y 0.1557344.

Tras el paso de los días, vemos que la probabilidad de que un individuo infeccioso contagie a individuos susceptibles va decreciendo, hasta que a partir del día 15, la probabilidad de contagio es insignificante.

Una aportación interesante que nos ofrece la distribución del intervalo de serie es la cuarentena que debe guardar cada individuo infectado para evitar la

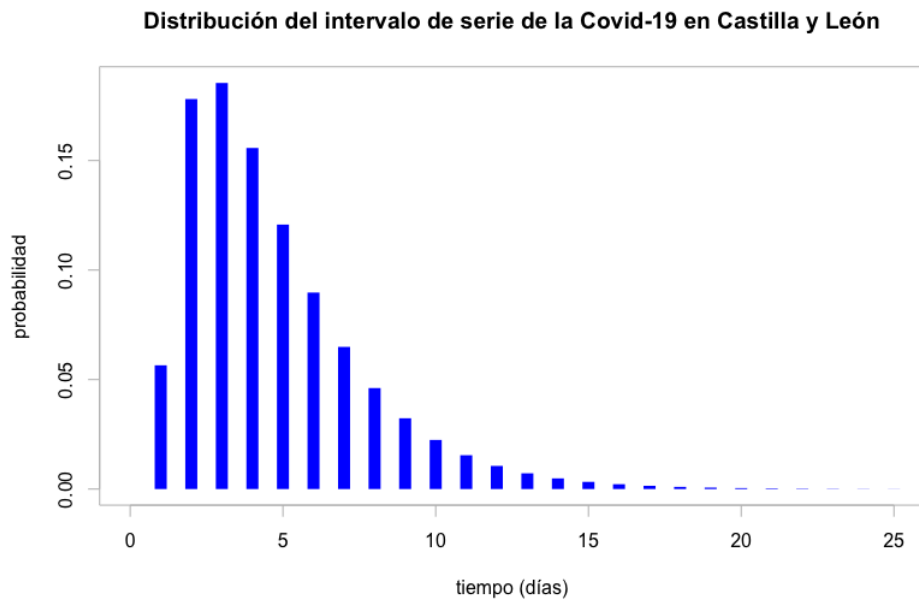


Figura 4.2: Distribución del intervalo de serie.

expansión. Cuando un individuo se contagia de Covid-19 las autoridades sanitarias obligaban a realizar un periodo de cuarentena de 10 días a todos los contactos estrechos con el individuo infectado. Al principio de la primera oleada, las cuarentenas de los individuos infectados no tenían un día preciso de fin, pero tras este tipo de estudios, se comprobó que no era necesario un periodo de cuarentena de 10 días, ya que a partir de ese día la probabilidad de contagio es mínima y no provocaría una expansión considerada de la enfermedad.

4.3. Estimación del número básico de reproducción $R(t)$

El parámetro más importante y que presenta un mayor uso en el ámbito de la Epidemiología es el número básico de reproducción $R(t)$. Como ya sabemos $R(t)$ es el número medio de infectados producidos por una persona infectada durante su periodo infeccioso en una población de individuos susceptibles a la enfermedad.

Como podemos ver en la Figura 4.3 existe una oscilación constante del valor del número $R(t)$, viendo que toma valores mayores o menores que uno. Fijándonos en la trayectoria que describe el número $R(t)$, podemos relacionarla con la curva epidémica, ya que cuando mayor es el número de infectados, mayor es $R(t)$. Como vemos en la Figura 4.3, en el inicio de dicha gráfica observamos el máximo

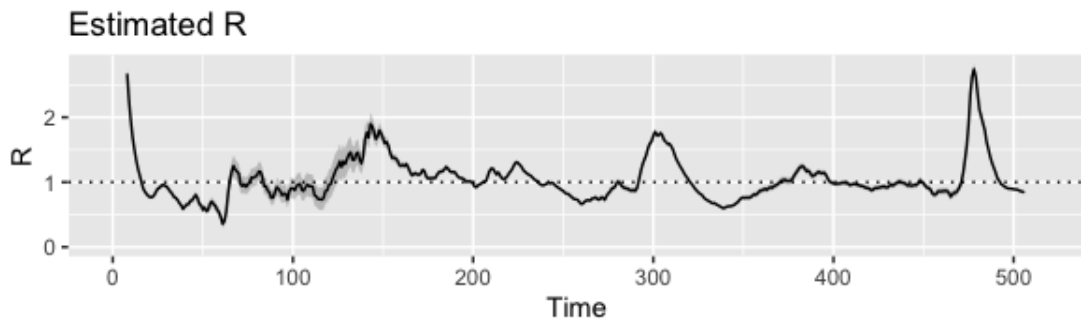


Figura 4.3: Estimación de $R(t)$ para la epidemia de Covid-19 en la comunidad de Castilla y León desde el día 11/03/2020 hasta el día 08/08/2021.

valor de $R(t)$, que hace referencia al momento en el que se produjo el confinamiento domiciliario de la población de estudio. En dicho momento la situación epidemiológica estaba descontrolada, pero tras la medida restrictiva impuesta se produjo un descenso considerado en el número $R(t)$, situándose durante un largo periodo por debajo de uno. Cuando $R(t)$ se sitúa por debajo de uno, la situación se vuelve controlada y la expansión del virus se erradicará con el paso del tiempo. Son necesarias las medidas de restricción del virus en el momento que $R(t)$ supera el valor uno, ya que si dichas medidas no se toman, la expansión del virus será incontrolable. Como última aclaración, tras el fin del estado de alarma las medidas se eliminaron y provocaron un gran aumento del número $R(t)$, pero esta vez la existencia de vacunas contra la Covid-19 fueron las causantes del descenso, situando el valor de $R(t)$ por debajo de uno de forma constante.

Además de esta gráfica, R nos proporciona un conjunto de datos formado por: los tiempo de inicio y fin de cada intervalo de tiempo considerado; los cuantiles 0.025, 0.05, 0.25, 0.5, 0.75, 0.95 del número de reproducción asociado a cada intervalo de tiempo y la distribución discreta del intervalo de serie. Mostraremos los diez primeros valores que toman cada variable.

\$R						
	t_start	t_end	Mean(R)	Std(R)	Quantile.0.025(R)	Quantile.0.05(R)
1	2	8	2.6726983	0.04328859	2.5885202	2.6018957
2	3	9	2.2617294	0.03446301	2.1946819	2.2053429
3	4	10	1.9549404	0.02850664	1.8994631	1.9082886
4	5	11	1.6967015	0.02415494	1.6496853	1.6571666
5	6	12	1.5081791	0.02111252	1.4670799	1.4736209
6	7	13	1.3410557	0.01877118	1.3045144	1.3103300
7	8	14	1.2064076	0.01703055	1.1732566	1.1785321
8	9	15	1.1039565	0.01579177	1.0732197	1.0781104
9	10	16	0.9946309	0.01469702	0.9660315	0.9705806
10	11	17	0.8948711	0.01380981	0.8680068	0.8722778

	Quantile.0.25(R)	Median(R)	Quantile.0.75(R)	Quantile.0.95(R)
1	2.6433746	2.6724646	2.7017672	2.7442980
2	2.2383901	2.2615544	2.2848780	2.3187130
3	1.9356381	1.9548018	1.9740915	2.0020647
4	1.6803474	1.6965869	1.7129307	1.7366274
5	1.4938857	1.5080806	1.5223651	1.5430733
6	1.3283474	1.3409681	1.3536685	1.3720802
7	1.1948774	1.2063274	1.2178504	1.2345563
8	1.0932645	1.1038812	1.1145665	1.1300594
9	0.9846789	0.9945585	1.0045041	1.0189282
10	0.8855182	0.8948001	0.9041466	0.9177068

\$si_distr						
	t0	t1	t2	t3	t4	t5
0.000000e+00	5.650079e-02	1.780743e-01	1.854180e-01	1.557344e-01	1.207514e-01	
	t6	t7	t8	t9	t10	
8.971764e-02	6.490842e-02	4.611124e-02	3.232565e-02	2.243368e-02		

4.4. Infectividad global $\lambda^*(t)$

La infectividad es la capacidad de un agente infeccioso de invadir un organismo y producir una infección. Una expresión que nos ayuda a calcular la infectividad global $\lambda^*(t)$ está constituida por la suma de los individuos infectados con anterioridad ponderados por su respectiva infectividad en el instante t (esta ponderación hace referencia a la distribución discreta del intervalo de serie $\omega(s)$); todo esto podemos resumirlo de la siguiente forma

$$\lambda^*(t) = \sum_{s=1}^t I(t-s)\omega(s)$$

La representación de la infectividad utilizando los datos de Covid-19 en la comunidad de Castilla y León viene reflejada en la siguiente figura:

Si observamos la figura anterior y revisamos los valores de $\lambda(t)$ proporcionados

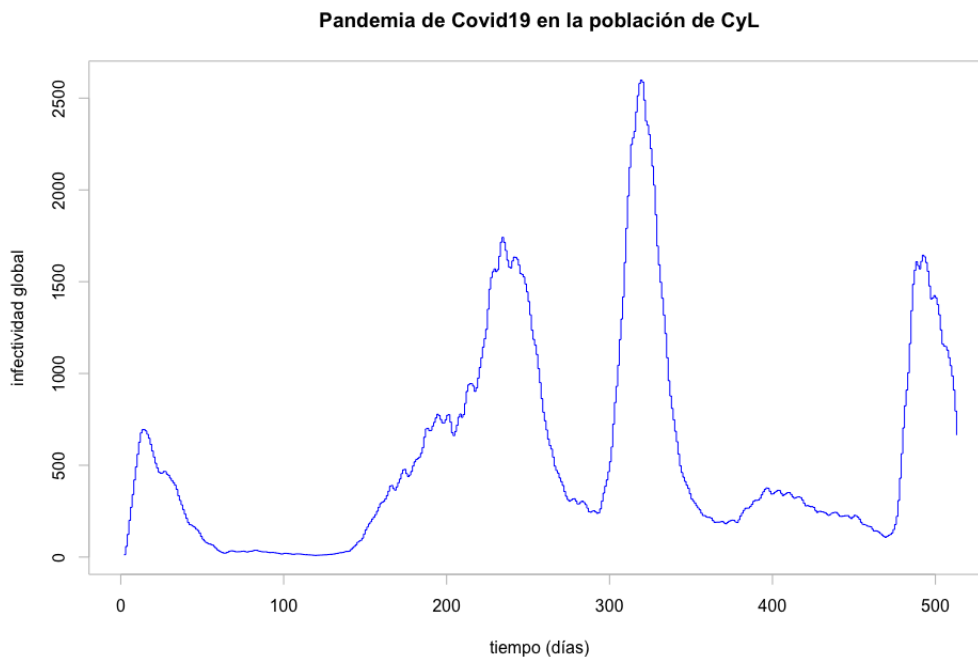


Figura 4.4: Infectividad de la epidemia de Covid-19 en la comunidad de Castilla y León.

por el software R, podemos destacar que el día en el que la epidemia presenta una mayor capacidad para producir nuevas infecciones es el día 306 con una tasa de 1741.297504 posibles nuevas infecciones. Un dato que cabe destacar es la baja infectividad a lo largo de la cuarta ola, durante el transcurso de esta ola se logró un destacado control debido a las restricciones existentes y el pico más alto solo fue de 376.460474 que se alcanzó el día 396 tras el comienzo de la epidemia. La infectividad comparada con la máxima alcanzada durante la tercera ola presenta una gran diferencia y esto es gracias a la aplicación de restricciones que durante el periodo de Navidad se eliminaron dando lugar a la trágica tercera ola.

A partir de la quinta ola, donde la infectividad vuelve a ser alta 1635.580785 el día 493, podemos concluir aunque no con total probabilidad que la epidemia de Covid-19 estaría llegando a su fin; gracias a la efectividad de las vacunas y a la reducción de la expansión de la Covid-19.

Por último, si comparamos la gráfica de la curva epidémica con la figura que estamos describiendo, podemos observar la enorme semejanza que presentan; y esto nos ayuda a concluir que el modelo matemático utilizado es una descripción razonable de la realidad.

Bibliografía

- [1] BILLINGSLEY, P. (1995), *Probability and Measure. 3a ed.*, Wiley. New York.
- [2] ANDERSEN P.K., BORGAN O., GILL R. D., KEIDING N.(1993) *Statistical Models Based on Counting Processes*, Springer Series in Statistics.
- [3] DIEKMANN O., HEESTERBEEK H., BRITTON T.(2013) *Mathematical Tools for Understanding Infectious Disease Dynamics*, Princeton Series in Theoretical and Computational Biology.
- [4] KERMACK W.O. AND A.G.MCKENDRICK. (1927) *Contributions to the mathematical theory of epidemics*. Proceedings of the Royal Society,115A:700-721.
- [5] MURRAY J.D.(1993) *Mathematical Biology I: An introduction*. Springer-Verlag , Third Edition
- [6] FRASER C.(2020) *Estimating Individual and Household Reproduction Numbers in an Emerging Epidemic*. PLoS ONE 2(8): e758.
- [7] FLAXMAN S., T UNWI H. J., COUPLAND HELEN.(2020) *Report 13: Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries*.Nature 584, 257–261 .
- [8] DALEY S. J. , VERE-JONES D.(1998) *An introduction to the theory of point processes*. Springer-Verlag , Second Edition.
- [9] COURRÈGE, P. AND PRIOURET, P. (1965) *Temps d'arrêt d'un fonction aléatoire* Publ. Inst. Stat. Univ. Paris 14,245-274.
- [10] BREMAUD P. (1999) *Markov Chains Gibbs Fields, Monte Carlo Simulation, and Queues*. New York, NY : Springer New York : Imprint: Springer
- [11] GÓMEZ MALDONADO L.(2012) *Modelos epidemiológicos en R: El libro EpiEstim*. Universidad de Granada.

-
- [12] FLAXMAN S., MISHRA S., GANDY A. (2020) *Report 13: Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries*. Imperial College COVID-19 Response Team
- [13] LLOYD, A. L. (2004) *Estimating variability in models for recurrent epidemics: assessing the use of moment closure techniques*. *Theor. Popul. Biol.*, 65, 49–65