



Centro de Investigación en Matemáticas, A.C.

---

---

CIMAT

# Inferencia Bayesiana en el Modelo SIR

**T E S I S**

Que para obtener el grado de

**Maestro en Ciencias**

con Especialidad en

**Probabilidad y Estadística**

P r e s e n t a

**Juan Carlos Durán Aguilar**

Director de Tesis:

**Dr. José Andrés Christen Gracia**

Guanajuato, Gto. Septiembre de 2014



**CIMAT**  
CENTRO DE INVESTIGACION  
EN MATEMÁTICAS A. C.

# Centro de Investigación en Matemáticas, A.C.

## Acta de Examen de Grado

Acta No.: 074  
Libro No.: 002  
Foja No.: 074

En la Ciudad de Guanajuato, Gto., siendo las 13:00 horas del día 12 de septiembre del año 2014, se reunieron los miembros del jurado integrado por los señores:

**DR. MARCOS AURELIO CAPISTRÁN OCAMPO** (CIMAT)  
**DRA. MARÍA DE LOS DOLORES SÁNCHEZ CASTAÑEDA** (CIMAT)  
**DR. JOSÉ ANDRÉS CHRISTEN GRACIA** (CIMAT)

bajo la presidencia del primero y con carácter de secretario el segundo, para proceder a efectuar el examen que para obtener el grado de

**MAESTRO EN CIENCIAS  
CON ESPECIALIDAD EN PROBABILIDAD Y ESTADÍSTICA**

sustenta

**JUAN CARLOS DURÁN AGUILAR**

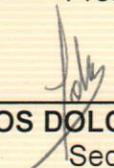
en cumplimiento con lo establecido en los reglamentos y lineamientos de estudios de posgrado del Centro de Investigación en Matemáticas, A.C., mediante la presentación de la tesis

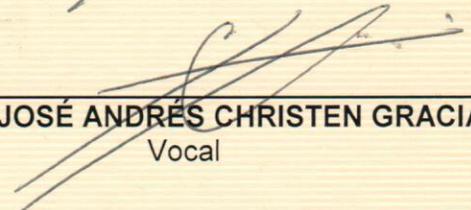
**“INFERENCIA BAYESIANA EN EL MODELO SIR”.**

Los miembros del jurado examinaron alternadamente al (la) sustentante y después de deliberar entre sí resolvieron declararlo (a):

APROBADO

  
\_\_\_\_\_  
**DR. MARCOS AURELIO CAPISTRÁN OCAMPO**  
Presidente

  
\_\_\_\_\_  
**DRA. MARÍA DE LOS DOLORES SÁNCHEZ CASTAÑEDA**  
Secretario

  
\_\_\_\_\_  
**DR. JOSÉ ANDRÉS CHRISTEN GRACIA**  
Vocal



**CIMAT**  
DIRECCION  
GENERAL

  
**DR. JOSÉ ANTONIO STEPHAN DE LA PEÑA MENA**  
Director General

CENTRO DE INVESTIGACIÓN EN  
MATEMÁTICAS, A.C.

TESIS DE MAESTRÍA

---

# Inferencia Bayesiana en el Modelo SIR.

---

*Autor:*

JUAN CARLOS DURÁN  
AGUILAR

*Director de Tesis:*

DR. JOSÉ ANDRÉS  
CHRISTEN GRACIA

*Tesis sustentada de conformidad con los requerimientos  
para el grado de Maestría en Ciencias*

*con Especialidad en*

Probabilidad y Estadística

*“La ciencia es la progresiva aproximación del hombre al mundo real.”*

Max Planck

CIMAT

# *Resumen*

Centro de Investigación en Matemáticas, A.C.

Probabilidad y Estadística

Maestría en Ciencias

## **Inferencia Bayesiana en el Modelo SIR.**

por JUAN CARLOS DURÁN AGUILAR

Las epidemias han tenido un impacto significativo en la historia de la humanidad, por lo que el conocimiento oportuno de los parámetros que producen los reportes de infectados puede ser de gran utilidad para la rápida y eficaz implementación de un plan de acción para atenuar o aplicar la respuesta más pertinente para combatir las epidemias.

Es costumbre utilizar la función de verosimilitud para identificar los parámetros que mejor describan o se acerquen a la información observada; no obstante, en ocasiones la forma analítica de esta expresión resulta intratable siquiera computacionalmente.

Se plantea el problema desde su definición más básica, en la cual se tiene información perfecta y se realiza la inferencia los parámetros deseados, se procederá a hacer supuestos acerca de información incompleta y realizar diversas pruebas de inferencia utilizando técnicas de Computación Bayesiana Aproximada o por sus siglas en inglés ABC (Approximate Bayesian Computation) variando: El estadístico utilizado para resumir la información observada, la función Kernel utilizada para la evaluación de los estadísticos y el método Monte Carlo que permita obtener la muestra de la distribución aproximada.

Se desarrollan distintos ejercicios de simulación y se compara entre las metodologías para determinar la opción más eficiente computacionalmente para aplicarlos a datos reales en un caso de Influenza española en San Francisco en 1918. Se realiza además, un análisis exhaustivo sobre los aspectos, ventajas y desventajas de cada una de las metodologías.

# *Agradecimientos*

A la vida, porque me ha provisto de todo lo que me ha llevado a ser la persona en la que me he convertido. Como un proceso estocástico, soy resultado de la suma de una infinidad de eventos y decisiones que me han moldeado.

A mis padres y mi hermano, porque me han dado las herramientas para salir adelante en la vida, por enseñarme el valor del trabajo y por haber soportado mi ausencia.

A CIMAT por darme la oportunidad de demostrar que no hay límite de edad, por enseñarme humildad y porque aprendí mucho más de lo que inicialmente imaginé.

A CONACYT por el apoyo No. 344786, sin el cual, mi estancia en Guanajuato no hubiera sido posible.

A mi director de Tesis, el Dr. Andrés Christen por permitirme trabajar con él, guiarme y aguantar mis múltiples dudas.

A mis sinodales, por permitirme robarles un poco de su tiempo.

A mis profesores, porque me enseñaron más allá de las materias.

A mis amigos y compañeros, por el apoyo, alegrías y vivencias que pasamos juntos, me hubiera gustado convivir más, pero sé que nos estaremos viendo y que la amistad que hemos formado perdurará.

A los que creyeron en mí, porque me alentaron a seguir adelante.

---

# ÍNDICE GENERAL

<b>Resumen</b>	<b>II</b>
<b>Agradecimientos</b>	<b>III</b>
<b>Contenido</b>	<b>IV</b>
<b>Abreviaciones</b>	<b>VII</b>
<b>Introducción</b>	<b>1</b>
<b>1. Modelo SIR</b>	<b>4</b>
1.1. Introducción . . . . .	4
1.2. Modelo de propagación de infecciones SIR . . . . .	4
1.2.1. Supuestos Básicos de la Propagación de la Enfermedad. . . . .	6
1.3. Modelo SIR Determinístico . . . . .	6
1.4. Modelo SIR Estocástico . . . . .	9
1.4.1. Procesos Estocásticos, la Ecuación Maestra y el Algoritmo Gillespie . . . . .	10
1.4.2. Ecuación maestra. . . . .	10
1.4.3. Algoritmo Gillespie . . . . .	12
1.4.4. Ejemplo con datos sintéticos . . . . .	13
<b>2. Inferencia Bayesiana sin Verosimilitud: ABC</b>	<b>16</b>
2.1. Introducción . . . . .	16
2.2. Inferencia libre de Verosimilitud . . . . .	17
2.2.1. El Estadístico . . . . .	18
2.2.2. Estimación Kernel . . . . .	19
2.3. Esquema General para Estimación Libre de Verosimilitud . . . . .	21

2.4.	Inferencia sin Verosimilitud: ABC . . . . .	22
2.4.1.	Muestreo de Aceptación Exacto . . . . .	23
2.4.2.	Muestreo de Aceptación Aproximado: ABC Clásico . . . . .	23
2.4.3.	Algoritmo Metropolis-Hastings . . . . .	24
2.5.	MCMC sin Verosimilitud: ABC - MCMC . . . . .	25
2.6.	Método Monte Carlo Secuencial libre de Verosimilitud: ABC - SMC . . . . .	26
<b>3.</b>	<b>Inferencia en el modelo SIR</b>	<b>28</b>
3.1.	Inferencia con información perfecta en el modelo SIR . . . . .	28
3.1.1.	Inferencia con enfoque clásico . . . . .	28
3.1.2.	Inferencia con enfoque bayesiano . . . . .	29
3.2.	Inferencia en el modelo SIR con información parcial . . . . .	29
3.3.	Características del modelo . . . . .	30
3.4.	Consideraciones . . . . .	31
3.4.1.	Estadístico . . . . .	32
3.4.2.	Estimación Kernel y Ancho de Banda . . . . .	32
3.4.3.	Método de simulación de la muestra . . . . .	33
3.5.	Implementación del Método de Inferencia . . . . .	34
3.5.1.	Recursos para la Simulación . . . . .	35
3.6.	Muestreo de Aceptación . . . . .	35
3.6.1.	Rejection Sampling-Kernel Uniforme-Estadístico Univariado . . . . .	36
3.7.	Caminata Aleatoria con Metropolis-Hastings: MCMC ABC . . . . .	42
3.7.1.	El caso Multidimensional . . . . .	43
3.7.2.	RWMH-Kernel Uniforme-Estadístico Multivariado . . . . .	44
3.7.3.	RWMH-Kernel Epanechnikov-Estadístico Multivariado . . . . .	46
3.7.4.	RWMH-Kernel Normal-Estadístico Multivariado . . . . .	47
3.7.5.	Caso del Estadístico Univariado . . . . .	49
3.7.5.1.	RWMH-Kernel Uniforme-Estadístico Univariado . . . . .	49
3.7.5.2.	RWMH-Kernel Epanechnikov-Estadístico Univariado . . . . .	51
3.7.5.3.	RWMH-Kernel Normal-Univariado . . . . .	53
3.8.	ABC SMC . . . . .	60
3.9.	Comparación entre Métodos . . . . .	62
<b>4.</b>	<b>Inferencia con Datos Reales</b>	<b>64</b>
4.1.	Inferencia de los parámetros con Datos Reales . . . . .	64
<b>5.</b>	<b>Conclusiones</b>	<b>71</b>
5.0.1.	Datos de Epidemias . . . . .	72



---

# ABREVIACIONES

$N$	Población total.
$S$	Cantidad de Individuos susceptibles.
$I$	Cantidad de Individuos infecciosos.
$R$	Cantidad de Individuos removidos
$s$	Proporción de Individuos susceptibles $S/N$ .
$i$	Proporción de Individuos infecciosos $I/N$ .
$r$	Proporción de Individuos removidos $R/N$ .
$\beta$	Probabilidad de contacto
$\lambda$	Fuerza de infección per cápita, $\lambda = \beta I$
$\gamma$	Tasa de recuperación.
$1/\gamma$	Tiempo promedio de infección.
$\mu$	Tasa promedio de defunciones
$R'_0$	Tasa de reproducción básica $R'_0 = \beta/\gamma$ .
$y$	Datos observados.
$x$	Simulación de datos (pseudo-observaciones).
$\theta = (\beta, \gamma)$	Parámetros del modelo .

*A mis padres*

---

# INTRODUCCIÓN

Las epidemias han tenido un impacto significativo en la historia de la humanidad, pues han afectado directamente la economía, la demografía, la política. Influyendo de manera importante el curso de las guerras, las creencias e incluso en el comportamiento religioso. Ejemplos de estos se tienen desde la peste negra, siendo ésta la pandemia más devastadora de la historia de la humanidad la cual afectó durante el siglo XIV; la viruela que fue una catástrofe demográfica en América con la llegada de los europeos; así, existen muchos ejemplos y entre los más recientes, están los causados por los virus de influenza.

Los modelos matemáticos son una herramienta indispensable para el entendimiento y la representación del comportamiento de los fenómenos naturales, en particular, en este trabajo se está interesado en la propagación de una enfermedad. Es importante notar que se trata de sistemas dinámicos, los cuales evolucionan a través del tiempo y que además no son exactos, pues tienen un componente aleatorio en la ocurrencia de los eventos.

El conocimiento oportuno de los parámetros que rigen el fenómeno puede ser de gran utilidad para la rápida y eficaz implementación de un plan de acción para atenuar o aplicar la respuesta más pertinente para combatir las epidemias.

El modelo SIR es de los más comunes para modelar el fenómeno de las epidemias, este modelo se ha atacado de diversas formas, desde el método determinista resolviendo una serie de ecuaciones diferenciales, hasta la utilización de procesos

estocásticos con diversas consideraciones que van haciendo más complejo el modelo al suponer la interacción de más variables involucradas. A grandes rasgos, el modelo SIR presenta una población cerrada en la que se presenta una enfermedad transmisible y la población se divide en tres segmentos, aquellos susceptibles de contagio, los que se han contagiado y aquellos que después de estar contagiados han adquirido inmunidad o han muerto a causa de la enfermedad. Los tiempos de infección se suponen como variables aleatorias y se desea conocer la evolución de la enfermedad a través del tiempo.

El primer problema que se presenta al modelar este tipo de fenómeno es el desconocimiento de la información completa del proceso, ya que no se conocen los tiempos exactos de ocurrencia de las infecciones.

Si se conociera el momento exacto de la ocurrencia de cada uno de los eventos y su tipo, ya sea infección o remoción, el problema se reduce en gran medida y se trata de un problema clásico de inferencia bayesiana. Sin embargo, la estimación se complica debido a que se conoce información parcial de individuos infectados en tiempos determinados.

Es costumbre utilizar la función de verosimilitud para identificar los parámetros que mejor describan o se acerquen a la información observada; no obstante, en ocasiones la forma analítica de esta expresión resulta intratable siquiera computacionalmente. La utilización de un espacio de tiempo continuo lleva a considerar una infinidad de posibilidades entre dos datos en dos instantes de tiempo; por lo que se considerarán variaciones de técnicas novedosas para trabajar el problema cuando no se tiene una función de verosimilitud. Otro problema que se presenta es establecer o fijar el punto inicial en el tiempo a partir del cual se inicia la epidemia, puesto que es esencial para la modelación.

El problema se complica cuando se tienen dos enfermedades que compiten dentro de una misma población y la incidencia de una puede generar inmunidad en ambas enfermedades. Actualmente existen diversas variaciones del modelo SIR, para un caso específico se plantea el problema de inferir los parámetros tanto de contagio como de recuperación.

En el presente trabajo se desarrollan diversos programas computacionales utilizando software libre como Python para programar los algoritmos que calculen la distribución de los parámetros, se opta por esta opción al ser un lenguaje numéricamente poderoso y eficiente para el cálculo de un gran número de simulaciones.

Se plantea el problema desde su definición más básica, en la cual se tiene información perfecta y se realiza la inferencia los parámetros deseados, se procederá a hacer supuestos acerca de información incompleta y realizar diversas pruebas de inferencia utilizando técnicas de Computación Bayesiana Aproximada o por sus siglas en inglés ABC (*Approximate Bayesian Computation*) variando: El estadístico utilizado para resumir la información observada, la función Kernel utilizada para la evaluación de los estadísticos y el método Monte Carlo que permita obtener la muestra de la distribución aproximada.

El objetivo del primer capítulo es mostrar el planteamiento del problema básico para la inferencia del modelo SIR en el cual se describen a grandes rasgos los componentes involucrados y la teoría desarrollada para su tratamiento.

El segundo capítulo explicará las herramientas teóricas que se utilizarán para el desarrollo de las simulaciones y las ventajas y dificultades de aplicar dichas técnicas.

En el capítulo 3 y 4 se desarrollan e implementan los métodos del capítulo 2 para la estimación de los parámetros del modelo SIR. Donde en el capítulo 3 se hará la inferencia en el caso de una enfermedad utilizando datos sintéticos y en el capítulo 4 se hará utilizando datos reales de la pandemia de influenza española en la ciudad de San Francisco en el año de 1918.

---

---

# CAPÍTULO 1

---

## MODELO SIR

### 1.1. Introducción

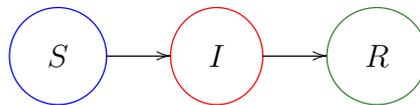
Los modelos matemáticos ayudan a comprender el comportamiento de los fenómenos que nos rodean y buscan explicar de manera práctica las leyes que los gobiernan, así como predecir el comportamiento de dichos fenómenos; en cuanto a las ciencias de la salud, estos modelos son de gran importancia, ya que permiten la evaluación de los factores de riesgo dentro de una población que pueden tener un efecto considerable en la mortalidad, razón por la cual han sido estudiados extensamente.

### 1.2. Modelo de propagación de infecciones SIR

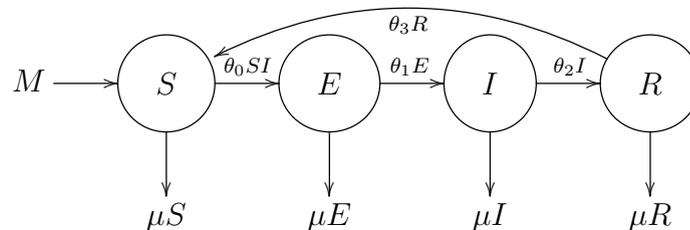
El modelo de propagación de enfermedades SIR (Susceptible-Infecioso-Removido) con el cual se trabajará se centra en determinar la evolución en el tiempo de una enfermedad específica en una población donde el total de habitantes se mantiene constante durante el periodo de análisis. Dicha población se divide en tres grupos ajenos: aquellos individuos libres de la enfermedad que son susceptibles de adquirirla ( $S$ ); los individuos que se han contagiado y que pueden transmitir la

enfermedad al entrar en contacto con algún individuo susceptible ( $I$ ); y los individuos que tuvieron la enfermedad pero que se recuperaron y adquirieron inmunidad ante la enfermedad se identifican como removidos ( $R$ ).

El siguiente diagrama muestra el flujo que tiene cualquier individuo en dicho proceso de infección.



Este modelo es de los más sencillos para epidemias y se deriva de un conjunto más general, donde se divide a la población en estados mutuamente ajenos y que puede determinarse por el modelo  $MSEIRS$  ( $M \rightarrow S \rightarrow E \rightarrow I \rightarrow R \rightarrow S$ )



donde

- $M$  = Nacidos con inmunidad temporal.
- $S$  = Susceptibles. Libres de la enfermedad pero que pueden contraerla.
- $E$  = Expuestos. Tienen la enfermedad pero no son capaces de infectar.
- $I$  = Infecciosos. Tienen la enfermedad y son capaces de infectar.
- $R$  = Recuperados. Adquieren inmunidad temporal ante la enfermedad.
- $\theta_i$  = tasas de incidencia.
- $\mu$  = tasa de mortalidad.

Adicionalmente, existen variantes en el modelo, como tiempos de retardo entre las transiciones de estados, tasas de nacimiento, mortalidad natural de la población ( $\mu$ ), efectos de vacunación e incluso la interacción entre más de un tipo de enfermedad.

El caso que se estudiará del modelo *SIR* supone una población libre de entradas o salidas por nacimientos, mortalidad o migración, con lo cual se definen los estados  $S =$  Susceptibles,  $I =$  Infecciosos,  $R =$  Removidos, en cada uno de estos estados  $S(t) = S_t =$  Número de individuos susceptibles de la población al tiempo  $t$ , análogamente para los individuos infecciosos y los removidos, de tal manera que  $S_t + I_t + R_t = N =$  población total, la cual permanece fija a lo largo del tiempo.

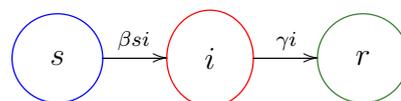
### 1.2.1. Supuestos Básicos de la Propagación de la Enfermedad.

Para establecer los lineamientos que regirán el modelo se establecen una serie de supuestos:

- 1 Los individuos susceptibles no son inmunes a la enfermedad.
- 2 Los individuos infectados dispersan la enfermedad hacia los susceptibles.
- 3 El estado de remoción asume inmunidad de por vida.
- 4 Para tener un nuevo infectado, es necesario tener al menos un individuo infectado y uno susceptible.
- 5 Se establecen condiciones iniciales para que el proceso tenga efecto, tanto susceptibles como infecciosos al tiempo cero, son mayores que cero y por comodidad se asume que los removidos son cero.
- 6 Supone una población grande, homogénea y distribuida uniformemente, de tal manera que cualquiera dos individuos puedan entrar en contacto.

## 1.3. Modelo SIR Determinístico

El modelo inicialmente propuesto por O. Kermack y A. G. McKendrick [12] en 1927 para estudio de la evolución de la propagación de enfermedades se representa en el siguiente diagrama.



Este modelo supone que la población es homogénea en el sentido de que cualquiera dos individuos tienen la misma probabilidad de entrar en contacto  $\beta$ , por lo que la razón de cambio de la población susceptible se ve disminuída por dicho parámetro, así como la proporción de individuos infectados y de los susceptibles como lo expresa la siguiente ecuación:

$$\frac{ds}{dt} = -\beta si \quad (1.1)$$

De la misma manera, se supone que pasado un tiempo, un individuo infectado adquiere inmunidad o muere a causa de la enfermedad, la tasa  $\gamma$  se aplica a la proporción de la población infectada, resultando la siguiente ecuación:

$$\frac{dr}{dt} = \gamma i \quad (1.2)$$

Finalmente, la razón de cambio de los individuos infectados varía conforme cambian la de los susceptibles y los removidos de acuerdo con la siguiente ecuación.

$$\frac{di}{dt} = \beta si - \gamma i \quad (1.3)$$

Las expresiones 1.1, 1.2 y 1.3 determinan un sistema no lineal de ecuaciones diferenciales y por consiguiente, no admite una solución analítica genérica, pero se rescata el término  $R'_0$  conocido como el *número básico de reproducción* y que es resultado de la razón siguiente.

$$R'_0 = \frac{\beta}{\gamma} \quad (1.4)$$

La interpretación de éste es que representa el número esperado de casos nuevos que producirá una persona infectada durante su periodo de contagio, y es de gran importancia, pues ayuda a determinar si una infección se propagará en una

población, cuando  $R'_0 < 1$  la enfermedad se extinguirá eventualmente; pero cuando es mayor que uno, la enfermedad puede ser capaz de propagarse en la población.

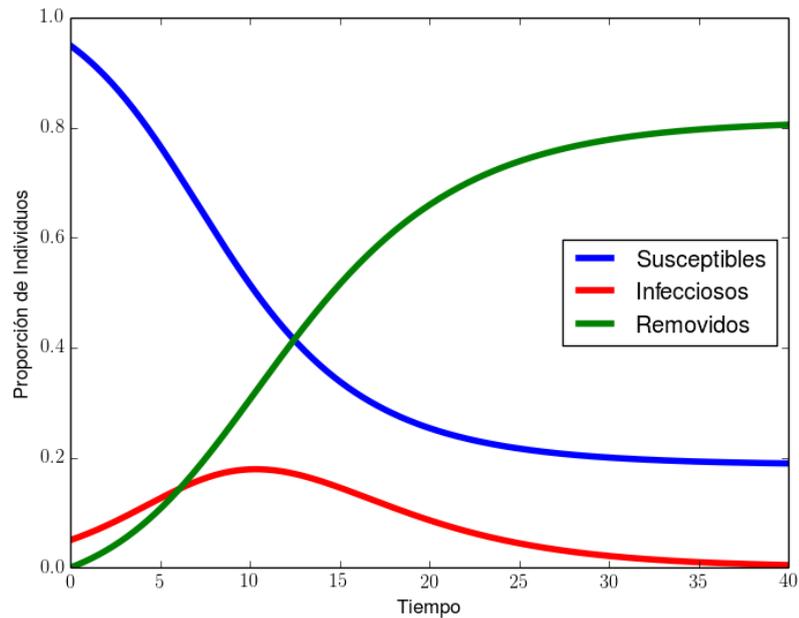


FIGURA 1.1: Modelo SIR Determinístico con parámetros  $\beta = 0.5$ ,  $\gamma = 0.25$  y condiciones iniciales  $(s_0, i_0, r_0)' = (0.95, 0.05, 0.0)'$

En la Figura 1.2 se observa el efecto que tiene el variar cada parámetro alrededor de un 20% por separado conservando el resto de las condiciones del modelo. Se puede percibir que el efecto del parámetro de infección tiene un efecto más violento en la rapidez con la que se propaga la infección, el efecto del parámetro  $\gamma$  acentúa la duración de la enfermedad.

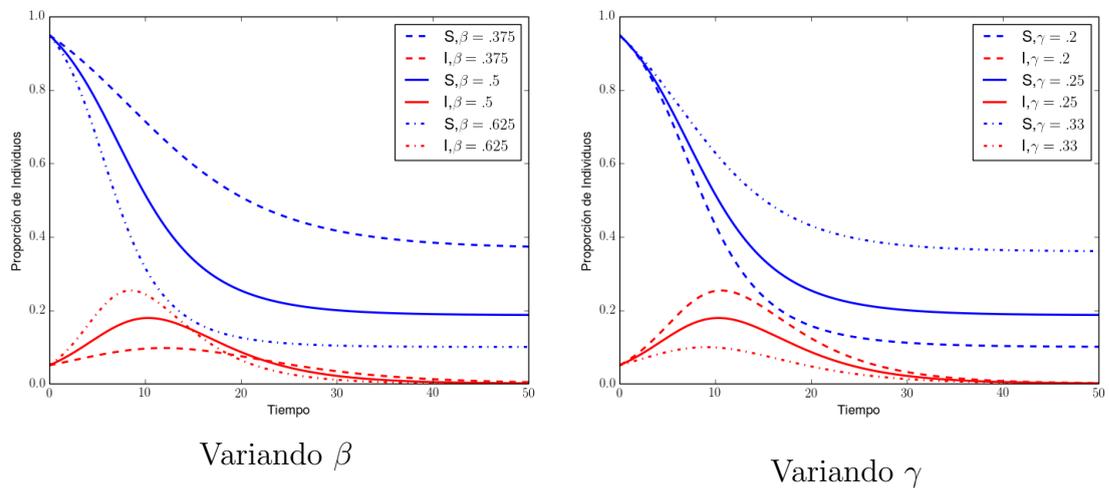


FIGURA 1.2: Modelo SIR Determinístico variando los parámetros.

## 1.4. Modelo SIR Estocástico

El paradigma que se pretende utilizar en este trabajo está ligado con el concepto de aleatoriedad y existen argumentos fuertes para utilizar modelos estocásticos en lugar de determinísticos como:

- Los procesos en la vida real son estocásticos, las ocurrencias se suceden de acuerdo con una naturaleza estocástica pues existen muchas fuentes de incertidumbre al tratar de modelar este fenómeno como la dinámica de contacto entre la población, las vías de comunicación, eventos de aproximación social, temporalidades en actividades; añadir aleatoriedad al modelo le da flexibilidad y el modelo se ajusta mejor a los datos reales.
- Una estructura estocástica permite identificar parámetros que no podrían ser identificados en el problema determinista.
- Se pueden asociar distribuciones con características del proceso de gran interés.
- La simulación puede ser fácil aunque pesada computacionalmente.

Para realizar las simulaciones del modelo SIR se escogió hacerlo de acuerdo con un proceso de Markov de saltos puros, el cual se basa en las ecuaciones de Kolmogorov y su implementación computacional mediante el algoritmo Gillespie.

### 1.4.1. Procesos Estocásticos, la Ecuación Maestra y el Algoritmo Gillespie

A continuación se muestra una línea de tiempo en la que se pueden apreciar los antecedentes del algoritmo heurístico implementado y popularizado por Daniel T. Gillespie en su trabajo “*Exact Stochastic Simulation of Coupled Chemical Reactions*” [7], pero que fue creado originalmente por Joseph L. Doob [4], [5].



FIGURA 1.3: Línea del tiempo del algoritmo Gillespie

### 1.4.2. Ecuación maestra.

Una versión más simplificada de las ecuaciones de Chapman-Kolmogorov es la conocida como Ecuación Maestra o *Chemical Master Equation (CME)*, muy utilizada en ciencias físicas y químicas para describir la evolución en el tiempo de un sistema compuesto por partículas en las que cada una puede encontrarse en uno de una cantidad numerable de estados diferentes en un momento específico, y

algún cambio de estado se hace de forma probabilística, estos cambios se llaman reacciones.

Una ecuación maestra consiste en un conjunto de ecuaciones diferenciales de primer orden que describen algún fenómeno que cambia a través del tiempo, la descripción más común es de acuerdo con la ecuación:

$$\frac{dX}{dt} = \mathbf{R}X \quad (1.5)$$

donde  $X$  es un vector columna que contiene los estados posibles y  $\mathbf{R}$  es una matriz que relaciona los estados con un conjunto de reacciones. Si las reacciones dependen del tiempo  $\mathbf{R} = \mathbf{R}(t)$  en la ecuación 1.5 el proceso es no estacionario; por el contrario, si las tasas de reacción son independientes del tiempo, la ecuación maestra representa un esquema cinético y el proceso es Markoviano, en consecuencia, cualquier tiempo de salto del sistema tiene una distribución exponencial.

Sea  $u$  la cantidad de estados posibles que se pueden visualizar en el vector  $X = (X_1, \dots, X_u)$  y  $v$  las posibles reacciones, la ecuación que describe a la  $k$ -ésima reacción del sistema se representa como:



donde  $p_{kj}$  representa el estado inicial del  $j$ -ésimo reactante  $X_j$  y  $q_{kj}$  representa el producto final de la reacción en el componente, así el efecto de la reacción en la  $j$ -ésima componente es la diferencia  $q_{kj} - p_{kj}$

Cada reacción tiene asociado una tasa de reacción, de tal manera que el tiempo de la siguiente reacción se distribuye exponencialmente con parámetro  $h_k(\mathbf{X}\theta_k)$ , es decir, el parámetro de la reacción es función de  $\theta_k$ , el cual es el vector de tasas de incidencia del sistema.

De lo anterior se obtienen dos condiciones:

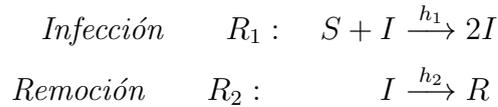
- Puesto que el tiempo de cualquier reacción se distribuye exponencial, el tiempo a la primer reacción se distribuye exponencialmente con parámetro

$$\sum_{i=1}^v h_i(\mathbf{X}\theta_i) \quad (1.7)$$

- La probabilidad de que la  $k$ -ésima reacción sea la primera en ocurrir tiene probabilidad de ocurrencia

$$\frac{h_k(\mathbf{X}\theta_k)}{\sum_{i=1}^v h_i(\mathbf{X}\theta_i)} \quad (1.8)$$

El modelo SIR se puede describir mediante la Ecuación Maestra o CME por sus siglas en inglés (*Chemical Master Equation*), donde el parámetro  $\theta = (\theta_1, \dots, \theta_v)$  depende del tiempo, esto es,  $\theta = \theta(t)$ , en el caso del modelo SIR, existen  $u = 3$  reactantes:  $(S, I, R)$  y  $v = 2$  reacciones posibles: Infección ( $R_1$ ) y Remoción ( $R_2$ ), y se representan como sigue:



### 1.4.3. Algoritmo Gillespie

El algoritmo fue creado en 1945 por Doob, pero popularizado en 1977 por Dan Gillespie [7] para la simulación de reacciones bioquímicas utilizando recursos computacionales limitados, involucra una variación de un método Monte Carlo Dinámico similar a métodos Monte Carlo Cinéticos.

El algoritmo permite una simulación discreta en el espacio de estados con saltos estocásticos, en la que cada una de las reacciones posibles es simulada explícitamente, pero donde los saltos en el tiempo son en un espacio continuo, el algoritmo se describe a continuación.

**Algoritmo**

- 1 Inicializar los componentes que intervienen: número de elementos de cada clase, tasas.
- 2 Cálculo del tiempo en que sucederán los posibles eventos y determinar el tipo de evento, como lo indican las ecuaciones 1.7 y 1.8.
- 3 Actualizar, aumentar el tiempo que tomó el proceso y las modificaciones hechas por la reacción efectuada.
- 4 Iterar hasta una condición de paro.

Puesto que la demanda computacional es considerable, existen técnicas adaptadas para acelerar la simulación tales como Tau-leaping. Sin embargo, se pierde exactitud teórica. La finalidad de este trabajo no es obtener un método óptimo para la simulación que sacrifique un grado de precisión en la simulación, por lo que no se consideran dichas técnicas de aceleración.

**1.4.4. Ejemplo con datos sintéticos**

Con fines de comparación posterior entre los métodos de simulación para inferencia de los parámetros se considerará una muestra sintética, proveniente de parámetros conocidos  $\beta = 0.5, \gamma = 0.25$ , a partir de los cuales se pretende realizar la inferencia.

En la Figura 1.4 se puede comparar el comportamiento de 200 trayectorias estocásticas contra el modelo determinístico con los parámetros fijos establecidos anteriormente.

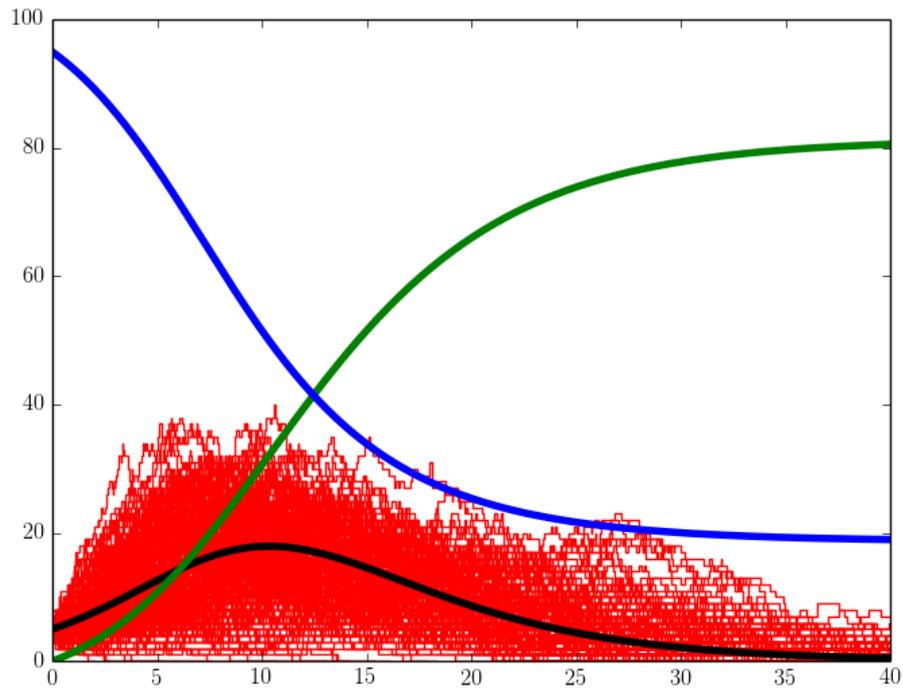


FIGURA 1.4: Simulación del número de individuos conforme pasa el tiempo en el modelo SIR estocástico con parámetros  $\beta = 0.5$  y  $\gamma = 0.25$

El resultado de una simulación se aprecia en las Figuras 1.5 y 1.6, en las que se varía la cantidad de individuos de la población, considerándose 100 y 1,000 individuos respectivamente; en ambas se observa la relación con el modelo determinístico. Los datos sintéticos producidos con 100 habitantes se utilizarán como referencia sobre todo el análisis para comparación de los resultados entre las metodologías utilizadas.

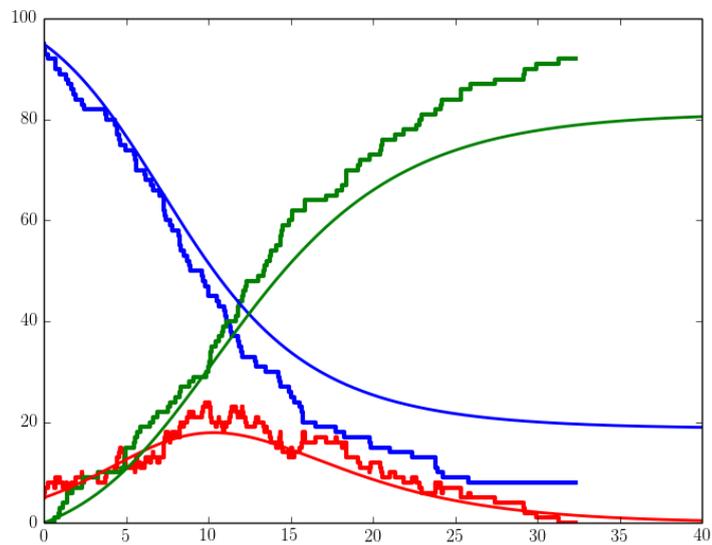


FIGURA 1.5: Simulación del modelo SIR estocástico con parámetros  $\beta = 0.5$  y  $\gamma = 0.25$  y una población  $N = 100$

Al incrementar el número de individuos de la población a 1,000, se observa que las curvas se parecen más al modelo determinístico; sin embargo el costo computacional de dicha simulación también se incrementa.

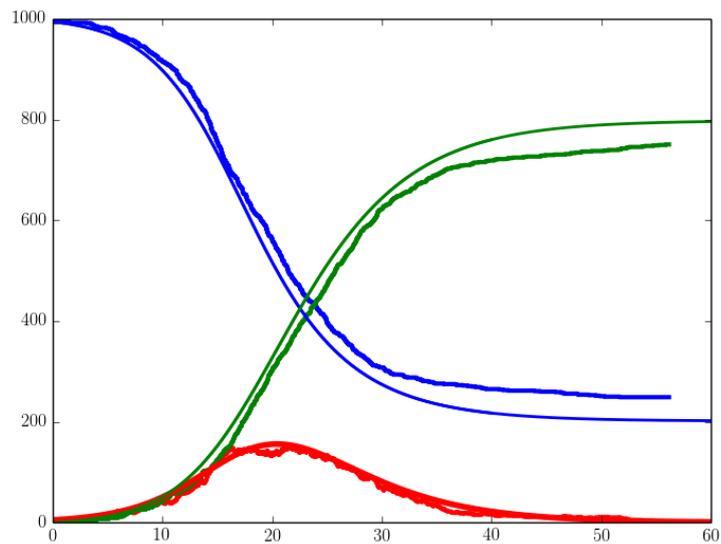


FIGURA 1.6: Simulación del modelo SIR estocástico con parámetros  $\beta = 0.5$  y  $\gamma = 0.25$  y una población  $N = 1,000$

---

---

# CAPÍTULO 2

---

## INFERENCIA BAYESIANA SIN VEROSIMILITUD: ABC

### 2.1. Introducción

En este capítulo se exponen de manera breve las herramientas que serán de utilidad para el desarrollo de la inferencia en el modelo SIR.

La aplicación de la inferencia en la Estadística permite obtener estimadores para los parámetros que rigen ciertos modelos, su desarrollo depende en gran medida de la función de verosimilitud. No obstante, en ocasiones las expresiones pueden ser muy complicadas de calcular o simplemente no existe una forma analítica que las defina, es por eso que se han desarrollado técnicas para tratar de resolver este problema. Los requerimientos del procedimiento desarrollado implican el poder proponer parámetros y ser capaces de producir datos simulados de acuerdo con un modelo.

Una vez obtenidos los datos simulados se requerirá un estadístico que permita resumir la información tanto de las simulaciones como de los datos observados del fenómeno. También se requerirá un mecanismo que permita evaluar aquellos parámetros que produzcan resultados más cercanos a los observados, para lo cual

será de utilidad la teoría de Estimación por Kernel para aproximar la distribución objetivo, considerándose en el estudio los kernels Uniforme, Epanechnikov y Normal. Se mostrará cómo el primero coincide con el algoritmo clásico de la metodología ABC y se expondrá la robustez del kernel Normal ante los demás.

Se requerirá en gran medida la teoría de simulación Monte Carlo, para ello se necesitará del algoritmo Metropolis-Hastings y se explicará la modificación para la aplicación del algoritmo en un contexto libre de verosimilitud.

## 2.2. Inferencia libre de Verosimilitud

El problema clásico de inferencia en Estadística hace gran uso de la función de verosimilitud, tanto en el paradigma Frecuentista como en el Bayesiano. En el segundo, la inferencia se basa en encontrar la llamada distribución posterior, la cual toma en consideración la información previa que se tiene del parámetro  $\theta$ , así como de la información que aportan los datos observados  $y$ :

$$\underbrace{\pi(\theta|y)}_{\text{Posterior}} \propto \underbrace{\pi(\theta)}_{\text{a priori}} \underbrace{f(y|\theta)}_{\text{Verosimilitud}} \quad (2.1)$$

El caso que se presenta en el modelo SIR con información parcial refleja una función de verosimilitud que es intratable computacionalmente, pues no existe una forma analítica cerrada para esta expresión.

Las herramientas que han favorecido a los estadísticos para generar muestras de distribuciones complejas son los métodos Monte Carlo. Para realizar inferencia, se recurre a métodos bayesianos computacionalmente intensivos, es por eso que es posible apoyarse en estos métodos para dar solución al problema, el cual se basa en encontrar una distribución posterior aproximada  $\pi^*(\theta|y) \approx \pi(\theta|y)$  de la distribución real.

Para plantear una solución a este tipo de problemas se considera un modelo que rige al fenómeno, en el cual, dado un parámetro  $\theta \in \Theta$  se generan datos u observaciones  $y = f(\theta) \in \mathcal{D}$  que posiblemente contienen ruido. El problema de inferencia consiste en que se tienen observaciones  $y$  del fenómeno y a partir de estas se desea obtener

el parámetro  $\theta$  que las produce. Este esquema es el mismo que relaciona a los llamados Problemas Inversos.

$$y \longleftarrow \theta$$

El problema clásico en inferencia se ataca mediante el uso de un estadístico suficiente  $t$ . Sin embargo, cuando no existe una expresión de verosimilitud, tampoco se tiene un estadístico suficiente, entonces es posible recurrir a un estimador de la densidad posterior  $\pi^*$  utilizando alguna estadística sobre los datos.

### 2.2.1. El Estadístico

Como medida para resumir la información que proporcionan los datos, se utilizan funciones de dichas observaciones, en este caso el estadístico que se utiliza es aquel que mide de cierta forma la discrepancia entre las observaciones reales ( $y$ ) y las simulaciones obtenidas ( $x$ ) dado un parámetro fijo  $\theta$ .

En primer lugar se realiza una transformación  $\eta : \mathcal{D} \rightarrow \mathcal{S}$  que es utilizada para resumir la información, la cual se aplica tanto a observaciones como a simulaciones y mediante la transformación  $\rho : \mathcal{S} \times \mathcal{S} \rightarrow \mathbb{R}^p$ , que proporciona una medida de comparación entre las observaciones reales y las simulaciones del modelo. Lo anterior se puede resumir mediante una función  $t$  que no depende de cantidades desconocidas, donde  $t : \mathcal{D} \times \mathcal{D} \rightarrow \mathbb{R}^p$  de tal manera que  $t(x, y) = \rho(\eta(x), \eta(y))$

La alta dimensionalidad de algunos problemas ha generado variaciones en las que se recurre a estadísticos que resumen demasiado la información. La desventaja inmediata es la pérdida de información, sobre todo en modelos complejos, por lo que resulta importante determinar el cómo se establece dicho estadístico.

El sentido de aproximación a la distribución posterior se cuantifica en cierta medida considerando algunos aspectos de la información obtenida, por lo que es necesario obtener un estadístico que extraiga la información y algún tipo de comparación entre los datos y el modelo.

Una vez que se tiene el estadístico, se debe cuantificar entre las distintas propuestas, lo cual se hará mediante la estimación kernel. Siendo la estimación kernel

herramienta fundamental de este análisis, a continuación se dará una breve explicación de la teoría de estimación kernel tanto para el caso univariado como para el multivariado.

### 2.2.2. Estimación Kernel

Durante la realización de este trabajo se han encontrado publicaciones de Fukumizu [8] y Nakagome [16] en los cuales se fomenta el uso de la estimación Kernel.

Puesto que los estadísticos  $t$  que se utilizarán pueden ser tanto univariados como multivariados, se expondrá el caso general multivariado.

Sea una muestra  $\mathbf{x}_1, \dots, \mathbf{x}_n$  donde cada uno de los elementos es de dimensión  $d$ , proveniente de una densidad  $f$ , el estimador de la densidad kernel  $\hat{f}$  para  $\mathbf{x} \in \mathbb{R}^d$  se define como:

$$\begin{aligned}\hat{f}_{\mathbf{H}}(\mathbf{x}) &= \frac{1}{n} \sum_{i=1}^n K_{\mathbf{H}}(\mathbf{x} - \mathbf{x}_i) \\ &= \frac{1}{n|\mathbf{H}|^{1/2}} \sum_{i=1}^n K(\mathbf{H}^{-1/2}(\mathbf{x} - \mathbf{x}_i))\end{aligned}\quad (2.2)$$

donde  $\mathbf{x} = (x_1, x_2, \dots, x_d)^T$  y las observaciones  $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{id})^T$ ,  $i = 1, 2, \dots, n$  y  $\mathbf{H}$  representa el ancho de banda, el cual consiste de una matriz  $d \times d$ , simétrica y definida positiva.  $K$  es la función Kernel, la cual cumple ser una función de densidad multivariada de dimensión  $d$ , simétrica y  $K_{\mathbf{H}}(\mathbf{x}) = |\mathbf{H}|^{-1/2}K(\mathbf{H}^{-1/2}\mathbf{x})$  es la función kernel escalada simétrica y  $|\mathbf{H}|$  el determinante de la matriz  $\mathbf{H}$ .

Existen dos tipos de kernel que se pueden crear a partir de un kernel  $k$  univariado simétrico:

Un kernel producto, el cual se utilizará en los casos Uniforme y Epanechnikov

$$K^P(\mathbf{x}) = \prod_{i=1}^d k_{x_i}$$

y un kernel esférico simétrico, el cual será de utilidad en el caso Gaussiano

$$K^S(\mathbf{x}) = \frac{k\left(\sqrt{\mathbf{x}^T \mathbf{x}}\right)}{\int k\left(\sqrt{\mathbf{x}^T \mathbf{x}}\right)}$$

La diferencia en la elección entre estos dos tipos de kernel no es tan significativa como la elección del ancho de banda  $\mathbf{H}$ . Sea  $\mathcal{H}$  la clase de todas las matrices positivas definidas simétricas de  $d \times d$ , la matriz completa  $\mathbf{H} \in \mathcal{H}$  tiene  $d(d+1)/2$  posibles entradas, lo cual se puede simplificar imputando restricciones como ser matriz diagonal o aún más, que  $\mathbf{H} = h^2 \mathbf{I}_d$ .

Entre los distintos tipos de Kernel que se utilizarán para la aproximación están:

- General  $K(u) = \frac{(1-u^2)^p}{2^{2p+1} B(p+1, p+1)} \mathbf{1}_{\{|u| \leq 1\}}$
- Uniforme  $K(u) = \frac{1}{2} \mathbf{1}_{\{|u| \leq 1\}}$
- Epanechnikov  $K(u) = \frac{3}{4} (1-u^2) \mathbf{1}_{\{|u| \leq 1\}}$
- Normal  $K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$

Un problema con el que se debe lidiar es encontrar el ancho de banda óptimo para la estimación; sin embargo, esto depende de cada problema en particular y es difícil dar una regla general, pero la mejor opción es evaluar distintas opciones hasta encontrar la que se ajuste mejor a las necesidades y que obtenga un balance entre la precisión y el esfuerzo computacional.

Como ejemplo, en el caso de utilizar un Kernel Normal, aplicado a un estadístico multivariado se obtiene

$$K_{\mathbf{H}}(x) = (2\pi)^{-d/2} |\mathbf{H}|^{-1} \exp(-1/2 x^T \mathbf{H}^{-1} x)$$

y en el caso univariado

$$K_H(x) = \frac{1}{(2\pi)^{1/2} |H|} \exp\left(-\frac{1}{2} \frac{x^2}{H}\right)$$

en este caso, resulta claro que el ancho de banda asume el papel del parámetro de dispersión en la densidad Normal.

## 2.3. Esquema General para Estimación Libre de Verosimilitud

Una alternativa para la estimación de la densidad se propone vía **estimación de densidad por Kernel**; dicha estimación estará muy ligada tanto a la función kernel utilizada  $K$ , como al ancho de banda  $H$  asociado, lo anterior se muestra esquemáticamente en el siguiente diagrama.

$$\underbrace{\pi(\theta|y)}_{\text{Posterior}} \longrightarrow \pi(\theta|t(y)) \longrightarrow \underbrace{\pi^*(\theta|t(y))}_{\text{Aproximación}} \longrightarrow \pi_{K,H}^*(\theta|t(y)) \longrightarrow \underbrace{\theta^{(i)}}_{\text{muestra}}$$

Se introducen pseudo-observaciones  $x \in \mathcal{D}$  que corresponden a una simulación del modelo, que provienen de utilizar un parámetro  $\theta$  propuesto, de tal manera que se realiza la aproximación que se expresa modificando la notación que establece Marin [9], retomada por Del Moral [2] como:

$$\begin{aligned} \pi(\theta|y) &\approx \pi_{K,H}^*(\theta|y) = \int \pi_{K,H}(\theta, x|y) dx \\ &= \int \frac{\pi(\theta) f(x|\theta) K_{H,y}(x)}{\int_{K_{H,y} \times \Theta} \pi(\theta) f(x|\theta) dx d\theta} dx \\ &\propto \int \pi(\theta) f(x|\theta) K_{H,y}(x) dx \\ &\propto \int \pi(\theta) f(x|\theta) [|H|^{-1/2} K(H^{-1/2}t(x, y))] dx \end{aligned} \tag{2.3}$$

El estadístico  $t$  puede ser univariado ( $p = 1$ ) o multivariado ( $p > 1$ ), está definido por  $t : \mathcal{D} \times \mathcal{D} \rightarrow \mathbb{R}^p$  y representa un estadístico aplicado a los datos  $y$  y las simulaciones  $x$  para ser evaluados por el estimador kernel.

Dicho método permite la obtención de una muestra independiente de la distribución:

$$\pi_{K,H}^*(\theta|y)$$

Cuando  $|H| \rightarrow 0$ , el ancho de banda tiende a cero, entonces

$$\int f(x|\theta)|H|^{-1/2}K(H^{-1/2}t(x,y))dx \longrightarrow f(y|\theta)$$

tal como lo indica la teoría de Estimación de densidad por kernel o KDE (*Kernel Density Estimation*) y por eso

$$\int K_{H,y}(\theta, x|y)dx \longrightarrow \pi(\theta|y)$$

## 2.4. Inferencia sin Verosimilitud: ABC

Los métodos ABC (Tavare [17], Beaumont [1]) o también llamados métodos libres de verosimilitud, son una serie de métodos que proponen una aproximación de la verosimilitud en procesos complejos que resultan tener una verosimilitud intratable.

Para establecer lo que significa intratable se hará referencia a una clasificación hecha por Murray [11], que distinguió tres niveles de intratabilidad de verosimilitudes: la primera tiene que ver con que la verosimilitud es conocida salvo una constante de proporcionalidad  $P(y|\theta) = \frac{f(y|\theta)}{C}$  donde  $f$  se puede calcular, y  $C$  es una constante desconocida independiente del parámetro  $\theta$ . El segundo tipo se refiere con que la constante  $C$  en el punto anterior dependa del parámetro  $\theta$ , para estos dos primeros casos se necesita la función de verosimilitud y es posible utilizar métodos Monte Carlo para encontrar las distribuciones posteriores. El tercer tipo, y motivador para el desarrollo del método ABC consiste en conocer el modelo estocástico del cual se originan los datos; sin embargo, no se conoce parte alguna de la verosimilitud.

El método ABC originalmente tratado por Diggle y Graiton [3] en 1984 consideró la inferencia estadística desde el enfoque frecuentista, condicionando al parámetro  $\theta$  y generando los datos bajo el modelo  $P(X|\theta)$ . La intuición que subyace es que entre las simulaciones de distribuciones conjuntas de los parámetros y los datos, aquellos parámetros que corresponden a observaciones del fenómeno similares a los datos observados  $y$ , son aproximadamente proporcionales a la verosimilitud y por ende a la distribución posterior. Si se pidiera que los datos simulados fueran exactamente iguales a los datos observados, entonces la inferencia sería exacta, la complicación

para que esto ocurra está ligada con el modelo mismo, pues puede ser que se tengan datos continuos y/o que la dimensión de los parámetros y observaciones sea muy elevada.

### 2.4.1. Muestreo de Aceptación Exacto

Conocido en inglés como *Rejection Sampling*. Si los parámetros  $\theta^{(1)}, \dots, \theta^{(k)}$  propuestos son aplicados al modelo y éstos resultan en datos simulados que coinciden exactamente con las observaciones reales del fenómeno, se obtiene una muestra de la distribución posterior exacta;

El algoritmo se puede resumir en los siguientes pasos:

- 1 Generar  $\theta$  de una distribución a priori

$$\theta \longleftarrow \pi(\cdot)$$

- 2 Simular datos sintéticos del modelo con parámetro  $\theta$  obtenido

$$x \longleftarrow f(\cdot|\theta)$$

- 3 Aceptar el parámetro  $\theta$  propuesto si  $x = y$  y regresar al paso uno hasta obtener el tamaño de muestra deseado

Esto produce una muestra exacta de la posterior  $\pi(\theta|y)$ . Sin embargo, en espacios discretos, conforme se tenga un mayor número de observaciones del fenómeno, el obtener valores simulados que coincidan exactamente con los observados se vuelve más improbable, este problema deja de ser manejable cuando se trabaja con espacios de estados continuos, por lo que se recurre a aproximaciones de la distribución verdadera.

### 2.4.2. Muestreo de Aceptación Aproximado: ABC Clásico

Un caso particular de la idea de utilizar una distribución aproximada con estimación kernel, es utilizar una función kernel uniforme, la cual produce el llamado ABC clásico.

De la expresión mostrada por la ecuación 2.3 utilizando un kernel uniforme se obtiene

$$\pi_{K,H}(\theta|y) \propto \int \pi(\theta) f(x|\theta) K_{H,y}(x) dx = \int \pi(\theta) f(x|\theta) \mathbb{I}_{H,y}(x) dx$$

donde

$$\mathbb{I}_{H,y}(x) = \{x \in \mathcal{D} : \rho(\eta(y), \eta(x)) < H\}$$

y que establece la idea de que usando un estadístico  $t$  acompañado de kernel con una tolerancia  $H$  lo suficientemente pequeña, se puede producir una aproximación suficientemente buena de la distribución posterior

$$\begin{aligned} \pi_H(\theta|y) &= \int_{\mathcal{D}} \pi_H(\theta, x|y) dx \\ &= \int_{\mathcal{D}} \frac{\pi(\theta) f(x|\theta) K_H(t(x, y))}{\int_{K_{H,y} \times \Theta} \pi(\theta') f(x|\theta') K_H(t(z, y)) dz d\theta'} dx \\ &\approx \pi(\theta|y) \end{aligned}$$

conforme la tolerancia referida aumenta  $H \rightarrow \infty$  se obtienen observaciones de la distribución a priori y si  $H \rightarrow 0$  se llega a la situación en la que  $x = y$ , por lo que las observaciones provendrán de la distribución  $\pi(\theta|y)$ .

### 2.4.3. Algoritmo Metropolis-Hastings

El algoritmo Metropolis-Hastings se puede resumir en:

- Simular  $\theta \sim q(\cdot|\theta^{(t-1)})$
- Calcular  $\alpha(\theta|\theta^{(t-1)}) = \min \left\{ 1, \frac{f(\theta)}{f(\theta^{(t-1)})} \frac{q(\theta^{(t-1)}|\theta)}{q(\theta|\theta^{(t-1)})} \right\}$
- Con probabilidad  $\alpha = \alpha(\theta|\theta^{(t-1)})$  asignar  $\theta^{(t)} = \theta$ , de otra manera  $\theta^{(t)} = \theta^{(t-1)}$

En el caso que se plantea para este trabajo, se utiliza una aproximación de la distribución posterior exacta mediante estimación kernel, por lo que la probabilidad de aceptación se modifica como lo muestra la ecuación 2.4.

Estando en un valor del parámetro  $\theta'$  se propone moverse a  $\theta$  mediante un kernel de transición  $q(\theta|\theta')$  y se calcula la razón de Metropolis-Hastings para evaluar la calidad del siguiente punto propuesto.

$$\begin{aligned} \alpha &= \min \left\{ 1, \frac{\pi(\theta') f(x'|\theta') K_H(t(x, y))}{\pi(\theta) f(x|\theta) K_H(t(x', y))} \frac{q(\theta|\theta') f(x|\theta)}{q(\theta'|\theta) f(x'|\theta')} \right\} \\ &= \min \left\{ 1, \frac{K_H(t(x, y))}{K_H(t(x', y))} \frac{q(\theta|\theta') \pi(\theta')}{q(\theta'|\theta) \pi(\theta)} \right\} \end{aligned} \quad (2.4)$$

En el caso en el cual se tiene una propuesta  $q(\theta'|\theta)$  simétrica, éste término se cancela y la expresión de la ecuación 2.4 se reduce aún más.

En el caso en el que, en la estimación kernel se utiliza un kernel uniforme, se obtiene la reciente clase de métodos computacionales conocido como Computación Bayesiana Aproximada (ABC).

## 2.5. MCMC sin Verosimilitud: ABC - MCMC

En la práctica utilizar propuestas provenientes directamente de la distribución a priori  $\pi(\cdot)$  resulta ineficiente, porque no contempla los datos en la fase de la propuesta y produce valores en regiones de baja probabilidad. Marjoram [13] propone una solución al introducir el algoritmo ABC-MCMC, el cual hace referencia a la simulación mediante *Markov Chain Monte Carlo* (MCMC) y que se describe a continuación:

### Algoritmo ABC-MCMC

- Utilizando el algoritmo de Rejection Sampling, obtenga una primera realización  $(\theta^{(0)}, x^{(0)})$  de la distribución  $\pi_H(\theta, x|y)$
- Para  $t = 1 : N$ 
  - Generar  $\theta'$  de un Kernel de transición  $q(\cdot|\theta^{(t-1)})$
  - Generar  $x'$  de  $f(\cdot|\theta')$
  - Generar  $u \sim \mathcal{U}_{[0,1]}$

- Si  $u \leq \frac{\pi(\theta^{(t-1)})q(\theta'|\theta^{(t-1)})K_H(t(x',y))}{\pi(\theta')q(\theta^{(t-1)}|\theta')K_H(t(x,y))}$ , de acuerdo con la ecuación 2.4 entonces  $(\theta^{(t)}, x^{(t)}) = (\theta', x')$

donde  $x$  y  $x'$  son las pseudo observaciones generadas por los parámetros  $\theta$  y  $\theta'$  respectivamente. Cuando no se cumple esta desigualdad, entonces

$$(\theta^{(t)}, x^{(t)}) = (\theta^{(t-1)}, x^{(t-1)})$$

La naturaleza del algoritmo es realizar una caminata aleatoria o bien un *Random Walk Metropolis-Hastings*. En el caso que se abordará en este trabajo, se utilizará un kernel de transición con distribución de perturbación normal multivariada como propuesta para la caminata, al no tener información previa se supondrá una matriz de covarianza de tipo diagonal, la cual considera independencia entre los parámetros.

## 2.6. Método Monte Carlo Secuencial libre de Verosimilitud: ABC - SMC

Toni [18] propone un método que consiste en un procedimiento para mejorar las implementaciones de MCMC en el ABC, aprendiendo secuencialmente sobre la distribución objetivo.

Existen dos desventajas con los métodos MCMC, el primero consiste en determinar el momento en el cual la cadena ha convergido a la distribución objetivo, el segundo es que los métodos MCMC propuesto por Del Moral [14] no pueden ser usados en un contexto de estimación bayesiana secuencial. La alternativa viene de la propuesta basada en métodos Monte Carlo Secuenciales, conocidos en inglés como *Sequential Monte Carlo (SMC)*.

La idea básica es obtener una gran colección de muestras aleatorias ponderadas, las cuales se denominan partículas, y cuya distribución empírica converge asintóticamente conforme el tamaño de estas partículas  $N \rightarrow \infty$  a  $\pi_n$ .

Las partículas evolucionan a través del tiempo haciendo uso de ideas de remuestreo y Muestreo por Importancia o IS (*Importance Sampling*); sin embargo, las ideas estándares de algoritmos SMC no aplican para métodos libres de verosimilitud, por lo que aquí se realizan ajustes que permitan su tratamiento.

A continuación se describe el algoritmo propuesto por Toni [18] para realizar SMC.

**Algoritmo ABC-SMC**

1 Inicializar las tolerancias  $\epsilon_1, \dots, \epsilon_T$  y fijar el indicador de la población en  $t = 0$ .

2 Fijar el indicador de partícula en  $i = 1$

2.1 Si  $t = 0$  simular de forma independiente  $\theta^{**}$  proveniente de  $\pi(\theta)$ .

De otra manera, simular  $\theta^*$  de la población anterior  $\{\theta_{t-1}^{(i)}\}$  con pesos  $w_{t-1}$  y perturbar la partícula para obtener  $\theta^{**} \sim K_t(\theta|\theta^*)$  donde  $K_t$  es el Kernel de perturbación.

Si  $pi(\theta^{**}) = 0$  regresar a [2.1]

Simular una pseudo-observación  $x^* \sim f(x|\theta^{**})$ .

Si  $d(x^*, x_0) \geq \epsilon_t$ , regresar a [2.1]

2.2 Fijar  $\theta_t^{(i)} = \theta^{**}$  y calcular los pesos para las partículas  $\theta_t^{(i)}$  de acuerdo con:

$$w_t^{(i)} = \begin{cases} 1 & \text{si } t=0 \\ \frac{\pi(\theta_t^{(i)})}{\sum_{j=1}^N w_{t-1}^{(j)} K(\theta_{t-1}^{(j)}, \theta_t^{(i)})} & \text{si } t > 0 \end{cases}$$

Si  $i < N$ , asignar  $i = i + 1$  y regresar al paso [2.1]

3 Normalizar los pesos. Si  $t < T$ , asignar  $t = t + 1$  e ir al paso [2]

---

---

# CAPÍTULO 3

---

## INFERENCIA EN EL MODELO SIR

### 3.1. Inferencia con información perfecta en el modelo SIR

En el Capítulo 1 se generaron datos en forma sintética, por lo que se conocen los tiempos exactos de ocurrencia de eventos y los eventos ocurridos en cada salto, entonces se cuenta con información perfecta, en este caso la inferencia se puede calcular de manera explícita, pues al considerar una distribución a priori Gamma para cada uno de los parámetros y de acuerdo con la CME, se sabe que las observaciones saltan de acuerdo con una distribución exponencial, por lo que la distribución posterior resultará en una distribución conjugada Exponencial-Gamma.

#### 3.1.1. Inferencia con enfoque clásico

Al presentarse el modelo como un proceso de saltos puros, los eventos se presentan de acuerdo con un Proceso Poisson no homogéneo, en el cual se sabe que para un momento determinado, entonces, es posible encontrar expresiones cerradas que ayuden a determinar la inferencia de los parámetros.

$$L(\mathbf{t}|\beta, \gamma) = \prod_{i=1}^n \left( \beta \frac{S_{t_i}}{N} I_{t_i} + \gamma I_{t_i} \right) \exp \left\{ -t_i \left( \beta \frac{S_{t_i}}{N} I_{t_i} + \gamma I_{t_i} \right) \right\}$$

Al saber cuál fue el evento en cada instante, se discriminan los eventos de un sólo tipo

$$L(\mathbf{t}|\beta) = \prod_{i \in I_\beta} \left( \beta \frac{S_{t_i}}{N} I_{t_i} \right) \exp \left\{ -t_i \left( \beta \frac{S_{t_i}}{N} I_{t_i} + \gamma I_{t_i} \right) \right\}$$

obteniéndose una expresión de la cual se pueden obtener los parámetros.

### 3.1.2. Inferencia con enfoque bayesiano

Para el desarrollo de esta inferencia se necesita una distribución a priori que representa el conocimiento previo respecto de los parámetros que se pretende actualizar con la información proveniente de los datos observados  $y$ . Al ser los tiempos distribuidos de acuerdo con una ley exponencial lo más conveniente para obtener una distribución conjugada sería utilizar distribuciones a priori gamma, pues concuerda con la naturaleza de los parámetros al abarcar valores positivos.

## 3.2. Inferencia en el modelo SIR con información parcial

Se ha establecido que para hacer una inferencia de los parámetros de infección y remoción se necesitan los momentos exactos de ocurrencia de cada uno de los eventos a lo largo de su historia, en la realidad esto no es posible y sólo se pueden detectar en algunos puntos en el tiempo, el número de infectados que son reportados.

Al analizar una enfermedad que se propaga de acuerdo con el modelo SIR, es necesario contar con al menos algunos registros de pacientes infecciosos; no obstante se deben considerar algunas deficiencias de registro como:

- No toda la población que se ve infectada por cierta enfermedad acude a los centros de salud.

- No toda persona que acude al centro de salud puede ser diagnosticada correctamente a tiempo.
- No se conoce la cantidad de personas susceptibles en una población.
- Se desconocen las condiciones iniciales en las cuales se inicia el proceso de infección.

Puesto que se desea en primera instancia obtener resultados comparativos, se harán algunos supuestos básicos, al considerarse periodos de análisis de un año, se considera que durante ese intervalo la población es cerrada.

### 3.3. Características del modelo

Con fines comparativos se realizó una simulación con datos sintéticos a partir de parámetros que coinciden con Fuchs [6], donde en el caso del parámetro de propagación de la infección se considera una tasa de contacto de  $\beta = 0.5$ ; mientras que para el parámetro de remoción se utiliza una tasa de  $\gamma = 0.25$ , se trabaja con una población fija de  $N = 100$  individuos, los cuales se dividen en  $S = 95$  susceptibles,  $I = 5$  infecciosos y  $R = 0$  removidos.

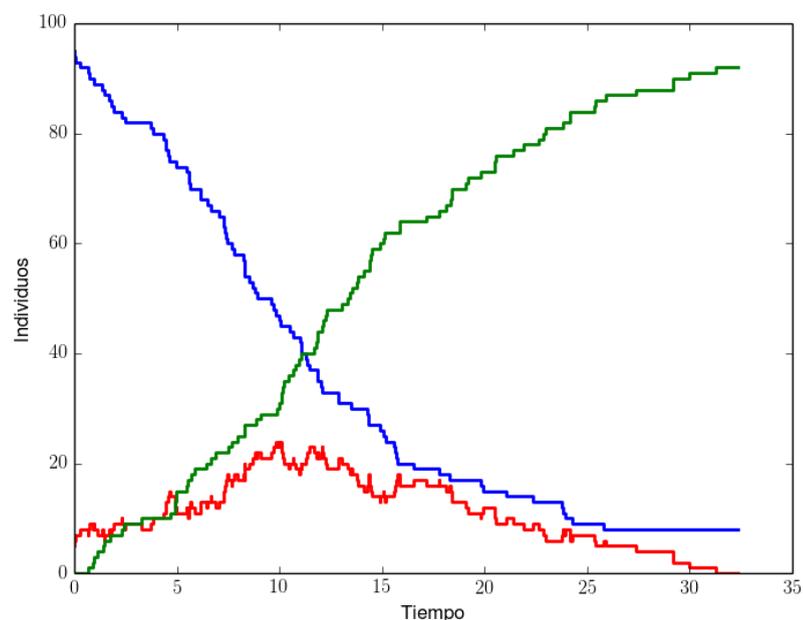


FIGURA 3.1: Distribución de curvas generadas con parámetro  $\theta = (\beta, \gamma) = (0.5, 0.25)$ .

Resulta interesante, antes de abordar el problema directamente, explorar la variabilidad de la distribución de las curvas de los procesos de infección con los parámetros anteriores fijos, por lo que se realizó una simulación con 100 curvas, la cual se muestra en la Figura 3.2

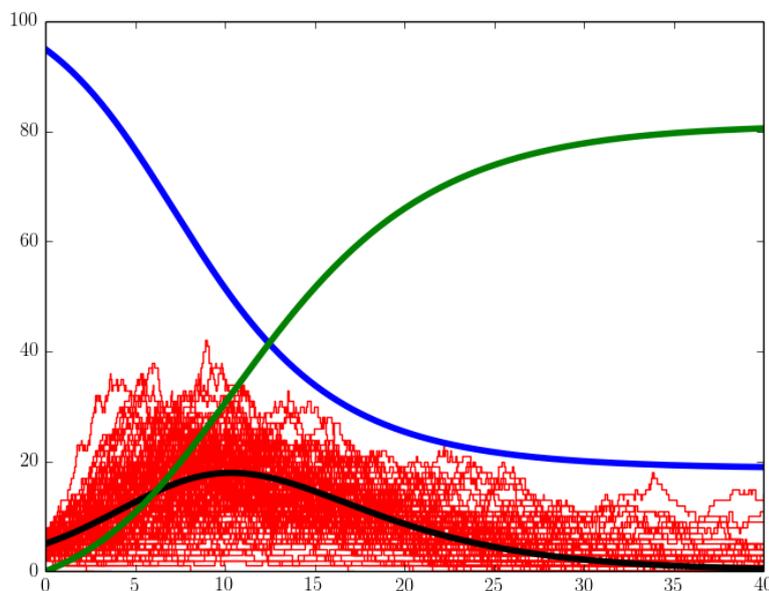


FIGURA 3.2: Distribución de curvas generadas con parámetro  $\theta = (\beta, \gamma) = (0.5, 0.25)$ .

La Figura muestra que las curvas representan un patrón común, además se incorpora la curva correspondiente al modelo determinístico con los mismos parámetros y resulta claro que las curvas estocásticas se distribuyen cercanas al modelo determinístico.

### 3.4. Consideraciones

En las siguientes secciones se exponen los resultados al llevar a cabo la inferencia de los parámetros, tanto de infección como de remoción en un modelo SIR clásico, el cual se aborda desde un esquema de un proceso de saltos puros con datos sintéticos. Este análisis sirve para entender la sensibilidad de los parámetros involucrados. En este caso, se asume una población total inicial  $N = 100$  conocida de individuos que se supondrá constante durante el estudio.

### 3.4.1. Estadístico

Se consideran dos tipos de funciones de distancia aplicados a los estadísticos utilizados para evaluar la discrepancia entre los datos observados y los simulados: la primera, corresponde a una estadística multidimensional que considera la diferencia absoluta en cada componente de las observaciones contra las simulaciones  $I_{Obs} - I_{Sim}$ , por lo que la dimensión correspondiente es igual a la cantidad de observaciones obtenidas. La segunda, corresponde a una comparación que resume la información conocida como la distancia de Pearson, definida por  $\sum_i \frac{(I_{Sim} - I_{Obs})^2}{I_{Obs}}$  para aquellas observaciones que son distintas de cero, y donde por facilidad en cálculos posteriores se obtiene la raíz  $\sqrt{\sum_i \frac{(I_{Sim} - I_{Obs})^2}{I_{Obs}}}$ .

### 3.4.2. Estimación Kernel y Ancho de Banda

La notación utilizada es de acuerdo con lo expuesto por Marin [9], en el que referencia a Wilkinson [15] donde se propone un cambio de perspectiva y reemplazar un error de aproximación con una inferencia exacta sobre una aproximación controlada a la distribución objetivo basada en una función kernel  $K_H$  arbitraria, la cual se considera como una generalización del ABC clásico.

Asociado con el estimador kernel utilizado, explícitamente incluye un ancho de banda  $H$ , el cual tomará el papel de la tolerancia  $\epsilon$ , pues en el caso de utilizar un kernel uniforme, corresponde al caso del ABC con muestreo por aceptación; por lo que la distribución objetivo se expresa como.

$$\pi_{K,H}^*(\theta|y) = \int \pi_{K,H}(\theta, x|y) = \frac{\int \pi(\theta) f(x|\theta) K_H(t(x, y)) dx}{\int \pi(\theta) f(x|\theta) K_H(t(x, y)) dx d\theta}$$

En el presente análisis se consideran las funciones kernel:

- Uniforme
- Epanechnikov
- Normal

### 3.4.3. Método de simulación de la muestra

Para hacer inferencia eficiente en el caso en el que se utilizarán datos reales, se evaluarán diferentes alternativas utilizando datos sintéticos y se elegirá la combinación que produzca mejores resultados.

Entre los métodos a utilizar se consideran:

- ABC Muestreo por Aceptación
- ABC MCMC
- ABC SMC

En resumen, se analizarán distintas combinaciones en la implementación de la inferencia de los parámetros en el modelo de infección, que se dividen en tres elementos:

- El estadístico que compare las discrepancias
  - Multidimensional:  $|I_{Obs} - I_{Sim}|$
  - Unidimensional:  $\sqrt{\sum_i \frac{(I_{Obs_i} - I_{Sim_i})^2}{I_{Obs_i}}}$
- La función kernel
  - Uniforme
  - Epanechnikov
  - Normal
- El método de muestreo por Monte Carlo
  - Muestreo por aceptación
  - MCMC
  - SMC

### 3.5. Implementación del Método de Inferencia

En el caso que se ha venido trabajando se establecen algunos tiempos de observación. En el ejemplo siguiente se escogieron en particular  $t = (t_1, t_2, \dots, t_{15}) = (1.5, 3.5, 5.5, 7.5, 9.5, 11.5, 13.5, 15.5, 17.5, 19.5, 21.5, 23.5, 25.5, 27.5, 29.5)$ , con la única característica de ser equidistantes, sin tener ésto importancia relevante en el análisis.

La Figura 3.3 muestra los puntos donde se obtuvieron dichas observaciones provenientes de parámetros  $\beta = 0.5$  y  $\gamma = 0.25$ , la cantidad de individuos infecciosos en este caso es de  $\{I_{t_j}\}_{j=1}^{15} = \{7, 8, 11, 18, 21, 23, 19, 14, 16, 11, 9, 6, 5, 4, 2\}$ . **Estos datos serán utilizados en todas las combinaciones de metodologías para comparar la que tenga un mejor desempeño.**

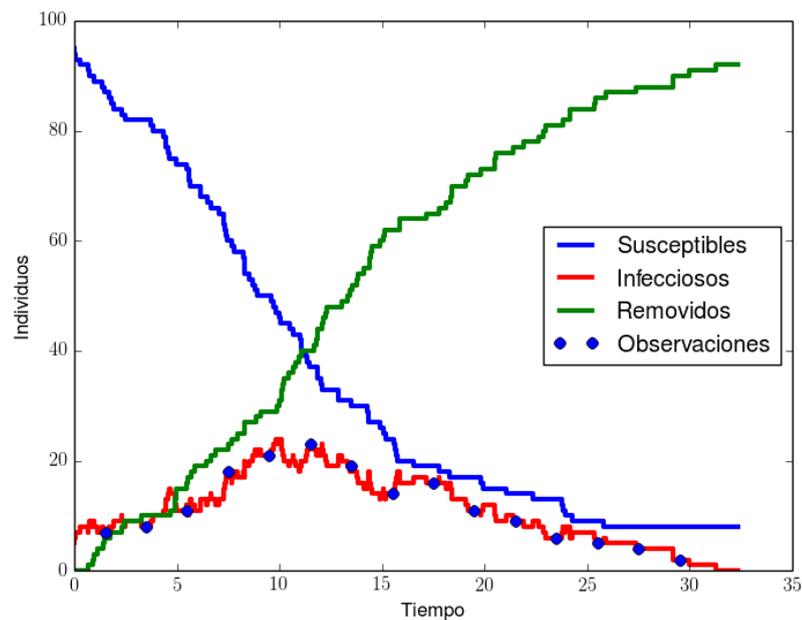


FIGURA 3.3: Ejemplo de un modelo SIR estocástico con parámetro  $\theta = (0.5, 0.25)$ .

En este caso se sabe que los parámetros reales que producen los datos observados son  $\beta = 0.5$  y  $\gamma = 0.25$ , además, al ser esta una simulación se conocen los momentos exactos en los cuales ocurre cada infección y remoción del proceso.

Los estimadores de máxima verosimilitud obtenidos suponiendo información perfecta son los siguientes:

$$\hat{\beta} = \frac{\sum_{i \in I} N_i}{\sum_{i \in I} T_i * S_i * I_i} = 0.622654$$
$$\hat{\gamma} = \frac{\sum_{i \in R} N_r}{\sum_{r \in R} T_r * I_r} = 0.25678$$

donde  $N_i, N_r$  corresponden al número total de observaciones de infección y remoción respectivamente.

Se observa que los resultados, en este caso son muy parecidos a los parámetros reales, adicionalmente, se puede hacer un análisis bayesiano que podría incluir información que se tenga del fenómeno.

### 3.5.1. Recursos para la Simulación

La simulación se programó en el lenguaje Python 2.7 y se corrió en una máquina Desktop, con procesador AMD FX, Quad-Core, 8Gb de memoria RAM en un sistema operativo Windows 7.

## 3.6. Muestreo de Aceptación

En este método las propuestas provienen de una distribución a priori, para ello se trabaja con una distribución Wishart independiente a priori, lo que para cada parámetro corresponde a distribuciones gamma con parámetros  $\Gamma_i(1, 1)$ , cada uno de ellos coincide con una distribución exponencial. Lo anterior con finalidad de tener una distribución a priori que contenga la menor información posible acerca de los parámetros.

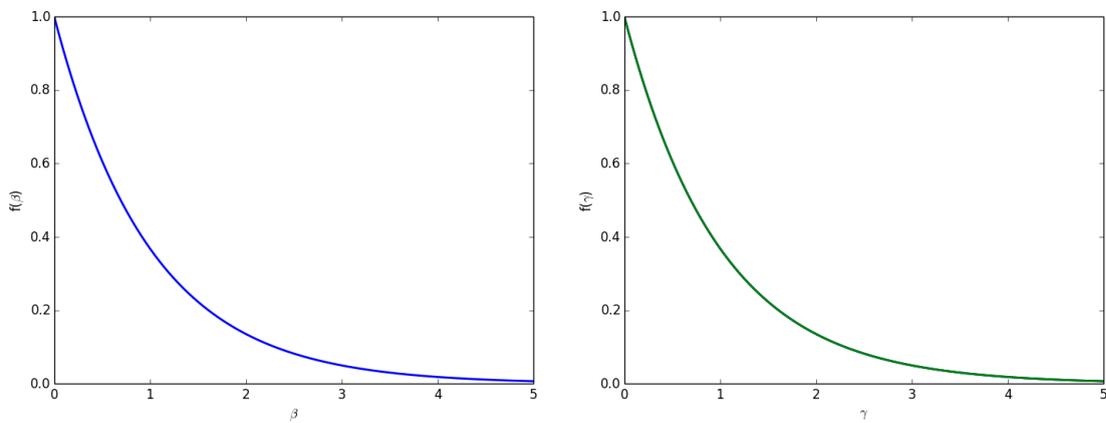


FIGURA 3.4: Distribuciones a priori de los parámetros.

### 3.6.1. Rejection Sampling-Kernel Uniforme-Estadístico Univariado

En el caso donde se establece una estadística univariada como medida de comparación, por cada punto observado se establece una cantidad  $h'$ , la cual servirá como medida de comparación entre simulaciones, de modo que el ancho de banda total permitido para el ejercicio con 15 datos sintéticos tomando  $h' = 5$  resulta  $h = nh' = 15 * 5 = 75$ .

El generar 100 valores tomó 2,362 iteraciones y 48.65 segundos lo que equivale a un 95.76% de rechazos con un ancho de banda de  $h = 75$ , en la Figura 3.5 se observa la distribución de las curvas aceptadas.

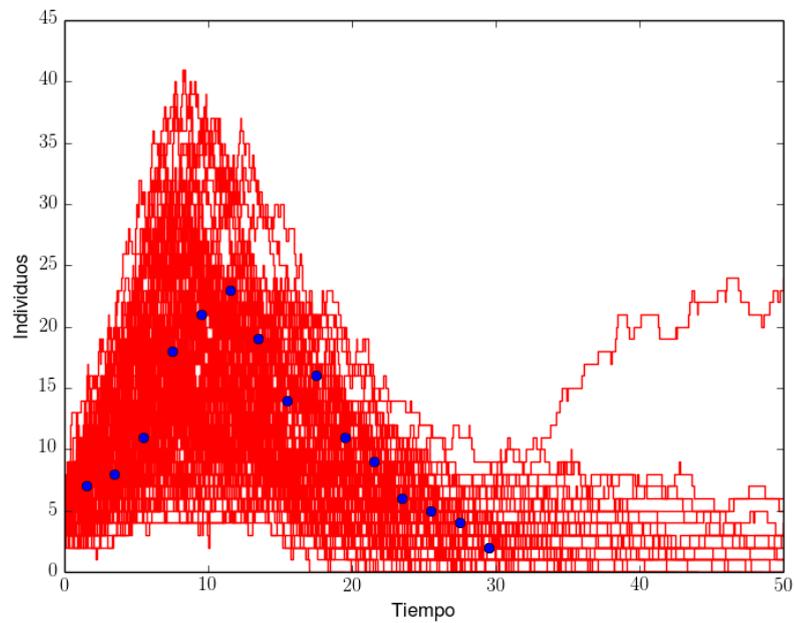


FIGURA 3.5: Distribución de las curvas que fueron aceptadas.

En la Figura 3.6 se puede apreciar los puntos que fueron propuestos y aquellos que fueron aceptados con un ancho de banda univariado  $h = 75$ .

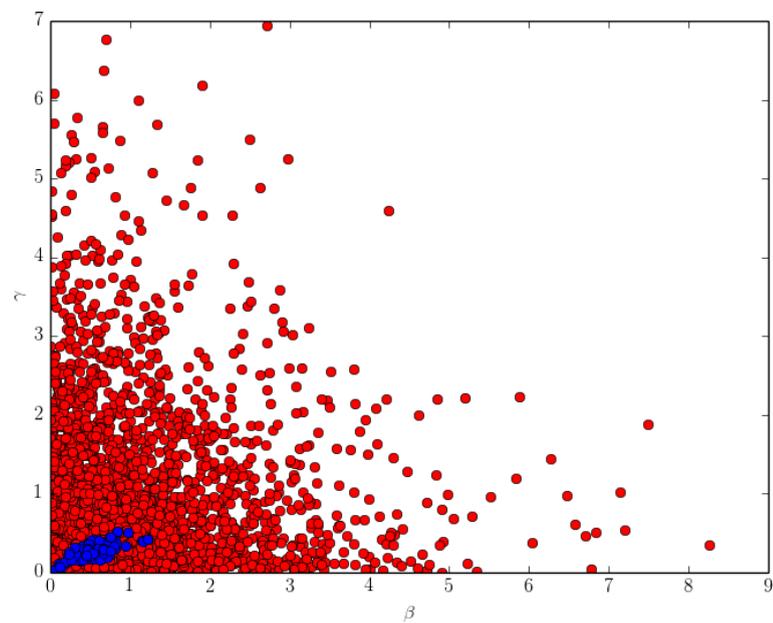


FIGURA 3.6: Puntos rechazados y propuestos.

En la Figura 3.7 se puede observar solamente los parámetros  $\theta^{(i)}$  que producen curvas aceptadas y adicionalmente se grafica con un diamante el punto que corresponde a la media marginal en cada componente del vector de parámetros, dicho estimador resulta en un valor muy próximo al parámetro real que produce los datos; adicionalmente se observa una estructura de correlación entre los parámetros de infección y remoción.

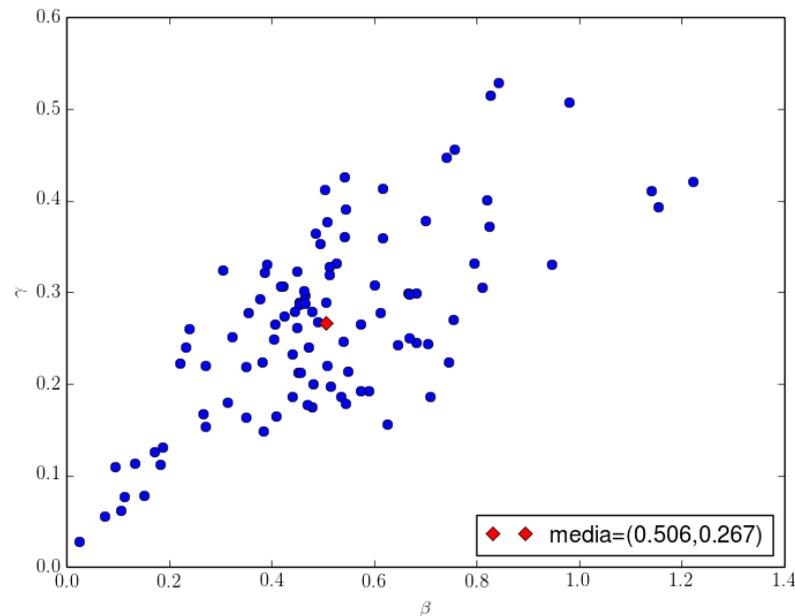


FIGURA 3.7: Distribución de los puntos aceptados.

Al conservar las mismas condiciones y solamente reducir el valor  $h' = 3$ , el ancho de banda se reduce y se produce una mejor aproximación a los datos reales mostrado en la Figura 3.8; sin embargo, como es de esperarse, se produjo una mayor cantidad de rechazos 98.65%, en esta ocasión fue necesario producir 7,417 propuestas, las cuales se evaluaron en un tiempo de 137.04 segundos.

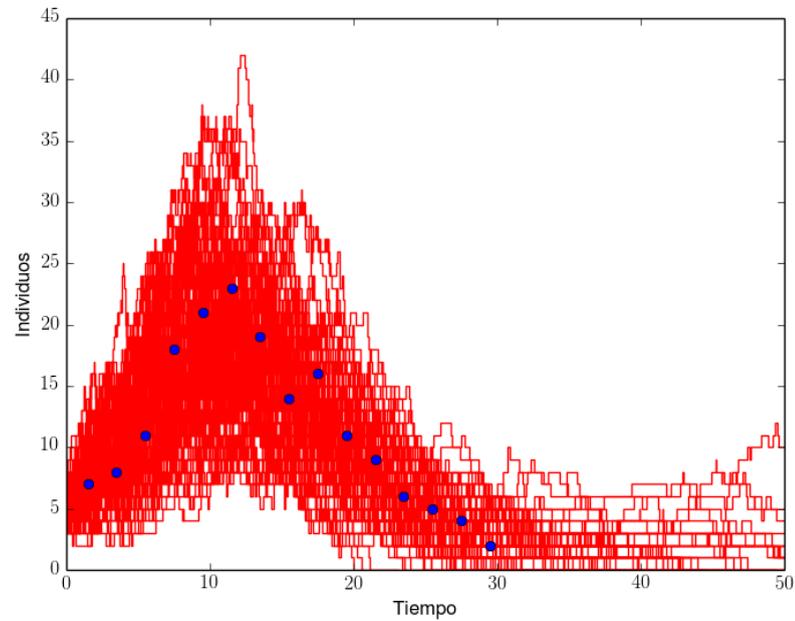


FIGURA 3.8: Distribución de las curvas que fueron aceptadas.

Se observa que las curvas generadas se encuentran más cercanas a los puntos observados, lo que significa que se tiene una mejor distribución posterior aproximada  $\pi_{K,H}(\theta|y)$ . En la Figura 3.9 se puede apreciar los puntos que fueron propuestos y aquellos que fueron aceptados

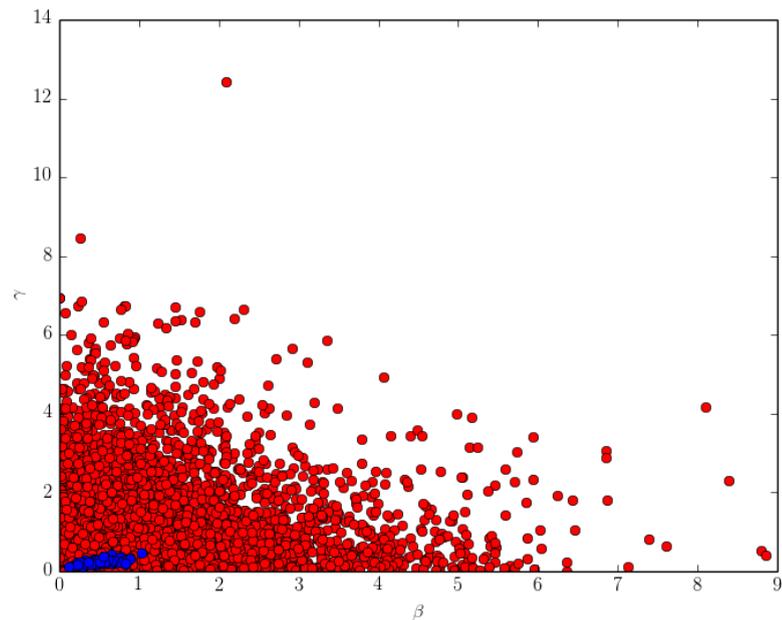


FIGURA 3.9: Puntos rechazados y propuestos.

Al comparar la distribución de los puntos entre los distintos anchos de banda se puede observar que en la de menor ancho de banda, los puntos presentan ligeramente menor dispersión que los de la Figura 3.6.

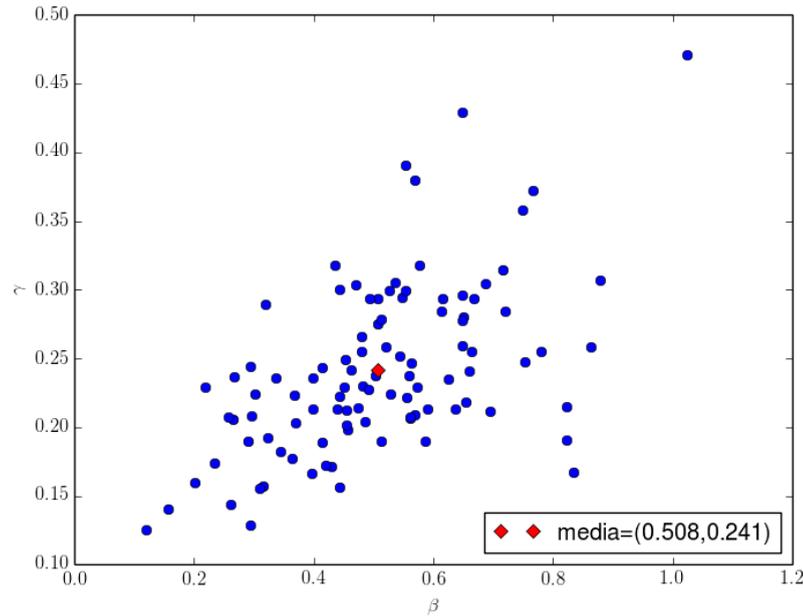


FIGURA 3.10: Distribución de los puntos propuestos.

Una vez que se obtiene una muestra de tamaño suficiente, esto último depende de las características de cada problema, entonces es posible obtener los estimadores que uno desee, ya sean medidas de centralidad, de dispersión o correlación entre variables. En el ejercicio de simulación se extrajo una muestra de tamaño 10,000, para un parámetro bidimensional.

Se puede hacer inferencia sobre cada parámetro, al tomar las densidades marginales, lo cual se observa en la Figura 3.11. Sin embargo, la muestra misma revela una ligera correlación entre los elementos del parámetro.

El tiempo que requirió el algoritmo para generar la muestra de tamaño 10,000 fue de 6,688.61 segundos, el número de parámetros propuestos fue 382,197, resultando en un porcentaje de rechazos del 97.38 %.

La media de cada componente de la distribución marginal para la infección es  $\hat{\beta} = 0.5252$  y para la remoción  $\hat{\gamma} = 0.25493$ , las correspondientes varianzas muestrales marginales son  $\hat{S}_{\beta} = 0.044742993$  y  $\hat{S}_{\gamma} = 0.006864013$ , además calculando el coeficiente de correlación se obtiene  $\hat{\rho}_{\beta,\gamma} = 0.684295$ .

En la Figura 3.11 se observa que los datos han generado aprendizaje en el modelo, las curvas representan la distribución a priori, mientras que la región sombreada es la distribución posterior.

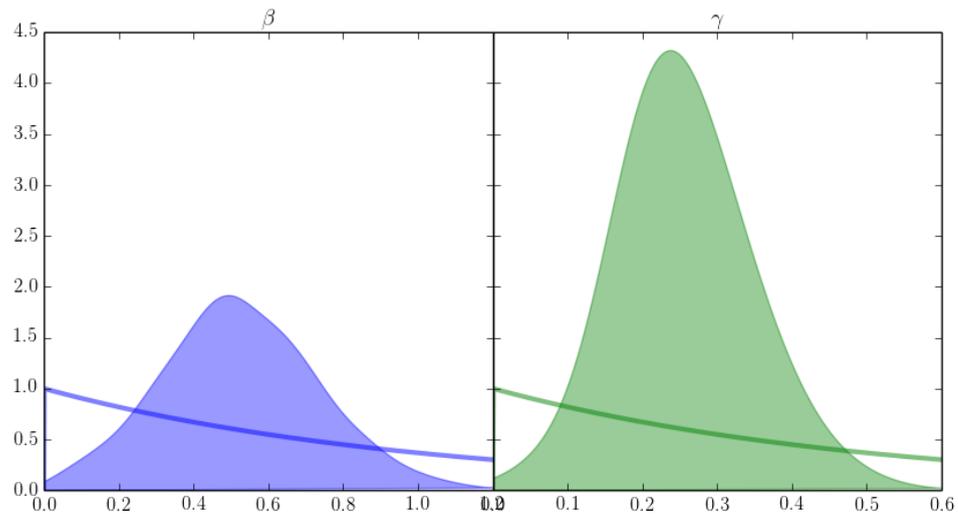


FIGURA 3.11: Distribución a priori y posterior marginal de los parámetros.

La distribución conjunta de los parámetros se puede observar en la siguiente Figura 3.12

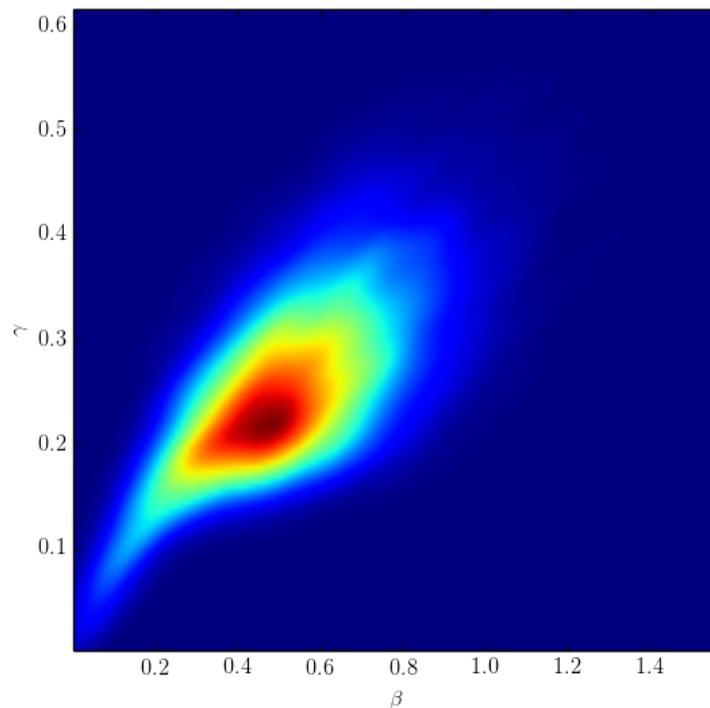


FIGURA 3.12: Distribución conjunta de los parámetros.

El método de Muestreo de Aceptación utilizando kernel Uniforme o Epanechnikov, resulta similar, por lo que no se explota mejoría del segundo en puntos de mayor cercanía con los observados, en el caso del kernel Normal, se muestra una inconsistencia, ya que el pedir que se rechacen ciertos valores después de un umbral, conlleva a utilizar un kernel uniforme con un ancho de banda posiblemente distinto.

### 3.7. Caminata Aleatoria con Metropolis-Hastings: MCMC ABC

Utilizando una modificación del algoritmo para la simulación en el cual se pueda aprovechar la estimación kernel y comparando entre las funciones kernel se desarrolla el siguiente análisis.

Se desarrolló un algoritmo en el que las propuestas fueran caminatas aleatorias con dos tipos de kernel de transición Normal. El primero con una dispersión en la que cada componente explore en puntos lejanos, mientras que el segundo, con una dispersión menor, sea capaz de obtener puntos cercanos y enriquecer la muestra.

En todas las simulaciones Caminata Aleatoria con Metropolis-Hastings o RWMH (*Random Walk Metropolis-Hastings*) se generaron 5,000 propuestas para el algoritmo, posteriormente se considera el tiempo de burn-in y la extracción de una muestra independiente calculando el Tiempo de Autocorrelación Integrada o IAT (*Integrated Autocorrelated Time*).

Se consideran pasos en los cuales no hay correlación en los parámetros, por lo que la matriz de dispersión del kernel de transición se puede ver como:

$$\Sigma_i = \begin{bmatrix} \sigma_{i,1} & 0 \\ 0 & \sigma_{i,2} \end{bmatrix}$$

De tal forma que las propuestas del parámetro  $\theta$  se obtienen mediante:

$$\theta' = \mathbb{I}_{x=0} \mathcal{N}(\theta^{(i-1)}, \Sigma_1) + (1 - \mathbb{I}_{x=0}) \mathcal{N}(\theta^{(i-1)}, \Sigma_2)$$

donde  $x$  se distribuye Bernoulli con probabilidad de éxito  $p = 0.5$ , por lo que se tienen dos tipos de pasos y es posible realizar una exploración del espacio más eficiente. Inicialmente se propone que no haya correlación entre los datos, pero esto se podría corregir conforme se conoce más del proceso.

### 3.7.1. El caso Multidimensional

En el caso donde se utiliza un estadístico multidimensional entre las observaciones y las simulaciones, se utiliza la diferencia elemento a elemento,  $(|I_{Obs,1} - I_{Sim,1}|, \dots, |I_{Obs,p} - I_{Sim,p}|)$ , donde  $p$  es la cantidad de observaciones del fenómeno. Al ser el estadístico resultante de carácter multidimensional, la estimación por kernel utiliza también las formas multivariadas de los mismos, en los cuales se supondrá la matriz de ancho de banda de la siguiente forma.

$$H = \begin{bmatrix} h_{11} & 0 & \cdots & 0 \\ 0 & h_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & h_{pp} \end{bmatrix} = \text{diag}(h_{11}, h_{22}, \dots, h_{pp}), \quad h_{ii} > 0$$

Recordando la distribución de curvas generadas por el verdadero parámetro en la Figura 3.2 se observa que en los puntos más altos se concentra una mayor dispersión de las curvas, por lo que se establecerá que el ancho de banda sea proporcional a la magnitud de cada valor observado en  $I_{Obs}$ .

Al utilizar un kernel uniforme, la propuesta inicial es muy importante, puesto que si está muy alejado del verdadero parámetro y el kernel de transición Markoviano no es lo suficientemente extenso en el soporte, de tal manera que propicie una exploración del espacio, será difícil que se acerque a los verdaderos parámetros, por lo que se opta por los kernels de transición de tipo gaussiano para generar propuestas. Además, se consideran dos kernels de transición, gaussianos, uno con parámetro de dispersión  $\sigma_1 = 0.5$  y el otro con  $\sigma_2 = 0.02$ , de tal manera que se pueda explorar el espacio y también obtener muestras de puntos con probabilidad alta.

### 3.7.2. RWMH-Kernel Uniforme-Estadístico Multivariado

Se realizaron diversas pruebas para comprobar la sensibilidad de los componentes involucrados en la estimación.

En todos los casos se realizan 5,000 valores propuestos sin quitar los puntos correspondientes al burn-in, utilizando un ancho de banda igual a  $H = \text{diag}(0.5I_{Obs})$ , con punto inicial precisamente en los valores exactos del parámetro (0.5,0.25), se obtuvieron muy pocos puntos que aparecen en la Figura 3.14, que producen las curvas de la Figura 3.13.

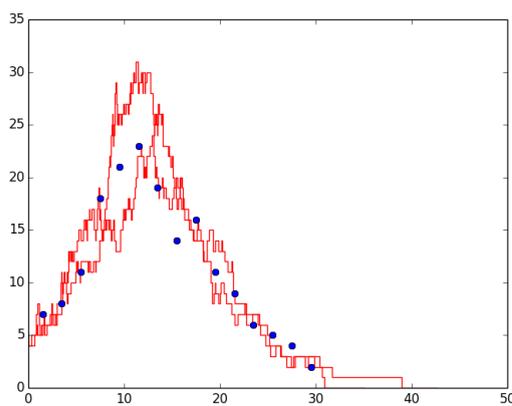


FIGURA 3.13: Curvas de parámetros aceptados.

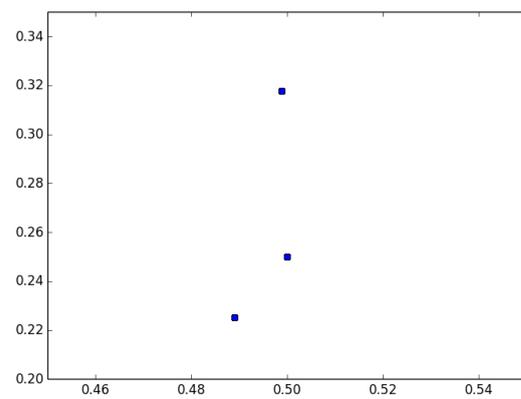


FIGURA 3.14: Parámetros aceptados.

Lo cual indica que el valor 0.5 es muy restrictivo, por lo que se amplía  $H = \text{diag}(0.6I_{Obs})$ , de igual manera se considera que el punto inicial es el parámetro exacto (0.5,0.25), resultando las distribuciones de la Figura 3.15

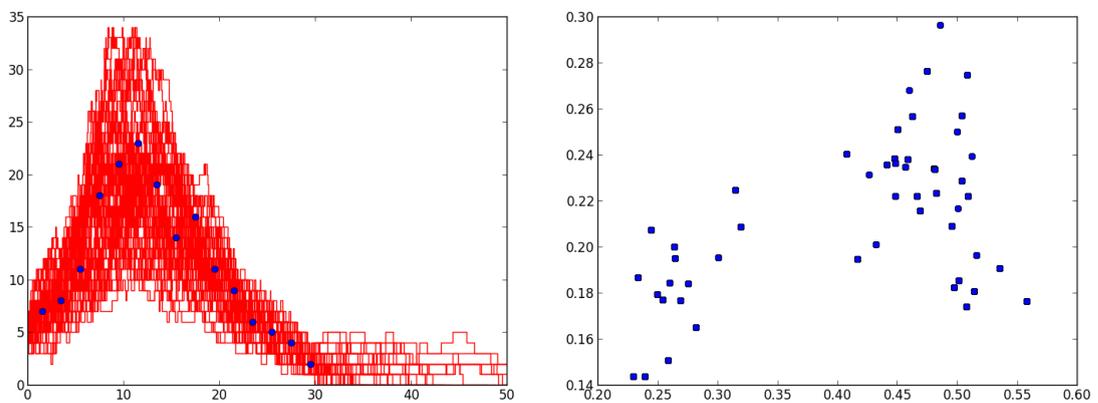


FIGURA 3.15: Distribución de las curvas aceptadas (izquierda) y Parámetros aceptados (derecha).

Bajo las mismas condiciones, solamente al cambiar el punto inicial por  $(5,5)$ , y al realizar varias veces el ejercicio no se obtuvieron valores aceptados, adicionalmente se realizó el ejercicio con punto inicial en  $(2,1)$ , con los mismos resultados, esto indica la gran sensibilidad del modelo a iniciar en puntos alejados.

Se exploró la opción de cambiar a un kernel  $H = \text{diag}(0.75I_{Obs})$ , con punto inicial el parámetro exacto. El resultado se observa en las Figuras 3.16 y 3.17.

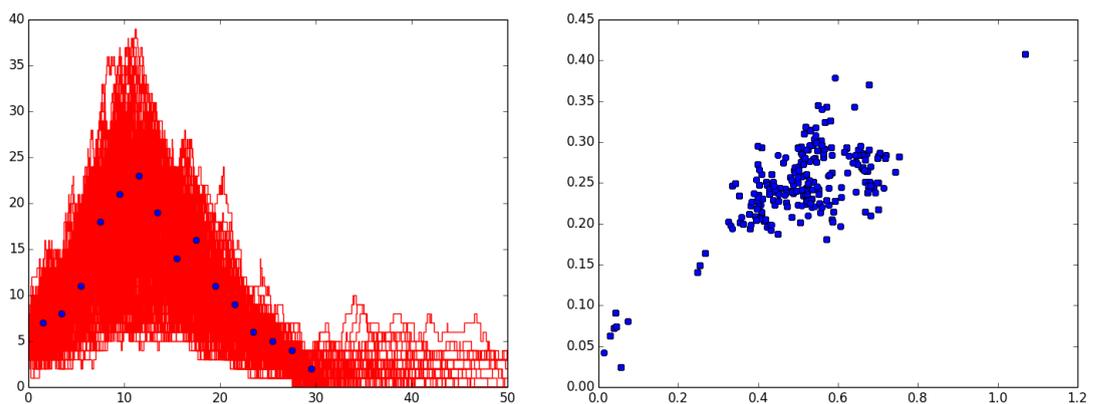


FIGURA 3.16: Distribución de las curvas aceptadas y Parámetros aceptados.

En la Figura 3.17 el punto inicial es  $(1,1)$  donde se observa que conforme el punto inicial esté más alejado del verdadero parámetro, en promedio se necesitará de

una mayor cantidad de simulaciones para alcanzar el burn-in, esto se nota en la densidad de las curvas, respecto de las de las generadas por la Figura 3.16.

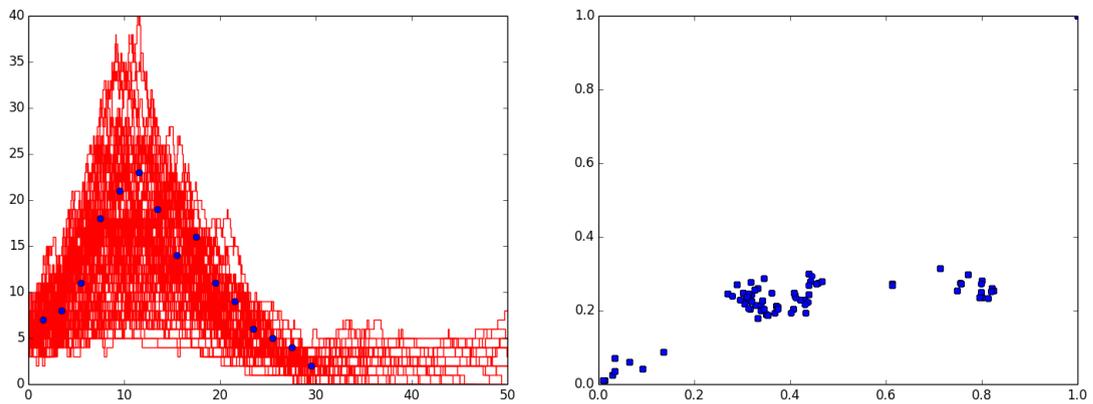


FIGURA 3.17: Distribución de las curvas aceptadas y Parámetros aceptados.

### 3.7.3. RWMH-Kernel Epanechnikov-Estadístico Multivariado

Este kernel tiene problemas parecidos al kernel uniforme para alcanzar la región que abarca el ancho de banda, pero una vez que alcanza dicha región, es más fácil acercarse a puntos de probabilidad alta, cosa que no pasa con el kernel uniforme, en el cual sólo importa que las discrepancias estén en la región que abarca el ancho de banda.

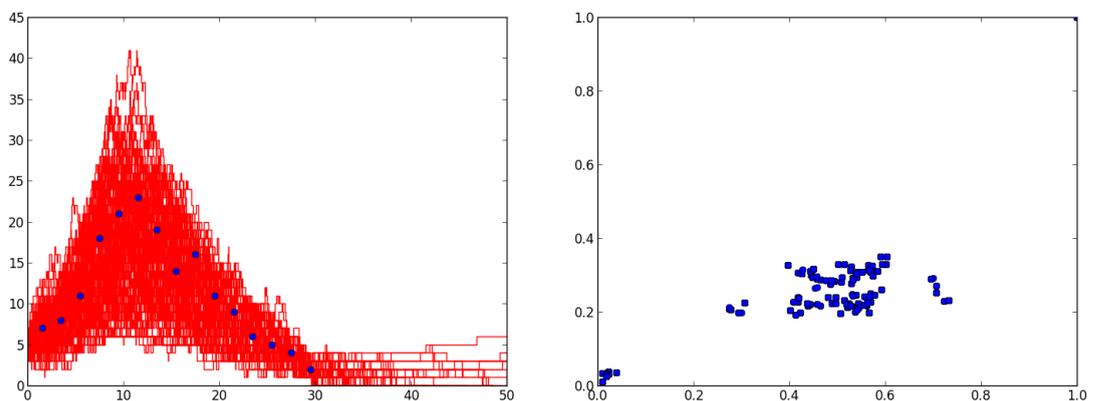


FIGURA 3.18: Distribución de las curvas aceptadas y Parámetros aceptados.

En la Figura 3.18 se observa la simulación con instancias similares a las vistas para el kernel uniforme, en este caso el ancho de banda se establece por  $H = \text{diag}(0.75I_{Obs})$  y el punto inicial es (1,1).

En el experimento de cambiar el punto inicial por el (2,2), todas las propuestas fueron rechazadas.

En la Figura 3.19 se muestra que tampoco es bueno permitir un ancho de banda demasiado grande,  $H = \text{diag}(1.5I_{Obs})$ , en este caso el punto inicial es (2,2).

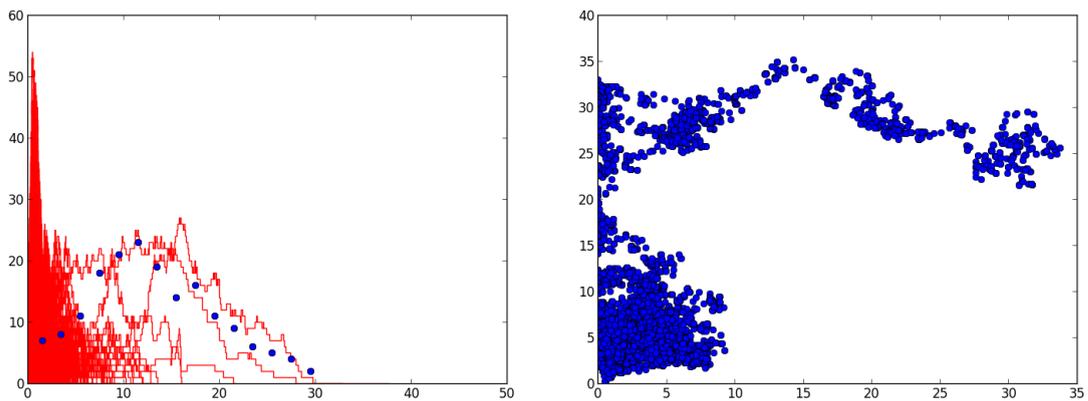


FIGURA 3.19: Distribución de las curvas aceptadas y Parámetros aceptados.

Algo conveniente sería que una vez alcanzada la región y se obtengan puntos de alta probabilidad, se reduzca el tamaño del ancho de banda.

### 3.7.4. RWMH-Kernel Normal-Estadístico Multivariado

Este kernel resulta más robusto puesto que se tiene un soporte más extenso y es posible comparar puntos en cualquier parte del espacio, la ventaja es que por las propiedades del recocido simulado involucrado en el algoritmo Metropolis-Hastings, es posible acercarse a puntos de probabilidad alta de manera eficiente.

En la siguiente Figura se observa la evolución de las propuestas, en las que se comienza con el punto inicial (5,5), el cual es muy lejano del valor real, el ancho de banda se fija por  $H = \text{diag}(0.75I_{Obs})$ .

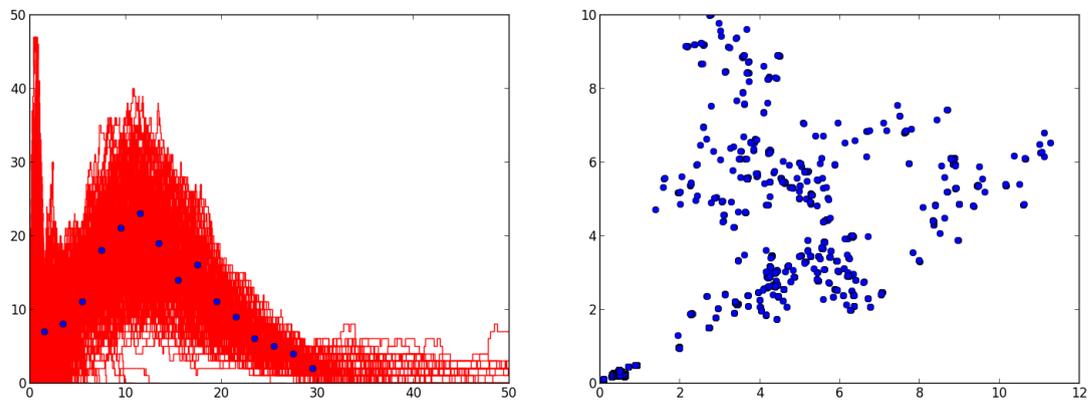


FIGURA 3.20: Distribución de las curvas aceptadas y Parámetros aceptados.

En la siguiente Figura se observa un ancho de banda más reducido  $H = \text{diag}(0.6I_{Obs})$ , con punto inicial (1,1). Resalta la lejanía de una de las curvas en el cuadro de la izquierda de la Figura 3.21 la distribución del punto inicial.

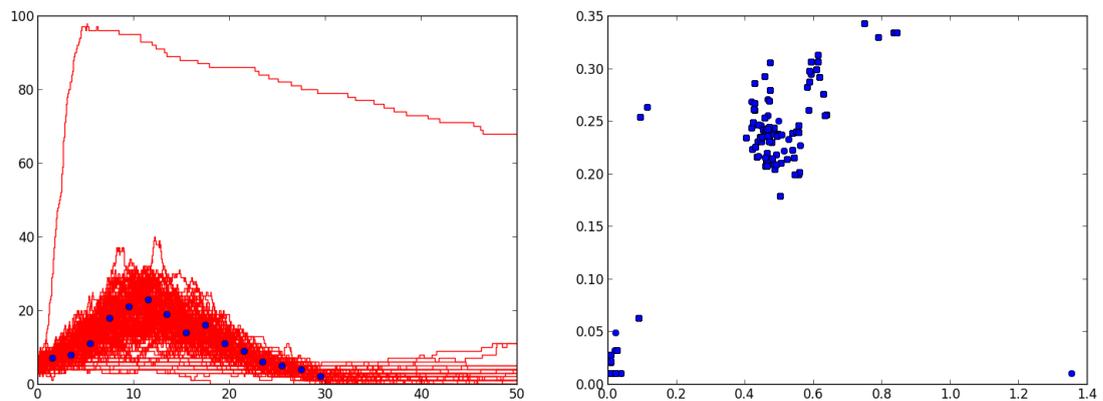


FIGURA 3.21: Distribución de las curvas aceptadas y Parámetros aceptados.

Otra ventaja de utilizar kernel normal se puede observar al reducir el ancho de banda de tal manera que  $H = \text{diag}(0.1I_{Obs})$ , con punto inicial (1,1). Recordando que con el kernel uniforme el factor más chico explorado fue 0.5 y casi todo fue rechazado, en este caso se tiene un valor de 0.1.

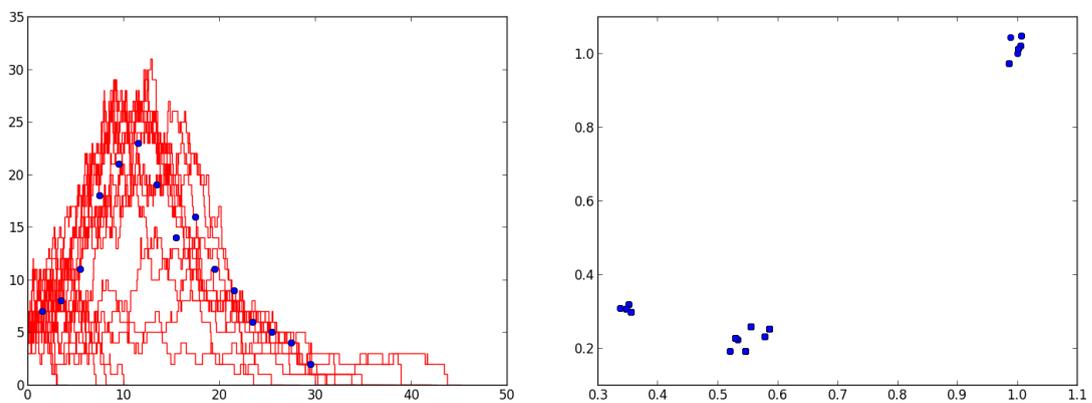


FIGURA 3.22: Distribución de las curvas aceptadas y Parámetros aceptados.

### 3.7.5. Caso del Estadístico Univariado

El estadístico utilizado para llevar la comparación de forma univariada es la siguiente  $\sum_{i=1}^p \frac{(I_{Obs,i} - I_{Sim,i})^2}{I_{Obs,i}}$ , donde  $p$  es la cantidad de observaciones.

El ancho de banda se establece como un factor que se multiplicará por el número de observaciones, esto es,  $H = fp$ , en este caso  $p = 15$ .

#### 3.7.5.1. RWMH-Kernel Uniforme-Estadístico Univariado

Como en el caso multidimensional se exploró la sensibilidad del ancho de banda y el punto inicial en el cual se comienza la simulación.

En primer lugar se probó utilizando un ancho de banda  $H = 0.5 * d = 0.5 * 15 = 7.5$ , y comenzando en el punto (1,1). No se obtuvo ninguna aceptación. Al incrementar el ancho de banda a  $H = 1 * 15 = 15$  tampoco se aceptaron puntos; sin embargo, para el ancho de banda  $H = 1.5 * 15 = 22.5$  se obtuvieron los siguientes resultados:

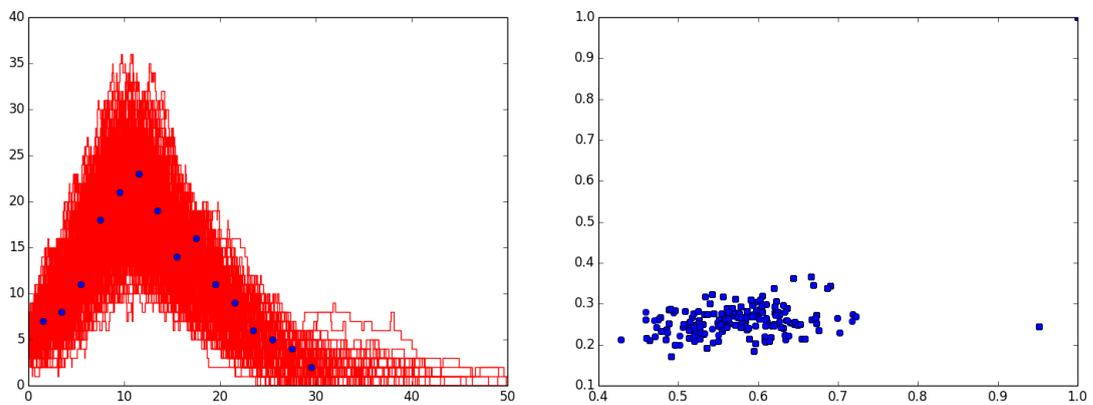


FIGURA 3.23: Distribución de las curvas aceptadas y Parámetros aceptados.

Es de notar que la dispersión comparada con el kernel multivariado resulta visualmente menor.

Al cambiar el ancho de banda a  $H = 2.5 * 15 = 37.5$  y conservar el punto inicial en (1,1), se presentó el resultado de la Figura 3.24.

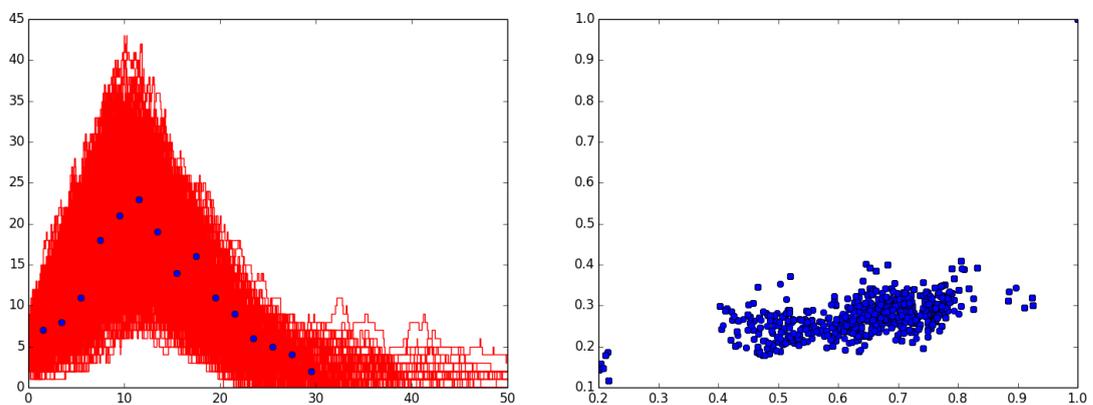


FIGURA 3.24: Distribución de las curvas aceptadas y Parámetros aceptados.

Lo cual resulta en una mayor cantidad de puntos aceptados respecto de la forma multivariada, al conservar el ancho de banda y explorar la convergencia en un punto más lejano (2,2), el resultado fue el que se muestra en la Figura 3.25.

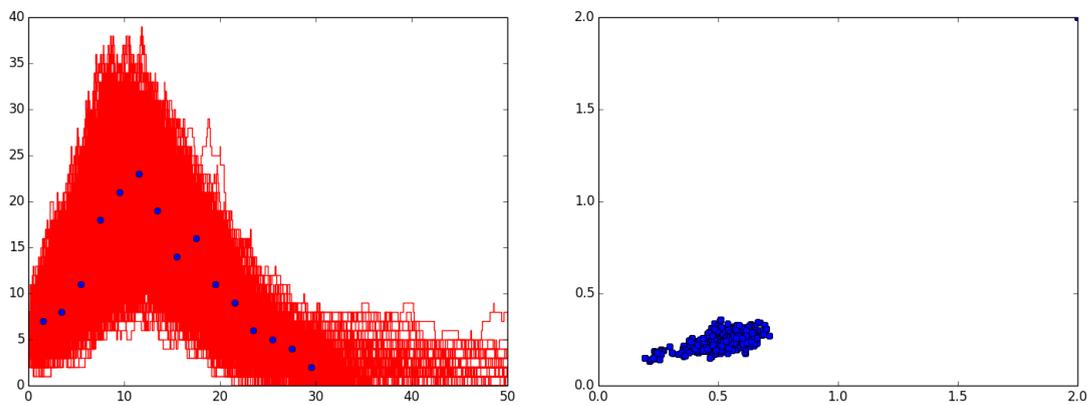


FIGURA 3.25: Distribución de las curvas aceptadas y Parámetros aceptados.

El utilizar un estadístico univariado permite una mejor aproximación a los datos, puesto que se evalúa de manera global la dispersión total permitida que evalúa el ancho de banda.

### 3.7.5.2. RWMH-Kernel Epanechnikov-Estadístico Univariado

Los resultados de esta implementación resultan parecidos a los de utilizar un kernel uniforme; sin embargo, una vez que se encuentra dentro del ancho de banda, las propuestas se enfocan en puntos de mayor probabilidad.

Considerando un ancho de banda  $H = 0.5 * 15 = 7.5$  y punto inicial (1,1) se obtuvo la siguiente distribución, que arroja pocos datos aceptados.

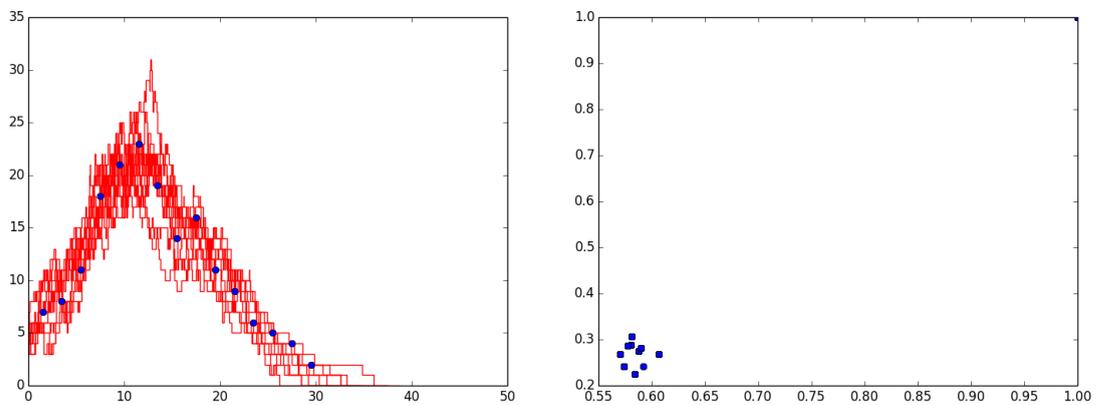


FIGURA 3.26: Distribución de las curvas aceptadas y Parámetros aceptados.

La cantidad de puntos aceptados es muy poca, por lo que se aumenta el ancho de banda a  $H = 15$ , conservando el punto inicial  $(1,1)$ , el resultado se observa en la Figura 3.27

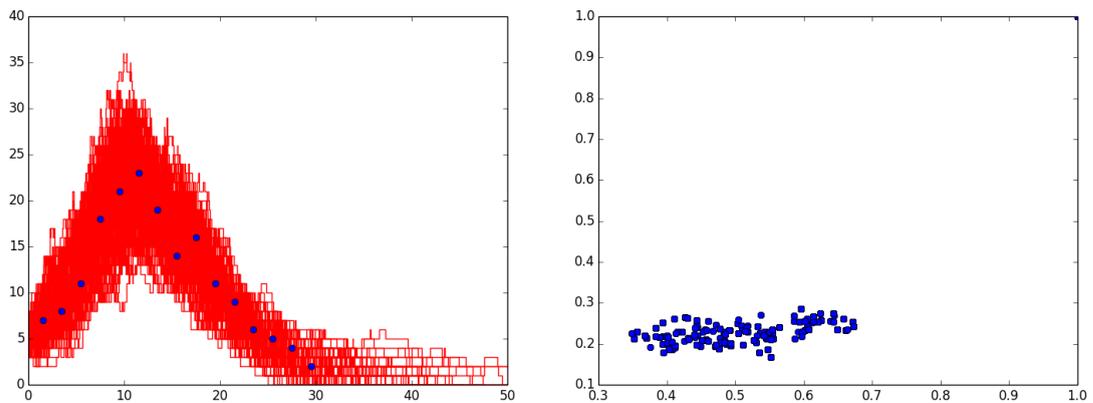


FIGURA 3.27: Distribución de las curvas aceptadas y Parámetros aceptados.

En la Figura 3.28 se observa una mejoría en la obtención de la muestra, de manera similar con el kernel uniforme se propone un ancho de banda  $H = 1.5 * 15 = 22.5$ .

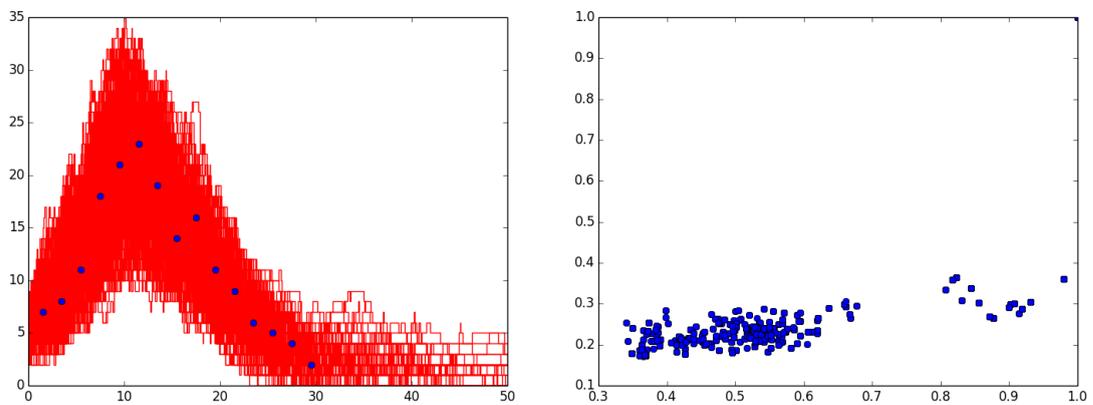


FIGURA 3.28: Distribución de las curvas aceptadas y Parámetros aceptados.

Finalmente se compara con un ancho de banda  $H = 2.5 * 15 = 37.5$ , el resultado mostrado en la Figura 3.29 presenta mayor dispersión como se espera.

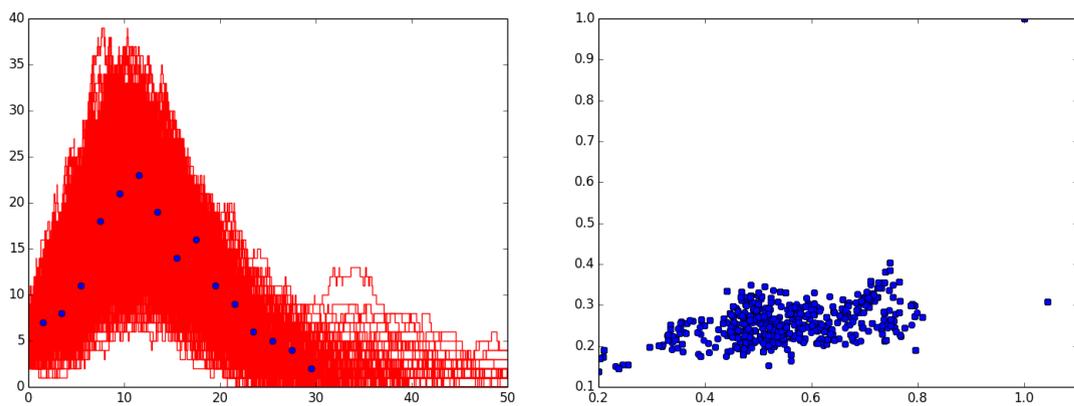


FIGURA 3.29: Distribución de las curvas aceptadas y Parámetros aceptados.

### 3.7.5.3. RWMH-Kernel Normal-Univariado

En el caso multivariado se mostró que el utilizar el kernel normal resulta más robusto para encontrar puntos que reflejen curvas que se parezcan a las observaciones.

Explorando para un ancho de banda pequeño en el contexto de este problema  $H = 0.1 * 15 = 1.5$  con punto inicial para la simulación (1,1), se obtiene la Figura 3.30.

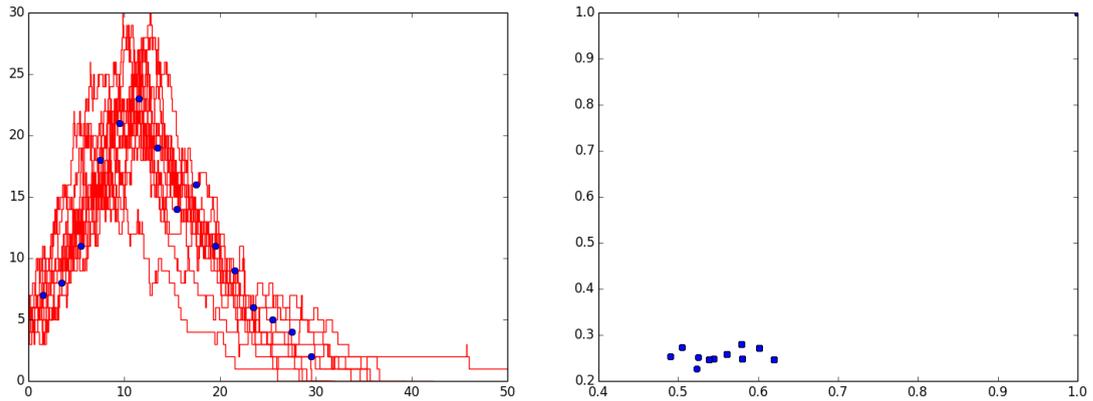


FIGURA 3.30: Parámetros aceptados  $H = 0.1 * 15 = 1.5$ ,  $\theta^{(0)} = (1, 1)$ .

Al experimentar con puntos iniciales que se encuentran alejados se observa que el método se desempeña muy bien, acercando rápidamente a curvas con valores muy cercanos a los datos observados.

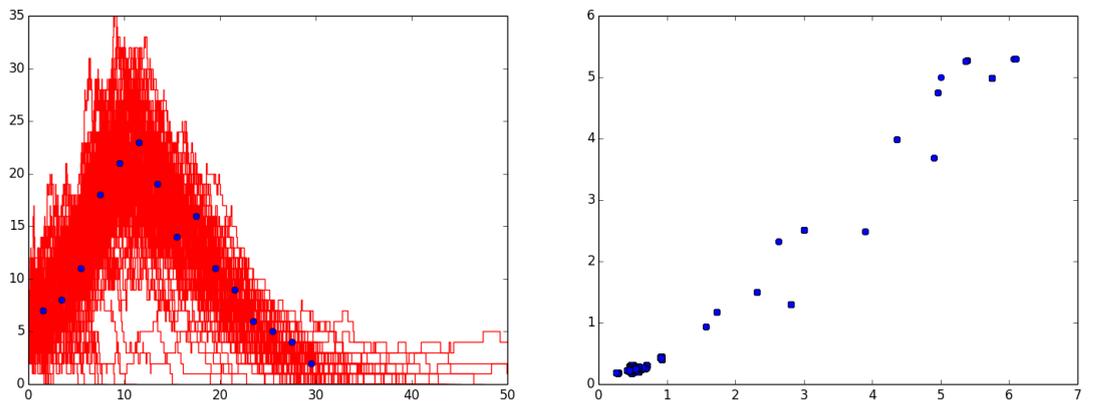


FIGURA 3.31: Parámetros aceptados  $H = 0.5 * 15 = 7.5$ ,  $\theta^{(0)} = (5, 5)$ .

Después de las Figuras anteriores, resulta claro que con un ancho de banda más holgado y un punto inicial no tan alejado, se obtiene un buen desempeño del algoritmo.

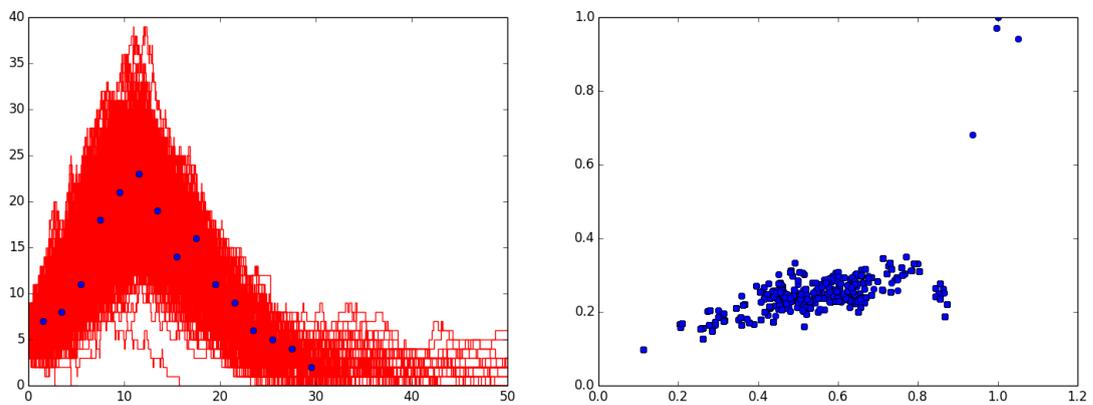


FIGURA 3.32: Parámetros aceptados  $H = 1 * 15 = 15$ ,  $\theta^{(0)} = (1, 1)$ .

En la siguiente Figura se experimenta con un ancho de banda más grande y el mismo punto inicial; es de notar que se preserva la forma y la distribución es muy parecida a la de las curvas generadas por el parámetro exacto.

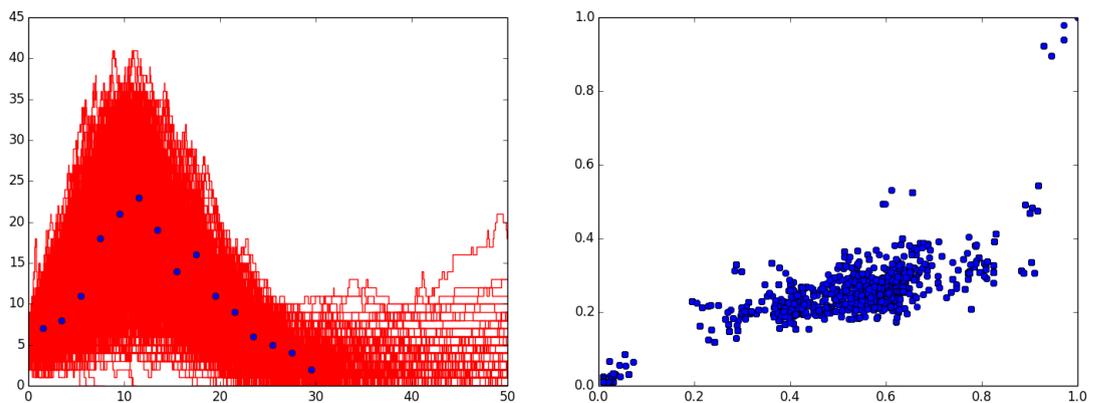


FIGURA 3.33: Parámetros aceptados  $H = 2 * 15 = 30$ ,  $\theta^{(0)} = (1, 1)$ .

Contemplando en este caso un punto relativamente lejano  $\theta^{(0)} = (5, 5)$  se puede observar en la Figura 3.34 de las curvas descritas por los parámetros aceptados que a la izquierda se generan curvas delgadas, las cuales provienen de los parámetros aceptados en el burn-in. Una vez que la distribución converge, se observa que las curvas se ajustan de manera cercana a los puntos observados.

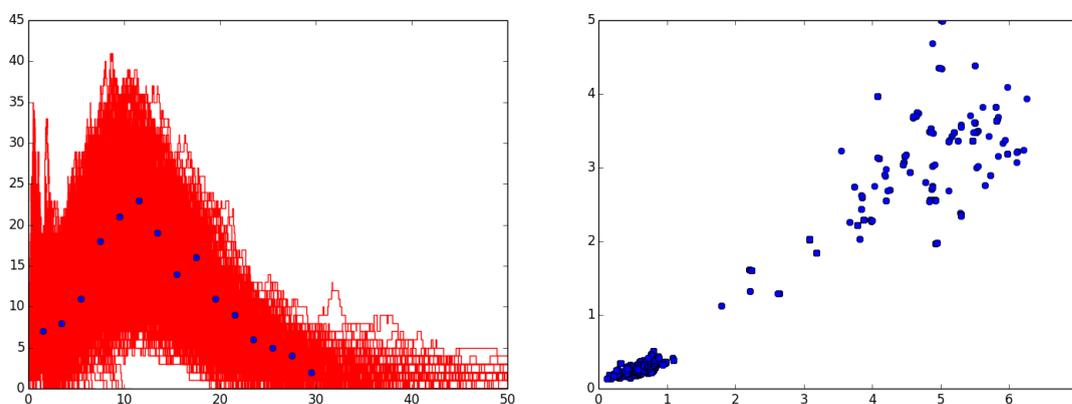


FIGURA 3.34: Parámetros aceptados  $H = 2 * 15 = 30$ ,  $\theta^{(0)} = (5, 5)$ .

El resultado es considerablemente mejor que las otras funciones kernel vistas anteriormente por ser más robusto y permitir anchos de banda más reducidos que los anteriores.

Con el fin de mostrar la robustez de hacer uso de un kernel normal, utilizando un estadístico univariado y un kernel Gaussiano se presenta una simulación utilizando una Caminata Aleatoria con Metropolis-Hastings, en el cual el punto inicial se encuentra muy lejano al parámetro que origina las observaciones, el puntón inicial  $\theta^{(0)} = (50, 50)$ . El ancho de banda utilizado fue de  $H = 1.75 * 15$ , como resultado se obtuvo la caminata aleatoria que se muestra en la Figura 3.35.

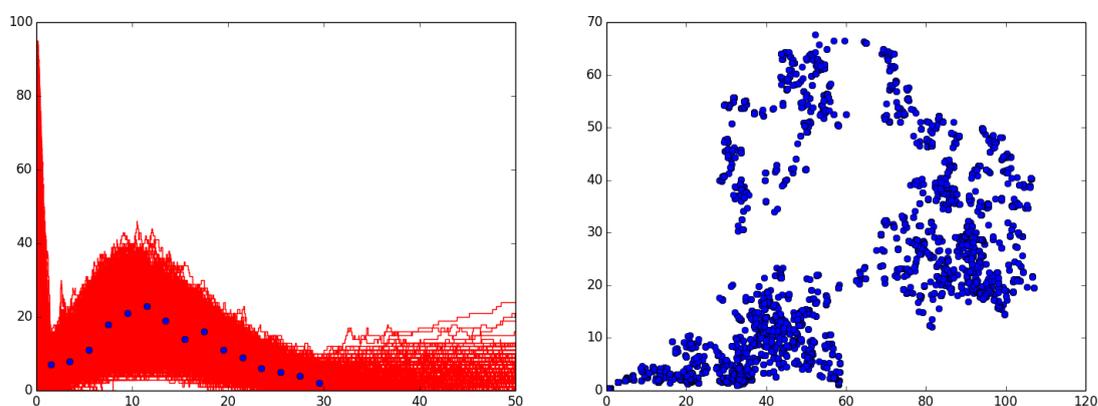


FIGURA 3.35: Parámetros aceptados  $H = 1.75 * 15 = 26.25$ ,  $\theta^{(0)} = (50, 50)$  con burn-in.

Como puede observarse en la gráfica de la izquierda de la Figura 3.35 hay algunas trayectorias que no se parecen a los datos observados; no obstante, estos representan las simulaciones antes de quitar el burn-in, después de quitar los parámetros propuestos antes de la convergencia se obtienen las gráficas de la Figura 3.36.

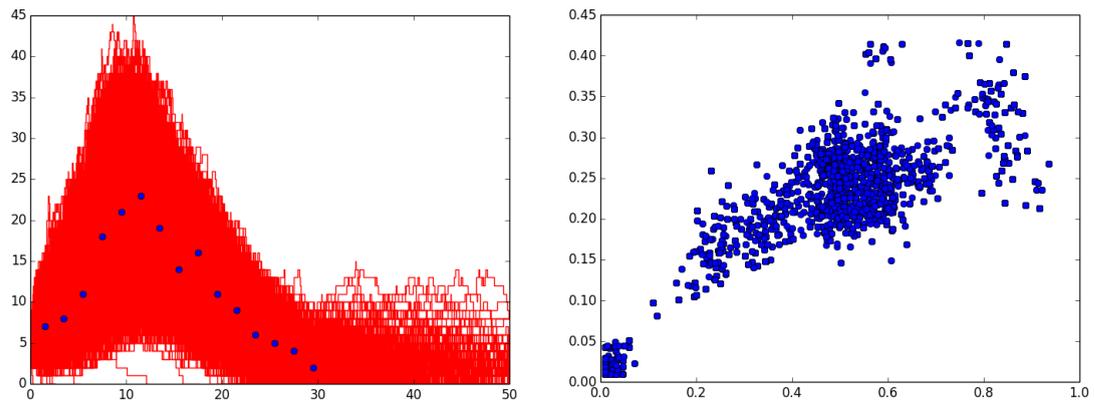


FIGURA 3.36: Parámetros aceptados  $H = 1.75 * 15 = 26.25$ ,  $\theta^{(0)} = (50, 50)$  quitando el burn-in.

En la Figura 3.37 se observa la distribución conjunta antes de quitar el burn-in a la izquierda y a la derecha después del burn-in, en la que los puntos blancos corresponden a las propuestas aceptadas por el algoritmo, a pesar de la larga trayectoria, se observa que la masa de los puntos se concentra cerca del parámetro real.

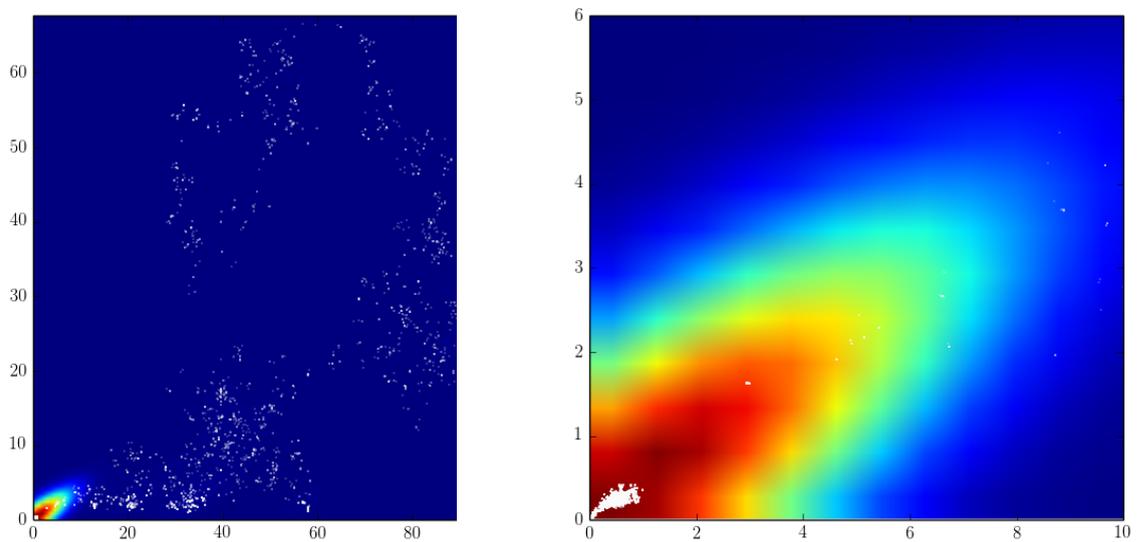


FIGURA 3.37: Distribución conjunta antes y después del burn-in.

Al quitar el burn-in se observa con mayor detalle la densidad acumulada de las observaciones, la cual se centra cerca de los parámetros reales.

A continuación se desea mostrar una muestra considerable de la distribución posterior aproximada, utilizando un estadístico univariado, kernel Normal, mediante el método de simulación de Caminata Aleatoria con Metropolis-Hastings, el punto inicial que se elige es (1,1) y el valor  $\alpha = 1.75$ , los resultados se muestran a continuación.

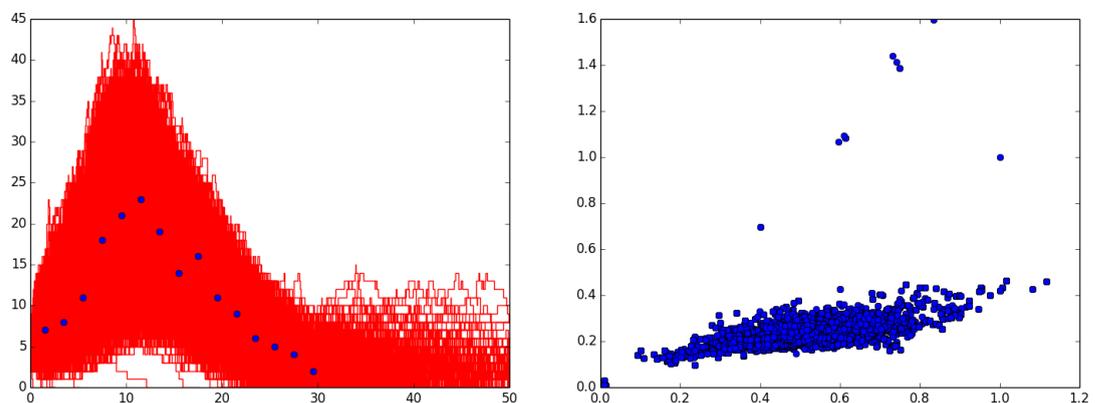


FIGURA 3.38: Parámetros aceptados  $H = 2 * 15 = 30$ ,  $\theta^{(0)} = (5, 5)$ .

Es muy notorio que los puntos se sitúan alrededor del parámetro que generó los datos, dicho parámetro es  $(0.5, 0.25)$ . En la Figura 3.39 se muestra la distribución conjunta de la muestra.

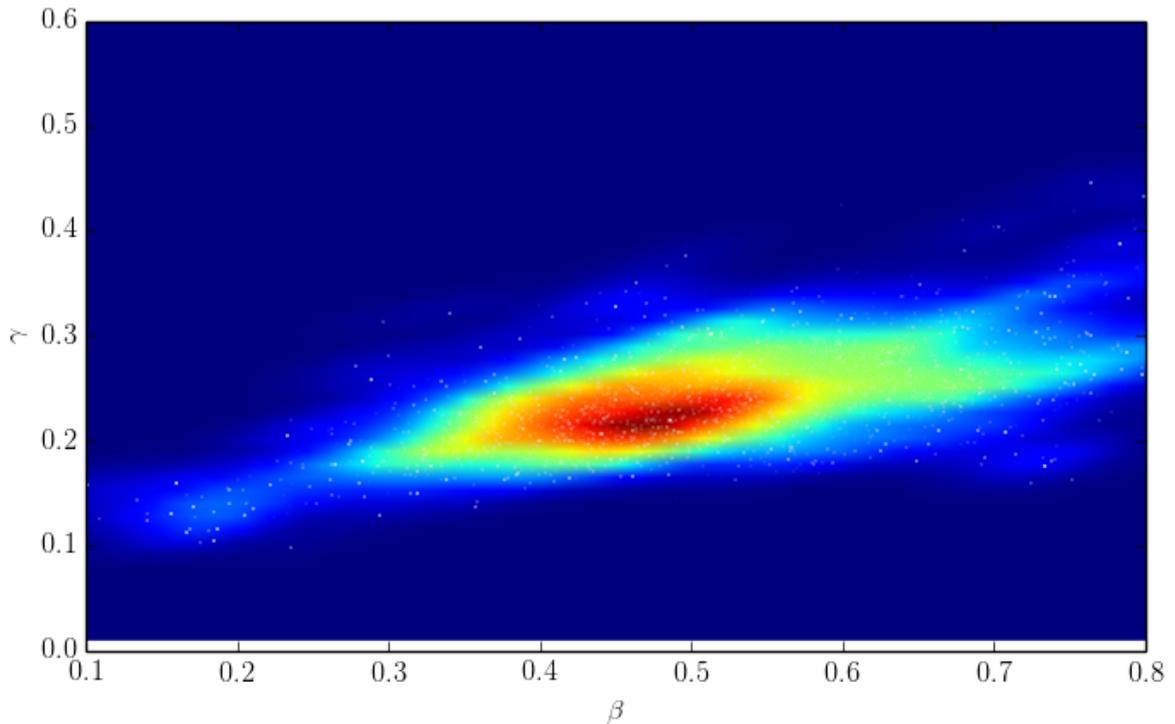


FIGURA 3.39: Distribución conjunta de la muestra

Al graficar las densidades marginales de cada parámetro se generan las gráficas de la Figura 3.40, donde la gráfica de la izquierda corresponde al parámetro de infección, en el cual la media de la muestra es  $\hat{\beta} = 0.5323$ , valor muy cercano al verdadero parámetro  $\beta = 0.5$ . Por otro lado, el parámetro de recuperación se muestra en la gráfica de la derecha, el promedio de la muestra es  $\hat{\gamma} = 0.2452$ , el cual también es muy cercano al verdadero valor  $\gamma = 0.25$ .

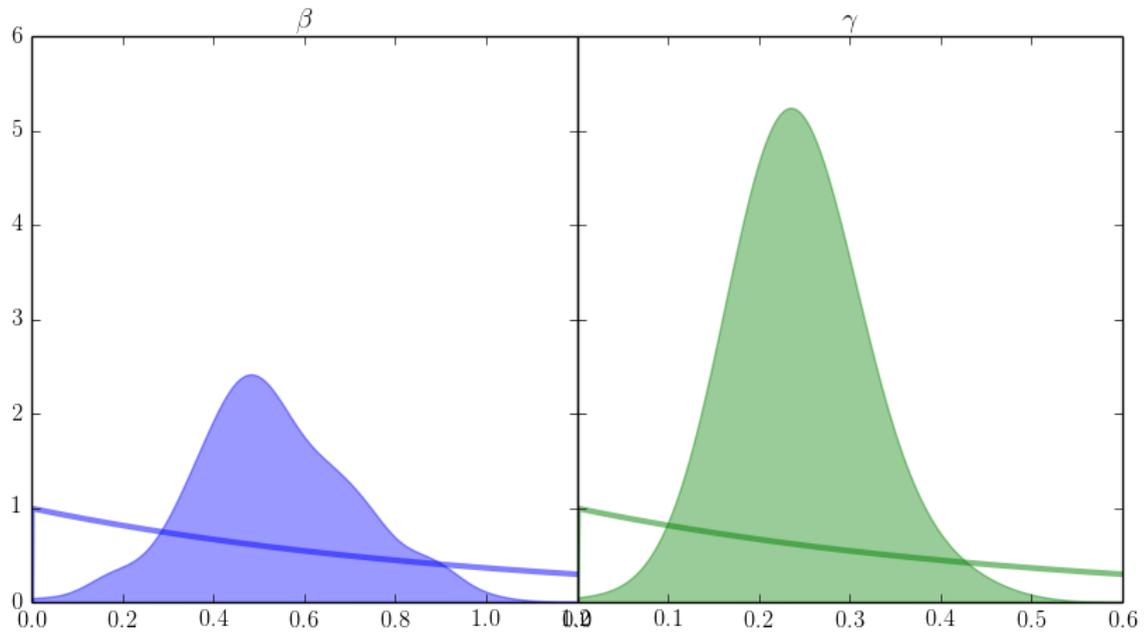


FIGURA 3.40: Distribución conjunta de la muestra.

### 3.8. ABC SMC

Una alternativa para mejorar la estimación de los métodos anteriores correspondería en tratar de combinar ambos métodos. Esta idea se explota secuenciando, puesto que se desearía tomar puntos de acuerdo con cierta propuesta, discriminando los que no son de interés y, en aquellos que sí son puntos que presentan similitud con las observaciones, realizar una exploración en vecindades de dichos puntos. Además, se desearía que el ancho de banda se ajustara conforme evoluciona el algoritmo, de tal manera que cada vez se tenga una mejor aproximación a la distribución posterior real.

Inicialmente se consideran parámetros provenientes de una distribución a priori, dichos puntos se evalúan y resultan en realizaciones del proceso, de estas realizaciones se obtienen aquellas que cumplan con un ancho de banda inicial  $H_1$ , las curvas se ponderan de acuerdo con la estimación kernel y de ellas se remuestrea para obtener las propuestas que contengan mayor probabilidad.

En la Figura 3.41 se observan los puntos por los que pasan las distribuciones propuestas, y se marcan las trayectorias que quedan dentro del ancho de banda inicial.

Posteriormente se toman los parámetros aceptados y se perturban en regiones cercanas y se ajusta el ancho de banda. Lo anterior se realiza de forma secuencial, para finalmente obtener algo como se muestra en las Figuras 3.41 y 3.42.

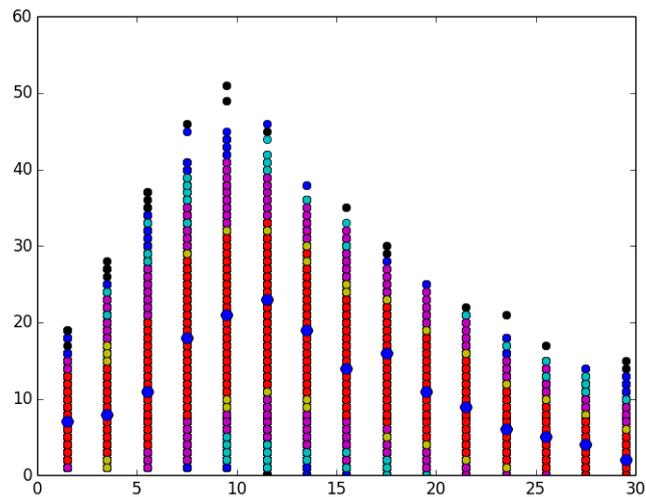


FIGURA 3.41: Distribución de las curvas con SMC.

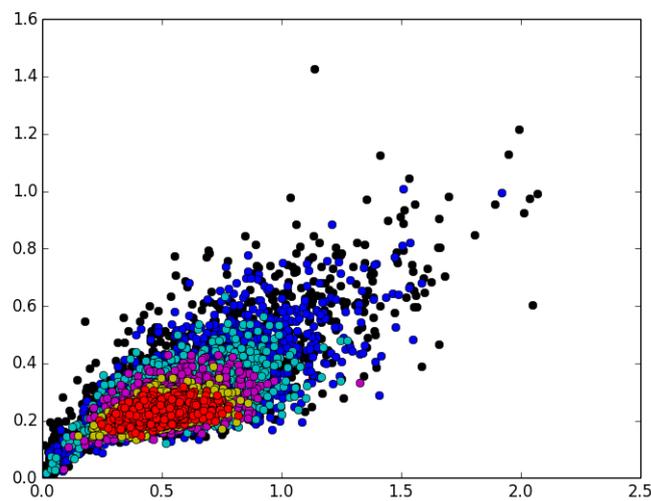


FIGURA 3.42: Distribución del parámetro.

En la Figura 3.42 se puede observar cómo va evolucionando la distribución tanto de los puntos como de las curvas y se perciben como capas en las que se mejora la precisión de la distribución aproximada  $\pi_H(\theta|y)$ .

### 3.9. Comparación entre Métodos

El utilizar un estadístico univariado simplifica el hecho de observar las discrepancias una a una, como lo hace el estadístico multivariado, además se observa una discrepancia global, en lugar de discrepancias individuales, lo cual es bueno, ya que permite cierta flexibilidad en el algoritmo. Resulta más natural el poder tener discrepancias globales, pues cada una de nuestras observaciones pesa lo mismo y el que se cometa cierto error en una de ellas, a pesar de que el resto tuviera una muy buena aproximación, resultando en descartar la propuesta, se distingue por ser un criterio muy restrictivo. Además, el establecer los tamaños de las componentes de la matriz de ancho de banda en kernel multivariado, resultan un trabajo extenuante. Es común reducir la dimensionalidad de algunos problemas, aunque ello implique en algunos casos la pérdida de información, en este caso, lo que se gana es un poco de eficiencia, pues se simplifica el trabajo de simulación.

Se realizaron distintos ejercicios, en los cuales se comprobó la robustez de utilizar estimación de densidad por Kernel, y sobre todo, el hacerlo utilizando un kernel que tuviera un soporte más extenso, la eficacia radica en el hecho de que es posible comparar las propuestas respecto al parámetro en el cual se encuentra en el algoritmo, no importando la lejanía entre la propuesta y el punto actual. En el caso de distribuciones que están truncadas no es posible comparar los puntos cuando la propuesta genera valores fuera del soporte de la distribución. Siendo ésta una dificultad, lo más conveniente es utilizar un kernel que tenga un soporte lo más amplio posible, en el caso abordado en este trabajo, utilizar un kernel Normal generó muy buenos resultados.

Finalmente, el método de muestreo, en el cual se mostró que el de Muestreo por Aceptación está muy limitado y no permite la utilización de un kernel más extenso; al comparar éste método con la Caminata Aleatoria con Metropolis-Hastings, se observó que la caminata tiene un mejor desempeño, se explotó el utilizar diferentes tipos de pasos como propuestas, tanto para explorar el espacio como para obtener valores que tuvieran alta probabilidad, lo que permitió una mejor tasa de aceptación en éste algoritmo. El utilizar SMC, obtiene muy buenos resultados; sin embargo, en este tipo de problema puede tardar en converger, puesto que consiste en la generación de una cadena completa para cada ancho de banda (tolerancia); algoritmos como este pueden ser utilizados para distribuciones multimodales o

aquellas en las cuales los parámetros evolucionan a través del tiempo. En cuestión de eficiencia, se prefiere el desempeño del método utilizando Caminatas Aleatorias.

---

---

# CAPÍTULO 4

---

## INFERENCIA CON DATOS REALES

### 4.1. Inferencia de los parámetros con Datos Reales

En este capítulo se implementa el método teórico de inferencia desarrollado en los capítulos previos con información real, para lo cual se utilizaron los datos provenientes de la pandemia de influenza española que se desató en 1918 y que se estima que afectó a alrededor de 500 millones de personas alrededor del mundo.

El número de muertes debidas a la pandemia fue mucho mayor que el número de muertes ocasionados por la recién terminada Primera Guerra Mundial que tuvo un impacto de 9 millones de pérdidas humanas.

Los datos a los que se hace referencia corresponden a la distribución temporal de influenza española en San Francisco, Estados Unidos en 1918. <sup>1</sup>

---

<sup>1</sup>Departamento de Higiene, 1922. Capítulo 7, Sección 2. de los Registros Epidémicos y métodos preventivos de influenza en los Estados Unidos de América En: Influenza (Ryukousei Kanbou). pag. 431-484. Tokyo, Japón: Ministerio del Interior.

CUADRO 4.1: Distribución temporal de influenza española en San Francisco, EEUU, 1918

Fecha	Tiempo	Casos	Fecha	Tiempo	Casos
23-sep	0	4	25-oct	32	2319
24-sep	1	5	26-oct	33	1916
25-sep	2	5	27-oct	34	1443
26-sep	3	7	28-oct	35	1424
27-sep	4	9	29-oct	36	1450
28-sep	5	10	30-oct	37	1481
29-sep	6	4	31-oct	38	672
30-sep	7	13	01-nov	39	557
01-oct	8	9	02-nov	40	467
02-oct	9	20	03-nov	41	385
03-oct	10	18	04-nov	42	379
04-oct	11	30	05-nov	43	457
05-oct	12	29	06-nov	44	305
06-oct	13	36	07-nov	45	337
07-oct	14	57	08-nov	46	221
08-oct	15	48	09-nov	47	234
09-oct	16	58	10-nov	48	115
10-oct	17	124	11-nov	49	132
11-oct	18	150	12-nov	50	185
12-oct	19	201	13-nov	51	240
13-oct	20	320	14-nov	52	130
14-oct	21	438	15-nov	53	104
15-oct	22	510	16-nov	54	100
16-oct	23	698	17-nov	55	36
17-oct	24	689	18-nov	56	38
18-oct	25	856	19-nov	57	26
19-oct	26	1104	20-nov	58	34
20-oct	27	1189	21-nov	59	36
21-oct	28	1190	22-nov	60	50
22-oct	29	1143	23-nov	61	43
23-oct	30	2058	24-nov	62	19
24-oct	31	1943			

En el Cuadro 4.1 se puede observar el registro de la evolución del número de personas infectadas de forma diaria, la cual, revela la gráfica de la figura 4.1, donde resultan muy notorios los saltos en el rango de tiempo 30 a 39, permitiendo la pregunta sobre el criterio que se utilizó para determinar si una persona aún seguía enferma y si dicho criterio fue homogéneo.

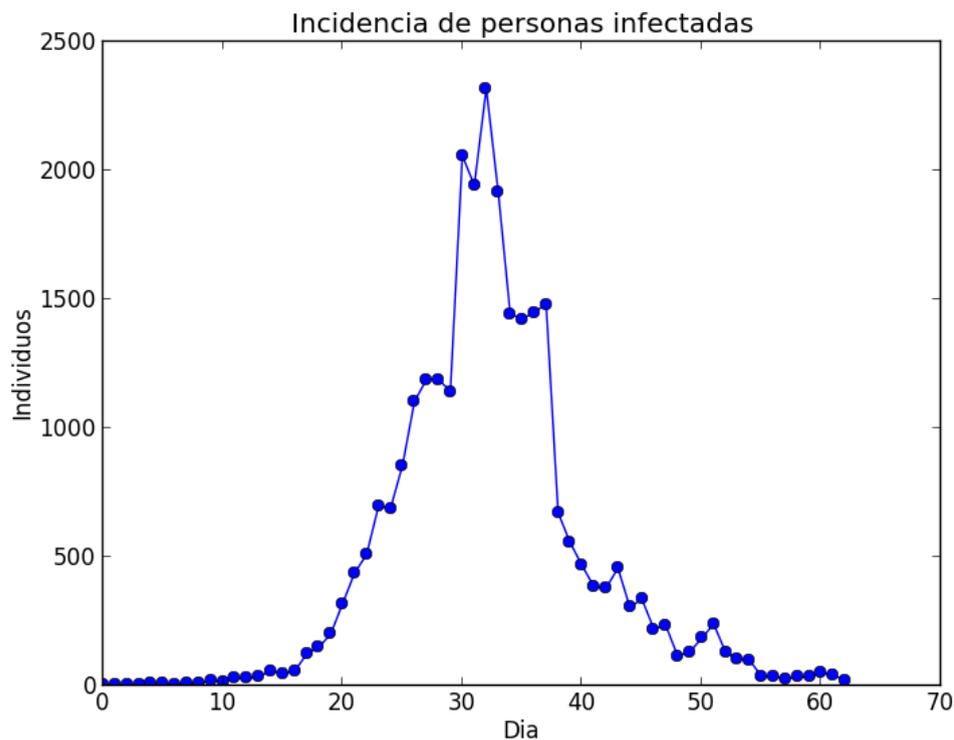


FIGURA 4.1: Casos de Influenza española en San Francisco, 1918.

En el capítulo anterior se analizó que para este tipo de problema la combinación de Estadístico, Kernel y Método Monte Carlo que resultó más eficiente en la generación de una muestra de la distribución de aproximada fue el utilizar un estadístico univariado, porque resume la información global; un Kernel Normal, el cual le proporciona robustez al algoritmo; y un método de muestreo basado en Random-Walk Metropolis Hastings, en el cual, se consideraron dos tipos de pasos para explorar de mejor forma el espacio utilizando dos tipos de Kernel de transición Normal, que se diferenciaron por el tamaño del parámetro de dispersión.

El primer problema al tomar estos datos, es enfrentarse a la población en la ciudad de San Francisco en 1918, la cual se estima en alrededor de medio millón de personas; debido al volumen de población y a que el Algoritmo Gillespie simula el

comportamiento de cada una de las partículas a la vez, el costo computacional es demandante. Otro problema es el del subregistro, al ser una enfermedad nueva, los métodos de registro al inicio de la enfermedad no son tan precisos; sin embargo, conforme evoluciona la enfermedad y al ser de alto riesgo, se destinan mayores recursos para obtener una mayor precisión de los datos.

Al realizarse las simulaciones correspondientes para obtener la distribución aproximada, se obtuvieron las curvas descritas en la gráfica de la Figura 4.2. Al ser una enfermedad nueva, es posible que toda la población se halla visto afectada, pero sólo algunos tuvieron registro de haberse enfermado, como lo muestra el total de casos a lo largo de la pandemia en el Cuadro 4.1, en el cual se registraron 28,310 casos, lo que representa aproximadamente el 5% de la población en dicha ciudad. Esto permite establecer una estimación del número de personas susceptibles en el modelo al inicio de la simulación, estableciéndose como población inicial  $S_0$  un total de 25,000 personas, se comienza con un total inicial de infectados de  $I_0 = 4$  y ningún individuo recuperado  $R_0 = 0$ .

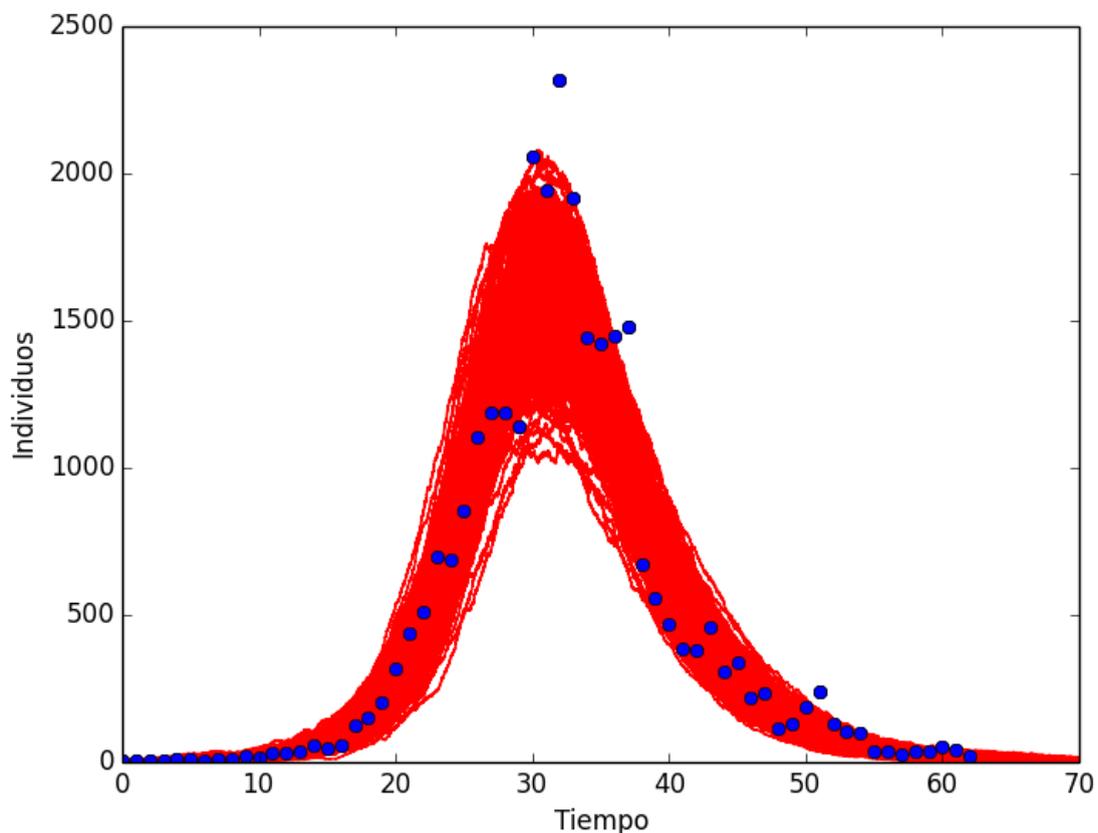


FIGURA 4.2: Distribución de curvas del parámetro

Se puede notar que las simulaciones se ajustan de manera cercana a los datos observados, aunque existen puntos que se alejan de las curvas simuladas, esto podría ser debido a diversos factores, que van desde la precisión con la que se hace el registro de los individuos, la concientización de la población a prevenir y no exponerse a la enfermedad, el subregistro, el criterio de acuerdo con el cual se diagnostica a las personas, el que el modelo que se haya escogido sea el adecuado, etcétera. Todo ello implicando cambios en la medida con la cual se desarrolla el fenómeno. Sin embargo, es notorio que aún así, el modelo se ajusta muy bien y cubre la mayoría de los puntos observados.

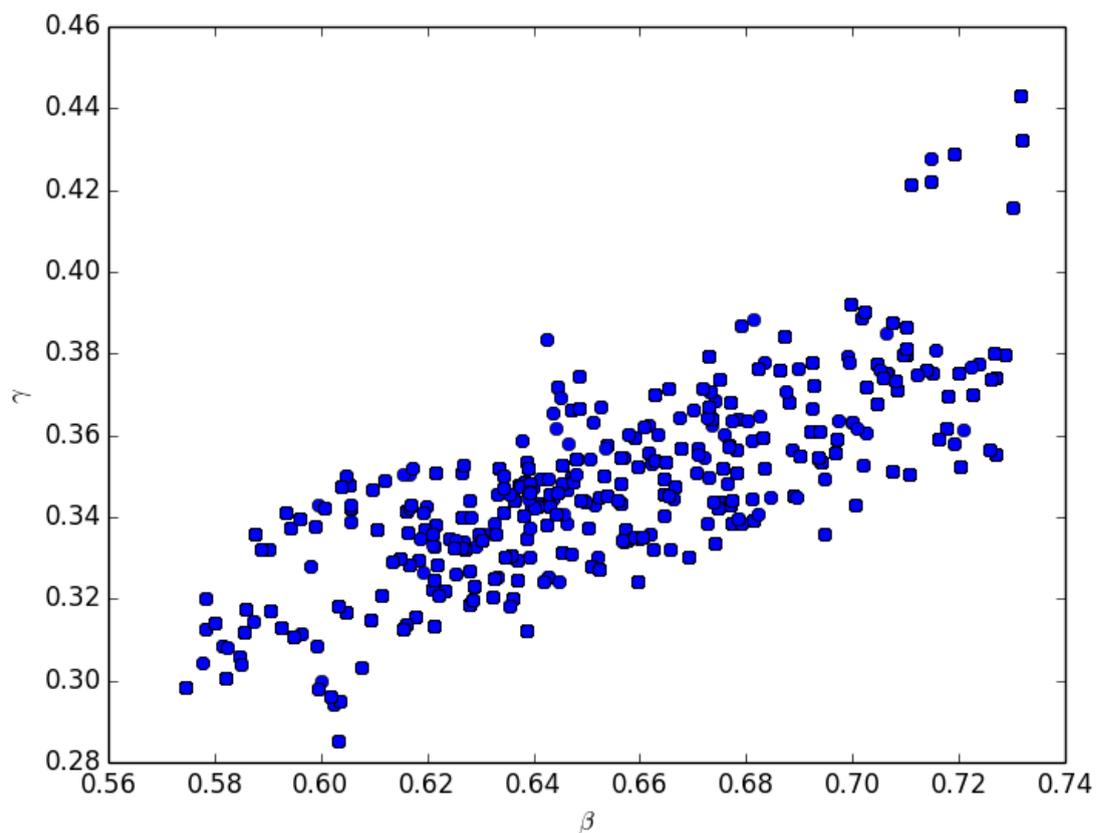


FIGURA 4.3: Distribución del parámetro

El costo computacional fue enorme pues el volumen de la población susceptible demanda recursos tanto de memoria, como de proceso para llevar a cabo cada una de las simulaciones. Se realizaron 10,000 propuestas, en el cual se permitió un ancho de banda de  $H = 40(\#I_{Obs})$ , recordando que en los datos reales, entre el 22

y el 23 de octubre existe un salto de casi mil casos, por lo que se considera una buena aproximación a la distribución aproximada.

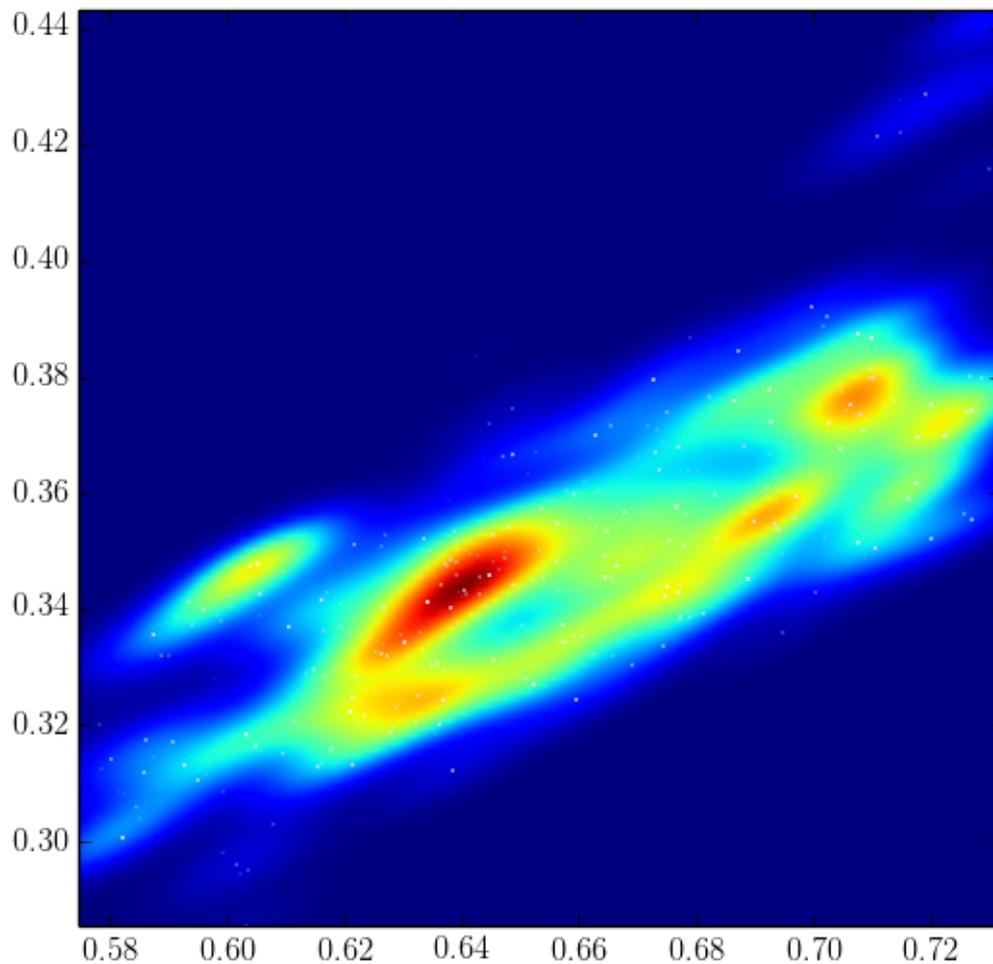


FIGURA 4.4: Distribución del parámetro

Al calcular la media en cada una de las componentes de  $\theta$ , se obtuvo que para el parámetro de contagio  $\hat{\beta} = 0.654$ , mientras que que en el caso del parámetro de recuperación se obtuvo  $\hat{\gamma} = 0.348$ , lo cual tiene asociado un número de reproducción básica de  $R'_0 = 1.879$ . Este número es muy cercano a 2 e implica, al ser mayor que 1, que la infección puede llegar a propagarse extensamente en la población y que concuerda con el rango que expone Mills [10].

De acuerdo con lo anterior, se ha reforzado la utilidad de los métodos de aproximación bayesiana aplicados a problemas en los cuales no se cuenta con una función de verosimilitud tratable computacionalmente y en los cuales, hasta la fecha no se ha ideado alguna alternativa diferente que permita realizar la inferencia.

---

---

# CAPÍTULO 5

---

## CONCLUSIONES

Los modelos son caricaturas inteligentes de la realidad; es decir, son visiones en gran medida simplificadas que intentan capturar la esencia de fenómenos reales, puesto que existen infinidad de variables involucradas que tienen un efecto sobre el fenómeno. Las menos significativas provocando pequeñas desviaciones que percibimos como ruido en las mediciones del fenómeno. Cuando dichas desviaciones presentan un tipo de comportamiento estable, se puede construir una teoría que involucre estos factores que se perciben como estocásticos.

Entre mayor sea la capacidad del ser humano por entender el medio que lo rodea, mayor será la capacidad para enfrentar los posibles riesgos a los que se pueda enfrentar. Es por tal motivo que el ser humano busca constantemente la explicación de los mecanismo y las leyes que gobiernan su entorno, ayudándose de todas las herramientas disponibles a su alcance. En ocasiones las expresiones matemáticas de las cuales nos valemos para desarrollar una teoría de los modelos son insuficientes, pero la incesante necesidad de explicación vence la impotencia, es entonces cuando se recurre a herramientas que provean algún tipo de aproximación a los resultados.

De manera similar, la inferencia estadística que se ha construido depende en gran medida de la función de Verosimilitud. Sin embargo, en ocasiones ésta resulta intratable. Es por eso que se han desarrollado métodos que puedan ofrecer resultados aproximados para obtener resultados útiles. La técnica desarrollada hasta el

momento es lo que se conoce como Computación Bayesiana Aproximada, la cual aún sigue definiéndose, puesto que no existen métodos estándares o definiciones que se puedan aplicar de manera general. Los problemas se abordan de manera muy particular, lo cual ha ampliado la diversidad de metodologías para mejorar la eficiencia de los algoritmos desarrollados para su implementación.

La aplicación de éstos modelos en salud resulta útil como ya se expuso para las ciencias básicas, pero aún más, para el desarrollo de políticas públicas en salud que permitan controlar, prevenir y atender emergencias epidemiológicas oportunamente. Otra aplicación es en el estudio de la evolución de la resistencia a medicamentos.

El modelo de compartimentos que se abordó es sencillo, pero aún así, presenta muchas dificultades para su desarrollo. El problema se complica conforme aumentan el número de observaciones y la definición de factores involucrados, la definición de los anchos de banda adecuados, o la del kernel de transición para realizar nuevas propuestas son decisivos en la eficiencia del modelo.

El modelo se puede especializar, ya que se hicieron algunos supuestos inicialmente, y que bien podrían incluirse en un modelo más robusto para obtener una mejor aproximación de la realidad. Por ejemplo, se hicieron supuestos sobre que se conocía la cantidad inicial de individuos en cada compartimento; se supuso que no había efectos de mortalidad, natalidad ni migración, se asumió que sólo habían tres tipos de compartimentos y que no había tiempos de latencia para cambiar la transición entre dichos compartimentos. Más aún, se supuso un modelo específico para la inferencia, pero se descartó hacer una comparación o un método secuencial que revelara un modelo matemático que se ajustara mejor a los datos.

Otro problema concerniente al modelo que se utiliza para tratar los datos es que en inferencia estadística usual, a pesar de que los datos provengan de un modelo en particular, quizá éstos datos se parezcan más a los de otro modelo, por lo que siempre cabe la posibilidad de cometer un error de especificación.

### **5.0.1. Datos de Epidemias**

Es conocido que los registros de personas con infección estén subestimados, por múltiples circunstancias. En primer lugar, muchas personas interpretan síntomas y los relacionan con base en su experiencia en alguna otra enfermedad de menor

riesgo o simplemente no asiste a un centro de salud donde se pueda realizar atención médica y obtener registro del caso. En algunas otras ocasiones, el diagnóstico podría no ser preciso y provocar que la enfermedad no sea detectada a tiempo. Finalmente, los sistemas de salud en muchas ocasiones carecen de los recursos suficientes, ya sean técnicos, humanos, tecnológicos o de alguna otra índole para llevar a cabo sus labores, por lo que también el registro se puede ver afectado por no existir los medios para su difusión, o bien para su captura. El problema de subestimación del volumen de información es todo un tema, para el cual también se desarrolla teoría y para el cual se pueden extender resultados.

Si bien es posible utilizar modelos determinísticos para la inferencia de los parámetros de infección y contagio utilizando métodos de optimización de mínimos cuadrados, el objetivo de este trabajo se centró en utilizar un tipo de modelo estocástico y existen razones fuertes para introducir un modelo de este tipo. Aunque es posible realizar una combinación, en la que mediante éstos modelos determinísticos se propongan parámetros que aproximen trayectorias a las observaciones obtenidas. En la vida real, los procesos ocurren de forma estocástica, lo cual le proporciona flexibilidad y se ajusta mejor a los datos. El modelo estocástico permite trabajar con conteos pequeños, ya que el determinístico se basa en conteos grandes en promedio.

La inferencia estadística para estos modelos puede resultar desafiante, puesto que las verosimilitudes asociadas a estos fenómenos resultan computacionalmente caras, e incluso intratables.

Existe más de una forma de plantearse un modelo estocástico de propagación de enfermedades. El que se abordó en el presente trabajo es utilizando un proceso de saltos puros, en el cual el principal supuesto es que los saltos corresponden a un Proceso Poisson no Homogéneo en el cual las tasas dependen de las cantidades de individuos en cada compartimento. La desventaja de este método es que conforme el número total de individuos crece, el tiempo computacional requerido para realizar cada simulación es más demandante. La ventaja es que se tiene detalle del salto de cada individuo entre los compartimentos en el tiempo. Otro modelo estocástico posible el cual puede manejar el problema de tener poblaciones grandes es utilizando procesos de difusión.

Siempre que se realiza una investigación de este tipo es necesario recibir una retroalimentación por parte de expertos de la salud, los cuales indicarán relaciones

efectos y consecuencias, en este caso los virólogos son parte indispensable para el entendimiento del problema, pues conocen más de la dinámica de los fenómenos relacionados como coinfección o superinfección. Se debe reconocer que la base de un buen modelo está directamente relacionada con el claro entendimiento del fenómeno, los factores principales que lo afectan y el reconocimiento de los elementos que intervienen en él.

En el caso de las IRA's (Infecciones Respiratorias Agudas), se tiene una especial vigilancia por parte de las autoridades, haciendo uso de reportes en hospitales para monitoreo y determinar cuál es el virus dominante.

En los resultados expuestos se comprobó que es posible realizar inferencia sobre los parámetros de infección y remoción, se verificó que el costo computacional tanto en memoria como en proceso crece demasiado conforme la población aumenta.

Se observó que utilizar métodos de estimación kernel proporcionan un enfoque más robusto a la utilización de métodos ABC estándar.

Para este tipo de problema se mostró que la mejor combinación se obtuvo haciendo uso de un estadístico univariado, pues es capaz de resumir la información y permite mayor holgura en el error de pocas observaciones sin descuidar las discrepancias globales respecto de los valores observados. También el utilizar un kernel que tuviera un soporte extenso para permitir la exploración y comparación de puntos en el espacio de parámetros, en el caso tratado aquí, el kernel Normal, fue el que presentó mayores ventajas. En cuanto al método para generar la muestra, utilizar una Caminata Aleatoria con Metropolis-Hastings, en el que se proponen diferentes tipos de pasos, conlleva una más eficiente exploración del espacio.

La inferencia misma no es un proceso que permita la determinación exacta de los parámetros a menos que se posea una cantidad infinita de información, representa el problema mismo de la modelación matemática, al llevar un problema real al lenguaje de las matemáticas de manera simplificada para que pueda ser interpretada y calculada.

Los métodos ABC permiten abordar problemas de inferencia de una extensa cantidad de modelos, los cuales no se pueden abordar con métodos tradicionales. Por lo tanto, es una alternativa cuando se han agotado los métodos para atacar el problema. Es por eso que se ha abierto el interés en temas teóricos y aplicaciones,

la eficiencia del método puede ser mejorada mediante técnicas secuenciales y regresión post-proceso en la salida de datos.

Se espera que avances futuros se enfoquen en herramientas que hagan útil el ABC en una más amplia gama de modelos paralelamente al pre y post métodos de procesamiento.

La utilización de la Computación Bayesiana Aproximada es muy empírica, es necesario establecer algunos principios para el análisis de datos y una suficiencia aproximada.

En estudios posteriores se puede abordar temas como:

- Métodos relacionados con la eficiencia de los algoritmos.
- Velocidad de convergencia y creación de estadísticos suficientes.
- Generalización del modelo SIR con distintas cepas interactuando en una población.
- Utilizar distintos tipos de modelos de compartimentos y comparación entre ellos.
- Modelos que permitan la evolución de los parámetros en el tiempo.

---

# BIBLIOGRAFÍA

- [1] Mark A. Beaumont. Adaptive approximate bayesian computation. *Infection, Genetics and Evolution*, 10:826–833, 2009.
- [2] Pierre et all Del Moral. An adaptive sequential monte carlo method for approximate bayesian computation. *Statistics and Computing*, 22:pp 1009–1020, 2012.
- [3] J. Graiton. Diggle P.J. Monte carlo methods of inference for implicit statistical models. *Journal of the Royal Statistical Society, Series B* 46:193–227, 1984.
- [4] J. L. Doob. Topics in the theory of markoff chains. *Transactions of the American Mathematical Society*, 52:37 – 64, 1942.
- [5] J. L. Doob. Markoff chains, denumerable case. *Transactions of the American Mathematical Society*, 58:455–473, 1945.
- [6] Christiane Fuchs. *Inference for Diffusion Processes, with Applications in Life Sciences*. Springer, 2013.
- [7] Danlel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, Vol. 81, No.25:2340, 1977.
- [8] Arthur Gretton Kenji Fukumizu, Le Song. Kernel bayes rule, bayesian inference with positive definite kernels. *Journal of Machine Learning Research*, 14:3753–3783, 2013.
- [9] Jean-Michel et all Marin. Approximate bayesian computational methods. *Statistics and Computing*, 22:pp 1167–1180, 2011.

- 
- [10] Lipsitch M Mills CE, Robins JM. Transmissibility of 1918 pandemic influenza. *US National Library of Medicine National Institutes of Health*, 6:904, 2004.
- [11] Zoubin; MacKay David J. C. Murray, Iain; Ghahramani. MCMC for doubly-intractable distributions. *Proceedings of the 22nd Annual Conference on Uncertainty in Artificial Intelligence (UAI-06)*, 22:359–366, 2006.
- [12] Kermack W. O. and A. G. McKendrick. Contributions to the mathematical theory of epidemics. *Proceedings of the Royal Society*, 115A:700–721, 1927.
- [13] Vincent Plagnol Paul Marjoram, John Molitor and Simon Tavaré. Markov chain monte carlo without likelihoods. *PNAS*, 100:15324–15328, 2003.
- [14] Ajay Jasra Pierre Del Moral, A. D. Arnaud Doucet. Sequential monte carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68:411–436, 2006.
- [15] T. Kirkwood R. Boys, D. Wilkinson. Bayesian inference for a discretely observed stochastic kinetic model, statistics and computing. *Statistics and Computing*, 18:125, 2008.
- [16] Kenji Fukumizu Shigeki Nakagome and Shuhei Mano. Kernel approximate bayesian computation in population genetic inferences. *Statistical Applications in Genetics and Molecular Biology*, 12(6):667–678, 2012.
- [17] Tavaré S, Balding DJ, Griffiths. Inferring Coalescence Times From DNA Sequence Data. *Genetics Society of America*, 145:505–518, 1997.
- [18] Natalja Strelkova Andreas Ipsen Tina Toni, David Welch and Michael P.H Stumpf. Approximate bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society*, 6:187–202, 2009.