

UNIVERSIDAD AUTÓNOMA METROPOLITANA-IZTAPALAPA  
DIVISIÓN DE CIENCIAS BÁSICAS E INGENIERÍA

# ESTIMACIÓN BAYESIANA EN MODELOS FARMACOCINÉTICOS

Tesis que presenta  
**Alejandro Nieto Ramos**  
Para obtener el grado de  
**Maestro en Ciencias Matemáticas**  
**Aplicadas e Industriales**

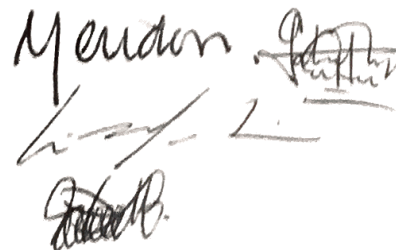
---

Asesores:

Dr. Gabriel Núñez Antonio  
Dr. José Héctor Morales Bárcenas

Jurado calificador:

Presidente: Dr. Manuel Mendoza Ramírez  
Secretario: Dr. Gabriel Núñez Antonio  
Vocal: Dr. Luis Alberto Medina Velázquez  
Vocal: Dr. José Héctor Morales Bárcenas



Manuel Mendoza Ramírez  
Gabriel Núñez Antonio  
Luis Alberto Medina Velázquez  
José Héctor Morales Bárcenas

Ciudad de México, a 21 de febrero de 2017

UNIVERSIDAD AUTÓNOMA METROPOLITANA-IZTAPALAPA  
DIVISIÓN DE CIENCIAS BÁSICAS E INGENIERÍA

# **ESTIMACIÓN BAYESIANA EN MODELOS FARMACOCINÉTICOS**

Tesis que presenta  
**Alejandro Nieto Ramos**  
Para obtener el grado de  
**Maestro en Ciencias Matemáticas  
Aplicadas e Industriales**

---

Asesores:  
Dr. Gabriel Núñez Antonio  
Dr. José Héctor Morales Bárcenas

Jurado calificador:  
Presidente: Dr. Manuel Mendoza Ramírez  
Secretario: Dr. Gabriel Núñez Antonio  
Vocal: Dr. Luis Alberto Medina Velázquez  
Vocal: Dr. José Héctor Morales Bárcenas

Ciudad de México, a 21 de febrero de 2017

*A Juan Carlos, por todo lo que hemos vivido juntos*



It all began the day I found  
That from my window  
I could only see  
A piece of sky.  
I stepped outside and looked around  
I never dreamed I was so wide  
Or even half as high  
  
The time had come  
To try my wings  
And even though it seemed at any  
Moment I could fall,  
I felt the most  
Amazing things,  
The things you can't imagine if  
You've never flown at all.  
  
Though it's safer to stay on  
The ground,  
Sometimes where danger lies  
There the sweetest of pleasures  
Are found.  
No matter where I go—there'll be  
Mem'ries that tug at my sleeve  
But there will also be  
more to question yet more to believe  
  
The more I live—the more I learn  
The more I learn—the more I realize  
The less I know.  
Each step I take—  
Each I page I turn—  
Each mile I travel only means  
The more I have to go  
What's wrong with wanting more?  
If you can fly—then soar!  
With all there is—why settle for  
Just a piece of sky?

*Alan & Marilyn Bergman*



# Índice general

<b>Agradecimientos</b>	<b>v</b>
<b>Prólogo</b>	<b>vii</b>
<b>Introducción</b>	<b>ix</b>
<b>1. Farmacocinética</b>	<b>1</b>
1.1. Preliminares . . . . .	1
1.2. Caracterización del proceso ADME . . . . .	3
1.2.1. Administración intravenosa . . . . .	3
1.2.2. Administración oral . . . . .	5
1.3. Incorporación de la dosis . . . . .	6
<b>2. Estadística computacional</b>	<b>9</b>
2.1. Muestreo de variables aleatorias . . . . .	9
2.1.1. Funciones de distribución estándar . . . . .	9
2.1.2. Funciones de distribución no estándar . . . . .	10
2.1.3. Muestreo por rechazo . . . . .	10
2.2. Monte Carlo vía cadenas de Markov . . . . .	12
2.2.1. Integración Monte Carlo . . . . .	13
2.2.2. Cadenas de Markov . . . . .	14
2.2.3. Métodos de Monte Carlo vía cadenas de Markov . . . . .	15
2.2.4. El algoritmo de Metropolis . . . . .	15
2.2.5. El algoritmo de Metropolis-Hastings . . . . .	19
2.2.6. El algoritmo de Metropolis-Hastings para distribuciones multivariadas . . . . .	19
2.2.7. El muestreo de Gibbs . . . . .	21
2.3. Análisis de datos bayesiano . . . . .	22

2.3.1.	Estrategias de la estimación de parámetros . . . . .	23
2.3.2.	Máxima Verosimilitud . . . . .	23
2.3.3.	El máximo a posteriori . . . . .	24
2.3.4.	El muestreo a posteriori . . . . .	25
<b>3.</b>	<b>Aplicaciones</b>	<b>27</b>
3.1.	Primer ejemplo . . . . .	27
3.1.1.	Metodología . . . . .	28
3.1.2.	Implementación . . . . .	33
3.1.3.	Resultados . . . . .	34
3.1.4.	Simulación . . . . .	35
3.2.	Segundo ejemplo . . . . .	36
3.2.1.	Metodología . . . . .	38
3.2.2.	Implementación . . . . .	42
3.2.3.	Resultados . . . . .	42
3.3.	Tercer ejemplo . . . . .	43
3.3.1.	Metodología . . . . .	44
3.3.2.	Implementación . . . . .	46
3.3.3.	Resultados . . . . .	46
<b>A.</b>	<b>Modelos estocásticos de ecuaciones diferenciales ordinarias</b>	<b>51</b>
A.1.	El modelo SIR determinista . . . . .	52
A.1.1.	Modelo SIR con demografía . . . . .	53
A.2.	Diferentes modelos SIR estocásticos . . . . .	54
A.2.1.	Ruido en las observaciones . . . . .	56
A.2.2.	Ruido del proceso . . . . .	56
A.2.3.	Parámetros aleatorios . . . . .	57
A.2.4.	Comparación entre los distintos tipos de ruido . . . . .	58
A.2.5.	Ventajas y desventajas entre los diferentes tipos de modelos . . . . .	59
<b>B.</b>	<b>Aproximación normal bayesiana</b>	<b>61</b>
	<b>Índice alfabético</b>	<b>67</b>



# Agradecimientos

*Cuánto trabajo consumí allí, qué dificultades soporté, cuántas veces desesperé y cuántas desistí y comencé nuevamente el empeño de aprender; testigo es la conciencia, tanto la mía, pues yo he padecido, tanto la de aquéllos que conmigo vivieron.*

*Sor Juana Inés de la Cruz.*

Quiero agradecer a uno de dos mis asesores, al doctor J. Héctor Morales Bárcenas, por haberme sugerido hacer un trabajo interdisciplinario con quien más tarde se convirtió en mi asesor también: el doctor Gabriel Núñez Antonio. A ambos les agradezco su tiempo y todas sus enseñanzas. Al doctor Manuel Mendoza Ramírez le agradezco haber aceptado amablemente revisar mi tesis, darme sugerencias para mejorarla y haberse sentado conmigo a explicarme algunas cosas que no me eran claras. Al doctor Luis Alberto Medina Velázquez, por habernos proporcionado los datos para llevar a cabo este trabajo tan interesante.

A todas los profesores que hicieron que mi experiencia en la UAM fuera muy grata, en particular al doctor L. Héctor Juárez Valencia, a la doctora Patricia Saavedra Barrera y al doctor Mario G. Medina Valdez, porque siempre tuvieron tiempo de escucharme y guiarme con sus consejos. A mi amigo Charles, que desde el principio hizo que adaptarme a la UAM fuera más fácil; a Érika por habernos presentado y haber hecho lo propio, a mi Tía por su chispa y a mis queridos amigos y compañeros Pablo y Manuel.

A Marc Lavielle, por su generosidad y por todo lo que me enseñó en tan poco tiempo.

Finalmente quiero agradecerle a quien ha sido mi mentora desde la licenciatura, la doctora Beatriz Rumbos Pellicer, y a mi querida maestra Alma J. Sánchez Porras, que nunca me ha dejado solo; a Françoise y a mi familia: mi mamá, July, y Juan Carlos, por su cariño.



# Prólogo

El proceso de desarrollo de nuevos fármacos es algo que además de ser costoso, lleva mucho tiempo[10]. En un estudio de 2003[7], se indica que lanzar un medicamento al mercado cuesta aproximadamente USD \$800 000. Se estima que cerca del noventa por ciento de los compuestos que son candidatos, fallen durante este proceso. Por lo tanto, la industria farmacéutica está siempre en la búsqueda de nuevas herramientas que ayuden a que este proceso sea más eficiente. Según la Administración de Alimentos y Medicamentos de los Estados Unidos[4], la simulación y el modelado matemático y computacional son herramientas útiles para mejorar la eficiencia en el desarrollo seguro y efectivo de los fármacos.

En esta obra, se desarrollan modelos de ecuaciones diferenciales ordinarias -llamados farmacocinéticos- los cuales están basados en principios biológicos y farmacológicos. La metodología que se emplea consiste en describir la evolución de la concentración de un medicamento en cierto organismo a través de un modelo farmacocinético, para después estimar las distribuciones de los parámetros de dicho modelo por medio de inferencia bayesiana; específicamente lo último se hace empleando métodos Monte Carlo vía cadenas de Markov.

Particularmente, la técnica desarrollada se ejemplifica con la cinética de la teofilina en el cuerpo humano y con la concentración en el tiempo de cisplatino y tecnecio encapsulados en liposomas<sup>1</sup>, administrados, respectivamente, en ratones atímicos<sup>2</sup> y ratas Wistar.

En cuanto a la escritura de este trabajo, cabe señalar que la mayor parte de las veces que se compiló información de una obra -como es el caso de los dos primeros capítulos- las ideas y las aseveraciones encontradas aquí

---

<sup>1</sup>Un liposoma es una pequeña burbuja elaborada del mismo material que constituye una membrana celular.

<sup>2</sup>Que carecen de timo.

corresponden a las de los autores originales: [10] y [11] del capítulo 1, [1], [22] y [23] del capítulo 2 y [9] del apéndice A. Mi aportación en este sentido es la de haber reunido toda la información encontrada de manera que la obra tuviera coherencia, y de producir una traducción que en español se leyera natural; todas las veces que hice una traducción textual, se debió a que consideré que no pude haber presentado mejor los conceptos, en particular porque estoy muy lejos de ser un experto en la materia. Finalmente, aparte de lo descrito anteriormente, la principal colaboración de mi parte se encuentra en el capítulo 3.

# Introducción

En general, se puede decir que el proceso completo del desarrollo de un medicamento está formado por una parte **preclínica** y otra **clínica**. En la parte preclínica, se prueba el efecto en animales de diferentes tipos de compuestos.

Ahora bien, la parte clínica se divide en tres fases. En la primera, la medicina se administra en humanos saludables para averiguar la compatibilidad fisiológica; en la segunda, se investiga el efecto farmacológico y terapéutico de la misma; en la última, el fármaco se prueba en miles de pacientes.

Un experimento en la fase preclínica consta de dos partes. La primera estudia la concentración del medicamento en el organismo en el transcurso del tiempo. En esta parte, el interés se centra en la distribución del fármaco en el cuerpo, mas no se considera a la enfermedad misma ni al efecto que la medicina pudiera tener en la enfermedad. En pocas palabras, se observa *lo que el cuerpo le hace a la medicina*. A esta etapa se le conoce como **farmacocinética**, denotada **PK**.

En la segunda parte, se observa el desarrollo de la enfermedad y de los efectos farmacológicos del medicamento en la enfermedad; a esta última se le conoce como **respuesta farmacológica**. En esta etapa se observa *lo que el medicamento le hace al cuerpo*. Esta parte se conoce como **farmacodinámica**, denotada **PD**. En la figura 1 se muestra un esquema del procedimiento completo.

Al combinar la farmacocinética y la farmacodinámica, conocidas en conjunto como **farmacometría**, se obtiene un panorama general de la respuesta del fármaco. En ambas etapas (PKPD), se asume que la concentración del fármaco es la causa principal que provoca el efecto farmacológico en la enfermedad. En un experimento PKPD se hacen mediciones farmacocinéticas y farmacodinámicas en una población de individuos. A menudo, los datos farmacocinéticos son escasos, porque las muestras de sangre se tienen que

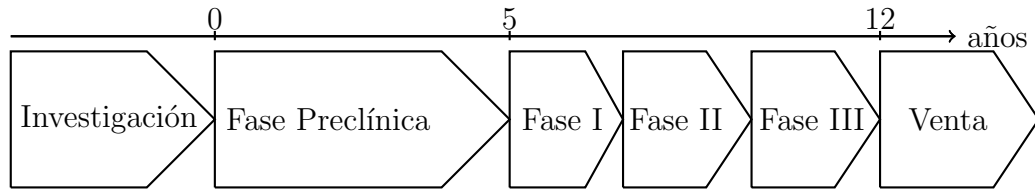


Figura 1: Resumen esquemático del desarrollo de un medicamento.

tomar de los individuos cada determinado tiempo. En cuanto a la farmacodinámica, el desarrollo de la enfermedad se describe por medio de la toma de lecturas apropiadas.

Para tener una visión general del efecto del medicamento, se deben administrar diferentes dosis durante un experimento. A los datos PD que describen la enfermedad con un fármaco administrado se les llama **datos perturbados**. También se administra un **placebo** para describir el desarrollo de la enfermedad sin el efecto del medicamento. A los datos que vienen de un tiempo en específico en el cual se aplicó cierta dosis se les conoce como **grupo de dosificación**.

Este trabajo se enfoca en los llamados **modelos matemáticos semi-mecánicos**. Este tipo de modelos describen la situación biológica subyacente por medio de principios básicos y como resultado de los datos que se tiene disponibles; no se considerarán modelos que solo caractericen a los datos sin hacer suposiciones biológicas.

# Capítulo 1

## Farmacocinética

### 1.1. Preliminares

Comúnmente, un medicamento se puede administrar a un paciente de manera intravenosa, oral o subcutánea; de maneras menos habituales, por medio de un parche en la piel o por inhalación. Habiendo administrado el medicamento, existen procesos subsecuentes dentro del organismo, a los cuales usualmente se les describe por medio del proceso farmacocinético conocido como **ADME**: absorción, distribución, metabolismo y excreción.

- La **absorción** es el paso del medicamento al torrente sanguíneo desde un punto situado fuera de los vasos sanguíneos. Ahora bien, al grado de absorción de un medicamento en el sistema circulatorio se le conoce como **biodisponibilidad**.
- Después de que ocurre la absorción, la mayoría de los medicamentos se distribuyen en los tejidos del cuerpo a través de la sangre. La **distribución** describe la transferencia en ambos sentidos, de un medicamento entre la sangre, el tejido y los órganos.
- El **metabolismo** se refiere a la biotransformación del medicamento dentro del cuerpo en otras moléculas conocidas como **metabolitos**; desde el punto de vista farmacológico, muchas veces se encuentran inactivos, pero pueden estar activos o inclusive ser tóxicos. Algunos medicamentos, llamados **profármacos**, están inactivos hasta que el organismo los metaboliza.

- La **excreción** es la eliminación del medicamento del cuerpo. Los riñones, los cuales excretan sustancias solubles con la orina, son los órganos principales de este proceso. El flujo biliar del hígado también es una ruta principal para la eliminación en las heces. Un medicamento puede también ser liberado a través de otras rutas naturales, como la respiración, las lágrimas, el sudor o la saliva, entre otros. A los procesos de metabolismo y excreción a menudo se les llama conjuntamente **procesos de eliminación**.

La farmacocinética de un medicamento es un proceso biológico muy complejo, que además es difícil de describir cuantitativamente. Por lo tanto, los estudiosos de la farmacología simplifican muchos procesos fisiológicos al crear un **modelo por compartimentos**. Es decir, el cuerpo humano se describe como una serie de compartimentos en los cuales se distribuye el medicamento.

El **compartimento de depósito** es el lugar en el cual se deposita el medicamento: el estómago, en el caso de administración oral, el torrente sanguíneo en el caso de administración intravenosa, la piel, en el caso del parche y así sucesivamente. El **compartimento central** está formado por la sangre y por los órganos altamente irrigados por ella, como es el caso del hígado, los riñones o los pulmones. Los **compartimentos periféricos** son aquéllos cuyo tejido está menos irrigado por la sangre. Lo cierto es que tomar lecturas del cuerpo que no sean de sangre, es imposible en la mayoría de los casos. Además, es un hecho que son suficientes dos compartimentos para describir apropiadamente el curso del tiempo de los fármacos en el organismo [10].

Este esquema simplificado se puede transformar en un modelo matemático, el cual intenta dar una buena aproximación descriptiva de los procesos farmacocinéticos, y con el cual se pueden calcular concentraciones de la droga en cualquier momento y en cualquier parte del cuerpo. *Cuando se usan modelos farmacocinéticos, se asume que la concentración del medicamento es perfectamente homogénea en cada compartimento del cuerpo y en cualquier momento.* Esta fuerte suposición tiene la ventaja de permitir una descripción cualitativa del modelo PK. De hecho, describir cuantitativamente un modelo PK es equivalente a describir la variación de la concentración del medicamento en los compartimentos. Esto significa que típicamente es necesario describir tres componentes:

- La **tasa de entrada**, que describe cómo pasa el medicamento del compartimento de depósito al compartimento central.



- La **tasa de distribución**, que describe los intercambios entre el compartimento central y los compartimentos periféricos.
- La **tasa de eliminación**, que describe cómo se elimina el medicamento del compartimento central.

El modelo final une estos tres componentes en un solo modelo, como se puede ver en la figura 1.1.

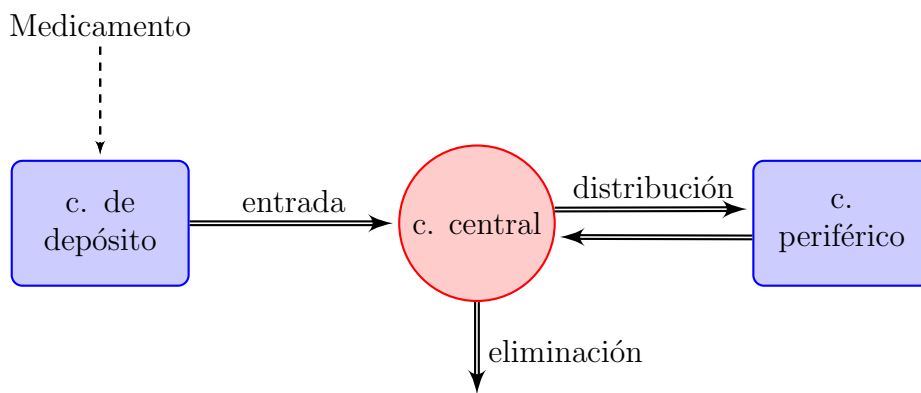


Figura 1.1: Esquema general de un modelo por compartimentos.

## 1.2. Caracterización del proceso ADME

Un modelo PK es un sistema dinámico matemáticamente representado por un sistema de ecuaciones diferenciales ordinarias (EDO) que describen las transferencias entre los compartimentos y la eliminación del compartimento central. A continuación se verán diferentes ejemplos de sistemas dinámicos para diversos tipos de administración.

### 1.2.1. Administración intravenosa

Considérese un sistema donde la administración es intravenosa y donde solo existe un compartimento. El único proceso que se necesita describir en este caso es la eliminación. Lo cierto es que existen diferentes procesos para describir el **proceso de eliminación**. Por ejemplo, la **eliminación lineal**,

también llamada de **eliminación de primer orden**, significa que la tasa de eliminación es directamente proporcional a la cantidad de concentración; es decir

$$A'_c(t) = -k_e A_c(t), \quad (1.1)$$

donde  $A_c(t)$  y  $A'_c(t)$  son, respectivamente, la cantidad de medicamento en el compartimento central y su derivada al tiempo  $t$ . La constante de proporcionalidad  $k_e$  es la **tasa de eliminación constante**. Esta tasa está relacionada con la constante conocida como **semivida de eliminación** o sencillamente **semivida**, que se define como el tiempo requerido para que la mitad del medicamento se haya eliminado del compartimento:

$$t_{\frac{1}{2}} = \frac{\log(2)}{k_e}. \quad (1.2)$$

Si ahora se denota a  $V$  como el **volumen de distribución** del compartimento central, se define entonces a la **remoción** como  $Cl = k_e V$ . La remoción relaciona a la tasa de eliminación con la concentración, es decir,  $C(t) = \frac{A_c(t)}{V}$ . De este modo,

$$A'_c(t) = -(Cl) C(t).$$

La remoción se expresa en volumen por unidad de tiempo.

Para medicamentos que se metabolizan, la tasa de eliminación no se incrementa proporcionalmente a la concentración del medicamento. En este caso se habla de **eliminación saturada**, lo que significa que más allá de una determinada concentración, la tasa de eliminación tiende a alcanzar un valor máximo. Esta capacidad limitada de la eliminación se puede explicar por medio de las ecuaciones de **Michaelis-Menten**:

$$\begin{aligned} A'_c(t) &= -\frac{V_m}{V K_m + A_c(t)} A_c(t) \\ &= \frac{V_m}{K_m + C(t)} C(t). \end{aligned} \quad (1.3)$$

$V_m$  es la capacidad de eliminación máxima, es decir, la cantidad total de medicamento que se puede eliminar por unidad de tiempo durante la saturación, y  $K_m$  es la concentración del medicamento eliminado cuando la capacidad máxima  $V_m$  se encuentra exactamente a la mitad. Cuando la concentración del medicamento es muy pequeña con respecto a  $K_m$ , este proceso mixto se vuelve similar al de primer orden y la eliminación es directamente

proporcional a la concentración. Cuando la concentración del medicamento es grande en relación a  $K_m$ , la tasa de eliminación alcanza el valor constante  $V_m$ .

Ahora se añade un compartimento periférico al modelo y se considera que existe una transferencia lineal entre el compartimento central y el periférico. Si además se asume que haya eliminación lineal del compartimento central, la representación matemática de este modelo es ahora un sistema de dos ecuaciones diferenciales ordinarias lineales:

$$\begin{aligned} A'_c(t) &= k_{21}A_p(t) - k_{12}A_c(t) - k_eA_c(t) \\ A'_p(t) &= -k_{21}A_p(t) + k_{12}A_c(t), \end{aligned} \quad (1.4)$$

donde  $A_p$  es la cantidad de medicamento en el compartimento periférico y  $k_{12}$  y  $k_{21}$  son las tasas de distribución constantes. La eliminación lineal se puede reemplazar por otro tipo de proceso de eliminación, como el Michaelis-Menten. También es directo ampliar este modelo a uno con más de un compartimento periférico. Por ejemplo, se presenta a continuación un **modelo con dos compartimentos periféricos**  $p$  y  $q$ :

$$\begin{aligned} A'_c(t) &= -(k_{12} + k_{13} + k_e)A_c(t) + k_{21}A_p(t) + k_{31}A_q(t) \\ A'_p(t) &= k_{12}A_c(t) - k_{21}A_p(t) \\ A'_q(t) &= k_{13}A_c(t) - k_{31}A_q(t). \end{aligned}$$

### 1.2.2. Administración oral

Una vez que el medicamento se hubo tragado, éste llega al tracto gastrointestinal y se absorbe en el torrente sanguíneo. Lo cierto es que solo una fracción  $F$  de una dosis administrada de manera oral, puede alcanzar el sistema circulatorio debido a diversos factores tales como cierta descomposición en el intestino, una absorción pobre o una extracción relacionada con el **metabolismo de primer paso**<sup>1</sup>. A  $F$  se le conoce como **biodisponibilidad**.

En un proceso de **absorción de orden cero** se asume que el medicamento es transferido del compartimento de depósito con tasa constante  $R_0$ , es decir,

$$A'_d(t) = -R_0 \mathbb{I}_{A_d(t) > 0}, \quad (1.5)$$

---

<sup>1</sup>Fenómeno relacionado con el metabolismo de los medicamentos debido al cual la concentración de un medicamento se reduce considerablemente antes de llegar al sistema circulatorio.

donde  $\mathbb{I}_{A_d(t)>0}$  es una función indicadora (es 0 cuando  $A_d(t) \leq 0$  y 1 en cualquier otro caso).

En un proceso de **absorción de primer orden** se asume que la tasa de absorción es proporcional a la cantidad de medicamento en el compartimento de depósito, es decir,

$$A'_d(t) = -k_a A_d(t). \quad (1.6)$$

Del mismo modo que en la ecuación (1.2), la tasa de absorción constante  $k_a$  está directamente relacionada con la correspondiente semivida de absorción:

$$t_{abs, \frac{1}{2}} = \frac{\log(2)}{k_a}. \quad (1.7)$$

De manera más general, la absorción de orden cero se puede definir como el límite del proceso de **absorción de orden  $\alpha$**

$$A'_d(t) = -R_\alpha A_d^\alpha(t), \quad (1.8)$$

cuando  $\alpha \rightarrow 0$ . Los modelos PK para administración oral están formados a su vez de tres modelos: uno de absorción, otro de distribución y un último de eliminación. Por ejemplo, un **modelo de un compartimento con un proceso de absorción de primer orden y eliminación lineal** es el siguiente:

$$\begin{aligned} A'_d(t) &= -k_a A_d(t) \\ A'_c(t) &= k_a A_d(t) - k_e A_c(t). \end{aligned} \quad (1.9)$$

### 1.3. Incorporación de la dosis

Los modelos PK no describen cómo se administra un medicamento. Las dosis administradas son términos fuente, es decir, entradas que dinámicamente modifican el estado del sistema. Se hace notar que el mismo modelo PK se puede usar para diferentes regímenes de dosis.

La concentración plasmática pronosticada por un modelo es la solución de un sistema de ecuaciones diferenciales ordinarias, dada una serie de entradas y condiciones iniciales. En este caso, *se supone que todos los compartimentos están vacíos antes de la primera dosis*. Cuando el sistema de ecuaciones diferenciales ordinarias es lineal, se puede hallar una solución analítica. De otro modo, se requiere de métodos numéricos para aproximar la solución del sistema.

En el caso de **administración intravenosa por bolo**, el medicamento se administra en un periodo de tiempo despreciable y se distribuye instantáneamente en todo el compartimento central. Si  $D$  es la cantidad de medicamento administrado en el tiempo  $\tau$ , entonces

$$A_c(\tau^+) = A_c(\tau^-) + D,$$

donde  $\tau^-$  y  $\tau^+$  son, respectivamente, instantes justo antes y después de la administración. Si el compartimento contiene una cantidad nula del fármaco antes de  $\tau$ , la solución es

$$C(t) = \frac{D}{V} e^{-k_e(t-\tau)} \mathbb{I}_{t>\tau}. \quad (1.10)$$

Si un número  $N$  de dosis se administran a los tiempos  $\tau_1, \tau_2, \dots, \tau_N$ , se aplica el principio de superposición de las ecuaciones diferenciales debido a que el sistema es lineal. En este caso se tiene que la concentración es

$$C(t) = \frac{1}{V} \sum_{i=1}^N D_i e^{-k_e(t-\tau_i)} \mathbb{I}_{t>\tau_i}. \quad (1.11)$$

En el caso de **administración por infusión**, se aplica una dosis a una tasa constante  $R_{inf}$  durante un cierto periodo de tiempo de infusión de tamaño  $T_{inf} = \frac{D}{R_{inf}}$ . Suponiendo que una única dosis se administra al tiempo  $\tau$ , la solución es

$$C(t) = \begin{cases} 0 & \text{si } t < \tau, \\ \frac{R}{V k_e} (1 - e^{-k_e(t-\tau)}), & \text{si } \tau \leq t < \tau + T_{inf} \\ \frac{R}{V k_e} (1 - e^{-k_e T_{inf}}) e^{k_e(t-\tau-T_{inf})} & \text{si } t \geq \tau + T_{inf} \end{cases} \quad (1.12)$$

Se hace notar que en esta sección, se usó el mismo modelo para los diferentes ejemplos. Cambiar las diferentes entradas, es decir, las dosis, es lo que hace que se tengan diferentes soluciones.

Considérese ahora administración oral y sea  $A_d$  la cantidad de medicamento en el compartimento de depósito. Si una cantidad  $D$  se deposita instantáneamente al tiempo  $\tau$ , entonces, el estado variable  $A_d$  se modifica al tiempo  $\tau$  como sigue

$$A_c(\tau^+) = A_c(\tau^-) + F D,$$

donde  $\tau^-$ ,  $\tau^+$  y  $F$  se definen como antes. Si el compartimento central y el de depósito contienen una cantidad nula del fármaco antes de  $\tau$ , la solución es

$$C(t) = \frac{F D k_a}{V(k_a - k_e)} (e^{-k_e(t-\tau)} - e^{-k_a(t-\tau)}) \mathbb{I}_{t>\tau}. \quad (1.13)$$

Se pueden obtener otros tipos de soluciones similares cuando la absorción es de orden cero o cuando se tienen dos o tres compartimentos; las soluciones son combinaciones lineales de exponenciales que decaen.

Antes de terminar este capítulo es importante hacer notar que los parámetros de los diferentes tipos de modelos farmacocinéticos que se vieron se consideran fijos. Pero, ¿qué pasa cuando se tienen datos a los cuales se les desea ajustar un modelo farmacocinético y los valores de los parámetros del modelo elegido se desconocen? Pues bien, en el capítulo siguiente se desarrollará la metodología estadística que se emplea para encontrar los valores de los parámetros que mejor se ajustan a un conjunto de datos, e inclusive se encontrarán distribuciones de probabilidad de los primeros cuando a la incertidumbre se le describe mediante un modelo de probabilidad.

# Capítulo 2

## Estadística computacional

### 2.1. Muestreo de variables aleatorias

Muchas veces, los modelos probabilísticos que aparecen en cualquier problema, son muy complejos como para abordarse analíticamente. Los investigadores, cada vez más, se valen de métodos numéricos computacionales cuando trabajan con modelos probabilísticos intrincados. Con el uso de la computadora, los científicos no necesitan hacer suposiciones poco realistas, como que exista normalidad o independencia, cuando se usan técnicas analíticas.

La clave para la mayoría de las técnicas de aproximación es la habilidad de obtener muestras de las funciones de distribución. Se pueden obtener muestras para predecir el comportamiento que tendrá un modelo en ciertas circunstancias, o para pronosticar los valores de las variables latentes -los parámetros- cuando se aplican modelos a los datos experimentales. Muchas estrategias de muestreo computacional hacen una simplificación: convierten el problema de muestrear de una distribución compleja en subproblemas que a su vez hacen uso de distribuciones de muestreo más simples.

#### 2.1.1. Funciones de distribución estándar

Algunas distribuciones se usan tanto, que ya forman parte de los programas o lenguajes de programación del cómputo científico. Tal es el caso de R, MATLAB o Mathematica, por nombrar algunos. Con este tipo de software es prácticamente inmediato evaluar la densidad de probabilidad o la distribución acumulada de estas distribuciones en un punto, en conjunto de puntos,

u obtener muestras de dichas distribuciones. Inclusive, hoy en día también es fácil encontrar cómo trabajar con otras distribuciones comunes buscando en Internet.

### 2.1.2. Funciones de distribución no estándar

Supóngase ahora que se desea obtener una muestra de una distribución que no está predefinida en el paquete de cómputo con el que se esté trabajando. Cuando se modela, esto es algo que sucede a menudo, porque un investigador puede proponer, por ejemplo, combinar distribuciones preexistentes para resolver un problema determinado. Es común que los métodos computacionales para resolver problemas de muestreo complejos, se basen en distribuciones de muestreo de las cuales ya se sabe cómo obtener muestras de manera eficiente. Los valores aleatorios de estas distribuciones simples se pueden entonces transformar, ya sea aplicándoles la distribución objetivo o comparándolos con ella. De hecho, algunas de las estrategias que se verán en esta sección se usan internamente en algunos paquetes de cómputo para el muestreo de distribuciones como la normal o la exponencial.

### 2.1.3. Muestreo por rechazo

Este tipo de técnica, en general es apropiada para el caso univariado. La ventaja principal del método de **muestreo por rechazo** es que no necesita de un *periodo de calentamiento*<sup>1</sup> para obtener la muestra final. En vez de eso, todos los valores obtenidos durante el muestreo, precisamente los que no son rechazados, se pueden usar inmediatamente como una muestra de la distribución objetivo.

Para ilustrar la idea general del muestro por rechazo, también llamado **algoritmo de aceptación-rechazo**, obsérvese la figura 2.1.<sup>2</sup> Supóngase que se desea obtener una muestra de puntos que se distribuyan uniformemente en un círculo de radio 1 centrado en  $(0, 0)$ . De entrada, parecería difícil obtener directamente la muestra de puntos que se distribuyan de esta forma dentro de la circunferencia. Sin embargo, se puede aplicar el muestro por rechazo, primero obteniendo una muestra uniforme  $(x, y)$  del cuadrado que encierra al círculo, y rechazando después todos los valores para los cuales  $x^2 + y^2 > 1$ .

---

<sup>1</sup>Término informal que describe la práctica de deshacerse de un cierto número de las primeras iteraciones resultantes de ejecutar un algoritmo MCMC (sección 2.2.3).

<sup>2</sup>Las ilustraciones presentadas en esta sección fueron tomadas de [23].



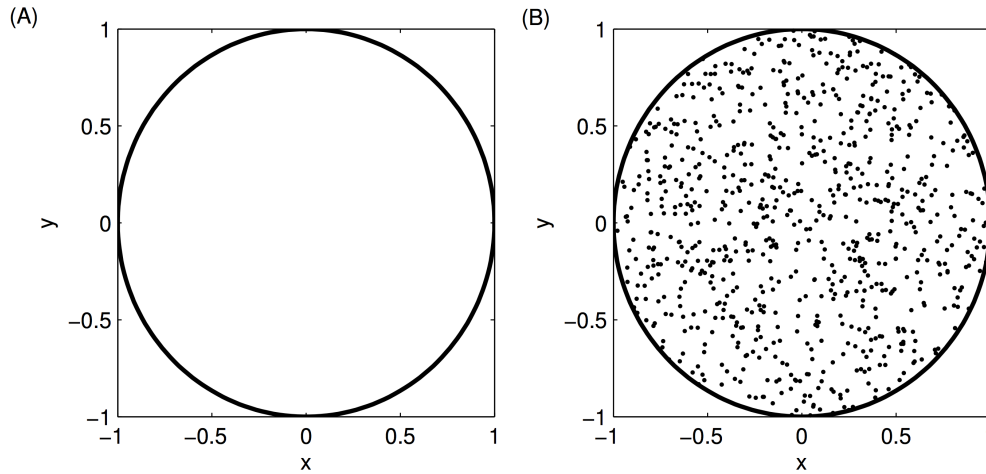


Figura 2.1: Muestra obtenida uniformemente del círculo de radio 1 con el método de muestreo por rechazo.

En este ejemplo, la distribución que se propuso como base del muestreo para obtener una muestra de otra distribución mucho más compleja es una muy simple: la uniforme. El muestreo por rechazo permite generar observaciones de una distribución de la que es difícil obtener una muestra, pero con la cual se puede evaluar la probabilidad de un valor muestra en particular. Es decir, que se tiene una distribución  $p(\theta)$  conocida como **distribución objetivo**, de la cual es difícil obtener una muestra directamente, pero la cual es posible evaluar en un valor particular de  $\theta$ .

Lo primero que se necesita hacer es *proponer* una distribución simple  $q(\theta)$  de la cual se pueda obtener una muestra. Ésta es la razón por la cual a esta distribución se le conoce como **distribución de propuesta**. La idea es evaluar la probabilidad de los valores obtenidos de la muestra tanto en la distribución de propuesta como en la objetivo, y rechazar los valores que sean poco probables tomando en cuenta la distribución objetivo *relativa* a la inicial.

Por otro lado, es necesario encontrar una constante  $c$  tal que

$$c q(\theta) \geq p(\theta) \quad (2.1)$$

para todas los valores de  $\theta$  posibles. A la distribución de propuesta  $q(\theta)$  multiplicada por la constante  $c$  se le conoce como **distribución de comparación**. Encontrar la constante  $c$  puede ser una tarea poco sencilla, pero por ahora se supondrá que es factible hallarla. Después, se obtiene un número  $u$  de una

distribución uniforme entre  $[0, cq(\theta)]$ . Como puede observarse, este valor se encuentra en el segmento de recta entre 0 y la altura de la distribución de comparación evaluado en el valor propuesto  $\theta$ . Dicho valor se rechazará si  $u > p(\theta)$  y se aceptará en caso contrario. Si se acepta, el valor  $\theta$  forma parte de la muestra de la distribución objetivo  $p(\theta)$ . Ver la figura 2.2.

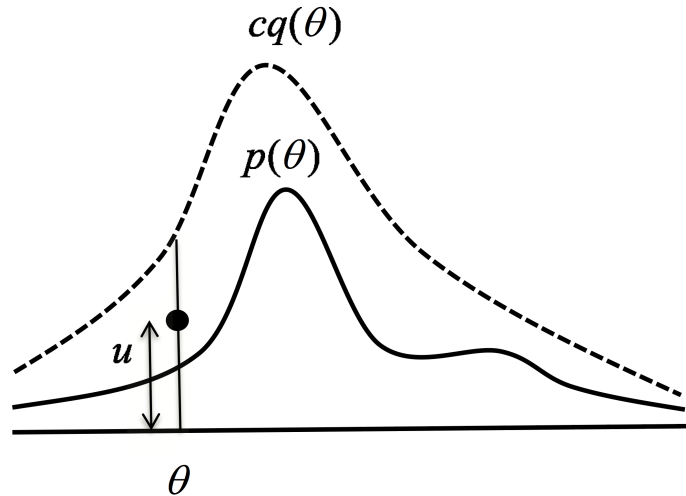


Figura 2.2: Ilustración del muestreo por rechazo. El valor obtenido en este caso, será rechazado.

La clave para que este algoritmo sea eficiente depende de la elección de la distribución de propuesta. Es decir, si ésta es muy diferente de la distribución objetivo, habrá muchos valores rechazados y el proceso de muestreo se hará lento.

## 2.2. Monte Carlo vía cadenas de Markov

Los modelos probabilísticos aplicados a un conjunto de datos, muchas veces dan origen a problemas de inferencia para los cuales se requiere que se integren distribuciones multidimensionales complejas. Los métodos de Monte Carlo vía cadenas de Markov (MCMC) pueden verse como un enfoque computacional en el que la integración se reemplaza por sumas de muestras generadas por algoritmos iterativos. Muchos problemas que prácticamente son intratables si se usan estrategias analíticas, pueden resolverse usando algún método MCMC, aún cuando se trate de distribuciones multidimensionales. De hecho, el desarrollo de los métodos MCMC ha sido clave en el avance

de la estadística computacional. Ahora bien, aunque se sigue haciendo investigación sobre estos métodos, existen algunas técnicas estandarizadas que se usan ampliamente. A continuación se discutirá una estrategia de MCMC conocida como el algoritmo de Metropolis-Hastings, no sin antes entender dos conceptos concernientes a los métodos MCMC: la integración de Monte Carlo y las cadenas de Markov.

### 2.2.1. Integración Monte Carlo

En muchos problemas de inferencia probabilística se requiere de calcular integrales difíciles o sumas en espacios de eventos muy grandes. Por ejemplo, un problema frecuente consiste en hallar la esperanza de una función  $g$  que depende de la variable aleatoria  $X$ .

Si la variable aleatoria es continua y  $p(x)$  es su función de densidad, el problema se reduce a calcular la integral

$$\int g(x)p(x)dx. \quad (2.2)$$

Si  $X$  es discreta,

$$E[g(X)] = \sum g(x)p(x). \quad (2.3)$$

De manera más general, este tipo de problemas aparecen en muchas situaciones donde se desea calcular algún estadístico de una distribución. En el caso de ciertas distribuciones, integrar o sumar usando técnicas analíticas puede llegar a ser una tarea retardadora. Por ejemplo, en el caso de las distribuciones continuas, la función de densidad de la variable aleatoria en cuestión podría tener una forma funcional que no se pudiera integrar analíticamente. En el caso de las distribuciones discretas, el espacio de eventos podría llegar a ser tan grande que sumar sobre todos los valores posibles fuera una tarea imposible.

La idea general de la **integración de Monte Carlo** es usar muestras para aproximar la esperanza de una distribución cuya forma es compleja. Específicamente, se obtiene una muestra  $x^{(t)}$ ,  $t = 1, 2, 3, \dots, N$ , de valores independientes obtenida de la distribución  $p(x)$ . En este caso, se puede aproximar la esperanza tanto en el caso continuo como en el discreto, por una suma finita:

$$E[g(X)] = \frac{1}{N} \sum_{t=1}^N g(x^{(t)}). \quad (2.4)$$

En este procedimiento, se ha reemplazado una integral analítica por una suma sobre un conjunto finito. Generalmente, la aproximación se puede hacer tan exacta como se desee al incrementar el valor de  $N$ . Lo cierto es que la precisión de la aproximación también depende de la independencia de los valores de la muestra. Cuando los valores están correlacionados, el tamaño efectivo de la muestra se hace más pequeño. Éste es un problema potencial del enfoque MCMC.

### 2.2.2. Cadenas de Markov

Una **cadena de Markov** es un proceso estocástico donde un estado cambia de uno a otro mediante un procedimiento secuencial. Una cadena de Markov empieza en algún estado  $x^{(i)}$ , y a través de una función de transición  $p(x^{(t)}|x^{(t-1)})$ , se determina el siguiente estado  $x^{(i+1)}$ . El nuevo estado depende únicamente del estado anterior. Al continuar iterando, se obtiene una sucesión de estados

$$x^{(1)}, x^{(2)}, \dots, x^{(t)}, \dots \quad (2.5)$$

A cualquier sucesión obtenida de este modo se le conoce como cadena de Markov.

Debido a que las cadenas de Markov van de un estado a otro y la transición a un nuevo estado solo depende del último, esta dependencia local es la que hace que este procedimiento no tenga memoria o sea un **proceso de Markov**. Ésta es una propiedad importante de las cadenas de Markov cuando se implementan métodos MCMC.

Cuando se inicializa una cadena de Markov, la cadena se moverá en el espacio de estados alrededor del estado inicial. Por lo tanto, si se genera un cierto número de estas cadenas cada una con estado inicial diferente, en un principio, los estados en los cuales se encuentren las cadenas serán cercanos al estado inicial. Al periodo en el cual sucede lo anterior se le conoce como **periodo de calentamiento**. Una propiedad crucial de las cadenas de Markov es que el estado inicial no afecta al estado de la cadena después de una sucesión de transiciones suficientemente larga (asumiendo que ciertas condiciones acerca de la cadena de Markov se cumplan<sup>3</sup>). Al alcanzar este estado, se dice que la cadena ha alcanzado el **estado estacionario** y que los estados reflejan muestras de la distribución estacionaria. La propiedad de

---

<sup>3</sup>En realidad, gracias al *Teorema Ergódico* es que se puede asegurar que si una cadena de Markov es homogénea, irreducible y aperiódica, ésta alcanza una distribución de equilibrio.

que las cadenas de Markov converjan a un estado estacionario independientemente de dónde empiecen (si se satisfacen ciertas condiciones de la función de transición llamadas de regularidad), es algo decisivo. Al aplicar métodos MCMC, se pueden obtener muestras de una distribución haciendo uso de un procedimiento secuencial en el cual el estado inicial de la sucesión no afecta el proceso de estimación.

### 2.2.3. Métodos de Monte Carlo vía cadenas de Markov

Los **métodos de Monte Carlo vía cadenas de Markov** son particularmente eficientes cuando se tienen distribuciones multivariadas. El objetivo de los métodos MCMC es diseñar una cadena de Markov de modo que la distribución estacionaria de la cadena, coincida con la distribución de la que se desea muestrear originalmente: la **distribución objetivo**; esta última puede ser la distribución final de los parámetros de un modelo, la distribución posterior predictiva en el caso del enfoque bayesiano o cualquier otra distribución de interés. Dicho de otro modo, lo que se busca es que la muestra obtenida con la cadena de Markov también sea una muestra de la distribución objetivo. Ahora bien, con este tipo de muestreo es posible estimar varias características de una distribución como podrían ser la media, la varianza o cualquier otro estadístico de interés.

Pues bien, la idea es usar algunos métodos ingeniosos para definir la función de transición de modo que sin importar cómo se inicialice la cadena, ésta converja a la distribución objetivo. Existen algunos métodos para lograr este objetivo usando procedimientos relativamente sencillos. Se verán los algoritmos de Metropolis, de Metropolis-Hastings y el muestreo de Gibbs.

### 2.2.4. El algoritmo de Metropolis

A continuación se ilustrará el algoritmo más sencillo de los métodos MCMC: el **algoritmo de Metropolis**. Éste es un caso especial del algoritmo de Metropolis-Hastings, el cual se presentará en la siguiente sección.

Por conveniencia, el algoritmo se presentará primero para el caso de una variable. Supóngase entonces que lo que se busca es obtener una muestra de la densidad objetivo  $p(\theta)$ , donde  $\theta \in \mathbb{R}$ . El algoritmo de Metropolis crea una cadena de Markov la cual produce una sucesión de valores  $\theta^{(1)}, \theta^{(1)}, \dots, \theta^{(t)}, \dots$ , donde  $\theta^{(t)}$  representa un estado de la cadena en la iteración  $t$ . La muestra de la cadena, después de un periodo de calentamiento, refleja una muestra

correlacionada de la distribución objetivo. Por esta razón, si se desea obtener una muestra independiente, solo se toma el  $k$ -ésimo elemento de la muestra correlacionada para algún  $k$  natural, el cual se conoce como **retardo**.

En el algoritmo de Metropolis, a  $\theta^{(1)}$  se le asigna un valor inicial. Después, se usa una distribución de propuesta  $q(\theta|\theta^{(t-1)})$  de la que es fácil muestrear, para generar un candidato  $\theta^*$  cuya probabilidad es condicional al estado previo obtenido con el algoritmo. Un requerimiento clave para este algoritmo de es que la distribución de propuesta sea *simétrica*, de modo que

$$q(\theta = \theta^{(t)}|\theta^{(t-1)}) = q(\theta = \theta^{(t-1)}|\theta^{(t)}). \quad (2.6)$$

Por lo tanto, la probabilidad de proponer algún estado nuevo dado el anterior es la misma que proponer ir del estado nuevo y de regreso al estado anterior. Ciertas funciones como la normal, la Cauchy o la  $t$  de student, así como las distribuciones uniformes, son simétricas. Si la simetría no se cumple, entonces se aplica el algoritmo de Metropolis-Hastings, el cual generaliza al algoritmo de Metropolis.

Cabe señalar que la distribución de propuesta se elige por el investigador y *una buena elección de la misma depende del problema*. Una restricción a considerar es la de que la distribución de propuesta debería cubrir el espacio de estados, de modo que cada resultado posible en ese mismo espacio tenga probabilidad distinta de cero bajo la distribución de propuesta.

El siguiente paso consiste en aceptar o rechazar al candidato, es decir, al valor propuesto. La probabilidad de aceptarlo es

$$\alpha = \min \left( 1, \frac{p(\theta^*)}{p(\theta^{(t-1)})} \right). \quad (2.7)$$

Para tomar la decisión de si en realidad se acepta o se rechaza este valor, se genera un valor aleatorio  $u$  de una distribución uniforme  $(0, 1)$ . Si  $u \leq \alpha$ , se acepta el valor propuesto y éste se convierte en el siguiente estado de la cadena:  $\theta^{(t)} = \theta^*$ . En caso contrario, es decir, si  $u > \alpha$ , se rechaza la propuesta y el siguiente estado se define como el anterior:  $\theta^{(t)} = \theta^{(t-1)}$ . Se siguen generando nuevos valores de propuesta condicionales al valor actual del algoritmo, aceptándolos o rechazándolos de acuerdo a la regla anteriormente descrita. El procedimiento continua hasta que el algoritmo converja, es decir, hasta que la cadena de Markov generada alcance una única distribución estacionaria, misma que coincide con la distribución objetivo. En ese momento, la muestra  $\theta^{(1)}, \theta^{(1)}, \dots, \theta^{(t)}, \dots$  reflejará una muestra de la distribución  $p(\theta)$ . Ver la figura 2.3.

**Algoritmo de Metropolis**

Dado  $\theta^{(t)}$ ,

1. Generar  $\theta^* \sim q(\theta|\theta^{(t)})$ , con  $q$  simétrica
2. Tomar  $\theta^{(t+1)} = \begin{cases} \theta^* & \text{con probabilidad } \alpha, \\ \theta^t & \text{con probabilidad } 1 - \alpha, \end{cases}$

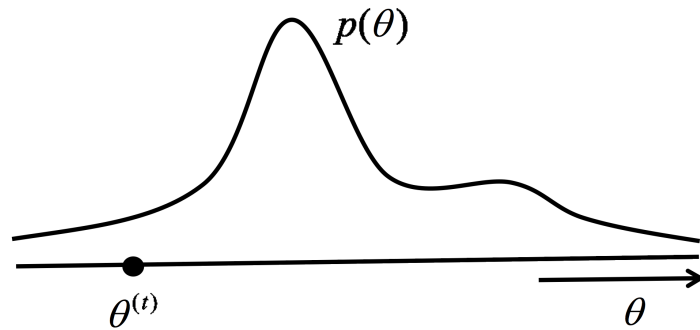
donde

$$\alpha = \min \left( 1, \frac{p(\theta^*)}{p(\theta^{(t)})} \right).$$

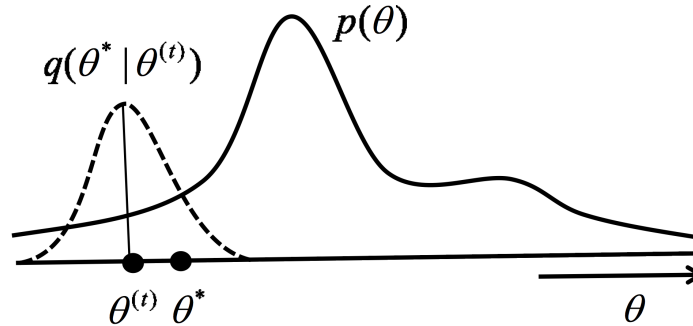
Como puede observarse, la idea del algoritmo de Metropolis para elegir a un candidato es similar a la que se vio en la sección 2.1.3. Lo cierto es que, a diferencia de los algoritmos de aceptación-rechazo, el algoritmo de Metropolis también se puede aplicar a problemas de más de una dimensión.

Para entender intuitivamente por qué se logra obtener una muestra de la distribución objetivo con este proceso, es preciso notar que en la ecuación 2.7, siempre se acepta el valor propuesto si éste tiene una probabilidad mayor de ocurrir que la del estado anterior (en ambos casos se considera a la densidad de la distribución objetivo para calcular dicha probabilidad). Por lo tanto, el algoritmo se moverá hacia las regiones del espacio de estados en las cuales la función objetivo tenga una densidad mayor. Sin embargo, hay que notar que si es menos probable que el nuevo valor propuesto tenga una menor probabilidad que el estado actual, es posible aceptar este "peor" valor propuesto y moverse hacia él. Este proceso de aceptar siempre una propuesta "buena" y una "mala" de vez en cuando, asegura que el algoritmo explore por completo el espacio de estados, y obtenga valores de la muestra de todas las partes de la distribución, incluidas las colas.

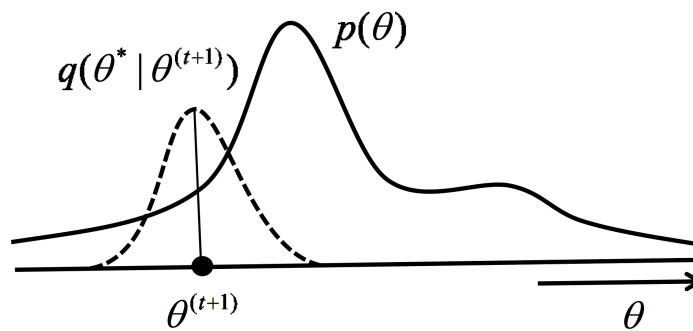
Una gran ventaja del algoritmo de Metropolis es que la ecuación 2.7 requiere de solo un cociente de densidades. Por lo tanto, cualquier término que no dependa de  $\theta$  en la forma funcional de  $p(\theta)$ , es irrelevante. Esto significa que *no se necesita de una constante de normalización de la función de densidad de  $\theta$* . De hecho, esta característica de poder usar funciones de densidad sin sus constantes de normalización es uno de los principales atractivos de este algoritmo. Querer calcular una muestra de una distribución sin normalizar es algo que sucede frecuentemente con los modelos bayesianos, donde calcular la constante de normalización es difícil.



(a) Inicio.



(b) Propuesta.



(c) Aceptación.

Figura 2.3: Ilustración del algoritmo de Metropolis para obtener una muestra de una densidad  $p(\theta)$ , no necesariamente normalizada. En 2.3a, el estado actual de la cadena es  $\theta^{(t)}$ . En 2.3b, se utiliza una distribución de propuesta cerca del estado actual para generar una propuesta  $\theta^*$ . Finalmente en 2.3c, la propuesta se acepta en este caso y el nuevo estado se define igual a ésta; la distribución de propuesta ahora se centra en el nuevo estado.



### 2.2.5. El algoritmo de Metropolis-Hastings

El **algoritmo de Metropolis-Hastings** (MH) es una versión generalizada del algoritmo de Metropolis, porque a diferencia del último, con éste se puede hacer uso de distribuciones de propuesta que no sean simétricas. Este algoritmo opera exactamente de la misma forma que el algoritmo de Metropolis, aunque requiere de una probabilidad de aceptación diferente:

$$\alpha = \min \left( 1, \frac{p(\theta^*)}{p(\theta^{(t-1)})} \frac{q(\theta^{(t-1)}|\theta^*)}{q(\theta^*|\theta^{(t-1)})} \right). \quad (2.8)$$

El cociente adicional  $\frac{q(\theta^{(t-1)}|\theta^*)}{q(\theta^*|\theta^{(t-1)})}$  en la ecuación 2.8, corrige cualquier asimetría en la distribución de propuesta. Por ejemplo, supóngase que se tiene una distribución de propuesta que está sesgada en una dirección, con una media centrada en el estado actual. Si la distribución de propuesta dicta moverse, por decir algo, a la derecha en vez de a la izquierda, el cociente de las densidades corregirá esta asimetría.

#### Algoritmo de Metropolis-Hastings

Dado  $\theta^{(t)}$ ,

1. Generar  $\theta^* \sim q(\theta|\theta^{(t)})$
2. Tomar  $\theta^{(t+1)} = \begin{cases} \theta^* & \text{con probabilidad } \alpha, \\ \theta^{(t)} & \text{con probabilidad } 1 - \alpha, \end{cases}$   
donde

$$\alpha = \min \left( 1, \frac{p(\theta^*)}{p(\theta^{(t)})} \frac{q(\theta^{(t)}|\theta^*)}{q(\theta^*|\theta^{(t)})} \right).$$

Cuando se tienen variables acotadas, se debe tener cuidado al construir una distribución de propuesta adecuada. En general, *una regla apropiada consiste en usar una distribución de propuesta que tenga el mismo soporte que la distribución objetivo.*

### 2.2.6. El algoritmo de Metropolis-Hastings para distribuciones multivariadas

Es relativamente directo generalizar el algoritmo de Metropolis-Hastings para distribuciones multivariadas. Lo cierto es que existen dos modos diferen-

tes de generalizar el procedimiento para ahora obtener muestras de variables aleatorias en espacios multidimensionales.

### Actualización por bloque

Con este enfoque, se usa una distribución de propuesta de la *misma dimensión* que la distribución objetivo. Esto significa que si se desea obtener una muestra de una distribución que esté formada por  $N$  variables, se designa una distribución de propuesta de dimensión  $N$ , y la propuesta se acepta o rechaza como un bloque, el cual está conformado por todas las variables.

Sea  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_N)$  un vector que representa una variable aleatoria de  $N$  componentes y  $\boldsymbol{\theta}^{(t)}$  el estado  $t$  de esa variable. Se puede generalizar el algoritmo de Metropolis-Hastings, donde las variables escalares  $\theta$  se reemplazan por los vectores  $\boldsymbol{\theta}$ .

#### Algoritmo MH por bloque

Dado  $\boldsymbol{\theta}^{(t)}$ ,

1. Generar  $\boldsymbol{\theta}^* \sim q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$

2. Tomar  $\boldsymbol{\theta}^{(t+1)} = \begin{cases} \boldsymbol{\theta}^* & \text{con probabilidad } \alpha, \\ \boldsymbol{\theta}^{(t)} & \text{con probabilidad } 1 - \alpha, \end{cases}$

donde

$$\alpha = \min \left( 1, \frac{p(\boldsymbol{\theta}^*) q(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^*)}{p(\boldsymbol{\theta}^{(t)}) q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t)})} \right).$$

### Actualización componente a componente

El primer problema que puede presentarse con el enfoque anterior es que podría ser difícil encontrar una distribución de propuesta multidimensional adecuada. Otro problema de la actualización por bloque es que la tasa de rechazo puede ser alta [23]. Por estas dos razones, en vez de aceptar o rechazar una propuesta para  $\boldsymbol{\theta}$  que involucre todas sus componentes a la vez, podría ser más sencillo computacionalmente, proponer valores para cada una de las componentes de  $\boldsymbol{\theta}$ , una a la vez. Esto da origen a la *actualización componente a componente*.

Con esta técnica, al igual que en la actualización por bloque, se le asigna un valor inicial a  $\boldsymbol{\theta} \in \mathbb{R}^N$ . Después, lo que se hace es que simplemente para

cada componente del vector  $\theta$ , se aplica el algoritmo de Metropolis-Hastings univariado a dicha componente manteniendo a todas las demás constantes.

**Algoritmo MH componente a componente**

Dado  $\theta^{(t)} \in \mathbb{R}^N$ , para cada  $i = 1, 2, \dots, N$

1. Generar  $\theta_i^* \sim q(\theta_i | \theta_i^{(t)})$

2. Tomar  $\theta_i^{(t+1)} = \begin{cases} \theta_i^* & \text{con probabilidad } \alpha, \\ \theta_i^{(t)} & \text{con probabilidad } 1 - \alpha, \end{cases}$

donde

$$\alpha = \min \left( 1, \frac{p(\theta_i^*) q(\theta_i^{(t)} | \theta_i^*)}{p(\theta_i^{(t)}) q(\theta_i^* | \theta_i^{(t)})} \right).$$

Existe otra versión de este mismo algoritmo en la cual las componentes se toman en orden aleatorio.

### 2.2.7. El muestreo de Gibbs

Una desventaja tanto del algoritmo de Metropolis-Hastings como del muestreo por rechazo, es que puede ser difícil calibrar la distribución de propuesta. Esto significa que si en un principio, la media de la distribución se encontrara lejos de los valores de la distribución objetivo que tienen mayor probabilidad, o que si por la elección que se haya hecho de la varianza, ésta fuera muy pequeña o muy grande, entonces el algoritmo convergería más lentamente. Por otro lado, una buena parte de los cálculos se llevan a cabo produciendo muestras que se rechazan y no se usan en la aproximación.

El **muestreo de Gibbs** es un procedimiento en el cual todas las muestras se aceptan, lo que da lugar a una mejora en la eficiencia computacional. Una ventaja adicional de este método es que el investigador no tiene que especificar una distribución de propuesta, lo que permite que el procedimiento MCMC haga ciertas estimaciones.

Sin embargo, el procedimiento únicamente se puede aplicar en situaciones en las cuales se conocen todas las distribuciones condicionales de cada componente de la distribución multivariada dadas todas las demás -su complemento-. En algunos casos, estas distribuciones condicionales sencillamente no se conocen, y por lo tanto, no es posible el aplicar el muestreo

de Gibbs. Lo cierto es que en muchos modelos bayesianos, sí se tienen las distribuciones que se prestan para este tipo de muestreo.

Para ilustrar el caso del muestreo de Gibbs, se tomará una distribución bivariada  $f(\theta_1, \theta_2)$ . El requerimiento clave para este tipo de algoritmo es que se muestrean fácilmente de las dos distribuciones condicionales  $f(\theta_1|\theta_2 = \theta_2^{(t)})$ ,  $f(\theta_2|\theta_1 = \theta_1^{(t)})$ , es decir, la distribución de cada variable dada una realización específica de la *otra* variable. Al principio, el muestreo se inicializa con algunos valores apropiados  $\theta_1^{(1)}$  y  $\theta_2^{(1)}$ . En cada iteración  $t$ , se siguen pasos muy similares a los del algoritmo de Metropolis-Hastings componente a componente. Es decir, primero se muestrea un nuevo valor  $\theta_1^{(t)}$  dado el estado previo de la otra componente  $\theta_2^{(t-1)}$ ; esto se lleva a cabo obteniendo un valor muestral de  $f(\theta_1|\theta_2 = \theta_2^{(t-1)})$ . A diferencia del algoritmo de MH, esta propuesta siempre se acepta, de modo que el nuevo estado se actualiza inmediatamente. En el segundo paso, se muestrea un nuevo valor para  $\theta_2^{(t)}$  dado el estado actual de la otra componente  $\theta_1^{(t)}$ ; esto se hace obteniendo un valor muestral de  $f(\theta_2|\theta_1 = \theta_1^{(t)})$ . Por lo tanto, el procedimiento da lugar a un muestreo condicional iterativo en el cual se sigue yendo y viniendo al obtener un valor muestral para el nuevo estado de una variable dados los valores actuales de la otra componente. Para distribuciones de tres o más variables se tiene el algoritmo general que aparece a continuación.

#### Muestreo de Gibbs

Dado  $\boldsymbol{\theta}^{(t)}$ , generar

1.  $\theta_1^{(t+1)} \sim q(\theta_1|\theta_2^{(t)}, \dots, \theta_N^{(t)})$
- ⋮
- $i$ .  $\theta_i^{(t+1)} \sim q(\theta_i|\theta_1^{(t+1)}, \dots, \theta_{i-1}^{(t+1)}, \theta_{i+1}^{(t)}, \dots, \theta_N^{(t)})$
- ⋮
- $N$ .  $\theta_N^{(t+1)} \sim q(\theta_N|\theta_1^{(t+1)}, \dots, \theta_{N-1}^{(t+1)})$

### 2.3. Conceptos del análisis de datos bayesiano

Hasta ahora se han ilustrado dos métodos Monte Carlo vía cadenas de Markov, pero probablemente no sea muy claro cómo es que estos algoritmos puedan ser útiles para modelar o para analizar los datos. En esta sección, se

verá cómo es que los métodos MCMC se pueden usar en el contexto del análisis de datos bayesiano. El objetivo será encontrar los valores de los parámetros en un modelo probabilístico que mejor expliquen a los datos.

Sea  $\mathbf{y}$  el conjunto de datos que se desean analizar<sup>4</sup> y  $\boldsymbol{\theta}$  el de los parámetros del modelo. Supóngase además que se tiene un modelo diseñado para explicar los datos en el sentido de que éstos se pueden generar con dicho modelo. Entonces se puede definir la verosimilitud del modelo

$$p(\mathbf{y}|\boldsymbol{\theta}). \quad (2.9)$$

La verosimilitud básicamente refleja qué tan probable es obtener los datos con el modelo habiendo elegido un conjunto específico de parámetros. Es decir, explica los datos en términos de los parámetros. Pues bien, también se supondrá que es posible evaluar la verosimilitud.

El objetivo principal de cualquier proceso de modelado consiste en escoger los parámetros del modelo de modo que con éste se puedan explicar mejor los datos que con cualquier otro conjunto de parámetros.

### 2.3.1. Estrategias de la estimación de parámetros

#### 2.3.2. Máxima Verosimilitud

Un enfoque básico que se usa para la estimación de parámetros es el **método de máxima verosimilitud**. El objetivo de esta técnica consiste en encontrar la estimación de los parámetros que maximice la probabilidad de los datos, dado el modelo con esa estimación de los parámetros en específico. Es decir,

$$\boldsymbol{\theta}_{MV} = \arg \min_{\theta_i} p(\mathbf{y}|\theta_i),$$

donde los valores  $\theta_i$  varían en el rango de todos los valores posibles. El estimador de máxima verosimilitud corresponde con los valores de los parámetros que mejor coinciden con los datos observados. Aunque éste parezca un modo muy directo para estimar los parámetros de un modelo probabilístico, en realidad tiene algunos inconvenientes serios. Por ejemplo, si se tuviera un modelo con muchos parámetros, podría ser difícil encontrar el máximo de la

---

<sup>4</sup>Propiamente se tiene un conjunto de pares ordenados, pero se eligió a al conjunto de datos que representan a la respuesta  $\mathbf{y}$  para simplificar la notación.

función de verosimilitud. En este caso, los métodos numéricos podrían quedarse atrapados en máximos locales, o ser muy sensibles a la elección de los valores iniciales necesarios para dichos métodos.

### 2.3.3. El máximo a posteriori

En este caso, el valor de  $\boldsymbol{\theta}$  es desconocido pero fijo. Una estrategia bayesiana para la estimación de parámetros consiste en asignarle a esta incertidumbre un modelo de probabilidad, tomando a  $\boldsymbol{\theta}$  como una variable aleatoria. A esta variable se le asocia una **distribución a priori**  $p(\boldsymbol{\theta})$ . Esta distribución, también llamada **distribución inicial**, denota el conocimiento general que se tiene de los parámetros antes de observar los datos. En este contexto, uno de los objetivos consiste en encontrar la distribución **final** o **a posteriori** de los parámetros dados los datos<sup>5</sup>. Es decir, uno desearía conocer  $p(\boldsymbol{\theta}|\mathbf{y})$ . Aplicando el teorema de Bayes, se puede evaluar esta probabilidad condicional de la siguiente manera:

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{y})}. \quad (2.10)$$

Como  $p(\boldsymbol{\theta}|\mathbf{y})$  es una función que depende de la variable  $\boldsymbol{\theta}$ ,  $p(\mathbf{y})$  se puede ver como una constante, y por lo tanto se escribe

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}). \quad (2.11)$$

Es preciso notar que la distribución a posteriori está especificada cuando se conoce el producto de la verosimilitud (qué tan probables son los datos dado el modelo con esos parámetros estimados en específico) y la distribución a priori (qué tan probables son esos parámetros al inicio). En el análisis de datos, un modo de ajustarles un modelo consiste en encontrar los valores máximos de la distribución final de los parámetros. La meta es encontrar los estimadores de los parámetros que maximicen la probabilidad final de los

---

<sup>5</sup>La distribución final o a posteriori es la misma *distribución objetivo* de la que se habla en la sección 2.2.3.

parámetros dados los datos. En otras palabras,

$$\begin{aligned}
 \boldsymbol{\theta}_{MAP} &= \arg \min_{\theta_i} p(\theta_i | \mathbf{y}), \\
 &= \arg \min_{\theta_i} \frac{p(\mathbf{y} | \theta_i) p(\theta_i)}{p(\mathbf{y})}, \\
 &= \arg \min_{\theta_i} p(\mathbf{y} | \theta_i) p(\theta_i).
 \end{aligned} \tag{2.12}$$

Este proceso se parece al del cálculo de máxima verosimilitud. La diferencia se encuentra en que la distribución inicial de los parámetros influirá en la estimación de los mismos.

### 2.3.4. El muestreo a posteriori

La estrategia bayesiana de nuestro interés es el **muestreo a posteriori**. Con el enfoque de máxima verosimilitud, cuando la distribución es unimodal, se obtiene un único conjunto de parámetros para un modelo. Por lo tanto, se caracteriza a la distribución objetivo precisamente con la *moda* de la misma. Aunque este hecho sea conceptualmente claro, tiene algunas desventajas. Por ejemplo, si existieran diferentes conjuntos de valores de los parámetros, los cuales tienen una probabilidad a posteriori alta, probablemente uno debería tratar de obtener esos otros conjuntos; o qué tal si ahora se tienen intercambios entre los valores de los parámetros. Por ejemplo, podría tenerse un modelo con parámetros  $a$  y  $b$  con el cual, ya sea dándole a  $a$  un valor grande y a  $b$  uno pequeño o viceversa, se lograra una probabilidad a posteriori alta.

Teniendo un enfoque bayesiano, el objetivo es caracterizar la distribución a posteriori y no simplemente encontrar la moda de la misma. En algunos casos se podría encontrar una expresión analítica de la distribución final. Sin embargo, en muchos casos, hay que recurrir a métodos como el de Monte Carlo vía cadenas de Markov para obtener muestras de esta distribución. Estas muestras se pueden usar para calcular una variedad de características, como lo son las medias, las varianzas u otros momentos de la distribución, además de que se puede verificar si existen o no correlaciones entre los parámetros.

Para finalizar, es importante hacer notar que los métodos de Monte Carlo vía cadenas de Markov se pueden usar para obtener muestras de una distribución, y que en particular, estos métodos se pueden aplicar para obtener muestras de la distribución final. Por lo tanto, los dos métodos MCMC que se

discutieron anteriormente se pueden aplicar para el problema de la inferencia a posteriori.



# Capítulo 3

## Aplicaciones

En los capítulos anteriores se definieron, primero, el tipo de problemas que se modelan en farmacocinética. En el siguiente, se mostró la teoría estadística para obtener una muestra de una distribución compleja -dado un conjunto de datos- y un modelo que es adecuado para describirlos. En este capítulo se aplicarán estos conceptos en conjunto para resolver tres problemas concretos. Para todos los casos, se tienen datos de un experimento realizado en cierto tipo de pacientes a los cuales se les administró una sustancia en particular. El propósito que se busca para cada uno es describir la concentración del medicamento a través del tiempo. Esto se llevará a cabo a través de un modelo farmacocinético por compartimentos, en el cual, se lleva a cabo inferencia bayesiana para conocer las distribuciones de los parámetros de dicho modelo. Cabe señalar que solo se ejemplificará detalladamente la metodología en el caso del primer problema, ya que ésta no cambia. Posteriormente únicamente se especificarán los modelos utilizados y los resultados.

### 3.1. Primer ejemplo

En este caso se tomaron los datos de un experimento realizado por Boeckmann et al (1994). Los datos contienen, inicialmente, la información de 12 pacientes a los cuales se les administró una dosis de teofilina, un medicamento usado en el tratamiento de las enfermedades respiratorias. A este grupo de pacientes se les tomaron 11 mediciones de la concentración del medicamento en la sangre en un periodo de 25 horas. Dado que previamente a la administración de la dosis se observó que tres pacientes tenían rastros del fármaco

en el organismo, solo se consideró a los 9 restantes, es decir, a los pacientes cuya concentración de teofilina en la sangre al momento de empezar el experimento fuera de 0. Los datos se pueden consultar en el sitio del proyecto R (R Development Team, 2016). Una representación gráfica de los datos se puede ver en la figura 3.1.

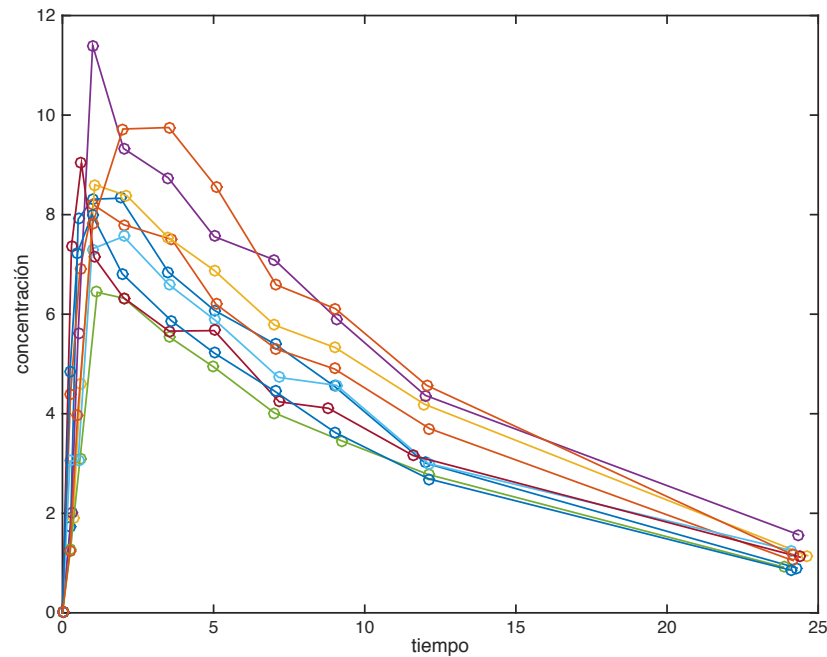
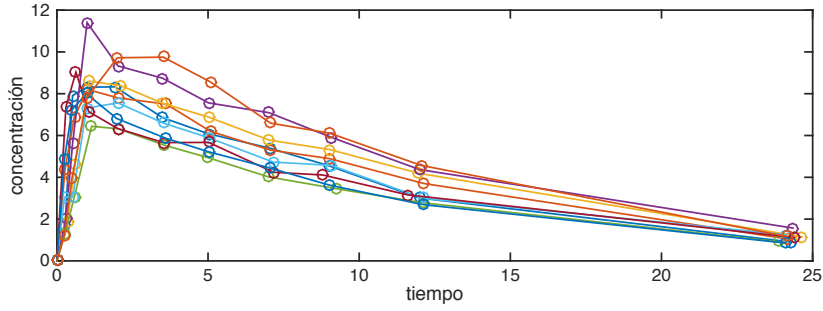


Figura 3.1: Perfiles de concentración de teofilina en 9 pacientes.

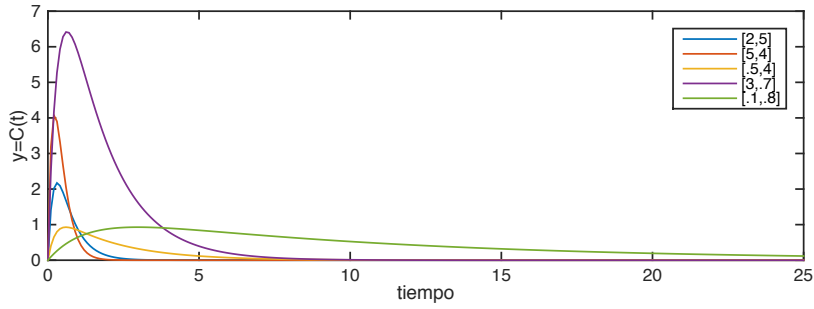
### 3.1.1. Metodología

#### Modelo PK por compartimentos

Primeramente, para modelar los perfiles de concentración se utilizó un *modelo PK de administración oral de un compartimento con absorción de primer orden y eliminación lineal* (sección 1.9). Antes de continuar, uno debería estar seguro de que usar un modelo como éste es adecuado para ajustar los datos. Lo cierto es que así es, como puede verse en la figura 3.2, donde se tomaron distintos valores para los parámetros  $\alpha$  y  $\beta$  de la solución 3.2. El esquema de este modelo puede verse en la figura 3.3.



(a) Datos del experimento.



(b) Gráficas de la solución  $C(t)$  para diferentes pares de parámetros  $(\alpha, \beta)$ .

Figura 3.2: Comparación entre la gráfica que describe a los datos 3.2a y de la concentración  $C(t)$  para diferentes valores de los parámetros  $\alpha$  y  $\beta$ , donde de acuerdo a [17], se tomó  $\frac{F \cdot D}{V} = 10$ .

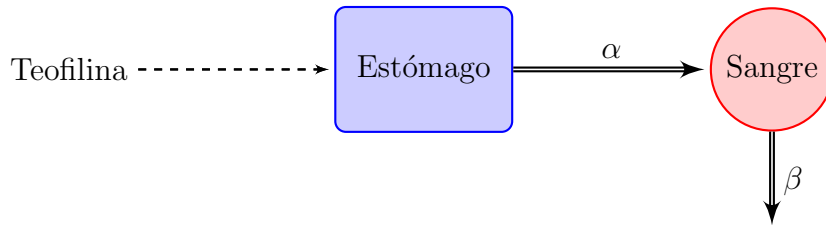


Figura 3.3: Esquema del modelo de administración oral de un compartimento.

Ahora bien, dado que al momento del iniciar el tratamiento no existía concentración del fármaco en el compartimento central, el modelo farmacocinético con condiciones iniciales es

$$\begin{aligned}x_1'(t) &= -\alpha x_1(t), & x_1(0) &= D, \\x_2'(t) &= \alpha x_1(t) - \beta x_2(t), & x_2(0) &= 0.\end{aligned}\tag{3.1}$$

La concentración en el compartimento central ( $x_2$ ) está dada por

$$C(t) = \frac{c\alpha}{\alpha - \beta}(e^{-\beta t} - e^{-\alpha t}), \quad c = \frac{FD}{V}.\tag{3.2}$$

### Modelo estadístico

Dado el conjunto de datos  $\{(t_{ijk}, y_{ijk})\}$  y el modelo PK dado por 3.1, se considera el siguiente modelo estadístico<sup>1</sup>

$$y_{ijk} = x_j(t_{ijk}|\boldsymbol{\theta}) + \epsilon_{ijk},\tag{3.3}$$

donde

- $y_{ijk}$  es la  $k$ -ésima observación de la concentración de la  $i$ -ésima unidad experimental en la  $j$ -ésima función de estado registrada en el tiempo  $t_{ijk}$ ,
- $j \in C \subset \{1, 2, 3, \dots, d\}$ ,
- $N$  es el número de unidades experimentales,
- $n_{ij}$  es el número de observaciones de la  $i$ -ésima unidad experimental en la  $j$ -ésima función de estado,
- $\epsilon_{ijk} \sim N(0, \tau_j^{-\frac{1}{2}} x_j(t_{ijk}|\boldsymbol{\theta}))$ ,  $\tau_j > 0$  y son independientes entre sí.

Este modelo estadístico está pensado como el resultado de un proceso determinista representado por la solución del modelo PK, más una componente aleatoria representada por el error<sup>2</sup>. En otras palabras, se supone que todos

<sup>1</sup>La formulación del modelo estadístico es idéntica a la desarrollada en [17].

<sup>2</sup>En el apéndice A se pueden ver distintos modos de agregar aleatoriedad a un sistema de ecuaciones diferenciales. En este caso se trata del mismo ruido aditivo simple del que se habla en la sección A.2.2; la diferencia en este caso radica en que el ruido se añade a una de las soluciones del sistema de ecuaciones.

los pacientes tienen el mismo comportamiento, y la variabilidad entre cada paciente y entre cada instante se representa por medio del error. Ahora bien, la desviación estándar del error se encuentra así definida, porque como se puede observar en la figura 3.1, la variabilidad del proceso decrece a medida que transcurre el tiempo. En realidad ésta es una característica de los modelos farmacocinéticos, ya que el medicamento es eliminado por el organismo después de un cierto periodo de tiempo<sup>3</sup>.

### Inferencia estadística

Los parámetros del modelo estadístico son entonces  $\boldsymbol{\theta} = (\alpha, \beta)$  y  $\tau$ . Como se ha dicho antes, estos parámetros son fijos pero desconocidos. Eligiendo un enfoque bayesiano, la incertidumbre se describe por medio de un modelo de probabilidad. Entonces se consideró a los parámetros como variables aleatorias, para después aplicar el muestreo a posteriori visto en las secciones 2.3.3 y 2.3.4.

De este modo, por el teorema de Bayes se tiene que

$$p(\boldsymbol{\theta}, \tau | \mathbf{y}) \propto p(\mathbf{y} | \boldsymbol{\theta}, \tau) p(\boldsymbol{\theta}, \tau),$$

donde  $p(\boldsymbol{\theta}, \tau | \mathbf{y})$  es la distribución final o a posteriori,  $p(\mathbf{y} | \boldsymbol{\theta}, \tau)$  es la verosimilitud y  $p(\boldsymbol{\theta}, \tau)$  es la distribución inicial. La verosimilitud está dada por

$$p(\mathbf{y} | \boldsymbol{\theta}, \tau) = \prod_{i=1}^N \prod_{j \in C} \prod_{k=1}^{n_{ij}} \frac{\sqrt{\tau}}{\sqrt{2\pi} x_j(t_{ijk} | \boldsymbol{\theta})} e^{-\frac{\tau}{2} \left[ \frac{y_{ijk} - x_j(t_{ijk} | \boldsymbol{\theta})}{x_j(t_{ijk} | \boldsymbol{\theta})} \right]^2}.$$

Específicamente, el compartimento de interés en este caso es el central ( $j = 2$ ); lo que se busca es describir la concentración del fármaco en la sangre. Tomando a  $n_{ij} = n$ , la verosimilitud en este caso está dada por

$$p(\mathbf{y} | \alpha, \beta, \tau) = \prod_{i=1}^N \prod_{k=1}^n \frac{\sqrt{\tau}}{\sqrt{2\pi} x_2(t_{ik} | \alpha, \beta)} e^{-\frac{\tau}{2} \left[ \frac{y_{ik} - x_2(t_{ik} | \alpha, \beta)}{x_2(t_{ik} | \alpha, \beta)} \right]^2}. \quad (3.4)$$

Luego, la distribución de propuesta  $p(\alpha, \beta, \tau)$  depende de los parámetros positivos  $\alpha$ ,  $\beta$  y  $\tau$ , ya que los dos primeros representan tasas y el tercero

---

<sup>3</sup>Ésta es la razón por la cual, cuando se prescribe un tratamiento que no es de una sola toma, la dosis tiene que ser administrada cada determinado tiempo; de este modo, se garantiza que el organismo tenga al menos cierta concentración del medicamento en cualquier momento mientras dure el tratamiento.

calibra la variabilidad de la solución del modelo PK. Por lo tanto, en [17] se propuso

$$P(\alpha, \beta, \tau) = P(\alpha; a_\alpha, b_\alpha)P(\beta; a_\beta, b_\beta)P(\tau; a_\tau, b_\tau), \quad (3.5)$$

donde cada distribución del lado derecho es una gamma de la forma

$$P(x; a, b) = \frac{1}{\Gamma(a)b^a} x^{a-1} e^{-\frac{x}{b}} \mathbb{I}_{(0, \infty)}(x). \quad (3.6)$$

Es evidente que por haber definido de este modo la distribución de propuesta, se consideró que los parámetros  $\alpha$ ,  $\beta$  y  $\tau$ , vistos como variables aleatorias, son independientes entre sí. Los parámetros de una distribución inicial representan el conocimiento que se tenga del fenómeno que se intenta describir. Si no se tiene conocimiento de éste o el que se tiene resulta ser vago, entonces las distribuciones iniciales deberían reflejar este hecho. A las distribuciones de este tipo se les conoce como **no informativas**. En específico, en [17] se tomaron  $(a_\alpha, b_\alpha) = (4, 1)$ ,  $(a_\beta, b_\beta) = (1, 2)$  y  $(a_\tau, b_\tau) = (2, 160)$ . La gráfica de las densidades de probabilidad se puede ver en la figura 3.4.

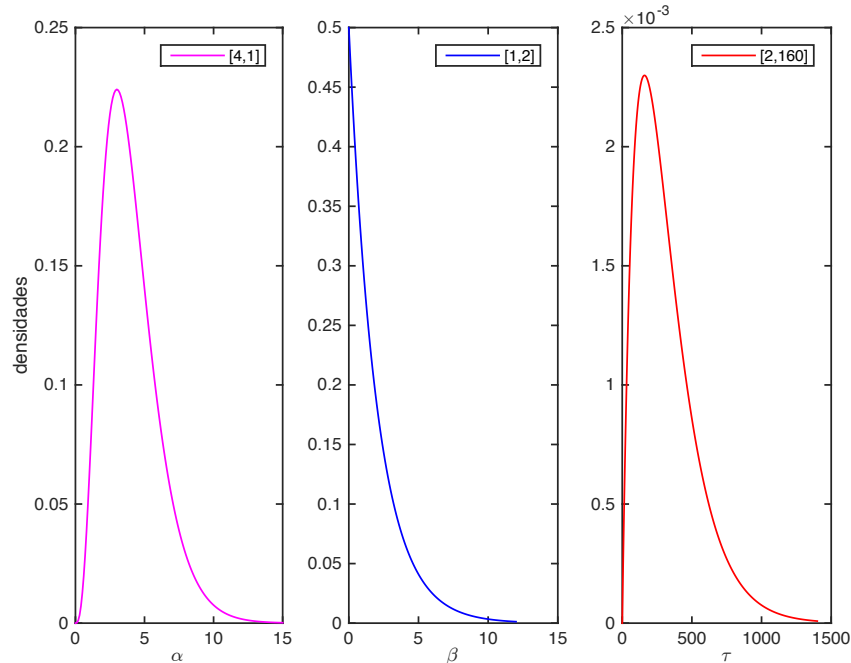


Figura 3.4: Gráficas de las funciones de densidad referentes a la distribución de propuesta 3.5.

Por último, la distribución a posteriori queda como sigue

$$p(\alpha, \beta, \tau | \mathbf{y}) \propto \left[ \prod_{i=1}^N \prod_{k=1}^n x_2(t_{ik} | \alpha, \beta) \right]^{-1} \tau^{\frac{N*n}{2}} e^{-\frac{1}{2}\tau \sum_{i=1}^N \sum_{k=1}^n \left[ \frac{y_{ik} - x_2(t_{ik} | \alpha, \beta)}{x_2(t_{ik} | \alpha, \beta)} \right]^2} \cdot P(\alpha; 4, 1) P(\beta; 1, 2) P(\tau; 2, 160). \quad (3.7)$$

Ahora bien, sería preciso obtener una muestra de esta distribución mediante algún método MCMC. Lo cierto es que para simplificar los cálculos, se obtuvo primero la densidad marginal  $p(\alpha, \beta, \tau | \mathbf{y}) d\tau = \int_0^\infty p(\alpha, \beta, \tau | \mathbf{y}) d\tau$ . Después de haber hecho los cálculos se observó que

$$p(\alpha, \beta | \mathbf{y}) \propto [P(\alpha, \beta)]^{-1} \alpha^{a_\alpha - 1} \beta^{a_\beta - 1} e^{-\frac{\alpha}{b_\alpha} - \frac{\beta}{b_\beta}} \cdot B_\tau(\alpha, \beta)^{A_\tau(\alpha, \beta)}, \quad (3.8)$$

donde

$$P(\alpha, \beta) = \prod_{i=1}^N \prod_{k=1}^n c(e^{-\beta t_{ik}} - e^{-\alpha t_{ik}}),$$

$$B_\tau(\alpha, \beta) = \left( \frac{1}{2} \sum_{i=1}^N \sum_{k=1}^n \left[ \frac{y_{ik} - c(e^{-\beta t_{ik}} - e^{-\alpha t_{ik}})}{c(e^{-\beta t_{ik}} - e^{-\alpha t_{ik}})} \right]^2 + \frac{1}{b_\tau} \right)^{-1},$$

y

$$A_\tau(\alpha, \beta) = \frac{Nn}{2} + a_\tau.$$

Lo que es más, se observó que

$$\tau | \alpha, \beta \sim \text{Gamma}(A_\tau(\alpha, \beta), B_\tau(\alpha, \beta)), \quad (3.9)$$

lo que implica que obtener una muestra de la distribución  $p(\tau | \alpha, \beta)$  se reduce a obtener primero una muestra de la distribución marginal 3.8. Además, dado que lenguajes como R o MATLAB tienen ya implementadas funciones para obtener muestras de funciones de densidad conocidas, la tarea de obtener la segunda muestra es algo sencillo.

### 3.1.2. Implementación

Para obtener la muestra de la distribución 3.8, primero se llevo a cabo la transformación

$$Z_1 = \log \alpha, \quad Z_2 = \log \beta. \quad (3.10)$$

Habiendo hecho lo anterior, se aplicó el algoritmo de Metropolis por bloque a la distribución transformada  $p(\mathbf{z})$ , con  $\mathbf{z} = (z_1, z_2)^T$ , tomando como distribución de propuesta una normal multivariada. Por el teorema de aproximación normal (ver el apéndice B), se tomó como la media de esta distribución a la moda de  $p(\mathbf{z})$  y como matriz de varianza-covarianza a  $V = -\left[\frac{\partial^2}{\partial \mathbf{z}^T \partial \mathbf{z}} \log p(\mathbf{z})\right]^{-1}$ . El cálculo de la moda se hizo con el módulo de optimización de MATLAB utilizando la función *fminunc*, y el de la matriz, mediante un programa implementado igualmente en MATLAB aplicando diferencias hacia adelante. Después de haber obtenido la muestra anterior, se obtuvo la muestra de 3.9 con la función *rgamma* del lenguaje R. Esto significa que para este último paso se aplicó el muestreo de Gibbs.

### 3.1.3. Resultados

Se obtuvo una muestra de 5000 elementos, para la cual se tomó un periodo de calentamiento igual a 100,000, y a partir de entonces se tomaron los elementos de la muestra dejando pasar 25 entre uno y otro, con el propósito de disminuir la correlación entre ellos<sup>4</sup>. Los resultados de la convergencia y de la autocorrelación de la muestra obtenida se pueden ver en la figura 3.5. En cuanto a la concentración, al igual que en [17], se eligió una constante  $\frac{FD}{V} = 10$ . Los histogramas para cada uno de los parámetros y la distribución predictiva final se muestran en la figura 3.6. Las modas de las muestras obtenidas para cada uno de los parámetros coinciden con las reportadas en [17], en una décima para el caso de  $\alpha$  y  $\tau$ , y en tres décimas para el caso de  $\beta$  (ver tabla 3.1).

Resultados	$\alpha$	$\beta$	$\tau$
Este trabajo	1.858398	0.09478814	15.0942
Montoya et al	1.8276	0.0949	15.0796

Cuadro 3.1: Comparación entre los resultados obtenidos en este trabajo y los obtenidos en [17].

<sup>4</sup>Para elegir el periodo de calentamiento y el retardo, tanto en este ejemplo como en los subsecuentes, se corrió el programa donde se implementó el algoritmo de Metropolis una primera vez para calcular muestras de tamaño  $m$  y se graficaron los promedios ergódicos  $(\frac{1}{n} \sum_{i=1}^n x_i, n = 1, \dots, m)$  para cada muestra  $\{x_i\}_{i=1}^m$ .



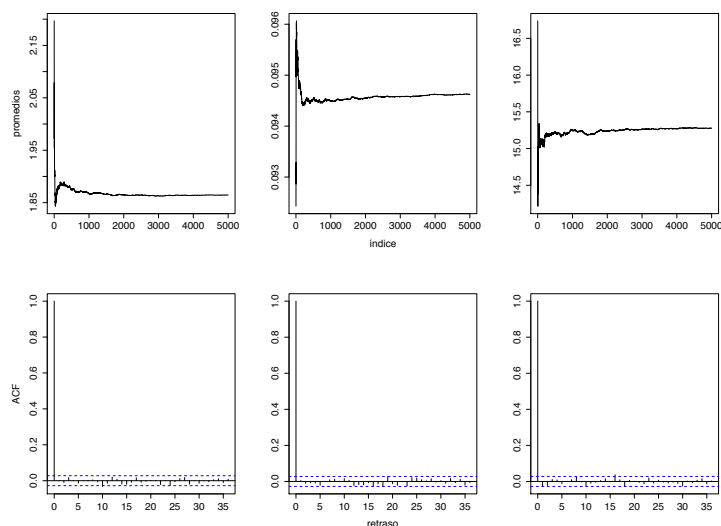


Figura 3.5: Gráficas de los promedios ergódicos y de la función de autocorrelación de las muestras obtenidas para cada uno de los parámetros  $\alpha$ ,  $\beta$  y  $\tau$ , respectivamente, del ejemplo 1.

### 3.1.4. Simulación

Dado que para el ejemplo 1 las distribuciones iniciales en 3.5 fueron tomadas de un artículo, y dado que para los siguientes ejemplos, el investigador desconoce los valores que pueden tomar los parámetros de los modelos farmacocinéticos, se llevó a cabo una simulación en la cual se generaron el doble de datos que en el ejemplo anterior. Es decir, se creó una muestra que pretende representar la concentración del medicamento en 18 pacientes; el número de mediciones para cada uno y el periodo de tiempo en el que se tomaron éstas es el mismo que antes: 11 mediciones en un periodo de 25 horas. Por otro lado, se tomaron diferentes parámetros de las distribuciones y las modas de las muestras obtenidas -las cuales resultaron unimodales-, se compararon con los valores verdaderos.

Los datos originales y los obtenidos con la metodología desarrollada en las secciones anteriores se pueden ver en el cuadro 3.2.

Al ver estos resultados, se puede observar que al menos con el doble de la cantidad de datos, es preciso tener una idea de los valores que toman los parámetros a estimar, porque de lo contrario, las aproximaciones obtenidas pueden ser engañosas. En el caso del escenario 1, por ejemplo, la tasa de absorción es tan grande que prácticamente se tendría un perfil de eliminación,

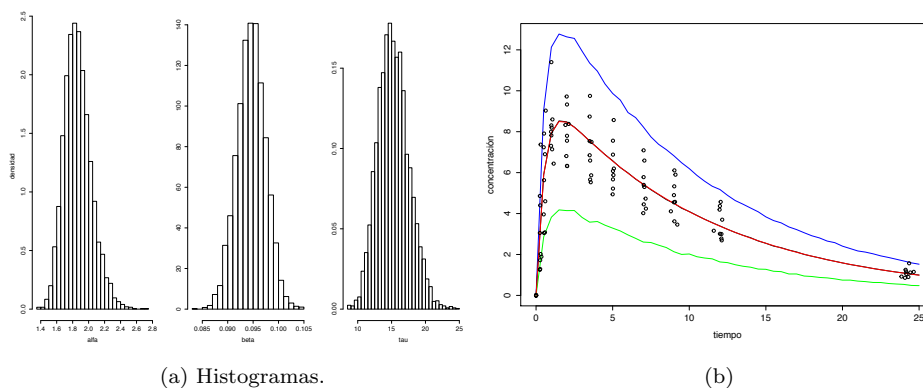


Figura 3.6: En la figura 3.6a se observan los histogramas obtenidos de las muestras de  $\alpha$ ,  $\beta$  y  $\tau$ . En 3.6b, se tiene en rojo la gráfica de la función que representa la concentración del medicamento evaluada en las modas de cada una de las muestras de los parámetros estimados; en negro se observa la distribución predictiva final entre bandas con una probabilidad de 0.95. Los círculos representan las mediciones de concentración de teofilina del experimento.

	$(a_\alpha, b_\alpha)$	$(a_\beta, b_\beta)$	$(a_\tau, b_\tau)$	$\alpha$	$\beta$	$\tau$
Valores verdaderos	-	-	-	2	0.1	16
Escenario 1	(50,100)	(2,100)	(2,160)	4981.975	0.09972562	10.78511
Escenario 2	(2,2)	(2,2)	(2,2)	2.1385860	0.1008269	16.1295097

Cuadro 3.2: Diferentes parámetros de las distribuciones en 3.5 y valores obtenidos con la metodología desarrollada en esta sección.

lo que dista de la realidad como se puede ver en la figura 3.7.

## 3.2. Segundo ejemplo

En la tesis de licenciatura de Laura Mondragón (2014), realizada en el Instituto Nacional de Cancerología, se llevó a cabo un experimento preclínico; en éste, se emplearon ratones atímicos calvos<sup>5</sup> a los cuales se les trasplantaron células HeLa -derivadas de cáncer cérvicouterino humano- y se les administró un fármaco utilizado en quimioterapia llamado cisplatino, el cual se administró de dos formas:

- a) libre (convencional)

<sup>5</sup>Debido a que este tipo de ratones carecen de timo, su organismo se encuentra inmunodeprimido; esto provoca que al ser trasplantados con células de otra especie, su organismo no las rechace.

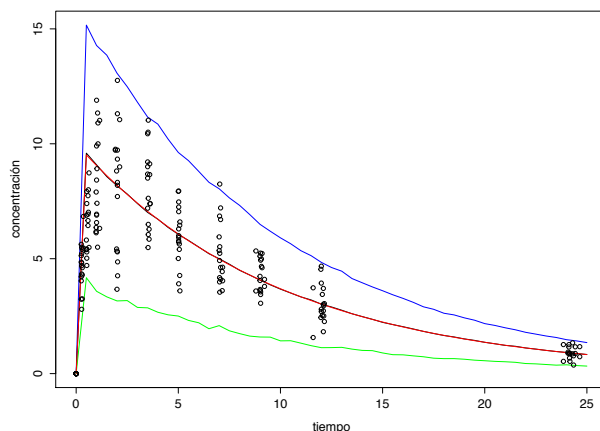


Figura 3.7: . En la gráfica se observa en rojo, la gráfica de la función que representa la concentración del medicamento evaluada en las modas de cada una de las muestras de los parámetros estimados para el caso del escenario 1; en negro se observa la distribución predictiva final entre bandas con una probabilidad de 0.95. Lo cierto es que se puede observar que con las distribuciones iniciales elegidas, el ajuste obtenido no es adecuado: no se trata de un modelo de eliminación únicamente.

### b) encapsulado en liposomas

Para determinar la concentración del fármaco en el tumor, se tomaron 5 muestras de tejido<sup>6</sup> en un periodo de 24 horas.

La relevancia de este experimento consiste en que el tratamiento correspondiente a administrar el cisplatino encapsulado, se pueda posteriormente administrar en humanos. La razón es que, como es sabido, la quimioterapia es altamente tóxica para el organismo, además de que se calcula que únicamente el 1% de los fármacos administrados llegan al tumor cancerígeno[16]. En la misma fuente puede consultarse que la administración del mismo medicamento encapsulado en liposomas reduce aún más el tamaño del tumor que en el caso de la quimioterapia convencional; permanece más tiempo en el organismo, llegando mayor cantidad del fármaco al tumor (por la naturaleza del mismo), y el daño renal y hepático es casi nulo. Los datos de este experimento pueden verse representados en la figura 3.8.

<sup>6</sup>Como se mencionó en la sección 1.1, es muy difícil tomar muestras de los tejidos. Lo cierto es que se usaron 10 ratones de aproximadamente la misma edad y el mismo peso, a los cuales se les sacrificó en el momento en el que se tomó cada medición.

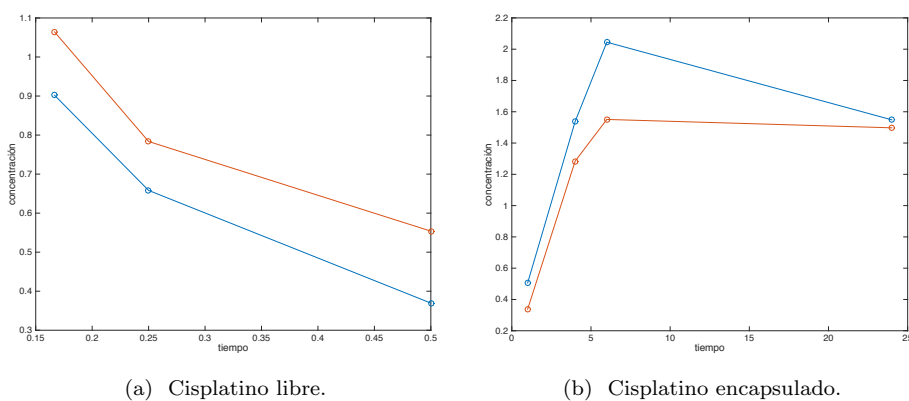


Figura 3.8: Concentraciones de cisplatino para los dos tipos de tratamiento.

### 3.2.1. Metodología

#### Modelo PK por compartimentos

##### Cisplatino encapsulado

Los perfiles de concentración se modelaron utilizando dos compartimentos, con eliminación de primer orden y tasas de distribución iguales entre los mismos. La representación esquemática del modelo se encuentra en la figura 3.9.

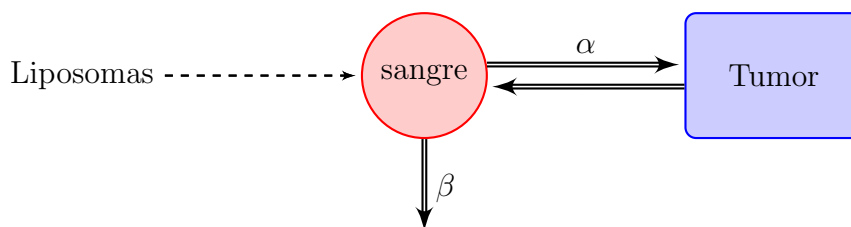


Figura 3.9: Dos compartimentos con eliminación de primer orden.

El modelo farmacocinético con condiciones iniciales es en este caso

$$\begin{aligned} x_1'(t) &= -(\alpha + \beta)x_1(t) - \alpha x_2(t), & x_1(0) &= D, \\ x_2'(t) &= \alpha x_1(t) - \alpha x_2(t), & x_2(0) &= 0, \end{aligned} \quad (3.11)$$

y la concentración en el tumor está dada por

$$C(t) = \frac{x_2(t)}{V} = \frac{c\alpha}{\sqrt{(2\alpha + \beta)^2 - 4\alpha\beta}}(e^{\lambda_1 t} - e^{\lambda_2 t}), \quad c = \frac{DF}{V}, \quad (3.12)$$

donde

$$\lambda_1, \lambda_2 = \frac{1}{2} \left( -(2\alpha + \beta) \pm \sqrt{(2\alpha + \beta)^2 - 4\alpha\beta} \right).$$

### Cisplatino libre

Los perfiles de concentración se modelaron con un compartimento de depósito coincidente con el compartimento central y eliminación de primer orden. El esquema del modelo se puede ver a continuación.

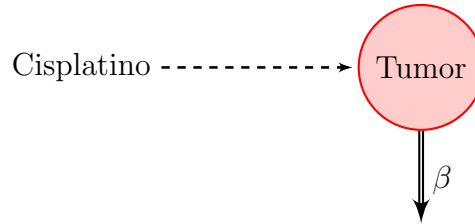


Figura 3.10: Eliminación de primer orden solo con compartimento central.

El modelo farmacocinético con condiciones iniciales es

$$x'(t) = -\beta x(t), \quad x(0) = D, \quad (3.13)$$

y la concentración está dada por

$$C(t) = \frac{x(t)}{V} = ce^{-\beta t}, \quad c = \frac{DF}{V}. \quad (3.14)$$

### Modelo estadístico

Para ambos casos -de hecho para toda esta obra-, el modelo estadístico está representado por la ecuación 3.3, y es exactamente el mismo descrito en la sección 3.1.1. Pues bien, para el caso de cisplatino encapsulado, el compartimento de interés es el del tumor  $x_2$ ; en el caso del cisplatino libre, el espacio de estados está formado únicamente por un elemento, ya que solo existe el compartimento central. Para ambos casos, el número de unidades experimentales es  $N = 2$ ; lo cierto es que el número de mediciones es  $n = 3$  para el caso del cisplatino libre, y de  $n = 4$  para el caso del cisplatino liposomal.

## Inferencia estadística

### Cisplatino encapsulado

Los parámetros del modelo estadístico son, igual que en el ejemplo 1,  $\theta = (\alpha, \beta)$  y  $\tau$ . Por lo tanto, la verosimilitud está dada por la siguiente ecuación:

$$p(\mathbf{y}|\alpha, \beta, \tau) = \prod_{i=1}^N \prod_{k=1}^n \frac{\sqrt{\tau}}{\sqrt{2\pi} x_2(t_{ik}|\alpha, \beta)} e^{-\frac{\tau}{2} \left[ \frac{y_{ik} - x_2(t_{ik}|\alpha, \beta)}{x_2(t_{ik}|\alpha, \beta)} \right]^2}. \quad (3.15)$$

Como se ha visto antes, no se tiene información de los valores que puedan tomar las tasas, salvo que los valores son positivos. Lo cierto es que a través de las ecuaciones 1.2 y 1.7 que describen respectivamente la semivida de eliminación y la de absorción, es posible aproximar los valores que pueden tomar las tasas. Es decir, dado que el periodo de observación está entre  $[0, 24]$ , las tasas toman valores, en número redondos, entre  $[0, 3]$  (basta despejar cualquiera de las dos tasas en cualquiera de las dos ecuaciones que representan la semivida). Por esta razón, en lo que resta de este trabajo, la distribución de propuesta para cada tasa se tomó como una gamma(2,2) independiente. Su gráfica se representa a continuación.

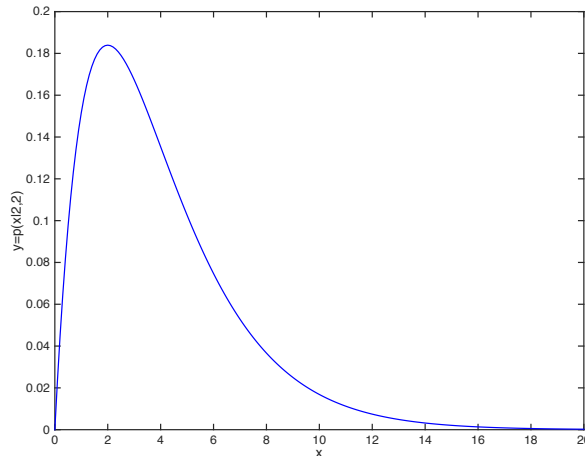


Figura 3.11: Gráfica de la función de densidad  $P(x; a, b) = \frac{1}{\Gamma(a)b^a} x^{a-1} e^{-\frac{x}{b}} \mathbb{I}_{(0, \infty)}(x)$  con  $(a, b) = (2, 2)$ .

En cuanto al parámetro de calibración, se siguió tomando una gamma no informativa, esta vez con parámetros  $(2, 100)$  en éste y en el ejemplo posterior.

La distribución a posteriori queda de la siguiente manera

$$p(\alpha, \beta, \tau | \mathbf{y}) \propto \left[ \prod_{i=1}^N \prod_{k=1}^n x_2(t_{ik} | \alpha, \beta) \right]^{-1} \tau^{\frac{Nn}{2}} e^{-\frac{1}{2}\tau \sum_{i=1}^N \sum_{k=1}^n \left[ \frac{y_{ik} - x_2(t_{ik} | \alpha, \beta)}{x_2(t_{ik} | \alpha, \beta)} \right]^2} \cdot P(\alpha; 2, 2)P(\beta; 2, 2)P(\tau; 2, 100). \quad (3.16)$$

Al igual que en el ejemplo 1, se marginalizó  $p(\alpha, \beta, \tau | \mathbf{y})$  integrando con respecto a  $\tau$  y se usó el algoritmo de Metropolis para obtener una muestra de

$$p(\alpha, \beta | \mathbf{y}) \propto [P(\alpha, \beta)]^{-1} \alpha^{a_\alpha - 1} \beta^{a_\beta - 1} e^{-\frac{\alpha}{b_\alpha} - \frac{\beta}{b_\beta}} \cdot B_\tau(\alpha, \beta)^{A_\tau(\alpha, \beta)}, \quad (3.17)$$

donde

$$P(\alpha, \beta) = \prod_{i=1}^N \prod_{k=1}^n \frac{c\alpha}{\sqrt{(2\alpha + \beta)^2 - 4\alpha\beta}} (e^{\lambda_1 t} - e^{\lambda_2 t}),$$

$$B_\tau(\alpha, \beta) = \left( \frac{1}{2} \sum_{i=1}^N \sum_{k=1}^n \left[ \frac{y_{ik} - \frac{c\alpha}{\sqrt{(2\alpha + \beta)^2 - 4\alpha\beta}} (e^{\lambda_1 t} - e^{\lambda_2 t})}{\frac{c\alpha}{\sqrt{(2\alpha + \beta)^2 - 4\alpha\beta}} (e^{\lambda_1 t} - e^{\lambda_2 t})} \right]^2 + \frac{1}{b_\tau} \right)^{-1},$$

y

$$A_\tau(\alpha, \beta) = \frac{Nn}{2} + a_\tau,$$

para posteriormente obtener una muestra de

$$\tau | \alpha, \beta \sim \text{Gamma}(A_\tau(\alpha, \beta), B_\tau(\alpha, \beta)). \quad (3.18)$$

### Cisplatino libre

Los parámetros del modelo estadístico son ahora  $\theta = \beta$  y  $\tau$ . La verosimilitud en este caso es

$$p(\mathbf{y} | \beta, \tau) = \prod_{i=1}^N \prod_{k=1}^n \frac{\sqrt{\tau}}{\sqrt{2\pi} x(t_{ik} | \beta)} e^{-\frac{\tau}{2} \left[ \frac{y_{ik} - x(t_{ik} | \beta)}{x(t_{ik} | \beta)} \right]^2}, \quad (3.19)$$

y la distribución final es

$$p(\beta, \tau | \mathbf{y}) \propto \tau^{\frac{Nn}{2}} \left[ \prod_{i=1}^N \prod_{k=1}^n c e^{-\beta t} \right]^{-1} e^{\tau S(\beta)} \beta^{a_\beta - 1} \tau^{a_\tau - 1} e^{-\frac{\beta}{b_\beta} - \frac{\tau}{b_\tau}}, \quad (3.20)$$

donde

$$S(\beta) = \left( \frac{1}{2} \sum_{i=1}^N \sum_{k=1}^n \left[ \frac{y_{ik} - ce^{-\beta t}}{ce^{-\beta t}} \right]^2 \right).$$

Se empleó el algoritmo de Metropolis para obtener una muestra de esta distribución.

### 3.2.2. Implementación

#### Cisplatino encapsulado

La muestra de la distribución final de 3.17, se obtuvo exactamente como se describe en la sección 3.1.2. Después de haber obtenido esta muestra, la misma se usó para obtener la muestra de la distribución 3.18 que depende de  $(\alpha, \beta)$  por medio del muestreo de Gibbs. Esto se llevó a cabo con la función *rgamma* de R.

#### Cisplatino libre

Para obtener la muestra de la distribución final de 3.20, nuevamente se aplicó el procedimiento descrito en la sección 3.1.2.

### 3.2.3. Resultados

#### Cisplatino encapsulado

Se obtuvo una muestra de 5000 elementos, para la cual se tomó un periodo de calentamiento igual a 800,000, y a partir de entonces se tomaron los elementos de la muestra dejando pasar 45 entre uno y otro<sup>7</sup>.

#### Cisplatino libre

Se obtuvo una muestra de 5000 elementos, para la cual se tomó un periodo de calentamiento igual a 600,000, y a partir de entonces se tomaron los elementos de la muestra dejando pasar 60 entre uno y otro, como se ha dicho antes, para disminuir la correlación entre ellos<sup>8</sup>.

Para ambos casos, la convergencia y la autocorrelación de la muestra obtenida, además de los histogramas para cada uno de los parámetros y la distribución predictiva final se muestran respectivamente en las figuras 3.12, 3.13 y 3.14.

---

<sup>7</sup>Se eligió una constante  $\frac{FD}{V} = 9.8$  calculada con la función de mínimos cuadrados no lineales de MATLAB

<sup>8</sup>Se eligió una constante  $\frac{FD}{V} = 1.4$  calculada con la función de mínimos cuadrados no lineales de MATLAB



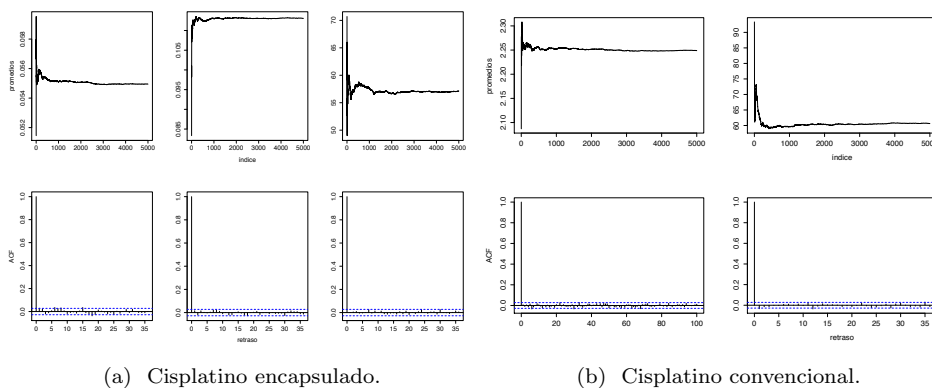


Figura 3.12: Gráficas de los promedios acumulados y de la función de autocorrelación de las muestras obtenidas para los parámetros  $\alpha$ ,  $\beta$  y  $\tau$  en el caso de 3.12a, y de  $\beta$  y  $\tau$  en el de 3.12b.

### 3.3. Tercer ejemplo

Los datos para este ejemplo son de la tesis de licenciatura de Kevin López (2016) de otro experimento también llevado a cabo en el Instituto Nacional de Cancerología. En este experimento se utilizaron seis ratas Wistar sanas. Por vía intravenosa, a la mitad de ellas se les administró el material radiactivo conocido como tecnecio, mientras que al resto se les administraron liposomas radiomarcados<sup>9</sup> con el mismo material radiactivo. Los liposomas son del mismo tipo que se usaron en el experimento del ejemplo 2 para encapsular el cisplatino. El propósito del experimento, es decir, el de radio marcar la formulación de cisplatino liposomal con tecnecio, es determinar la farmacocinética plasmática y la biodistribución en las ratas.

Para determinar la concentración de tecnecio en la sangre, se tomaron 10 muestras a distintos tiempos en un periodo de 24 horas. El investigador desea saber si hay una diferencia entre la farmacocinética descrita por los diferentes esquemas aplicados. Los perfiles de concentración se presentan en la gráfica 3.15.

<sup>9</sup>El radiomarcaje consiste en incorporar un radionúclido -una especie nuclear, que tiene un valor específico para el número de protones y neutrones, radiactiva- en una molécula.

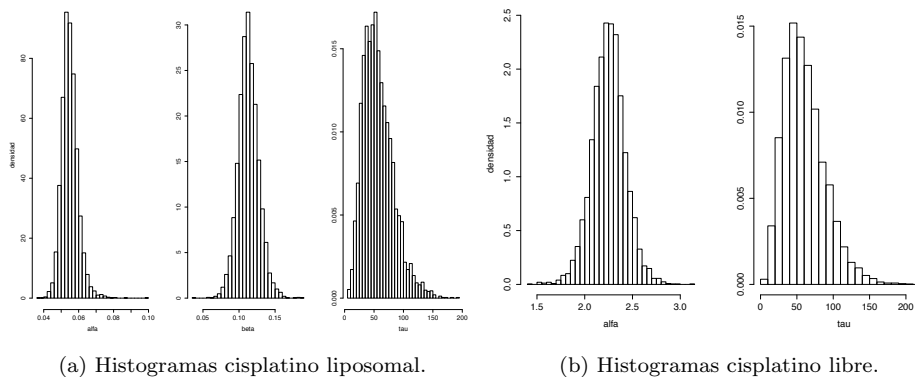


Figura 3.13: Histogramas respectivos de los parámetros para cada uno de los tratamientos. Para el caso del cisplatino liposomal, se tienen tres distribuciones unimodales con modas respectivas de  $\alpha$ ,  $\beta$  y  $\tau$ , 0.05494371, 0.1130818 y 57.13327. Para el caso del cisplatino convencional, también se tienen distribuciones unimodales pero con modas respectivas de  $\alpha$  y  $\tau$  iguales a 2.248947 y 60.65841.

### 3.3.1. Metodología

#### Modelo PK por compartimentos

Esta vez, los perfiles de concentración de ambos tratamientos se modelaron con un mismo modelo: de un compartimento de depósito y **eliminación de segundo orden**. De hecho, esquemáticamente, este modelo es idéntico al del cisplatino libre, solo que en este caso, el compartimento de depósito -que en este caso también es el central-, solo está formado por el torrente sanguíneo (se puede asegurar en este caso que la biodisponibilidad es  $F=1$ ). El esquema del modelo es el que se presenta en la figura 3.16.

El modelo farmacocinético con condiciones iniciales es

$$x'(t) = -\beta x^2(t), \quad x(0) = D, \quad (3.21)$$

y la concentración está dada por

$$C(t) = \frac{x(t)}{V} = \frac{D}{V(D\beta x(t) + 1)}. \quad (3.22)$$

#### Modelo estadístico

Como se mencionó anteriormente, el modelo estadístico no cambia. Además, como se verá a continuación, este modelo comparte diferentes características

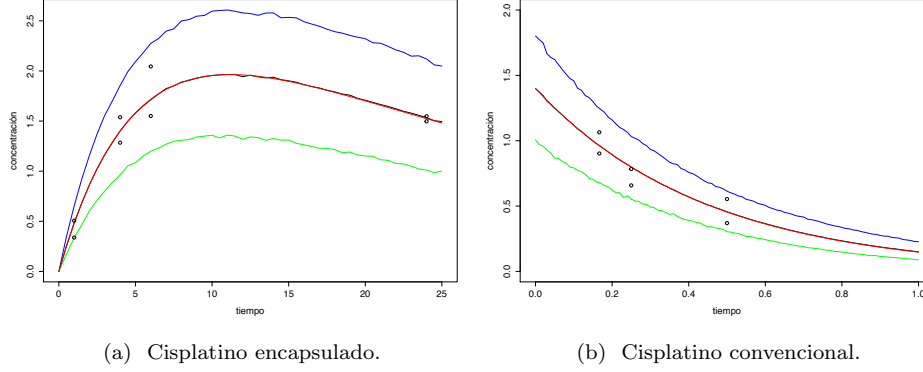


Figura 3.14: Para cada uno de los tratamientos se tiene en rojo la gráfica de la función que representa la concentración del medicamento evaluada en las modas de cada una de las muestras de los parámetros estimados; en negro se observa la distribución predictiva final entre bandas con una probabilidad de 0.95. Los círculos representan las mediciones de concentración de cisplatino del experimento.

con el modelo farmacocinético del cisplatino libre visto en la sección anterior. Ahora bien, para los dos esquemas de este experimento con tecnecio, el número de unidades experimentales es  $N = 2^{10}$  y el número de mediciones es  $n = 10$ .

### Inferencia estadística

Los parámetros del modelo estadístico son ahora  $\theta = \beta$  y  $\tau$ . La verosimilitud en este caso es del mismo tipo que en 3.19, es decir

$$p(\mathbf{y}|\beta, \tau) = \prod_{i=1}^N \prod_{k=1}^n \frac{\sqrt{\tau}}{\sqrt{2\pi} x(t_{ik}|\beta)} e^{-\frac{\tau}{2} \left[ \frac{y_{ik} - x(t_{ik}|\beta)}{x(t_{ik}|\beta)} \right]^2}; \quad (3.23)$$

la distribución final es

$$p(\beta, \tau|\mathbf{y}) \propto \tau^{\frac{Nn}{2}} \left[ \prod_{i=1}^N \prod_{k=1}^n \frac{D}{V(D\beta x(t) + 1)} \right]^{-1} e^{\tau S(\beta)} \beta^{a\beta-1} \tau^{a\tau-1} e^{-\frac{\beta}{b\beta} - \frac{\tau}{b\tau}}, \quad (3.24)$$

<sup>10</sup>Aunque en el experimento se trabajó con tres ratas, existen discrepancias en los datos, por lo que el investigador desechó las mediciones correspondientes a un individuo.

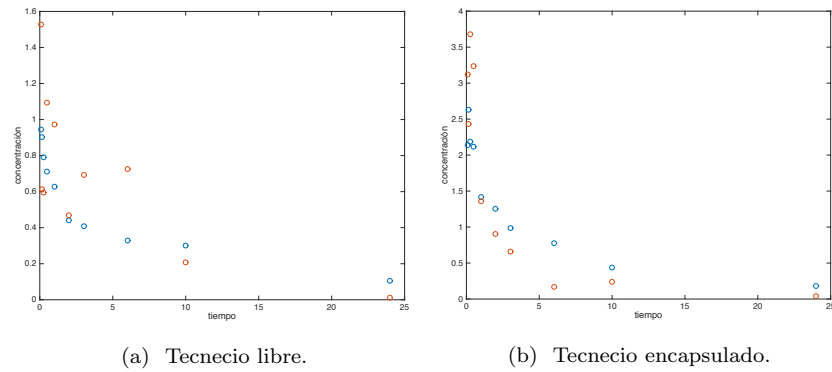


Figura 3.15: Concentraciones de tecnecio para los dos tipos de tratamiento.

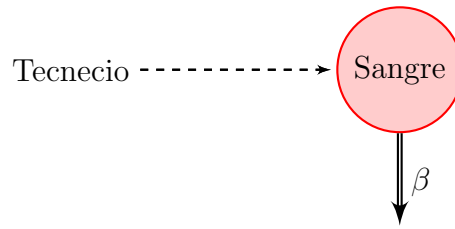


Figura 3.16: Eliminación de segundo orden con compartimento central únicamente.

donde

$$S(\beta) = \left( \frac{1}{2} \sum_{i=1}^N \sum_{k=1}^n \left[ \frac{y_{ik} - \frac{D}{V(D\beta x(t)+1)}}{\frac{D}{V(D\beta x(t)+1)}} \right]^2 \right).$$

### 3.3.2. Implementación

Para obtener la muestra de la distribución final de 3.24, nuevamente se aplicó el algoritmo de Metropolis como se describió en la sección 3.1.2.

### 3.3.3. Resultados

Para el caso del tecnecio libre, se obtuvo una muestra de 5000 elementos, para la cual se tomó un periodo de calentamiento igual a 800,000, y a partir de entonces se tomaron los elementos de la muestra dejando pasar 45 entre

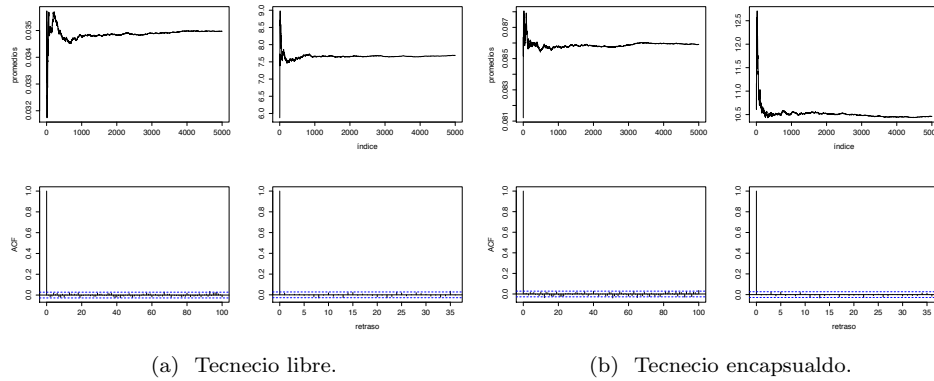


Figura 3.17: Gráficas de los promedios acumulados y de la función de autocorrelación de las muestras obtenidas para los parámetros  $\beta$  y  $\tau$ , respectivamente. Para el caso del tecnecio libre, se tienen dos distribuciones unimodales con modas respectivas de la tasa de eliminación y  $\tau$ , 0.03497279 y 7.687558. Para el caso del tecnecio encapsulado, también se tienen distribuciones unimodales pero con modas respectivas de la tasa de eliminación y  $\tau$  iguales a 2.80866 y 10.46297.

uno y otro<sup>11</sup>. La convergencia y la autocorrelación de la muestra obtenida, además de los histogramas para cada uno de los parámetros y la distribución predictiva final se muestran respectivamente en las figuras 3.17, 3.18 y 3.19.

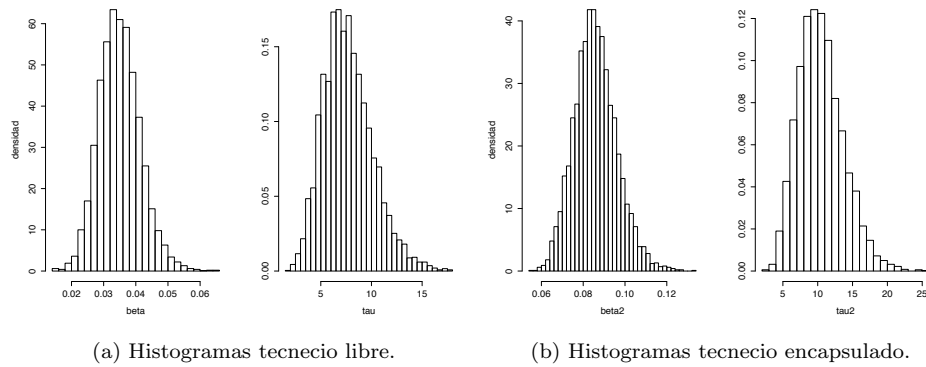


Figura 3.18: Histogramas respectivos de los parámetros  $\beta$  y  $\tau$  para cada uno de los tratamientos.

<sup>11</sup>Se eligieron constantes  $DV = 62.6166$  calculadas con la función de mínimos cuadrados no lineales de MATLAB

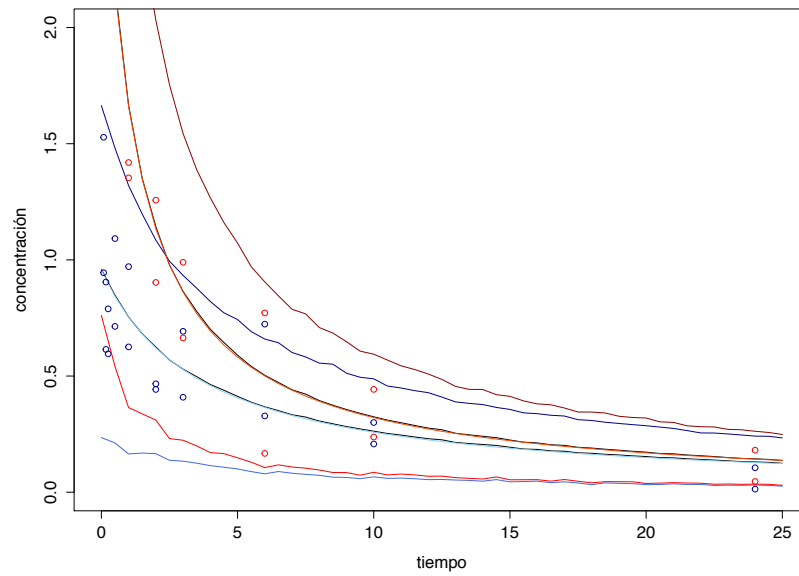


Figura 3.19: Gráficas sobrepuestas para los tratamientos en los cuales se aplicó tecneio libre (en azul) y encapsulado (en rojo). Para cada uno, se tiene en el centro, en azul o en rojo, la gráfica de la función que representa la concentración del medicamento evaluada en las modas de cada una de las muestras de los parámetros estimados; en negro se observa la distribución predictiva final entre bandas con una probabilidad de 0.95. Los círculos azules y rojos representan respectivamente las mediciones de concentración de tecneio libre y tecneio encapsulado.

# Conclusiones

En los ejemplos que se implementaron en el capítulo anterior se hizo evidente el alcance de la metodología matemática utilizada. Es decir, los modelos farmacocinéticos y los métodos de Monte Carlo vía Cadenas de Markov pueden ser extremadamente útiles para describir el comportamiento de la concentración de un medicamento a través del tiempo para una población, a pesar de la variabilidad que pueda existir entre los individuos. Además, si este trabajo se llegara a complementar con modelos farmacodinámicos, los modelos en conjunto, a través de modificar las dosis y simular diferentes tiempos en la dosificación, podrían ayudar a entender el efecto que los fármacos tienen en las enfermedades y los mecanismos subyacentes de las mismas.

En particular se pudo observar que se pueden usar varios tipos de modelos farmacocinéticos dependiendo del tratamiento que se estudie, y si éstos llegaran a ser más complejos -si las ecuaciones diferenciales ya no fueran lineales, por ejemplo- se podrían aplicar métodos numéricos para encontrar la solución que representan la concentración.

El propósito de haber usado estadística bayesiana es el de que, al ser una Teoría de Inferencia y no un conjunto de métodos, no presenta ciertos inconvenientes. En particular, se hizo patente que el algoritmo de Metropolis, el más sencillo de los métodos MCMC, es una herramienta muy poderosa para obtener muestras de distribuciones multidimensionales complejas.

Lo cierto es que este trabajo muestra que este tipo de modelos deberían estar basados en fundamentos biológicos y principios farmacológicos, para así poder entender mejor el problema en cuestión y describir la realidad de manera más efectiva. Esto implica que el desarrollo de nuevos procedimientos debería llevarse a cabo de manera interdisciplinaria.





# Apéndice A

## Modelos estocásticos de ecuaciones diferenciales ordinarias

En los modelos con ecuaciones diferenciales deterministas, dadas las mismas condiciones iniciales, exactamente la misma trayectoria se observa siempre. Pero esta forma de ver el mundo no se aplica a la dinámica de muchos procesos reales. Por ejemplo, en el caso de la epidemiología, si fuera posible volver a ocasionar cierta epidemia mundial, no se esperaría que exactamente la misma gente se infectara en precisamente los mismos tiempos. La razón es que en fenómenos como éste, existe un elemento aleatorio.

Dado que en este capítulo se tomará como base un modelo epidemiológico, se puede decir en general, que el papel de la aleatoriedad será más importante cuando el número de individuos infecciosos sea relativamente pequeño, lo que es factible en diversos casos. En circunstancias como éstas, es muy importante que se tome en cuenta la estocasticidad y que ésta se incorpore en los modelos.

Se detallarán tres métodos distintos de aproximar el elemento aleatorio en la transmisión de una enfermedad y en la recuperación: introduciendo la aleatoriedad en las variables de la población; por medio de la variación de parámetros aleatorios, y mediante el modelado explícito de eventos aleatorios a nivel individual. Todos estos ejemplos tienen dos elementos en común. En primer lugar, predicen diferentes resultados con las mismas condiciones iniciales, y por lo tanto, se requieren muchas simulaciones para determinar el comportamiento esperado. En segundo lugar, todos requieren el uso de una rutina que genere números aleatorios.

Para hacer una comparación clara y razonable entre los diferentes métodos que consideran la aleatoriedad, se usará como base el siguiente modelo epidemiológico.

## A.1. El modelo SIR determinista

En Epidemiología, la propagación de las enfermedades infecciosas agudas, es decir, las no crónicas como la influenza o la varicela, se puede modelar con un sistema de ecuaciones diferenciales no lineales. El modelo más sencillo es el **modelo SIR**. Para tal efecto, se considera primero que se tiene una población de tamaño constante. Luego, si un grupo pequeño de individuos infectados se introduce en una población mucho más grande, se desearía poder describir cómo se propaga la infección dentro de la población a través del tiempo. Es evidente que el fenómeno depende de muchas circunstancias, empezando por el tipo de infección en consideración, pero se pueden hacer suposiciones generales que sean razonables, para tener un modelo sencillo que también sea eficiente.

Considérese entonces una enfermedad en la cual, después de haber estado infectado, se alcance la inmunidad. Esto significa que ya sea que un individuo viva o muera después de pasar por el proceso infeccioso, se le considera inmune. Por lo tanto, las muertes debidas a la infección están consideradas en el modelo. Por otro lado, se considera que la población total se puede dividir en tres clases diferentes: los **susceptibles**  $S$ , los **infecciosos**  $I$  y los **removidos**  $R$ . Los primeros son los que pueden contraer la infección; los segundos tiene la enfermedad y la pueden transmitir; los últimos son aquéllos que, ya sea que hayan tenido la enfermedad, o se recuperaron, o están aislados mientras se recuperan, o son inmunes. Las transiciones entre estos tres estados se suelen representar como

$$S \rightarrow I \rightarrow R.$$

Las suposiciones que se hacen de la transmisión de la infección y el periodo de incubación son decisivas en cualquier modelo, y se observan en los parámetros y en los términos de las ecuaciones. Siendo  $S(t)$ ,  $I(t)$  y  $R(t)$  el número de individuos en cada clase al tiempo  $t$ , se hacen las siguientes suposiciones:

- i*) La **tasa de incremento de los individuos infecciosos** es proporcional al número de infecciosos y susceptibles. Esto se puede representar

como  $\beta SI$  donde  $\beta > 0$ . El **decremento de susceptibles** ocurre a la misma tasa.

- ii) La **tasa de remoción** de la clase de infecciosos a removidos es proporcional al número de infecciosos, es decir,  $\gamma I$  con  $\gamma > 0$ ;  $\frac{1}{\gamma}$  es una medida del tiempo que un individuo permanece en el estado infeccioso.
- iii) Se considera que el **tiempo de incubación** es tan corto que es insignificante. Por lo tanto, un individuo susceptible que contrae la enfermedad pasa a ser infeccioso de manera inmediata.

Finalmente, se considera que las clases descritas anteriormente se encuentran mezcladas uniformemente, lo que significa que dado cualquier par de individuos, cualquiera de ellos tiene la misma probabilidad de tener contacto con el otro<sup>1</sup>. El modelo resultante es el siguiente:

$$\frac{dS}{dt} = -\beta SI, \quad (\text{A.1})$$

$$\frac{dI}{dt} = \beta SI - \gamma I, \quad (\text{A.2})$$

$$\frac{dR}{dt} = \gamma I. \quad (\text{A.3})$$

La suposición de que la población es constante se vuelve evidente al sumar las ecuaciones (A.1-A.3):

$$\frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0.^2 \quad (\text{A.4})$$

Estas ecuaciones tienen como condiciones iniciales que  $S(0) > 0$ ,  $I(0) > 0$  y  $R(0) = 0$ , ya que para que una infección se empiece a propagar, es preciso que exista al menos una persona que la pueda transmitir y otra que se pueda infectar.

### A.1.1. Modelo SIR con demografía

En el modelo SIR descrito anteriormente, el periodo de tiempo en el cual se estudió la dinámica de la infección, se consideró tan corto que no se

<sup>1</sup>Esta consideración podría no cumplirse; considérese, por ejemplo, el caso de las enfermedades de transmisión sexual.

<sup>2</sup>Esta ecuación implica que la suma  $S(t) + I(t) + R(t)$  es constante.

tomaron en cuenta los nacimientos ni las muertes que no fueran causadas por la enfermedad. Pero si ahora se deseara estudiar la persistencia de la epidemia a largo plazo y la dinámica propia de la enfermedad infecciosa en cuestión, el papel de la demografía ciertamente sería diferente. La manera más sencilla de incorporar este cambio en el modelo SIR es la de suponer que existe un **periodo de vida medio** para los huéspedes de  $\frac{1}{\mu}$  años. De este modo, la **tasa natural de mortalidad** de cualquier clase epidemiológica de individuos es  $\mu$ . Esta tasa es independiente de la enfermedad y no intenta reflejar la capacidad del agente patógeno de causar la muerte. El **modelo SIR con demografía** en el que la tasa de nacimientos  $\nu$  es igual a  $\mu$  queda entonces como sigue:

$$\frac{dS}{dt} = \mu - \beta SI - \mu S, \quad (\text{A.5})$$

$$\frac{dI}{dt} = \beta SI - \gamma I - \mu I, \quad (\text{A.6})$$

$$\frac{dR}{dt} = \gamma I - \mu R. \quad (\text{A.7})$$

## A.2. Diferentes modelos SIR estocásticos

Se tomará como ejemplo al modelo SIR con demografía. Este modelo, en el esquema determinista, siempre converge a un equilibrio (de manera oscilatoria). Se considerará también que tanto los parámetros como el tamaño de la población son constantes. Para los modelos estocásticos, es necesario que se consideren, tanto al comportamiento a largo plazo y a la dinámica transitoria a corto plazo, como a la media y a la varianza del número de casos infectados. Puesto que los modelos estocásticos tienen una relación estrecha con los modelos con valores enteros o con los modelos basados en individuos, generalmente se trabaja con el número *entero* de **susceptibles** ( $X$ ), **infectados** ( $Y$ ) y **recuperados** ( $Z$ ), a diferencia de usar la proporción como en las secciones previas. Por esta razón se usan letras diferentes en uno y otro caso.

Existen cinco características que distinguen a los modelos estocásticos de los deterministas:

1. **Variabilidad entre simulaciones.** El elemento más obvio de cualquier modelo aleatorio es que simulaciones diferentes dan lugar a dife-

rentes resultados. Esto implica que aunque las propiedades estadísticas se pueden predecir con exactitud, en general, es imposible predeterminar exactamente la prevalencia de la enfermedad en un tiempo futuro.

2. **Varianzas y covarianzas.** Las perturbaciones constantes causadas por la naturaleza aleatoria de las ecuaciones, origina una variación en la prevalencia de la enfermedad y el número de susceptibles. Además, la interacción entre la dinámica determinista y la estocástica, generalmente provoca que exista una covarianza negativa entre el número de infecciosos y susceptibles, lo que a su vez puede ocasionar que los niveles medios de la población se desvíen del equilibrio determinista.
3. **Incrementos temporales.** Las perturbaciones estocásticas alejadas de la solución de equilibrio, generalmente van en contra de la conducta subyacente de la dinámica determinista que en general es convergente. Cuando se encuentran lejos del punto de equilibrio, esta conducta usualmente domina, de modo que el modelo actúa en gran medida como las ecuaciones deterministas. Por lo tanto, la dinámica de los modelos estocásticos, se puede pensar como el resultado de perturbaciones aleatorias que se encuentran lejos del atractor determinista, o de tipo temporal que regresan a él.
4. **Resonancia estocástica.** Como la mayoría de los modelos tipo SIR alcanzan el equilibrio en una serie de epidemias que pierden fuerza, el aumento de la dinámica temporal de los modelos estocásticos origina oscilaciones cercanas a la frecuencia natural. Por lo tanto, la estocasticidad puede provocar oscilaciones epidémicas cerca de la prevalencia normal.
5. **Extinciones.** Los modelos aleatorios de valores enteros pueden originar extinciones. En poblaciones cerradas, las fluctuaciones aleatorias tarde o temprano originarán la extinción de la enfermedad, independientemente del tamaño de la población. La persistencia a largo plazo requiere de patógenos que provengan de una fuente externa. Pueden ocurrir extinciones similares en las primeras etapas de la infección.

### A.2.1. Ruido en las observaciones

La forma más simple de ruido que puede incluirse en el modelado de enfermedades infecciosas es el ruido en las observaciones. Al abordar el problema de esta forma, la dinámica de la epidemia subyacente conserva las ecuaciones diferenciales deterministas, pero se asume que existe algo de incertidumbre en los datos registrados. En general, la información de las enfermedades que tienen que reportarse obligatoriamente, como es el caso del sarampión, la fiebre aftosa o la fiebre porcina clásica, es muy detallada y precisa. Sin embargo, las enfermedades infecciosas que son altamente asintomáticas, como la meningococemia o las enfermedades que padece la fauna silvestre, pueden estar sujetas a una mayor incertidumbre a la hora de que se reportan. Para una enfermedad infecciosa en específico, la estocasticidad debida al ruido en las observaciones puede cuantificarse fácilmente, tanto en el registro incompleto de casos verdaderos, o mediante diagnósticos erróneos.

Si  $R$  es una variable aleatoria que representa el número de infecciosos reportados y  $I$  es el número de individuos infecciosos de hecho, entonces

$$\mathbb{P}(R) = \mathbb{P}(R | I)\mathbb{P}(I) + \mathbb{P}(R | I^c)\mathbb{P}(I^c).$$

Aunque evidentemente estas consideraciones son importantes, cuando se interpretan datos de casos reportados, éstos no interactúan con la dinámica subyacente de la epidemia. El verdadero número de casos siempre tiende al punto fijo determinista y las observaciones están determinadas enteramente por las probabilidades  $\mathbb{P}(R | I)$ ,  $\mathbb{P}(I)$ ,  $\mathbb{P}(R | I^c)$  y  $\mathbb{P}(I^c)$ . En las epidemias recurrentes, inclusive, se ha visto que el error en las observaciones afecta la amplitud de los datos, mas no la periodicidad. El ruido en las observaciones no tiene impacto en la dinámica epidemiológica, sino en los datos reportados.

### A.2.2. Ruido del proceso

Un modo más intuitivo y esencialmente diferente de considerar ruido, consiste en incorporarlo directamente en las ecuaciones deterministas. La dinámica en cada punto en el tiempo como tal, está sujeta a alguna variabilidad aleatoria, y esta variabilidad se propaga en tiempos posteriores por las ecuaciones subyacentes. Aquí se estudia la interacción entre las fuerzas deterministas y estocásticas, es decir, se observa cómo se cancelan o se amplifican entre sí. Para empezar, se incorporará el ruido únicamente en el término de

transmisión. Sea  $\xi(t)$  una serie de tiempo de desviaciones aleatorias que provienen de una distribución normal con  $\mu = 0$  y  $\sigma^2 = 1$ . Si se supone que la transmisión depende de la frecuencia, las ecuaciones básicas se transforman en:

$$\begin{aligned}\frac{dX}{dt} &= \nu N - \left[ \beta \frac{XY}{N} + f(X, Y)\xi \right] - \mu X, \\ \frac{dY}{dt} &= \left[ \beta \frac{XY}{N} + f(X, Y)\xi \right] - \gamma Y - \mu Y, \\ \frac{dZ}{dt} &= \gamma Y - \mu Z.\end{aligned}\tag{A.8}$$

Para hacer el modelo más general, se incluye una función  $f(X, Y)$  para escalar la aleatoriedad en respuesta al tamaño actual de las variables. Si  $f$  es constante, por ejemplo, se tiene **ruido aditivo simple**. Si  $f = \sqrt{\frac{\beta XY}{N}}$ , se tiene **ruido aditivo calibrado**.

### A.2.3. Parámetros aleatorios

Una técnica de modelado comparada con la de introducir términos de ruido es la de considerar que éste es inherente a los parámetros. Esta alternativa, aunque es matemáticamente equivalente a la anterior, considera una hipótesis subyacente muy diferente. Los modelos previos suponen que los niveles de la población presentan fluctuaciones aleatorias, mientras que estos modelos consideran que los parámetros varían. Esto puede deberse a dos razones. Primero, a que puede existir variabilidad en los parámetros debido a una fuerza externa e impredecible. Por ejemplo, la tasa de transmisión puede verse afectada por las condiciones climáticas; muchos virus sobreviven mejor en condiciones de frío y humedad con poca luz ultravioleta. Por lo tanto, los cambios en la temperatura o en la humedad pueden traducirse en fluctuaciones en  $\beta$ . Esta variabilidad es independiente de la prevalencia de la enfermedad o del número de susceptibles.

Una segunda fuente de fluctuaciones de los parámetros se debe a la variación individual, tanto en términos de los contactos, como en los de la respuesta inmunológica a la infección, lo que resulta en tasas de transmisión variables. En este caso, las fluctuaciones estarán determinadas por los individuos que se encuentren infectados en ese momento. Se espera entonces que esta forma de estocasticidad disminuya cuando haya más individuos infecta-

dos, porque la variabilidad entre ellos llegará a ser la misma. Tomando solo en cuenta la variabilidad en la tasa de transmisión, las ecuaciones básicas del modelo SIR son:

$$\begin{aligned}\frac{dX}{dt} &= \nu N - \beta(1 + F\xi)\frac{XY}{N} - \mu X, \\ \frac{dY}{dt} &= \beta(1 + F\xi)\frac{XY}{N} - \gamma Y - \mu Y,\end{aligned}\tag{A.9}$$

Cuando la variabilidad se debe a una fuerza externa,  $F$  es constante en general. Sin embargo, cuando el ruido viene de la heterogeneidad de la población en la transmisión, se espera que  $F$  sea proporcional a  $\sqrt{Y}$ ; esto se debe a que la variabilidad en la tasa de transmisión promedio se reduce cuando muchos individuos se infectan. Si la estocasticidad se debe únicamente a la heterogeneidad de la susceptibilidad, entonces  $F$  es proporcional a  $\sqrt{X}$ . La constante de proporcionalidad en ambos casos se determina respectivamente por la magnitud de las fluctuaciones externas o por la variabilidad de la población.

#### A.2.4. Comparación entre los distintos tipos de ruido

Tomando en cuenta los ejemplos para los dos tipos de ruido descritos anteriormente, se puede considerar cómo cambian los términos de ruido  $f$  cuando el tamaño de población  $N$  cambia, ya que en general  $X$  y  $Y$  son proporcionales a  $N$ . En el caso de ruido constante el tamaño de la población no influye; en el caso del ruido aditivo calibrado y en el de que exista un parámetro de ruido externo,  $f$  se ajusta a la raíz cuadrada del tamaño de la población. Finalmente, en el caso de ruido del parámetro relativo a la heterogeneidad,  $f$  es proporcional al tamaño de la población. Esto significa que cuando el tamaño de la población es grande, el término de error dominante se debe a fluctuaciones externas en los parámetros. Además, cuando hay mucha variabilidad en la respuesta individual a la infección, el parámetro de ruido debido a este factor puede sobrepasar con facilidad al ruido calibrado debido a la variabilidad dinámica. Sin embargo, a pesar de la importancia potencial del ruido de los parámetros, éste en general ha recibido poca atención en la literatura epidemiológica.

En conclusión, el ruido puede ser generado por diversas fuentes. La magnitud relativa de estos términos de ruido depende del tamaño de la población. El ruido de los parámetros externos domina cuando la población es grande.



### A.2.5. Ventajas y desventajas entre los diferentes tipos de modelos

Incorporar ruido en las ecuaciones diferenciales estándar tiene diferentes ventajas. Por un lado, la modificación de las ecuaciones básicas es relativamente directa; por lo tanto, se pueden usar las mismas técnicas que se usan para iterar las ecuaciones y determinar la dinámica. Por otro lado, la correspondencia entre los modelos deterministas y las ecuaciones estocásticas es evidente, ya que mientras el término de ruido se reduzca a cero, se recupera la dinámica determinista. Finalmente, el costo computacional relativo a esta forma de modelo estocástico, es pequeño si se compara con el estudio que tendría que hacerse si se considerara la aleatoriedad de la población. En particular, los tiempos de simulación son independientes del tamaño de la población. Sin embargo, estas ecuaciones estocásticas tienen una gran desventaja: no incorporan la naturaleza discreta e individual de las poblaciones, y por lo tanto, no son un modelo apropiado cuando los niveles de la población son pequeños. Las técnicas de ruido calibrado o de los parámetros aleatorios cuando existe heterogeneidad en la población, quedan atrapadas entre dos elementos opuestos. Esos modelos difieren mucho de sus contrapartes deterministas solo cuando los niveles de la población son grandes. Por lo tanto, aunque estos modelos basados en el ruido permiten la extinción del patógeno, la exactitud que otorgan cuando se trabaja con procesos que se basan en el comportamiento individual son cuestionables.

En resumen, estas modificaciones basadas en el ruido hechas a las ecuaciones diferenciales estándar, son más apropiadas cuando se consideran los efectos de variaciones de parámetros aplicados a poblaciones grandes debido a fuentes externas. También pueden ayudar a entender mejor los efectos de la estocasticidad en general. Sin embargo, cuando se trabaja con un número pequeño de individuos, se necesita usar una estrategia diferente.



## Apéndice B

### Aproximación normal bayesiana

Esta aproximación se basa en resultados obtenidos cuando el tamaño de nuestra  $n$  crece. Considérese un vector de parámetros  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_d)^T \in \mathbb{R}^n$  con distribución final  $\pi$ . Si se toma la expansión en serie de Taylor del logaritmo de la distribución final alrededor de la moda  $\mathbf{m}$  -que en este caso se considera única-, se tiene

$$\begin{aligned}\log \pi(\boldsymbol{\theta}) &= \log \pi(\mathbf{m}) + \left[ \frac{\partial \log \pi(\mathbf{m})}{\partial \boldsymbol{\theta}} \right]^T (\boldsymbol{\theta} - \mathbf{m}) \\ &\quad - \frac{1}{2} (\boldsymbol{\theta} - \mathbf{m})^T \left[ -\frac{\partial^2 \log \pi(\mathbf{m})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right] (\boldsymbol{\theta} - \mathbf{m}) + R(\boldsymbol{\theta}), \\ &\doteq \log \pi(\mathbf{m}) - \frac{1}{2} (\boldsymbol{\theta} - \mathbf{m})^T \left[ \frac{\partial^2 \log \pi(\mathbf{m})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right] (\boldsymbol{\theta} - \mathbf{m}),\end{aligned}$$

donde el residuo  $R(\boldsymbol{\theta})$  está formado por términos de la forma  $(\boldsymbol{\theta} - \mathbf{m})$  de orden mayor o igual a tres, tales que los en la aproximación se desprecian. En general, la distribución a posteriori se conoce salvo cierta constante de proporcionalidad; es decir, solo se conoce  $\pi^*(\boldsymbol{\theta})$ , donde si  $l(\boldsymbol{\theta})$  representa la verosimilitud, la primera se define como  $\pi^*(\boldsymbol{\theta}) = l(\boldsymbol{\theta})p(\boldsymbol{\theta}) = k\pi(\boldsymbol{\theta})$ , donde  $k = \int \pi^*(\boldsymbol{\theta})d\boldsymbol{\theta}$ . De igual manera, la expansión de  $\pi^*(\boldsymbol{\theta})$  alrededor de  $\log \mathbf{m}$

es

$$\begin{aligned}\pi^*(\boldsymbol{\theta}) &\doteq \pi^*(\mathbf{m}) \exp \left\{ -\frac{1}{2}(\boldsymbol{\theta} - \mathbf{m})^T \left[ \frac{\partial^2 \log \pi^*(\mathbf{m})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right] (\boldsymbol{\theta} - \mathbf{m}) \right\} \\ &= \pi^*(\mathbf{m}) \exp \left\{ -\frac{1}{2}(\boldsymbol{\theta} - \mathbf{m})^T V^{-1}(\boldsymbol{\theta} - \mathbf{m}) \right\}\end{aligned}$$

y por lo tanto

$$k = \pi^*(\mathbf{m})(2\pi)^{\frac{d}{2}}|V|^{\frac{1}{2}}$$

donde

$$V = \left[ -\frac{\partial^2 \log \pi(\mathbf{m})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right]^{-1} = \left[ -\frac{\partial^2 \log \pi^*(\mathbf{m})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right]^{-1}$$

es el negativo de la inversa de la matriz hessiana de  $\log \pi$  y  $\log \pi^*$  evaluadas en la moda. Por lo tanto, a posteriori,  $\boldsymbol{\theta} \sim N(\mathbf{m}, V)$ .

# Bibliografía

- [1] C. Bailer-Jones. (2013). Notas del curso *Statistical Methods*. Capítulo 8: *Bayesian modelling using Monte Carlo methods for sampling and integration*.
- [2] K. Barrios. (2016). *Biodistribución de una formulación de una formulación de cisplatino liposomal radiomarcado con Tc-99m en un xenotrasplante de cáncer*. Tesis de licenciatura, UNAM.
- [3] S. Chib & E. Greenberg. (1995). *Understanding the Metropolis-Hastings Algorithm*. *The American Statistician*, Vol. 49, No. 4, pp. 327-335.
- [4] J. A. DiMasi & R. W. Hansen & H. G. Grabowski. (2003). *The Price of Innovation: New Estimates of Drug Development Costs*. Health Economics, 22.
- [5] D. Gamerman & H. Freitas. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. 2nd. Edition. Chapman & Hall/CRC.
- [6] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, D. B. Rubin. (2014). *Bayesian Data Analysis*. Third Edition. Chapman & Hall/CRC.
- [7] J. V. S. Gobburu & P. J. Marroum. *Utilisation of Pharmacokinetic-Pharmacodynamic Modelling and Simulation in Regulatory Decision-Making*. *Clinical Pharmacokinetics*, 40, 2001.
- [8] W. K. Hastings. (1970). *Monte Carlo Sampling Methods Using Markov Chains and Their Applications*. *Biometrika*, Vol. 57, No. 1. Abril, pp. 97-109.

- [9] M. J. Keeling & P. Rohani. (2008). *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press.
- [10] G. Koch. (2012). *Modeling of Pharmacokinetics and Pharmacodynamics with Application to Cancer and Arthritis*. Tesis de doctorado, Universität Konstanz, Mayo.
- [11] M. Lavielle. (2015). *Mixed Effects Models for the Population Approach Models, Tasks, Methods and Tools*. Chapman & Hall/CRC Biostatistics Series.
- [12] A. D. Martin & K. M. Quinn & J. H. Park. (2011). *MCMCpack: Markov Chain Monte Carlo in R*. Journal of Statistical Software, Junio, volumen 42, número 9.
- [13] C. McRae. (2014). *Bayesian inference in nonlinear differential equation models*. Australian Mathematical Sciences Institute.
- [14] Merriam-Webster. (2016). *English Online Dictionary*. <http://www.merriam-webster.com>
- [15] F. Mittelbach & M. Goossens. (2004). *The L<sup>A</sup>T<sub>E</sub>X Companion*. Addison Wesley.
- [16] L. Mondragón. (2014). *Evaluación de la eficacia de una formulación liposomal de cisplatino en xenotransplantes de cáncer cérvicouterino*. Tesis de Licenciatura, UNAM.
- [17] J. A. Montoya & G. Figueroa & G. Dávila. (2013). *Estimación de parámetros en sistemas de ecuaciones diferenciales ordinarias en modelos farmacocinéticos de dos compartimentos*. Epistemus, 15: 13-19.
- [18] A. M. Mood & F. A. Graybill & D. C. Boes. (1974). *Introduction to the Theory of Statistics*. McGraw Hill International Editions.
- [19] J. C. Nash. (2014). *On Best Practice Optimization Methods in R*. Journal of Statistical Software, Septiembre, volumen 60, número 2.
- [20] T. Oetiker. (2008). *The not so Short Introduction to L<sup>A</sup>T<sub>E</sub>X 2<sub>ε</sub>*. Version 4.26.

- [21] R Development Core Team (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, <http://www.R-project.org>
- [22] C. P. Robert & G. Casella. (2010). *Introducing Monte Carlo Methods with R*. Springer.
- [23] M. Steyvers. (2011). Notas del curso: *Computational Statistics with Matlab*.
- [24] T. Tantau. (2008). *The Tikz and PGF Packages (Manual for version 2.00)*. 2008.
- [25] WordReference. (2016). *English-Spanish Dictionary*. <http://www.wordreference.com>





# Índice alfabético

- absorción, 1
  - de orden  $\alpha$ , 6
  - de orden cero, 5
  - de primer orden, 6, 28
- actualización
  - por bloque, 20
- ADME, 1
- administración
  - intravenosa, 3
  - intravenosa por bolo, 7
  - oral, 5, 6, 28
  - por infusión, 7
- algoritmo
  - de aceptación rechazo, 10
  - de Metropolis, 15
  - de Metropolis Hastings, 19
  - de Metropolis por bloque, 34
- biodisponibilidad, 1, 5, 34
- cadena de Markov, 14
  - estado estacionario, 14
  - estado inicial, 14
  - periodo de calentamiento, 14
- capacidad de eliminación máxima, 4
- cisplatino, 36
- compartimento
  - central, 2
  - de depósito, 2
  - periférico, 2, 5
- distribución, 1
  - a posteriori*, 24
  - a priori*, 24
  - de comparación, 11
  - de propuesta, 11, 31
  - final, 24
  - gamma, 32
  - inicial, 24
  - no informativa, 32
  - normal multivariada, 34
  - objetivo, 11, 15, 24, 25
  - predictiva final, 36, 37, 45, 48
- dosis, 6
- eliminación
  - de primer orden, 4
  - de segundo orden, 44
  - lineal, 3, 6, 28
  - saturada, 4
- eliminación lineal, 5
- excreción, 2
- infecciosos, 52
- integración Monte Carlo, 13
- liposoma, 37

- método
  - de máxima verosimilitud, 23
- matriz de varianza-covarianza, 34
- MCMC, 12, 15
- metabolismo, 1
  - de primer paso, 5
- metabolitos, 1
- Michaelis-Menten, 4, 5
- moda, 34
- modelo
  - SIR, 52
  - de absorción, 6
  - de distribución, 6
  - de dos compartimentos periféricos, 5
  - de eliminación, 6
  - farmacocinético, 3
  - PK, 3, 30, 32
  - por compartimentos, 2
- Monte Carlo vía cadenas de Markov, 12, 15
- muestreo
  - a posteriori, 25, 31
  - de Gibbs, 21
  - por rechazo, 10
- periodo
  - de calentamiento, 10, 34, 42, 46
  - de vida medio, 54
- proceso
  - de absorción, 5
  - de eliminación, 2, 3
  - de Markov, 14
- profármaco, 1
- quimioterapia, 37
- remoción, 4
- retardo, 16
- ruido
  - aditivo calibrado, 57
  - aditivo simple, 30
- ruido aditivo
  - calibrado, 57
- semivida, 4
  - de absorción, 6
  - de eliminación, 4
- susceptibles, 52
- tasa
  - de decremento de susceptibles, 53
  - de distribución, 3
  - de distribución constante, 5
  - de eliminación, 3
  - de eliminación constante, 4
  - de entrada, 2
  - de incremento de infecciosos, 52
  - de mortalidad, 54
  - de remoción, 53
- tecnecio, 43
- teofilina, 27
- teorema
  - de aproximación normal, 34
- tumor cancerígeno, 37
- verosimilitud, 31
- volumen de distribución, 4