



FACULTAD DE ESTUDIOS ESTADÍSTICOS

MASTER EN BIOESTADÍSTICA

CURSO: 2016/2017

TRABAJO FINAL DE MASTER

**ESTUDIO EN TIEMPO
DISCRETO DE LA EXPANSIÓN
DE UNA EPIDEMIA**

Alumno: María Gamboa Pérez

**Tutores: Julia Amador Pacheco y
María Jesús López Herrero**



Agradecimientos:

Quiero dar las gracias a mis padres y a Vicen por apoyarme siempre en todo y en especial a mis directoras de trabajo por su ayuda; sin ellas no hubiera sido posible.

Índice general

Índice de figuras	IV
Índice de tablas	V
Resumen y objetivos	1
1. Introducción	2
1.1 Preliminares	2
1.2 Marco teórico	5
2. Modelización matemática de epidemias	16
2.1 Enfoque determinista	18
2.1.1 Medidas de transmisión de una enfermedad contagiosa	20
2.2 Enfoque estocástico	23
2.2.1 Medidas de transmisión en modelos markovianos . .	24
2.2.2 Comportamiento de la epidemia en su fase activa . .	31
3. Estudio en tiempo discreto de una epidemia	33
4. Ilustración numérica de los resultados teórico algorítmicos obtenidos	47
5. Conclusiones	58
6. Anexos	60
7. Bibliografía	65

Índice de figuras

1.	Mortalidad durante la peste de Bombay	5
2.	Modelo SIR	17
3.	Modelo SIS.	18
4.	Principales R_0 de enfermedades contagiosas actuales	21
5.	Simulación de las primeras 100 etapas de la epidemia. . . .	35
6.	Simulación hasta la erradicación	36
7.	Diagrama de transición	47
8.	Funciones de distribución para $N = 5$	49
9.	Función de masa para $N = 50$	49
10.	Gráficos caja para $\gamma = 2$	51
11.	Gráficos caja para $\gamma = 5$	52
12.	Valores esperados m_1 variando N , $\alpha = 1,6$ y $\gamma = 0,8$. . .	54
13.	Valores esperados m_i con $N = 20$, $\alpha = 1,6$ y $\gamma = 0,8$. . .	55
14.	Funciones de masa al variar Δt	57

Índice de tablas

1.	Probabilidad de que la epidemia finalice antes de la observación n	50
2.	Número medio de etapas hasta la extinción	53
3.	Amplitud y número esperado de etapas al variar N y α	56

Resumen y objetivos

A lo largo de toda la historia, la Humanidad se ha visto afectada por grandes epidemias como la peste, la viruela o el sida. Estas epidemias se han estudiado matemáticamente dando lugar a diversos modelos epidemiológicos. Estos modelos tienen dos posibles enfoques, el determinista que se modeliza mediante sistemas de ecuaciones diferenciales y el enfoque estocástico que nosotros lo modelizaremos mediante cadenas de Markov y cuyo modelo subyacente es el modelo SIS estocástico. En ambos casos disponemos de ciertos indicadores los cuales nos permitirán cuantificar la epidemia. Estos son el número reproductivo básico, el número reproductivo exacto y el número reproductivo poblacional.

El objetivo principal de este trabajo es el estudio en tiempo discreto de una epidemia en poblaciones de tamaño moderado sin inmunidad y aunque la mayor parte de los estudios son en tiempo continuo estos no se ajustan bien a los datos disponibles en la vida real porque se suelen tener datos medidos sobre intervalos de tiempo como por ejemplo días o meses y no en cualquier instante (inspecciones).

La metodología usada para el estudio del modelo epidemiológico, objeto de este trabajo, es la que concierne a las cadenas de Markov en tiempo discreto, en concreto a los procesos de nacimiento y muerte. Al modelizarse mediante cadenas de Markov se garantiza que todos los brotes epidémicos finalicen en un tiempo esperado finito y por ello se estudia el número de etapas hasta la finalización de la epidemia. Para su análisis hemos usado ecuaciones en diferencias finitas, método de la primera transición. Su posterior análisis numérico para evaluar la influencia de los parámetros del modelo se ha realizado con el software *R*.

Palabras clave: *epidemiología, modelo SIS, cadenas de Markov, número reproductivo básico, tiempo hasta la extinción.*

1. Introducción

En este capítulo hablaremos sobre el origen de la epidemiología y las principales figuras que iniciaron el desarrollo gracias a la aplicación de modelos matemáticos. La situaremos históricamente y expondremos algunas de las enfermedades contagiosas más importantes de la historia. También describiremos y expondremos las técnicas más importantes en las que se apoya.

1.1 Preliminares

Las epidemias han sido uno de los grandes temores del ser humano ya que acabaron con la vida de numerosas personas a lo largo de la historia. Nos encontramos con epidemias tan importantes como la peste bubónica, que asoló la población de la Edad Media, cobrándose como víctimas a un tercio de la población, el cólera o la sífilis, que afectaron al 15 % de la población europea en el siglo XX.

El número de víctimas cobradas por estas enfermedades era tan elevado que provocaba que hubiera grandes cambios demográficos. Por ello fue de vital importancia estudiar y sobre todo modelizar el comportamiento de estas epidemias, lo que dio lugar a la aparición de la epidemiología.

La epidemiología estudia la evolución de una epidemia a lo largo del tiempo, con el fin de poder predecir su comportamiento, explicar su dinámica y así poder desarrollar planes de prevención. La palabra epidemia hace referencia a la aparición de una enfermedad, por lo general de tipo infeccioso, que afecta simultáneamente a un gran número de personas de una misma región.

El origen del estudio de las enfermedades infecciosas se remonta a la Prehistoria. El Papiro de Ebers, es quizás el documento más antiguo en el que se habla sobre enfermedades de tipo infeccioso. En él se mencionan unas fiebres pestilentes, probablemente de malaria, que asolaron a la población cercana del Nilo alrededor del año 2000 a. C., quizás sea el texto más antiguo en el que se hace referencia a un padecimiento colectivo.

La aparición de plagas a lo largo de la historia también fue registrada en la mayor parte de los libros sagrados en especial en la Biblia, el Talmud y el Corán, en los

cuales podemos encontrar además los primeros planes de prevención de enfermedades contagiosas [1].

La palabra epidemiología que proviene de los términos griegos “epi” (sobre), “demo” (pueblo) y “logos” (estudio), etimológicamente significa “el estudio sobre las poblaciones”.

Tras el Papiro de Ebers, la primera referencia propiamente médica se debe a Hipócrates (460-385 a. C.) quien usó las expresiones epidémico y endémico para referirse a los padecimientos según fueran o no propios de un determinado lugar. Su postura profundamente racionalista sobre el desarrollo de las enfermedades (ninguno de sus trabajos menciona curas sobrenaturales) y sus afirmaciones sobre la influencia del modo de vida y el ambiente en la salud de la población, hacen de este médico el principal representante de la epidemiología antigua [2].

En 1580, el médico francés Guillaume de Baillou (1538-1616) publicó el libro *Epidemiorum II* (“Sobre las epidemias”) que contiene una relación completa de las epidemias de sarampión, difteria y peste bubónica ocurridas en Europa entre 1570 y 1579, sus características y modo de propagación.

En castellano, la primera referencia al término epidemiología se la debemos a Quinto Tiberio Angelerio, en Madrid en 1598.

Según señala J. Ruiz [3], el proceso matemático que condujo a la elaboración de leyes de la enfermedad se inició con el análisis de la distribución de los nacimientos con J. Graunt y W. Petty.

En 1710 J. Arbuthnot, continuador de los trabajos de Graunt y Petty, había demostrado que la razón de nacimientos entre varones y mujeres era siempre de 13 a 12, independientemente de la sociedad y país en el que se estudiaran. Para Arbuthnot, esta regularidad no podía deberse al azar y tenía que ser una disposición divina, encaminada a equilibrar el exceso de muertes masculinas debidas a la violencia y la guerra. Entre 1741 y 1775, el sacerdote alemán J.P. Süßmilch, escribió varios tratados que seguían los métodos de enumeración propuestos por Graunt, Petty y Arbuthnot. Para Süßmilch, la regularidad encontrada en el volumen de nacimientos por sexo era toda una “ley estadística” (como las leyes naturales de la física), y debían existir leyes similares capaces de explicar el desarrollo de toda la sociedad.

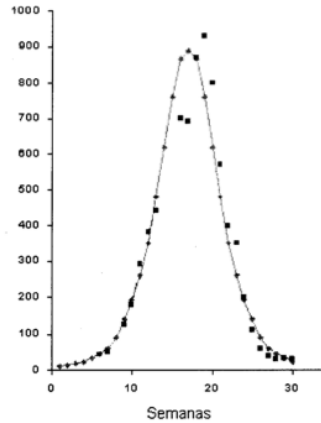
Pronto nació la idea de una “ley de mortalidad” y poco más tarde la convicción de que habría leyes para todas las desviaciones sociales: el suicidio, el crimen, la vagancia, la locura y naturalmente la enfermedad [4]. Como podemos ver, las matemáticas han sido de gran importancia para el desarrollo de la epidemiología. Uno de los primeros en desarrollar un modelo matemático fue el gran Daniel Bernouilli con un estudio sobre la técnica de variolación, en relación a la viruela. Aunque la Academia de Ciencias de Paris publicó su trabajo en 1760, el método nunca fue adoptado de forma oficial y el propio Rey Luis XV de Francia moriría de viruela años después, en 1774.

Otro precursor relevante fue W. Farr, quien en 1840 modelizó la evolución de una enfermedad que tenía el ganado vacuno y que en nuestros días ha sido aplicado a la evolución del sida. Solo a principios del siglo XX resurgió con vigor la idea de modelizar matemáticamente la propagación de epidemias.

Tal y como cuenta J. Velasco [5], en el siglo pasado es cuando realmente aparece el interés por aplicar métodos cuantitativos a la biología. El motivo es que su uso en la física había resultado tan satisfactorio que decidieron aplicarlo a otros campos. Rápidamente se observó que las enfermedades infecciosas se transmiten por contacto entre un individuo susceptible y otro infectado. De esta idea W. Hamer en 1906 formuló la ley de acción de masas, donde establece que el número de contactos infecciosos por unidad de tiempo, es proporcional al número total de contactos entre individuos infecciosos y sanos.

Más tarde, en 1927, Kermack y McKendrick formularían el que es considerado el primer modelo general matemático en epidemiología para describir la epidemia de la peste en India de 1906. En la siguiente figura podemos ver la gráfica de mortalidad durante la epidemia de peste que se produjo en Bombay entre los años 1905 y 1906. Los puntos describen los casos observados y la línea continua las predicciones de Kermack y McKendrick, basadas en su modelo publicado.

Figura 1: Mortalidad durante la peste de Bombay.



Las técnicas estadísticas que más han influenciado a la biomedicina en los últimos años han sido los modelos lineales generalizados (incluida la regresión lineal múltiple), análisis de supervivencia, análisis de datos categóricos, estadística espacial y los métodos bayesianos. Los metaanálisis también han sido de gran interés en los últimos años. En 1994 surgieron nuevos métodos estadísticos que tendrían un rol importante, como son el método Bootstrap, los métodos bayesianos, los árboles de clasificación y regresión, los modelos para datos longitudinales GEE (general estimating equations), los modelos para datos jerárquicos y las redes neuronales.

1.2 Marco teórico

Las cadenas de Markov son probablemente uno de los procesos más estudiados en todo el ámbito de la epidemiología. Un libro clásico donde se puede encontrar un amplio desarrollo de esta teoría es el libro de Kenelly y Snell [6]; otras buenas referencias son los libros de Karlin y Taylor [7], o el de Schinazi [8].

A continuación describiremos las principales características y resultados de las cadenas de Markov en tiempo discreto.

Un proceso estocástico es una colección de variables aleatorias $\{X_t : t \in T\}$ parametrizada por un conjunto T , llamado espacio paramétrico, en donde las variables toman valores en un conjunto S llamado espacio de estados.

Si el proceso se controla u observa en instantes de tiempo concretos (días, cada vez que llueve...) se dice que es un proceso estocástico en tiempo discreto y se denotará por $\{X_n : n = 0, 1, 2, \dots\}$. En adelante diremos que X_n es la observación del fenómeno en estudio en la etapa n .

Un proceso estocástico es llamado cadena cuando el espacio de estados es discreto. El nombre de Markov es gracias a A.A. Markov, un matemático Ruso (1856-1922), quien contribuyó profundamente al desarrollo de los procesos estocásticos. Asociada a cada variable aleatoria X_n , denotamos su función de masa por la siguiente expresión:

$$p_i(n) = P\{X_n = i\}, i \geq 0, \forall n \geq 0. \quad (1)$$

En adelante la función de masa de la etapa 0 se llamará distribución inicial de la cadena y se denotará por $a_0 = (a_1^0, a_2^0, \dots)$.

Definición 1. *Un proceso estocástico discreto $\{X_n : n \geq 0\}$ se dice que tiene la propiedad Markoviana si $P\{X_n = i_n | X_0 = i_0, \dots, X_{n-1} = i_{n-1}\} = P\{X_n = i_n | X_{n-1} = i_{n-1}\}$, donde los valores de $i_k \in \{1, 2, \dots\}$, para $k = 0, 1, 2, \dots, n$.*

En este caso el proceso estocástico se llama *cadena de Markov* o más específicamente *cadena de Markov en tiempo discreto*.

Las probabilidad de transición denotada por $p_{ij}(n)$, se define como la probabilidad condicionada siguiente:

$$p_{ij}(n) = P\{X_{n+1} = j | X_n = i\}$$

es decir, la probabilidad de que el proceso se encuentre en el estado j en el instante $n + 1$, sabiendo que en el instante n , estaba en i , para $i, j = 1, 2, \dots$.

Si las probabilidades de transición $p_{ij}(n)$ de una cadena de Markov no dependen del instante de tiempo n , se dice que es homogénea. En lo que sigue utilizaremos cadenas homogéneas y por lo tanto eliminaremos 'n' de la notación.

Proposición 1. *En cualquier estado i del espacio de estados se verifica que*

$$p_{ij} \geq 0,$$

$$\sum_{j=1}^{\infty} p_{ij} = 1; \forall i, j = 1, 2, \dots$$

Lo que significa que, con probabilidad 1, el proceso en cualquier estado i en el instante n , se mueve a otro estado $j \neq i$, o permanece en el estado i , en la siguiente etapa $n+1$. Si fijamos i , $\{p_{ij}\}$ es una función de probabilidad. Las probabilidades de transición en una etapa se pueden expresar en forma matricial, a la que llamaremos matriz de transición P .

Definición 2. *La matriz de transición de una cadena de Markov en tiempo discreto $\{X_n\}_{n=0}^{\infty}$ con espacio de estados $S = \{1, 2, \dots\}$ y probabilidades de transición en una etapa $\{p_{ij}\}_{i,j=1}^{\infty}$, se denota por $P = (p_{ij})$ y es*

$$P = \begin{pmatrix} p_{11} & p_{12} & p_{13} & \dots \\ p_{21} & p_{22} & p_{23} & \dots \\ p_{31} & p_{32} & p_{33} & \dots \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix}.$$

Las probabilidades de transición denotadas por $p_{ij}^{(n)}$ son las probabilidades de ir del estado i al estado j en n etapas.

$$p_{ij}^{(n)} = P\{X_n = j | X_0 = i\}$$

y su matriz de transición en n etapas se denota por $P^{(n)}$.

Observamos que:

$$p_{ij}^{(0)} = \begin{cases} 1, & \text{si, } i = j, \\ 0, & \text{si, } j \neq i. \end{cases}$$

Teorema 1. *La relación entre las probabilidades de transición en n y $(n-s)$ pasos son las conocidas como Ecuaciones de Chapman-Kolmogorov y son las siguientes:*

$$p_{ij}^{(n)} = \sum_{k=1}^{\infty} p_{ik}^{(s)} p_{kj}^{(n-s)}, \quad 0 < s < n.$$

La demostración de las ecuaciones de Chapman-Kolmogorov las podemos encontrar en [9].

Teorema 2. *La matriz de transición en n etapas verifica:*

$$P^{(1)} = P$$

$$P^{(2)} = P^2$$

...

En general,

$$P^{(n)} = P^n.$$

Denotamos $p(n)$ el vector de probabilidades de la función de masa asociada a la variable X_n

$$p(n) = (p_1(n), p_2(n), \dots)^T.$$

$$p(n+m) = a_0 P^{n+m} = P^n(a_0 P^m) = p(m)P^n.$$

A la hora de estudiar el comportamiento a largo plazo de las cadenas de Markov es necesario conocer y clasificar sus estados.

Definición 3. *Se dice que el estado j es accesible desde el estado i si la probabilidad de ir desde i a j es mayor que 0, para alguna etapa, es decir :*

$$\exists n \geq 0, \text{ tal que } p_{ij}^{(n)} > 0.$$

La anterior relación entre los estados es una relación de equivalencia e induce una partición en el espacio de estados. A las clases de equivalencia las llamaremos clases de comunicación, o simplemente clases.

Definición 4. *Si existe una sola clase, entonces se dice que la cadena de Markov es irreducible; en caso contrario se dice que es reducible.*

Definición 5. *Se dice que un conjunto de estados de una cadena de Markov es cerrado si es imposible acceder a cualquiera de sus estados, desde un estado que no pertenezca a él.*

Asumiendo que el proceso empieza en un estado inicial $X_0 = i$, definiremos las probabilidades de retornar al estado i , y la probabilidad de pasar por primera vez al estado j .

Definición 6. Sea $f_{ii}^{(n)}$ la probabilidad de que empezando desde el estado i retorne por primera vez al estado i , en la etapa n , $n > 0$; es decir: $f_{ii}^{(n)} = P\{X_n = i, X_m \neq i, m = 1, 2, \dots, n-1 | X_0 = i\}$.

Estas probabilidades son conocidas como las probabilidades de primer retorno. Definimos $f_{ii}^{(0)} = 1$.

Como podemos observar las probabilidades $f_{ii}^{(1)} = p_{ii}$ pero en general $f_{ii}^{(n)}$ no es igual a $p_{ii}^{(n)}$. Como vemos estas probabilidades representan la primera vez que la cadena retorna al estado i , verificando:

$$0 \leq \sum_{n=1}^{\infty} f_{ii}^{(n)} \leq 1.$$

El valor del sumatorio permite distinguir si un estado es transitorio, o no.

Definición 7. El estado i se dice que es transitorio si $\sum_{n=1}^{\infty} f_{ii}^{(n)} < 1$. El estado i se dice que es recurrente si $\sum_{n=1}^{\infty} f_{ii}^{(n)} = 1$.

Si el estado i es recurrente, entonces el conjunto $\{f_{ii}^{(n)}\}_{n=0}^{\infty}$ define una distribución de probabilidad para la variable aleatoria que representa el tiempo de primer retorno, es decir de la variable:

$$T_{ii} = \inf_{m \geq 1} \{m | X_m = i, X_0 = i\}.$$

Por lo tanto $P\{T_{ii} = n\} = f_{ii}^{(n)}$, $n = 0, 1, 2, \dots$.

A partir de la distribución de T_{ii} podremos obtener su media, la cual indica el tiempo medio de recurrencia, o el tiempo medio de regresar al estado i por primera vez, y es denotado por

$\mu_{ii} = E(T_{ii})$. Para un estado recurrente i :

$$\mu_{ii} = \sum_{n=1}^{\infty} n f_{ii}^{(n)}.$$

Aunque T_{ii} no está definido para un estado transitorio, tomamos por convenio que el tiempo medio de recurrencia de un estado transitorio es siempre infinito.

En cambio para un estado recurrente puede ser finito o infinito y por lo tanto se introduce una nueva clasificación de los estados.

Definición 8. *Si un estado recurrente satisface que $\mu_{ii} < \infty$ se dice que es recurrente positivo.*

Si un estado recurrente satisface que $\mu_{ii} = \infty$ se dice que es recurrente nulo.

El estudio de los primeros retornos se puede extender a visitar por primera vez un estado desde cualquier otro. Las nuevas variables T_{ij} se apoyan en las probabilidades $f_{ij}^{(n)} = P\{X_m = j, X_m \neq j, m = 1, 2, \dots, n-1 | X_0 = i\}, i \neq j$.

Definimos $f_{ij}^{(0)} = 0$ si $i \neq j$. Además:

$$0 \leq \sum_{n=1}^{\infty} f_{ij}^{(n)} \leq 1.$$

Si $\sum_{n=1}^{\infty} f_{ij}^{(n)} = 1$, el conjunto $\{f_{ij}^{(n)}\}$ define la distribución de la variable T_{ij} .

Si $X_0 = i$ entonces el tiempo medio de primer paso al estado j , $E(T_{ij})$, se denota μ_{ij} y se define como:

$$\mu_{ij} = \sum_{n=1}^{\infty} n f_{ij}^{(n)}, i \neq j.$$

Esta definición se puede extender al caso en el que $f_{ij} = \sum_{n=1}^{\infty} f_{ij}^{(n)} < 1$ donde definen las probabilidades de no alcanzar el estado j desde i como $1 - f_{ij}$.

Si $f_{ij} < 1$, entonces el tiempo medio de primer paso de i hasta j es infinito.

Existe una relación entre las probabilidades de transición de una cadena de Markov y las probabilidades de retornar por primera vez a un estado. Esta relación es la siguiente:

$$p_{ii}^{(n)} = \sum_{k=1}^n f_{ii}^{(k)} p_{ii}^{(n-k)}$$

con $f_{ii}^{(0)} = 1, p_{ii}^{(0)} = 1$.

Una expresión similar para $f_{ij}^{(n)}$ y $p_{ij}^{(n)}$ es:

$$p_{ij}^{(n)} = \sum_{k=1}^n f_{ij}^{(k)} p_{jj}^{(n-k)}.$$

Teorema 3. *Un estado es recurrente si y solo si $\sum_{n=0}^{\infty} p_{ii}^{(n)}$ diverge, es decir*

$$\sum_{n=0}^{\infty} p_{ii}^{(n)} = \infty.$$

Un estado es transitorio si y solo si $\sum_{n=0}^{\infty} p_{ii}^{(n)}$ converge, es decir

$$\sum_{n=0}^{\infty} p_{ii}^{(n)} < \infty.$$

Corolario 1. *Asumiendo que i y j intercomunican, podemos encontrarnos las siguientes relaciones:*

- i) El estado i es recurrente si y solo si j es recurrente.*
- ii) El estado i es transitorio si y solo si j es transitorio.*

Corolario 2. *Cada clase de estados recurrentes de una cadena de Markov en tiempo discreto es un conjunto cerrado.*

Definición 9. *Un estado es ergódico si es aperiódico y recurrente positivo. Una cadena de Markov es ergódica si es irreducible, aperiódica y positiva recurrente.*

Definición 10. *Un estado i es absorbente si y solamente si ningún estado de la cadena es accesible desde él, es decir si $p_{ii} = 1$.*

Cuando estudiamos las probabilidades de transición en n etapas $p_{ij}^{(n)}$ podemos hablar, bajo algunos supuestos, de su distribución estacionaria.

Se dice que una cadena de Markov en tiempo discreto admite una distribución estacionaria cuando las probabilidades de ir a largo plazo de un estado i a otro j existen y son independientes de la distribución inicial; es decir si

$$\lim_{n \rightarrow \infty} p_{ij}^{(n)} = p(j).$$

Teorema 4. *Suponemos una cadena de Markov discreta, irreducible, recurrente positiva, aperiódica, con espacio de estados $S=\{1,2,.. \}$ y matriz de transición P . Entonces existe una única distribución estacionaria positiva $\pi = (\pi_1, \pi_2, ...)^T$, de manera que*

$$\pi = \pi P$$

verificando

$$\lim_{n \rightarrow \infty} p_{ij}^{(n)} = \pi_j; \forall i, j \geq 1.$$

Teorema 5. *Para una cadena de Markov ergódica e irreducible existe el $\lim_{n \rightarrow \infty} p_{ij}^{(n)}$ y es independiente de i . Además, haciendo $\pi_j = \lim_{n \rightarrow \infty} p_{ij}^{(n)}$, entonces π_j es la única solución no negativa de*

$$\pi_j = \sum_{i=1}^{\infty} \pi_i p_{ij}, j \geq 1$$

con $\sum_{j=1}^{\infty} \pi_j = 1$.

Por lo tanto si la distribución límite es independiente de la distribución inicial la matriz de transición satisface

$$\lim_{n \rightarrow \infty} P^n = \begin{pmatrix} \pi_1 & \pi_2 & \pi_3 & \dots \\ \pi_1 & \pi_2 & \pi_3 & \dots \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \ddots \end{pmatrix}.$$

Es decir:

$$\lim_{n \rightarrow \infty} a_0 P^n = \pi$$

con $\pi = (\pi_1, \pi_2, \dots)^T$.

Se puede comprobar que cuando nos encontramos con cadenas de Markov, discretas, finitas, el comportamiento a largo plazo se puede obtener de una manera muy sencilla. En ese caso, nos vamos a encontrar sólo cuatro diferentes tipos de estados, atendiendo a su periodicidad y recurrencia.

Lema 1. *Si j es un estado transitorio de una cadena de Markov entonces para cualquier estado i se verifica $\lim_{n \rightarrow \infty} p_{ij}^{(n)} = 0$.*

Teorema 6. *En una cadena de Markov finita, una clase es recurrente si es cerrada.*

Teorema 7. *En una cadena de Markov finita no todos los estados pueden ser transitorios y no hay estados recurrentes nulos. En particular, una cadena de Markov finita e irreducible es recurrente positiva.*

Corolario 3. *Sea una cadena de Markov finita, irreducible y aperiódica, entonces existe una única distribución estacionaria de probabilidad $\pi = (\pi_1, \pi_2, \dots, \pi_N)^T$ tal que*

$$\lim_{n \rightarrow \infty} p_{ij}^{(n)} = \pi_i \quad i, j = 1, 2, \dots, N.$$

Cuando nos encontramos con las condiciones descritas en el Corolario 3, podemos determinar de manera rápida la matriz de transición en n etapas usando álgebra elemental. Tendremos que la matriz en n etapas satisface lo siguiente:

$$P^n = \sum_{i=1}^N \lambda_i^n x_i y_i^T,$$

donde λ_i son los autovalores asociados a la matriz P , x_i es autovector por el lado derecho asociado al autovalor λ_i e y_i es autovector por la izquierda asociado a ese mismo autovalor.

Su distribución estacionaria se podrá calcular tomando como autovalor $\lambda = 1$ como se sigue a continuación y cuya demostración se puede consultar en [7]:

$$\lim_{n \rightarrow \infty} P^n = x_1 y_1^T = \begin{pmatrix} \pi_1 & \pi_2 & \dots & \pi_N \\ \pi_1 & \pi_2 & \dots & \pi_N \\ \cdot & \cdot & \dots & \cdot \\ \pi_1 & \pi_2 & \dots & \pi_N \end{pmatrix}$$

con $x_1 = \pi$, $y_1^T = (1, 1, \dots, 1)$.

Cuando nos encontramos con cadenas de Markov con estados absorbentes la distribución estacionaria asigna probabilidad solo a los estados absorbentes y es necesario estudiar los tiempos necesarios hasta la absorción. Para ello es preciso reordenar la matriz de transición de forma que las filas correspondientes a los estados absorbentes aparezcan en primer lugar. Así ordenada se dirá que la matriz de transición está en la forma canónica. Podemos dividir la matriz en forma canónica en cuatro submatrices. La primera es la matriz unidad, del orden correspondiente. La segunda la matriz nula. La tercera contiene las probabilidades de paso de estados transitorios a estados absorbentes. La cuarta contiene las probabilidades de estados transitorios a estados transitorios.

Teorema 8. *Consideramos una cadena de Markov recurrente, irreducible, y aperiódica. Entonces*

$$\lim_{n \rightarrow \infty} p_{ij}^{(n)} = \frac{1}{\mu_{ii}}$$

donde μ_{ii} , es el tiempo medio de recurrencia del estado i .

A continuación explicaremos un método para calcular el tiempo medio de recurrencia μ_{ii} y tiempos de primer paso μ_{ij} para el caso particular de una cadena de Markov finita e irreducible con N estados.

Para ello utilizaremos la siguiente matriz, cuyos elementos son justamente esos tiempos medios de recurrencia y tiempos medios de primer paso:

$$M = (\mu_{ij}) = \begin{pmatrix} \mu_{11} & \mu_{12} & \mu_{13} & \dots & \mu_{1N} \\ \mu_{21} & \mu_{22} & \mu_{23} & \dots & \mu_{2N} \\ \mu_{31} & \mu_{32} & \mu_{33} & \dots & \mu_{3N} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \mu_{N1} & \mu_{N2} & \mu_{N3} & \cdot & \mu_{NN} \end{pmatrix}.$$

En este caso en vez de calcular los elementos de la matriz usando las probabilidades de primer paso $\{f_{ij}^{(n)}\}$ utilizaremos otro método alternativo, el llamado método de la primera transición. Este método utiliza la relación existente entre los tiempos medios de recurrencia y tiempos medios de primer paso, para así poder construir un sistema lineal, cuya solución es justamente M .

El proceso para su construcción considera qué es lo que ocurre en la primera etapa. Cualquier estado j puede ser alcanzado desde i con probabilidad p_{ij} , o necesitará más de una etapa para alcanzarlo, siendo ese el caso, se alcanzará otro estado k , $k \neq j$ con probabilidad p_{ik} . Entonces el tiempo en llegar al estado j es $1 + \mu_{kj}$, es decir una etapa más el tiempo medio de ir por primera vez desde j hasta k .

Esta relación viene dada por :

$$\mu_{ij} = p_{ij} + \sum_{k=1, j \neq k}^N p_{ik}(1 + \mu_{kj}) = 1 + \sum_{k=1, j \neq k}^N p_{ik}\mu_{kj}. \quad (2)$$

La relación (2) asume que la matriz P es irreducible, es decir cada estado j puede ser alcanzado desde cualquier otro i . Las ecuaciones (2) pueden expresarse de matricialmente de la siguiente forma:

$$M = E + (M - \text{diag}(M))P. \quad (3)$$

Donde E es una matriz de dimensión $N \times N$ de unos.

Como la cadena de Markov es irreducible, en particular es recurrente positiva, garantizamos que los tiempos de retorno y de primer paso sean finitos en todos los casos.

Por consiguiente el sistema de ecuaciones (3) tiene solución. En concreto en [9] se muestra un método iterativo para resolver dicho sistema de ecuaciones lineales.

2. Modelización matemática de epidemias

La construcción de modelos matemáticos para estudiar y predecir el comportamiento de enfermedades infecciosas es uno de los principales objetivos de los científicos en la actualidad. La principal función de estos modelos es mejorar la comprensión de un sistema para prevenir futuras situaciones de enfermedad, determinar la prevalencia e incidencia, y ayudar a tomar decisiones para controlar o erradicar esas enfermedades.

Existen dos tipos de modelos matemáticos: deterministas y estocásticos. En un modelo determinista las mismas entradas producirán invariablemente las mismas salidas, no contemplándose la existencia del azar ni el principio de incertidumbre. En cambio, en un modelo estocástico, el azar interviene en el modelo, de modo que una misma entrada puede producir diversas salidas de manera impredecible y por lo tanto los resultados posibles se generarán con una función de probabilidad.

Para el modelo determinístico, se genera un único valor posible, mientras que en el estocástico se permiten diferentes valores adjudicando una cierta probabilidad a cada uno de estos sucesos.

Supongamos una población de tamaño inicial N afectada por una enfermedad de tipo contagioso, de los cuáles tenemos S sanos, e I infectados. Los infectados, lo están por una enfermedad de tipo contagiosa, y por lo tanto podrán transmitir esa enfermedad. Mientras el procesos infeccioso esté activo y dependiendo del tipo de enfermedad, los individuos pueden pasar por todos o algunos de los siguientes estados:

- Susceptible (S): estado en el cual el individuo puede ser contagiado por otro individuo que esté infectado.

- Infectado (I): estado durante el cual el individuo se halla infectado y puede además infectar.

- Recuperado o Removido (R): estado durante el cual el individuo no puede ser infectado por haber adquirido inmunidad (temporal o permanente), ni infectar a los demás (por haber recuperado o haber pasado la etapa contagiosa de la enfermedad).

Entre las enfermedades infecto-contagiosas encontramos dos grupos principales:

-Las que confieren inmunidad (temporal o permanente) al infectado; es decir una vez recuperado no podrá volver a ser infectado o infectar en el tiempo que le dure su inmunidad. La mayoría son de origen viral (sarampión, varicela...).

-Las que una vez recuperado, el individuo vuelve a ser susceptible inmediatamente, estas suelen ser de tipo bacteriano (enfermedades venéreas, peste..) o protozoos (malaria).

Teniendo en cuenta los distintos estadios relacionados con un proceso infeccioso, podemos encontrarnos diversos modelos epidemiológicos matemáticos como SIS, SIR, SI...

El modelo de Kermack y McKendrick ha sido probablemente el más influyente en las últimas décadas. Es un modelo SIR (Susceptible-Infeccioso-Recuperado) y sus variantes se han convertido en prototipos de sistemas no lineales para explicar el comportamiento de enfermedades infecciosas.

Según este modelo, un individuo susceptible puede entrar en contacto con un infectado, y con una cierta probabilidad será infectado, y por lo tanto su estadio pasará a ser infectado. Por otra parte con la ayuda de su sistema inmunológico, cada individuo infectado entrará en el estadio recuperado, dejando de ser un posible transmisor de la enfermedad.

Figura 2: Modelo SIR.



En el modelo SIS tenemos dos tipos de estadios para los individuos. Los individuos susceptibles que pueden ser infectados pero no adquieren la inmunidad tras su recuperación y por lo tanto vuelven a ser susceptibles.

Figura 3: Modelo SIS.



La mayoría de los estudios epidemiológicos se basan en procesos continuos, yo consideraré un modelo estocástico en tiempo discreto.

2.1 Enfoque determinista

La evolución de la epidemia se representa a través de funciones que muestran la variación de la proporción de individuos en cada situación clínica, en cada instante de tiempo. Dichas funciones se obtienen como la solución de un sistema de ecuaciones diferenciales. Así, en el modelo SIS intervendrán $S(t)$ e $I(t)$ que son respectivamente, las proporciones de individuos susceptibles e infectados presentes en la población en un instante t . En el modelo SIR es preciso incluir también la proporción de individuos recuperados: $R(t)$.

Dada una población finita, cerrada, de tamaño N definimos Δt como un intervalo de tiempo en el que a lo sumo pueda ocurrir un único evento. Denotaremos α al número de contactos hechos por un infeccioso que resulta en infección en Δt ; es decir es la tasa de contagio y γ a la tasa de recuperación de un individuo infectado.

El modelo SIS determinista tiene la siguiente forma:

$$\begin{aligned} S(t + \Delta t) &= S(t) - \frac{\alpha S(t)I(t)}{N} + I(t)\gamma \\ I(t + \Delta t) &= \frac{\alpha S(t)I(t)}{N} + I(t)(1 - \gamma) \end{aligned} \tag{4}$$

donde $t \geq 0$, $0 < \alpha \leq 1$ y $0 < \gamma \leq 1$.

Buscamos una solución a (4) bajo las siguientes condiciones frontera:

$$S(0), I(0) > 0,$$

$$S(t) + I(t) = 1, \forall t \geq 0.$$

$S(0) > 0$, indica que el número de susceptibles en el instante inicial tiene que ser mayor que 0.

$I(0) > 0$, indica que el número de infectados en el instante inicial tiene que ser de al menos un individuo.

$S(0)+I(0)=1$ y $S(t)+I(t)=1, \forall t$, indican que la población inicial es constante a lo largo del tiempo.

En cambio en el modelo SIR determinista los individuos adquieren la inmunidad una vez superada la fase infecciosa. Si asumimos tamaño poblacional constante N y manteniendo la notación del modelo SIS, el modelo SIR clásico se puede expresar de la siguiente manera:

$$\begin{aligned} S(t + \Delta t) &= S(t) - \frac{\alpha S(t)I(t)}{N} \\ I(t + \Delta t) &= I(t)(1 - \gamma) + \frac{\alpha S(t)I(t)}{N} \\ R(t + \Delta t) &= R(t) + \gamma I(t) \end{aligned} \tag{5}$$

junto a las siguientes condiciones frontera:

$$S(0), I(0) > 0, R(0) \geq 0,$$

$$S(0) + I(0) + R(0) = 1,$$

$$S(t) + I(t) + R(t) = 1, t \geq 0.$$

En ambos modelos según sea la combinación de nuestros parámetros α y γ , podremos obtener diferentes conclusiones respecto a la extinción de la epidemia.

Intuitivamente es admisible que si $\frac{\alpha}{\gamma} \leq 1$ la epidemia se extinguirá; mientras que si $\frac{\alpha}{\gamma} > 1$ no podrá erradicarse.

Es decir cuando la tasa de contagio es menor o igual a la de recuperación la epidemia se extinguirá en un número de etapas, mientras que cuando es mayor no podrá erradicarse y la enfermedad será endémica. A la vista de estos resultados es necesario apoyarse en algún indicador que cuantifique la epidemia estudiada.

2.1.1 Medidas de transmisión de una enfermedad contagiosa

El número reproductivo básico, R_0 , es quizás uno de los indicadores más estudiados en epidemiología. Aunque el origen de la modelización matemática con este cuantificador fue propiamente determinista y aplicado a la demografía [10], también fue aplicado posteriormente para enfermedades contagiosas como la malaria. Más tarde se aplicaría a numerosas infecciones humanas como la peste, la viruela... Véase [11]. Siendo probablemente el cuantificador más importante en epidemiología.

Podemos definir R_0 como el número esperado de casos nuevos que produciría una persona infectada durante su período de contagio, en una población de personas susceptibles a la enfermedad; es decir sin resistencia específica frente a ese patógeno [12]. El valor de R_0 en los modelos SIS y SIR, descritos anteriormente se calcula como:

$$R_0 = \frac{\alpha}{\gamma}.$$

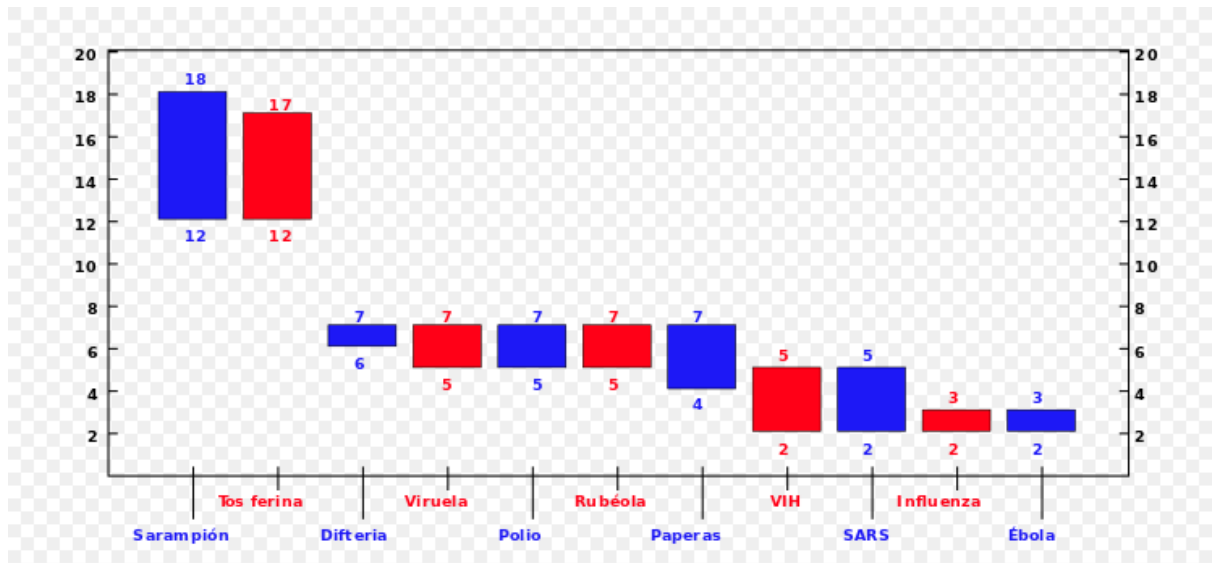
Sin embargo esta definición puede variar dependiendo del contexto epidémico o poblacional considerado.

La propiedad más importante de R_0 es la posibilidad de medir la transmisión potencial de una epidemia durante la fase inicial. En enfermedades con $R_0 \leq 1$, cada individuo infectado contagia en promedio como mucho a un individuo y por lo tanto la epidemia se extinguirá. En cambio si $R_0 > 1$ el patógeno puede establecerse en toda la población y permanecer en ella para siempre.

Por lo tanto es muy importante conocer el valor de R_0 en enfermedades infecciosas, para poder controlar factores que intervengan en él y así poder tomar medidas frente a esa epidemia. En pandemias y epidemias el conocimiento de R_0 permite poder establecer un nivel de alerta y el riesgo que tenemos frente a esa enfermedad.

El matemático Klaus Dietz, del Instituto de Epidemiología Clínica de la Universidad de Tubinga (Alemania), lleva desde la década de los 70 estudiando la propagación de una enfermedad contagiosa con la ayuda de las matemáticas. A continuación, bándonos en sus estudios y apoyándonos también en [13], mostramos en la Figura 4 los valores de R_0 para las principales enfermedades contagiosas.

Figura 4: Principales R_0 de enfermedades contagiosas actuales [14].



Como podemos ver el sarampión y la tos ferina son enfermedades cuyos números reproductivos básicos oscilan entre 12 y 18. Esto quiere decir que un individuo con sarampión puede infectar entre 12 y 18 personas durante su periodo infeccioso. En cambio la enfermedad ébola lleva asociado un R_0 entre 2 y 3 y por lo tanto una persona infectada con ébola contagiará entre 2 y 3 personas antes de su recuperación o muerte.

Comparado con otras enfermedades infecciosas, el ébola tiene un factor de transmisión muy bajo, sin embargo en los últimos años ha ocasionado grandes temores a la población ya que su cura es muy improbable, pues no existe tratamiento efectivo al 100 %, y es mortal. En cambio para el sarampión, aunque tiene un R_0 más alto y podría infectar a más personas, contamos con fármacos para su tratamiento y a día de hoy es muy improbable que alguien muera de esta enfermedad.

Son muchos los factores, entre ellos el temporal, que los biomatemáticos tienen

que tener en cuenta en el cálculo de R_0 . Por ejemplo, aunque el VIH es menos contagioso que el ébola, el primero tiene un periodo de infección de muchos años y el segundo solo puede infectar durante las pocas semanas que el paciente con síntomas pelea con la enfermedad hasta que la vence o sucumbe.

El factor de transmisión R_0 presenta un comportamiento distinto en cada brote de una misma enfermedad, aunque dentro de unos umbrales mínimos y máximos.

Otros elementos que intervienen en su estimación son el número inicial de infectados, población total de la zona afectada, si se trata de zonas aisladas o densamente pobladas o la redes de comunicaciones, que funcionan como vectores.

Pero el factor que más puede hacer variar esta ratio de contagio es el propio ser humano. Los números mostrados hasta aquí solo son ciertos en el caso de que no se haga nada para combatir la enfermedad.

A continuación mostraremos algunos de los diferentes métodos de cálculo de R_0 . El cuantificador R_0 , es fundamentalmente determinista. El primero en realizar un cálculo detallado fue H. Rietze en 1996 con un método basado en la función de supervivencia. Consideramos una población grande y sea $F(a)$ la probabilidad de que un nuevo individuo infectado permanezca al menos infectado un tiempo a y denotamos por $b(a)$ al promedio de nuevos infectados que un individuo infectado produce por unidad de tiempo cuando ha estado infectado al menos un tiempo a .

Con esta descripción, R_0 viene dado por

$$R_0 = \int_0^{\infty} b(a)F(a)da.$$

Esta definición de R_0 es muy general y se puede aplicar a muchas enfermedades contagiosas. No obstante hay casos como la malaria, en los que la transmisión se puede producir además por mosquitos y por lo tanto la función de supervivencia debe de incluir también ese factor de transmisión. Este problema se soluciona gracias a los “next generation methods”, que se adaptan a epidemias en las que existen más de una clase de elementos infecciosos transmisores. Este último método, que se deriva del anterior, fue propuesto por Dieckman en 1990. Más información acerca de otros métodos de cálculo los cuales incluyen jacobianos, autovalores, ... se pueden encontrar en [15].

Pese a que R_0 es fundamental en epidemiología, ya que cuantifica la propagación de una enfermedad, se puede ver que hay casos que ocurre justamente lo contrario a lo esperado.

Hay situaciones en las que la enfermedad persiste aunque la transmisión esperada sea inferior a 1 y se erradica en caso contrario.

Al existir diferentes métodos de cálculo para R_0 , obtendremos diferentes valores de nuestro cuantificador y por lo tanto la interpretación variará según el método usado.

Además R_0 no es el promedio exacto buscado ya que existe una sobreestimación de contagios porque contabiliza como nuevo contagio los que ya han sido infectados con anterioridad. Este cuantificador es más acertado en poblaciones grandes ya que cuando tenemos poblaciones como hospitales [16], prisiones [17]..., las poblaciones son más pequeñas y el valor de R_0 sobreestima la transmisión.

Para solventar este problema se deben de usar otros cuantificadores que no tengan en cuenta aquellos contagios secundarios.

2.2 Enfoque estocástico

En este caso la evolución de la epidemia se representa mediante procesos estocásticos. Suponiendo que el número de individuos de la población no varía durante la fase activa de la epidemia en cada instante $t > 0$, y denotando ahora a $S(t)$, $I(t)$ y $R(t)$ como el número de individuos susceptibles, infectados y recuperados en el instante t respectivamente, se verifica que $S(t) + I(t) = N$, en el modelo SIS o bien $S(t) + I(t) + R(t) = N$, en el modelo SIR.

Por consiguiente, la evolución de una epidemia en un modelo SIS se representa por el proceso univariante $\{I(t), t > 0\}$ que informa de la variación del número de individuos infectados a lo largo del tiempo. Mientras que la evolución de un modelo SIR viene dada por un proceso bidimensional $\{S(t), I(t), t > 0\}$, que proporciona información sobre la cantidad de susceptibles e infectados que conviven en la población en un mismo instante de tiempo.

Una hipótesis frecuente en estos modelos es suponer que los tiempos entre contactos o de recuperación sean variables aleatorias exponenciales, lo que da lugar a procesos markovianos. En concreto el modelo SIS es un proceso de nacimiento y muerte. Conocidas las tasas de contagio entre dos individuos α y de recuperación de un individuo γ , obtendremos directamente las tasas de nacimiento λ_i y de muerte μ_i como

$$\alpha_i = \frac{\alpha}{N}i(N-i), 0 \leq i \leq N,$$

$$\gamma_i = \gamma i, 0 \leq i \leq N.$$

En este tipo de modelos es necesario, al igual que en el determinista, cuantificar la epidemia.

2.2.1 Medidas de transmisión en modelos markovianos

En toda esta sección explicaremos los cuantificadores utilizados en los modelos estocásticos Markovianos según los estudios realizados por J.R. Artalejo y M.J. López-Herrero [18]. El cuantificador determinista R_0 conduce a valores de transmisión aceptables cuando se estudian en poblaciones grandes, pero no siempre tenemos estas condiciones. Por ello introduciremos el número reproductivo exacto, al que denotaremos por R_{e0} y el número de transmisión poblacional denotado por R_p . A diferencia del número reproductivo básico que sobreestimaba los contagios, estas medidas no cuentan aquellos contagios secundarios que R_0 sí que hacía. De esta manera se da una mejor cuantificación de la propagación de la enfermedad.

El número reproductivo exacto R_{e0} es una variable aleatoria que se define como el número exacto de contagios producidos por un individuo infectado durante su periodo de infección.

Cabe recalcar que aquí no están incluidas las infecciones a individuos que anteriormente habían sido infectados y ahora eran susceptibles. El efecto de esos contagios son ignorados.

Para poder estudiar la distribución de R_{e0} en el modelo SIS, la idea que proponen J.R. Artalejo y M.J. López-Herrero es marcar al individuo causante de la epidemia y distinguir en la tasa de contacto quién causa cada nueva infección. Para ello descomponen α_i como

$$\alpha_i = \alpha_i^* + \alpha_{i-1}^{\sim}, \quad 1 \leq i \leq N$$

con $\alpha_i^* = \frac{\alpha}{N}(N-i)$ tasa de contacto del individuo marcado y $\alpha_{i-1}^{\sim} = \frac{\alpha}{N}(i-1)(N-i)$

que lo asocian al promedio de superposiciones de tasas de contagio del resto de $(i-1)$ infectados.

A continuación definen $\varphi_i(z)$ como la función generatriz de probabilidad para R_{e0} , condicionada a i infectados iniciales, es decir:

$$\varphi_i(z) = E[z^{R_{e0}} | I(0) = i] = \sum_{k=0}^{\infty} z^k P\{R_{e0} = k | I(0) = i\}, \text{ si } |z| \leq 1.$$

El objetivo es determinar, en cada z , el valor de las funciones generatrices. La metodología que ellos han utilizado se basa en argumentos de primer paso y contempla las posibles transiciones de la cadena $\{I(t) : t > 0\}$ desde el estado i , que son las siguientes:

- i) Recuperación del individuo marcado.
- ii) Recuperación de otro que no esté en el caso anterior.
- iii) Contagio entre un individuo marcado y un susceptible.
- iv) Contagio entre otro infectado y un susceptible.

Por lo tanto las funciones generatrices verifican para $i, 1 \leq i \leq N$

$$\varphi_i(z) = \frac{\gamma}{\alpha_i + \gamma_i} + \frac{\gamma(i-1)}{\alpha_i + \gamma_i} \varphi_{i-1}(z) + \frac{\alpha_{i-1}^*}{\alpha_i + \gamma_i} z \varphi_{i+1}(z) + \frac{\alpha_{i-1}^{\sim}}{\alpha_i + \gamma_i} \varphi_{i+1}(z) \quad (6)$$

con $1 \leq i \leq N$.

Resolviendo el sistema de ecuaciones (6), en cada punto z , se conseguiría el valor numérico de las transformadas (funciones generatrices) y, aplicando métodos de inversión, se obtienen los valores numéricos de la función de distribución de la variable R_{e0} condicionada a cualquier cantidad de infectados.

Sea k un valor entero no negativo, se definen los momentos factoriales de orden k de R_{e0} condicionado a i infectados iniciales como

$$m_i^k = E[R_{e0}(R_{e0} - 1) \dots (R_{e0} - k + 1) | I(0) = i], 1 \leq i \leq N, k \geq 0.$$

Nótese que $m_i^0 = 1, \forall i = 1, 2, \dots, N$.

Para obtener los momentos tomaremos derivadas en la expresión (6), ya que $m_i^k = \frac{d^k[\varphi(z)]}{dz^k} \Big|_{z=1}$. Derivando k veces la expresión (6) y haciendo $z = 1$, se obtiene la siguiente relación entre los momentos factoriales.

$$m_i^k = \frac{\gamma(i-1)}{\alpha_i + \gamma_i} m_{i-1}^k + \frac{\alpha_i}{\alpha_i + \gamma_i} m_{i+1}^k + \frac{\alpha_i^*}{\alpha_i + \gamma_i} k m_{i+1}^{k-1}, \text{ para } 1 \leq i \leq N, k \geq 1. \quad (7)$$

La ecuación (7) muestra que es posible obtener recursivamente los momentos de orden k en función de los de orden $k-1$.

Nótese que el valor esperado $m_1^1 = \overline{R_{e0}}$ es conceptualmente el factor de transmisión clásico de R_0 .

Los métodos para resolver sistemas de ecuaciones como (7) se desarrollan en libros como [19], [20] o [21].

J.R. Artalejo y M.J. López-Herrero presentan un método directo para resolver (7) que se recoge en el siguiente teorema.

Teorema 9. *Los valores esperados $\{m_i^1; 1 \leq i \leq N\}$ pueden calcularse mediante la siguiente expresión:*

$$m_N^1 = \frac{(N-1)D_{N-1}}{Na_{N-1} + \alpha_{N-1}},$$

$$m_i^1 = \frac{D_i + \alpha_i m_{i+1}^1}{a_i + \alpha_i}, i = N-1, \dots, 1$$

los coeficientes $a_i, D_i, 1 \leq i \leq N-1$ vienen dados por las siguientes expresiones:

$$a_1 = \gamma,$$

$$a_i = \frac{\gamma(\alpha_{i-1} + \alpha_{i-1})}{a_{i-1} + \alpha_{i-1}}, 2 \leq i \leq N-1,$$

$$D_1 = \alpha_1^*,$$

$$D_i = \alpha_i^* + \frac{\gamma(i-1)D_{i-1}}{a_{i-1} + \alpha_{i-1}}, 2 \leq i \leq N-1.$$

Se puede comprobar numéricamente que

$$\lim_{N \rightarrow \infty} \overline{R_{e0}} = R_0 \quad (8)$$

y también que

$$\lim_{\alpha \rightarrow \infty} \overline{R_{e0}} = 1 + \sum_{k=1}^{N-1} \frac{1}{k}. \quad (9)$$

El resultado obtenido en (8) indica que los contactos múltiples entre dos individuos son insignificantes cuando el tamaño poblacional es grande, y por ello el número reproductivo exacto tiende al número reproductivo básico.

El resultado de (9) se puede interpretar del siguiente modo; cuando α tiende a infinito los contagios se vuelven casi instantáneos y por lo tanto el número de infectados, si partimos de uno solo al inicio de la epidemia, pasaría de 1 a N casi inmediatamente. De los $N - 1$ casos infectados, el k -ésimo es debido a un contacto directo con el infectado inicial y ocurre con probabilidad $\frac{1}{k}$.

Cuando toda la población está infectada estamos en el caso $I(0) = N$.

Sea $(\xi|I(0) = N)$ el número de recuperaciones ocurridos mientras el período de infección del individuo inicial. El comportamiento probabilístico de esta variable aleatoria se recoge en el siguiente resultado.

Proposición 2. *Si $\alpha \rightarrow \infty$ entonces $(\xi|I(0) = N)$ es una variable geométrica de parámetro $p = \frac{1}{N}$*

$$P\{\xi = k|I(0) = N\} = (1 - p)^k p, k \geq 0.$$

Por lo tanto

$$E[R_{e0}|\xi = k, I(0) = N] = \frac{k}{N - 1}$$

ya que $(R_{e0}|\xi = k, I(0) = N)$ se distribuye como una $B(k, \frac{1}{N-1})$.

Finalmente, deducimos que

$$E[R_{e0}|I(0) = N] = \frac{\frac{1-p}{p}}{N - 1} = 1,$$

y queda explicado ese 1 que aparece en (9).

Como R_{e0} es una variable aleatoria adimensional todo lo anterior es análogo si $\gamma \rightarrow 0$ para α constante.

Al comenzar la sección se estudiaron las funciones generatrices de las que se puede conseguir la distribución de la variable, mediante un procedimiento de inversión. Sin embargo ellos obtienen la función de masa de la variable R_{e0} de modo directo.

Sean x_i^k las probabilidades definidas por

$$x_i^k = P\{R_{e0} = k | I(0) = i\}, 1 \leq i \leq N, k \geq 0,$$

que son las funciones de masa de R_{e0} condicionada a un número i de infectados iniciales. De nuevo un argumento de primer paso permite obtener la siguiente relación:

$$x_i^k = \delta_{k0} \frac{\gamma}{\alpha_i + \gamma_i} + \frac{\gamma(i-1)}{\alpha_i + \gamma_i} x_{i-1}^k + (1 - \delta_{k0}) \frac{\alpha_i^*}{\alpha_i + \gamma_i} x_{i+1}^{k-1} + \frac{\alpha_{i-1}^*}{\alpha_i + \gamma_i} x_{i+1}^k \quad (10)$$

con $1 \leq i \leq N$ y $k \geq 0$ donde δ_{ab} es la delta de Kronecker definida por:

$$\delta_{ab} = \begin{cases} 1, & \text{si } a = b, \\ 0, & \text{si } a \neq b. \end{cases}$$

Para obtener recursivamente las probabilidades x_i^k podemos utilizar el esquema del Teorema 9.

En el caso $k=0$ tendremos que:

- i) Intercambiar m_i^1 por x_i^0 .
- ii) Tomar $\eta_i = -\alpha_{i-1}^*$ y $\nu = \gamma$ para $1 \leq i \leq N$.

Para $k \geq 1$ tendremos que:

- i) Las incógnitas son x_i^k .
- ii) Tomar $\eta_i = -\alpha_{i-1}^*$ y $\nu_i = \alpha_i^* x_{i+1}^{k-1}$, para $1 \leq i \leq N$.

Como el soporte de la variable aleatoria R_{e0} es infinito es necesario establecer un criterio de parada y detener la computación al acumular la cantidad de probabilidad fijada de antemano.

Otra variable aleatoria, introducida en [18] como un cuantificador de la propagación de la enfermedad, es R_p . Su finalidad es medir el impacto de la epidemia a nivel global y por ello tiene en cuenta cualquier caso de contagio durante un periodo determinado.

Se define la variable aleatoria R_p como el número exacto de contagios exitosos de cualquiera de los infectados, sin distinguir cuál transmite la enfermedad.

Los individuos específicos son considerados como un total y por lo tanto R_p puede verse como una medida de la propagación global.

Además tanto R_p , como R_{e0} , nos permiten controlar R_0 .

Sea R_p la variable aleatoria que contabiliza el número exacto de contagios, debido a la totalidad de infectados, hasta que ocurre la primera recuperación. Denotamos por $\overline{R_p} = E[R_p | I(0) = 1]$ al valor esperado de la variable R_p .

A continuación analizaremos las funciones generatrices de la variable aleatoria R_p , para proseguir con el estudio de momentos y probabilidades puntuales.

De nuevo utilizaremos el proceso de nacimiento y muerte $\{I(t) : t \geq 0\}$, con estado absorbente 0, que informa de la cantidad de infectados en el instante t .

Sea $\psi_i(z)$ la función generatriz de R_p , suponiendo que el número de infectados es i , con $1 \leq i \leq N$. Un argumento de primer paso permite obtener la siguiente relación entre funciones generatrices.

$$\psi_i(z) = \frac{\gamma_i}{\alpha_i + \gamma_i} + (1 - \delta_{iN}) \frac{\alpha_i}{\alpha_i + \gamma_i} z \psi_{i+1}(z), 1 \leq i \leq N. \quad (11)$$

Los momentos factoriales M_i^k también se obtienen recursivamente a partir de la siguiente expresión.

$$M_i^k = (1 - \delta_{i,N-1}) k \sum_{l=i}^{N-2} M_{l+1}^{k-1} \prod_{j=1}^l \frac{\alpha_j}{\alpha_j + \gamma_j} + \delta_{k1} \prod_{j=i}^{N-1} \frac{\alpha_j}{\alpha_j + \gamma_j} \quad (12)$$

con $1 \leq i \leq N-1, k \geq 1$.

A consecuencia de que $P\{R_p < \infty | I(0) = i\} = 1, 1 \leq i \leq N$ y que $R_p = 0$ cuando $I(0) = N$, se tienen los resultados triviales $M_i^0 = 1$, para $1 \leq i \leq N$ y $M_N^k = 0$, para $k \geq 1$.

Para $1 \leq i \leq N-1$ definimos las siguientes probabilidades

$$y_i^k = P\{R_p = k | I(0) = i\}.$$

Al igual que en la discusión de la variable Re_0 un argumento de primer paso, y

manipulaciones algebraicas, proporciona la expresión explícita de las probabilidades puntuales

$$y_i^k = \frac{\gamma_{i+k}}{\alpha_{i+k} + \gamma_{i+k}} \prod_{l=i}^{i+k-1} \frac{\alpha_l}{\alpha_l + \gamma_l}, 0 \leq k \leq N - i$$

donde $\prod_{l=i}^{i-1} \frac{\alpha_l}{\alpha_l + \gamma_l} \equiv 1$. Además $y_N^0 = 1$.

Con la elección de las tasas del modelo SIS clásico los valores esperados M_i^1 y las varianzas $\sigma_i^2 = M_i^2 + M_i^1 - (M_i^1)^2, 1 \leq i \leq N$ presentan las siguientes expresiones explícitas.

$$M_i^1 = \frac{(N - i)R_0}{R_0 + N}, 1 \leq i \leq N, \quad (13)$$

$$\sigma_i^2 = \frac{N(N - i)R_0((N - i + 1)R_0 + N)}{(2R_0 + N)(R_0 + N)^2}, 1 \leq i \leq N. \quad (14)$$

Como podemos observar en (13) y (14), ambas características disminuyen en cuanto el número de infectados aumenta.

Además podemos observar que $R_0 \in (0, \infty)$, mientras que $\overline{R_p} \in (0, N - 1)$, es decir tienen dominios distintos, sin embargo al igual que ocurría con el número reproductivo exacto, $\overline{R_p}$ converge a R_0 cuando el tamaño de la población crece.

Es decir,

$$\lim_{N \rightarrow \infty} \overline{R_p} = \lim_{N \rightarrow \infty} \frac{(N - 1)R_0}{R_0 + N} = R_0.$$

Así pues, el promedio $\overline{R_p}$ cuantifica de una manera similar a R_0 en poblaciones grandes.

2.2.2 Comportamiento de la epidemia en su fase activa

La extinción de la epidemia en una población finita ocurre con probabilidad 1, ya que el proceso markoviano tiene un espacio de estados finito y un único estado absorbente que corresponde a la situación en la que no hay infectados en la población.

Tenemos interés en conocer el comportamiento probabilístico de la epidemia en su fase activa ya que la extinción es segura, pero el tiempo hasta dicha extinción puede ser largo, aunque finito en media. La distribución estacionaria modela el comportamiento a largo plazo. Cuando además el espacio de estados es finito con un único estado absorbente la distribución estacionaria asigna probabilidad únicamente a dicho estado. Por ello es necesario considerar distribuciones análogas a la distribución estacionaria y que describan el comportamiento de la cadena en sus estados transitorios.

La distribución RE, o de razón de esperanzas, es una alternativa con la que podremos entender y estudiar la dinámica del sistema antes de la absorción. Este estudio también se puede hacer observando la distribución quasi-estacionaria, pero nosotros nos limitaremos al estudio de la distribución RE según los estudios realizados por J.R. Artalejo y M.J. López-Herrero [22].

Sea $\{X(t) : t \geq 0\}$ una cadena de Markov en tiempo continuo, con espacio de estados numerables, S . Dicho espacio está formado por la unión de los subconjuntos disjuntos S_A y S_T que agrupan, respectivamente, a los estados absorbentes y a los transitorios. Esto es, $S = S_A \cup S_T$.

Definición 11. Sea T_j el tiempo que el proceso $\{X(t) : t \geq 0\}$ pasa en el estado $j \in S_T$, antes de la absorción. Dado $X(0) = i \in S_T$, denotaremos el tiempo esperado hasta la absorción condicionado al estado inicial i como $E_i[T]$. La distribución RE, condicionada al estado inicial i , viene dada por la siguiente expresión:

$$p_i(j) = \frac{E_i[T_j]}{E_i[T]}, \quad i, j \in S_T.$$

Como podemos ver este cociente de esperanzas siempre existe ya que el tiempo esperado hasta la absorción condicionado al estado inicial i es finito.

Observamos que la distribución RE asigna probabilidades positivas a todos los estados j accesibles desde i .

La distribución RE reformulada, sin condicionar a la información $X(0) = i$, se expresa de la siguiente manera

$$p(j) = \frac{\sum_{i \in S_T} a_i E_i[T_j]}{\sum_{i \in S_T} a_i E_i[T]}, \quad j \in S_T.$$

Como vemos la distribución RE depende de la distribución inicial $(a_i : i \in S_T)$.

Volviendo a las probabilidades RE, para calcularlas necesitamos determinar los promedios $E_i[T]$.

Por una parte, cuando $P\{T < \infty | X(0) = i\} < 1$ se tiene que $E_i[T] = \infty$. Pero en el caso de que la $P\{T < \infty | X(0) = i\} = 1$, y siguiendo la metodología detallada en el libro de Kulkarni [20], cada $E_i[T]$ se calcula como la solución del sistema de ecuaciones

$$E_i[T] = \frac{1}{q_i} + \sum_{j \in S_T, i \neq j} \frac{q_{ij}}{q_i} E_j[T], \quad i \in S_T.$$

Los valores q_{ij} son las probabilidades instantáneas con la que el proceso realiza una transición de los estados i a j , definidas por $q_{ij} = \lim_{t \rightarrow 0} \frac{p_{ij}(t)}{t}$ si $i \neq j$, y

$$q_i = -\sum_{i \neq j} q_{ij}.$$

3. Estudio en tiempo discreto de una epidemia sin inmunidad

En este capítulo estudiaremos el comportamiento de una enfermedad contagiosa, sin inmunidad, utilizando un modelo SIS y asumiendo un enfoque Markoviano. Los resultados obtenidos son originales y aún están sin publicar.

La mayoría de los estudios basados en modelos estocásticos publicados utilizan procesos en tiempo continuo. En la vida real esto es por una limitación ya que muchas veces nos encontramos con datos agrupados por unidades de tiempo. En nuestro contexto particular, vendría a significar que no se tiene información del instante concreto del día en la que se ha producido una nueva infección pero si se dispone del número de infecciones que han tenido lugar día a día, mes a mes...

En este trabajo se va a considerar un modelo estocástico en tiempo discreto para tratar de subsanar este inconveniente. La literatura de modelos epidémicos en tiempo discreto no es demasiado extensa aunque sí es heterogénea; en este trabajo seguiremos la descripción y notación empleada por Allen y Burgin en [23].

Supongamos que en una población de N individuos tiene lugar un brote epidémico de una enfermedad contagiosa.

Supongamos además que durante este brote epidémico la población permanece inalterada, es decir el tamaño poblacional no varía.

Para representar la evolución de la enfermedad en la población, durante el brote epidémico, consideramos el proceso estocástico discreto $\{I_n : n \geq 0\}$; que describe el número de infectados al comienzo de cada etapa. Las hipótesis de la modelización se refieren al tiempo transcurrido hasta una nueva infección o una recuperación y son las siguientes:

i) Sea Y_i la variable aleatoria tiempo hasta un nuevo caso de infección, sabiendo que hay i infectados; asumimos que la variable $Y_i \sim \exp(\alpha_i)$, donde $\alpha_i = \frac{\alpha i(N-i)}{N}$, siendo α la tasa de contagio; es decir el número medio de contagios ocurridos por unidad de tiempo, i el número de individuos infectados y N el tamaño poblacional.

ii) Asumimos que los individuos se recuperan independientemente y que cada uno de ellos necesita un tiempo exponencial de tasa γ en conseguirlo. Por consiguiente,

la variable Z_i , asociada al tiempo transcurrido hasta que uno de los i individuos infectados se recupera, sigue una distribución exponencial de tasa $i\gamma$, es decir

$Z_i \sim \exp(i\gamma)$ para $i \geq 1$.

Nuestro proceso estocástico junto a las hipótesis i) y ii), nos permite modelizar $\{I_n : n \geq 0\}$ como una cadena de Markov en tiempo discreto.

Sea la variable S_n definida como el número de individuos susceptibles al comenzar la etapa n . En este modelo matemático recordamos que la cura de la enfermedad no proporciona la inmunidad y por lo tanto cualquier infectado que se recupera pasa inmediatamente a ser susceptible otra vez.

Consideramos una población cerrada de tamaño N que sufre una enfermedad contagiosa donde los individuos pueden pasar por los estados susceptible o infectado únicamente. En cualquier etapa se verificará que $S_n + I_n = N$.

Como vamos a utilizar procesos en tiempo discreto, supondremos que la población se observa a intervalos regulares de amplitud Δt y llamaremos etapas a dichos intervalos.

Lo primero que tenemos que hacer es elegir Δt de manera que en ese intervalo de tiempo a lo sumo pueda ocurrir un evento; siendo estos eventos: nueva infección o recuperación. Es decir, no podemos observar, en el paso de una etapa a otra, dos eventos o más. De esta manera los cambios que se pueden observar en una etapa, son:

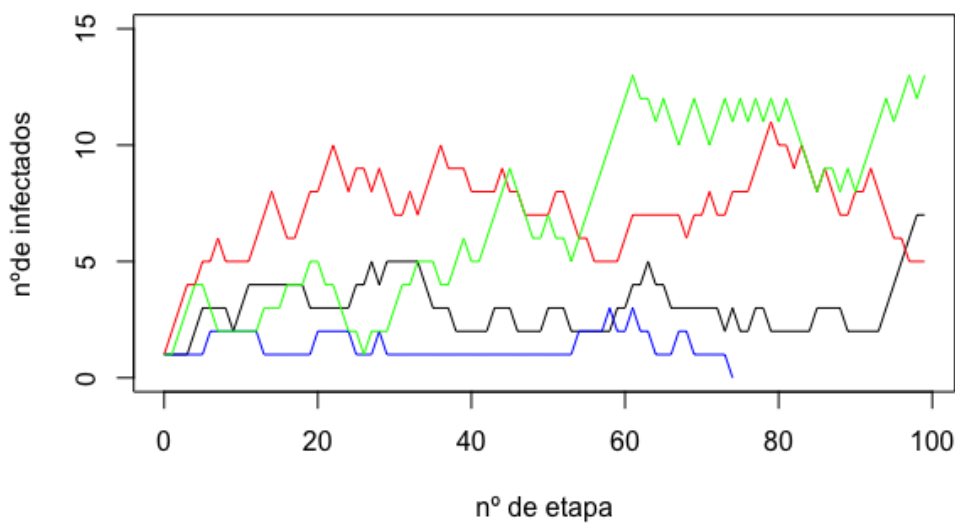
- Un susceptible es infectado, el número de susceptibles disminuye en una unidad y el de infectados aumenta una unidad.
- No ocurre nada, los valores de infectados y susceptibles se mantienen.
- Un infectado se recupera, el número de susceptibles aumenta en una unidad mientras que infectados disminuye en una unidad.

A continuación podemos ver una simulación de varias trayectorias de este tipo de modelo, asumiendo que inicialmente hay un único individuo infectado y transmisor de la enfermedad. Este modelo podría corresponderse con una epidemia de gripe en un barco pesquero, suponiendo que el barco parte del puerto con un único tripulante infectado.

Para la realización de la simulación, hemos considerado una población de 20 individuos, sometidos a una enfermedad contagiosa de la que cada individuo se recupera en media en una unidad de tiempo, con tasa de contagio 2 individuos por unidad de tiempo y que se realizan 25 observaciones del estado de los individuos por unidad de tiempo. Es decir los parámetros elegidos en este escenario son:

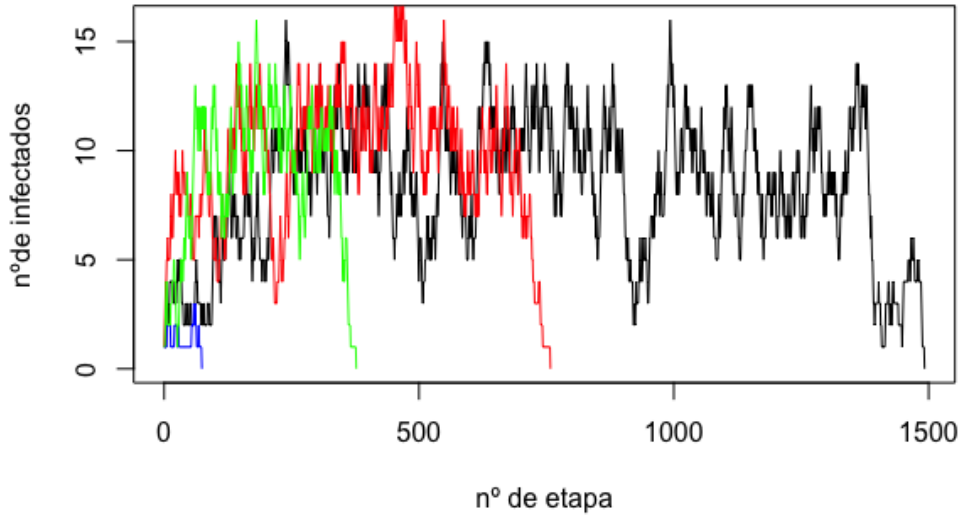
$N = 20$, $\gamma = 1$, $\alpha = 2$, y $\Delta t = 0,04$.

Figura 5: Simulación de las primeras 100 etapas de la epidemia.



Como vemos en la Figura 5, al ser un modelo estocástico, se pueden dar diferentes trayectorias. Partiendo de diferentes semillas, hemos obtenido cuatro trayectorias totalmente diferentes. Se observa que en la trayectoria azul la extinción de la epidemia se alcanza en un número de etapas inferior a 100, en concreto a las 72 etapas. En cambio las otras tres necesitarán más etapas para la erradicación de la epidemia, es decir la duración del brote epidémico es superior a 100 etapas.

Figura 6: Simulación hasta la erradicación.



En la Figura 6 aparecen las mismas trayectorias que en la Figura 5, pero esta vez no hemos restringido el número de etapas a 100, si no que hemos dejado evolucionar el proceso hasta llegar a la extinción de la epidemia. En la trayectoria presentada con color azul, como anteriormente habíamos dicho, el brote epidémico concluye sobre las 72 etapas, en la verde sobre los 400, en la roja sobre las 750 y en la negra sobre las 1500.

El proceso estocástico $\{I_n : n \geq 0\}$ asociado al modelo SIS tiene los enteros entre 0 y N como espacio de estados y, debido a la elección de las etapas, las únicas probabilidades de transición en una etapa, $P_{ij} = P\{I_{n+1} = j | I_n = i\}$, no nulas son las siguientes:

$$\begin{cases} p_{i,i+1} = P\{I_{n+1} = i+1 | I_n = i\} = (1 - \delta_{Ni}) \frac{\alpha}{N} i(N-i) \Delta t = \frac{\alpha}{N} i(N-i) \Delta t \\ p_{i,i-1} = P\{I_{n+1} = i-1 | I_n = i\} = (1 - \delta_{0i}) i \gamma \Delta t = i \gamma \Delta t \\ p_{i,i} = P\{I_{n+1} = i | I_n = i\} = 1 - (\frac{\alpha}{N} i(N-i) \Delta t + i \gamma \Delta t). \end{cases} \quad (15)$$

La matriz de transición obtenida es la siguiente:

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & . & . & . & 0 & 0 \\ p_{10} & p_{11} & p_{12} & 0 & 0 & . & . & . & 0 & 0 \\ 0 & p_{21} & p_{22} & p_{23} & 0 & . & . & . & 0 & 0 \\ 0 & 0 & p_{32} & p_{33} & p_{34} & 0 & . & . & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & . & . & . & p_{NN-1} & p_{NN} \end{pmatrix}.$$

Como podemos ver nos encontramos con un único estado absorbente en $\{0\}$ mientras que los demás son todos transitorios.

Atendiendo a la estructura de las probabilidades de transición debemos de elegir Δt , de modo que la matriz de transición sea estocástica para cada elección de parámetros α, γ y N . Esto equivale a imponer las siguientes condiciones:

- a) $0 \leq p_{ij} \leq 1, \forall i, j = 0, 1, \dots, N$.
- b) $\sum_{j=0}^N p_{ij} = 1, \forall i = 0, 1, \dots, N$.

Es decir, debemos de encontrar un tamaño de etapa positivo de manera que en la matriz de transición la suma por filas sea exactamente 1, esto es:

$$\Delta t > 0, \quad \frac{\alpha}{N} i(N-i) \Delta t + i \gamma \Delta t \leq 1, i = 0, 1, \dots, N. \quad (16)$$

Por lo tanto la condición que tiene que verificar Δt despejando en (16) es :

$$(\frac{\alpha i(N-i)}{N} + i \gamma) \Delta t \leq 1$$

o equivalentemente

$$\Delta t \leq \frac{1}{\frac{\alpha i(N-i)}{N} + i \gamma}. \quad (17)$$

A continuación vamos a analizar la función:

$$f(x) = \frac{1}{\frac{\alpha x(N-x)}{N} + x\gamma}, \quad x \in [0, N] \cap \mathbb{Z}^+.$$

La amplitud de la etapa, Δt , está acotada por el mínimo valor de la función $f(x)$, en el dominio de los valores enteros no negativos, inferiores o iguales al tamaño de la población. Obtendremos el mínimo de $f(x)$, ya que el tamaño de etapa tiene que ser inferior a esa expresión y por lo tanto si tomo el mínimo me aseguraré que Δt es menor que cualquier valor de esa función sea cual sea el valor de x .

Como podemos ver $f(x)$ es positiva en su dominio de definición, ya que tanto como dividendo como divisor son valores positivos, pues x es siempre menor o igual a N . Además $f(x)$ tiende a infinito si nos aproximamos a 0 por la derecha; y tiende a una constante cuando nos aproximamos a N por la izquierda.

Nótese que localizar el valor mínimo de $f(x)$, en $x > 0$, es equivalente a buscar el valor máximo de la función

$$g(x) = \frac{\alpha x(N-x)}{N} + x\gamma.$$

Derivando $g(x)$ obtenemos la siguiente expresión:

$$g'(x) = \frac{\alpha}{N} [N(1 + \frac{\gamma}{\alpha}) - 2x]$$

e igualando a 0, tenemos que el candidato a máximo es:

$$g'(x_0) = 0 \Leftrightarrow x_0 = \frac{N}{2} (1 + \frac{\gamma}{\alpha}).$$

Como en la población podemos tener a lo sumo N infectados, consideramos dos posibles casos:

- a) Si $x_0 \leq N \Rightarrow f(x_0) \leq f(x) \quad \forall x \in [0, N]$.
- b) Si $x_0 > N \Rightarrow f(N) \leq f(x) \quad \forall x \in [0, N]$.

Por lo tanto siempre podremos elegir un tamaño de etapa óptimo que sería:

$$\Delta t \leq \min\{f(x_0), f(N)\}.$$

Por analogía con los modelos clásicos denotaremos al cociente $\frac{\alpha}{\gamma}$ por R_0 y por lo tanto obtendremos la siguiente relación entre el tamaño de la etapa y el valor del número básico de reproducción del modelo SIS.

$$x_0 \leq N \Leftrightarrow \frac{n}{2}(1 + \frac{\gamma}{\alpha}) \leq N \Leftrightarrow 1 + \frac{\gamma}{\alpha} \leq 2 \Leftrightarrow \frac{\gamma}{\alpha} \leq 1 \Leftrightarrow R_0 \geq 1.$$

Si $x_0 \leq N$ el mínimo restringido a nuestros posibles valores de x , será justamente $f(x_0)$ y debemos por lo tanto elegir como Δt un valor inferior o igual a ese valor $f(x_0)$. Sustituyendo x_0 por $\frac{N}{2}(1 + \frac{\gamma}{\alpha})$ en la función $f(x)$ obtenemos :

$$Si R_0 \geq 1 \Rightarrow \Delta t \leq f(x_0) = \frac{N/\alpha}{\frac{N^2}{4}(1 + \frac{\gamma}{\alpha})^2} = \frac{4\alpha}{N(\alpha + \gamma)^2}. \quad (18)$$

Si $x_0 > N$ el mínimo restringido a nuestros posibles valores de x , será justamente $f(N)$ y debemos por lo tanto elegir como Δt un valor inferior o igual a $f(N)$. Sustituyendo N en la función obtenemos :

$$Si R_0 < 1 \Rightarrow \Delta t \leq f(N) = \frac{1}{\gamma N}. \quad (19)$$

Una vez elegido el tamaño de etapa atendiendo a las condiciones de (18) y (19), construiremos la matriz de transición del número de individuos infectados según las probabilidades de transición expresadas en (15).

Recordamos que $\{I_n : n \geq 0\}$ es una cadena de Markov con espacio de estados finito y un único estado absorbente en $\{0\}$, por lo tanto a largo plazo se verifica que

$$\lim_{n \rightarrow \infty} P^n = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 0 & 0 & \dots & 0 \\ 1 & 0 & 0 & \dots & 0 \\ \cdot & \cdot & \cdot & \dots & \cdot \\ 1 & 0 & 0 & \dots & 0 \end{pmatrix}.$$

La distribución es degenerada sobre el único estado absorbente, es decir $\lim_{n \rightarrow \infty} P\{I_n = j | I_0 = i\} = \delta_{0j}, \quad \forall i \in \{1, 2, \dots, N\}.$

Esta conclusión permite afirmar que con total seguridad a largo plazo la enfermedad contagiosa se extingue, en un tiempo medio finito, sean cuales sean los valores de nuestros parámetros usados.

Entonces, independientemente del valor de R_0 y a diferencia de los modelos deterministas, la epidemia se erradicará. Pero, aunque la epidemia finalice con total seguridad, no podemos predecir exactamente su duración. Por ello, a continuación estudiaremos el número de etapas transcurridas hasta la extinción de la epidemia.

Sea la variable $T = \min\{n \geq 0 : I_n = 0\}$, que denota el número mínimo de etapas transcurridas hasta finalizar la epidemia.

La variable T es una variable aleatoria discreta con infinitos puntos de masa, por lo que no se puede garantizar que exista un número máximo de etapas que aseguren la finalización del proceso infeccioso.

En primer lugar nos ocuparemos de estudiar la función de masa de T , suponiendo que hay 1 infectado inicial, esto es $I_0 = 1$; es decir

$$\alpha_1(n) = P\{T = n | I_0 = 1\}, n \geq 1. \quad (20)$$

En general podemos definir probabilidades similares suponiendo que observamos i infectados al comienzo del estudio. Esto es,

$$\alpha_i(n) = P\{T = n | I_0 = i\}, n \geq 1, i \geq 1. \quad (21)$$

Para poder determinar estas probabilidades es preferible trabajar con las probabilidades acumuladas

$$u_i(n) = P\{T \leq n | I_0 = i\}. \quad (22)$$

Conocidas estas probabilidades auxiliares determinamos la función de masa utilizando la relación

$$\alpha_i(n) = u_i(n) - u_i(n - 1), \forall n > 0. \quad (23)$$

con

$$u_i(0) = 0, \forall i > 0.$$

Teorema 10. *Las probabilidades auxiliares son la solución del sistema de ecuaciones*

$$u_i(n) = p_{i0} + \sum_{j=i-1, j \neq 0}^{i+1} p_{ij} u_j(n-1). \quad (24)$$

La demostración de (24) se puede consultar en la página 163 del libro de Kulkarni [20].

El sistema de ecuaciones (24) se resuelve recursivamente partiendo de las relaciones triviales

$$u_i(n) = 0, n+1 \leq i \leq N. \quad (25)$$

A modo de ejemplo muestro los resultados para las etapas 1 y 2.

$$\begin{aligned} u_1(1) &= p_{10} = \gamma \Delta t. \\ u_i(1) &= 0, i \geq 2. \\ u_1(2) &= p_{10} + p_{11} u_1(1) + p_{12} u_2(1) = \gamma \Delta t (2 - \frac{\alpha}{N} 2(N-2) \Delta t - \gamma \Delta t). \\ u_2(2) &= p_{20} + p_{21} u_1(1) + p_{23} u_3(1) = 2(\gamma \Delta t)^2. \\ u_i(2) &= 0, i \geq 3. \\ &\dots \end{aligned} \quad (26)$$

Como la variable T tiene soporte infinito numerable a la hora de resolver las ecuaciones (22) deberemos de establecer un criterio de parada, que depende de la probabilidad acumulada, como por ejemplo iterar la recursión hasta la etapa que acumule el 99 % de la distribución.

Calculadas las probabilidades acumuladas, $u_i(n)$, podremos calcular las probabilidades puntuales utilizando (21).

El siguiente algoritmo permite obtener numéricamente las probabilidades acumuladas en cada escenario concreto.

Algoritmo 1 *Las probabilidades acumuladas definidas en (22) se pueden calcular recursivamente según el siguiente esquema:*

Fijar $\epsilon > 0$

Paso 1: Tomamos $n=1$.

Paso 2: Tomamos $i=1$.

Paso 3: Calculamos $u_i(n)$ mediante (24) y usando (25).

Paso 4: comprobamos si $u_i(n) > \epsilon$, si es así paramos.

Paso 5: hacemos $i = i + 1$. Si $i \leq N$, vamos al paso 3.

Paso 6: hacemos $n = n + 1$, vamos al paso 2.

Fijado el valor de ϵ , con el algoritmo anterior se consigue determinar numéricamente cualquier probabilidad $u_i(n)$ y a partir de ellas obtendremos $\alpha_i(n)$, para cualquier $i \geq 1$.

Como mostraremos en la sección 4, para obtener todas las probabilidades, el criterio de parada a veces puede acarrear muchas iteraciones, y convertirse en un método largo en cuanto a la computación.

A continuación nos ocuparemos de estudiar los momentos de T . Como no podremos obtener de manera explícita su función de masa tenemos que ocuparnos de obtener dichos momentos sin contar con dicha función.

Para ello recordamos que la cadena $\{I_n : n \geq 0\}$ tiene espacio de estados finito, con un único estado absorbente y por lo tanto la probabilidad de que finalice la epidemia en un número finito de etapas es 1. Es decir, T es una variable aleatoria propia y satisface

$$P\{T < \infty | I_0 = i\} = 1, \forall i \geq 1.$$

Por consiguiente su esperanza es finita; es decir el número medio de etapas hasta la extinción de la epidemia es finito como habíamos afirmado.

Denotamos por m_i , al número medio de etapas hasta la erradicación de la epidemia, partiendo de i infectados, esto es

$$m_i = E[T | I_0 = i].$$

El cálculo de m_i se puede hacer a partir del Teorema 4.4 del libro de Kulkarni [20] y es el siguiente:

Teorema 11. *Los valores esperados m_i , son la menor solución de :*

$$m_i = 1 + \sum_{j=i-1, j \neq 0}^{i+1} p_{ij} m_j, \quad 1 \leq i \leq N. \quad (27)$$

Este sistema de ecuaciones puede expresarse matricialmente de la siguiente manera

$$\mathbf{m} = \mathbf{1} + \mathcal{P}_{\mathcal{T}}\mathbf{m} \quad (28)$$

donde $\mathbf{m} = [m_1, m_2, \dots, m_N]^T$, $\mathbf{1}$ es un vector de unos y $\mathcal{P}_{\mathcal{T}}$ es la parte de la matriz de transición que corresponde a los estados transitorios.

La ecuación (28) puede resolverse algebraicamente de la siguiente manera:

$$\mathbf{m} = -(\mathcal{P}_{\mathcal{T}} - \mathcal{I})^{-1}\mathbf{1}. \quad (29)$$

Dependiendo del tamaño poblacional elegido, el cálculo de esta matriz puede ser algo engorroso, y cuando estamos trabajando con poblaciones grandes o incluso no muy pequeñas, la inversión de esas matrices de dimensión $N \times N$ requiere de la utilización de softwares muy potentes.

Para evitar la dificultad computacional vamos a trabajar sobre las ecuaciones (27) de una manera más analítica y para ello utilizaremos la metodología de ecuaciones en diferencias finitas.

Teniendo en cuenta que $\{I_n : n \geq 0\}$ es un proceso de nacimiento y muerte, cada ecuación de la expresión (27) involucra como máximo tres probabilidades. Esto es,

$$\begin{aligned} (1 - p_{11})m_1 &= 1 + p_{12}m_2 \\ (1 - p_{22})m_2 &= 1 + p_{21}m_1 + p_{23}m_3 \\ &\dots \\ (1 - p_{ii})m_i &= 1 + p_{i,i-1}m_{i-1} + p_{i,i+1}m_{i+1} \\ &\dots \\ (1 - p_{NN})m_N &= 1 + p_{N,N-1}m_{N-1}. \end{aligned} \quad (30)$$

Para cualquier estado inicial i , con $1 \leq i \leq N$, se tiene que

$$1 - p_{ii} = p_{ii-1} + (1 - \delta_{iN})p_{ii+1}.$$

Podemos expresar las ecuaciones (30) como:

$$\begin{aligned}
 p_{10}m_1 &= 1 + p_{12}(m_2 - m_1) \\
 p_{21}m_2 + p_{23}m_3 &= 1 + p_{21}m_1 + p_{23}m_3 \\
 &\dots \\
 p_{ii-1}m_i + p_{ii+1}m_{i+1} &= 1 + p_{i,i-1}m_{i-1} + p_{i,i+1}m_{i+1} \\
 &\dots \\
 p_{NN}m_N &= 1 + p_{N,N-1}m_{N-1}.
 \end{aligned} \tag{31}$$

Definimos las diferencias

$$\begin{cases} d_1 = m_1, & \text{si } i = 1 \\ d_i = m_i - m_{i-1}, & \text{si } i = 2, \dots, N, \end{cases} \tag{32}$$

y expresamos el sistema (31) en función de ellas. Resultando el siguiente sistema de ecuaciones

$$\begin{aligned}
 p_{10}d_1 &= 1 + p_{12}d_2 \\
 p_{21}d_2 &= 1 + p_{23}d_3 \\
 &\dots \\
 p_{ii-1}d_i &= 1 + p_{i,i+1}m_{i+1} \\
 &\dots \\
 p_{NN-1}d_N &= 1.
 \end{aligned} \tag{33}$$

Nótese que la última ecuación tiene solución explícita

$$d_N = \frac{1}{p_{n,n-1}} = \frac{1}{N\gamma\Delta t}. \tag{34}$$

Y que todas las demás se solucionan recursivamente a partir de ella.

Dado k , con $1 \leq k \leq N - 1$, definimos

$$A_k = \frac{p_{k,k+1}}{p_{k,k-1}} \text{ y } B_k = \frac{1}{p_{k,k-1}}.$$

Reescribiendo (33) en función de ellas obtendremos una expresión más sencilla

$$d_k = A_k d_{k+1} + B_k, \quad 1 \leq j \leq N - 1. \tag{35}$$

Resolviendo en orden inverso se obtienen las siguientes expresiones:

$$\begin{aligned}
 d_{N-1} &= A_{N-1}d_N + B_{N-1}, \\
 d_{N-2} &= A_{N-2}d_{N-1} + B_{N-2} = A_{N-2}(A_{N-1}d_N + B_{N-1}) + B_{N-2} \\
 &= A_{N-2}A_{N-1}d_N + A_{N-2}B_{N-1} + B_{N-2} \\
 d_{N-3} &= A_{N-3}d_{N-2} + B_{N-3} = A_{N-3}(A_{N-2}A_{N-1}d_N + A_{N-2}B_{N-1}) + B_{N-2}) + B_{N-3} \\
 &= A_{N-3}A_{N-2}A_{N-1}d_N + A_{N-3}A_{N-2}B_{N-1} + A_{N-3}B_{N-2} + B_{N-3} \\
 &= (\prod_{j=1}^3 A_{N-j})d_N + \sum_{j=1}^3 B_{N-j} \\
 d_{N-4} &= A_{N-4}d_{N-3} + B_{N-4} \\
 &= A_{N-4}[\prod_{j=1}^3 A_{N-j}]d_N + A_{N-4} \sum_{j=1}^3 B_{N-j} (\prod_{i=j+1}^3 A_{N-i}) + B_{N-4} \\
 &= (\prod_{j=1}^4 A_{N-j})d_N + \sum_{j=1}^4 B_{N-j} (\prod_{i=j+1}^4 A_{N-i}). \\
 &\dots
 \end{aligned} \tag{36}$$

En general, si $1 \leq k \leq N-1$

$$d_k = (\prod_{j=1}^{N-k} d_N) + \sum_{j=1}^{N-k} B_{N-j} (\prod_{i=j+1}^{N-k} A_{N-i}).$$

Tras un cambio de variables, haciendo $m = N-j$ y $s = N-i$ tenemos:

$$d_k = (\prod_{m=k}^{N-1} d_N) + \sum_{m=k}^{N-1} B_m (\prod_{s=k}^{m-1} A_s). \tag{37}$$

Sustituyendo las expresiones A_m y B_m en (35) se tiene que:

$$d_k = \prod_{m=k}^{N-1} \left(\frac{p_{m,m+1}}{p_{m,m-1}} \right) \frac{1}{p_{N,N-1}} + \sum_{m=k}^{N-1} \frac{1}{p_{m,m-1}} \prod_{s=k}^{m-1} \left(\frac{p_{s,s+1}}{p_{s,s-1}} \right) \tag{38}$$

para $1 \leq k \leq N-1$.

Un poco de trabajo algebraico permite obtener una expresión explícita para las diferencias d_k

$$\begin{aligned}
 d_k &= \prod_{m=k}^{N-1} \frac{\frac{\alpha}{N} m(N-m)\Delta t}{m\gamma\Delta t} \frac{1}{N\gamma\Delta t} + \sum_{m=k}^{N-1} \frac{1}{m\gamma\Delta t} \prod_{s=k}^{m-1} \frac{\frac{\alpha}{N} s(N-s)\Delta t}{s\gamma\Delta t} \\
 &= \prod_{m=k}^{N-1} \left(\frac{\alpha}{\gamma} \frac{N-m}{N} \right) \frac{1}{N\gamma\Delta t} + \sum_{m=k}^{N-1} \frac{1}{m\gamma\Delta t} \prod_{s=k}^{m-1} \frac{\alpha}{\gamma} \frac{N-s}{N}
 \end{aligned}$$

$$\begin{aligned}
 &= \left(\frac{R_0}{N}\right)^{N-k} \frac{1}{N\gamma\Delta t} \prod_{m=k}^{N-1} (N-m) + \sum_{m=k}^{N-1} \frac{1}{m\gamma\Delta t} \prod_{s=k}^{m-1} R_0 \frac{N-s}{N} \\
 &= \left(\frac{R_0}{N}\right)^{N-k} \frac{1}{N\gamma\Delta t} (N-k)! + \frac{1}{\gamma\Delta t} \sum_{m=k}^{N-1} \left(\frac{R_0^{m-k}}{m} \prod_{s=k}^{m-1} \frac{N-s}{N}\right) \\
 &= \left(\frac{R_0}{N}\right)^{N-k} \frac{(N-k)!}{N\gamma\Delta t} + \frac{1}{\gamma\Delta t} \sum_{n=0}^{N-k-1} \frac{1}{k+n} \left(\frac{R_0}{N}\right)^n \frac{(N-k)!}{(N-k-n)!}.
 \end{aligned}$$

Finalmente, las diferencias son

$$d_k = \begin{cases} \left(\frac{R_0}{N}\right)^{N-k} \frac{(N-k)!}{N\gamma\Delta t} + \frac{1}{\gamma\Delta t} \sum_{n=0}^{N-k-1} \frac{1}{k+n} \left(\frac{R_0}{N}\right)^n \frac{(N-k)!}{(N-k-n)!}, & \text{si } 1 \leq k \leq N-1, \\ \frac{1}{N\gamma\Delta t}, & \text{si } k = N. \end{cases} \quad (39)$$

Sustituyendo la expresión (39) en (32) y despejando los valores m_i podemos conseguir el número medio de etapas hasta la extinción de la epidemia, partiendo de cualquier cantidad inicial de infectados.

4. Ilustración numérica de los resultados teórico-algorítmicos obtenidos

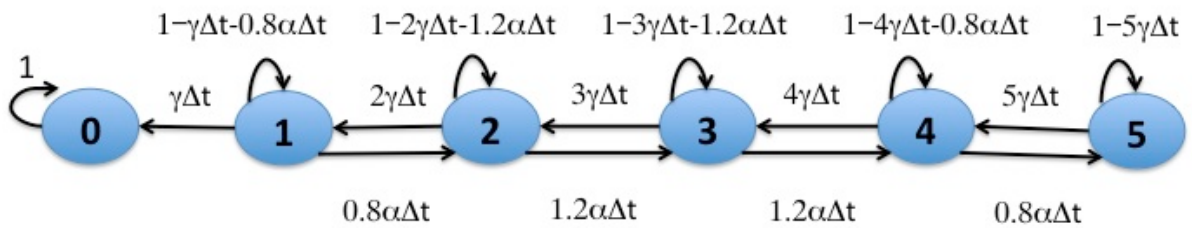
A continuación ilustraremos numéricamente el comportamiento probabilístico de la variable T . Esto incluye funciones de masa y momentos de la variable número de etapas hasta la erradicación de la epidemia. Recordamos que la función de masa de T condicionada a que partimos en el instante inicial con i infectados venía definida por

$$\alpha_i(n) = P\{T = n | I_0 = i\}, \forall i \geq 1, \forall n \geq 1.$$

El proceso de obtención de estas probabilidades acarrea en muchas ocasiones la necesidad de iterar el Algoritmo 1, un número de veces muy grande para poder satisfacer el criterio de parada elegido. Esto hace que el cálculo de la función de masa, para algunas combinaciones de parámetros, sea un proceso largo.

En la siguiente figura podemos ver el diagrama de transición entre los estados asociado a un modelo SIS con tamaño poblacional igual a 5 y parámetros α , Δt y γ genéricos.

Figura 7: Diagrama de transición.



A continuación se proponen dos posibles elecciones de parámetros y se comparan las matrices de transición y la distribución de T en ambos escenarios. En concreto consideramos:

- a) $\alpha = 2$, $\gamma = 1$ y $\Delta t = 0,01$.
- b) $\alpha = 20$, $\gamma = 10$ y $\Delta t = 0,01$.

En ambos se realizan 100 inspecciones de la población por unidad de tiempo y podemos ver que ambos comparten el mismo valor para el número básico de

reproducción $R_0 = 2$.

Las matrices de transición obtenidas para ambos casos son las siguientes:

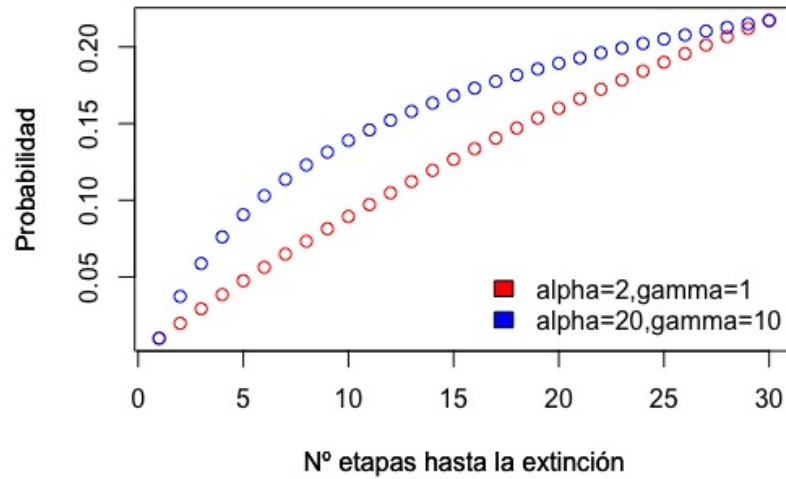
$$P_a = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0,01 & 0,974 & 0,016 & 0 & 0 & 0 \\ 0 & 0,02 & 0,956 & 0,024 & 0 & 0 \\ 0 & 0 & 0,03 & 0,946 & 0,024 & 0 \\ 0 & 0 & 0 & 0,04 & 0,944 & 0,016 \\ 0 & 0 & 0 & 0 & 0,05 & 0,95 \end{pmatrix}$$

$$P_b = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0,1 & 0,74 & 0,16 & 0 & 0 & 0 \\ 0 & 0,2 & 0,56 & 0,24 & 0 & 0 \\ 0 & 0 & 0,3 & 0,46 & 0,24 & 0 \\ 0 & 0 & 0 & 0,4 & 0,44 & 0,16 \\ 0 & 0 & 0 & 0 & 0,5 & 0,5 \end{pmatrix}$$

El tamaño de etapa seleccionado es $\Delta t = 0,01$, que es el máximo factible común a todas las combinaciones expuestas de nuestros parámetros tal y como se expuso en (18) y (19). Como podemos observar cuando multiplicamos ambos parámetros α y γ por la misma cantidad, en este caso en particular ha sido 10, las probabilidades de transición entre estados distintos también quedan multiplicadas por ese mismo valor.

En estos dos casos la tasa de contagio es el doble que la de recuperación y por lo tanto la extinción de la epidemia será un proceso largo. Las funciones de distribución de la variable aleatoria número de etapas hasta la extinción de la epidemia suponiendo que partimos de un único infectado inicial $I_0 = 1$ para cada uno de los casos anteriores tienen el siguiente aspecto:

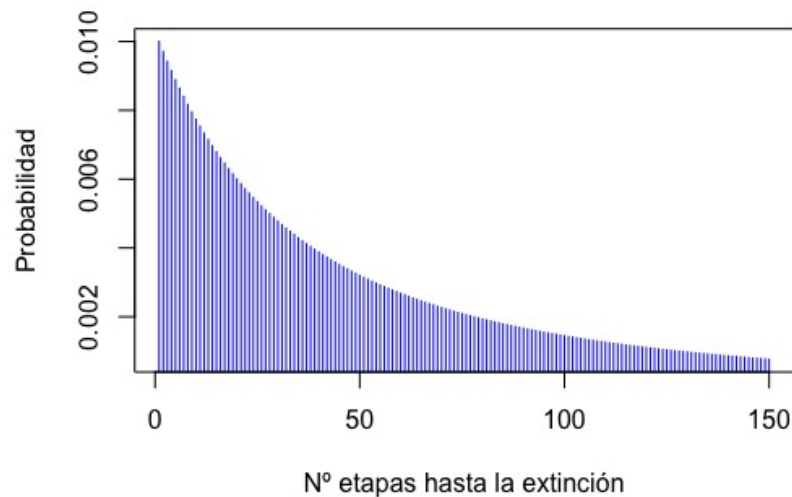
Figura 8: Funciones de distribución para $N = 5$.



En ambos casos el número de iteraciones requerido para acumular al menos el 99 % de probabilidad es muy elevado, por ello únicamente se muestra hasta la etapa 30.

A continuación mostraremos la distribución del tiempo hasta la extinción de la epidemia al variar la cantidad de individuos que integran la población:

Figura 9: Función de masa para $N = 50$.



En la Figura 9 representamos la función de masa asociada a la variable T cuando $N = 50$, $\alpha = 2$, $\gamma = 1$ y $\Delta t = 0,01$. Como podemos observar, la función de masa es una función monótona decreciente y ha sido representada únicamente hasta las primeras 150 etapas. Con esta combinación de parámetros se necesitan más de 30,000 etapas para poder acumular una probabilidad de al menos el 99% y por lo tanto la función de masa tendrá una larga cola a la derecha. Esto quiere decir que la extinción se puede dar en un número elevado de etapas.

Además a medida que aumentamos el tamaño poblacional el número de etapas requerido aumenta también. Esto queda ilustrado en la siguiente tabla, donde podemos ver las probabilidades acumuladas hasta la etapa n para distintos valores de N y manteniendo la misma elección para el resto de parámetros ($\alpha = 2$, $\gamma = 1$ y $\Delta t = 0,01$).

Tabla 1: Probabilidad de que la epidemia finalice antes de la observación n .

$P\{T \leq n\}$	$N = 5$	$N = 25$	$N = 50$	$N = 100$
$n = 1$	0,0100	0,0100	0,0100	0,0100
$n = 10$	0,0894	0,0882	0,0880	0,0879
$n = 20$	0,1599	0,1558	0,1553	0,1551
$n = 30$	0,2172	0,2093	0,2083	0,2078
$n = 50$	0,3047	0,2879	0,2858	0,2848
$n = 75$	0,3827	0,3536	0,3501	0,3483
$n = 100$	0,4400	0,3983	0,3932	0,3907
$n = 125$	0,4856	0,4302	0,4236	0,4204
$n = 150$	0,5226	0,4539	0,4458	0,4418

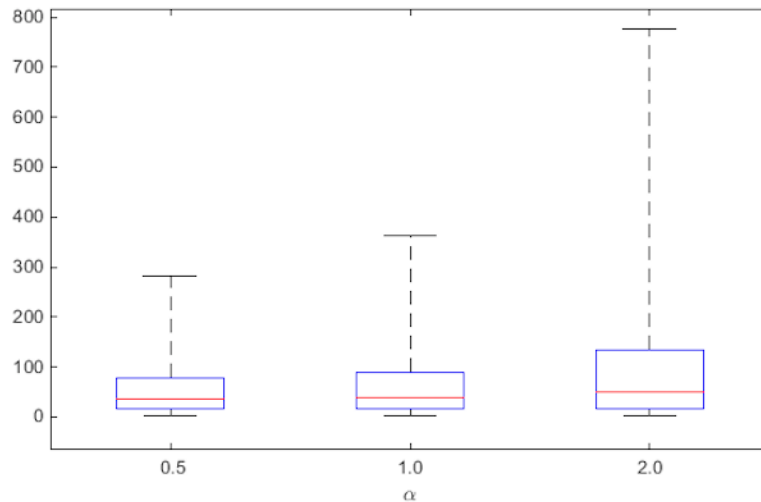
Tal y como se puede contemplar en la Tabla 1 a medida que aumentamos el tamaño poblacional acumulamos probabilidad más lentamente. En este caso en particular la extinción de la epidemia se produce en un número de etapas menor o igual a 1, con probabilidad igual a 0,01 en los 4 casos. En cambio al calcular las probabilidades acumuladas en etapas superiores se ve como esas probabilidades disminuyen

a medida que incrementamos el tamaño poblacional. En concreto para $N = 5$, acumulamos 0,5226 de probabilidad en la etapa 150, mientras que para $N = 100$ esa probabilidad es 0,4418.

El número de etapas necesarias hasta la extinción no solo se ve afectado por el incremento de tamaño poblacional, sino también por la combinación de las tasas de contagio y recuperación. Cuando tenemos tasas de contagio con valores más grandes que las de recuperación se requiere un mayor número de etapas hasta la erradicación. De igual manera, cuando vamos incrementando el valor de la tasa de contagio, manteniendo constante la tasa de recuperación, ese número de etapas también se ve aumentado.

Para ilustrarlo representaremos los diagramas de cajas de la distribución T para $N = 20$, $\gamma = 2$ y un tamaño de etapa de 0,01 combinándolos con los valores de $\alpha = 0,5, 1, 2$. El valor más alto en cada uno de los gráficos corresponde al cuantil que acumula un 99% de probabilidad.

Figura 10: Gráficos caja para $\gamma = 2$.

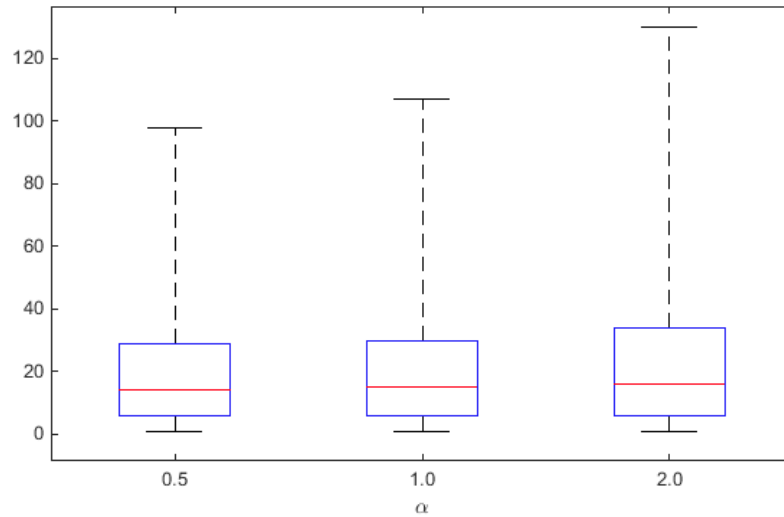


A la vista de los diagramas de la Figura 10 podemos ver como a medida que aumentamos la tasa de contagio, obtenemos una cola a la derecha cada vez mayor. Además las medianas se desplazan también a la derecha y esto es más acusado con el cuartil 3.

Por otra parte, en la Figura 11 presentamos los diagramas de cajas cuando la tasa de recuperación es $\gamma = 5$, observamos que las partes finales de las colas derechas de la distribución alcanzan valores más pequeños que si tomamos el caso anterior con $\gamma = 2$.

A grandes rasgos, el número de etapas hasta la extinción del brote epidémico, para $\alpha = 0,5$ no sobrepasan las 120 etapas mientras que en el anterior sobrepasaban las 200.

Figura 11: Gráficos caja para $\gamma = 5$.



Cuando hablamos del número medio de etapas hasta la erradicación partiendo de i infectados iniciales, $m_i = E[T|I_0 = i]$, el comportamiento al variar los distintos parámetros es similar.

En la Tabla 2 resumimos los valores del número medio de etapas transcurridas hasta la finalización de un brote epidémico, para distintas combinaciones de las tasas de contagio y recuperación. La población elegida tiene $N = 20$ individuos y la amplitud de etapa es $\Delta t = 0,01$ para todos los casos y un único infectado inicial. Los resultados obtenidos son los siguientes:

Tabla 2: Número medio de etapas hasta la extinción.

$E(T)$	$\alpha = 0,5$	$\alpha = 1$	$\alpha = 2$	$\alpha = 5$
$\gamma = 0,5$	444,73	6242,14	$1,294 \times 10^7$	$1,917 \times 10^{13}$
$\gamma = 1$	134,19	222,36	3121,07	$1,536 \times 10^8$
$\gamma = 2$	56,99	67,09	111,18	11325
$\gamma = 5$	21,01	22,16	25,02	44,47

Si comparamos el número medio de etapas hasta la extinción según la combinación de parámetros α y γ , podemos ver como estos valores aumentan según nos desplazamos de izquierda a derecha y de abajo a arriba. En concreto, el número medio de etapas durante las cuales la epidemia está activa aumenta al aumentar la tasa de contagio, α , y también cuando disminuye la tasa de recuperación γ . Lo que concuerda con la intuición, pues es esperable que la epidemia se observe durante más etapas cuando se produzcan más contagios o cuando los individuos tarden más tiempo en recuperarse.

Por ejemplo si $\alpha = 1$ y $\gamma = 1$ lo que viene siendo que un individuo infecta a una persona por unidad de tiempo hasta su recuperación, mientras que en esa unidad de tiempo hay una recuperación, tenemos que el número medio de etapas hasta la extinción con esa elección de 222,36, es decir el número esperado de etapas necesarias hasta la erradicación de la enfermedad partiendo de un único infectado es de 222,36.

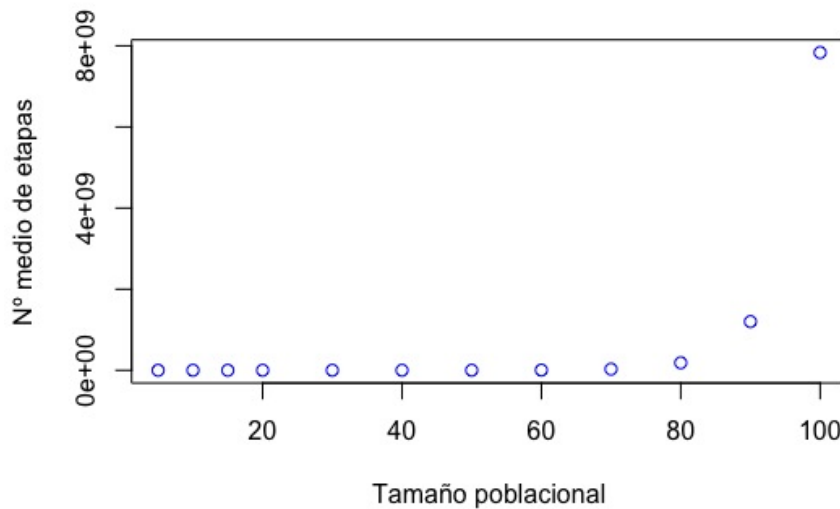
Si aumentamos γ a 5, obtenemos que el número esperado se reduce a 22,16 ya que ahora hay más recuperaciones por unidad de tiempo que antes. En cambio si modificamos α a 5 y mantenemos γ a 1 el valor obtenido de la esperanza es $1,536 \times 10^8$, ya que ahora hay muchísimos más contagios, para ser exactos 5 contagios más que recuperaciones por unidad de tiempo.

Como se puede observar el caso más favorable se da cuando tenemos la combinación de α 's bajos con γ 's altos; de hecho a la vista de nuestros datos esto se daría con valores de $\alpha = 0,5$ y $\gamma = 5$, los cuales nos dan una esperanza para T condicionada a $i_0 = 1$ de 21,01 etapas, si desplazamos hacia la derecha o hacia arriba en la tabla

veremos como esa esperanza vuelve a aumentar. Al igual el caso más desfavorable se produce con valores de $\alpha = 5$ y $\gamma = 0,5$ que proporciona un valor tan alto como $1,917 \times 10^{13}$.

Este promedio también se ve afectado, como era de esperar, por los distintos tamaños poblacionales. En el siguiente gráfico representamos el número medio de etapas hasta la extinción como función del tamaño de la población, en todos los casos $\Delta t = 0,01$, $\alpha = 1,6$ y $\gamma = 0,8$.

Figura 12: Valores esperados m_1 variando N con $\alpha = 1,6$ y $\gamma = 0,8$.



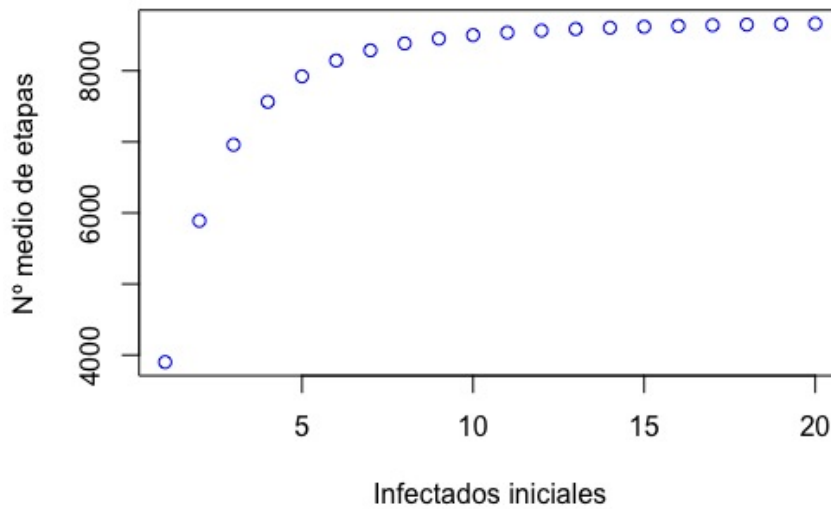
La función obtenida es creciente, con un comportamiento de tipo asintótico a medida que aumenta la cantidad de individuos. Como el promedio para $N = 100$ es bastante más grande que para valores inferiores a él, daremos los valores concretos de algunos de ellos para que se aprecie este aumento mejor.

Para un tamaño de $N = 5$, el momento de orden 1 es 368,36, para $N = 20$ es 3901,34, para $N = 50$ es de 728284,59 y para $N = 100$ el promedio obtenido es de 7834803778. Se puede ver que a medida que vamos aumentando la población estos promedios van creciendo, siendo este aumento mucho más acusado en valores superiores a 100.

Proseguimos observando la cantidad media de etapas que la infección permanece activa, pero teniendo en cuenta el número inicial de individuos infectados. Intuitivamente, al aumentar la cantidad inicial de infectados esperamos que la epidemia tenga una duración mayor.

Para ilustrarlo representaremos en la Figura 13 el número medio de etapas hasta la extinción como función de la cantidad inicial de infectados. Consideramos una población formada por 20 individuos, con parámetros $\alpha = 1,6$, $\gamma = 0,8$ y fijaremos etapas de amplitud $\Delta t = 0,01$.

Figura 13: Valores esperados m_i con $N = 20$, $\alpha = 1,6$ y $\gamma = 0,8$.



Observamos como a medida que aumenta la cantidad de infectados también crece la cantidad de etapas en las que se observa una epidemia activa. Este crecimiento es más acusado en los casos con pocos infectados iniciales y nos muestra la importancia de la detección temprana de las epidemias.

A continuación fijaremos $\gamma = 1$ para ver la variación que producen en la elección del tamaño de etapa y en la esperanza de la variable T distintas combinaciones de N y α .

Tabla 3: Amplitud y número esperado de etapas al variar N y α .

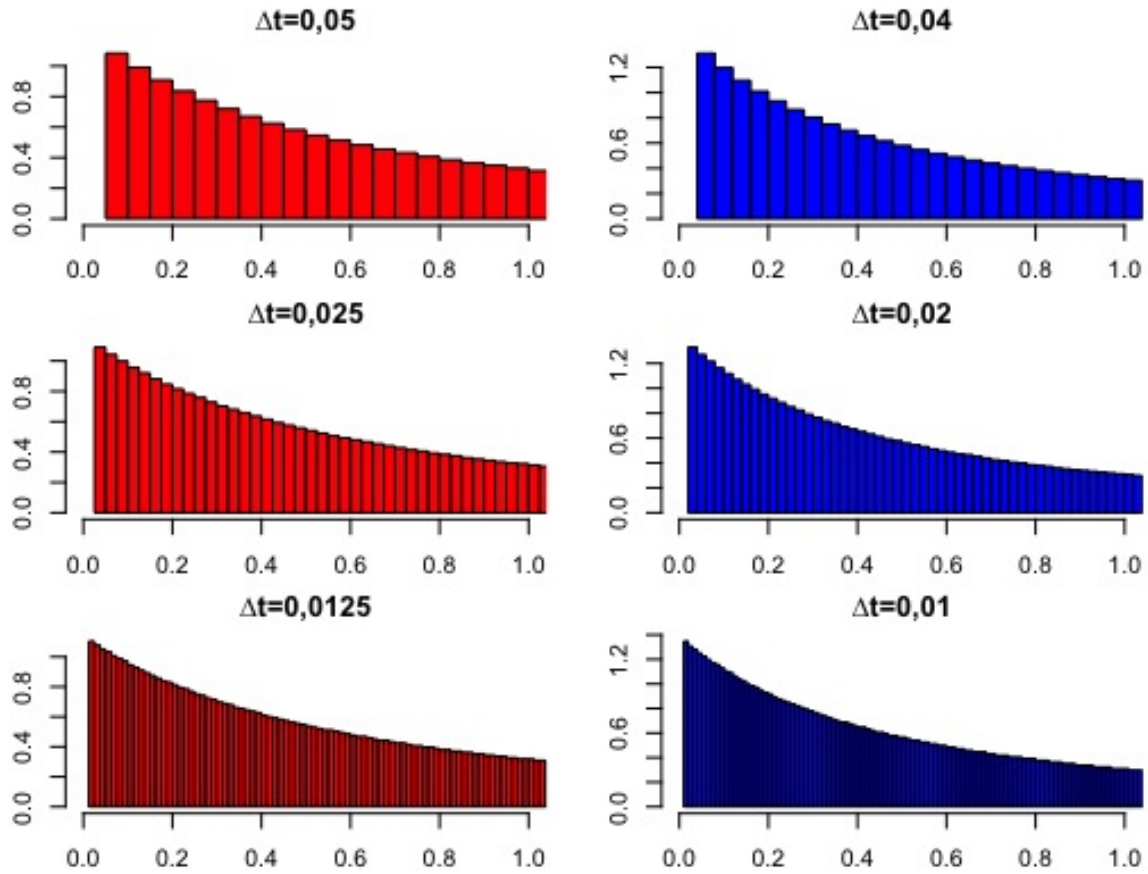
$\Delta t/E(T)$	$N = 5$	$N = 25$	$N = 50$	$N = 100$
$\alpha = 0,5$	0,2	0,04	0,02	0,01
	6,2324	33,751	68,3658	137,656
$\alpha = 1$	0,2	0,04	0,02	0,01
	8,0784	58,148	132,458	297,89
$\alpha = 2$	0,17	0,03	0,01	0,008
	17,334	2364,74	582627,67	$7,834 \times 10^9$
$\alpha = 5$	0,11	0,02	0,01	0,005
	161,81	$3,915 \times 10^9$	$3,371 \times 10^{18}$	$1,791 \times 10^{36}$

A la vista de los resultados de la Tabla 3 podemos decir que a medida que el tamaño poblacional crece también lo hace el número medio de etapas hasta la extinción, como ya habíamos comprobado antes. En cambio el tamaño de etapa se debe de tomar más pequeño para cumplir las hipótesis expuestas en (16).

Este resultado es razonable pues al aumentar la cantidad de individuos, los contactos aumentan y para que en la etapa de observación se produzca como mucho un contagio o recuperación debemos de tomar etapas cada vez más cortas.

Las variaciones del tamaño de etapa se ilustran en la siguiente figura, donde se ha considerado una población de $N = 20$ individuos y se compara la función de masa para dos tasas de contagio 0,75 y 1,25 respectivamente para tres amplitudes de etapa.

Figura 14: Funciones de masa al variar Δt .



Como podemos apreciar en las imágenes de la Figura 14, a medida que vamos disminuyendo el tamaño de etapa los saltos van disminuyendo proporcionalmente al valor Δt ; las funciones de masa se van aproximando a la función de densidad de la característica continua equivalente: el tiempo hasta la extinción de la epidemia.

5. Conclusiones

Sobre una población de tamaño moderado (por ejemplo una unidad de cuidados intensivos, una guardería...) hemos estudiado la evolución de una enfermedad contagiosa no inmunizante.

El modelo epidemiológico subyacente es el modelo SIS planteado desde el punto de vista estocástico. Si la epidemia tuviese lugar en una población de gran tamaño es preferible representar la evolución de la enfermedad mediante un modelo SIS determinista, ya que las diferencias individuales no afectan el comportamiento global de la epidemia.

En el caso determinista el conocimiento de R_0 , número reproductivo básico, juega un papel fundamental a la hora de cuantificar la epidemia. Definido R_0 como el cociente entre las tasas de contagio y recuperación, se observa que para $R_0 < 1$ la epidemia desaparece, y en otro caso la epidemia persiste en la población, sin llegar a desaparecer por completo. Si utilizáramos otra definición de R_0 lo anterior no tiene por qué ser cierto.

También se puede decir al respecto que aunque R_0 se defina como el número esperado de casos nuevos que producirá una persona infectada durante su período de contagio en una población de personas susceptibles a la enfermedad, el método de cálculo de R_0 no conduce a su valor exacto, ya que sobreestima el número de contagios porque no diferencia segundos contagios. Por este motivo en poblaciones moderadas es preferible el enfoque estocástico. En este tipo de modelos nos basamos en otros tipos de cuantificadores: R_{e0} y R_p . Ambos se definen como variables aleatorias y R_{e0} , a diferencia de R_0 , no tiene en cuenta esos contagios secundarios. Como mostramos en su momento cuando el tamaño poblacional es grande, este cuantificador R_{e0} tiende al número reproductivo básico R_0 . La variable R_p permite estudiar globalmente el impacto de toda la población; es decir el número exacto de contagios exitosos causados por todos los individuos infectados hasta que se produce la primera recuperación.

La particularidad de nuestro estudio es la hipótesis de que la población se supervisa a intervalos regulares. Esta condición implica la utilización de cadenas de

Markov en tiempo discreto para representar la evolución de la epidemia.

Al considerar una población finita y de tamaño constante (al menos durante el brote epidémico), las propiedades de las cadenas de Markov garantizan que todos los brotes epidémicos finalizan con total seguridad en un tiempo esperado finito. Es decir, en promedio hay una cantidad finita de inspecciones en las que se observa algún individuo infectado en la población.

Precisamente, la contribución original de este trabajo está relacionada con la cantidad de etapas (inspecciones) que transcurren hasta que la epidemia finaliza.

Esta variable aleatoria se ha estudiado analíticamente, se han presentado algoritmos que conducen a evaluar su distribución y características numéricas interesantes y se ha realizado un análisis numérico para observar la influencia de los parámetros del modelo: el tamaño de la población, las tasas de contagio y de recuperación, y la distancia entre las supervisiones.

Como hemos visto en la Sección 4, escenarios con tasa de contagio superior a la de recuperación dan lugar a funciones de masa con largas colas a la derecha y por lo tanto es muy probable que la epidemia tarde más en extinguirse que cuando la tasa de recuperación supera a la de contagio. Además, cuando se fijan las tasas de contagio y recuperación, a medida que aumentamos el tamaño poblacional también se incrementa el número de etapas requeridas hasta la finalización del brote epidémico. Este comportamiento es similar al estudiar los valores esperados condicionados a distintos infectados iniciales. Si mantenemos constantes todos los parámetros (N , α , γ y Δt) y vamos aumentando el número de infectados iniciales, el promedio de etapas requeridas hasta que la epidemia finaliza también se ve aumentado.

6. Anexos

1. Función que devuelve la matriz de transición dados los parámetros α , γ , N y etapa.

Función que devuelve el tamaño de tapa dados α , γ y N .

```
etapa<-function(a,g,N){  
  xo<-(N/2)*(1+(g/a))  
  fx<-((a*xo*(N-xo)/N)+xo*g)^(-1)  
  fN<-1/(N*g)  
  R<-a/g  
  if (R>=1){t<-fx}  
  if (R<1){t<-fN}  
  return(t)  
}
```

2. Función que devuelve la matriz de transición dados α , γ , N y etapa.

```
matriztransicion<-function(a,g,N,t){  
  P<-matrix(0,nrow = N+1,ncol = N+1)  
  for ( i in 1:(N+1)) {  
    for( j in 1:(N+1)){  
      if (j==i+1) {  
        P[i,j]<-(a/N)*(i-1)*(N-i+1)*t  
      }  
      if(j==i) {  
        P[i,j]<-1-(a/N)*(i-1)*(N-i+1)*t-(i-1)*g*t  
      }  
      if(j==i-1){  
        P[i,j]<-((i-1)*g*t)  
      }  
    }  
  }  
  return(P)  
}
```

3. Funciones que devuelve la función de masa de la variable T | $i_0=1$ hasta acumular el 99% o superar la etapa 150, dados α , γ , N y etapa a . Las condiciones de parada se modifican a lo largo del estudio.

`require(lattice)`

```
funcionalphas1<-function(a,g,N,t){
```

```

        if (j==i+1) {
            P[i,j]<-(a/N)*(i-1)*(N-i+1)*t
        }
        if(j==i) {
            P[i,j]<-1-(a/N)*(i-1)*(N-i+1)*t-(i-1)*g*t
        }
        if(j==i-1){
            P[i,j]<-((i-1)*g*t)
        }
    }
}
U<-matrix(0,nrow = N+1,ncol = 30000)
for (m in 1:(N+1)){
    U[m,1]<-0
}
U[1,1]<-1
n<-2
U[2,2]<-P[2,1]
while ((U[2,n]<0.99)&(n<=151)){
    n<-n+1

    for(i in 2:n){
        if (i< (N+1)){
            suma<-0
            suma2<-0
            for(j in (i-1):(i+1)){
                if (j==1){suma2<-0}
                if ((j!=1)&(j>(n-1))){suma2<-0}
                if((j!=1)&(j<=(n-1))){suma2<-P[i,j]*U[j,n-1]}
                suma<-suma+suma2
            }
            U[i,n]<-(P[i,1]+suma)
        }
    }
}
n<-n-1 #la parada se termina en n-1 pq n dentro del while vale n+1 q
es justo antes de ver si
#se cumple la condicion y no la cumple
cat("El criterio de parada se produce a las",n, "iteracciones, con ",
n-1," etapas")
#el vector siguiente son las alphas de longitud n pq incluye el 0 alph
a1(n)en R es alpha_1(n-1)teorico
alpha1<-vector(length = n)
alpha1[1]<-0
alpha1[2]<-U[2,2]
for (f in 3:n){
    alpha1[f]<-U[2,f]-U[2,f-1]
}

```

```
    return(alpha1)
}
```

4. Cálculo de la función de distribución y gráficos.

```
funcionmasaA<-funcionalphas1(a,g,N,t)
```

```
funcionmasaA<-funcionmasaA[-1]
```

```
funciondistribucionA<-cbind(1:150,cumsum(funcionmasaA))
```

```
plot(funciondistribucionA,type = "p",main = "funciones de Distribución  

N=5",sub="t=0.01",xlab = "nº etapas hasta la extinción",ylab = "probabi-  

lidad",col="red")
```

5. funcion que devuelve los momentos m_i (esperanza con i infectados i niciales).

```
funcionmomentos1<-function(a,g,N,t){
```

```
  R<-a/g
```

```
  #####
```

```
  #CÁLCULO DE LOS  $dk-mi$  ALGORITMO 2
```

```
  #####
```

```
  #forma directa del cálculo de la matriz de transición
```

```
  P<-matrix(0,nrow = N+1,ncol = N+1)
```

```
  for ( i in 1:(N+1)) {
```

```
    for( j in 1:(N+1)){
```

```
      if (j==i+1) {
```

```
        P[i,j]<-(a/N)*(i-1)*(N-i+1)*t
```

```
      }
```

```
      if(j==i) {
```

```
        P[i,j]<-1-(a/N)*(i-1)*(N-i+1)*t-(i-1)*g*t
```

```
      }
```

```
      if(j==i-1){
```

```
        P[i,j]<-((i-1)*g*t)
```

```
      }
```

```
    }
```

```
  }
```

#Me calculo todos los dk aplicando la fórmula descrita en el pdf.para los valores

```
#0<k<N y para  $d_N$  aplico la formula directa
```

```
d<-c()
```

```
for(k in 1:(N-1)){
```

```
  sumatotal<-0
```

```
  for(i in 0:(N-1-k)){
```

```
    suma2<-((R/N)^i)*(1/(k+i))*((factorial(N-k))/(factorial(N-k-i)))
```

```
    sumatotal<-sumatotal+suma2
```

```
  }
```

```

sumatotal2<-((g*t)^(-1))*sumatotal
sumatotal3<-((R/N)^(N-k))*((factorial(N-k))/(N*g*t))
d[k]<-sumatotal3+sumatotal2

}
d[N]<-(P[N+1,N])^(-1)

#Me calculo Los momentos a partir de Los dk.
m<-c()
m[1]<-d[1]
for (j in 2:N){
  m[j]<-d[j]+m[j-1]
}
return(m)
}

```

6. Función que dibuja un box-plot a partir de una función de probabilidad.

```

box<-function(a){
  datos<-c()
  for(i in 1:length(a)){
    veces<-10000*a[i]
    x<-rep(i,veces)
    datos<-c(datos,x)
  }
  return(datos)
}

```

7. Simulación de un modelo SIS y gráfico conocida la matriz P.

```

set.seed(112)
set.seed(788786)
set.seed(986756)
set.seed(93796754)
i0<-1
movimientos<-i0
infectados<-i0
infec<-0
while ( infectados > 0) {
  if(infectados==N){
    u<-runif(1,0,1)
    if (u<=P[infectados+1,infectados]) {
      cat("recuperado ")
      infec<-(infectados-1) }
    if((P[infectados+1,infectados]<u)&(P[infectados+1,infectados+1]+P[infectados+1,infectados]>u)){

```

```
    cat("nada ")
    infec<-infectados}
    infectados<-infec
    movimientos<-c(movimientos,infectados)
  }

  if(infectados<N){
    u<-runif(1,0,1)

    if (u<=P[infectados+1,infectados]) {
      cat("recuperado ")
      infec<-(infectados-1)}
    if((P[infectados+1,infectados]<u)&(P[infectados+1,infectados]+P[infec
tados+1,infectados+1]>u)){
      cat("nada ")
      infec<-infectados
    }
    if(u>=P[infectados+1,infectados]+P[infectados+1,infectados+1]){
      cat("infectado ")
      infec<-(infectados+1)}
    infectados<-infec
    movimientos<-c(movimientos,infectados)
  }
  movimientos
}  #final del while

trayectoria<-cbind(0:(length(movimientos)-1),movimientos)
plot(trayectoria,type = "l",xlab = "nº de etapa",ylab = "nºde infectado
s")
lines(trayectoria,col="red")
lines(trayectoria,col="blue")
lines(trayectoria,col="green")
```

7. Bibliografía

- [1] Bucaille M. *La Bible, le Coran et la science*. Editions Seghers. París; 1987.
- [2] Martínez C. Vigilancia Epidemiológica: Una herramienta que da respuestas a los problemas de Salud Pública. *Pediatría*. 2013 Abril; 40(1): 9.
- [3] Ruiz-Guzman J. Historia de las estadísticas de la salud. *Gaceta Médica Boliviana*. 2006; 29(2): 72-77.
- [4] López-Moreno S, Garrido-Latorre F, Hernández-Avila M. Desarrollo histórico de la epidemiología: su formación como disciplina científica. *Salud pública México*. 2000; 42(2): 133-143.
- [5] Velasco-Hernández JX. Modelos Matemáticos en epidemiología: enfoques y alcances. *Miscelánea Matemática*. 2007; 44: 11-27.
- [6] Kemeny JG, Laurie-Snell L. *Finite Markov Chains*. Van Nostrand. Princeton, NJ; 1960.
- [7] Karlin S, Taylor HM. *A First Course in Stochastic Processes*. Academic Press, 2nd Edition. York; 1975.
- [8] Schinazi RB. *Classical and Spatial Stochastic Processes*. Birkhäuser, Boston; 1999.
- [9] Stewart WJ. *Introduction to the Numerical Solution of Markov Chains*. Princeton University Press. Princeton; 1994.
- [10] Böckh R. *Statistisches FahrBuch der Stadt Berlin, Zxölfter Jahrgang*. Statistik des Jahres. P. Stankieewicz. Berlín; 1886. p. 30-31.
- [11] Heesterbeek JAP, Dietz K. The concept of R_0 in epidemic theory. *Statistica Neerlandica*. 1996; 50(1): 89-110. doi:10.1111/j.1467-9574.1996.tb01482
- [12] Hethcote HW. The Mathematics of Infectious Diseases. *SIAM Review*. 2000; 42(4): 599-653.
- [13] Ridenhour B, Kowalik JM, Shay DK. Unraveling R_0 : Considerations for Public Health Applications. *American Journal of Public Health*. 2014; 104(2): 32-41. doi: 10.2105/AJPH.2013.301704

- [14] Número básico de reproducción r0.svg. Wikimedia Commons contributors; 2016. Disponible en: https://commons.wikimedia.org/w/index.php?title=File:N%C3%BAmero_b%C3%A1sico_de_reproducci%C3%B3n_r0.svg&oldid=224468872
- [15] Heffernan JM, Smith RJ, Wahl LM. Perspectives on the Basic Reproductive Ratio. *Journal of the Royal Society Interface*. 2005; 2(4): 281-293.
- [16] Artalejo JR, López-Herrero MJ. The SIS and SIR stochastic epidemic models: a maximum entropy approach. *Theoretical Population Biology*. 2011; 80(4): 256-264.
- [17] Hotta LK. Bayesian melding estimation of a stochastic SEIR model. *Mathematical Population Studies*. 2010; 17(2): 101-111.
- [18] Artalejo JR, López-Herrero MJ. On the Exact Measure of Disease Spread in Stochastic Epidemic Models. *Bulletin of Mathematical Biology*. 2013; 75(7): 1031-1050.
- [19] Ciarlet PG. *Introduction to numerical linear algebra and optimisation*. Cambridge University Press. Cambridge; 1989.
- [20] Kulkarni VG. *Modeling and analysis of stochastic systems*. Chapman and Hall. Londres; 1995.
- [21] Allen LJS. *An introduction to stochastic processes with applications to biology*. Prentice-Hall. Englewood Cliffs; 2003.
- [22] Artalejo JR, López-Herrero MJ. Quasi-stationary and ratio of expectations distributions: A comparative study. *Journal of Theoretical Biology*. 2010; 266(2): 264-74.
- [23] Allen LJS, Burgin AM. A comparison of deterministic and stochastic SIS and SIR model in discrete time. *Mathematical Biosciences*. 2000; 163(1): 1-33.